MARTINGALIZED U-STATISTICS-BASED SEQUENTIAL INFERENCE: A BLESSING OF DEGENERACY

Anonymous authors
Paper under double-blind review

ABSTRACT

Degenerate degree-two U-statistics have non-Gaussian limits, making sequential inference substantially more difficult than in the nondegenerate case. However, row normalization preserves the target parameter and makes the centered numerator a martingale on a harmonic variance clock, at the cost of logarithmic variance inflation. We establish a strong Gaussian approximation for this process and consistently estimate its variance through online updates, allowing spectrum-free calibration. Combining this strong approximation with Gaussian-mixture (GM) and stitched law-of-the-iterated-logarithm (LIL) boundaries yields confidence sequences with asymptotic time-uniform coverage and corresponding sequential tests for null hypotheses under which the kernel is degenerate. We further extend the framework to multi-dimensional vector-valued kernels, yielding Mahalanobis confidence ellipsoids based on consistent covariance estimation. Numerical experiments illustrate sequential type-I error control and favorable detection power in two-sample and independence testing against spectral baselines, and in three-sample homogeneity testing using vector confidence ellipsoids.

1 INTRODUCTION

Many kernel- and distance-based tests are based on degree-two U- or V-statistics that become degenerate under the null hypothesis. Prominent examples include the maximum mean discrepancy (MMD) (Gretton et al., 2012), the Hilbert–Schmidt independence criterion (HSIC) (Gretton et al., 2005), energy statistics, and distance covariance (Székely et al., 2007; Székely & Rizzo, 2013; Sejdinovic et al., 2013). For degree-two U-statistics, the Hoeffding decomposition yields fundamentally different asymptotic behavior according to whether the first-order projection vanishes. In the nondegenerate case, the first-order term gives the usual Gaussian limit, whereas under degeneracy the normalized statistic typically converges to a quadratic Gaussian chaos whose law depends on the spectrum of the associated kernel operator (Hoeffding, 1948; Serfling, 1980; Neuhaus, 1977; Dehling et al., 1984; Arcones & Giné, 1993; 1995; Kanagawa & Yoshihara, 1994). The spectral dependence of the degenerate case complicates fixed-sample calibration and has motivated, among other approaches, permutation procedures and modifications of the original statistic that restore Gaussian limits (Kim & Ramdas, 2024; Shekhar et al., 2022; 2023).

Sequential monitoring, in which inference may be updated after each new observation, requires a stronger form of error control than fixed-sample inference. Rather than controlling the statistic at a single deterministic sample size, one must control the probability that it crosses a prescribed boundary at any monitoring time. Classical sequential analysis studies such boundary-crossing events through mixture constructions and related methods (Robbins, 1970; Robbins & Siegmund, 1970), while modern anytime-valid inference is often based on nonnegative supermartingales, e-processes, mixture boundaries, stitching, and betting (Howard et al., 2020; 2021; Ramdas et al., 2023).

A complementary asymptotic approach derives time-uniform inference from process-level approximations. If a statistic can be coupled to a Gaussian process with an error negligible relative to the relevant boundary scale, Gaussian boundary-crossing results can be transferred to the original process. This principle underlies the time-uniform central limit theory and asymptotic confidence-sequence framework of Waudby-Smith et al. (2024); related sharp asymptotic constructions are developed by Gnettner & Kirch (2025). For ordinary degenerate U-statistics, however, the natural approximating process is a spectrum-dependent quadratic Gaussian chaos rather than a Gaussian process. Cai et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(2026) address this difficulty directly by coupling the U-statistic to its Gaussian-chaos approximation and constructing time-uniform boundaries for the resulting chaos process. This yields sharp sequential calibration, but estimating the spectral quantities that determine the Gaussian-chaos law involves eigendecomposition of empirical Gram matrices, which can be computationally demanding under dense monitoring.

One way to avoid this obstruction is to introduce a statistic or stochastic structure with a more tractable sequential law. Cross U-statistics use sample splitting and self-normalization to recover Gaussian calibration (Kim & Ramdas, 2024); this idea has led to permutation-free versions of MMD, HSIC, and distance covariance (Shekhar et al., 2022; 2023). Martingale and betting methods provide a different route to sequential nonparametric inference, including two-sample testing (Shekhar & Ramdas, 2024), kernelized independence testing (Podkopaev et al., 2023), interactive rank and conditional-independence testing (Duan et al., 2022; Grünwald et al., 2024), and sequential kernelized goodness-of-fit testing (Martinez-Taboada & Ramdas, 2025).

More recently, Chatterjee & Ramdas (2026) introduce a row-weighted MMD statistic with an asymptotically Gaussian null distribution, yielding fixed-sample tests without spectral estimation. Their multi-kernel extension also yields a Mahalanobis-type test based on a chi-squared limit. These fixed-sample results do not by themselves provide time-uniform error control under repeated monitoring. Building on this row-weighted construction, we develop asymptotic time-uniform inference for general degenerate degree-two kernels. At each time point, the pairwise terms involving the new observation are averaged over the preceding sample and then accumulated across rows. This reweighting preserves the target parameter, while degeneracy makes the centered row averages a martingale-difference sequence. The resulting martingale evolves on the harmonic variance scale $T_n = H_{n-1} \asymp \log n$. The row normalization changes the variance from $O(n^{-2})$ for the ordinary degenerate U-statistic to $O(\log n/n^2)$. Despite this logarithmic inflation, the resulting confidence sequences remain nearly n^{-1}: their radii are of order $\sqrt{\log n \, \log\log\log n}/n$ for the stitched LIL boundary and $\sqrt{\log n \, \log\log n}/n$ for the GM boundary. At the same time, the only nuisance quantity to be estimated is a scalar variance, rather than the spectrum of the kernel operator.

We establish a strong Gaussian approximation on the harmonic clock and control the approximation and scale-estimation errors uniformly over an unbounded monitoring horizon. Combining these results with Gaussian-mixture (GM) and stitched law-of-the-iterated-logarithm (LIL) boundaries yields feasible confidence sequences with asymptotic time-uniform coverage and corresponding one-sided sequential tests. The scalar intervals also form asymptotic confidence sequences in the sense of Waudby-Smith et al. (2024). The same framework extends to multi-dimensional vector-valued kernels: a strong Gaussian approximation for the vector martingale and consistent covariance estimation lead to Mahalanobis confidence ellipsoids calibrated by radial GM and LIL-type Brownian boundaries (de la Peña et al., 2007; Whitehouse et al., 2026).

Numerical experiments examine the finite-sample behavior of the resulting scalar and vector procedures. For MMD and HSIC, we compare the one-sided sequential tests with SAGE (Cai et al., 2026) as a spectral-calibration baseline, focusing on sequential type-I error control and power. The vector experiments assess the same metrics for three-sample homogeneity testing over the monitoring windows considered.

The rest of the paper is organized as follows. Section 2 presents the proposed method and representative examples. Section 3 establishes the main theoretical results, and Section 4 develops an extension to multi-dimensional kernels. Section 5 presents simulations, and Section 6 concludes. Proofs and additional results are collected in the appendix.

2 METHODOLOGY

Let $X_1, X_2, \ldots$ be i.i.d. random elements on $(\mathcal{X}, \mathcal{A})$ with distribution P. Let $h : \mathcal{X}^2 \to \mathbb{R}$ be a symmetric measurable kernel, and let $\theta = \mathbb{E}h(X_1, X_2)$ be the parameter of interest. Write $h_0(x, y) = h(x, y) - \theta$. We assume that the centered kernel is first-order degenerate (also called canonical), $\mathbb{E}\{h_0(X_1, x)\} = 0$ for P-a.e. x. The ordinary degree-two U-statistic is $U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j)$. Under square integrability and degeneracy, $n(U_n - \theta)$ has a quadratic Gaussian-chaos limit. We instead change the weighting of the rows while retaining the target θ. For

$n \geq 2$, define

$$\widetilde{U}_n = \frac{1}{n-1}\sum_{i=2}^{n}\frac{1}{i-1}\sum_{j<i}h(X_i,X_j)$$
$$= \frac{1}{n-1}\left[\sum_{i=2}^{n}\frac{1}{i-1}\sum_{j<i}\{h(X_i,X_j)-\theta\}\right]+\theta := \frac{M_n}{n-1}+\theta.$$

Degeneracy implies that M_n is a martingale with respect to $\mathcal{F}_n = \sigma(X_1,\ldots,X_n)$. If $\tau^2 := \mathbb{E}h_0^2(X_1,X_2) < \infty$, then the centered row averages $Y_i = (i-1)^{-1}\sum_{j<i}h_0(X_i,X_j)$ satisfy $\mathbb{E}Y_i^2 = \tau^2/(i-1)$. Write $T_n = H_{n-1} = \sum_{r=1}^{n-1} r^{-1}$ and set $M_1 = 0$. The resulting orthogonality gives $\mathrm{Var}(\widetilde{U}_n) = \tau^2 T_n/(n-1)^2$, while $\mathrm{Var}(U_n) = 2\tau^2/n(n-1)$. Thus row normalization incurs only a logarithmic variance inflation.

The martingale construction above applies whenever the centered kernel is degenerate. This condition arises naturally under the null hypothesis in a number of kernel- and distance-based testing problems. We record three representative examples, two of which will also be used in the numerical experiments.

Maximum mean discrepancy (MMD). Let $Z_i = (X_i, Y_i)$, where $X_i \sim P$, $Y_i \sim Q$, and the pairs $Z_1, Z_2, \ldots$ are i.i.d. Let $k : \mathcal{X}^2 \to \mathbb{R}$ be a symmetric positive definite kernel. The squared maximum mean discrepancy is estimated by the kernel $h(z,z') = k(x,x') + k(y,y') - k(x,y') - k(x',y)$. Its target parameter is $\theta = \mathbb{E}h(Z_1,Z_2) = \mathrm{MMD}_k^2(P,Q)$. Under the null hypothesis $H_0 : P = Q$, $\theta = 0$ and $\mathbb{E}\{h(z,Z_1)\} = 0$ for P_Z-almost every z, so the kernel is degenerate. For bounded kernels, such as the Gaussian kernel $k(x,y) = \exp\{-\|x-y\|^2/(2\sigma^2)\}$, the fourth-moment condition imposed in Section 3 is immediate.

Hilbert–Schmidt independence criterion (HSIC). Let $Z_i = (X_i, Y_i) \sim P_{XY}$ be i.i.d. observations on $\mathcal{X}\times\mathcal{Y}$. Let k and ℓ be symmetric positive definite kernels on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Define the population-centered kernels $\widetilde{k}(x,x') = k(x,x') - \mathbb{E}k(x,X_1) - \mathbb{E}k(X_1,x') + \mathbb{E}k(X_1,X_2)$ and $\widetilde{\ell}(y,y') = \ell(y,y') - \mathbb{E}\ell(y,Y_1) - \mathbb{E}\ell(Y_1,y') + \mathbb{E}\ell(Y_1,Y_2)$. Then the population-centered HSIC kernel is $h(z,z') = \widetilde{k}(x,x')\widetilde{\ell}(y,y')$. Its target parameter is $\theta = \mathbb{E}h(Z_1,Z_2) = \mathrm{HSIC}(P_{XY}, P_X P_Y)$. Under the null hypothesis $H_0 : X \perp Y$, $\theta = 0$, and $\mathbb{E}\{h(z,Z_1)\} = \mathbb{E}\left\{\widetilde{k}(x,X_1)\widetilde{\ell}(y,Y_1)\right\} = \mathbb{E}\left\{\widetilde{k}(x,X_1)\right\}\mathbb{E}\left\{\widetilde{\ell}(y,Y_1)\right\} = 0$. Thus $h_0 = h$ is degenerate under H_0. If k and ℓ are bounded, then h is bounded. This construction uses population centering; replacing the population quantities by empirical estimates requires additional analysis beyond the fixed-kernel setting considered here.

Quadratic empirical-process functionals. A broader class is obtained from quadratic empirical-process discrepancies. Let $\{\phi_t : t \in \mathcal{T}\}$ be a measurable class of real-valued functions, and let ν be a finite measure on $\mathcal{T}$. Suppose the null hypothesis specifies the moment function $m_0(t) = \mathbb{E}_0\{\phi_t(X)\}$, $t \in \mathcal{T}$. Assume joint measurability in (t,x) and $\mathbb{E}\int|\phi_t(X) - m_0(t)|^2\,d\nu(t) < \infty$, which permit the expectation and integration steps below. A quadratic empirical-process discrepancy is $\theta = \int_{\mathcal{T}}\{\mathbb{E}\phi_t(X) - m_0(t)\}^2\,d\nu(t)$. It is estimated by the degree-two kernel $h(x,y) = \int_{\mathcal{T}}\{\phi_t(x) - m_0(t)\}\{\phi_t(y) - m_0(t)\}\,d\nu(t)$. Indeed, $\mathbb{E}h(X_1,X_2) = \int_{\mathcal{T}}\{\mathbb{E}\phi_t(X) - m_0(t)\}^2\,d\nu(t) = \theta$. Under the null hypothesis $H_0 : \mathbb{E}\phi_t(X) = m_0(t)$ for all $t \in \mathcal{T}$, we have $\theta = 0$, and $\mathbb{E}\{h(x,X_1)\} = 0$. Hence $h_0 = h$ is degenerate under H_0. Classical Cramér–von Mises statistics and related weighted quadratic goodness-of-fit discrepancies are included in this class.

The preceding examples illustrate how the degenerate condition arises under the null hypothesis. To formulate the corresponding sequential guarantees, we recall two related definitions of asymptotic confidence sequences discussed by Waudby-Smith et al. (2024). The first concerns asymptotic time-uniform coverage, while the second characterizes asymptotic equivalence to an exact confidence sequence.

Definition 1 (Asymptotic time-uniform coverage)**.** *For each deterministic integer $m \geq 2$, let $\{C_{m,n}\}_{n\geq m}$ be a sequence of random sets. The family is said to have asymptotic time-uniform*

$(1-\alpha)$-*coverage for* θ *if*

$$\liminf_{m\to\infty} \Pr(\theta \in C_{m,n} \text{ for every } n \geq m) \geq 1-\alpha.$$

Definition 1 requires simultaneous coverage over the entire tail $n \geq m$, with the coverage guarantee obtained as the deterministic monitoring start m tends to infinity.

Definition 2 (Asymptotic confidence sequences)**.** *For a fixed starting point* m*, intervals* $C_{m,n} = [\widehat{\theta}_n \pm w_{m,n}]$*,* $n \geq m$*, form an asymptotic confidence sequence if there exist intervals* $C^*_{m,n} = [\widehat{\theta}_n \pm w^*_{m,n}]$ *satisfying*

$$\mathbb{P}\{\theta \in C^*_{m,n} \text{ for every } n \geq m\} \geq 1-\alpha, \qquad w^*_{m,n}/w_{m,n} \longrightarrow 1 \quad \text{almost surely as } n \to \infty.$$

Definition 2 instead fixes m and requires the candidate intervals to be asymptotically equivalent in width to an exact confidence sequence. These two properties are distinct: asymptotic equivalence to an exact confidence sequence at a fixed starting point does not by itself imply the time-uniform coverage property in Definition 1.

We next record the Gaussian boundary results used in the subsequent analysis. In the scalar case, these are expressed through time-uniform boundaries for standard Brownian motion.

Definition 3 (Brownian motion boundary)**.** *For* $t_0 > 0$*, a nondecreasing function* $u_{\alpha,t_0} : [t_0, \infty) \to (0, \infty)$ *is a two-sided* $(1-\alpha)$ *Brownian motion boundary if, for standard Brownian motion* W*,* $\mathbb{P}\{\exists t \geq t_0 : |W(t)| \geq u_{\alpha,t_0}(t)\} \leq \alpha$.

Two standard choices are used throughout.

Robbins Gaussian-mixture boundary. Let $g(a) = 2\{1 - \Phi(a) + a\phi(a)\}$ for $a \geq 0$, where Φ and ϕ are the standard normal distribution and density functions. The scalar Gaussian-mixture boundary may be written as

$$u^{\mathrm{GM}}_{\alpha,t_0}(t) = \sqrt{t\left[\{g^{-1}(\alpha)\}^2 + \log\left(t/t_0\right)\right]}, \qquad t \geq t_0. \tag{1}$$

Stitched LIL boundary. For $\eta > 1$ and $s > 1$,

$$u^{\mathrm{LIL}}_{\alpha,t_0}(t) = \frac{\eta^{1/4} + \eta^{-1/4}}{\sqrt{2}} \sqrt{t\left[s \log\log\left(\max\left\{\frac{\eta t}{t_0}, e\right\}\right) + \log\left\{\frac{2\zeta(s)}{\alpha\,(\log\eta)^s}\right\}\right]}. \tag{2}$$

Here ζ is the Riemann zeta function. The one-sided stitched formula of Howard et al. (2021) has $\zeta(s)$ in place of $2\zeta(s)$. We allocate $\alpha/2$ to each sign in (2).

Lemma 3 establishes the two-sided validity of both displayed boundaries. For fixed t_0, the LIL boundary has order $O(\sqrt{t \log\log t})$, whereas the Gaussian-mixture boundary has order $O(\sqrt{t \log t})$.

