

Nonadditivity of the Mean Symmetric-Difference Risk for Kernel Level Sets

Alper Ferudun*

27 October 2026

Abstract

The mean symmetric-difference risk of kernel density level sets need not admit an additive variance–bias asymptotic with coefficients independent of the bandwidth. For the standard bivariate normal density, a compactly supported radial polynomial kernel, and bandwidths $h_n(t) \asymp n^{-1/6}$, the risk divided by $h_n(t)^2$ converges to a positive constant times $\mathbb{E}|1 + tZ|$, where Z is standard normal. Strict convexity in t rules out simultaneous universal limits $\sqrt{nh^2} I_1 \rightarrow A$ and $h^{-2} I_2 \rightarrow B$ for any exact decomposition of the risk as $I_1 + I_2$. Thus the unrestricted bias-limit proposal in the 2016 Oberwolfach discussion cannot hold together with its stated variance-limit premise. We also calculate two distinct limits for the exact decomposition obtained by centering the density estimator. The proof uses a pointwise triangular-array central limit theorem and an integrable Chebyshev bound. The nonlinear risk law is consistent with, and already encompassed by, the general asymptotic results of Doss–Weng and Qiao; the present result isolates its obstruction to the two bandwidth-independent constants.

1 The question and the counterexample

For a probability density f , a fixed level $c > 0$, and a kernel density estimator

$$\hat{f}_{n,h}(x) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}^2,$$

let $L = \{f \geq c\}$ and $\hat{L}_{n,h} = \{\hat{f}_{n,h} \geq c\}$. The observations X_i are independent with density f . We consider both Lebesgue and probability-weighted risk:

$$\mathcal{M}_w(n, h) = \mathbb{E} \int_{\mathbb{R}^2} w(x) \mathbf{1}_{\hat{L}_{n,h} \Delta L}(x) dx, \quad w \equiv 1 \quad \text{or} \quad w = f.$$

Chacón’s contribution, joint with Polonik, to Oberwolfach Report 32/2016 [1, p. 1833] describes a decomposition $\mathcal{M}_w = I_1 + I_2$ and a variance asymptotic $\sqrt{nh^d} I_1 \rightarrow A(f, K, c)$. It asks whether $h^{-2} I_2 \rightarrow B(f, K, c)$ without the further undersmoothing condition $nh^{d+4} \rightarrow 0$. The decomposition itself is not given as a formula in that report. We will rule out the two stated universal limits simultaneously, independently of how the decomposition is chosen.

Fix

$$f(x) = \frac{1}{2\pi} e^{-\|x\|^2/2}, \quad c = \frac{e^{-2}}{2\pi}, \quad K(x) = \frac{4}{\pi} (1 - \|x\|^2)^3 \mathbf{1}_{\{\|x\| \leq 1\}}. \quad (1)$$

Then L is the closed disk of radius 2. For $w = 1$ or $w = f$, set

$$w_* = \begin{cases} 1, & w = 1, \\ c, & w = f, \end{cases} \quad C_w = \frac{\pi w_*}{5}, \quad \Lambda = \frac{3200e^2}{7}.$$

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

Write ϕ, Φ for the standard normal density and distribution function, respectively, and define

$$G(t) = \mathbb{E}|1 + tZ| = 2t\phi(1/t) + 2\Phi(1/t) - 1, \quad t > 0. \quad (2)$$

Theorem 1. *For every fixed $t > 0$, put*

$$h_n(t) = \left(\frac{\Lambda}{t^2 n} \right)^{1/6}.$$

For the density and kernel in (1),

$$\lim_{n \rightarrow \infty} \frac{\mathcal{M}_w(n, h_n(t))}{h_n(t)^2} = C_w G(t). \quad (3)$$

This holds for each of the two weights $w = 1, f$.

All these bandwidth sequences decrease to zero and satisfy $nh^2/\log n \rightarrow \infty$ and $nh^4/\log n \rightarrow \infty$. Their value of $nh^6 = \Lambda/t^2$ is a positive constant. The density is smooth with bounded derivatives, the level boundary is compact and regular, and the kernel is a nonnegative, radial, compactly supported C^2 probability density. Thus the example has none of the degeneracies associated with a critical level, a rough density, or an inconsistent bandwidth.

