LOCALLY DIFFERENTIALLY PRIVATE ONLINE QUANTILE INFERENCE WITH AN ADAPTIVE MECHANISM

Anonymous authors
Paper under double-blind review

ABSTRACT

A local privacy mechanism chosen for frequency estimation need not produce an efficient joint confidence set for quantiles. We select the reporting channel for ellipsoid volume or maximum simultaneous interval width, using one locally private categorical report per user. At the target quantiles, category probabilities are determined by the requested levels, making each candidate's reference score law public. Unknown quantile densities cancel from the volume comparison; an independent private probe supplies them for interval-width selection. Under local smoothness and bounded-library conditions, we establish exact library minimization for volume, a plug-in oracle inequality for width, and conditional asymptotic coverage after selection. A fixed-threshold correction separates localization error from the final score average. Independent strong-privacy experiments give selected joint coverage of 0.93–0.95 and 42%–43% smaller ellipsoid volumes than randomized response under the same correction, including the users reserved for localization. The comparison connects mechanism choice to confidence geometry and quantifies the precision cost of localization.

1 INTRODUCTION

Health surveillance uses quantiles to distinguish the center of a distribution from its tails. For example, Zhao et al. (2023) compare quantiles of self-reported and measured body mass index (BMI) to estimate reporting bias across the distribution. Joint uncertainty matters for such distributional summaries because their quantile estimates are dependent. We study this uncertainty when observations arrive sequentially and remain on users' devices: each user contributes one report satisfying ε-local differential privacy (LDP), and the server updates an estimate of

$$\boldsymbol{Q} = \{Q(\tau_1), \ldots, Q(\tau_k)\}^\top, \qquad Q_j = Q(\tau_j) = F^{-1}(\tau_j), \qquad 0 < \tau_1 < \cdots < \tau_k < 1.$$

Here F is the population distribution function and the requested levels are public. We seek a reporting channel that gives an efficient joint confidence set for this vector.

Private quantile inference and task-adapted mechanism design provide two starting points. Building on stochastic root finding and iterate averaging (Robbins & Monro, 1951; Ruppert, 1988; Polyak & Juditsky, 1992), Liu et al. (2023) obtain single-quantile LDP inference through self-normalization, without estimating the quantile density. Related methods address cumulative distribution function (CDF) inference (Liu et al., 2024), private confidence sets (Waudby-Smith et al., 2023), and time-uniform quantile confidence sequences (Cai et al., 2025). For distributed quantile estimation, Aamand et al. (2025) establish optimal user complexity at small ε for adaptive single-quantile protocols on a finite domain. These methods address estimation accuracy or validity under repeated observation; our objective is the geometry of a joint confidence set at a specified horizon.

Mechanism choice is already central to LDP estimation. Randomized response (Warner, 1965), optimized unary encoding and local hashing (Wang et al., 2017), and Hadamard response (Acharya et al., 2019) have different noise and communication costs. McKenna et al. (2020) optimize linear-query reconstruction, while staircase extremality provides structured candidates for channel optimization (Kairouz et al., 2016). Minimax theory characterizes privacy costs (Duchi et al., 2018; Duchi & Rogers, 2019); Steinberger (2024) develops efficiency theory and adaptive mechanisms for regular parametric models. These reporting and reconstruction principles provide candidate channels. Our

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.

selection criteria measure calibrated quantile uncertainty through ellipsoid volume or simultaneous interval width.

Two obstacles prevent a direct substitution of an existing mechanism selector. First, quantile uncertainty transforms the score covariance by unknown density values. Ellipsoid volume aggregates directions through a determinant, whereas maximum interval width depends on individual scales and correlations. A good frequency-estimation channel need not optimize either objective. Second, a private quantile recursion queries thresholds learned from earlier reports. Its score law changes along the trajectory. Calibrating the noise at the true quantiles does not by itself establish coverage for the moving-threshold estimator.