3 Main Results

Throughout this section, we retain the setup and degeneracy condition of Section 2 and impose the following additional condition:

(A1) The kernel satisfies $\mathbb{E}\,|h_0(X_1, X_2)|^4 < \infty$, and $\tau^2 := \mathbb{E}h_0^2(X_1, X_2) > 0$.

We estimate the unknown scale τ^2 by the rowwise plug-in estimator

$$\widehat{\tau}^2_n = \frac{1}{T_n} \sum_{i=2}^{n} \left\{ \frac{1}{i-1} \sum_{j<i} h(X_i, X_j) - \widetilde{U}_n \right\}^2.$$

Recursive implementations of the scalar and vector procedures, together with their computational costs, are given in Appendix B. The following theorem supplies both its consistency and the variance-clock control needed for the strong approximation.

Theorem 1. *Under (A1), $(Y_i, \mathcal{F}_i)$ is a martingale-difference sequence with the following properties.*

(a) The predictable quadratic variation satisfies $\langle M\rangle_n/T_n := \sum_{i=2}^n \mathbb{E}\left(Y_i^2 \mid \mathcal{F}_{i-1}\right)/T_n \xrightarrow{\text{a.s.}} \tau^2$. Moreover, $\langle M\rangle_n - \tau^2 T_n = o_{a.s.}\left(T_n/\log T_n\right)$.

(b) The optional quadratic variation satisfies $[M]_n/T_n := \sum_{i=2}^n Y_i^2/T_n \xrightarrow{\text{a.s.}} \tau^2$.

(c) The plug-in estimator is strongly consistent, that is $\widehat{\tau}_n^2 \xrightarrow{\text{a.s.}} \tau^2$.

Since $T_n \asymp \log n$, Theorem 1 identifies the harmonic clock as the natural variance scale of M_n. The next theorem establishes a strong Gaussian approximation on this time scale.

Theorem 2. *Under (A1), on an enriched probability space there exists a standard Brownian motion W such that $M_n = \tau W(T_n) + o_{a.s.}\left(\sqrt{T_n}\right)$. Equivalently,*

$$\widetilde{U}_n - \theta = \frac{\tau W(T_n)}{n-1} + o_{a.s.}\left(\frac{\sqrt{T_n}}{n}\right).$$

Combining the Brownian approximation with either boundary in Section 2 gives the following sequence of intervals:

$$C_{m,n}^{\mathrm{MG}}(\alpha) = \left[\widetilde{U}_n \pm \widehat{\tau}_n \frac{u_{\alpha,T_m}(T_n)}{n-1}\right], \qquad n \geq m. \tag{3}$$

The following theorem establishes asymptotic time-uniform coverage of (3) in the sense of Definition 1.

Theorem 3. *Under (A1), let u_{α,t_0} be either (1) or (2). Then*

$$\liminf_{m\to\infty} \mathbb{P}\left(\forall n \geq m : \theta \in C_{m,n}^{\mathrm{MG}}(\alpha)\right) \geq 1-\alpha.$$

Since $T_n \asymp \log n$ and $\widehat{\tau}_n \xrightarrow{\text{a.s.}} \tau$, the interval-width orders follow directly from the growth rates of the two Brownian boundaries. For fixed m and the boundary tuning parameters, the radii of $C_{m,n}^{\mathrm{MG,LIL}}$ and $C_{m,n}^{\mathrm{MG,GM}}$ are of order $O_{a.s.}\left(\sqrt{\log n \log\log\log n}/n\right)$ and $O_{a.s.}\left(\sqrt{\log n \log\log n}/n\right)$ respectively.

Remark 1. *For every fixed $m \geq 2$, the sequence $\{C_{m,n}^{\mathrm{MG}}(\alpha) : n \geq m\}$ also satisfies Definition 2. More precisely, under (A1) and for either boundary, there exists a nonasymptotic $(1-\alpha)$-confidence sequence $\{C_{m,n}^{*}(\alpha) : n \geq m\}$, centered at $\widetilde{U}_n$, such that $\mathrm{rad}\{C_{m,n}^{*}(\alpha)\}/\mathrm{rad}\{C_{m,n}^{\mathrm{MG}}(\alpha)\} \xrightarrow{\text{a.s.}} 1$, as $n \to \infty$. This property complements the time-uniform coverage guarantee of Theorem 3.*

Remark 2. *In many applications of degenerate U-statistics, degeneracy arises under a null hypothesis and the relevant alternative is one-sided. Examples include MMD and HSIC, for which the population discrepancy is nonnegative and the kernel is degenerate under the null hypothesis with $\theta_0 = 0$.*

For $\eta > 1$, $s > 1$, and $\alpha \in (0,1)$, the upper stitched LIL boundary is

$$u_{\alpha,t_0}^{+,\mathrm{LIL}}(t) = \frac{\eta^{1/4} + \eta^{-1/4}}{\sqrt{2}} \sqrt{t\left[s\log\log\left(\max\left\{\frac{\eta t}{t_0}, e\right\}\right) + \log\left\{\frac{\zeta(s)}{\alpha(\log\eta)^s}\right\}\right]}, \qquad t \geq t_0,$$

which differs from (2) only in that the full error probability α is allocated to the upper tail.

For the Gaussian-mixture construction, let $g_+(a) = 1 - \Phi(a) + a\phi(a) + \phi(a)^2/\Phi(a)$, $a \geq 0$, and, for $0 < \alpha < g_+(0) = 1/2 + 1/\pi$, define $g_+^{-1}(\alpha) = \inf\{a \geq 0 : g_+(a) \leq \alpha\}$. The corresponding upper boundary is

$$u_{\alpha,t_0}^{+,\mathrm{GM}}(t) = \sqrt{t\left\{[g_+^{-1}(\alpha)]^2 + \log(t/t_0)\right\}}, \qquad t \geq t_0.$$

Lemma 4 establishes that both boundaries satisfy $\mathbb{P}\left\{\exists t \geq t_0 : W(t) \geq u_{\alpha,t_0}^{+}(t)\right\} \leq \alpha$ for their respective parameter ranges.

Corollary 1. *Fix a distribution under $H_0 : \theta = \theta_0$ for which the centered kernel $h-\theta_0$ is degenerate and satisfies (A1). Let u^+_{α,t_0} be either boundary in Remark 2, with parameters in the stated range. Then*

$$\limsup_{m\to\infty} \mathbb{P}_{H_0}\left\{\exists n \geq m : (n-1)(\widetilde{U}_n - \theta_0) > \widehat{\tau}_n u^+_{\alpha,T_m}(T_n)\right\} \leq \alpha.$$

Thus, the event in the preceding display is the rejection region for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, with rejection occurring at the first boundary crossing.

Moreover, the same rejection rule is consistent under fixed alternatives satisfying the fourth-moment condition in (A1); see Appendix C.6.

4 Extension to multi-dimensional kernels

We now extend the martingaleized construction to multi-dimensional kernels. Let $\boldsymbol{h} : \mathcal{X}^2 \to \mathbb{R}^d$ be a symmetric measurable kernel, where d is fixed, and define $\boldsymbol{\theta} = \mathbb{E}\boldsymbol{h}(X_1, X_2) \in \mathbb{R}^d$ and $\boldsymbol{h}_0(x, y) = \boldsymbol{h}(x, y) - \boldsymbol{\theta}$. We assume that the centered kernel is degenerate componentwise: $\mathbb{E}\{\boldsymbol{h}_0(X_1, x)\} = \mathbf{0} \in \mathbb{R}^d$ for P-a.e. x. Define the corresponding statistic

$$\widetilde{\boldsymbol{U}}_n = \frac{1}{n-1}\sum_{i=2}^{n} \frac{1}{i-1}\sum_{j<i} \boldsymbol{h}(X_i, X_j),$$

and, analogously to the scalar case, denote $\boldsymbol{Y}_i = \sum_{j<i} \boldsymbol{h}_0(X_i, X_j)/(i-1)$ and $\boldsymbol{M}_n = \sum_{i=2}^{n} \boldsymbol{Y}_i$. Then $\boldsymbol{M}_n$ is an $\mathbb{R}^d$-valued martingale with respect to $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

Throughout this section, we impose the following additional condition:

(A2) The kernel satisfies $\mathbb{E}\,\|\boldsymbol{h}_0(X_1, X_2)\|_2^4 < \infty$, and $\boldsymbol{\Omega} := \mathbb{E}\left\{\boldsymbol{h}_0(X_1, X_2)\boldsymbol{h}_0(X_1, X_2)^\mathsf{T}\right\}$ is positive definite.

We estimate $\boldsymbol{\Omega}$ by the rowwise plug-in estimator

$$\widehat{\boldsymbol{\Omega}}_n = \frac{1}{T_n}\sum_{i=2}^{n}\left\{\frac{1}{i-1}\sum_{j<i}\boldsymbol{h}(X_i, X_j) - \widetilde{\boldsymbol{U}}_n\right\}\left\{\frac{1}{i-1}\sum_{j<i}\boldsymbol{h}(X_i, X_j) - \widetilde{\boldsymbol{U}}_n\right\}^\mathsf{T}.$$

Lemma 5 establishes that $\widehat{\boldsymbol{\Omega}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$ under (A2). We next give a strong Gaussian approximation for the vector martingale on the harmonic clock.

Proposition 1. *Under (A2), on an extension of the underlying probability space, there exist a standard d-dimensional Brownian motion $\boldsymbol{W}$ and a constant $\kappa > 0$ such that*

$$\left\|\boldsymbol{M}_n - \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(T_n)\right\|_2 = O_{a.s.}(T_n^{1/2-\kappa}) = o_{a.s.}(T_n^{1/2}).$$

To translate this strong Gaussian approximation into joint time-uniform inference, we use radial boundaries for the Euclidean norm of a d-dimensional Brownian motion. These are the vector counterparts of the scalar Brownian boundaries introduced in Section 2.

Definition 4 (Radial Brownian motion boundary)**.** *For $t_0 > 0$, a nondecreasing function $b_{d,\alpha,t_0} : [t_0, \infty) \to (0, \infty)$ is called a radial $(1-\alpha)$ Brownian motion boundary if, for a standard d-dimensional Brownian motion $\boldsymbol{W}_d$, $\mathbb{P}\{\exists t \geq t_0 : \|\boldsymbol{W}_d(t)\|_2 \geq b_{d,\alpha,t_0}(t)\} \leq \alpha$.*

We use the following Gaussian-mixture and LIL-type radial boundaries.

Gaussian-mixture radial boundary. An explicit radial boundary follows from a Gaussian mixture of Brownian exponential martingales; see de la Peña et al. (2007) and Whitehouse et al. (2026, Fact 4.3). For a standard d-dimensional Brownian motion, the boundary is

$$b^{\mathrm{GM}}_{d,\alpha,t_0}(t) = \left[(t+t_0)\left\{d\log\left(1+\frac{t}{t_0}\right) + 2\log(1/\alpha)\right\}\right]^{1/2}, \qquad t \geq t_0. \tag{4}$$

LIL-type radial boundary. We use the vector self-normalized inequality of Whitehouse et al. (2026, Theorem 4.1). Fix $a > 1$, $\beta > 1$, $\varepsilon \in (0, 1)$, and $q > 1$. For $t \geq t_0$, define

$$b^{\mathrm{LIL}}_{d,\alpha,t_0}(t) = \frac{\sqrt{2at}}{1-\varepsilon}\left[q \log\left\{1 + \log_a\left(\frac{t}{t_0}\right)\right\} + \log\frac{\zeta(q)\beta}{\alpha(1-\beta^{-1})} + d\log\left(\frac{3\beta}{\varepsilon}\right)\right]^{1/2}. \tag{5}$$

For fixed t_0, d, and tuning parameters, the Gaussian-mixture boundary has order $O(\sqrt{t \log t})$, whereas the LIL-type boundary has the sharper order $O(\sqrt{t \log\log t})$. Both choices are valid radial Brownian boundaries.

Using the radial boundary b_{d,α,t_0}, define, for $n \geq m$, the Mahalanobis confidence ellipsoid

$$C^{\mathrm{ell}}_{m,n}(\alpha) = \left\{\boldsymbol{\vartheta} \in \mathbb{R}^d : (n-1)^2 \left(\widetilde{\boldsymbol{U}}_n - \boldsymbol{\vartheta}\right)^{\mathsf{T}} \widehat{\boldsymbol{\Omega}}_n^{-1} \left(\widetilde{\boldsymbol{U}}_n - \boldsymbol{\vartheta}\right) \leq \{b_{d,\alpha,T_m}(T_n)\}^2\right\}, \tag{6}$$

with the convention that $C^{\mathrm{ell}}_{m,n}(\alpha) = \mathbb{R}^d$ whenever $\widehat{\boldsymbol{\Omega}}_n$ is singular.

The following theorem gives asymptotic time-uniform coverage of these ellipsoids in the sense of Definition 1.

Theorem 4. *Suppose that (A2) holds, and choose either (4) or (5). Then*

$$\liminf_{m\to\infty} \mathbb{P}\left\{\forall n \geq m : \boldsymbol{\theta} \in C^{\mathrm{ell}}_{m,n}(\alpha)\right\} \geq 1 - \alpha.$$

Remark 3. *Under (A2), for every fixed $m \geq 2$ and either radial boundary, the ellipsoids satisfy a multivariate analogue of Definition 2. Specifically, in the sense of Waudby-Smith et al. (2024, Definition B.6), there exists a nonasymptotic $(1-\alpha)$-confidence sequence $\{C^{*,\mathrm{ell}}_{m,n}(\alpha) : n \geq m\}$, centered at $\widetilde{\boldsymbol{U}}_n$, such that $\nu\left(C^{\mathrm{ell}}_{m,n}(\alpha) \triangle C^{*,\mathrm{ell}}_{m,n}(\alpha)\right)/\nu\left(C^{*,\mathrm{ell}}_{m,n}(\alpha)\right) \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. Here ν denotes Lebesgue measure on $\mathbb{R}^d$, and $\triangle$ denotes symmetric difference.*

5 Simulation

We study the finite-sample behavior of the scalar and vector procedures developed in Sections 3 and 4. For the scalar case, we evaluate the one-sided sequential tests in MMD and HSIC problems and compare them with the spectral SAGE procedures of Cai et al. (2026). For the vector case, we examine simultaneous coverage and rejection behavior in a three-sample kernel testing problem. Throughout, the nominal level is $\alpha = 0.05$.

5.1 One-Sided Sequential Testing

Setup and procedures. We consider the MMD and population-centered HSIC examples introduced in Section 2. For MMD, let $Z_i = (X_i, Y_i)$, where $X_i \sim \mathcal{N}(0,1)$ and $Y_i \sim N(\delta, 1)$ are independent. For $z = (x, y)$ and $z' = (x', y')$, use $k(x, x') = \exp\{-(x - x')^2/2\}$ and $h(z, z') = k(x, x') + k(y, y') - k(x, y') - k(x', y)$. The null corresponds to $\delta = 0$, and the alternatives are $\delta \in \{0.05, 0.10, \ldots, 0.45\}$.

For HSIC, we generate (X_i, Y_i) from a centered bivariate normal distribution with unit marginal variances and correlation ρ, and use $h(z, z') = \widetilde{k}(x, x')\widetilde{k}(y, y')$, where the population-centered Gaussian kernel is available analytically $\widetilde{k}(x, x') = e^{-(x-x')^2/2} - (e^{-x^2/4} + e^{-(x')^2/4})/\sqrt{2} + 1/\sqrt{3}$. The independence null corresponds to $\rho = 0$, and the alternatives are $\rho \in \{0.05, 0.10, \ldots, 0.50\}$. Both kernels are bounded and satisfy (A1).

We implement the one-sided MG-GM and MG-LIL tests from Section 3 using the harmonic clock T_n and the rowwise plug-in variance estimator. MG-GM uses the one-sided Gaussian-mixture boundary, while MG-LIL uses the one-sided stitched LIL boundary with $\eta = 2$ and $s = 1.4$.

As spectral benchmarks, we use the one-sided SAGE-GM and SAGE-LIL procedures of Cai et al. (2026), based on the ordinary U-statistic. Following their simulation implementation, we retain $L_n = \lfloor n^{1/4} \rfloor$ empirical eigenvalues and update the spectral allocation weights proportionally to

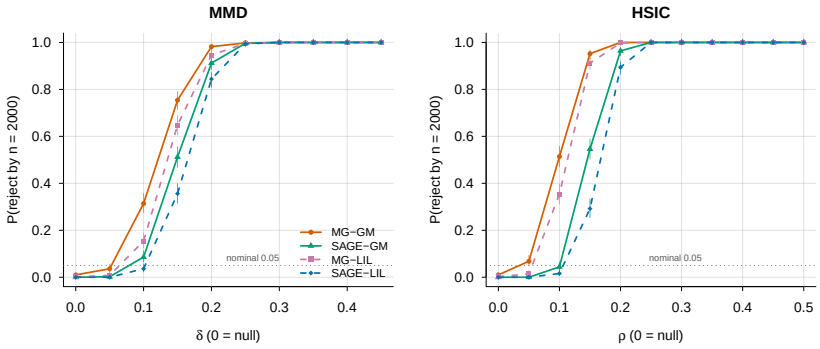


Figure 1: Cumulative rejection probability through $n = 2000$ for the one-sided scalar tests. The left panel varies the MMD mean shift δ, and the right panel varies the Gaussian correlation ρ for HSIC.