Corollary 2. *There are no constants $A, B \in \mathbb{R}$ and functions $I_1(n, h), I_2(n, h)$ such that $\mathcal{M}_w(n, h) = I_1(n, h) + I_2(n, h)$ and*

$$\sqrt{nh_n(t)^2} I_1(n, h_n(t)) \rightarrow A, \quad \frac{I_2(n, h_n(t))}{h_n(t)^2} \rightarrow B \quad (4)$$

for each of $t = 1, 2, 3$, with the same A, B . Consequently (4) cannot hold for all regular bandwidth sequences without an additional restriction.

Proof. Since $\sqrt{nh_n(t)^6} = \sqrt{\Lambda}/t$, the asserted limits would imply

$$C_w G(t) = \frac{A}{\sqrt{\Lambda}} t + B, \quad t = 1, 2, 3.$$

Direct differentiation of (2) gives

$$G'(t) = 2\phi(1/t), \quad G''(t) = \frac{2\phi(1/t)}{t^3} > 0. \quad (5)$$

It follows that $G(1) - 2G(2) + G(3) > 0$, contradicting affinity. $\square$

The corollary retains the variance-limit premise of the report. It is not a claim about an arbitrary term named I_2 after that premise is discarded. In particular, defining $I_2 = 0$ and $I_1 = \mathcal{M}_w$ would not satisfy the required bandwidth-independent variance limit.

2 An elementary proof of the risk limit

For a function on $\mathbb{R}^2$, integrals below are with respect to Lebesgue measure unless otherwise indicated. Put $f_h = \mathbb{E}\hat{f}_{n,h} = f * K_h$, where $K_h(x) = h^{-2}K(x/h)$. Polar integration and symmetry give

$$\int K = 1, \quad \int xK(x) dx = 0, \quad \int xx^\top K(x) dx = \frac{1}{10}I_2, \quad R_K := \int K^2 = \frac{16}{7\pi}. \quad (6)$$

Here I_2 in the matrix identity denotes the identity matrix, not a component of the risk. For example, the scalar integrals for mass, one coordinate second moment, and R_K are, respectively,

$$8 \int_0^1 r(1-r^2)^3 dr, \quad 4 \int_0^1 r^3(1-r^2)^3 dr, \quad \frac{32}{\pi} \int_0^1 r(1-r^2)^6 dr.$$

Because the fourth derivatives of f are bounded and all odd kernel moments vanish, Taylor expansion gives, uniformly in x ,

$$f_h(x) = f(x) + \frac{h^2}{20} \Delta f(x) + O(h^4). \quad (7)$$

In particular, $\sup_x |f_h(x) - f(x)| = O(h^2)$. Writing $V_{n,h}(x) = \text{Var}(\hat{f}_{n,h}(x))$, independence also gives

$$V_{n,h}(x) = \frac{1}{nh^2} \int K(u)^2 f(x-hu) du - \frac{1}{n} f_h(x)^2, \quad (8)$$

$$\sup_x V_{n,h}(x) \leq \frac{\|f\|_\infty R_K}{nh^2}, \quad \int V_{n,h}(x) dx \leq \frac{R_K}{nh^2}. \quad (9)$$

Fix $t > 0$, and write $h = h_n(t)$ for the rest of the proof. Then $1/(nh^2) = O(h^4)$. Set $b = c/10$. For $u \in \mathbb{R}$, consider the moving point $x_n(u) = (2 + h^2 u, 0)$, defined for all sufficiently large n . At the circle $\|x\| = 2$ we have

$$|\nabla f| = 2c, \quad \Delta f = 2c.$$

Consequently (7) implies

$$\frac{f_h(x_n(u)) - c}{h^2} \longrightarrow b - 2cu. \quad (10)$$

Since $nh^6 = \Lambda/t^2 = cR_K/(b^2 t^2)$, (8) gives

$$\frac{V_{n,h}(x_n(u))}{h^4} \longrightarrow b^2 t^2. \quad (11)$$

To justify the normal limit, express $(\hat{f}_{n,h}(x_n(u)) - f_h(x_n(u)))/h^2$ as a sum of independent centered random variables. Each summand is bounded in absolute value by $2\|K\|_\infty/(nh^4) = O(h^2)$. The total variance has the positive limit (11). Thus the Lindeberg condition holds, and the triangular-array central limit theorem yields

$$\frac{\hat{f}_{n,h}(x_n(u)) - c}{h^2} \xrightarrow{d} b(1 + tZ) - 2cu. \quad (12)$$

No uniform central limit theorem is needed.

