ASYMPTOTICALLY ANYTIME-VALID INFERENCE FOR FRÉCHET VARIANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Fréchet variance quantifies dispersion for random objects in metric spaces. Existing inference is mainly fixed-sample and does not justify repeated inspection or data-dependent stopping. We construct positive asymptotic confidence sequences by applying time-uniform normal approximations to log Fréchet variance and then exponentiating. The offline procedure recomputes empirical Fréchet means, while the online procedure uses predictable squared-distance losses based on Riemannian stochastic-gradient centres. Under local smoothness, curvature and moment conditions, estimating or tracking the Fréchet mean has a uniformly lower-order effect, yielding asymptotic anytime-valid coverage and boundary-dependent width rates. Simulations and an EEG covariance example illustrate the effect of repeated monitoring and the computational gain from the online construction.

1 INTRODUCTION

Statistical problems involve random objects, including shapes, covariance matrices, networks, phylogenetic trees, point clouds and probability distributions, whose geometry is described by a metric rather than vector operations. Fréchet's variational definition gives a centre by minimizing expected squared distance and a scalar dispersion by the corresponding minimum (Fréchet, 1948). This construction underlies large-sample theory for means on manifolds and in metric spaces (Ziezold, 1977; Bhattacharya & Patrangenaru, 2003; 2005; Pennec, 2006; Afsari, 2011; Patrangenaru & Ellingson, 2015; Marron & Alonso, 2014; Dryden & Mardia, 2016; Petersen & Müller, 2019; Schötz, 2019; Ahidar-Coutrix et al., 2020). In Euclidean space the Fréchet variance is the trace of the covariance matrix, but it remains defined without linear operations. Existing central-limit theory and variance estimation for its empirical version are fixed-sample results and therefore do not justify repeated monitoring or data-dependent stopping (Dubey & Müller, 2019; Song & Müller, 2026).

Confidence sequences cover simultaneously over time and remain valid at stopping times. Their martingale foundations are classical (Ville, 1939; Darling & Robbins, 1967; Robbins, 1970; Robbins & Siegmund, 1970), while modern constructions use mixture boundaries, betting arguments and empirical Bernstein methods (Howard et al., 2021; Waudby-Smith & Ramdas, 2024; Ramdas et al., 2023; Grünwald et al., 2024). Time-uniform central limit theory offers a complementary asymptotic route through strong Gaussian approximation (Waudby-Smith et al., 2024), but none of these tools handles a Fréchet variance profiled over an unknown centre: although the target is scalar, its empirical version involves data-dependent optimization rather than averaging losses at a fixed centre.

Sequential Fréchet-variance inference has two coupled difficulties. Offline, estimating the centre creates a profiling remainder that must be controlled uniformly in time rather than at a fixed sample size. Online, the data-adaptive centre moves on a curved space and must remain predictable when each loss is evaluated; pointwise convergence of the centre does not control the accumulated effect of tracking and curvature errors on the loss sequence. We evaluate squared-distance losses at the pre-update Riemannian stochastic-gradient centre, which depends only on past observations, and prove that tracking error, curvature bias and the associated stochastic remainder are uniformly lower order. These one-pass moving-centre losses therefore have the same leading partial-sum expansion as losses evaluated at the unknown population mean, so scalar time-uniform inference applies despite the evolving geometric estimate.

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.

We make three contributions. First, a time-uniform log-linearization of the empirical Fréchet variance shows that estimating the Fréchet mean affects the offline statistic only at second order, which provides the uniform remainder control needed to move from fixed-sample inference to repeated monitoring. Second, for streaming data we establish an almost-sure tracking rate for Riemannian stochastic gradient descent and use it to construct a positive, one-pass asymptotic confidence sequence with the same first-order limit as the offline procedure, so first-order accuracy is retained without recomputing the empirical Fréchet mean after each observation. Third, we translate normal-mixture and finite-LIL boundaries into explicit width rates, clarifying how the choice of boundary affects sequential precision, and assess sequential noncoverage, interval width and computation in simulations and an EEG covariance application. Positivity is enforced by working on the logarithmic scale and exponentiating.

The rest of the paper is organized as follows. Section 2 presents the offline and RSGD online procedures, Section 3 their assumptions and guarantees, Sections 4 and 5 the simulations and the EEG application, and Section 6 concludes; geometry-specific conditions and formulas, complete empirical results and all proofs are collected in the appendices.

2 METHODOLOGY

Let $X_1, X_2, \ldots$ be independent copies of a random object X in a separable metric space $(\mathcal{M}, d)$. The Fréchet functional $F(m) = \mathbb{E}d(X, m)^2$ has a unique minimizer μ, and the target is its minimum $V = F(\mu) > 0$. We work with $\vartheta = \log V$ and exponentiate the reported intervals to obtain positive variance-scale intervals. Write $\sigma^2 = \mathrm{Var}\{d(X, \mu)^2\}$ and $\tau^2 = \sigma^2/V^2$, with $0 < \sigma^2 < \infty$. A deterministic burn-in n_0 specifies when monitoring begins; the ridge $a_n = n^{-2}$ makes logarithms well defined and is eventually inactive almost surely.

2.1 BOUNDARY AND COVERAGE TARGET

We use the normal-mixture and finite-LIL boundaries (Howard et al., 2021; Waudby-Smith et al., 2024):

$$b_\alpha^{\mathrm{GM}}(n; \rho) = \{(n+\rho)[2\log(1/\alpha) + \log\{(n+\rho)/\rho\}]\}^{1/2},$$
$$b_\alpha^{\mathrm{LIL}}(n) = 1.7\{n[\log\log(2n) + 0.72\log(10.4/\alpha)]\}^{1/2}. \tag{1}$$

Here $\alpha \in (0, 1)$ and $\rho = r^{-2} > 0$, where r denotes the mixture-scale parameter in Waudby-Smith et al. (2024). Both boundaries satisfy

$$\log\log n = o\{b_\alpha(n)\}, \qquad \liminf_{n\to\infty} \frac{b_\alpha(n)}{\sqrt{n\log\log n}} > \sqrt{2}. \tag{2}$$

Their strict slack above the law-of-the-iterated-logarithm (LIL) scale absorbs the effect of estimating the centre and standard error.

Definition 1 (Asymptotic confidence sequence)**.** Intervals $I_n = [T_n \pm h_n]$ form a $(1-\alpha)$ asymptotic confidence sequence for θ if, on a possibly extended probability space, positive reference radii h_n^* satisfy

$$\mathbb{P}\{\theta \in [T_n \pm h_n^*] \text{ for every } n \geq n_0\} \geq 1 - \alpha, \qquad h_n/h_n^* \to 1 \quad \text{a.s.}.$$

For asymmetric intervals the criterion applies to each endpoint distance from the centre.

This is the strong-approximation criterion of Waudby-Smith et al. (2024), verified for (1) in Proposition C.3. The reference radii need not be computable; the reported radius is the estimated log-scale standard deviation times $b_\alpha(n)/n$.

2.2 OFFLINE ESTIMATION

The offline procedure recomputes the empirical Fréchet mean at each observation:

$$\widehat{\mu}_n \in \arg\min_m n^{-1}\sum_{i=1}^n d(X_i, m)^2, \qquad \widehat{V}_n = n^{-1}\sum_{i=1}^n d(X_i, \widehat{\mu}_n)^2,$$
$$\widehat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \{d(X_i, \widehat{\mu}_n)^2 - \widehat{V}_n\}^2. \tag{3}$$

Algorithm 1 Offline log-Fréchet-variance confidence sequence

For each $n \geq n_0$, compute the empirical centre, variance and squared-loss variance in (3). Apply the ridge and report (4).

Algorithm 2 Online log-Fréchet-variance confidence sequence

1. Input $c_0 \in K$, $\gamma_n = \gamma/(n + n_*)$, n_0, α and b_α. Initialize $A_0 = B_0 = 0$.
2. Observe X_n and update
$$Y_n = d(X_n, c_{n-1})^2, \qquad A_n = A_{n-1} + Y_n, \qquad B_n = B_{n-1} + Y_n^2.$$
3. Update $c_n = \Pi_K[\mathrm{Exp}_{c_{n-1}}\{\gamma_n \mathrm{Log}_{c_{n-1}} X_n\}]$.
4. For $n \geq n_0$, set $\bar{Y}_n = A_n/n$, $\widehat{s}_n^2 = B_n/n - \bar{Y}_n^2$, $\check{Y}_n = \max(\bar{Y}_n, n^{-2})$ and $\widehat{\tau}_n^{\mathrm{on}} = \widehat{s}_n/\check{Y}_n$. Report
$$\mathcal{C}_n^{\mathrm{on,log}} = [\log \check{Y}_n \pm \widehat{\tau}_n^{\mathrm{on}} b_\alpha(n)/n], \qquad \mathcal{C}_n^{\mathrm{on},+} = \exp(\mathcal{C}_n^{\mathrm{on,log}}).$$

Set $\check{V}_n = \max(\widehat{V}_n, n^{-2})$ and $\widehat{\tau}_n = \widehat{\sigma}_n/\check{V}_n$. For $n \geq n_0$, report

$$\mathcal{C}_n^{\log} = [\log \check{V}_n \pm \widehat{\tau}_n b_\alpha(n)/n], \qquad \mathcal{C}_n^{+} = \exp(\mathcal{C}_n^{\log}). \tag{4}$$

The notation $[u \pm r]$ denotes $[u - r, u + r]$. Equations (3)–(4) specify Algorithm 1.

2.3 Predictable online estimation

Riemannian stochastic gradient descent (RSGD) updates the centre using one observation at a time (Bonnabel, 2013; Arnaudon et al., 2012; Kushner & Yin, 2003). For $h_x(m) = d(x, m)^2/2$, its gradient is $G_x(m) = -\mathrm{Log}_m x$. We use a geodesically convex set K, projection Π_K, and step sizes $\gamma_n = \gamma/(n + n_*)$. The loss $Y_n = d(X_n, c_{n-1})^2$ is evaluated before the centre update. Conditional on the past, c_{n-1} is fixed and X_n is independent, which separates the accumulated loss into a population fluctuation, a tracking bias and a martingale remainder.

Algorithm 2 stores one centre and two scalar sums. In flat coordinates, including local torus coordinates and log-Euclidean covariance matrices, the centre update reduces to $c_n = c_{n-1} + \gamma_n(U_n - c_{n-1})$, where U_n is the coordinate representation of X_n; on smooth Riemannian regions it uses the exponential and logarithm maps. Two representative cases follow; further geometry-specific formulas appear in Appendix C.3.

2.4 Representative geometries

Euclidean dispersion. For $\mathcal{M} = \mathbb{R}^q$, the target is the total coordinate variance,

$$\mu = \mathbb{E}X, \qquad V = \mathbb{E}\|X - \mu\|^2 = \mathrm{tr}\{\mathrm{Var}(X)\}.$$

The offline centre is $\bar{X}_n$ and $\widehat{V}_n = n^{-1}\sum_i \|X_i - \bar{X}_n\|^2$. The online centre follows $c_n = c_{n-1} + \gamma_n(X_n - c_{n-1})$; initialized at $c_1 = X_1$, the choice $\gamma_n = 1/n$ for $n \geq 2$ gives the running mean. Even in this familiar case, the two variance estimates differ: the offline losses use the fitted centre $\bar{X}_n$, whereas the online loss uses the preceding centre c_{n-1}. A finite fourth moment, $V > 0$ and nondegenerate squared-distance variation suffice for the offline conditions below. The stated online result additionally uses compact support.

Affine-invariant covariance dispersion. For positive-definite matrices $\mathrm{SPD}(p)$, the affine-invariant metric is

$$d_{\mathrm{AI}}(P, Q) = \|\log(P^{-1/2} Q P^{-1/2})\|_F.$$

The Fréchet mean is the Karcher mean and V is the expected squared distance from it. Algorithm 2 becomes the matrix recursion

$$c_n = \Pi_K\left[c_{n-1}^{1/2} \exp\{\gamma_n \log(c_{n-1}^{-1/2} X_n c_{n-1}^{-1/2})\} c_{n-1}^{1/2}\right].$$

Thus the same two scalar loss sums produce the uncertainty interval after each matrix update. On compact spectral sets $K = \{P : aI \preceq P \preceq bI\}$, with $0 < a < b < \infty$ and the mean in the interior, the distance derivatives and losses are bounded. Hadamard geometry supplies a unique mean and nonexpansive projection onto closed geodesically convex sets, which together with nondegenerate squared-distance variation give a concrete regime for the theory below. Under isotropic tangent coordinates $X = \exp(RA)$, with A a unit symmetric direction, $\mu = I_p$ and $V = \mathbb{E}R^2$.