$(\widehat{\lambda}_{\ell,n})_+$ at each reporting time n. For SAGE-LIL, we use the same stitching parameters as for MG-LIL.

We begin monitoring at $m = 400$ over the grid $\mathcal{G} = \{400, 401, \ldots, 2000\}$. For each example, empirical type-I error is evaluated using 2,000 Monte Carlo replications, and power at each nonzero effect size is evaluated using 500 replications. Table 1 reports cumulative type-I error through $n = 2000$, while Figure 1 presents the corresponding power curves. To compare detection speed at effect size 0.15, let $\nu = \inf\{n \in \mathcal{G} : \text{the rejection criterion is satisfied at time } n\}$, with $\inf \varnothing = \infty$. Table 2 reports the empirical rejection probability by N and the mean capped rejection time $\min\{\nu, N + 1\}$; a path without rejection by N is assigned $N + 1 = 2001$ for this summary.

Table 1: Cumulative type-I error under monitoring over the grid $\mathcal{G} = \{400, 401, \ldots, 2000\}$.

Example	MG-GM	SAGE-GM	MG-LIL	SAGE-LIL
RBF-MMD	0.0100	0.0000	0.0015	0.0000
HSIC	0.0100	0.0000	0.0025	0.0000

Results. Table 1 shows that the martingale procedures are less conservative than SAGE while retaining empirical type-I error below the nominal level; neither SAGE procedure rejects under the null. Moreover, the martingale procedures have substantially higher empirical power against moderate alternatives, especially for HSIC: Table 2 gives the powers at effect size 0.15. By effect size 0.25, all four procedures have power close to one, as shown in Figure 1.

The mean capped first rejection times in Table 2 reinforce the power comparison. At effect size 0.15, the mean capped times for MG-GM and MG-LIL are approximately 367 and 306 observations lower than those for SAGE-GM and SAGE-LIL, respectively, for MMD. The corresponding differences for HSIC are approximately 794 and 779 observations. Together with the higher endpoint power, these shorter detection times provide additional evidence that the martingale procedures detect the alternatives more rapidly in the settings considered.

The proposed procedures also offer substantial computational advantages over SAGE; see Appendix B for timing comparisons.

Table 2: Rejection probability and mean capped first-rejection time at effect size 0.15.

Example	Measure	MG-GM	SAGE-GM	MG-LIL	SAGE-LIL
RBF-MMD, $\delta = 0.15$	Rejection by N	0.754	0.512	0.648	0.356
	Mean capped time	1239.8	1606.9	1460.6	1766.1
HSIC, $\rho = 0.15$	Rejection by N	0.952	0.546	0.912	0.292
	Mean capped time	881.9	1675.6	1068.8	1847.9

5.2 Three-Sample Kernel Testing with Vector Confidence Ellipsoids

Setup and procedures. We consider the three-sample homogeneity null $H_0 : P_1 = P_2 = P_3$. At each time i, we observe $\boldsymbol{Z}_i = (\boldsymbol{X}_i^{(1)}, \boldsymbol{X}_i^{(2)}, \boldsymbol{X}_i^{(3)})$, where $\boldsymbol{X}_i^{(a)} \in \mathbb{R}^2$ has distribution P_a. Observations are independent across groups and times. We use the fixed Gaussian kernel $k(\boldsymbol{x}, \boldsymbol{y}) = \exp\{-\|\boldsymbol{x} - \boldsymbol{y}\|_2^2/4\}$. For $\boldsymbol{z} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)})$, $\boldsymbol{z}' = (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \boldsymbol{y}^{(3)})$, and $(a, b) \in \{(1,2), (1,3), (2,3)\}$, define $h_{ab}(\boldsymbol{z}, \boldsymbol{z}') = k(\boldsymbol{x}^{(a)}, \boldsymbol{y}^{(a)}) + k(\boldsymbol{x}^{(b)}, \boldsymbol{y}^{(b)}) - k(\boldsymbol{x}^{(a)}, \boldsymbol{y}^{(b)}) - k(\boldsymbol{x}^{(b)}, \boldsymbol{y}^{(a)})$. The vector kernel $\boldsymbol{h} = (h_{12}, h_{13}, h_{23})^\top$ has target $\boldsymbol{\theta} = \mathbb{E}\{\boldsymbol{h}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)\} = \big(\mathrm{MMD}_k^2(P_1, P_2), \mathrm{MMD}_k^2(P_1, P_3), \mathrm{MMD}_k^2(P_2, P_3)\big)^\top$.

Under H_0, $\boldsymbol{\theta} = \mathbf{0}$ and $\boldsymbol{h}$ is componentwise degenerate. Under the null, we set $P_1 = P_2 = P_3 = \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ and corresponding $\boldsymbol{\Omega} = (3\mathbf{I}_3 + \mathbf{1}_3\mathbf{1}_3^\top)/20$ is positive definite. We set $P_1 = P_2 = \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ and $P_3 = \mathcal{N}(\boldsymbol{\mu}_\delta, \mathbf{I}_2)$, where $\boldsymbol{\mu}_\delta = (\delta, 0)^\top$ and $\delta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ under the alternatives. We use 2,000 Monte Carlo replications under the null and 500 at each nonzero effect size. Null coverage is evaluated for $m \in \{20, 30, 40\}$, while the power curves use $m = 30$; all monitoring ends at $N = 2000$.

We evaluate the ellipsoids in (6) using (4) (VEC-GM) and (5) (VEC-WWR-LIL), with $d = 3$, $\alpha = 0.05$, and $t_0 = T_m$. For VEC-WWR-LIL, we set $a = 2$, $\beta = 2$, $\varepsilon = 0.5$, and $q = 1.4$. Here we report simultaneous coverage over $n = m, \ldots, N$ under H_0 and the power across different δ.

m	VEC-GM	VEC-WWR-LIL
20	0.9620	1.0000
30	0.9780	1.0000
40	0.9810	1.0000

Table 3: Simultaneous coverage under the three-sample homogeneity null.

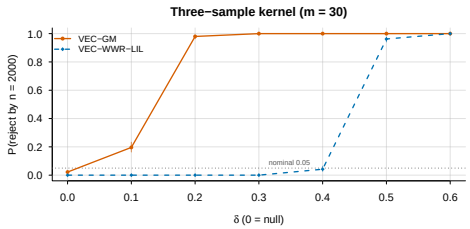


Figure 2: Rejection probability by $N = 2000$ against mean shift δ, with $m = 30$.

Results. Table 3 shows empirical simultaneous coverage between 0.9620 and 0.9810 for VEC-GM. VEC-WWR-LIL covers all 2,000 null paths at each monitoring start. Figure 2 shows a substantial difference in detection power: VEC-GM rejects on nearly all paths at $\delta = 0.2$, whereas VEC-WWR-LIL rarely rejects through $\delta = 0.4$. The two procedures use the same estimator and estimated covariance, so this difference reflects their radial boundaries. At $n = 2000$, the GM and LIL boundary values are approximately 10.65 and 42.76, respectively. VEC-GM is thus the more useful choice for detection within the monitoring window considered.

6 Concluding Remarks

We develop a martingale approach to sequential inference for degenerate degree-two kernels. Row normalization preserves the target parameter and enables a strong Gaussian approximation on the harmonic variance clock, at the cost of logarithmic variance inflation. Combined with consistent variance and covariance estimation, this approximation yields scalar confidence sequences and multi-dimensional Mahalanobis confidence ellipsoids with asymptotic time-uniform coverage, together with one-sided sequential tests. In the MMD and HSIC experiments, the martingale procedures are less conservative than SAGE and attain higher empirical power against moderate alternatives at the same nominal level.

The results also point to two directions for further research. The analysis covers fixed kernels; empirical centering, as in practical implementations of HSIC (Gretton et al., 2005), remains outside this scope and requires time-uniform control of the resulting centering error. The vector theory is also restricted to multi-dimensional kernels with fixed d. Extending it to high-dimensional settings may build on dimension-explicit martingale couplings (Cattaneo et al., 2025) and regularized covariance estimation (Bickel & Levina, 2008), but their integration into time-uniform inference remains open.

AI USE STATEMENT

Generative AI tools were used to revise wording and improve the clarity and readability of the manuscript; to identify potentially relevant literature, references, and search directions, and to assist with research execution, including drafting, debugging, and refining code. For mathematical proofs, AI was used primarily to review and check the authors' arguments for potential errors, inconsistencies, or gaps. AI-assisted manuscript revisions, literature-search suggestions, feedback on research ideas and proofs, and generated or modified code were reviewed and verified by the authors before incorporation into the final work. The authors independently assessed the accuracy and appropriateness of these outputs and retain full responsibility for the methodology, proofs, implementation, analysis, and conclusions presented in the paper.

REPRODUCIBILITY STATEMENT

All assumptions and complete proofs of the theoretical results are provided in the appendix. Section 5 describes the data-generating mechanisms, kernel choices, monitoring windows, boundary parameters, and numbers of Monte Carlo replications used in the numerical experiments. Code for reproducing the reported results is included in the anonymous supplementary material.

REFERENCES

Miguel A. Arcones and Evarist Giné. Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542, 1993. doi: 10.1214/aop/1176989128.

Miguel A. Arcones and Evarist Giné. On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Processes and their Applications*, 58(2):217–245, 1995. doi: 10.1016/0304-4149(94)00023-M.

Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008. doi: 10.1214/08-AOS600.

Leheng Cai, Qirui Hu, and Weijia Li. Asymptotic anytime-valid inference for U-statistics, 2026. URL https://arxiv.org/abs/2605.14692.

Matias D. Cattaneo, Ricardo P. Masini, and William G. Underwood. Yurinskii's coupling for martingales. *The Annals of Statistics*, 53(5):2179–2203, 2025. doi: 10.1214/25-AOS2538.

Anirban Chatterjee and Aaditya Ramdas. A martingale kernel two-sample test. In *Proceedings of the 37th International Conference on Algorithmic Learning Theory*, volume 313 of *Proceedings of Machine Learning Research*, pp. 1–44, 2026. URL https://proceedings.mlr.press/v313/chatterjee26a.html.

Victor H. de la Peña, Michael J. Klass, and Tze Leung Lai. Pseudo-maximization and self-normalized processes. *Probability Surveys*, 4:172–192, 2007. doi: 10.1214/07-PS119.

Herold Dehling, Manfred Denker, and Walter Philipp. Invariance principles for von Mises and U-statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67:139–167, 1984.

Boyan Duan, Aaditya Ramdas, and Larry Wasserman. Interactive rank testing by betting. In *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pp. 201–235. PMLR, 2022.

Felix Gnettner and Claudia Kirch. A new and flexible class of sharp asymptotic time-uniform confidence sequences. *Statistics & Probability Letters*, 226:110462, 2025. doi: 10.1016/j.spl.2025.110462.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pp. 63–77. Springer, 2005. doi: 10.1007/11564089_7.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Peter Grünwald, Alexander Henzi, and Tyron Lardy. Anytime-valid tests of conditional independence under Model-X. *Journal of the American Statistical Association*, 119(546):1554–1565, 2024. doi: 10.1080/01621459.2023.2205607.

Peter Hall and Christopher C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.

Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. doi: 10.1214/aoms/1177730196.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020. doi: 10.1214/18-PS321.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

S. Kanagawa and K. Yoshihara. The almost sure invariance principles of degenerate U-statistics of degree two for stationary random variables. *Stochastic Processes and their Applications*, 49 (2):347–356, 1994. ISSN 0304-4149. doi: 10.1016/0304-4149(94)90141-4. URL https://www.sciencedirect.com/science/article/pii/0304414994901414.

Ilmun Kim and Aaditya Ramdas. Dimension-agnostic inference using cross U-statistics. *Bernoulli*, 30(1):683–711, 2024. doi: 10.3150/23-BEJ1613.

Diego Martinez-Taboada and Aaditya Ramdas. Sequential kernelized Stein discrepancy. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1288–1296. PMLR, 2025.

Georg Neuhaus. Functional limit theorems for U-statistics in the degenerate case. *Journal of Multivariate Analysis*, 7(3):424–439, 1977. doi: 10.1016/0047-259X(77)90083-5.

Aleksandr Podkopaev, Patrick Blöbaum, Shiva Kasiviswanathan, and Aaditya Ramdas. Sequential kernelized independence testing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27957–27993. PMLR, 2023.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023. doi: 10.1214/23-STS894.

Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, 1970.

Herbert Robbins and David Siegmund. Boundary crossing probabilities for the Wiener process and sample sums. *The Annals of Mathematical Statistics*, 41(5):1410–1429, 1970. doi: 10.1214/aoms/1177696787.

Dino Sejdinovic, Bharath Sriperumbudur, Arthur Gretton, and Kenji Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5): 2263–2291, 2013. doi: 10.1214/13-AOS1140.

Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 1980. doi: 10.1002/9780470316481.

Shubhanshu Shekhar and Aaditya Ramdas. Nonparametric two-sample testing by betting. *IEEE Transactions on Information Theory*, 70(2):1178–1203, 2024. doi: 10.1109/TIT.2023.3305867.

Shubhanshu Shekhar, Ilmun Kim, and Aaditya Ramdas. A permutation-free kernel two-sample test. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18168–18180, 2022. doi: 10.52202/068431-1321.

Shubhanshu Shekhar, Ilmun Kim, and Aaditya Ramdas. A permutation-free kernel independence test. *Journal of Machine Learning Research*, 24(369):1–68, 2023.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013. doi: 10.1016/j.jspi.2013.03.018.

Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007. doi: 10.1214/009053607000000505.

Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H. Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *The Annals of Statistics*, 52 (6):2613–2640, 2024.

Justin Whitehouse, Zhiwei Steven Wu, and Aaditya Ramdas. Time-uniform self-normalized concentration for vector-valued processes. *The Annals of Applied Probability*, 36(3):1972–2013, 2026. doi: 10.1214/25-AAP2246.

Li-xin Zhang. Strong approximations of martingale vectors and their applications in Markov-chain adaptive designs. *Acta Mathematicae Applicatae Sinica, English Series*, 20(2):337–352, 2004. doi: 10.1007/s10255-004-0173-z.

A Additional Vector Coverage Results

We also evaluate the vector ellipsoids for the kernel $\boldsymbol{h}(x, y) = \mathbf{A}(x \odot y)$, where $\mathbf{A} = \mathrm{diag}(1, 2, 1/2)$ and $\odot$ denotes elementwise multiplication. The observations have mean zero and covariance matrix $\boldsymbol{\Sigma}$ with unit diagonal entries and off-diagonal entries equal to $1/2$. Consequently, $\boldsymbol{\theta} = \mathbf{0}$ and $\boldsymbol{\Omega} = \mathbf{A}(\boldsymbol{\Sigma} \odot \boldsymbol{\Sigma})\mathbf{A}$. We consider both $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ observations and a standardized multivariate t_5 distribution with covariance $\boldsymbol{\Sigma}$. For each distribution, 2,000 Monte Carlo paths are monitored from $m \in \{20, 30, 40\}$ through $N = 2000$, using the VEC-GM and VEC-WWR-LIL procedures specified in Section 5.2.

Table 4: Simultaneous coverage for the vector product kernel under Gaussian and standardized t_5 observations.

	Gaussian		t_5	
m	VEC-GM	VEC-WWR-LIL	VEC-GM	VEC-WWR-LIL
20	0.9640	1.0000	0.9750	1.0000
30	0.9795	1.0000	0.9855	1.0000
40	0.9875	1.0000	0.9885	1.0000

Table 4 shows empirical simultaneous coverage above the nominal level in all settings. VEC-WWR-LIL covers all 2,000 paths in each setting; the Gaussian and standardized t_5 results are similar.

B Online implementation and computational cost

Scalar updates. For each new observation, compute the uncentered row average $a_n = \sum_{j<n} h(X_n, X_j)/(n-1)$ where $n \geq 2$. The statistic and its scale estimator can be written as $\widetilde{U}_n = \sum_{i=2}^{n} a_i/(n-1)$ and $\widehat{\tau}_n^2 = S_n/T_n = \sum_{i=2}^{n}(a_i - \widetilde{U}_n)^2/T_n$.

Algorithm 1 updates the mean and centered sum of squares recursively. The identity $S_n = \sum_{i=2}^{n} a_i^2 - (n-1)\widetilde{U}_n^2$ verifies that these updates compute the original estimators exactly in exact arithmetic. Maintaining the centered sum of squares also avoids subtracting two potentially large raw second-moment terms. All observations, including those received before the monitoring start m, contribute to the updates.

Algorithm 1 Online scalar martingaleized inference

Require: Kernel h, observations $X_1, X_2, \ldots$, monitoring start $m \geq 2$, and level α.
Require: Selected two-sided boundary u_{α,t_0}; for one-sided testing, also select u^{+}_{α,t_0} and specify θ_0.