Our starting observation resolves much of the first obstacle. At the true quantiles, the $k+1$ interval probabilities are $\tau_1, \tau_2 - \tau_1, \ldots, 1 - \tau_k$, independent of the unknown distribution. Unbiased CDF decoding turns one categorical report into scores for all requested quantiles, so every candidate has a publicly computable truth-score law. For ellipsoids, the unknown density determinant is common to all candidates and cancels from the volume comparison. For simultaneous intervals, it does not cancel; a disjoint private probe estimates the coordinate scales. This yields exact volume minimization over a finite library and an oracle inequality for maximum width.

To address the second obstacle, we localize, probe, and then commit to the selected channel on fresh users. A conditional Polyak–Ruppert representation connects its public score covariance to the full online average, even when near-ties make the selected channel vary across pilot samples. We also study a fixed-threshold correction: part of the final stream localizes an anchor and the remaining reports estimate its score drift. Its error separates into centered score noise, a density–localization product, and curvature at that single anchor. Here *adaptive mechanism* means explore–then–commit selection; thresholds may continue to evolve, but the selected channel is fixed. The confidence sets are evaluated at a specified horizon.

The experiments separate the consequences of these two decisions. In strong-privacy normal and mixture designs, fixed-threshold inference has selected Wald coverage 0.938 and 0.930, with volumes 42%–43% smaller than randomized response under the same sample split. Wald selection chooses local hashing, whereas maximum-width selection also chooses a paired staircase channel in the mixture design. The correction reduces bias but uses fewer final inference reports and increases estimation error. Thus the evidence supports matching the mechanism to the requested confidence geometry, with the localization cost accounted for rather than attributed to mechanism selection.

Sections 2 and 3 present the procedures and theoretical guarantees. Section 4 reports numerical results, and Section 5 concludes. Implementation details, additional results, and proofs appear in the appendices.

2 Selecting a mechanism for joint inference

2.1 Quantile scores and mechanism decoding

Let $X_1, X_2, \ldots$ be i.i.d. with distribution function F, and fix public bounds $L < U$. The quantile recursion uses the indicator scores of online private quantile inference (Liu et al., 2023). For ordered thresholds $\boldsymbol{q} \in \mathcal{C} = \{q \in [L,U]^k : q_1 \le \cdots \le q_k\}$, a single category $C(\boldsymbol{q}; X) = \sum_{j=1}^{k} \mathbf{1}\{X > q_j\}$ encodes all k threshold comparisons. Put $d = k+1$ and define the cumulative-sum matrix $P \in \mathbb{R}^{k \times d}$ by $P_{j,c} = \mathbf{1}\{c \le j-1\}$ for $j = 1, \ldots, k$ and $c = 0, \ldots, k$. Thus P maps category probabilities to cumulative probabilities at the thresholds; for the category coordinate vector e_c, $Pe_{C(\boldsymbol{q};X)}$ has coordinates $\mathbf{1}\{X \le q_j\}$. The category probabilities are

$$\boldsymbol{\pi}(\boldsymbol{q}) = \{F(q_1), F(q_2) - F(q_1), \ldots, F(q_k) - F(q_{k-1}), 1 - F(q_k)\}^\top.$$

Write $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)^\top$. At the target,

$$\boldsymbol{\pi}^* = \boldsymbol{\pi}(\boldsymbol{Q}) = \{\tau_1, \tau_2 - \tau_1, \ldots, \tau_k - \tau_{k-1}, 1 - \tau_k\}^\top, \qquad P\boldsymbol{\pi}^* = \boldsymbol{\tau}. \tag{1}$$

The crucial feature is that $\boldsymbol{\pi}^*$ is public even though $\boldsymbol{Q}$ and the density values are unknown.

