# Location estimation at variance-gamma logarithmic poles

Scott Schubr
Luminary Oddities

November 2026

**Abstract**

We determine the sharp behaviour of location estimation for variance-gamma laws at the critical shape, where the density has a logarithmic pole, and settle the rate and limit-distribution questions for the leave-one-out estimator that Nitithumbundit and Chan studied by simulation. In one dimension, the global adjacent-midpoint leave-one-out estimator of Nitithumbundit and Chan converges at rate $(\log n)/n$ to a drifted Brownian argmax, with convergence of all fixed absolute moments; with $I = \pi/\sigma$ its limiting normalized absolute and squared risks are $3/I$ and $26/I^2$. The sharp minimax squared risk at this scale is $16\zeta(3)/I^2$: the flat-prior posterior mean attains it exactly at every sample size, and averaging the same leave-one-out likelihood instead of maximizing it attains it asymptotically, lowering the risk of the maximizer by the factor $8\zeta(3)/13$. In dimension at least two, shape below one gives a sharp Poisson shift rate, shape above one gives ordinary Gaussian local asymptotics, and shape one gives an exact critical Gaussian experiment at rate $(n \log\log n)^{-1/2}$ in the plane and $(n \log n)^{-1/2}$ in higher dimensions; a truncated score attains the critical minimax squared-risk constants. When shape, skew and scatter are unknown, the joint local experiment at shape one splits into independent critical location and ordinary nuisance Gaussian shifts, and an explicit pilot followed by a centered truncated score preserves the exact critical constant uniformly on compact brackets. Finally we resolve the crossover as the shape tends to one with the sample size: in dimension at least three a shape deficit of order $1/\log n$ multiplies the critical information by $(1 - e^{-\tau})/\tau$, while in the plane the transition from below occurs when the deficit times $\log n$ reaches $\log\log\log n$; one estimator attains the sharp constant in every regime without knowing which applies.

## Contents

**1 Introduction** **2**

**2 The leave-one-out location estimator at the variance-gamma logarithmic pole** **3**
2.1 Estimator and results . . . 3
2.2 Density estimates and a coarse bound . . . 4
2.3 Fine approximation and increment bounds . . . 4
2.4 Process convergence and maximization . . . 5
2.5 The limiting absolute and squared risks . . . 7

**3 Sharp location rates for multivariate variance-gamma laws** **7**
3.1 Model and phase diagram . . . 8
3.2 Density and weighted differences . . . 8
3.3 The subcritical Poisson phase . . . 9
3.4 Critical experiments . . . 10

3.5 Attaining the critical information . . . 11
3.6 Compact parameter sets and the minimax constant . . . 12

**4 Critical variance-gamma location with unknown shape, skew and scatter 12**
4.1 Model and results . . . 12
4.2 Density bounds near the pole . . . 13
4.3 A moment pilot for the full parameter . . . 14
4.4 The joint local experiment . . . 15
4.5 Estimating location with unknown nuisance . . . 16

**5 The sharp minimax risk at the variance-gamma logarithmic pole 17**
5.1 Model and results . . . 17
5.2 Capped comparison . . . 18
5.3 Fine likelihood integrals and moments . . . 20
5.4 Minimax identity and constant . . . 21

**6 Location inference as variance-gamma shape approaches one 21**
6.1 Results . . . 22
6.2 Annular information . . . 23
6.3 The local experiments . . . 24
6.4 An attaining estimator . . . 26

# 1 Introduction

Variance-gamma laws have a logarithmic density singularity at their centre when the shape parameter equals one, and a power singularity below one. Location estimation at such a singularity falls between the Poisson regime of a jump and the regular Gaussian regime. Podgórski and Wallin [4] introduced maximum leave-one-out likelihood for unbounded densities and proved consistency, super-efficiency and a lower bound on the rate for power poles. Nitithumbundit and Chan [6, 5] recorded that the optimal rate and the asymptotic distribution of this estimator had no theoretical treatment and studied both by simulation, including the critical variance-gamma case. Section 2 settles both questions at the logarithmic pole: it gives the exact one-dimensional limit of the leave-one-out estimator at the logarithmic pole. Section 3 determines the sharp rates and minimax constants in every dimension at least two. Section 4 shows that the critical constant is unaffected when shape, skew and scatter must also be estimated. The Brownian argmax constants are classical; see [3].

Section 5 identifies the sharp minimax constant in one dimension. 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 [2] for a discontinuous signal in white noise, and the posterior mean is exactly minimax at each sample size by Pitman's construction [1]. Since the leave-one-out maximizer has risk $26/I^2$, averaging the leave-one-out likelihood improves it by the factor $8\zeta(3)/13 \approx 0.740$. Section 6 lets the shape move with the sample size and derives the exact information in each window around the critical shape, connecting the Poisson, critical and regular regimes of Section 3.