3 Theoretical guarantees

The core results show that estimating the unknown centre preserves the leading scalar fluctuation of log Fréchet variance. We state the operative conditions here; Appendix C gives their complete envelope formulation, supporting results and all proofs.

In a finite-dimensional chart around μ, let $\theta = 0$ represent μ, $\ell_\theta(x) = d(x, m_\theta)^2$, $Q(\theta) = \mathbb{E}\ell_\theta(X)$ and $Q_n(\theta) = n^{-1}\sum_i \ell_\theta(X_i)$. Write $\widehat{\theta}_n$ for the chart coordinate of the selected empirical mean and $H = \nabla^2 Q(0)$. On a compact convex chart domain Θ containing 0 in its interior, Q has a unique separated minimum at 0 and a positive-definite Hessian there. The loss is three times continuously differentiable, with envelopes $M_0, \ldots, M_3$ for the loss and its first three derivatives satisfying $\mathbb{E}(M_0^2 + M_0M_1 + M_1^2 + M_2 + M_3) < \infty$. These are Assumption C.1; the selected empirical mean minimizes the criterion over this chart domain, and a global empirical minimizer has the same eventual localization when a uniform law and global separation place it there.

The conditions separate identification from fluctuation control: separation localizes the empirical mean, the positive Hessian turns score fluctuations into centre error, and the derivative envelopes provide uniform laws for the criterion, its squared loss and its Hessian, which respectively control the fitted variance, its standard error and the Taylor remainder. Compact support in a sufficiently small regular geodesic ball supplies bounded envelopes; for unbounded Euclidean observations the same requirements follow from the fourth-moment condition in Section 2.4.

Theorem 1 (Offline profiling and positive confidence sequence)**.** *Under the local smooth model and chart selection above, $\check{V}_n = \widehat{V}_n$ eventually almost surely,*

$$\log \check{V}_n - \log V = \frac{1}{nV}\sum_{i=1}^{n}\{d(X_i,\mu)^2 - V\} + O_{\mathrm{a.s.}}\{\log\log(e^e n)/n\}, \tag{5}$$

and $\widehat{\tau}_n^2 \to \tau^2$ almost surely. Either boundary in (1) *gives asymptotic confidence sequences for* $\log V$ *and* V*, with $\mathcal{C}_n^+ \subset (0,\infty)$. More generally,* (2) *gives eventual almost-sure containment.*

To see the profiling mechanism, put $r_n = \{\log\log(e^e n)/n\}^{1/2}$. The empirical score at μ and the chart error of $\widehat{\mu}_n$ are both $O_{\mathrm{a.s.}}(r_n)$. Taylor expansion therefore gives $\widehat{V}_n - n^{-1}\sum_i d(X_i,\mu)^2 = O_{\mathrm{a.s.}}(r_n^2)$: the empirical-score term is a product of two $O_{\mathrm{a.s.}}(r_n)$ quantities, and the quadratic term has the same order. The log transformation also has a quadratic remainder, leaving the influence function $\{d(X,\mu)^2 - V\}/V$. Multiplication by n gives a remainder of order $\log\log n$, smaller than the boundary. Proposition C.1 supplies the scale consistency. The profiling and coverage proofs are in Appendices C.4.2 and C.4.4.

More explicitly, define the empirical score $g_n = \nabla Q_n(0)$ and the integrated Hessian

$$H_n = \int_0^1 \nabla^2 Q_n(t\widehat{\theta}_n)\, dt.$$

The estimating equation gives $\widehat{\theta}_n = -H_n^{-1} g_n$ eventually. The uniform Hessian law and consistency imply $H_n \to H$, whereas the coordinatewise LIL gives $\|g_n\| = O_{\mathrm{a.s.}}(r_n)$. A second Taylor expansion yields

$$Q_n(\widehat{\theta}_n) - Q_n(0) = g_n^\top \widehat{\theta}_n + \frac{1}{2}\widehat{\theta}_n^\top \widetilde{H}_n \widehat{\theta}_n = O_{\mathrm{a.s.}}(r_n^2),$$

where $\widetilde{H}_n$ is a bounded averaged Hessian along the segment from 0 to $\widehat{\theta}_n$. The population derivative vanishes at the minimizer, so centre estimation enters the profiled variance through this quadratic term. The geometry affects localization and Hessian control, while the leading dispersion fluctuation is scalar.

Online geometric regime. In addition, K is compact and strongly geodesically convex, contains the support of X and has μ in its interior. Squared-distance losses are smooth on K, their gradients are bounded, and $q(m) = F(m)/2$ is geodesically λ-strongly convex. The geometric one-step bound (C.1) and projection nonexpansivity toward μ hold; the harmonic steps satisfy $2\lambda\gamma > 1$, with n_* large enough for the bound. Finally, $\mathbb{E}\{[d(X,m)^2 - d(X,\mu)^2]^2\} \le L^2 d(m,\mu)^2$. Assumption C.2 states these conditions in full. Compact regular geodesic balls give bounded envelopes and avoid cut-locus singularities (Afsari, 2011; Arnaudon et al., 2012).

Theorem 2 (Online profiling and positive confidence sequence)**.** *Under the local smooth model and online geometric regime, with a chart containing K, Algorithm 2 satisfies $\check{Y}_n = \bar{Y}_n$ eventually almost surely,*

$$\log \check{Y}_n - \log V = \frac{1}{nV}\sum_{i=1}^{n}\{d(X_i,\mu)^2 - V\} + o_{\mathrm{a.s.}}(n^{-1/2}), \tag{6}$$

and $(\widehat{\tau}_n^{\mathrm{on}})^2 \to \tau^2$ almost surely. Either boundary in (1) *gives a positive asymptotic confidence sequence for V;* (2) *more generally gives eventual almost-sure containment.*

The online proof uses predictability to recover the same leading partial sum as (5). Write $\ell_m(x) = d(x,m)^2$, $Pf = \mathbb{E}f(X)$ and $\mathcal{F}_i = \sigma(X_1,\dots,X_i)$. Then

$$\bar{Y}_n - V = n^{-1}\sum_{i=1}^{n}\{\ell_\mu(X_i) - V\} + \beta_n + \zeta_n, \quad \beta_n = n^{-1}\sum_{i=1}^{n}\{P\ell_{c_{i-1}} - P\ell_\mu\}, \quad \zeta_n = n^{-1}\sum_{i=1}^{n} D_i,$$

with $D_i = \ell_{c_{i-1}}(X_i) - \ell_\mu(X_i) - P\{\ell_{c_{i-1}} - \ell_\mu\}$. Evaluating the loss before updating the centre is essential here: it makes the D_i martingale increments, $\mathbb{E}(D_i \mid \mathcal{F}_{i-1}) = 0$. Since $\nabla F(\mu) = 0$, Taylor expansion, local curvature and loss Lipschitzness give

$$0 \le P\ell_{c_{i-1}} - P\ell_\mu \le C\, d(c_{i-1},\mu)^2, \qquad \mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1}) \le L^2 d(c_{i-1},\mu)^2,$$

so Proposition C.2 controls both remainders: $\beta_n = O_{\mathrm{a.s.}}(n^{-1+\varepsilon})$ for $\varepsilon \in (0,1/2)$, while a dyadic martingale maximal inequality gives $\sum_i D_i = o_{\mathrm{a.s.}}(\sqrt{n})$. Thus $\beta_n + \zeta_n = o_{\mathrm{a.s.}}(n^{-1/2})$, and both log-variance estimates have the same $N(0,\tau^2)$ limit after centering and $\sqrt{n}$ scaling; see Appendices C.4.6 and C.4.7.

The tracking rate follows from the curvature of the population criterion: the one-step geometric bound and strong convexity give

$$\mathbb{E}\{d(c_n,\mu)^2 \mid \mathcal{F}_{n-1}\} \le (1 - 2\lambda\gamma_n) d(c_{n-1},\mu)^2 + C\gamma_n^2.$$

The first term contracts the previous error and the second is the cost of the stochastic gradient. For harmonic steps, $2\lambda\gamma > 1$ makes this contraction strong enough to yield the almost-sure rates in Proposition C.2, and summing those rates controls both loss remainders without recomputing an empirical mean.

Standardization uses the same stored losses: smoothness and tracking imply that the averages of Y_i and Y_i^2 converge to V and $\mathbb{E}\ell_\mu(X)^2$, so the online sample variance estimates σ^2 and division by $\check{Y}_n^2$ estimates τ^2, the scale appearing in the confidence sequence.

Boundary widths and monitoring. Both procedures have log-scale radii of order $(\log n/n)^{1/2}$ for the normal-mixture boundary and $(\log\log n/n)^{1/2}$ for the finite-LIL boundary. The latter matches the order of the leading LIL fluctuation. Exponentiation preserves simultaneous coverage exactly and yields multiplicative intervals $\check{V}_n \exp\{\pm\widehat{\tau}_n b_\alpha(n)/n\}$, with the analogous online expression. Under (2), the proofs establish eventual almost-sure containment, so the probability of coverage for every $n \ge m$ tends to one as m increases; this is the conservative asymptotic meaning of the chosen boundaries.

Comparison with the Euclidean case. For $\mathcal{M} = \mathbb{R}^q$ the two theorems concern the sample variance $\mathrm{tr}\{\widehat{\mathrm{Var}}(X)\}$ and its running-mean analogue, and the mechanisms become explicit. The profiling remainder in (5) is $\widehat{V}_n - n^{-1}\sum_i \|X_i - \mu\|^2 = -\|\bar{X}_n - \mu\|^2$, so the score term $g_n^\top \widehat{\theta}_n = -2\|\bar{X}_n - \mu\|^2$ and the quadratic term $\frac{1}{2}\widehat{\theta}_n^\top \widetilde{H}_n \widehat{\theta}_n = \|\bar{X}_n - \mu\|^2$, with $H = 2I_q$, sum to this correction, and the LIL for $\bar{X}_n$ gives the order $\log\log n/n$ directly; the general result reproduces

this through the chart Hessian at the cost of localizing the empirical mean, and the fourth-moment condition of Section 2.4 replaces compact support. Online, the steps $\gamma_n = 1/n$ make $c_{n-1} = \bar{X}_{n-1}$, the one-step bound (C.1) holds with equality and $A = 1$, and the bias β_n is exactly the average of $\|\bar{X}_{i-1} - \mu\|^2$. Curvature therefore enters only through the constants A, λ and L of Assumption C.2, and the manifold procedures retain the Euclidean width rates. Beyond the two representative geometries of Section 2.4, Appendix C.3 verifies the conditions and gives the explicit centre updates for the flat torus, spherical caps, hyperbolic balls, log-Euclidean covariance matrices, rotations and product manifolds.

4 Simulation Study

We assess simultaneous coverage under repeated monitoring at nominal level 0.95. For each method, coverage is the proportion of the 500 replications in which every interval from $n_0 = 100$ through the terminal sample size contains V. We use b_α^{GM} with $\rho = 1$ and compare the offline and online log-CS with the fixed-time log-CLT interval monitored at the same times. The fixed-time interval uses the offline estimate and its standard error with the Gaussian quantile.

The three principal designs examine Theorems 1 and 2 through $N = 1000$ observations on spheres and $N = 500$ otherwise:

- **Spheres $\mathbb{S}^5, \mathbb{S}^{10}$.** The distribution is rotationally symmetric about the north pole, with geodesic radius $R = 0.72B$, $B \sim \mathrm{Beta}(2, 5)$, and target $V = \mathbb{E}R^2$; the focus is offline profiling and predictable tracking in positively curved, regular geodesic caps.
- **Hyperbolic spaces $\mathbb{H}^5, \mathbb{H}^{10}$.** Radial symmetry about the origin and radius $R \sim \mathrm{Unif}(0, 0.60)$ give $V = 0.60^2/3$, under negative curvature and compact radial support.
- **Affine-invariant $\mathrm{SPD}(p)$, $p = 2, 3, 4$.** The distribution is isotropic about I_p in tangent coordinates, with the same radius law as the spherical settings and $V = \mathbb{E}R^2$; in this covariance-matrix setting the nonlinear centre update enters the online decomposition.