Store X_1; initialize $\widetilde{U}_1 \leftarrow 0$, $S_1 \leftarrow 0$, and $T_1 \leftarrow 0$.
for $n = 2, 3, \ldots$ **do**
  Receive X_n and compute $a_n \leftarrow (n-1)^{-1} \sum_{j<n} h(X_n, X_j)$.
  $\Delta \leftarrow a_n - \widetilde{U}_{n-1}$.
  $\widetilde{U}_n \leftarrow \widetilde{U}_{n-1} + \Delta/(n-1)$.
  $S_n \leftarrow S_{n-1} + \{(n-2)/(n-1)\}\Delta^2$.
  $T_n \leftarrow T_{n-1} + 1/(n-1)$; $\widehat{\tau}_n^2 \leftarrow S_n/T_n$; $\widehat{\tau}_n \leftarrow \sqrt{\widehat{\tau}_n^2}$.
  if $n = m$ **then**
    Store $t_0 \leftarrow T_n$.
  end if
  if $n \geq m$ **then**
    Report $C^{\mathrm{MG}}_{m,n}(\alpha) = [\widetilde{U}_n \pm \widehat{\tau}_n u_{\alpha,t_0}(T_n)/(n-1)]$.
    For one-sided testing, reject at the first time satisfying
   $(n-1)(\widetilde{U}_n - \theta_0) > \widehat{\tau}_n u^{+}_{\alpha,t_0}(T_n)$.
  end if
  Store X_n for subsequent row evaluations.
end for

Vector updates. For a d-dimensional kernel, define $\mathbf{a}_n = (n-1)^{-1}\sum_{j<n}\mathbf{h}(X_n, X_j)$ and $\mathbf{\Delta}_n = \mathbf{a}_n - \widetilde{\mathbf{U}}_{n-1}$. Starting from $\widetilde{\mathbf{U}}_1 = \mathbf{0}$ and $\mathbf{S}_1 = \mathbf{0}_{d\times d}$, we use the update

$$\widetilde{\mathbf{U}}_n = \widetilde{\mathbf{U}}_{n-1} + \frac{\mathbf{\Delta}_n}{n-1},$$
$$\mathbf{S}_n = \mathbf{S}_{n-1} + \frac{n-2}{n-1}\mathbf{\Delta}_n\mathbf{\Delta}_n^{\mathsf{T}} \text{ and } \widehat{\mathbf{\Omega}}_n = \frac{\mathbf{S}_n}{T_n}.$$

These updates maintain $\mathbf{S}_n = \sum_{i=2}^{n}(\mathbf{a}_i - \widetilde{\mathbf{U}}_n)(\mathbf{a}_i - \widetilde{\mathbf{U}}_n)^{\mathsf{T}}$, and hence recover the covariance estimator in Section 4. At monitored times, use the selected radial boundary to construct the Mahalanobis ellipsoid. As in the main text, the ellipsoid is $\mathbb{R}^d$ when $\widehat{\mathbf{\Omega}}_n$ is singular. When it is positive definite, the Mahalanobis statistic can be evaluated by a Cholesky solve.

Computational cost. At time n, computing the new row average requires $n-1$ kernel evaluations. Given this average, updating the estimator, harmonic clock, and scalar variance estimate takes $O(1)$ arithmetic operations; the vector covariance update takes $O(d^2)$. A fresh Cholesky factorization for the Mahalanobis statistic costs $O(d^3)$ at each monitored time. Hence, for fixed d and constant-cost kernel evaluations, the total computational cost through time N is $O(N^2)$. The implementation retains past observations and uses $O(d^2)$ additional storage, without storing past kernel values.

Runtime comparison. We compare the per-update running times of the proposed MG procedure and SAGE under the MMD and HSIC null models of Section 5.1. At $n \in \{400, 800, 1200, 1600, 2000\}$, Table 5 reports the median time over ten independent paths to update both the GM and LIL decisions. Timings include kernel evaluations and, for SAGE, spectral computation, but exclude data generation and initialization. All procedures were timed on the same Apple M4. The $n = 400$ spectral update uses a direct eigendecomposition of a 200×200 matrix; the remaining checkpoints use a warm-started subspace eigensolver.

Table 5: Median time in milliseconds for one monitoring update under the null, based on ten independent paths.

	MMD		HSIC	
n	MG	SAGE	MG	SAGE
400	0.014	5.089	0.013	5.916
800	0.020	3.532	0.018	6.289
1200	0.027	6.959	0.024	12.325
1600	0.034	12.646	0.029	22.761
2000	0.043	19.197	0.035	33.734

C Proofs of scalar results

C.1 Variance clock

Proof of Theorem 1. Let

$$q(x, y) = \mathbb{E}\left\{h_0(Z, x)h_0(Z, y)\right\}, \qquad Z \sim P \text{ independent.}$$

Then q is symmetric. By Jensen's inequality,

$$\mathbb{E}q^2(X_1, X_2) \le \mathbb{E}\left\{h_0^2(Z, X_1)h_0^2(Z, X_2)\right\} \le \mathbb{E}h_0^4(X_1, X_2) < \infty,$$

and similarly $\mathbb{E}q^2(X_1, X_1) < \infty$. Moreover, q is canonical because

$$\mathbb{E}\left\{q(X_1, y)\right\} = \mathbb{E}\left[h_0(Z, y)\mathbb{E}\left\{h_0(Z, X_1) \mid Z\right\}\right] = 0.$$

Also

$$\mathbb{E}q(X_1, X_1) = \mathbb{E}h_0^2(X_1, X_2) = \tau^2.$$

Recall that

$$Y_i = \frac{1}{i-1}\sum_{j<i} h_0(X_i, X_j), \qquad v_i = \mathbb{E}\left(Y_i^2 \mid \mathcal{F}_{i-1}\right).$$

Since X_i is independent of $\mathcal{F}_{i-1}$ and the kernel is canonical,

$$\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1}\right) = \frac{1}{i-1}\sum_{j<i} \mathbb{E}\left\{h_0(X_i, X_j) \mid \mathcal{F}_{i-1}\right\} = \frac{1}{i-1}\sum_{j<i} \mathbb{E}\left\{h_0(X_i, X_j) \mid X_j\right\} = 0.$$

Thus $(Y_i, \mathcal{F}_i)$ is a martingale-difference sequence and also

$$\langle M\rangle_n = \sum_{i=2}^{n} \mathbb{E}\left(Y_i^2 \mid \mathcal{F}_{i-1}\right) = \sum_{i=2}^{n} \frac{1}{(i-1)^2} \sum_{1\le j,k<i} q(X_j, X_k) = \sum_{j,k=1}^{n-1} a_{jk}^{(n)} q(X_j, X_k),$$

where $a_{jk}^{(n)} = \sum_{r=\max(j,k)}^{n-1} r^{-2}$. By the canonicity of q,

$$\mathbb{E}\langle M\rangle_n = \tau^2 \sum_{j=1}^{n-1} a_{jj}^{(n)} = \tau^2 \sum_{r=1}^{n-1} \frac{1}{r} = \tau^2 T_n.$$

Write

$$\langle M\rangle_n = \sum_{j=1}^{n-1} a_{jj}^{(n)} q(X_j, X_j) + 2 \sum_{1\le j<k\le n-1} a_{jk}^{(n)} q(X_j, X_k) =: A_n + B_n.$$

For the diagonal part, $a_{jj}^{(n)} \lesssim j^{-1}$ and $\mathbb{E}q^2(X_1, X_1) < \infty$, so $\mathrm{Var}(A_n) \lesssim \sum_{j=1}^{\infty} j^{-2} = O(1)$. For the off-diagonal part, the canonicity of q gives the needed orthogonality, and

$$\mathrm{Var}(B_n) \lesssim \sum_{1\le j<k\le n-1} \left\{a_{jk}^{(n)}\right\}^2 \lesssim \sum_{k=2}^{n-1}\sum_{j=1}^{k-1} \frac{1}{k^2} = O(T_n).$$

For $j < k$,

$$\mathbb{E}\left\{q(X_j, X_j) q(X_j, X_k)\right\} = \mathbb{E}\left[q(X_j, X_j)\mathbb{E}\left\{q(X_j, X_k) \mid X_j\right\}\right] = 0$$

by canonicity of q. The same conditioning argument eliminates the other diagonal–off-diagonal cross terms, so

$$\mathrm{Cov}(A_n, B_n) = 0.$$

Consequently, with $\Delta_n = \langle M\rangle_n - \tau^2 T_n$,

$$\mathbb{E}\Delta_n = 0, \qquad \mathrm{Var}(\Delta_n) = \mathrm{Var}(A_n) + \mathrm{Var}(B_n) = O(T_n).$$

Choose a sparse grid $n_k = \lfloor \exp(k^p) \rfloor$ with $p > 2$. Since $T_n \sim \log n$, $T_{n_k} \asymp k^p$ and $\log T_{n_k} \asymp \log k$. Chebyshev's inequality and Borel–Cantelli give

$$\Delta_{n_k} = o_{a.s.}\left(\frac{T_{n_k}}{\log T_{n_k}}\right).$$

If $n_k \le n \le n_{k+1}$, then monotonicity gives

$$\Delta_{n_k} - \tau^2\left(T_n - T_{n_k}\right) \le \Delta_n \le \Delta_{n_{k+1}} + \tau^2\left(T_{n_{k+1}} - T_n\right).$$

After division by $T_n/\log T_n$, the grid errors vanish by the preceding display and the deterministic increments vanish because

$$T_{n_{k+1}} - T_{n_k} = O(k^{p-1}) = o\left(\frac{k^p}{\log k}\right) = o\left(\frac{T_{n_k}}{\log T_{n_k}}\right).$$

It follows that

$$\langle M\rangle_n - \tau^2 T_n = o_{a.s.}\left(\frac{T_n}{\log T_n}\right), \qquad \frac{\langle M\rangle_n}{T_n} \xrightarrow{\text{a.s.}} \tau^2.$$

For the optional quadratic variation, define

$$D_i = Y_i^2 - \mathbb{E}\left(Y_i^2 \mid \mathcal{F}_{i-1}\right), \qquad i \ge 2.$$

Then $(D_i, \mathcal{F}_i)$ is a martingale-difference sequence. Conditionally on X_i, the variables $h_0(X_i, X_j)$, $1 \le j < i$, are independent and centered. The fourth-moment formula for sums of independent centered random variables and Jensen's inequality give $\mathbb{E}Y_i^4 = O(i^{-2})$. Therefore $\sum_{i=2}^{\infty} \mathbb{E}D_i^2 < \infty$, so $\sum_{i=2}^{n} D_i$ converges almost surely and in L^2. Hence

$$[M]_n - \langle M \rangle_n = O_{a.s.}(1) = o_{a.s.}\left(\frac{T_n}{\log T_n}\right),$$

and

$$\frac{[M]_n}{T_n} \xrightarrow{\text{a.s.}} \tau^2.$$

Finally, since $M_n = \sum_{i=2}^{n} Y_i = (n-1)(\widetilde{U}_n - \theta)$,

$$\sum_{i=2}^{n} \left\{ \frac{1}{i-1} \sum_{j<i} h(X_i, X_j) - \widetilde{U}_n \right\}^2 = \sum_{i=2}^{n} \left(Y_i - \frac{M_n}{n-1} \right)^2 = [M]_n - \frac{M_n^2}{n-1}.$$

Thus

$$\widehat{\tau}_n^2 = \frac{[M]_n}{T_n} - \frac{M_n^2}{(n-1)T_n}.$$

Since

$$\sum_{i=2}^{\infty} \mathbb{E}\left(\frac{Y_i}{\sqrt{i}}\right)^2 = \tau^2 \sum_{i=2}^{\infty} \frac{1}{i(i-1)} < \infty,$$

the martingale $L_n = \sum_{i=2}^{n} Y_i/\sqrt{i}$ is L^2-bounded. The martingale convergence theorem gives $L_n \to L_\infty$ almost surely. Kronecker's lemma, applied with the increasing weights $\sqrt{i}$, yields

$$\frac{1}{\sqrt{n}} \sum_{i=2}^{n} Y_i = \frac{M_n}{\sqrt{n}} \to 0 \qquad \text{a.s.}$$

Hence

$$\frac{M_n^2}{(n-1)T_n} = o_{a.s.}(1),$$

and $\widehat{\tau}_n^2 \xrightarrow{\text{a.s.}} \tau^2$. □

C.2 Strong approximation

The following embedding is given in Hall & Heyde (1980, Appendix I).

Lemma 1 (Martingale Skorokhod embedding). *Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ be a square-integrable martingale with respect to $\{\mathcal{F}_n\}$. There exist a standard Brownian motion W, an increasing sequence of embedding times $0 = R_0 \le R_1 \le \cdots$, and a filtration $(\mathcal{G}_n)$ such that $S'_n = W(R_n)$ is a martingale with respect to $(\mathcal{G}_n)$ and $(S'_n)_{n\ge 0}$ has the same law as $(S_n)_{n\ge 0}$. Writing $\rho_i = R_i - R_{i-1}$ and $X'_i = S'_i - S'_{i-1}$, we have*

$$\mathbb{E}(\rho_i \mid \mathcal{G}_{i-1}) = \mathbb{E}\big((X'_i)^2 \mid \mathcal{G}_{i-1}\big), \qquad \mathbb{E}(\rho_i^2 \mid \mathcal{G}_{i-1}) \le C\,\mathbb{E}\big(|X'_i|^4 \mid \mathcal{G}_{i-1}\big)$$

for a universal constant C.

Lemma 2 (Random-time consistency). *Under the canonical condition and $E|h_0(X_1, X_2)|^4 < \infty$, the embedding times for $M_n = \sum_{i=2}^{n} Y_i$ satisfy*

$$\big|R_n - \tau^2 T_n\big| = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

In particular, $R_n/T_n \xrightarrow{\text{a.s.}} \tau^2$.

Proof. Apply Lemma 1 to the martingale differences Y_i. Let $M'_n = B(R_n)$ be the embedded martingale copy, where $R_n = \sum_{i=2}^n \rho_i$, and set $Y'_i = M'_i - M'_{i-1}$. In the embedded filtration $(\mathcal{G}_i)$, define

$$v'_i = \mathbb{E}\left\{(Y'_i)^2 \mid \mathcal{G}_{i-1}\right\}, \qquad \langle M' \rangle_n = \sum_{i=2}^n v'_i.$$

The equality in law of (Y'_i) and (Y_i) gives

$$\sum_{i=2}^\infty \mathbb{E}\left\{(Y'_i)^4\right\} = \sum_{i=2}^\infty \mathbb{E}Y_i^4 < \infty.$$

Write $D'_i = (Y'_i)^2 - v'_i$. Then $(D'_i, \mathcal{G}_i)$ is a martingale-difference sequence and

$$\sum_{i=2}^\infty \mathbb{E}(D'_i)^2 \le \sum_{i=2}^\infty \mathbb{E}\left\{(Y'_i)^4\right\} < \infty.$$

Thus

$$[M']_n - \langle M' \rangle_n = \sum_{i=2}^n D'_i = O_{a.s.}(1).$$

The proof of Theorem 1 establishes

$$[M]_n - \tau^2 T_n = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

Because (Y'_i) and (Y_i) have the same finite-dimensional distributions, this optional-quadratic-variation rate transfers to the embedded copy:

$$[M']_n - \tau^2 T_n = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

Consequently,

$$\langle M' \rangle_n - \tau^2 T_n = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

Moreover,

$$\mathbb{E}\left(\rho_i \mid \mathcal{G}_{i-1}\right) = v'_i,$$

so $\sum_{i=2}^n (\rho_i - v'_i)$ is a martingale. The conditional second-moment bound in the embedding lemma gives

$$\sum_{i=2}^\infty \mathbb{E}(\rho_i - v'_i)^2 \le C \sum_{i=2}^\infty \mathbb{E}(Y'_i)^4 < \infty.$$

Hence

$$R_n - \langle M' \rangle_n = O_{a.s.}(1).$$

Combining the last two displays gives the asserted rate. □

Proof of Theorem 2. By the Hall–Heyde Skorokhod embedding applied to the martingale differences Y_i, we may work on an enlarged probability space on which $M_n = B(R_n)$ where B is a standard Brownian motion and R_n is the embedding clock. By Lemma 2,

$$|R_n - \tau^2 T_n| = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

Let

$$\omega_B(T, h) = \sup_{\substack{0 \le s,t \le T \\ |s-t| \le h}} |B(s) - B(t)|.$$

The Brownian modulus of continuity gives

$$\omega_B(T, h) = O_{a.s.}\left(\sqrt{h\left\{\log(T/h) + \log\log T\right\}}\right)$$

uniformly as $T \to \infty$, for $0 < h \le T/2$. Since $R_n/(\tau^2 T_n) \to 1$ almost surely, with $h_n = |R_n - \tau^2 T_n|$, choose a deterministic constant $C > \tau^2$. Eventually,

$$R_n \vee (\tau^2 T_n) \le C T_n, \qquad h_n = o_{a.s.}\left(\frac{T_n}{\log T_n}\right).$$

Thus $h_n \le C T_n/2$ eventually, and the modulus bound, applied with $T = C T_n$, gives

$$h_n \left\{\log(C T_n/h_n) + \log\log(C T_n)\right\} = o_{a.s.}(T_n).$$

Here the display is interpreted trivially when $h_n = 0$. Therefore

$$|B(R_n) - B(\tau^2 T_n)| = o_{a.s.}(\sqrt{T_n}).$$

Defining $W(t) = \tau^{-1} B(\tau^2 t)$, we obtain a standard Brownian motion W. This is the standard martingale strong-approximation argument obtained from Skorokhod embedding plus Brownian modulus of continuity; see Hall & Heyde (1980, Appendix I). Hence

$$M_n = \tau W(T_n) + o_{a.s.}(\sqrt{T_n}).$$

Division by $n - 1$ gives the equivalent display for $\widetilde{U}_n - \theta$. □

C.3 Time-uniform coverage

Lemma 3 (Scalar Brownian calibration)**.** *Both (1) and (2) are two-sided $(1-\alpha)$ Brownian motion boundaries in the sense of Definition 3, for every $\alpha \in (0,1)$ and $t_0 > 0$. The LIL formula permits every fixed $\eta > 1$ and $s > 1$.*

Proof. For GM, let $c = g^{-1}(\alpha)$. On $t \ge t_0$, the process $L_t = t^{-1/2} \exp\left\{W(t)^2/2t\right\}$ is a nonnegative martingale conditional on $W(t_0)$.