This is the second edition of this article. The second edition adds Sections 5 and 6.

# 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 $X_i = \theta_0 + Z_i$, $i = 1, \dots, n$, 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 $\theta_0$, 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 [7], this is the model with dimension one, $\nu = 1/2$, $\gamma = 0$ and known $\Sigma = \sigma^2$. Their Section 5.3 maximizes over the adjacent midpoints with the other parameters fixed. Their simulated power-law fit does not determine the logarithmic factor. The theorem also supplies the limiting distribution they examined numerically. Podgórski and Wallin [8] 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 [11] 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

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

These and the exponential tails follow from the Bessel representations and expansions in [10]. The logarithmic derivative is bounded on $|x| \ge x_0$. Also $|\log f(x)| \le C(1 + |x|)$ 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 $r$,

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

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 $C_{n,r_g}(t) \ge -C_1 L$.

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 $x = tz$, $2 \le z \le 3$. 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 $H(f, g_{r_g}) \le C\sqrt{r_g}$. In particular for $|t| \ge W/2$, $nH^2 \ge cBL$ for sufficiently large $n$. Markov's inequality on a deterministic mesh of width $n^{-(D+3)}$ in the data range rules out $C_{n,r_g} \ge -C_1 L - 1$ 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

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

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

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

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 $\pi^2$ 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 $2\delta$, including enough overlap so any pair at distance at most $\delta$ 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 $\delta, c$ the term with $\epsilon_n$ tends to zero. Sending $\delta$ 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 $\sigma$, 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 Terentyev, as recorded in [9]. 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

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

This follows from the exponential Brownian martingale stopped at $\tau_m$ 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 $\tau_m$ gives

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

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 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.

### 3.1 Model and phase diagram

Observe $n$ independent copies of $X = \theta + Z$ for $\theta \in \mathbb{R}^d$, $d \geq 2$, 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 $\omega_{d-1}$ for the surface area of the unit sphere in $\mathbb{R}^d$.

**Theorem 3.1.** *The minimax expected Euclidean error and root mean squared Euclidean error for $\theta$ both have order $d_n$, where*

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

*For fixed $\theta_0$ 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 [7] with positive scatter $\Sigma$, the linear map from the normalization above is $(\Sigma/(2\nu))^{1/2}$. Thus at criticality the information in their coordinates is $\Sigma^{-1}$ in dimension two and $(d-2)\Sigma^{-1}/d$ in higher dimensions, at the corresponding rates. Their one-dimensional numerical study addresses a different local regime, with the logarithmic pole at shape one-half. Its sharp location rate and the actual leave-one-out limit are given in [11] and Section 2. 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 [12] in one-dimensional location and the Poisson models of Dachian [13] are related. Gupta, Lee and Price [14] 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.

### 3.2 Density and weighted differences

Let $G$ be gamma with shape $\nu$ 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 $|f''(r)| \le Cr^{-2}$. 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 [10]. 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 $0 < \nu \le 1$ and $\nu < d/2$, the part $|x| \le C_0 D$ 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 [10], 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 [15]. 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 3.1. We prove the remaining upper bounds and experiments below.

### 3.3 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 $\theta$. 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 $|z - \theta| \ge Bh$, 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 $O(\sum_{h \ge h_f} h^2 \exp(-cnh^{2\nu})) = O(h_f^2)$. 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.

### 3.4 Critical experiments

Take $\nu = 1$, and put $\Lambda_n = \log\log n$ if $d = 2$ and $\Lambda_n = \log n$ if $d \ge 3$. 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 $\Delta \le |x| \le r_0$, with a single censoring symbol for all other values. We first show that this retains the local experiments. If $d \ge 3$, 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

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

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 $d \ge 3$,

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

The Gamma relation for $\omega_{d-1}$ 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.

### 3.5 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 $J_\Delta$. 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 $J_\Delta$. 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 $C/\Delta$ 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 $\theta$.

### 3.6 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 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 $\tau$. 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 5.

## 4 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.

### 4.1 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 $\theta$ 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 $(\theta_0, \eta_0)$ with $\nu_0 = 1$, and put

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

Write $f_\eta$ 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 4.1.** *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 4.2.** *Fix compact brackets for $\theta, \gamma, S$ with the eigenvalues of $S$ bounded away from zero. For each fixed $H < \infty$, there is an estimator using unknown shape within $|\nu - 1| \leq H/\sqrt{n}$ and unknown $\gamma, S$ 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 [16, 7] study this normal mixture parameterization and construct likelihood algorithms for the skew multivariate model. Their simulations of the asymptotic location law and rate in [7, Section 5.3] use the univariate symmetric model with the other parameters fixed. Methods of moments for variance-gamma models are classical. Seneta [17] uses moment fitting in the skew univariate model. Fischer, Gaunt and Sarantsev [18] 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 3.