Figures 1, 2 and 3 show how simultaneous coverage evolves during monitoring; Appendix A gives the pointwise coverage trajectories and supporting summaries. Table 1 shows simultaneous coverage of 0.984–0.998 for offline log-CS and 0.962–0.992 for online log-CS, against 0.648–0.784 for repeatedly monitored log-CLT intervals; both confidence sequences are conservative, consistent with the eventual-coverage guarantees. The flat-torus experiments, where online coverage falls below 0.95, are reported completely in Appendix A.2.

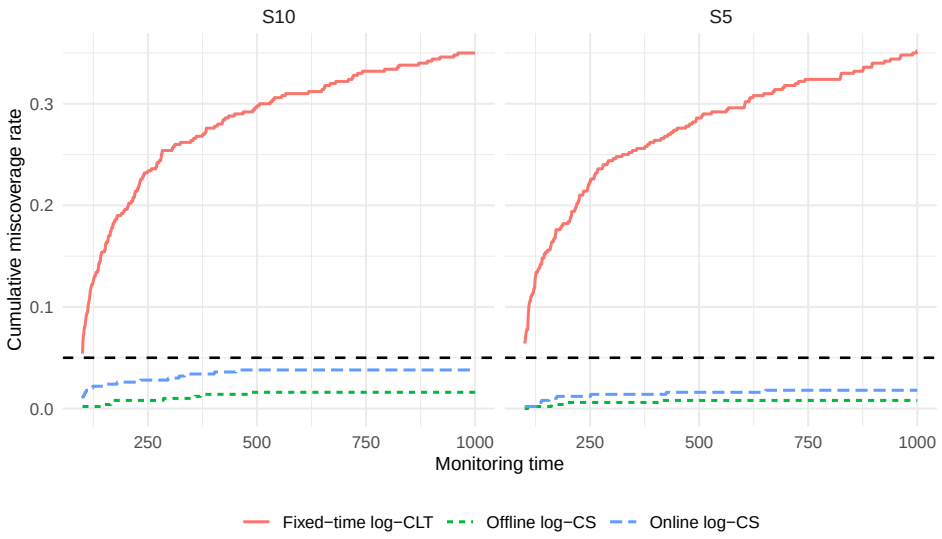

Figure 1: Cumulative miscoverage under repeated monitoring for the spherical simulation settings.

Table 1: Simultaneous coverage over all integer monitoring times; bold entries deviate less from the nominal 0.95 than the repeatedly monitored fixed-time log-CLT baseline.

Method	$\mathbb{S}^5$	$\mathbb{S}^{10}$	$\mathbb{H}^5$	$\mathbb{H}^{10}$	SPD(2)	SPD(3)	SPD(4)
Offline log-CS	**0.992**	**0.984**	**0.996**	**0.986**	**0.998**	**0.988**	**0.986**
Online log-CS	**0.982**	**0.962**	**0.978**	**0.974**	**0.992**	**0.984**	**0.974**
Fixed-time log-CLT	0.648	0.650	0.784	0.746	0.752	0.742	0.736

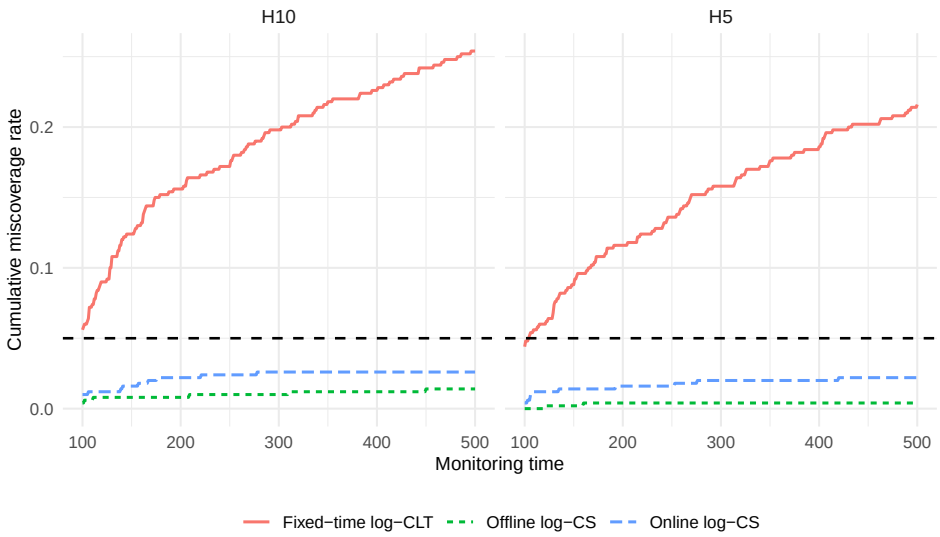

Figure 2: Cumulative miscoverage under repeated monitoring for the hyperbolic simulation settings.

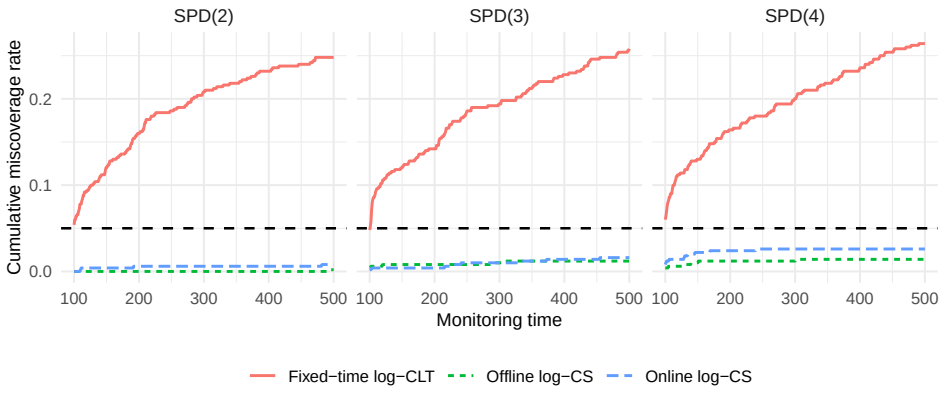

Figure 3: Cumulative miscoverage under repeated monitoring for affine-invariant SPD simulations. The dashed line marks $\alpha = 0.05$; the fixed-time log-CLT interval is monitored repeatedly as a pointwise baseline.

5 Real-data application

We apply the procedures to the PhysioNet EEG Motor Movement/Imagery data (Goldberger et al., 2000; Schalk et al., 2004). Spatial covariance matrices summarize dependence across EEG channels and form SPD-valued data, a representation used in brain–computer-interface analysis (Barachant

et al., 2012; 2013). We use subjects $1, \ldots, 30$, runs $4, 8, 12$, and channels FC3, FC4, C3, C4, CP3 and CP4. Each subject–run recording is split into nonoverlapping four-second epochs, at most four per run, and a regularized spatial correlation matrix per epoch gives 360 observations of dimension 6×6. They are processed in subject–run–epoch order with burn-in $n_0 = 40$, using the affine-invariant metric in Section 2.4. The offline procedure recomputes the Karcher mean, whereas the online procedure uses predictable losses from the affine-invariant RSGD centre; the subject–run–epoch observations carry dependence within subjects and runs.

Table 2: EEG covariance application, one SPD observation per selected PhysioNet subject–run–epoch segment. Estimates and final widths use the final observation; W_t is the interval width at observation index t.

Method	Final est.	Final width	Time (s)	W_{40}	W_{180}	W_{360}
Offline log-CS	6.460	1.805	35.110	1.817	2.285	1.805
Online log-CS	7.740	2.647	**0.076**	9.556	3.405	2.647
Fixed-time log-CLT	–	1.022	–	1.123	1.330	1.022

Table 2 and Figure 4 show appreciable dispersion in the selected EEG covariance patterns. The online trajectory is somewhat higher because losses are evaluated at recursively updated predictable centres; its interval is wider near burn-in but narrows substantially as more epochs arrive. The fixed-time log-CLT interval provides a pointwise benchmark. The main practical conclusion is computational: the offline method takes 35.110 seconds and the online method 0.076 seconds, so the proposed sequential intervals track covariance-dispersion uncertainty without repeatedly solving the full empirical Fréchet-mean problem.

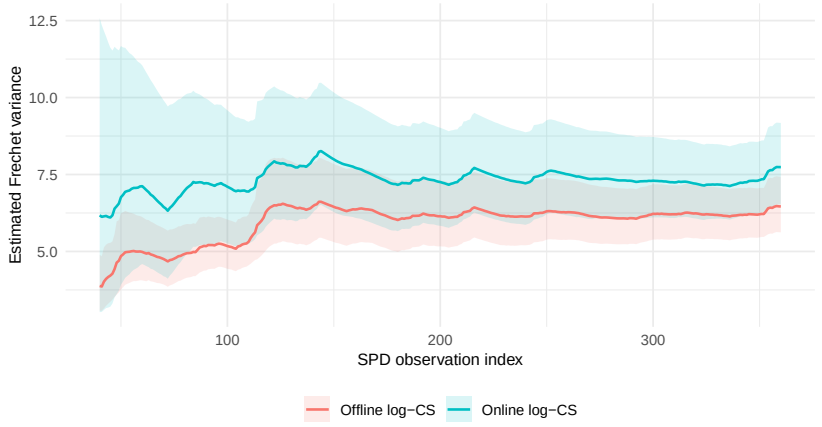

Figure 4: Offline and online log-CS trajectories for the EEG covariance application; the observation index is the processing order of the selected subject–run–epoch covariance matrices, not the time axis of a single recording.

Figure 5 complements the dispersion trajectories with their uncertainty widths. The online interval contracts markedly from burn-in to the final observation, while the offline width varies as the fitted centre and squared-loss variation change. The pointwise log-CLT interval is narrower throughout; its fixed-time coverage target differs from simultaneous monitoring.

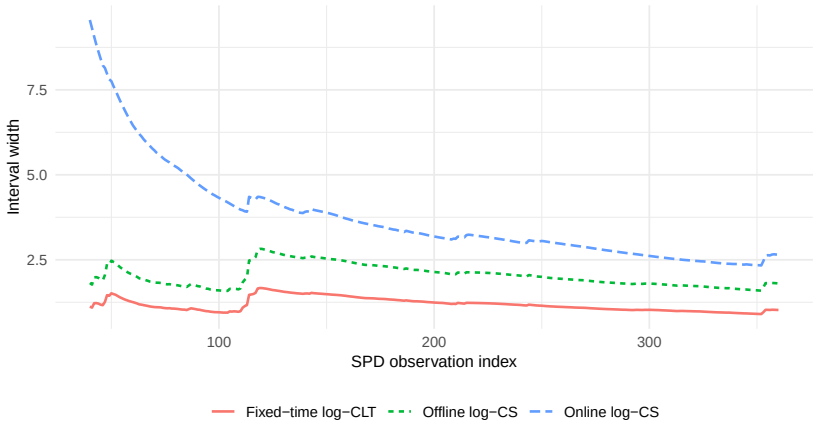

Figure 5: Interval widths for the EEG covariance application; the fixed-time log-CLT interval is narrower because it is pointwise, whereas the log-CS intervals are designed for sequential monitoring.

The multiplicative form of the intervals makes their relative precision explicit. For the offline procedure, the ratio of the upper to the lower endpoint is

$$\frac{\sup \mathcal{C}_n^+}{\inf \mathcal{C}_n^+} = \exp\{2\widehat{\tau}_n b_\alpha(n)/n\},$$

and the same identity holds online with $\widehat{\tau}_n^{\mathrm{on}}$. This ratio separates relative uncertainty from the current level of dispersion, complementing the absolute widths in Figure 5. For a chosen tolerance $\varepsilon > 0$, the condition $\widehat{\tau}_n b_\alpha(n)/n \leq \log(1+\varepsilon)$ places both endpoints within a factor $1+\varepsilon$ of the reported centre. It therefore provides a directly computable precision criterion for sequential monitoring, using the same quantities already stored by the procedure. Under the independent-sampling assumptions of the corresponding theorem, scale consistency and $b_\alpha(n)/n \to 0$ ensure that either specified boundary eventually meets any fixed positive tolerance. Eventual containment gives coverage at data-dependent reporting times with a growing starting time, rather than finite-sample calibration at the chosen burn-in.