A finite channel is a column-stochastic matrix $\mathsf{Q} = (\mathsf{Q}_{y,c})_{y \in \mathcal{Y}, 0 \le c \le k}$ with output alphabet $\mathcal{Y}$. It is ε-LDP if

$$\mathsf{Q}_{y,c} \le e^{\varepsilon} \mathsf{Q}_{y,c'} \qquad \text{for all } y, c, c'.$$

A decoder assigns $a_y \in \mathbb{R}^k$ to output y and is collected as $A = (a_y : y \in \mathcal{Y})$.

An unbiased CDF decoder satisfies

$$A\mathsf{Q} = P. \tag{2}$$

Equivalently, $\mathbb{E}(a_Y \mid C = c) = Pe_c$ for every category c. A channel–decoder pair is called unbiased when (2) holds.

Unbiased CDF decoding recovers all k CDF indicators in expectation from one report. At thresholds $\boldsymbol{q}$, the mean of the score $\boldsymbol{\tau} - a_Y$ has coordinates $\tau_j - F(q_j)$, which move each threshold toward its target quantile. Mechanism choice changes the score noise while preserving this common drift.

At the truth, Y^* is the channel output generated at $\boldsymbol{Q}$. The decoded score $\xi^* = \boldsymbol{\tau} - a_{Y^*}$ has mean zero and covariance $\Sigma = \mathrm{Var}(\xi^*)$. Its distribution is determined by the public vector $\boldsymbol{\pi}^*$ and the channel; candidate-specific quantities carry a subscript r.

Candidate mechanisms. Let

$$\mathfrak{M}_n = \{(\mathsf{Q}_r, A_r) : r = 1, \ldots, R_n\}$$

be a finite public or pilot-measurable library. We always include direct d-ary randomized response, denoted r_{RR}. It may additionally contain optimized unary encoding (OUE), optimized local hashing (OLH), Hadamard response (HR), and several staircase channels obtained from distinct public or pilot-weighted workload criteria. The decoder below puts these channels on the same CDF workload; only feasible, nonsingular candidates enter the comparison.

For a channel row $q_y = (\mathsf{Q}_{y,0}, \ldots, \mathsf{Q}_{y,k})^\top$, put

$$r_y = (\boldsymbol{\pi}^*)^\top q_y, \qquad R = \mathrm{diag}(r_y : y \in \mathcal{Y}), \qquad M(\mathsf{Q}) = \mathsf{Q}^\top R^{-1} \mathsf{Q},$$

after removing null outputs.

Theorem 2.1 (Loewner-optimal CDF decoder)**.** *Suppose a finite channel admits a decoder satisfying* (2)*. Then* $M(\mathsf{Q})$ *is positive definite and*

$$A^* = P\{\mathsf{Q}^\top R^{-1} \mathsf{Q}\}^{-1} \mathsf{Q}^\top R^{-1} \tag{3}$$

is the unique decoder minimizing $ARA^\top$ *in the Loewner order. Its truth-score covariance is*

$$\Sigma(\mathsf{Q}) = PM(\mathsf{Q})^{-1} P^\top - \boldsymbol{\tau}\boldsymbol{\tau}^\top. \tag{4}$$

This specializes weighted linear reconstruction for fixed private channels (McKenna et al., 2020) to the CDF workload and the target output probabilities. The decoder minimizes variance in every direction. Equation (4) then supplies the covariance for comparing confidence geometries across channels. Proofs of this result and the staircase reduction are in appendix G.

Staircase optimization supplies workload-adaptive candidates through the convex program in theorem C.1. The full program optimizes a weighted trace loss to generate candidates for the confidence-size comparison. The implementation uses the complementary-pair subclass (PairStair), whose restriction is explicit in appendix A.3. The final comparison uses confidence-set geometry and the quantile covariance $D^{-1}\Sigma_r D^{-1}$, where $D = \mathrm{diag}\{f(Q_1), \ldots, f(Q_k)\}$ and $f = F'$.