## 4.2 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

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

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

$$|\nabla_x \log f_\eta| \leq C/|x|, \qquad \frac{\|D_x^2 f_\eta\|}{f_\eta} \leq C/|x|^2, \tag{5}$$

$$|s_\eta| \leq C(1 + \log(1/|x|)), \qquad \|D_\eta \ell_\eta\| \leq C(1 + \log(1/|x|))/|x|. \tag{6}$$

To check these bounds, inserting $t^{-j}$ in the integral for $B$ costs at most $C_j r^{-2j}$. For $t \geq r^2$ this follows directly. For $t < r^2$ put $z = r^2/(4t)$ and compare the resulting bound to the part of $B$ from $r^2 \leq t \leq 2r^2$. A further $|\log t|$ costs at most $C(1 + \log(1/r))$ in the same bounds. For $r^2 \leq t \leq 1$ 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 $\nu$ 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 (4), including its smooth prefactors, proves (5) and (6). 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 (4) 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 $d \geq 3$, 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 $A_\Delta = \{x : \Delta \leq |x| \leq r_0\}$ and $\Delta = n^{-1/2}(\log n)^2$, at $\nu = 1$

$$\Lambda_n^{-1} J_\Delta(\eta) \;\longrightarrow\; \mathcal{K}(S), \qquad J_\Delta(\eta) = \int_{A_\Delta} \ell_\eta \ell_\eta^{\mathsf{T}} f_\eta. \tag{7}$$

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.

### 4.3 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 $\theta_j$ factors as

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

To determine $\nu$ 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 $\pm 1$. Thus the inverse recovers $w, \rho$, then $\nu = c_2/\rho^2$, with uniform Lipschitz constants.

Given $\nu$, 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 $\pm 1$. This recovers $\gamma$ 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 $\nu$ 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 $\Delta/C$ 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 $\theta$ 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 [19]. This is $s_\eta$ from (6). 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 $\mathcal{J}$.

### 4.4 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 $\eta_0$ 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 $\Delta$ 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 (4). 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 $2\Delta$ 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 (6) 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 (5), $|k| \le C\delta_n/\Delta = o(1)$. On the annulus the density ratios $f_h/f_0$ tend uniformly to one by (6). Taylor expansion, (5) and the critical density bounds give

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

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, (6) 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 (6), the bound $|k(x)| \le C\delta_n/|x|$ on $A$ and the integrability following (7). 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 (7) 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 [19]. 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 (6) 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 $\Delta$ 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 4.1.

### 4.5 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-$\Delta$ 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 $m_\Delta$ 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 (5), 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 (6) 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 (7), 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 $\Lambda_n$ 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 $J_\Delta$ on the good event, so the normalized variance converges also in expectation. On the bad event the score is still deterministically bounded by $C/\Delta$, 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 4.1. 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 $\mathrm{tr}(\mathcal{K}(S_0)^{-1})$. The estimator therefore attains the sharp local constant also when the nuisance is unknown.

# 5 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.

## 5.1 Model and results

Observe independent $X_i = \theta + Z_i$, $i = 1, \dots, n$, $n \ge 2$, 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 studied by Nitithumbundit and Chan [7, Section 5.3] 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 5.1.** *Under any $\theta$, 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 $(\Pi, \Pi)$ 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 [20] gives the best equivariant location estimator by averaging the likelihood. Strasser [21] 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 [22, 9] in the change-point setting. The Brownian argmax has second moment 26, as recorded in [9]. Section 2 proves that the unmodified variance-gamma leave-one-out maximizer has that canonical argmax limit and its associated risk. Here the new step is global weighted integral control for the original likelihood despite its infinite spikes, and for the leave-one-out likelihood despite the deletion changing with $t$. This proves the canonical risk transfer to the actual variance-gamma likelihood averages.

## 5.2 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

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

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,

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

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

$$\int_{|x|\leq r} \{f(x) - f(r)\}\,dx \leq Cr. \tag{8}$$

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 $p \geq 1$. 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 $B$,

$$\sup_{BL^2/n \leq |t| \leq (C_0+1)L} C_{r_g}(t) \leq -c_0 BL \tag{9}$$

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 $-c_0 BL$ 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 $|t| \leq (C_0 + 1)L$ this is $O(L)$ with a constant independent of $B$. Thus (9) 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 (8). 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 $(\log K_0)'(x) \leq -1$ for $x > 0$ by its integral as a mixture of $e^{-xu}$, $u \geq 1$. Along each outer ray the omitted index is fixed. Combining this fact with (9), the integrated spike bound, and $M/d_n = O(n)$, gives for either likelihood ratio $Z_n^*(u)$, true or leave-one-out,

$$\int_{|u|>BL} (1 + |u|^p) Z_n^*(u)\, du \leq \exp\{-[c_1 B - (p+2)]L\} \tag{10}$$

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 (10) and sufficiently large $B$, this makes both scaled likelihood means at most $BL + 1$ in absolute value on the coarse good event.