6 Concluding Remarks

Positive asymptotic confidence sequences make Fréchet dispersion amenable to repeated monitoring. The offline profiling expansion and the online predictable-loss decomposition isolate the same leading scalar fluctuation, while RSGD supplies a practical one-pass centre update. Their combination links uncertainty quantification for non-Euclidean data to established scalar sequential boundaries.

The theory identifies local smoothness, curvature and tracking as the quantities governing this transfer. Geometry-specific concentration and subgradient methods offer directions toward finite-sample calibration and nonsmooth spaces; Appendix B develops these directions. For finite-sample calibration, the online decomposition identifies two quantities to bound explicitly: the accumulated squared centre error and the conditional variation of the centred loss differences. The same tracking error controls both, suggesting that geometric concentration bounds could be incorporated without changing the one-pass loss construction. Dependent covariance streams raise a complementary question. A sequential analysis for epoch-level observations would need to account jointly for centre tracking and serial or within-subject dependence in the squared losses. For a stationary stream, serial covariances contribute to the fluctuation scale of accumulated squared losses, while their marginal expectation continues to define the same Fréchet dispersion target. Combining such control with an appropriate long-run variance estimator would extend the dispersion-monitoring interpretation to the ordered structure illustrated by the EEG application.

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

The code and an accompanying `README` file are submitted with the manuscript. We provide a dedicated reproduction script for each figure and table, allowing the reported results to be reproduced individually. The `README` provides instructions for running the code.

REFERENCES

B. Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139:655–673, 2011.

A. Ahidar-Coutrix, T. Le Gouic, and Q. Paris. Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probability Theory and Related Fields*, 177:323–368, 2020.

M. Arnaudon, C. Dombry, A. Phan, and L. Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122:1437–1455, 2012.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178, 2013.

R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31:1–29, 2003.

R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds—II. *The Annals of Statistics*, 33:1225–1259, 2005.

S. Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58:2217–2229, 2013.

D. A. Darling and H. Robbins. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America*, 58:66–68, 1967.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R*. Wiley, Chichester, 2 edition, 2016.

P. Dubey and H.-G. Müller. Fréchet analysis of variance for random objects. *Biometrika*, 106: 803–821, 2019.

M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.

David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118, 1975.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

P. Grünwald, R. de Heide, and W. Koolen. Safe testing. *Journal of the Royal Statistical Society: Series B*, 86:1091–1128, 2024.

S. R. Howard, A. Ramdas, J. D. McAuliffe, and J. S. Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49:1055–1080, 2021.

H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York, 2 edition, 2003.

J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56: 732–753, 2014.

V. Patrangenaru and L. Ellingson. *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, 2015.

X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.

A. Petersen and H.-G. Müller. Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47:691–719, 2019.

A. Ramdas, P. Grünwald, V. Vovk, and G. Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38:576–601, 2023.

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

H. Robbins and D. Siegmund. Boundary crossing probabilities for the Wiener process and sample sums. *The Annals of Mathematical Statistics*, 41:1410–1429, 1970.

H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In J. S. Rustagi (ed.), *Optimizing Methods in Statistics*, pp. 233–257. Academic Press, New York, 1971.

Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R. Wolpaw. BCI2000: A general-purpose brain-computer interface system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.

C. Schötz. Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electronic Journal of Statistics*, 13:4280–4345, 2019.

W. Song and H.-G. Müller. Inference for dispersion and curvature of random objects. *Journal of the American Statistical Association*, 121:729–740, 2026.

J. Ville. *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.

I. Waudby-Smith and A. Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society: Series B*, 86:1–27, 2024.

I. Waudby-Smith, D. T. Arbour, R. Sinha, E. H. Kennedy, and A. Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *The Annals of Statistics*, 52:2613–2640, 2024.

H. Ziezold. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians*, volume A, pp. 591–602, Dordrecht, 1977. Reidel.

A SIMULATION SETTINGS AND ADDITIONAL RESULTS

The coverage experiments use 500 replications, $\alpha = 0.05$, b_α^{GM} with $\rho = 1$, and monitoring at every integer time from $n_0 = 100$. Theorem 1 concerns offline profiling; Theorem 2 concerns predictable online losses. The following settings examine these two mechanisms across geometries.

The principal distributions, targets and terminal sample sizes are specified alongside their results in Section 4. Here we give complementary pointwise trajectories, width and computation summaries, and the complete flat-torus results in Appendix A.2.

Coverage trajectories. Cumulative miscoverage at n is the fraction of replications with at least one exclusion of V between n_0 and n; simultaneous coverage is its complement. Pointwise miscoverage records exclusion at n alone. The main manuscript presents the cumulative trajectories for the principal settings, and the figures below give their pointwise trajectories.

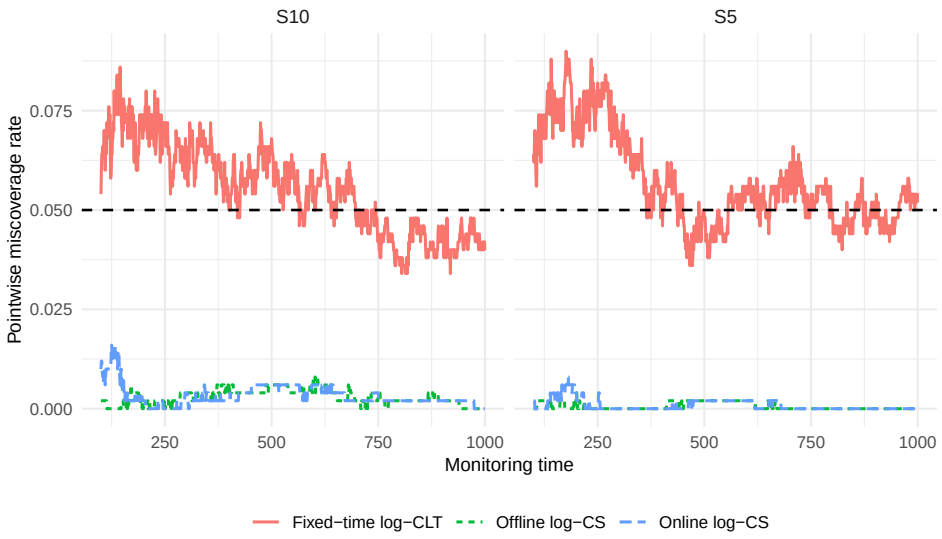

Figure 6: Pointwise miscoverage under repeated monitoring for the spherical simulation settings.

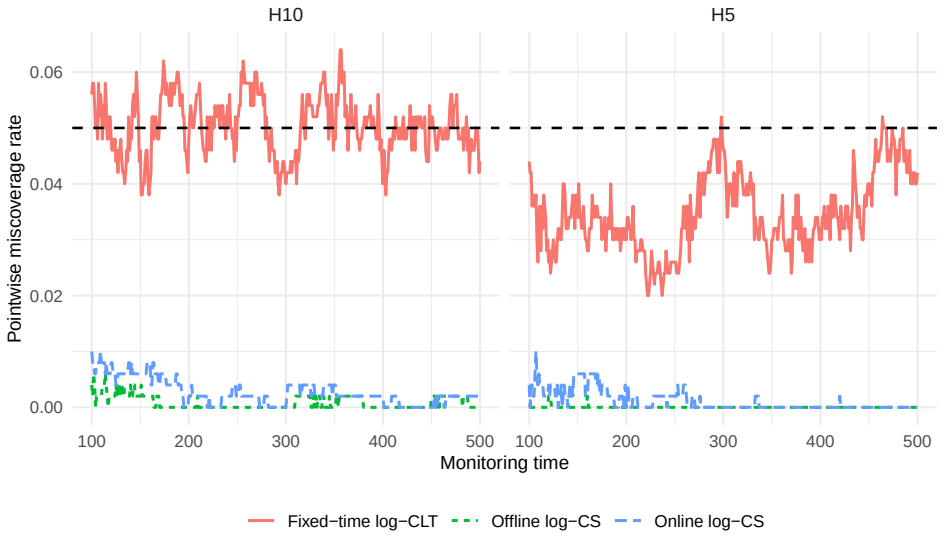

Figure 7: Pointwise miscoverage under repeated monitoring for the hyperbolic simulation settings.

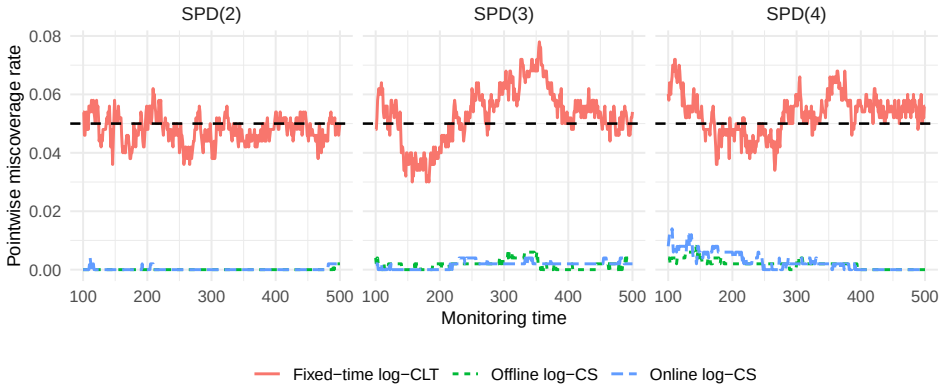

Figure 8: Pointwise miscoverage under repeated monitoring for the affine-invariant SPD simulation settings.

A.1 SUPPORTING WIDTH AND COMPUTATION SUMMARIES

The original half-width and runtime summaries complement the coverage results. Half-widths refer to the reported early and terminal monitoring times; runtime is measured in seconds. The pointwise interval and confidence sequences have different coverage targets. Bold runtime entries are lower than the offline computation time.

Table 3: Supporting diagnostics. Half-width entries are (offline / online) [fixed-time log-CLT]; runtime entries are offline / online.

Setting	Half-width, early	Half-width, final	Runtime (s)
$\mathbb{S}^5$	(0.018 / 0.019) [0.011]	(0.007 / 0.007) [0.004]	0.886 / **0.033**
$\mathbb{S}^{10}$	(0.018 / 0.019) [0.011]	(0.007 / 0.007) [0.004]	1.447 / **0.033**
$\mathbb{H}^5$	(0.035 / 0.037) [0.021]	(0.017 / 0.017) [0.009]	15.841 / **0.011**
$\mathbb{H}^{10}$	(0.035 / 0.036) [0.021]	(0.017 / 0.017) [0.009]	29.664 / **0.011**
$\mathrm{SPD}(2)$	(0.019 / 0.020) [0.011]	(0.009 / 0.009) [0.005]	23.302 / **0.075**
$\mathrm{SPD}(3)$	(0.018 / 0.019) [0.011]	(0.009 / 0.009) [0.005]	24.859 / **0.082**
$\mathrm{SPD}(4)$	(0.019 / 0.019) [0.011]	(0.009 / 0.009) [0.005]	25.675 / **0.086**

A.2 FLAT-TORUS MONITORING

On $\mathbb{T}^q$, $q = 5, 10$, independent local angular coordinates follow $\mathrm{Unif}(-0.60, 0.60)$. The population mean is zero and $V = q\,0.60^2/3$. These settings examine offline profiling in Theorem 1 and predictable online tracking in Theorem 2 within a fixed flat chart. Both centre sequences stay in $[-0.8, 0.8]^q$, whose side lengths are less than π; the offline centre is the coordinatewise arithmetic mean. We use 500 replications and monitor every integer time from $n_0 = 100$ through $N = 1000$, with $\alpha = 0.05$ and the normal-mixture boundary at $\rho = 1$.

Table 4 gives the complete summaries. Both log-CS procedures are closer to nominal coverage than the repeatedly monitored log-CLT baseline, but online coverage is 0.948 in dimension five and 0.888 in dimension ten. The latter shows a substantial coverage shortfall over the finite monitoring horizon, even in this flat geometry. The asymptotic results do not guarantee nominal coverage from $n_0 = 100$.

Table 4: Flat-torus results. Coverage concerns every monitoring time from 100 to 1000. Half-widths are medians of $(U_n - L_n)/2$ across replications at the indicated times; runtimes are median seconds per replication. Bold coverage entries are closer to 0.95 than log-CLT; bold runtimes are lower than offline runtime.