A closed crossing of the GM boundary is equivalent to $L_t \ge t_0^{-1/2} e^{c^2/2}$ for some $t \ge t_0$. With $Z = W(t_0)/\sqrt{t_0}$, the conditional maximal inequality and then Gaussian integration give

$$\begin{aligned}\mathbb{P}\left\{\exists t \ge t_0 : |W(t)| \ge u^{\mathrm{GM}}_{\alpha,t_0}(t)\right\} &= \mathbb{E}\left[\mathbb{P}\left\{\sup_{t\ge t_0} L_t \ge t_0^{-1/2} e^{c^2/2} \mid W(t_0)\right\}\right] \\ &\le \mathbb{E}\min\left\{1, e^{(Z^2-c^2)/2}\right\} = 2\{1 - \Phi(c) + c\phi(c)\} = \alpha.\end{aligned}$$

This is the classical mixture-martingale boundary-crossing argument of Robbins & Siegmund (1970).

For LIL, Brownian scaling reduces the proof to $t_0 = 1$. For $k \ge 0$, set $A_k = \log\left\{2\zeta(s)(k+1)^s/\alpha\right\}$ and $\lambda_k = \sqrt{2A_k/\eta^{k+1/2}}$. Consider the exponential martingale $\exp\left\{\lambda_k W(t) - \lambda_k^2 t/2\right\}$ and let $L_k(t) := A_k/\lambda_k + \lambda_k t/2$, then by Ville's inequality,

$$\mathbb{P}\left\{\exists t \ge 0 : W(t) \ge L_k(t)\right\} \le \mathbb{P}\left\{\sup_{t\ge 0} \exp\left(\lambda_k W(t) - \frac{\lambda_k^2 t}{2}\right) \ge e^{A_k}\right\} \le e^{-A_k}.$$

For $t \in [\eta^k, \eta^{k+1}]$,

$$\frac{L_k(t)}{\sqrt{A_k t}} = \frac{\sqrt{\eta^{k+1/2}/t} + \sqrt{t/\eta^{k+1/2}}}{\sqrt{2}} \le \frac{\eta^{1/4} + \eta^{-1/4}}{\sqrt{2}}.$$

Moreover, the expression in square brackets in (2) is at least A_k on this epoch. Indeed, since $t \ge \eta^k$,

$$\begin{aligned}s\log\log\{\max(\eta t, e)\} + \log\frac{2\zeta(s)}{\alpha(\log\eta)^s} &\ge s\log\{(k+1)\log\eta\} + \log\frac{2\zeta(s)}{\alpha(\log\eta)^s} \\ &= \log\left\{\frac{2\zeta(s)(k+1)^s}{\alpha}\right\} = A_k.\end{aligned}$$

Thus, on the k-th epoch, $u^{\mathrm{LIL}}_{\alpha,1}(t) \geq L_k(t)$. Therefore a crossing of the upper LIL boundary during this epoch implies a crossing of L_k, and

$$\mathbb{P}\left\{\exists t \in [\eta^k, \eta^{k+1}] : W(t) \geq u^{\mathrm{LIL}}_{\alpha,1}(t)\right\} \leq e^{-A_k}.$$

Summing over $k \geq 0$,

$$\mathbb{P}\left\{\exists t \geq 1 : W(t) \geq u^{\mathrm{LIL}}_{\alpha,1}(t)\right\} \leq \sum_{k=0}^{\infty} e^{-A_k} = \frac{\alpha}{2\zeta(s)} \sum_{k=0}^{\infty} \frac{1}{(k+1)^s} = \frac{\alpha}{2}.$$

By symmetry, the same bound holds for crossing the lower boundary. Hence

$$\mathbb{P}\left\{\exists t \geq 1 : |W(t)| \geq u^{\mathrm{LIL}}_{\alpha,1}(t)\right\} \leq \alpha.$$

Brownian scaling gives the same conclusion for every $t_0 > 0$. □

Proof of Theorem 3. By Theorem 2, we have $M_n = \tau W(T_n) + r_n = \tau W(T_n) + o_{a.s.}(\sqrt{T_n})$. For either boundary, $u_{\alpha,t_0}(t_0 r) = \sqrt{t_0}\, u_{\alpha,1}(r)$ for $r \geq 1$ and $c_0 := \inf_{r\geq 1} u_{\alpha,1}(r)/\sqrt{r} > 0$. Hence

$$u_{\alpha,T_m}(T_n) \geq c_0 \sqrt{T_n}, \qquad n \geq m.$$

Define

$$\Delta_m = \sup_{n\geq m} \left[\frac{|r_n|}{\tau u_{\alpha,T_m}(T_n)} + \left| \frac{\widehat{\tau}_n}{\tau} - 1 \right| \right].$$

Since $\widehat{\tau}_n/\tau \xrightarrow{\text{a.s.}} 1$, we have

$$\Delta_m \leq \frac{1}{\tau c_0} \sup_{n\geq m} \frac{|r_n|}{\sqrt{T_n}} + \sup_{n\geq m} \left| \frac{\widehat{\tau}_n}{\tau} - 1 \right| \xrightarrow{\text{a.s.}} 0.$$

Consider the feasible and ideal crossing events $E_m = \{\exists n \geq m : |M_n| > \widehat{\tau}_n u_{\alpha,T_m}(T_n)\}$ and $A_m = \{\exists t \geq T_m : |W(t)| \geq u_{\alpha,T_m}(t)\}$. By Lemma 3, $\mathbb{P}(A_m) \leq \alpha$ for every m.

If E_m occurs, then for some $n \geq m$,

$$\begin{aligned} |W(T_n)| &\geq \frac{|M_n| - |r_n|}{\tau} > \frac{\widehat{\tau}_n}{\tau} u_{\alpha,T_m}(T_n) - \frac{|r_n|}{\tau} \\ &\geq (1 - \Delta_m) u_{\alpha,T_m}(T_n). \end{aligned}$$

Thus, for any fixed $\delta \in (0,1)$,

$$\mathbb{P}(E_m \setminus A_m) \leq \mathbb{P}(\Delta_m > \delta) + \mathbb{P}\left\{A_m^c,\ \exists t \geq T_m : |W(t)| > (1-\delta) u_{\alpha,T_m}(t)\right\}.$$

By Brownian scaling and the scaling identity of the boundary, the second probability equals

$$\mathbb{P}\left\{\forall r \geq 1 : |W(r)| < u_{\alpha,1}(r),\ \exists r \geq 1 : |W(r)| > (1-\delta) u_{\alpha,1}(r)\right\}.$$

We next show that this probability vanishes as $\delta \downarrow 0$. For the Gaussian-mixture boundary, $u^{\mathrm{GM}}_{\alpha,1}(r)/\sqrt{2r \log\log r} \to \infty$, whereas for the stitched-LIL boundary, $\{u^{\mathrm{LIL}}_{\alpha,1}(r)\}^2/2r \log\log r \to s(\eta^{1/4} + \eta^{-1/4})^2/4 > 1$. Therefore, by the Brownian law of the iterated logarithm, we have

$$\limsup_{r\to\infty} \frac{|W(r)|}{u_{\alpha,1}(r)} < 1 \qquad \text{a.s.}$$

On the event $\{|W(r)| < u_{\alpha,1}(r) \text{ for all } r \geq 1\}$, the preceding limsup bound implies that there exist $\varepsilon > 0$ and $R < \infty$ such that $|W(r)|/u_{\alpha,1}(r) \leq 1 - \varepsilon$ for $r \geq R$. On the compact interval $[1, R]$, continuity of W and $u_{\alpha,1}$ gives $\sup_{1\leq r\leq R} |W(r)|/u_{\alpha,1}(r) < 1$, since the ratio is strictly below 1 at every point. Therefore on this event

$$\sup_{r\geq 1} \frac{|W(r)|}{u_{\alpha,1}(r)} < 1 \qquad \text{a.s..}$$

Consequently, as $\delta \downarrow 0$ we have

$$\mathbb{P}\left\{\forall r \geq 1 : |W(r)| < u_{\alpha,1}(r),\ \exists r \geq 1 : |W(r)| > (1-\delta) u_{\alpha,1}(r)\right\} \to 0.$$

Since $\Delta_m \xrightarrow{\text{a.s.}} 0$, it follows that for every fixed $\delta \in (0,1)$,

$$\limsup_{m\to\infty} \mathbb{P}(E_m) \le \alpha + \mathbb{P}\left\{\forall r \ge 1 : |W(r)| < u_{\alpha,1}(r),\ \exists r \ge 1 : |W(r)| > (1-\delta)u_{\alpha,1}(r)\right\}.$$

Letting $\delta \downarrow 0$ gives $\limsup_{m\to\infty} \mathbb{P}(E_m) \le \alpha$.

This is exactly

$$\liminf_{m\to\infty} \mathbb{P}\left\{\forall n \ge m : \theta \in C^{\mathrm{MG}}_{m,n}(\alpha)\right\} \ge 1-\alpha,$$

which proves the theorem. □

C.4 Asymptotic confidence sequences

Proof of Remark 1. Fix $m \ge 2$. On the coupling in Theorem 2, as $M_n = \tau W(T_n) + r_n$ where $r_n = o_{a.s.}(\sqrt{T_n})$, we have

$$A_m := \sup_{n\ge m} \frac{|r_n|}{\sqrt{T_n}} < \infty \qquad \text{a.s..}$$

Fix any $\delta \in (0,\alpha)$. Since $A_m < \infty$ almost surely, there exists a deterministic $\varepsilon > 0$ such that

$$\mathbb{P}(A_m > \varepsilon) \le \delta.$$

Put $a = \alpha - \delta$ and define

$$w^*_{m,n} = \frac{\tau u_{a,T_m}(T_n) + \varepsilon\sqrt{T_n}}{n-1}, \qquad n \ge m,$$

and $C^*_{m,n}(\alpha) = [\,\widetilde{U}_n \pm w^*_{m,n}\,]$. On the event $\{A_m \le \varepsilon\} \cap \left\{|W(T_n)| \le u_{a,T_m}(T_n) \text{ for all } n \ge m\right\}$, we have for every $n \ge m$,

$$|M_n| \le \tau u_{a,T_m}(T_n) + \varepsilon\sqrt{T_n},$$

and therefore $\theta \in C^*_{m,n}(\alpha)$ for all $n \ge m$. By Lemma 3 and a union bound,

$$\mathbb{P}\{\theta \in C^*_{m,n}(\alpha) \text{ for all } n \ge m\} \ge 1 - a - \delta = 1 - \alpha.$$

Thus $\{C^*_{m,n}(\alpha) : n \ge m\}$ is a nonasymptotic $(1-\alpha)$-confidence sequence. The constants τ and ε may depend on the underlying law and need not be available for implementation.

It remains to compare its radius with that of the feasible sequence. Recall that $w_{m,n} = \widehat{\tau}_n u_{\alpha,T_m}(T_n)/(n-1)$, we have

$$\frac{w^*_{m,n}}{w_{m,n}} = \frac{\tau}{\widehat{\tau}_n}\frac{u_{a,T_m}(T_n)}{u_{\alpha,T_m}(T_n)} + \frac{\varepsilon}{\widehat{\tau}_n}\frac{\sqrt{T_n}}{u_{\alpha,T_m}(T_n)}$$

almost surely for all sufficiently large n. By Theorem 1, $\widehat{\tau}_n \xrightarrow{\text{a.s.}} \tau > 0$. For either the Gaussian-mixture or stitched-LIL boundary, with m fixed, $u_{a,T_m}(T_n)/u_{\alpha,T_m}(T_n) \to 1$ and $\sqrt{T_n}/u_{\alpha,T_m}(T_n) \to 0$. Consequently,

$$\frac{w^*_{m,n}}{w_{m,n}} \xrightarrow{\text{a.s.}} 1, \qquad n \to \infty.$$

Hence $\{C^{\mathrm{MG}}_{m,n}(\alpha) : n \ge m\}$ satisfies Definition 2. □

C.5 One-sided sequential testing

Lemma 4 (One-sided Brownian boundaries)**.** *Let $t_0 > 0$. The boundaries in Remark 2 satisfy*

$$\mathbb{P}\left\{\exists t \ge t_0 : W(t) \ge u^+_{\alpha,t_0}(t)\right\} \le \alpha,$$

where W is a standard Brownian motion. For the LIL boundary, this holds for every $\alpha \in (0,1)$, $\eta > 1$, and $s > 1$; for the GM boundary, it holds for $0 < \alpha < g_+(0) = 1/2 + 1/\pi$.

Proof. For the GM boundary, write $c = g_+^{-1}(\alpha) > 0$. The function g_+ is continuous and strictly decreasing on $[0, \infty)$, since

$$g'_+(a) = -\phi(a)\left\{a + \frac{\phi(a)}{\Phi(a)}\right\}^2 < 0,$$

and $g_+(a) \to 0$ as $a \to \infty$. Hence $g_+(c) = \alpha$. For $t \geq t_0$, define

$$L_t^+ = t^{-1/2}\exp\left\{\frac{W(t)^2}{2t}\right\}\Phi\left(\frac{W(t)}{\sqrt{t}}\right) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \exp\left\{\lambda W(t) - \frac{\lambda^2 t}{2}\right\} d\lambda.$$

Conditionally on $W(t_0)$, this is a nonnegative martingale on $[t_0, \infty)$. A crossing of the upper GM boundary implies

$$L_t^+ \geq t_0^{-1/2} e^{c^2/2}\Phi(c)$$

at the crossing time. With $Z = W(t_0)/\sqrt{t_0} \sim N(0,1)$, the conditional maximal inequality therefore yields

$$\begin{aligned}\mathbb{P}\left\{\exists t \geq t_0 : W(t) \geq u_{\alpha,t_0}^{+,\mathrm{GM}}(t)\right\} &\leq \mathbb{E}\min\left\{1, e^{(Z^2-c^2)/2}\frac{\Phi(Z)}{\Phi(c)}\right\} \\ &= 1 - \Phi(c) + \frac{\phi(c)}{\Phi(c)}\int_{-\infty}^c \Phi(z)\, dz \\ &= 1 - \Phi(c) + c\phi(c) + \frac{\phi(c)^2}{\Phi(c)} = \alpha.\end{aligned}$$

The second line uses the strict monotonicity of $z \mapsto e^{z^2/2}\Phi(z)$, which follows from its positive-integral representation, and the third uses $\int_{-\infty}^c \Phi(z)\, dz = c\Phi(c) + \phi(c)$.

For the LIL boundary, Brownian scaling reduces the proof to $t_0 = 1$. For $k \geq 0$, put

$$A_k = \log\{\zeta(s)(k+1)^s/\alpha\}, \qquad \lambda_k = \sqrt{2A_k/\eta^{k+1/2}}, \qquad L_k(t) = A_k/\lambda_k + \lambda_k t/2.$$

Ville's inequality for $\exp\{\lambda_k W(t) - \lambda_k^2 t/2\}$ gives

$$\mathbb{P}\{\exists t \geq 0 : W(t) \geq L_k(t)\} \leq e^{-A_k}.$$

For $t \in [\eta^k, \eta^{k+1}]$,

$$\frac{L_k(t)}{\sqrt{A_k t}} = \frac{\sqrt{\eta^{k+1/2}/t} + \sqrt{t/\eta^{k+1/2}}}{\sqrt{2}} \leq \frac{\eta^{1/4} + \eta^{-1/4}}{\sqrt{2}}.$$

Moreover,

$$s\log\log\{\max(\eta t, e)\} + \log\frac{\zeta(s)}{\alpha(\log\eta)^s} \geq \log\frac{\zeta(s)(k+1)^s}{\alpha} = A_k.$$

Thus $u_{\alpha,1}^{+,\mathrm{LIL}}(t) \geq L_k(t)$ on this epoch. Summing the epochwise crossing probabilities gives

$$\mathbb{P}\left\{\exists t \geq 1 : W(t) \geq u_{\alpha,1}^{+,\mathrm{LIL}}(t)\right\} \leq \sum_{k=0}^\infty e^{-A_k} = \frac{\alpha}{\zeta(s)}\sum_{k=0}^\infty (k+1)^{-s} = \alpha.$$

Brownian scaling gives the conclusion for arbitrary $t_0 > 0$. □

Proof of Corollary 1. Work under the fixed null distribution in the statement. By Theorem 2,

$$M_n = (n-1)(\widetilde{U}_n - \theta_0) = \tau W(T_n) + r_n, \qquad r_n = o_{a.s.}(\sqrt{T_n}),$$

and Theorem 1 gives $\widehat{\tau}_n/\tau \xrightarrow{\text{a.s.}} 1$. For either one-sided boundary,

$$u_{\alpha,t_0}^+(t_0 r) = \sqrt{t_0}\, u_{\alpha,1}^+(r), \qquad c_0 := \inf_{r \geq 1}\frac{u_{\alpha,1}^+(r)}{\sqrt{r}} > 0.$$

Consequently,

$$\Delta_m := \sup_{n \geq m} \left\{ \frac{|r_n|}{\tau u^+_{\alpha,T_m}(T_n)} + \left| \frac{\widehat{\tau}_n}{\tau} - 1 \right| \right\} \xrightarrow{\text{a.s.}} 0.$$

If the test rejects at some $n \geq m$, then $W(T_n) > (1 - \Delta_m) u^+_{\alpha,T_m}(T_n)$. Lemma 4 bounds the ideal closed-crossing probability by α. To transfer this bound, note that

$$\frac{u^{+,\mathrm{GM}}_{\alpha,1}(r)}{\sqrt{2r \log\log r}} \longrightarrow \infty, \qquad \frac{\{u^{+,\mathrm{LIL}}_{\alpha,1}(r)\}^2}{2r \log\log r} \longrightarrow \frac{s(\eta^{1/4} + \eta^{-1/4})^2}{4} > 1.$$

The Brownian LIL therefore gives $\limsup_{r\to\infty} W(r)/u^+_{\alpha,1}(r) < 1$ almost surely. On the event of no ideal closed crossing, continuity on compact intervals then implies $\sup_{r\geq 1} W(r)/u^+_{\alpha,1}(r) < 1$ almost surely. The same scaling and vanishing-perturbation argument as in the proof of Theorem 3, with W in place of $|W|$, now yields the asserted upper bound on the limiting rejection probability. □

C.6 Fixed-alternative consistency

The following result imposes the fourth-moment condition in (A1) under the alternative distribution, without assuming degeneracy.