We now justify integrating these pointwise limits. The distribution of $\hat{f}_{n,h}(x)$ depends only on $r = \|x\|$, by rotation invariance of both f and K . Let $p_n(r)$ be the probability of misclassifying the point $(r, 0)$ relative to the level c . For $r < 2$ it is $\mathbb{P}(\hat{f}_{n,h}(r, 0) < c)$, and for $r > 2$ it is $\mathbb{P}(\hat{f}_{n,h}(r, 0) \geq c)$. Tonelli's theorem gives

$$\mathcal{M}_w(n, h) = 2\pi \int_0^\infty r w(r) p_n(r) dr, \quad (13)$$

where $w(r) = 1$ or $f(r, 0)$.

First consider points outside the annulus $1 \leq r \leq 3$. There $|f(x) - c|$ has a positive lower bound δ . For sufficiently small h , $f_h(x)$ is on the same side of c as $f(x)$, at distance at least $\delta/2$. By Chebyshev's inequality and (9), their total contribution to the risk is at most

$$\frac{4\|w\|_\infty}{\delta^2} \int_{\mathbb{R}^2} V_{n,h}(x) dx \leq \frac{4\|w\|_\infty R_K}{\delta^2 n h^2} = O(h^4) = o(h^2).$$

This estimate also handles the infinite-area tail when $w = 1$.

On the annulus, the radial derivative of f is negative and bounded away from zero. Choose $m > 0$ such that $|f(r, 0) - c| \geq m|r - 2|$ for $1 \leq r \leq 3$. If $|r - 2| \geq Ch^2$, with C sufficiently large, (7) shows that $f_h(r, 0) - c$ has the same sign as $f(r, 0) - c$ and absolute value at least $m|r - 2|/2$. Another application of Chebyshev's inequality therefore gives, for a constant D depending on the fixed t ,

$$p_n(2 + h^2u) \leq \begin{cases} 1, & |u| \leq C, \\ D/u^2, & |u| > C, \end{cases} \quad 1 \leq 2 + h^2u \leq 3.$$

Extend the annular integrand by zero elsewhere. Since $rw(r)$ is bounded on the annulus, this is an integrable bound on the u -axis, independent of n .

Let $W_t = (1 + tZ)/20$. Equation (12) implies that, for $u \neq 0$,

$$p_n(2 + h^2u) \longrightarrow \begin{cases} \mathbb{P}(W_t < u), & u < 0, \\ \mathbb{P}(W_t \geq u), & u > 0. \end{cases}$$

The limit law is continuous, so strict versus non-strict inequalities do not affect these limits. Changing variables in (13) and using dominated convergence now yields

$$\frac{\mathcal{M}_w(n, h)}{h^2} \longrightarrow 4\pi w_* \left[\int_{-\infty}^0 \mathbb{P}(W_t < u) du + \int_0^{\infty} \mathbb{P}(W_t \geq u) du \right] = 4\pi w_* \mathbb{E}|W_t| = C_w G(t).$$

This proves Theorem 1.

3 An exact centered decomposition

The obstruction is not merely formal. Define the centered field and its level-set risk by

$$\tilde{f}_{n,h}(x) = f(x) + \hat{f}_{n,h}(x) - f_h(x), \quad \tilde{\mathcal{M}}_w(n, h) = \mathbb{E} \int w(x) \mathbf{1}_{\{\tilde{f}_{n,h} \geq c\} \triangle L}(x) dx.$$

This field need not be a nonnegative density; its role is to define an exact mean-centered comparison. Let

$$I_1^\circ = \tilde{\mathcal{M}}_w, \quad I_2^\circ = \mathcal{M}_w - \tilde{\mathcal{M}}_w.$$

We introduce this notation here and do not identify it with an unwritten formula in the Oberwolfach report.