Quantity	Method	$\mathbb{T}^5$	$\mathbb{T}^{10}$
Simultaneous coverage	Offline log-CS	**0.992**	**0.986**
	Online log-CS	**0.948**	**0.888**
	Fixed-time log-CLT	0.700	0.646
Cumulative miscoverage	Offline log-CS	0.008	0.014
	Online log-CS	0.052	0.112
	Fixed-time log-CLT	0.300	0.354
Half-width at $n = 100$	Offline log-CS	0.079	0.110
	Online log-CS	0.091	0.131
	Fixed-time log-CLT	0.047	0.066
Half-width at $n = 1000$	Offline log-CS	0.027	0.039
	Online log-CS	0.028	0.040
	Fixed-time log-CLT	0.015	0.021
Runtime (s)	Offline log-CS	0.068	0.100
	Online log-CS	**0.026**	**0.026**

Figures 9 and 10 show the cumulative and pointwise miscoverage trajectories, respectively, for the same replications.

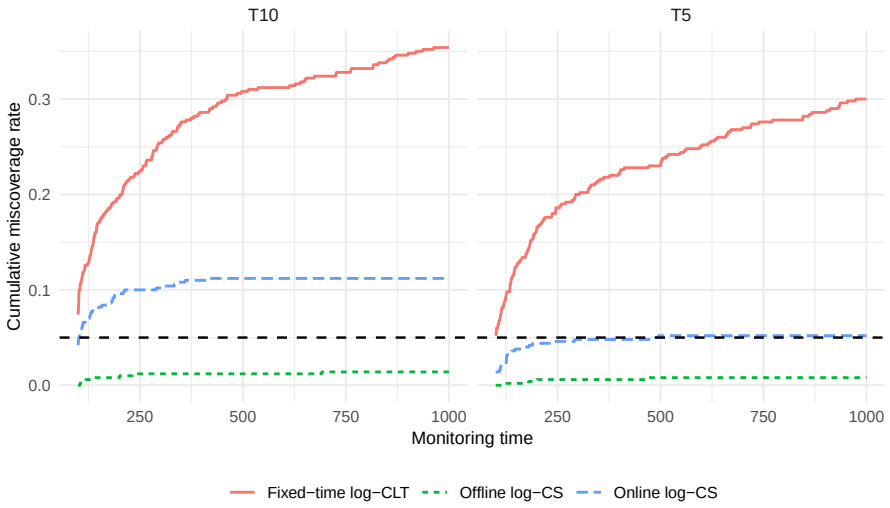

Figure 9: Cumulative miscoverage under repeated monitoring for the flat-torus settings.

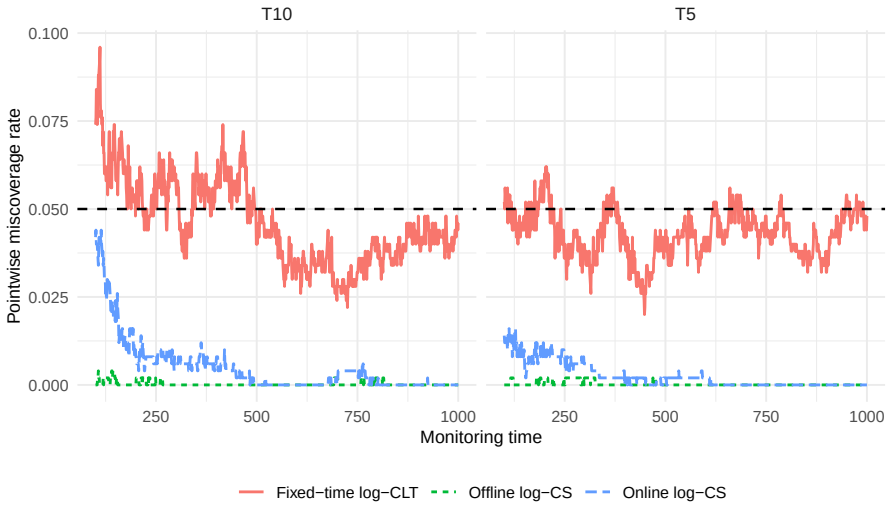

Figure 10: Pointwise miscoverage for the flat-torus settings over the same monitoring times.

B ADDITIONAL DISCUSSION

B.1 FURTHER METHODOLOGICAL DIRECTIONS

The log transformation links dispersion estimation to positive multiplicative intervals. A common scalar influence function connects the offline profiled statistic and the online predictable-loss statistic, while the different remainder bounds describe how each procedure accommodates centre estimation. This separation suggests combining geometry-specific concentration with the same log-scale construction to obtain finite-sample calibration.

For nonsmooth metric spaces, local growth conditions and subgradient tracking may play the roles of Hessian curvature and Riemannian gradient tracking. For dependent streams, a corresponding long-run variance and predictable-loss analysis would connect sequential dispersion monitoring to temporal structure. The EEG application illustrates the latter motivation through an ordered sequence of covariance observations.

B.2 DEPENDENT COVARIANCE STREAMS

The EEG analysis in Section 5 uses epoch-level matrices in subject–run–epoch order. Its trajectory and interval-width plots retain that ordering and display the transient effects of recursive centre estimation. A theory for such streams would combine control of those tracking errors with a variance estimator that reflects serial and within-subject dependence.

C THEORETICAL SUPPLEMENT AND FURTHER THEORY

C.1 REGULARITY CONDITIONS

Let ϕ be a chart from a neighborhood of μ into $\mathbb{R}^p$ with $\phi(\mu) = 0$. For a compact convex set $\Theta \subset \mathbb{R}^p$ containing 0 in its interior, write $m_\theta = \phi^{-1}(\theta)$, $\ell_\theta(x) = d(x, m_\theta)^2$, $Q(\theta) = \mathbb{E}\ell_\theta(X)$ and $Q_n(\theta) = n^{-1}\sum_{i=1}^n \ell_\theta(X_i)$.

Assumption C.1 (Local smooth Fréchet model)**.** The following conditions hold on Θ.

(F1) *Identification.* 0 is the unique minimizer of Q on Θ, $V = Q(0) > 0$, and for every $\eta > 0$, $\inf_{\theta\in\Theta:\|\theta\|\geq\eta} Q(\theta) > Q(0)$.

(F2) *Smooth dominated losses.* For P-almost every x, $\theta \mapsto \ell_\theta(x)$ is three times continuously differentiable on Θ. There are measurable envelopes M_0, M_1, M_2, M_3 such that $\sup_\Theta |\ell_\theta(X)| \leq M_0(X)$, $\sup_\Theta \|\nabla\ell_\theta(X)\| \leq M_1(X)$, $\sup_\Theta \|\nabla^2\ell_\theta(X)\| \leq M_2(X)$ and $\sup_\Theta \|\nabla^3\ell_\theta(X)\| \leq M_3(X)$, with $\mathbb{E}\{M_0^2 + M_0M_1 + M_1^2 + M_2 + M_3\} < \infty$.

(F3) *Local strong convexity.* Q is twice continuously differentiable at 0, $\nabla Q(0) = 0$, and $H = \nabla^2 Q(0)$ is positive definite.

(F4) *Nondegenerate squared-distance variation.* With $\psi(X) = \ell_0(X) - V$, $0 < \sigma^2 = \mathrm{Var}\{\psi(X)\} < \infty$.

Condition (F1) is the separated local identifiability condition that selects the population mean. It is weaker than global convexity. In smooth Riemannian models, it follows locally when the distribution is concentrated away from the cut locus and the expected Hessian of the squared-distance loss is positive definite; sufficient geometric conditions are given in work on intrinsic means and Riemannian centres of mass (Bhattacharya & Patrangenaru, 2003; 2005; Afsari, 2011). Condition (F2) is a finite-dimensional smoothness-and-envelope condition. It is the primitive source of the uniform laws of large numbers used below. Compact support in a strongly convex ball implies bounded envelopes; unbounded models may also satisfy the condition when the distance and its local derivatives have the displayed moments. Condition (F3) is a local property at the mean. The limiting Hessian is H at 0 and $\nabla^2 Q(\theta)$ elsewhere. Condition (F4) ensures positive squared-distance variation.

For the RSGD assumptions, write $h_x(m) = d(x,m)^2/2$, $q(m) = \mathbb{E}h_X(m)$ and $G_x(m) = \nabla_m h_x(m) = -\operatorname{Log}_m x$ whenever the logarithm is defined.

Assumption C.2 (Primitive RSGD regime)**.** The following conditions hold.

(R1) *Geometric localization.* $\mathcal{M}$ is a complete finite-dimensional Riemannian manifold and $K \subset \mathcal{M}$ is compact and strongly geodesically convex. The distribution of X is supported in K, μ lies in the interior of K, and $m \mapsto d(x,m)^2$ is three times continuously differentiable on K for every $x \in K$.

(R2) *Strong convexity and smoothness.* There are constants $\lambda > 0$, $A < \infty$ and $G < \infty$ such that q is geodesically λ-strongly convex on K, $\|G_x(m)\| \leq G$ for all $x, m \in K$, and for all sufficiently small γ,

$$d\{\mathrm{Exp}_m(-\gamma v), \mu\}^2 \leq d(m,\mu)^2 - 2\gamma\langle v, -\operatorname{Log}_m \mu\rangle_m + A\gamma^2\|v\|_m^2 \tag{C.1}$$

whenever $m \in K$ and $\mathrm{Exp}_m(-\gamma v) \in K$.

(R3) *Step sizes.* The RSGD recursion is initialized at $c_0 \in K$ and uses $\gamma_n = \gamma/(n + n_*)$, where $2\lambda\gamma > 1$ and n_* is a large enough constant that (C.1) applies to all steps before projection. Projection onto K is nonexpansive for squared distance to μ.

(R4) *Loss Lipschitzness.* There is $L < \infty$ such that for all $m \in K$, $\mathbb{E}[\{d(X,m)^2 - d(X,\mu)^2\}^2] \leq L^2 d(m,\mu)^2$. Also $d(X,m)^2$ is uniformly bounded on $K \times K$.

These conditions are primitive in the following sense. Compact support in a strongly convex ball gives bounded gradients and bounded losses. The L^2 Lipschitz bound in (R4) follows from the

bounded-gradient part of (R2) by the mean-value theorem along geodesics. The one-step inequality (C.1) follows from a second-order Taylor expansion of the smooth function $z \mapsto d(z,\mu)^2$ on the compact set K. Strong geodesic convexity of q is the population curvature condition for the Fréchet mean; on a sufficiently small regular geodesic ball it is implied by standard curvature-radius restrictions for Riemannian centres of mass (Afsari, 2011; Arnaudon et al., 2012). The step-size requirement is the usual harmonic-step condition for strongly convex stochastic approximation. Proposition C.2 derives the almost-sure tracking rate from this step-size requirement and (R1)–(R4).

C.2 Supporting results

Proposition C.1 (Offline standard-error consistency)**.** *Under Assumption C.1 and the chart selection of Theorem 1, $\widehat{\tau}_n^2 \to \tau^2$ almost surely, where $\tau^2 = \sigma^2/V^2$.*

Proposition C.2 (RSGD tracking rate)**.** *Under Assumption C.2, for every $s < 1$, $d(c_n,\mu)^2 = O_{\mathrm{a.s.}}(n^{-s})$. In particular, for every $\varepsilon > 0$, $d(c_n,\mu) = O_{\mathrm{a.s.}}\left(n^{-1/2+\varepsilon}\right)$ and $n^{-1}\sum_{i=1}^n d(c_{i-1},\mu)^2 = O_{\mathrm{a.s.}}\left(n^{-1+\varepsilon}\right)$.*

Lemma C.1 (Consequences of the primitive offline assumptions)**.** *Under Assumption C.1, the following statements hold almost surely.*

(i) $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \to 0$ *and* $\sup_{\theta\in\Theta} |P_n\ell_\theta^2 - P\ell_\theta^2| \to 0$.

(ii) $\sup_{\theta\in\Theta} \|\nabla^2 Q_n(\theta) - \nabla^2 Q(\theta)\| \to 0$.

(iii) Any measurable global minimizer over Θ, $\widehat{\theta}_n \in \arg\min_{\theta\in\Theta} Q_n(\theta)$, satisfies $\widehat{\theta}_n \to 0$. For all sufficiently large n, $\widehat{\theta}_n$ is an interior point of Θ, $\nabla Q_n\left(\widehat{\theta}_n\right) = 0$, and $\|\widehat{\theta}_n\| = O_{\mathrm{a.s.}}\{[\log\log(e^e n)/n]^{1/2}\}$.