Proposition 2 (Fixed-alternative consistency)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with distribution P, and let h be a fixed symmetric measurable kernel with $\theta = \mathbb{E}h(X_1, X_2) > \theta_0$. Suppose that $\mathbb{E}|h_0(X_1, X_2)|^4 < \infty$, where $h_0 = h - \theta$. Fix a deterministic integer $m \geq 2$, and let u^+_{α,t_0} be either one-sided boundary in Remark 2, with all parameters fixed within their stated ranges. Then, almost surely,*

$$(n-1)(\widetilde{U}_n - \theta_0) > \widehat{\tau}_n u^+_{\alpha,T_m}(T_n) \qquad \textit{for all sufficiently large } n \geq m.$$

In particular, the first rejection time is finite almost surely, and the probability of rejection by time N tends to one as $N \to \infty$.

Proof of Proposition 2. Define $\mu(x) = \int h(x,y)\, P(dy)$, $a_i = \sum_{j<i} h(X_i, X_j)/(i-1)$ and $e_i = a_i - \mu(X_i)$ for $i \geq 2$. Conditional Jensen's inequality gives $\mathbb{E}|\mu(X_1) - \theta|^4 \leq \mathbb{E}|h_0(X_1, X_2)|^4 < \infty$. Writing $\xi = h(X_1, X_2) - \mu(X_1)$, we also have

$$K_4 := \mathbb{E}|\xi|^4 \leq 16\mathbb{E}|h_0(X_1, X_2)|^4 < \infty.$$

Conditionally on X_i, the variables $h(X_i, X_j) - \mu(X_i)$, $j < i$, are independent and centered. Expanding the conditional fourth moment and then taking expectations yields

$$\mathbb{E}|e_i|^4 = \frac{K_4}{(i-1)^3} + \frac{3(i-2)}{(i-1)^3}\mathbb{E}\big[\{\mathbb{E}(\xi^2 \mid X_1)\}^2\big] \leq \frac{3K_4}{(i-1)^2},$$

where the inequality follows from conditional Jensen's inequality. Thus, for every $\varepsilon > 0$,

$$\sum_{i=2}^{\infty} \mathbb{P}(|e_i| > \varepsilon) \leq \frac{3K_4}{\varepsilon^4} \sum_{i=2}^{\infty} \frac{1}{(i-1)^2} < \infty.$$

The Borel–Cantelli lemma implies $e_i \xrightarrow{\text{a.s.}} 0$; independence between rows is not needed. The strong law for $\mu(X_i)$ and Cesàro averaging therefore give

$$\widetilde{U}_n = \frac{1}{n-1}\sum_{i=2}^{n} \mu(X_i) + \frac{1}{n-1}\sum_{i=2}^{n} e_i \xrightarrow{\text{a.s.}} \theta.$$

To control the scale estimator, note that $\sum_{i=2}^n e_i^2/(n-1) \xrightarrow{\text{a.s.}} 0$ and $\sum_{i=2}^n \mu(X_i)^2/(n-1) \xrightarrow{\text{a.s.}} \mathbb{E}\{\mu(X_1)^2\} < \infty$ by Cesàro averaging and the strong law, respectively. Since $a_i = \mu(X_i) + e_i$, the Cauchy–Schwarz inequality yields

$$\frac{1}{n-1}\sum_{i=2}^{n} |a_i^2 - \mu(X_i)^2| \leq 2\left\{\frac{1}{n-1}\sum_{i=2}^{n}\mu(X_i)^2\right\}^{1/2}\left\{\frac{1}{n-1}\sum_{i=2}^{n} e_i^2\right\}^{1/2} + \frac{1}{n-1}\sum_{i=2}^{n} e_i^2 \xrightarrow{\text{a.s.}} 0.$$

Consequently,

$$\frac{T_n}{n-1}\widehat{\tau}_n^2 = \frac{1}{n-1}\sum_{i=2}^{n}(a_i - \widetilde{U}_n)^2 = \frac{1}{n-1}\sum_{i=2}^{n} a_i^2 - \widetilde{U}_n^2 \xrightarrow{\text{a.s.}} \mathrm{Var}\{\mu(X_1)\}.$$

In particular, $\widehat{\tau}_n^2 = O_{\text{a.s.}}(n/T_n)$, including when $\mathrm{Var}\{\mu(X_1)\} = 0$.

For fixed m and fixed boundary parameters, the formulas in Remark 2 imply

$$\frac{\{u^+_{\alpha,T_m}(T_n)\}^2}{T_n} = \begin{cases} O(\log T_n), & \text{GM}, \\ O(\log\log T_n), & \text{LIL}. \end{cases}$$

Since $T_n \sim \log n$, it follows that

$$\begin{aligned} \frac{\widehat{\tau}_n u^+_{\alpha,T_m}(T_n)}{n-1} &= \left(\frac{T_n\widehat{\tau}_n^2}{n-1}\right)^{1/2} \frac{u^+_{\alpha,T_m}(T_n)}{\sqrt{(n-1)T_n}} \\ &= \begin{cases} O_{\text{a.s.}}\left(\sqrt{\dfrac{\log\log n}{n}}\right), & \text{GM}, \\ O_{\text{a.s.}}\left(\sqrt{\dfrac{\log\log\log n}{n}}\right), & \text{LIL}, \end{cases} \xrightarrow{\text{a.s.}} 0. \end{aligned}$$

Combining this with $\widetilde{U}_n \xrightarrow{\text{a.s.}} \theta$ gives

$$\widetilde{U}_n - \theta_0 - \frac{\widehat{\tau}_n u^+_{\alpha,T_m}(T_n)}{n-1} \xrightarrow{\text{a.s.}} \theta - \theta_0 > 0.$$

Hence the rejection inequality holds at every sufficiently large monitoring time, proving the assertion. The argument also covers times at which $\widehat{\tau}_n = 0$, since it never divides by the estimated scale. □

D Proofs of vector results

D.1 Covariance clock

Lemma 5 (Vector covariance clock and plug-in consistency)**.** *Under (A2),*

$$\|\langle \boldsymbol{M}\rangle_n - T_n\boldsymbol{\Omega}\|_{\text{op}} = O_{a.s.}(T_n^{3/4}) = o_{a.s.}\left(\frac{T_n}{\log T_n}\right),$$

where $\langle \boldsymbol{M}\rangle_n := \sum_{i=2}^{n}\mathbb{E}\left(\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathsf{T}} \mid \mathcal{F}_{i-1}\right)$ *and hence* $\langle \boldsymbol{M}\rangle_n/T_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$ *in the operator norm. Moreover, the plug-in estimator is strongly consistent* $\widehat{\boldsymbol{\Omega}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$*.*

Proof. Conditionally on X_i, the vectors $\boldsymbol{h}_0(X_i, X_j)$, $j < i$, are independent. They are also conditionally centered, since symmetry and canonicity give $\mathbb{E}\{\boldsymbol{h}_0(X_i, X_j) \mid X_i\} = \mathbf{0}$. Thus, for P-a.e. fixed x, applying Rosenthal inequality with exponent four coordinatewise and using $\|\boldsymbol{u}\|_2^4 \le d\sum_{r=1}^{d}|u_r|^4$, we obtain

$$\begin{aligned} \mathbb{E}\left\|\sum_{j<i}\boldsymbol{h}_0(x, X_j)\right\|_2^4 \le C_d\Big[&(i-1)\mathbb{E}\,\|\boldsymbol{h}_0(x, X_1)\|_2^4 \\ &+ (i-1)^2\{\mathbb{E}\,\|\boldsymbol{h}_0(x, X_1)\|_2^2\}^2\Big]. \end{aligned}$$

By Jensen's inequality, $\{\mathbb{E}\,\|\boldsymbol{h}_0(x, X_1)\|_2^2\}^2 \le \mathbb{E}\,\|\boldsymbol{h}_0(x, X_1)\|_2^4$. Evaluating the preceding bound at $x = X_i$ and using independence of X_i and $(X_1, \ldots, X_{i-1})$ gives

$$\mathbb{E}\left\{\left\|\sum_{j<i}\boldsymbol{h}_0(X_i, X_j)\right\|_2^4 \;\middle|\; X_i\right\} \le C_d\{(i-1) + (i-1)^2\}\mathbb{E}\left(\|\boldsymbol{h}_0(X_i, X_1)\|_2^4 \mid X_i\right).$$

Taking expectations gives $\mathbb{E}\,\|\boldsymbol{Y}_i\|_2^4 \le C(i-1)^{-2}$, so $\sum_{i=2}^{\infty}\mathbb{E}\,\|\boldsymbol{Y}_i\|_2^4 < \infty$.

Moreover, by Tonelli's theorem,

$$\mathbb{E}\left\{\sum_{i=2}^{\infty}\mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1})\right\} = \sum_{i=2}^{\infty}\mathbb{E}\left\{\mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1})\right\} = \sum_{i=2}^{\infty}\mathbb{E}\,\|\boldsymbol{Y}_i\|_2^4 < \infty.$$

Thus $\sum_{i=2}^{\infty}\mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1}) < \infty$ almost surely.

We next establish the covariance-clock bound. Fix $\boldsymbol{a} \in \mathbb{R}^d$, let $h_{\boldsymbol{a}}(x,y) = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{h}_0(x,y)$, and define $q_{\boldsymbol{a}}(x,y) = \mathbb{E}\{h_{\boldsymbol{a}}(Z,x)h_{\boldsymbol{a}}(Z,y)\}$ where $Z \sim P$ is independent. The kernel $q_{\boldsymbol{a}}$ is symmetric. It is also canonical, as for P-a.e. y,

$$\mathbb{E}q_{\boldsymbol{a}}(X_1,y) = \mathbb{E}\,[h_{\boldsymbol{a}}(Z,y)\mathbb{E}\{h_{\boldsymbol{a}}(Z,X_1) \mid Z\}] = 0.$$

Also, we have

$$\mathbb{E}q_{\boldsymbol{a}}(X_1,X_1) = \mathbb{E}h_{\boldsymbol{a}}(X_1,X_2)^2 = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{\Omega}\boldsymbol{a}.$$

Assumption (A2) ensures that both the diagonal and off-diagonal second moments of $q_{\boldsymbol{a}}$ are finite. Indeed, conditional Jensen's inequality gives

$$\mathbb{E}q_{\boldsymbol{a}}(X_1,X_1)^2 \le \mathbb{E}h_{\boldsymbol{a}}(Z,X_1)^4 \le \|\boldsymbol{a}\|_2^4\,\mathbb{E}\,\|\boldsymbol{h}_0(X_1,X_2)\|_2^4 < \infty.$$

Similarly, applying Jensen's inequality first with respect to Z, we obtain

$$\begin{aligned}\mathbb{E}q_{\boldsymbol{a}}(X_1,X_2)^2 &\le \mathbb{E}\{h_{\boldsymbol{a}}(Z,X_1)^2h_{\boldsymbol{a}}(Z,X_2)^2\}\\ &= \mathbb{E}\left(\left[\mathbb{E}\{h_{\boldsymbol{a}}(Z,X_1)^2 \mid Z\}\right]^2\right)\\ &\le \mathbb{E}h_{\boldsymbol{a}}(Z,X_1)^4 \le \|\boldsymbol{a}\|_2^4\,\mathbb{E}\,\|\boldsymbol{h}_0(X_1,X_2)\|_2^4 < \infty.\end{aligned}$$

Write $V_{\boldsymbol{a},n} = \boldsymbol{a}^{\mathsf{T}}\langle\boldsymbol{M}\rangle_n\boldsymbol{a}$ and $\mu_{\boldsymbol{a}} = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{\Omega}\boldsymbol{a}$. Then

$$\begin{aligned}V_{\boldsymbol{a},n} &= \sum_{i=2}^{n}\mathbb{E}\left\{(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{Y}_i)^2 \mid \mathcal{F}_{i-1}\right\}\\ &= \sum_{i=2}^{n}\frac{1}{(i-1)^2}\sum_{1\le j,k<i}q_{\boldsymbol{a}}(X_j,X_k)\\ &= \sum_{j,k=1}^{n-1}w_{jk}^{(n)}q_{\boldsymbol{a}}(X_j,X_k).\end{aligned}$$

where $w_{jk}^{(n)} = \sum_{r=\max(j,k)}^{n-1}1/r^2$. Only the diagonal terms contribute to the expectation, so $\mathbb{E}V_{\boldsymbol{a},n} = \mu_{\boldsymbol{a}}T_n$. Moreover,

$$V_{\boldsymbol{a},n} - \mu_{\boldsymbol{a}}T_n = \sum_{j=1}^{n-1}w_{jj}^{(n)}\{q_{\boldsymbol{a}}(X_j,X_j) - \mu_{\boldsymbol{a}}\} + 2\sum_{1\le j<k\le n-1}w_{jk}^{(n)}q_{\boldsymbol{a}}(X_j,X_k).$$

Consider two distinct unordered pairs $\{j,k\} \ne \{r,s\}$, with $j<k$ and $r<s$. If the two pairs are disjoint, the independence of X_i's and $\mathbb{E}q_{\boldsymbol{a}}(X_j,X_k) = 0$ imply $\mathbb{E}\{q_{\boldsymbol{a}}(X_j,X_k)q_{\boldsymbol{a}}(X_r,X_s)\} = 0$; If the pairs share exactly one index, for example $j = r$ and $k \ne s$, then

$$\mathbb{E}\{q_{\boldsymbol{a}}(X_j,X_k)q_{\boldsymbol{a}}(X_j,X_s)\} = \mathbb{E}\,[\mathbb{E}\{q_{\boldsymbol{a}}(X_j,X_k) \mid X_j\}\mathbb{E}\{q_{\boldsymbol{a}}(X_j,X_s) \mid X_j\}] = 0.$$

Thus distinct off-diagonal terms are orthogonal in L^2.

For $j \ne l$, by canonicity

$$\mathbb{E}\,[\{q_{\boldsymbol{a}}(X_j,X_j) - \mu_{\boldsymbol{a}}\}q_{\boldsymbol{a}}(X_j,X_l)] = \mathbb{E}\,[\{q_{\boldsymbol{a}}(X_j,X_j) - \mu_{\boldsymbol{a}}\}\mathbb{E}\{q_{\boldsymbol{a}}(X_j,X_l) \mid X_j\}] = 0.$$

Consequently,

$$\mathrm{Var}(V_{\boldsymbol{a},n}) = \sum_{j=1}^{n-1}\{w_{jj}^{(n)}\}^2\,\mathrm{Var}\{q_{\boldsymbol{a}}(X_j,X_j)\} + 4\sum_{1\le j<k\le n-1}\{w_{jk}^{(n)}\}^2\mathbb{E}q_{\boldsymbol{a}}(X_j,X_k)^2.$$

Since $w_{jj}^{(n)} \le C/j$ and $w_{jk}^{(n)} \le C/k$ for $j < k$, it follows that there exists constant $C_{\boldsymbol{a}}$ such that

$$\mathrm{Var}(V_{\boldsymbol{a},n}) \le C_{\boldsymbol{a}} \sum_{j=1}^{n-1} \frac{1}{j^2} + C_{\boldsymbol{a}} \sum_{k=2}^{n-1} \frac{k-1}{k^2} \le C_{\boldsymbol{a}} T_n.$$

Let $n_\ell = \lfloor \exp(\ell^3) \rfloor$. Since $T_{n_\ell} \asymp \ell^3$, Chebyshev's inequality gives

$$\sum_{\ell=1}^{\infty} \mathbb{P}\left(|V_{\boldsymbol{a},n_\ell} - \mu_{\boldsymbol{a}} T_{n_\ell}| > T_{n_\ell}^{3/4} \right) < \infty.$$

The Borel–Cantelli lemma therefore yields $V_{\boldsymbol{a},n_\ell} - \mu_{\boldsymbol{a}} T_{n_\ell} = O_{a.s.}(T_{n_\ell}^{3/4})$.