### 5.3 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 (10) 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 $L^q$.

### 5.4 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 $\mu_w$ 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 $16\zeta(3)$ [22, 9]. 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 5.1.

## 6 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.

## 6.1 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 6.1.** *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 $\eta$, independent of the choice among these regimes, with*

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

*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 [7] 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 3, 4. The phenomenon here is caused by the reference shape moving with $n$. Bakirov [23] 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.

### 6.2 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 $\nu = 1$,

$$|\ell_\eta| \le C/|x|, \quad \|D_x^2 f_\eta\|/f_\eta \le C/|x|^2, \quad |s_\eta| \le C(1 + \log(1/|x|)), \quad \|D_\eta \ell_\eta\| \le C(1 + \log(1/|x|))/|x|. \tag{11}$$

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 4, give (11). 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 $\nu$ 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 (11) 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 $\Delta, r_0$. 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 $\nu$ 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

$$B(a, r) = (w - 1)/a + O(1 + w), \tag{12}$$

$$-r\partial_r B(a, r) = 2\, 4^a \Gamma(1 + a) w[1 + O(r^\kappa)] \tag{13}$$

for some fixed $\kappa > 0$ and all sufficiently small $|a|$. At $a = 0$ the quotient is $2\log(1/r)$. To prove (12), 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 (13). 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 (12) 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 $a \neq 0$ 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 $\pi$ 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 $a_n \ge 0$ and the ratio defining $\lambda$ 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 $a_n \le 0$, 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 $\Lambda_n$ 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

$$J_n(\eta_n)/\Lambda_n \longrightarrow \mathcal{K}. \tag{14}$$

The same computations apply after any bounded root-$n$ nuisance perturbation. In the last case, if $\nu_n$ 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.

### 6.3 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 $\eta_n$, uniformly for bounded $u$, with $D = d_n$. For $d \ge 3$, the inner ball $|x| \le C_0 D$ costs $O(D^{2\nu_n}) = O(D^2)$ in squared Hellinger distance, by the two local masses. From $C_0 D$ to $\Delta$ 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 $|x| \le C_0 D$ both untilted radial density factors are bounded below by a positive constant times $b_D$, using (12) and monotonicity. Set $x = Dz$ and let $\rho_1, \rho_2$ be the whitened distances of $z$ to the two shifted pole centers. By (13), the radial difference is bounded by

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

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

In the planar annulus $C_0 D \le |x| \le \Delta$, 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 (12) and (13). 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 $\eta_n$ 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 $3\Delta$ 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 (11) 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 (12), 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 4, Taylor expansion and (11) 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 (14), 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 $\eta_n$. One way to verify the needed uniform quadratic mean expansion is to use the gamma-normal complete-data density. Its derivatives in $\eta$ 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 [19] and Section 4. 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 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.

### 6.4 An attaining estimator

Use the cumulant pilot of Section 4 on one half of the data, recovering $\theta, \nu, \gamma, S$ 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 $\Lambda_n$ 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, (11) 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 $CM_n|e|/\Delta$ and at the outer radius by $CM_n|e|$. 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 $C/\Delta$. 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 (14) 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 3. Their Bayes risks tend to $\mathrm{tr}(\mathcal{K}^{-1})$. This proves Theorem 6.1.

## Statements and Declarations

**Funding.** This research was self-funded.

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

**AI assistance.** OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

## References

[1] 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. https://doi.org/10.1093/biomet/30.3-4.391.

[2] 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. https://doi.org/10.1214/aos/1176324618.

[3] S. Dachian, On limiting likelihood ratio processes of some change-point type statistical models, *J. Statist. Plann. Inference* 140 (2010), 2682–2692. https://arxiv.org/abs/0907.0440.

[4] K. Podgórski and J. Wallin, Maximizing leave-one-out likelihood for the location parameter of unbounded densities, *Ann. Inst. Stat. Math.* (2015). https://doi.org/10.1007/s10463-013-0437-6.

[5] 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.* (2023). https://doi.org/10.1080/00949655.2023.2202399.

[6] 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.

[7] 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.

[8] 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.

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

[10] NIST Digital Library of Mathematical Functions, Chapter 10, Section 10.31, Section 10.32, and Section 10.40.

[11] S. Schubr, *Calibrating small noise from exteriors, frontiers, cusps and density singularities*, the part titled *Calibrating a noise endpoint from a logarithmic density singularity*, Luminary Research Archive, 2026.

[12] 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.

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

[14] 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.

[15] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (1998), Theorem 7.2.

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

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

[18] 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.

[19] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (1998), Chapter 7.

[20] 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.

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

[22] 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.

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