The loss and derivative envelopes imply uniform laws through finite covering arguments. The finite score variance permits the coordinatewise LIL. These facts connect the primitive assumptions to the profiling and tracking mechanisms in the main theorems.

Proposition C.3 (Gaussian coupling and asymptotic confidence sequences)**.** *Under Theorem 1, put $T_n = \log \check{V}_n$ and $s_n = \widehat{\tau}_n$; under Theorem 2, take $T_n = \log \check{Y}_n$ and $s_n = \widehat{\tau}_n^{\mathrm{on}}$ instead. On an extension of the probability space there are independent standard normal variables Z_i, with $G_n = \sum_{i=1}^n Z_i$, such that*

$$T_n - \log V = \tau G_n/n + e_n, \qquad e_n = o_{\mathrm{a.s.}}\{(\log\log n/n)^{1/2}\}. \tag{C.2}$$

For either boundary in (1)*, define $h_n = s_n b_\alpha(n)/n$. There exist positive, possibly unobservable radii h_n^* on this coupled space satisfying*

$$\mathbb{P}\{\log V \in [T_n \pm h_n^*] \text{ for all } n \geq n_0\} \geq 1-\alpha, \qquad h_n/h_n^* \longrightarrow 1 \quad \text{a.s.}.$$

Thus the reported log-scale intervals satisfy Definition 1. Exponentiation gives the corresponding positive reference sequences, with asymptotically equivalent distances from the estimated centre to each endpoint. This approximation does not calibrate the uncorrected reported intervals at a fixed burn-in.

The proof is in Appendix C.4.8.

C.3 Geometric examples and further theoretical discussion

Euclidean space. The target, sample variance and running-centre recursion are given in Section 2.4. The finite-fourth-moment and nondegeneracy conditions there imply Assumption C.1; compact support on a convex update domain also gives the stated online regime.

Flat torus away from wrap-around. Let $\mathcal{M} = \mathbb{T}^q$ with angular coordinates in a chart $(-\pi,\pi)^q$. Fix a closed rectangle $B = \prod_{j=1}^q [a_j, b_j] \subset (-\pi,\pi)^q$ with $b_j - a_j < \pi$ for every j, containing the lifted support and with $\mathbb{E}U$ in its interior. Write $U_i \in B$ for the lifted observations and restrict both

centre sequences to B. Every coordinate difference between observations and centres then lies in $(-\pi, \pi)$, so

$$d(x,m)^2 = \|U - M\|^2, \qquad \mu = \operatorname{wrap}(\mathbb{E}U), \qquad V = \mathbb{E}\|U - \mathbb{E}U\|^2.$$

The offline and online estimators are the Euclidean formulas applied to the lifted coordinates, followed by wrapping the mean back to the torus. The RSGD update is

$$C_n = C_{n-1} + \gamma_n (U_n - C_{n-1}), \qquad c_n = \operatorname{wrap}(C_n),$$

with $C_0 \in B$ and $0 < \gamma_n \le 1$, so the convex-combination update stays in B; otherwise project the lifted update onto B. The empirical mean also lies in B, and all squared distances use these fixed lifts.

Spherical geodesic balls. For $\mathbb{S}^q$ with intrinsic distance $d(x,m) = \arccos\left(x^\top m\right)$, work in a geodesic cap $K = \{x : d(x,a) \le r\}$ that avoids the antipodal cut locus and is small enough to ensure strong convexity of the Fréchet functional. For $m, x \in K$, let $\theta = \arccos\left(m^\top x\right)$. The logarithm and exponential are

$$\operatorname{Log}_m x = \frac{\theta}{\sin\theta}(x - \cos\theta\, m), \qquad \operatorname{Exp}_m v = \cos\|v\|\, m + \sin\|v\|\, v/\|v\|.$$

The target is $V = \mathbb{E}\{\arccos\left(X^\top \mu\right)^2\}$, and the empirical mean is any chart-local minimizer of $\sum_i \arccos\left(X_i^\top m\right)^2$. The online update is

$$c_n = \Pi_K\left[\operatorname{Exp}_{c_{n-1}}\{\gamma_n \operatorname{Log}_{c_{n-1}} X_n\}\right].$$

Compact support inside the cap gives smooth envelopes; nondegenerate radial variation is still required. The simulation design in the main paper uses $X = \operatorname{Exp}_\mu(RA)$, so $V = \mathbb{E}R^2$ when A is a unit tangent direction.

Hyperbolic space and the Poincaré ball. In the Poincaré ball model $\mathbb{B}^q = \{x : \|x\| < 1\}$,

$$d_{\mathbb{H}}(x,y) = \operatorname{arcosh}\left\{1 + \frac{2\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right\}.$$

The Fréchet mean is the minimizer of $\mathbb{E}d_{\mathbb{H}}(X,m)^2$ and is unique under the usual Hadamard nonpositive-curvature setting; on a compact ball $K \subset \mathbb{B}^q$ the loss derivatives are bounded. The offline estimator minimizes the empirical hyperbolic squared loss. The online update is again

$$c_n = \Pi_K\left[\operatorname{Exp}_{c_{n-1}}\{\gamma_n \operatorname{Log}_{c_{n-1}} X_n\}\right],$$

using the Poincaré exponential and logarithm maps. In the radial simulation $X = \operatorname{Exp}_0(RU)$ with U a unit tangent direction and $R \sim \operatorname{Unif}(0, 0.60)$, the population mean is 0 by symmetry and $V = \mathbb{E}R^2 = 0.60^2/3$. Compact regular sets supply the required local smoothness.

Affine-invariant SPD matrices. Section 2.4 gives the distance, target, update and compact spectral conditions. The corresponding logarithm and exponential maps are

$$\operatorname{Log}_P Q = P^{1/2}\log(P^{-1/2}QP^{-1/2})P^{1/2}, \qquad \operatorname{Exp}_P U = P^{1/2}\exp(P^{-1/2}UP^{-1/2})P^{1/2}.$$

These formulas yield the explicit recursion in the main text.

Log-Euclidean SPD matrices. Under the log-Euclidean metric,

$$d_{\mathrm{LE}}(P,Q) = \|\log P - \log Q\|_F,$$

the map $P \mapsto U = \log P$ identifies $\mathrm{SPD}(p)$ with the Euclidean space of symmetric matrices. Hence

$$\mu = \exp\{\mathbb{E}\log X\}, \qquad V = \mathbb{E}\|\log X - \mathbb{E}\log X\|_F^2,$$

$$\widehat{\mu}_n = \exp\left(n^{-1}\sum_{i=1}^n \log X_i\right).$$

The online recursion is Euclidean in the log domain,

$$C_n = C_{n-1} + \gamma_n(\log X_n - C_{n-1}), \qquad c_n = \exp(C_n).$$

The same log-scale confidence-sequence formulas apply with the Frobenius squared distance in log coordinates, under the corresponding Euclidean offline conditions and, for the stated online theorem, the compact-support and tracking conditions.

Rotations and product manifolds. For $\mathrm{SO}(3)$ with $d(R,S) = \|\log\left(R^\top S\right)\|_F/\sqrt{2}$, the local theory applies on balls avoiding rotations of angle π and small enough for strong convexity. The online update is

$$R_n = \Pi_K\{R_{n-1}\exp[\gamma_n \log\left(R_{n-1}^\top X_n\right)]\}.$$

For products $\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_J$ with $d(x,m)^2 = \sum_{j=1}^J w_j d_j(x_j, m_j)^2$, the Fréchet mean and RSGD update are componentwise, while the confidence sequence uses the scalar total squared distance. The assumptions follow if each component satisfies the corresponding local conditions and the total squared-distance variation is nondegenerate.

C.4 Proofs

C.4.1 Proof of Lemma C.1

Proof. We verify the three claims separately. The argument uses the finite-dimensional chart and its derivative envelopes.

Step 1. Uniform law for the loss. Fix $\delta > 0$ and choose a finite Euclidean δ-net $\{\theta_1, \ldots, \theta_N\}$ of the compact set Θ. For any $\theta \in \Theta$, choose $j(\theta)$ such that $\|\theta - \theta_{j(\theta)}\| \le \delta$. By the mean-value theorem and (F2),

$$|\ell_\theta(X) - \ell_{\theta_{j(\theta)}}(X)| \le M_1(X)\,\delta.$$

Consequently,

$$|(P_n - P)\left(\ell_\theta - \ell_{\theta_{j(\theta)}}\right)| \le \{P_n M_1 + P M_1\}\delta.$$

The strong law gives $P_n M_1 \to P M_1 < \infty$. Hence the preceding display is bounded by $C\delta$ eventually, uniformly in θ. At each net point, $(P_n - P)\,\ell_{\theta_j} \to 0$ by the strong law, because $|\ell_{\theta_j}| \le M_0$ and $P M_0 < \infty$. Therefore

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} |(P_n - P)\,\ell_\theta| \le C\delta.$$

Letting $\delta \downarrow 0$ proves $\sup_\Theta |Q_n - Q| \to 0$.

Step 2. Uniform law for the squared loss. The same net can be used for ℓ_θ^2. Since $|\ell_\theta| \le M_0$ and $|\ell_\theta - \ell_\eta| \le M_1\|\theta - \eta\|$,

$$\begin{aligned}|\ell_\theta(X)^2 - \ell_\eta(X)^2| &= |\ell_\theta(X) - \ell_\eta(X)|\,|\ell_\theta(X) + \ell_\eta(X)| \\ &\le 2M_0(X)\,M_1(X)\,\|\theta - \eta\|.\end{aligned}$$

The envelope $M_0 M_1$ is integrable by (F2). Applying the preceding finite-net argument to the functions ℓ_θ^2 gives $\sup_\Theta |P_n\ell_\theta^2 - P\ell_\theta^2| \to 0$.

Step 3. Uniform law for the Hessian. Apply the mean-value theorem componentwise to $\nabla^2\ell_\theta$. For a fixed matrix norm,

$$\|\nabla^2\ell_\theta(X) - \nabla^2\ell_\eta(X)\| \le C M_3(X)\,\|\theta - \eta\|.$$

At net points, the strong law applies to each matrix entry of $\nabla^2\ell_{\theta_j}(X)$ because $\|\nabla^2\ell_{\theta_j}(X)\| \le M_2(X)$ and $P M_2 < \infty$. The same equicontinuity bound with envelope M_3 gives

$$\sup_{\theta\in\Theta} \|P_n\nabla^2\ell_\theta - P\nabla^2\ell_\theta\| \to 0.$$

Dominated convergence also implies continuity of $\theta \mapsto P\nabla^2\ell_\theta$. Since differentiation may be interchanged with expectation under the envelope in (F2), $P\nabla^2\ell_\theta = \nabla^2 Q(\theta)$ and $P_n\nabla^2\ell_\theta = \nabla^2 Q_n(\theta)$. This proves the Hessian convergence.

Step 4. Localization of the empirical minimizer. Let $\eta > 0$ be such that the ball $\{\theta : \|\theta\| \le \eta\}$ is contained in Θ. By (F1),

$$\Delta_\eta = \inf_{\theta\in\Theta:\|\theta\|\ge\eta} Q(\theta) - Q(0) > 0.$$

On the event from Step 1, $\sup_\Theta |Q_n - Q| < \Delta_\eta/3$ for all sufficiently large n. For those n and every θ with $\|\theta\| \geq \eta$,

$$\begin{aligned} Q_n(\theta) &\geq Q(\theta) - \Delta_\eta/3 \\ &\geq Q(0) + 2\Delta_\eta/3 \\ &> Q_n(0). \end{aligned}$$

Thus no minimizer over Θ can lie outside the η-ball. Since η is arbitrary, every measurable global minimizer over Θ satisfying the displayed minimization relation converges to 0.