Corollary 3. *Along $h_n(t)$, for every fixed $t > 0$,*

$$\begin{aligned} \frac{I_1^\circ(n, h_n(t))}{h_n(t)^2} &\longrightarrow C_w t \sqrt{2/\pi}, \\ \frac{I_2^\circ(n, h_n(t))}{h_n(t)^2} &\longrightarrow C_w (G(t) - t \sqrt{2/\pi}). \end{aligned}$$

In particular, $\sqrt{nh_n(t)^2} I_1^\circ \rightarrow A_w := C_w \sqrt{\Lambda} \sqrt{2/\pi}$, independently of t , whereas the second limit is strictly decreasing in $t > 0$.

Proof. The proof of Theorem 1 applies to the centered field with mean f and unchanged variance. The mean in (10) is now $-2cu$, and W_t is replaced by $tZ/20$. All localization and tail estimates remain valid. The first limit follows from $\mathbb{E}|Z| = \sqrt{2/\pi}$, and subtraction gives the second. Finally,

$$\frac{d}{dt} (G(t) - t \sqrt{2/\pi}) = 2(\phi(1/t) - \phi(0)) < 0.$$

□

For example, for Lebesgue risk the limits of $h^{-2}I_2^\circ$ at $t = 1, 2, 3$ are approximately

$$0.23169018, \quad 0.12278419, \quad 0.08278914,$$

respectively. These numbers illustrate the strict analytic inequality; they are not used to prove it.

There is also a quantitative obstruction for any proposed constants. For $1 \leq t \leq 3$, (5) gives $G''(t) \geq 2\phi(1)/27$. The second-difference identity therefore implies

$$G(1) - 2G(2) + G(3) = \int_{-1}^1 (1 - |v|)G''(2 + v) dv \geq \frac{2\phi(1)}{27}.$$

For any affine function ℓ , at least one of $j = 1, 2, 3$ satisfies

$$|C_w G(j) - \ell(j)| \geq \frac{C_w \phi(1)}{54}.$$

Thus choosing different numerical values for A, B cannot repair the proposed approximation at all three scales.

4 Relation to existing asymptotic results

Cadre [2] established level-set asymptotics under undersmoothing conditions. The report [1] proposed retaining a bandwidth-independent variance coefficient while adding a bandwidth-independent coefficient of h^2 . Corollary 2 disproves that joint, unrestricted proposal. It does not challenge Cadre's undersmoothing theorem, nor does it rule out an expansion within one prescribed bandwidth regime.

The nonlinear Gaussian expression itself is established prior theory. Doss and Weng [3, Theorem 2.1] give a risk expansion for probability-weighted symmetric difference in dimension at least two. Qiao [4, Theorem 3.2 and Corollary 3.1] gives general weighted-risk expansions involving the absolute value of a normal variable plus the leading bias. Theorem 1 is an explicit special-case calculation of this established phenomenon, proved here directly. Our conclusion concerns its incompatibility with the two universal constants in the stated 2016 proposal. No new general risk expansion or bandwidth-selection method is claimed.

This unrefereed note was prepared with AI assistance in mathematical derivation, source comparison, and manuscript preparation. The written argument was self-audited; it has not been independently refereed or formalized in a proof assistant. The accompanying symbolic checks verify the stated constants and differentiation identities, not the probabilistic limit theorem.

References

- [1] J. E. Chacón (joint work with W. Polonik), Asymptotics and optimal bandwidth selection for level set estimation, in *Statistics for Shape and Geometric Features*, Oberwolfach Report 32/2016, pp. 1832–1833. [doi:10.4171/OWR/2016/32](https://doi.org/10.4171/OWR/2016/32).
- [2] B. Cadre, Kernel estimation of density level sets, *Journal of Multivariate Analysis* **97** (2006), 999–1023. [doi:10.1016/j.jmva.2005.05.004](https://doi.org/10.1016/j.jmva.2005.05.004).
- [3] C. R. Doss and G. Weng, Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions, *Electronic Journal of Statistics* **12** (2018), 4313–4376. [doi:10.1214/18-EJS1501](https://doi.org/10.1214/18-EJS1501); [arXiv:1806.00731v2](https://arxiv.org/abs/1806.00731).
- [4] W. Qiao, Asymptotics and optimal bandwidth for nonparametric estimation of density level sets, [arXiv:1707.09697v3](https://arxiv.org/abs/1707.09697) (2019), Theorem 3.2 and Corollary 3.1.