For $n_\ell \le n \le n_{\ell+1}$, monotonicity of $V_{\boldsymbol{a},n}$ and T_n gives

$$V_{\boldsymbol{a},n} - \mu_{\boldsymbol{a}} T_n \le V_{\boldsymbol{a},n_{\ell+1}} - \mu_{\boldsymbol{a}} T_{n_{\ell+1}} + \mu_{\boldsymbol{a}} (T_{n_{\ell+1}} - T_n)$$

and

$$V_{\boldsymbol{a},n} - \mu_{\boldsymbol{a}} T_n \ge V_{\boldsymbol{a},n_\ell} - \mu_{\boldsymbol{a}} T_{n_\ell} - \mu_{\boldsymbol{a}} (T_n - T_{n_\ell}).$$

As $T_{n_{\ell+1}} - T_{n_\ell} = O(\ell^2) = o(T_{n_\ell}^{3/4})$, it follows that $V_{\boldsymbol{a},n} - \mu_{\boldsymbol{a}} T_n = O_{a.s.}(T_n^{3/4})$ for every n.

Apply this result to the finite collection $\{\boldsymbol{e}_r : 1 \le r \le d\} \cup \{\boldsymbol{e}_r + \boldsymbol{e}_s : 1 \le r < s \le d\}$. Polarization shows that every element of $\langle \boldsymbol{M} \rangle_n - T_n \boldsymbol{\Omega}$ is $O_{a.s.}(T_n^{3/4})$. Since d is fixed, we have

$$\| \langle \boldsymbol{M} \rangle_n - T_n \boldsymbol{\Omega} \|_{\mathrm{op}} = O_{a.s.}(T_n^{3/4}),$$

which also implies

$$\| \langle \boldsymbol{M} \rangle_n - T_n \boldsymbol{\Omega} \|_{\mathrm{op}} = o_{a.s.}\left(\frac{T_n}{\log T_n} \right),$$

and division by T_n proves $T_n^{-1} \langle \boldsymbol{M} \rangle_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$ in operator norm.

The scalar optional quadratic-variation result applied to $\boldsymbol{e}_r^\top \boldsymbol{M}_n$ and $(\boldsymbol{e}_r + \boldsymbol{e}_s)^\top \boldsymbol{M}_n$, followed by the same finite polarization argument, gives $\sum_{i=2}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^\top / T_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$ in operator norm. Recall that the estimator is

$$\widehat{\boldsymbol{\Omega}}_n = \frac{1}{T_n} \sum_{i=2}^{n} \boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \frac{\boldsymbol{M}_n \boldsymbol{M}_n^\top}{(n-1) T_n}.$$

For every coordinate r, the scalar ASIP and the Brownian LIL give $M_{n,r}^2 = O_{a.s.}\{T_n \log\log T_n\}$. Since d is fixed, $\|\boldsymbol{M}_n\|_2^2 = O_{a.s.}\{T_n \log\log T_n\}$, and hence

$$\frac{\left\| \boldsymbol{M}_n \boldsymbol{M}_n^\top \right\|_{\mathrm{op}}}{(n-1) T_n} = O_{a.s.}\left(\frac{\log\log T_n}{n} \right) \longrightarrow 0.$$

The plug-in consistency now follows from the optional quadratic-variation convergence. □

D.2 Strong approximation

Lemma 6 (Theorem 1.1 of Zhang (2004)). *Let $\{(\boldsymbol{\xi}_i, \mathcal{G}_i) : i \ge 1\}$ be a square-integrable $\mathbb{R}^d$-valued martingale-difference sequence, and denote $\boldsymbol{\Sigma}_n = \sum_{i=1}^n \mathbb{E}(\boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top \mid \mathcal{G}_{i-1})$ and $V_n = \mathrm{tr}(\boldsymbol{\Sigma}_n)$. Let f be a nondecreasing positive function tending to infinity such that $f(t)(\log t)^{\alpha_0}/t$ is nonincreasing for some $\alpha_0 > 50d$, and $f(t)/t^{\varepsilon_0}$ is nondecreasing for some $\varepsilon_0 \in (0,1)$. Suppose that $V_n \to \infty$ a.s. and*

$$\sum_{n \ge 1} \mathbb{E}\left[\|\boldsymbol{\xi}_n\|_2^2 \, I\{\|\boldsymbol{\xi}_n\|_2^2 \ge f(V_n)\}/f(V_n) \mid \mathcal{G}_{n-1} \right] < \infty \qquad \text{a.s.},$$

Moreover, assume that there exists some covariance matrix $\boldsymbol{\Gamma}$, measurable with respect to $\mathcal{G}_k$ for some $k \ge 0$, such that

$$\| \boldsymbol{\Sigma}_n - \boldsymbol{\Gamma} V_n \|_{\max} = o_{a.s.}\{f(V_n)\}.$$

Then there exists a sequence $\{\boldsymbol{Z}_n : n \ge 1\}$ of i.i.d. $\mathbb{R}^d$-valued standard normal random vectors, independent of $\boldsymbol{\Gamma}$, such that

$$\sum_{m \ge 1} \boldsymbol{\xi}_m I\{V_m \le n\} - \boldsymbol{\Gamma}^{1/2} \sum_{m \le n} \boldsymbol{Z}_m = O_{a.s.}\left[n^{1/2} \{f(n)/n\}^{\frac{1}{50d}} \right].$$

Proof of Proposition 1. By the vector setup and Lemma 5, $\boldsymbol{Y}_i$ is a square-integrable $\mathbb{R}^d$-valued martingale-difference sequence. Apply Lemma 6 with $\mathcal{G}_0$ trivial, $\mathcal{G}_i = \mathcal{F}_i$ for $i \geq 1$, $\boldsymbol{\xi}_1 = \mathbf{0}$, and $\boldsymbol{\xi}_i = \boldsymbol{Y}_i$ for $i \geq 2$. Then $\boldsymbol{\Sigma}_n = \langle \boldsymbol{M} \rangle_n$ and $V_n = \operatorname{tr}(\boldsymbol{\Sigma}_n)$. Take $c = \operatorname{tr}(\boldsymbol{\Omega})$ and $\boldsymbol{\Gamma} = \boldsymbol{\Omega}/c$, then

$$\boldsymbol{\Sigma}_n - \boldsymbol{\Gamma} V_n = \{\boldsymbol{\Sigma}_n - \boldsymbol{\Omega} T_n\} - \boldsymbol{\Gamma} \operatorname{tr}\{\boldsymbol{\Sigma}_n - \boldsymbol{\Omega} T_n\}.$$

Hence, since d is fixed,

$$\begin{aligned} \|\boldsymbol{\Sigma}_n - \boldsymbol{\Gamma} V_n\|_{\max} &\leq \|\boldsymbol{\Sigma}_n - \boldsymbol{\Omega} T_n\|_{\max} + \|\boldsymbol{\Gamma}\|_{\max} |\operatorname{tr}\{\boldsymbol{\Sigma}_n - \boldsymbol{\Omega} T_n\}| \\ &\leq \{1 + d \|\boldsymbol{\Gamma}\|_{\max}\} \|\boldsymbol{\Sigma}_n - \boldsymbol{\Omega} T_n\|_{\mathrm{op}} = O_{a.s.}(T_n^{3/4}). \end{aligned}$$

Taking traces in Lemma 5 gives $V_n / T_n \to c > 0$ almost surely, and therefore

$$\|\boldsymbol{\Sigma}_n - \boldsymbol{\Gamma} V_n\|_{\max} = O_{a.s.}(V_n^{3/4}).$$

Choose $f(t) = t^{7/8}$ for all sufficiently large t. Then we obtain

$$\|\boldsymbol{\Sigma}_n - \boldsymbol{\Gamma} V_n\|_{\max} = o_{a.s.}\{f(V_n)\}.$$

For example, with $\alpha_0 = 51d$ and $\varepsilon_0 = 1/2$, the two monotonicity conditions in Lemma 6 hold for all sufficiently large t; extend f positively over the remaining bounded interval. For the conditional Lindeberg condition, $f(V_i) \geq 1$ eventually, and

$$\frac{\mathbb{E}\left[\|\boldsymbol{Y}_i\|_2^2 I\{\|\boldsymbol{Y}_i\|_2^2 \geq f(V_i)\} \mid \mathcal{F}_{i-1}\right]}{f(V_i)} \leq \frac{\mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1})}{f(V_i)^2} \leq \mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1}).$$

The series on the right is finite almost surely by Lemma 5. Thus all conditions of Lemma 6 are satisfied.

Define

$$\nu(m) = \max\big(\{1\} \cup \{r \geq 2 : V_r \leq m\}\big).$$

There exists a sequence of independent standard normal vectors $\{\boldsymbol{Z}_\ell\}$ such that

$$\|\boldsymbol{M}_{\nu(m)} - \boldsymbol{\Gamma}^{1/2} \sum_{\ell=1}^{m} \boldsymbol{Z}_\ell\|_2 = O_{a.s.}(m^{1/2-\kappa_0}), \qquad \kappa_0 = \frac{1 - 7/8}{50d} = \frac{1}{400d}.$$

Enlarging the probability space if necessary, construct a standard d-dimensional Brownian motion $\boldsymbol{B}$ satisfying $\boldsymbol{B}(m) = \sum_{\ell=1}^{m} \boldsymbol{Z}_\ell$ for every integer m. Thus

$$\|\boldsymbol{M}_{\nu(m)} - \boldsymbol{\Gamma}^{1/2} \boldsymbol{B}(m)\|_2 = O_{a.s.}(m^{1/2-\kappa_0}).$$

Write

$$v_i = V_i - V_{i-1} = \mathbb{E}(\|\boldsymbol{Y}_i\|_2^2 \mid \mathcal{F}_{i-1}).$$

Since V_{r+1} is $\mathcal{F}_r$-measurable, $\{\nu(m) \leq r\} = \{V_{r+1} > m\} \in \mathcal{F}_r$; hence $\nu(m)$ is a stopping time. Moreover, $I\{\nu(m-1) < i \leq \nu(m)\} = I\{m - 1 < V_i \leq m\}$ is $\mathcal{F}_{i-1}$-measurable.

Let $D_m = \max_{\nu(m-1) \leq r \leq \nu(m)} \left\|\boldsymbol{M}_r - \boldsymbol{M}_{\nu(m-1)}\right\|_2$, and $Q_m = V_{\nu(m)} - V_{\nu(m-1)}$. If the m-th block is nonempty, then $Q_m \leq 1 + v_{\nu(m-1)+1}$. Conditional Jensen's inequality gives $v_i^2 \leq \mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1})$, and therefore

$$\sup_m \mathbb{E} Q_m^2 \leq 2 + 2 \sum_{i=2}^{\infty} \mathbb{E} \|\boldsymbol{Y}_i\|_2^4 < \infty.$$

Since $I\{\nu(m-1) < i \leq \nu(m)\}$ is $\mathcal{F}_{i-1}$-measurable, write $Y_{i,r}$ for the r-th coordinate of $\boldsymbol{Y}_i$. Then the stopped increments

$$Y_{i,r} I\{\nu(m-1) < i \leq \nu(m)\}, \qquad i \geq 2,$$

form a scalar martingale-difference sequence for each $r = 1, \ldots, d$.

Using $\|\boldsymbol{x}\|_2^4 \le d\sum_{r=1}^d |x_r|^4$, and applying Doob's L^4 maximal inequality followed by the scalar martingale Rosenthal inequality (Hall & Heyde, 1980, Theorem 2.12) coordinatewise, we obtain

$$\mathbb{E}D_m^4 \le C_d \sum_{r=1}^{d} \mathbb{E}\left[\left\{ \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(Y_{i,r}^2 \mid \mathcal{F}_{i-1}) \right\}^2 + \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(|Y_{i,r}|^4 \mid \mathcal{F}_{i-1}) \right].$$

Since all conditional second moments are nonnegative,

$$\sum_{r=1}^{d} \left\{ \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(Y_{i,r}^2 \mid \mathcal{F}_{i-1}) \right\}^2 \le \left\{ \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(\|\boldsymbol{Y}_i\|_2^2 \mid \mathcal{F}_{i-1}) \right\}^2 = Q_m^2.$$

Moreover,

$$\sum_{r=1}^{d} \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(|Y_{i,r}|^4 \mid \mathcal{F}_{i-1}) \le \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1}).$$

Consequently,

$$\mathbb{E}D_m^4 \le C_d \mathbb{E}\left[Q_m^2 + \sum_{i=\nu(m-1)+1}^{\nu(m)} \mathbb{E}(\|\boldsymbol{Y}_i\|_2^4 \mid \mathcal{F}_{i-1}) \right].$$

The preceding bound on Q_m and Lemma 5 therefore imply

$$\sup_{m\ge 1} \mathbb{E}D_m^4 < \infty.$$

Hence, for every $\eta > 0$,

$$\sum_{m=1}^{\infty} \mathbb{P}\{D_m > m^{1/4+\eta}\} \le C \sum_{m=1}^{\infty} m^{-1-4\eta} < \infty,$$

and the Borel–Cantelli lemma gives $D_m = O_{a.s.}(m^{1/4+\eta})$.

For $m_n = \lceil V_n \rceil$, we have $m_n - 1 < V_n \le m_n$, and hence $\nu(m_n - 1) < n \le \nu(m_n)$. Therefore,

$$\begin{aligned} \left\|\boldsymbol{M}_n - \boldsymbol{M}_{\nu(m_n)}\right\|_2 &\le \left\|\boldsymbol{M}_n - \boldsymbol{M}_{\nu(m_n-1)}\right\|_2 + \left\|\boldsymbol{M}_{\nu(m_n)} - \boldsymbol{M}_{\nu(m_n-1)}\right\|_2 \\ &\le 2D_{m_n} = O_{a.s.}(V_n^{1/4+\delta}) \end{aligned}$$

for every $\delta > 0$, where we used $m_n/V_n \to 1$ almost surely.

Moreover, the Brownian unit-interval oscillations satisfy

$$\sup_{m-1\le t\le m} \|\boldsymbol{B}(t) - \boldsymbol{B}(m)\|_2 = O_{a.s.}(\sqrt{\log m}).$$

Combining this bound with the approximation at $\nu(m_n)$ gives

$$\left\|\boldsymbol{M}_n - \boldsymbol{\Gamma}^{1/2}\boldsymbol{B}(V_n)\right\|_2 = O_{a.s.}\left(V_n^{1/4+\delta} + V_n^{1/2-\kappa_0} + \sqrt{\log V_n}\right).$$

Choosing $0 < \delta < 1/4$, we obtain

$$\left\|\boldsymbol{M}_n - \boldsymbol{\Gamma}^{1/2}\boldsymbol{B}(V_n)\right\|_2 = O_{a.s.}(V_n^{1/2-\kappa_1}), \qquad \kappa_1 = \min\left\{\kappa_0, \frac{1}{4} - \delta\right\} > 0.$$

Finally, Lemma 5 gives

$$|V_n - cT_n| = O_{a.s.}(T_n^{3/4}).$$

A standard Brownian modulus estimate on dyadic time blocks implies that, for every $\varepsilon > 0$,

$$\|\boldsymbol{B}(V_n) - \boldsymbol{B}(cT_n)\|_2 = O_{a.s.}(T_n^{3/8+\varepsilon}).$$

Since $V_n \asymp T_n$ almost surely, choosing $0 < \varepsilon < 1/8$ gives

$$\left\|\boldsymbol{M}_n - \boldsymbol{\Gamma}^{1/2}\boldsymbol{B}(cT_n)\right\|_2 = O_{a.s.}(T_n^{1/2-\kappa}),$$

where

$$\kappa = \min\left\{\kappa_1, \frac{1}{8} - \varepsilon\right\} > 0.$$

Define $\boldsymbol{W}(t) = c^{-1/2}\boldsymbol{B}(ct)$. By Brownian scaling, $\boldsymbol{W}$ is a standard d-dimensional Brownian motion, and since $\boldsymbol{\Gamma} = \boldsymbol{\Omega}/c$,

$$\boldsymbol{\Gamma}^{1/2}\boldsymbol{B}(cT_n) = \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(T_n).$$

This completes the proof. □

D.3 Radial Brownian boundaries

Lemma 7 (Vector radial Brownian boundaries)**.** *For either* $b_{d,\alpha,t_0} \in \left\{b^{\mathrm{GM}}_{d,\alpha,t_0}, b^{\mathrm{LIL}}_{d,\alpha,t_0}\right\}$*, the bound satisfies*

$$\mathbb{P}\left\{\exists t \geq t_0 : \|\boldsymbol{W}_d(t)\|_2 \geq b_{d,\alpha,t_0}(t)\right\} \leq \alpha. \tag{7}$$

Equally, both choices are radial $(1-\alpha)$ *Brownian motion boundaries in the sense of Definition 4. Moreover, for every fixed* $t_0 > 0$*,*

$$\limsup_{t\to\infty} \frac{\|\boldsymbol{W}_d(t)\|_2}{b_{d,\alpha,t_0}(t)} < 1 \qquad \text{a.s.} \tag{8}$$

Proof. For the Gaussian-mixture boundary, let $\mu = \mathcal{N}_d(0, t_0^{-1}I_d)$ and consider

$$\begin{aligned} L_t &= \int_{\mathbb{R}^d} \exp\left\{\lambda^\top \boldsymbol{W}_d(t) - \frac{t\|\lambda\|_2^2}{2}\right\} \mu(d\lambda) \\ &= \left(\frac{t_0}{t+t_0}\right)^{d/2} \exp\left\{\frac{\|\boldsymbol{W}_d(t)\|_2^2}{2(t+t_0)}\right\}, \qquad t \geq 0. \end{aligned}$$

For each fixed λ, the integrand is a mean-one exponential martingale. Since μ is a probability measure, conditional Tonelli's theorem shows that $\{L_t\}_{t\geq 0}$ is a nonnegative martingale with $L_0 = 1$. For $t \geq t_0$, a crossing of $b^{\mathrm{GM}}_{d,\alpha,t_0}$ implies $L_t \geq 1/\alpha$. Ville's inequality therefore gives

$$\mathbb{P}\left\{\exists t \geq t_0 : \|\boldsymbol{W}_d(t)\|_2 \geq b^{\mathrm{GM}}_{d,\alpha,t_0}(t)\right\} \leq \mathbb{P}\left\{\sup_{t\geq 0} L_t \geq 1/\alpha\right\} \leq \alpha.$$

This proves (7) for the Gaussian-mixture boundary.