Step 5. First-order condition and score size. Because 0 is an interior point of Θ and $\widehat{\theta}_n \to 0$, the minimizer is eventually interior. Differentiability of Q_n then yields $\nabla Q_n\left(\widehat{\theta}_n\right) = 0$. At 0,

$$\nabla Q_n(0) = \frac{1}{n}\sum_{i=1}^n \nabla \ell_0(X_i), \qquad \mathbb{E}\nabla \ell_0(X) = \nabla Q(0) = 0.$$

The envelope condition gives $\mathbb{E}\|\nabla \ell_0(X)\|^2 < \infty$. Applying the scalar law of the iterated logarithm to each coordinate and combining the coordinates gives

$$\|\nabla Q_n(0)\| = O_{\text{a.s.}}\left\{\left(\frac{\log\log(e^e n)}{n}\right)^{1/2}\right\}.$$

Step 6. Rate of the empirical minimizer. Use the integral form of the mean-value theorem for the score:

$$0 = \nabla Q_n\left(\widehat{\theta}_n\right) = \nabla Q_n(0) + \bar{H}_n\widehat{\theta}_n, \qquad \bar{H}_n = \int_0^1 \nabla^2 Q_n\left(t\widehat{\theta}_n\right) dt.$$

Step 3 and the continuity of $\nabla^2 Q$ imply $\bar{H}_n \to H$. Since H is positive definite, the smallest eigenvalue of $\bar{H}_n$ is bounded away from zero for all large n. Hence

$$\|\widehat{\theta}_n\| \leq \|\bar{H}_n^{-1}\|\,\|\nabla Q_n(0)\| = O_{\text{a.s.}}\left\{\left(\frac{\log\log(e^e n)}{n}\right)^{1/2}\right\}.$$

This completes the proof of the lemma.

□

C.4.2 Proof of Theorem 1: offline linearization

Proof. Set $r_n = \{\log\log(e^e n)/n\}^{1/2}$. By Lemma C.1, $\|\widehat{\theta}_n\| = O_{\text{a.s.}}(r_n)$ and $\nabla Q_n\left(\widehat{\theta}_n\right) = 0$ eventually.

Step 1. Profile the criterion at the empirical minimizer. Taylor's theorem with integral remainder gives

$$\begin{aligned} Q_n\left(\widehat{\theta}_n\right) &= Q_n(0) + \nabla Q_n(0)^\top \widehat{\theta}_n + \frac{1}{2}\widehat{\theta}_n^\top \widetilde{H}_n \widehat{\theta}_n, \\ \widetilde{H}_n &= 2\int_0^1 (1-t)\,\nabla^2 Q_n\left(t\widehat{\theta}_n\right) dt. \end{aligned}$$

The uniform Hessian convergence in Lemma C.1 implies $\sup_{\|u\| \leq \|\widehat{\theta}_n\|} \|\nabla^2 Q_n(u)\| = O_{\text{a.s.}}(1)$, so the quadratic term is $O_{\text{a.s.}}\left(\|\widehat{\theta}_n\|^2\right) = O_{\text{a.s.}}\left(r_n^2\right)$.

For the linear term, Step 5 in the proof of Lemma C.1 gives $\|\nabla Q_n(0)\| = O_{\text{a.s.}}(r_n)$. Therefore

$$|\nabla Q_n(0)^\top \widehat{\theta}_n| \leq \|\nabla Q_n(0)\|\,\|\widehat{\theta}_n\| = O_{\text{a.s.}}\left(r_n^2\right).$$

Combining the two bounds,

$$Q_n\left(\widehat{\theta}_n\right) = Q_n(0) + O_{\text{a.s.}}\left(r_n^2\right).$$

Step 2. Separate the empirical fluctuation from the profiling error. Since $Q_n\left(\widehat{\theta}_n\right) = \widehat{V}_n$ and $Q_n(0) = P_n\ell_0$,

$$\begin{aligned}\widehat{V}_n - V &= P_n\ell_0 - V + O_{\text{a.s.}}\left(r_n^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{\ell_0(X_i) - V\} + O_{\text{a.s.}}\left(r_n^2\right). \end{aligned} \tag{C.3}$$

The scalar law of the iterated logarithm applied to $\psi = \ell_0 - V$ gives $P_n\ell_0 - V = O_{\text{a.s.}}(r_n)$. Hence $\widehat{V}_n \to V$ almost surely.

Step 3. Remove the ridge. Because $V > 0$ and $\widehat{V}_n \to V$, there is an almost surely finite random integer N such that $\widehat{V}_n > V/2$ for all $n \geq N$. Since $a_n = n^{-2} \to 0$, increasing N if necessary gives $a_n < V/2$ for all $n \geq N$. Therefore $\check{V}_n = \max\left(\widehat{V}_n, a_n\right) = \widehat{V}_n$ eventually almost surely.

Step 4. Pass to the logarithmic scale. On the same eventual event, Taylor's theorem for $g(x) = \log x$ around V yields

$$\log\widehat{V}_n - \log V = \frac{\widehat{V}_n - V}{V} - \frac{\left(\widehat{V}_n - V\right)^2}{2\xi_n^2}$$

for some ξ_n between $\widehat{V}_n$ and V. Since $\xi_n \geq V/2$ eventually and $\widehat{V}_n - V = O_{\text{a.s.}}(r_n)$,

$$\log\widehat{V}_n - \log V = \frac{\widehat{V}_n - V}{V} + O_{\text{a.s.}}\left(r_n^2\right).$$

Substituting (C.3) and using $r_n^2 = \log\log\left(e^e n\right)/n$ proves (5).

□

C.4.3 Proof of Proposition C.1

Proof. We show convergence of the empirical first and second moments at the random centre.

Step 1. First empirical moment. By Lemma C.1, $\widehat{\theta}_n \to 0$ and $\sup_\Theta |P_n\ell_\theta - P\ell_\theta| \to 0$. Hence

$$|P_n\ell_{\widehat{\theta}_n} - P\ell_0| \leq \sup_{\theta\in\Theta}|P_n\ell_\theta - P\ell_\theta| + |P\ell_{\widehat{\theta}_n} - P\ell_0|.$$

The first term tends to zero. For the second term, the mean-value theorem gives $|\ell_{\widehat{\theta}_n}(X) - \ell_0(X)| \leq M_1(X)\|\widehat{\theta}_n\|$, and dominated convergence gives $P\ell_{\widehat{\theta}_n} \to P\ell_0 = V$. Thus $P_n\ell_{\widehat{\theta}_n} \to V$.

Step 2. Second empirical moment. Similarly,

$$|P_n\ell_{\widehat{\theta}_n}^2 - P\ell_0^2| \leq \sup_{\theta\in\Theta}|P_n\ell_\theta^2 - P\ell_\theta^2| + |P\ell_{\widehat{\theta}_n}^2 - P\ell_0^2|.$$

The uniform term tends to zero by Lemma C.1. For the population term,

$$|\ell_{\widehat{\theta}_n}(X)^2 - \ell_0(X)^2| \leq 2M_0(X)M_1(X)\|\widehat{\theta}_n\|,$$

with integrable envelope $2M_0M_1$. Therefore $P\ell_{\widehat{\theta}_n}^2 \to P\ell_0^2$.

Step 3. Variance and log-scale standard error. The variance estimator satisfies

$$\widehat{\sigma}_n^2 = P_n\ell_{\widehat{\theta}_n}^2 - \left(P_n\ell_{\widehat{\theta}_n}\right)^2 \longrightarrow P\ell_0^2 - V^2 = \sigma^2.$$

Theorem 1 gives $\check{V}_n \to V > 0$. Dividing by $\check{V}_n^2$ gives $\widehat{\tau}_n^2 \to \sigma^2/V^2$.

□

C.4.4 PROOF OF THEOREM 1: OFFLINE COVERAGE

Proof. Let $\vartheta = \log V$ and $\tau = \sigma/V$. We prove eventual almost-sure containment under (2); Proposition C.3 establishes Definition 1 for the two boundaries in (1).

Step 1. Write the centred statistic. Theorem 1 gives

$$n\left(\log \check{V}_n - \vartheta\right) = V^{-1}\sum_{i=1}^{n}\psi\left(X_i\right) + R_n, \qquad R_n = O_{\text{a.s.}}\{\log\log\left(e^e n\right)\}. \tag{C.4}$$

Step 2. Control the leading partial sum. The scalar law of the iterated logarithm gives

$$\limsup_{n\to\infty} |\sum_{i=1}^{n}\psi\left(X_i\right)| / \{2\sigma^2 n\log\log n\}^{1/2} = 1 \quad \text{almost surely.}$$

For the finite-LIL boundary in (1), the constant 1.7 is larger than $\sqrt{2}$; hence there exists a deterministic $\eta > 0$ and an almost surely finite random integer N_1 such that

$$|V^{-1}\sum_{i=1}^{n}\psi\left(X_i\right)| \le (1-3\eta)\,\tau b_\alpha\left(n\right), \qquad n \ge N_1. \tag{C.5}$$

For the normal-mixture boundary, $b_\alpha^{\mathrm{GM}}\left(n;\rho\right)/\{n\log\log n\}^{1/2} \to \infty$, so (C.5) holds as well, for any fixed small $\eta > 0$. More generally, the strict lower limit in (2) gives such a deterministic η for every boundary covered by the theorem.

Step 3. Absorb the deterministic remainder and the estimated standard error. For either boundary in (1), $\log\log n = o\{b_\alpha\left(n\right)\}$. Therefore (C.4) and (C.5) imply

$$|n\left(\log \check{V}_n - \vartheta\right)| \le (1-2\eta)\,\tau b_\alpha\left(n\right)$$

for all sufficiently large n. Proposition C.1 gives $\widehat{\tau}_n \to \tau$, so $\widehat{\tau}_n \ge (1-\eta)\,\tau$ eventually. Hence

$$|\log \check{V}_n - \vartheta| \le \frac{\widehat{\tau}_n b_\alpha\left(n\right)}{n}$$

for all sufficiently large n. Thus $\vartheta \in \mathcal{C}_n^{\log}$ eventually almost surely, proving the tail-coverage assertion.

Step 4. Exponentiate. Because $\exp$ is strictly increasing,

$$V \in \mathcal{C}_n^{+} \quad \Longleftrightarrow \quad \log V \in \mathcal{C}_n^{\log}.$$

Moreover $\mathcal{C}_n^{+} \subset (0,\infty)$ for every n. This proves the theorem.

□

C.4.5 PROOF OF PROPOSITION C.2

Proof. Let $R_n = d\left(c_n,\mu\right)^2$ and let $\mathcal{F}_n = \sigma\left(X_1,\ldots,X_n\right)$. The proof follows the classical Robbins–Siegmund route, but we spell out the conditional recursion because this is where the Riemannian geometry enters.

Step 1. One-step conditional contraction. In Algorithm 2, the unprojected step is

$$\mathrm{Exp}_{c_{n-1}}\{-\gamma_n G_{X_n}\left(c_{n-1}\right)\},$$

because $G_x\left(m\right) = -\operatorname{Log}_m x$. Projection is nonexpansive for squared distance to μ by (R3). Thus (C.1), applied with $m = c_{n-1}$ and $v = G_{X_n}\left(c_{n-1}\right)$, gives

$$\begin{aligned} R_n \le R_{n-1} &- 2\gamma_n\langle G_{X_n}\left(c_{n-1}\right), -\operatorname{Log}_{c_{n-1}}\mu\rangle_{c_{n-1}} \\ &+ A\gamma_n^2\|G_{X_n}\left(c_{n-1}\right)\|^2. \end{aligned}$$

On the other hand, geodesic λ-strong convexity of q by (R2) and the fact that μ minimizes q imply the one-sided curvature inequality

$$\langle \nabla q(m), -\operatorname{Log}_m \mu \rangle_m \geq \lambda d(m,\mu)^2, \qquad m \in K. \tag{C.6}$$

Conditioning on $\mathcal{F}_{n-1}$ and using $\mathbb{E}\{G_X(c_{n-1}) \mid \mathcal{F}_{n-1}\} = \nabla q(c_{n-1})$, (C.6), and $\|G_x(m)\| \leq G$, we obtain

$$\mathbb{E}(R_n \mid \mathcal{F}_{n-1}) \leq (1 - 2\lambda\gamma_n) R_{n-1} + AG^2\gamma_n^2. \tag{C.7}$$

Step 2. Introduce the polynomial weight. Fix any $s < 1$. Since $2\lambda\gamma > 1$, $s < 2\lambda\gamma$. Let $w_n = (n + n_*)^s$ and $Z_n = w_n R_n$. Multiplying (C.7) by w_n gives

$$\mathbb{E}(Z_n \mid \mathcal{F}_{n-1}) \leq \frac{w_n}{w_{n-1}} (1 - 2\lambda\gamma_n) Z_{n-1} + AG^2 w_n \gamma_n^2. \tag{C.8}$$

Because $\gamma_n = \gamma/(n + n_*)$ and $w_n/w_{n-1} = 1 + s/(n + n_* - 1) + O\left(n^{-2}\right)$,

$$\frac{w_n}{w_{n-1}} (1 - 2\lambda\gamma_n) \leq 1 - \frac{b}{n}$$

for some $b > 0$ and all sufficiently large n. Also $w_n\gamma_n^2 \leq Cn^{s-2}$ and $\sum_n n^{s-2} < \infty$.