For the LIL-type boundary, for every $v \in \mathbb{S}^{d-1}$ and $\lambda \in \mathbb{R}$, $\exp\left\{\lambda v^\top \boldsymbol{W}_d(t) - \lambda^2 t/2\right\}$ is a nonnegative martingale for $t \geq 0$. Hence $\{(\boldsymbol{W}_d(t), tI_d)\}_{t\geq 0}$ is a vector sub-ψ_N process with $\psi_N(\lambda) = \lambda^2/2$.

Apply Whitehouse et al. (2026, Theorem 4.1) with regularization parameter $\rho = t_0$, confidence level $\delta = \alpha$, and $h(x) = \zeta(q)(x+1)^q$. Since $\sum_{k=0}^{\infty} 1/h(k) = 1$, the theorem applies.

For $t \geq t_0$, $tI_d \vee t_0 I_d = tI_d, \gamma_{\min}(tI_d) = \gamma_{\max}(tI_d) = t, \kappa(tI_d) = 1$, and $(\psi_N^*)^{-1}(x) = \sqrt{2x}$. Therefore, Theorem 4.1 yields that

$$\mathbb{P}\left\{\exists t \geq t_0 : \|\boldsymbol{W}_d(t)\|_2 \geq \frac{\sqrt{2at\, L_{t_0}(tI_d)}}{1-\varepsilon}\right\} \leq \alpha,$$

where $L_{t_0}(tI_d) = \log\zeta(q) + q\log\left\{1 + \log_a(t/t_0)\right\} + \log\{1/\alpha(1-\beta^{-1})\} + \log\left\{\beta N_{d-1}(\varepsilon/\beta)\right\}$. Using $N_{d-1}(u) \leq (3/u)^d$, we obtain

$$\begin{aligned} L_{t_0}(tI_d) \leq\ & q\log\left\{1 + \log_a\left(\frac{t}{t_0}\right)\right\} + \log\zeta(q) \\ & + \log\frac{1}{\alpha(1-\beta^{-1})} + \log\beta + d\log\left(\frac{3\beta}{\varepsilon}\right). \end{aligned}$$

Enlarging the boundary therefore gives (7) with $b_{d,\alpha,t_0} = b^{\mathrm{LIL}}_{d,\alpha,t_0}$.

From the (4) and (5), we know that $b_{d,\alpha,t_0}(t_0 r) = \sqrt{t_0}\, b_{d,\alpha,1}(r)$ for $r \geq 1$. Moreover, for the Gaussian-mixture boundary, since $t \geq t_0$, $b^{\mathrm{GM}}_{d,\alpha,t_0}(t) \geq \sqrt{d\log 2 + 2\log(1/\alpha)}\sqrt{t}$. For the LIL-type boundary, the expression inside the square brackets in (5) is bounded below by $\log\zeta(q) + \log\{1/\alpha(1-\beta^{-1})\} + \log\beta + d\log(3\beta/\varepsilon) > 0$. Thus there exists a constant $c_0 > 0$, independent of t_0, such that

$$b_{d,\alpha,t_0}(t) \geq c_0\sqrt{t}, \qquad t \geq t_0 > 0.$$

Finally, for fixed t_0, $b^{\mathrm{GM}}_{d,\alpha,t_0}(t)/\sqrt{2t\log\log t} \to \infty$, whereas $b^{\mathrm{LIL}}_{d,\alpha,t_0}(t)/\sqrt{2t\log\log t} \to \sqrt{aq}/(1-\varepsilon) > 1$. The fixed-dimensional Brownian law of the iterated logarithm gives

$$\limsup_{t\to\infty} \frac{\|\boldsymbol{W}_d(t)\|_2}{\sqrt{2t\log\log t}} = 1 \qquad \text{a.s.}$$

Hence (8) follows for both boundaries. □

D.4 Time-uniform coverage

Proof of Theorem 4. By Proposition 1, on an enriched probability space supporting a standard d-dimensional Brownian motion $\boldsymbol{W}_d$, $\boldsymbol{\Omega}^{-1/2}\boldsymbol{M}_n = \boldsymbol{W}_d(T_n) + \boldsymbol{e}_n$ where $\|\boldsymbol{e}_n\|_2 = o_{a.s.}(\sqrt{T_n})$. By Lemma 5, $\widehat{\boldsymbol{\Omega}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$. Since $\boldsymbol{\Omega}$ is positive definite, $\widehat{\boldsymbol{\Omega}}_n$ is almost surely positive definite for all sufficiently large n.

For every n, define

$$\boldsymbol{A}_n = \begin{cases} \boldsymbol{\Omega}^{1/2}\widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{\Omega}^{1/2}, & \widehat{\boldsymbol{\Omega}}_n \succ 0, \\ \boldsymbol{0}_{d\times d}, & \widehat{\boldsymbol{\Omega}}_n \text{ is singular.} \end{cases}$$

Then $\|\boldsymbol{A}_n - \boldsymbol{I}_d\|_{\mathrm{op}} \xrightarrow{\text{a.s.}} 0$.

By Lemma 7, for some $c_0 > 0$, $b_{d,\alpha,T_m}(T_n) \geq c_0\sqrt{T_n}$ for all $n \geq m$. Define

$$\eta_m = 1 - \left\{1 + \sup_{n\geq m}\|\boldsymbol{A}_n - \boldsymbol{I}_d\|_{\mathrm{op}}\right\}^{-1/2} + \frac{1}{c_0}\sup_{n\geq m}\frac{\|\boldsymbol{e}_n\|_2}{\sqrt{T_n}}.$$

Then $\eta_m \xrightarrow{\text{a.s.}} 0$.

Suppose that, for some $n \geq m$, $\boldsymbol{\theta} \notin C^{\mathrm{ell}}_{m,n}(\alpha)$. By the convention in (6), this can occur only if $\widehat{\boldsymbol{\Omega}}_n \succ 0$, in which case

$$\boldsymbol{M}_n^{\mathsf{T}}\widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{M}_n > \{b_{d,\alpha,T_m}(T_n)\}^2.$$

Using Proposition 1,

$$\begin{aligned} b_{d,\alpha,T_m}(T_n) &< \left[\{\boldsymbol{W}_d(T_n) + \boldsymbol{e}_n\}^{\mathsf{T}}\boldsymbol{A}_n\{\boldsymbol{W}_d(T_n) + \boldsymbol{e}_n\}\right]^{1/2} \\ &\leq \sqrt{1 + \|\boldsymbol{A}_n - \boldsymbol{I}_d\|_{\mathrm{op}}}\,\{\|\boldsymbol{W}_d(T_n)\|_2 + \|\boldsymbol{e}_n\|_2\}. \end{aligned}$$

Therefore

$$\begin{aligned} \|\boldsymbol{W}_d(T_n)\|_2 &> \left\{1 + \|\boldsymbol{A}_n - \boldsymbol{I}_d\|_{\mathrm{op}}\right\}^{-1/2} b_{d,\alpha,T_m}(T_n) - \|\boldsymbol{e}_n\|_2 \\ &\geq (1-\eta_m) b_{d,\alpha,T_m}(T_n). \end{aligned}$$

Define the feasible crossing events $E_m = \left\{\exists n \geq m : \boldsymbol{\theta} \notin C^{\mathrm{ell}}_{m,n}(\alpha)\right\}$ and ideal crossing events $A_m = \{\exists t \geq T_m : \|\boldsymbol{W}_d(t)\|_2 \geq b_{d,\alpha,T_m}(t)\}$. By Lemma 7, $\mathbb{P}(A_m) \leq \alpha$.

Fix $\delta \in (0,1)$. On $\{\eta_m \leq \delta\}$, occurrence of E_m implies that, for some $n \geq m$,

$$\|\boldsymbol{W}_d(T_n)\|_2 > (1-\delta) b_{d,\alpha,T_m}(T_n).$$

Hence

$$\mathbb{P}(E_m \setminus A_m) \leq \mathbb{P}(\eta_m > \delta) + \mathbb{P}\left\{A_m^c,\ \exists t \geq T_m : \|\boldsymbol{W}_d(t)\|_2 > (1-\delta) b_{d,\alpha,T_m}(t)\right\}.$$

Since $\eta_m \xrightarrow{\text{a.s.}} 0$, $\mathbb{P}(\eta_m > \delta) \to 0$.

By $b_{d,\alpha,t_0}(t_0 r) = \sqrt{t_0}\, b_{d,\alpha,1}(r)$ and Brownian scaling, the second probability equals

$$\mathbb{P}\left\{\|\boldsymbol{W}_d(r)\|_2 < b_{d,\alpha,1}(r) \text{ for all } r \geq 1,\ \exists r \geq 1 : \|\boldsymbol{W}_d(r)\|_2 > (1-\delta) b_{d,\alpha,1}(r)\right\}.$$

From Lemma 7, (8) gives that

$$\limsup_{r\to\infty} \frac{\|\boldsymbol{W}_d(r)\|_2}{b_{d,\alpha,1}(r)} < 1 \qquad \text{a.s.}$$

Thus, on the event $\left\{\|\boldsymbol{W}_d(r)\|_2 < b_{d,\alpha,1}(r) \text{ for all } r \geq 1\right\}$, there exist $\varepsilon > 0$ and $R < \infty$ such that

$$\frac{\|\boldsymbol{W}_d(r)\|_2}{b_{d,\alpha,1}(r)} \leq 1 - \varepsilon, \qquad r \geq R.$$

On the compact interval $[1, R]$, continuity of $\boldsymbol{W}_d$ and $b_{d,\alpha,1}$ gives

$$\sup_{1\leq r\leq R} \frac{\|\boldsymbol{W}_d(r)\|_2}{b_{d,\alpha,1}(r)} < 1.$$

Consequently,

$$\sup_{r\geq 1} \frac{\|\boldsymbol{W}_d(r)\|_2}{b_{d,\alpha,1}(r)} < 1$$

almost surely on this event. It follows that as $\delta \downarrow 0$,

$$\mathbb{P}\left\{\|\boldsymbol{W}_d(r)\|_2 < b_{d,\alpha,1}(r) \text{ for all } r \geq 1,\ \exists r \geq 1 : \|\boldsymbol{W}_d(r)\|_2 > (1-\delta) b_{d,\alpha,1}(r)\right\} \to 0$$

We therefore obtain, for every fixed $\delta \in (0,1)$,

$$\limsup_{m\to\infty} \mathbb{P}(E_m) \leq \alpha + \mathbb{P}\left\{\|\boldsymbol{W}_d(r)\|_2 < b_{d,\alpha,1}(r) \text{ for all } r \geq 1,\ \exists r \geq 1 : \|\boldsymbol{W}_d(r)\|_2 > (1-\delta) b_{d,\alpha,1}(r)\right\}.$$

Letting $\delta \downarrow 0$ yields

$$\limsup_{m\to\infty} \mathbb{P}\left\{\exists n \geq m : \boldsymbol{\theta} \notin C^{\text{ell}}_{m,n}(\alpha)\right\} \leq \alpha.$$

Equivalently,

$$\liminf_{m\to\infty} \mathbb{P}\left\{\forall n \geq m : \boldsymbol{\theta} \in C^{\text{ell}}_{m,n}(\alpha)\right\} \geq 1 - \alpha.$$

□

D.5 Multivariate asymptotic confidence sequences

Proof of Remark 3. Fix m. By Proposition 1,

$$\boldsymbol{\Omega}^{-1/2}\boldsymbol{M}_n = \boldsymbol{W}_d(T_n) + \boldsymbol{e}_n, \qquad \|\boldsymbol{e}_n\|_2 = o_{a.s.}(\sqrt{T_n}).$$

Hence

$$\sup_{n\geq m} \frac{\|\boldsymbol{e}_n\|_2}{\sqrt{T_n}} < \infty \qquad \text{a.s.}$$

Choose $\delta \in (0, \alpha)$ and a deterministic $\varepsilon > 0$ such that

$$\mathbb{P}\left\{\sup_{n\geq m} \frac{\|\boldsymbol{e}_n\|_2}{\sqrt{T_n}} > \varepsilon\right\} \leq \delta,$$

and put $a = \alpha - \delta$. Define

$$C^{*,\text{ell}}_{m,n}(\alpha) = \left\{\boldsymbol{\vartheta} \in \mathbb{R}^d : \left\|\boldsymbol{\Omega}^{-1/2}(n-1)(\widetilde{\boldsymbol{U}}_n - \boldsymbol{\vartheta})\right\|_2 \leq b_{d,a,T_m}(T_n) + \varepsilon\sqrt{T_n}\right\}.$$

By Lemma 7,

$$\mathbb{P}\left\{\|\boldsymbol{W}_d(t)\|_2 < b_{d,a,T_m}(t) \text{ for all } t \geq T_m\right\} \geq 1 - a.$$

On this event and

$$\left\{\sup_{n\geq m} \frac{\|\boldsymbol{e}_n\|_2}{\sqrt{T_n}} \leq \varepsilon\right\},$$

we have, for every $n \geq m$,

$$\left\|\boldsymbol{\Omega}^{-1/2}\boldsymbol{M}_n\right\|_2 \leq \|\boldsymbol{W}_d(T_n)\|_2 + \|\boldsymbol{e}_n\|_2 < b_{d,a,T_m}(T_n) + \varepsilon\sqrt{T_n}.$$

Since $\boldsymbol{M}_n = (n-1)(\widetilde{\boldsymbol{U}}_n - \boldsymbol{\theta})$, a union bound gives

$$\mathbb{P}\left\{\boldsymbol{\theta} \in C^{*,\mathrm{ell}}_{m,n}(\alpha) \text{ for all } n \geq m\right\} \geq 1 - a - \delta = 1 - \alpha.$$

Thus $\{C^{*,\mathrm{ell}}_{m,n}(\alpha) : n \geq m\}$ is a nonasymptotic $(1-\alpha)$-confidence sequence.

It remains to compare this sequence with $C^{\mathrm{ell}}_{m,n}(\alpha)$. For fixed m, the explicit forms of either radial boundary give

$$\frac{b_{d,a,T_m}(T_n) + \varepsilon\sqrt{T_n}}{b_{d,\alpha,T_m}(T_n)} \longrightarrow 1.$$

Indeed, as $b_{d,a,T_m}(T_n)/b_{d,\alpha,T_m}(T_n) \to 1$, $\sqrt{T_n}/b_{d,\alpha,T_m}(T_n) \to 0$ and $\widehat{\boldsymbol{\Omega}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}$ there exists a random sequence $\rho_n \xrightarrow{\text{a.s.}} 0$ such that, almost surely for all sufficiently large n,

$$(1-\rho_n)\left\{C^{*,\mathrm{ell}}_{m,n}(\alpha) - \widetilde{\boldsymbol{U}}_n\right\} \subseteq C^{\mathrm{ell}}_{m,n}(\alpha) - \widetilde{\boldsymbol{U}}_n \subseteq (1+\rho_n)\left\{C^{*,\mathrm{ell}}_{m,n}(\alpha) - \widetilde{\boldsymbol{U}}_n\right\}.$$

Therefore,

$$\frac{\nu\left(C^{\mathrm{ell}}_{m,n}(\alpha) \triangle C^{*,\mathrm{ell}}_{m,n}(\alpha)\right)}{\nu\left(C^{*,\mathrm{ell}}_{m,n}(\alpha)\right)} \leq (1+\rho_n)^d - (1-\rho_n)^d \xrightarrow{\text{a.s.}} 0.$$

Hence the ellipsoids satisfy the multivariate AsympCS criterion of Waudby-Smith et al. (2024, Definition B.6). □