Step 3. Apply Robbins–Siegmund. For all large n, (C.8) has the form

$$\mathbb{E}(Z_n \mid \mathcal{F}_{n-1}) \leq (1 - b/n) Z_{n-1} + Cn^{s-2}.$$

Since $\sum_n Cn^{s-2} < \infty$ and $\sum_n b/n = \infty$, the almost-supermartingale theorem of Robbins & Siegmund (1971) implies that Z_n converges to a finite random limit. In particular, $Z_n = O_{\text{a.s.}}(1)$ and

$$d(c_n, \mu)^2 = R_n = O_{\text{a.s.}}\left(n^{-s}\right). \tag{C.9}$$

Step 4. Derive the two displayed consequences. Taking $s = 1 - 2\varepsilon$ in (C.9) gives $d(c_n, \mu) = O_{\text{a.s.}}\left(n^{-1/2+\varepsilon}\right)$. For the averaged squared error, take $s = 1 - \varepsilon/2$. Then $R_i = O_{\text{a.s.}}\left(i^{-1+\varepsilon/2}\right)$, so

$$\frac{1}{n}\sum_{i=1}^{n} d(c_{i-1}, \mu)^2 \leq \frac{C}{n}\sum_{i=1}^{n} i^{-1+\varepsilon/2} = O_{\text{a.s.}}\left(n^{-1+\varepsilon/2}\right) = O_{\text{a.s.}}\left(n^{-1+\varepsilon}\right).$$

□

C.4.6 Proof of Theorem 2: online linearization and scale consistency

Proof. Write $\ell_m(x) = d(x,m)^2$ and $\ell_\mu(x) = d(x,\mu)^2$. Define the losses evaluated at predictable centres by $Y_i = \ell_{c_{i-1}}(X_i)$. The proof decomposes the moving-centre average into the ordinary empirical average at the population mean plus two negligible terms.

Step 1. Exact decomposition. For each i,

$$\begin{aligned} Y_i - V = {} & \{\ell_\mu(X_i) - V\} + \{P\ell_{c_{i-1}} - P\ell_\mu\} \\ & + \{\ell_{c_{i-1}}(X_i) - \ell_\mu(X_i) - P\left(\ell_{c_{i-1}} - \ell_\mu\right)\}. \end{aligned}$$

Averaging gives

$$\bar{Y}_n - V = P_n\ell_\mu - V + \beta_n + \zeta_n, \tag{C.10}$$

$$\beta_n = \frac{1}{n}\sum_{i=1}^{n}\{P\ell_{c_{i-1}} - P\ell_\mu\},$$

$$\zeta_n = \frac{1}{n}\sum_{i=1}^{n} D_i,$$

$$D_i = \ell_{c_{i-1}}(X_i) - \ell_\mu(X_i) - P\left(\ell_{c_{i-1}} - \ell_\mu\right).$$

Since c_{i-1} is $\mathcal{F}_{i-1}$-measurable and X_i is independent of $\mathcal{F}_{i-1}$, $\mathbb{E}(D_i \mid \mathcal{F}_{i-1}) = 0$.

Step 2. Control the predictable curvature bias. The population functional $F(m) = P\ell_m$ is twice continuously differentiable on the compact regular set. Since μ is its minimizer, $\nabla F(\mu) = 0$. Taylor's theorem along the geodesic from μ to m and boundedness of the Hessian on K give

$$0 \le F(m) - F(\mu) \le C d(m, \mu)^2, \qquad m \in K.$$

Using Proposition C.2, for every $\varepsilon \in (0, 1/2)$,

$$0 \le \beta_n \le \frac{C}{n} \sum_{i=1}^{n} d(c_{i-1}, \mu)^2 = O_{\text{a.s.}}\left(n^{-1+\varepsilon}\right) = o_{\text{a.s.}}(n^{-1/2}). \tag{C.11}$$

Step 3. Bound the conditional variance of the martingale part. By (R4),

$$\begin{aligned} \mathbb{E}\left(D_i^2 \mid \mathcal{F}_{i-1}\right) &\le \mathbb{E}\{[\ell_{c_{i-1}}(X_i) - \ell_\mu(X_i)]^2 \mid \mathcal{F}_{i-1}\} \\ &\le L^2 d(c_{i-1}, \mu)^2. \end{aligned} \tag{C.12}$$

The bounded-loss part of (R4) also gives $|D_i| \le C$. Combining (C.12) with Proposition C.2, for every $\varepsilon > 0$,

$$V_n^D := \sum_{i=1}^{n} \mathbb{E}\left(D_i^2 \mid \mathcal{F}_{i-1}\right) = O_{\text{a.s.}}(n^{\varepsilon}). \tag{C.13}$$

Step 4. Show that the martingale average is negligible. Let $M_m = \sum_{i=1}^{m} D_i$. Fix $\varepsilon \in (0, 1/2)$ and a deterministic $C_0 > 0$. On the event $E_k(C_0) = \{V_{2^k}^D \le C_0 2^{k\varepsilon}\}$, Freedman's maximal inequality for bounded martingales (Freedman, 1975) gives, for any $\eta > 0$,

$$\begin{aligned} &\mathbb{P}\left(\max_{m \le 2^k} |M_m| > \eta 2^{k/2},\ E_k(C_0)\right) \\ &\quad \le 2\exp\left[-\frac{\eta^2 2^k}{2(C_0 2^{k\varepsilon} + C\eta 2^{k/2}/3)}\right]. \end{aligned}$$

Because $\varepsilon < 1/2$, the right-hand side is summable in k. Borel–Cantelli implies that, on the event where $E_k(C_0)$ holds eventually,

$$\max_{m \le 2^k} |M_m| \le \eta 2^{k/2}$$

for all sufficiently large k. Intersecting these probability-one conclusions over positive integers C_0 and positive rational η, and using (C.13) to select a finite pathwise C_0, gives

$$\max_{m \le 2^k} |M_m| = o_{\text{a.s.}}(2^{k/2}). \tag{C.14}$$

If $2^{k-1} < n \le 2^k$, the left side of (C.14) bounds $|M_n|$ and $2^{k/2} \le \sqrt{2n}$. Hence

$$\zeta_n = \frac{M_n}{n} = o_{\text{a.s.}}(n^{-1/2}). \tag{C.15}$$

Step 5. Obtain the variance-scale expansion. Combining (C.10), (C.11) and (C.15),

$$\bar{Y}_n - V = P_n\ell_\mu - V + o_{\text{a.s.}}(n^{-1/2}). \tag{C.16}$$

The scalar law of the iterated logarithm gives $P_n\ell_\mu - V = O_{\text{a.s.}}\{(\log\log n/n)^{1/2}\}$. Thus $\bar{Y}_n \to V > 0$, so $\check{Y}_n = \bar{Y}_n$ eventually almost surely.

Step 6. Pass to the log scale. Taylor's theorem for $\log x$ around V gives, eventually,

$$\log \bar{Y}_n - \log V = \frac{\bar{Y}_n - V}{V} + O_{\text{a.s.}}\{\left(\bar{Y}_n - V\right)^2\}.$$

By (C.16), the square is $O_{\text{a.s.}}(\log\log n/n) = o_{\text{a.s.}}(n^{-1/2})$. Substituting (C.16) proves (6).

Step 7. First moment consistency for the online losses. The convergence $n^{-1}\sum_i Y_i = \bar{Y}_n \to V$ was established in Step 5.

Step 8. Second moment consistency for the online losses. On the compact set K, boundedness and the mean-value theorem give

$$|\ell_{c_{i-1}}(X_i)^2 - \ell_\mu(X_i)^2| \le Cd(c_{i-1},\mu).$$

For any $\varepsilon \in (0,1/2)$, Proposition C.2 yields $d(c_{i-1},\mu) = O_{\text{a.s.}}\left(i^{-1/2+\varepsilon}\right)$. Therefore

$$\frac{1}{n}\sum_{i=1}^{n}|\ell_{c_{i-1}}(X_i)^2 - \ell_\mu(X_i)^2| \le \frac{C}{n}\sum_{i=1}^{n} i^{-1/2+\varepsilon} \longrightarrow 0.$$

The ordinary strong law gives $n^{-1}\sum_i \ell_\mu(X_i)^2 \to P\ell_\mu^2$. Hence $n^{-1}\sum_i Y_i^2 \to P\ell_\mu^2$.

Step 9. Conclude standard-error consistency. Using Steps 7 and 8,

$$\widehat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}_n^2 \longrightarrow P\ell_\mu^2 - V^2 = \sigma^2.$$

Since $\check{Y}_n \to V > 0$, $\widehat{\tau}_n^{\text{on}\,2} = \widehat{s}_n^2/\check{Y}_n^2 \to \sigma^2/V^2$. This completes the proof.

□

C.4.7 Proof of Theorem 2: online coverage

Proof. Let $\vartheta = \log V$ and $\tau = \sigma/V$. The online linearization gives

$$n(\log\check{Y}_n - \vartheta) = V^{-1}\sum_{i=1}^{n}\psi(X_i) + r_n, \qquad r_n = o_{\text{a.s.}}(\sqrt{n}).$$

The scalar LIL and the strict boundary slack in (2) imply that, for some deterministic $\eta > 0$, the absolute value of this expression is at most $(1-2\eta)\tau b_\alpha(n)$ eventually almost surely. Scale consistency gives $\widehat{\tau}_n^{\text{on}} \ge (1-\eta)\tau$ eventually. Consequently,

$$|\log\check{Y}_n - \vartheta| \le \widehat{\tau}_n^{\text{on}} b_\alpha(n)/n$$

eventually almost surely. The probability of containment for every $n \ge m$ therefore tends to one as $m \to \infty$. Since exponentiation is strictly increasing, $\log V \in \mathcal{C}_n^{\text{on,log}}$ is equivalent to $V \in \mathcal{C}_n^{\text{on},+}$, and the latter interval is positive for every n. □

C.4.8 Proof of Proposition C.3

Proof. The variables $\psi(X_i)/\sigma$ are independent, centered and have unit variance. The finite-variance strong approximation described in Waudby-Smith et al. (2024, Section 2.2.2) gives, on an extended space,

$$\sum_{i=1}^{n}\psi(X_i)/\sigma - G_n = o_{\text{a.s.}}\{(n\log\log n)^{1/2}\}.$$

Combining this with (5) or (6) proves (C.2). The same Gaussian sequence works for both procedures when both sets of assumptions hold. For standard Gaussian partial sums, each boundary in (1) satisfies

$$\mathbb{P}\{|G_n| \le b_\alpha(n) \text{ for all } n \ge 1\} \ge 1-\alpha$$

by the normal-mixture and stitched bounds of Howard et al. (2021); Waudby-Smith et al. (2024). The constants 1.7, 0.72 and 10.4 are those of Howard et al. (2021, Eq. (2)); their Eq. (14) gives the normal-mixture boundary. Set $h_n^* = \tau b_\alpha(n)/n + |e_n|$. On this Gaussian coverage event, (C.2) and the triangle inequality give $|T_n - \log V| \le h_n^*$ for every n. Moreover,

$$\frac{h_n}{h_n^*} = \frac{s_n/\tau}{1 + n|e_n|/\{\tau b_\alpha(n)\}} \longrightarrow 1 \quad \text{a.s.},$$

since $s_n \to \tau > 0$ and $e_n = o_{\text{a.s.}}\{b_\alpha(n)/n\}$. This constructs the reference sequence and proves the relative-radius approximation. Both radii tend to zero; hence the variance-scale endpoint distances $e^{T_n}(e^{h_n}-1)$ and $e^{T_n}(1-e^{-h_n})$ have ratio one to their reference counterparts. Monotonicity of exponentiation preserves the reference coverage event. □