2.2 The public frozen-score experiment and inferential losses

For candidate r, generate the truth score from the public report law:

$$\mathbb{P}(Y_r^* = y) = (\boldsymbol{\pi}^*)^\top q_{r,y}, \qquad y \in \mathcal{Y}_r, \qquad \xi_r^* = \boldsymbol{\tau} - a_{r,Y_r^*}. \tag{5}$$

It is fully determined by $(\boldsymbol{\tau}, \varepsilon, \mathsf{Q}_r, A_r)$. This public experiment holds the category probabilities at their target values. The conditional online limit in theorem 3.2 links its covariance to the averaged estimator. For the planned inference report count N, the frozen endpoint is

$$S_{r,N}^\circ = N^{-1/2} \sum_{i=1}^{N} \xi_{r,i}^\circ, \qquad \xi_{r,i}^\circ \overset{\text{i.i.d.}}{\sim} \xi_r^*. \tag{6}$$

All candidates use the same N and confidence level $1 - \alpha$, with $0 < \alpha < 1$. At atoms, an exact reference rule accepts $T < c$ and accepts $T = c$ with public probability η (appendix A.2). The frozen experiment has exact acceptance under this rule; its transfer to the online estimator is established in theorem 3.2.

Calibrated Wald-volume selector. Let

$$T_{r,N}^{W,\circ} = S_{r,N}^{\circ\top} \Sigma_r^{-1} S_{r,N}^{\circ}, \tag{7}$$

and let $(c_{r,N,\alpha}^W, \eta_{r,N,\alpha}^W)$ satisfy (30). The local frozen confidence ellipsoid for the quantile vector has volume

$$\mathrm{Vol}_{r,N}^W(D) = v_k N^{-k/2} \det(D^{-1}) \{c_{r,N,\alpha}^W\}^{k/2} \det(\Sigma_r)^{1/2}, \qquad v_k = \frac{\pi^{k/2}}{\Gamma(k/2+1)}. \tag{8}$$

The unknown factor $\det(D^{-1})$ is common to all candidates. Define

$$J_{r,N}^W = \{c_{r,N,\alpha}^W\}^{k/2} \det(\Sigma_r)^{1/2}, \qquad \widehat{r}^W = \arg\min_{r \in \mathfrak{M}_n} J_{r,N}^W, \tag{9}$$

with deterministic tie breaking. This criterion compares ellipsoid volumes before intersection with $\mathcal{C}$. Density estimation is therefore unnecessary for choosing the Wald mechanism, but is still needed to express its confidence ellipsoid on the quantile scale.

The reference simulation uses only the requested levels, the candidate channel, and the planned report count. For each simulated report, draw its output from $\mathsf{Q}_r \boldsymbol{\pi}^*$ and decode all coordinates together. Repeating this experiment supplies a critical value for each candidate while preserving dependence between quantiles. These are public computations: increasing the number of reference draws improves calibration accuracy without querying additional users. The selected channel is then applied to fresh observations through their current threshold categories.

Rectangle selector. For coordinatewise simultaneous inference, let $(c_{r,N,\alpha}^{\max}, \eta_{r,N,\alpha}^{\max})$ satisfy (30) for

$$\max_{1 \le j \le k} \frac{|S_{r,N,j}^{\circ}|}{\sqrt{(\Sigma_r)_{jj}}}. \tag{10}$$

The largest leading half-width is proportional to

$$J_{r,N}^{\max}(D) = c_{r,N,\alpha}^{\max} \max_{1 \le j \le k} \frac{\sqrt{(\Sigma_r)_{jj}}}{f(Q_j)}. \tag{11}$$

The loss in (11) depends on D and is evaluated at the independent density estimate described next.

For example, a low density at an upper quantile inflates that coordinate's interval and can determine the maximum width even when its score variance is moderate. Rectangle selection then favors channels that reduce uncertainty in that coordinate. Ellipsoid volume aggregates all directions through a determinant, so its ranking is unchanged by the common transformation from score units to quantile units.

As $N \to \infty$, the Wald critical values share the χ_k^2 limit, so the limiting volume criterion is proportional to $\det(\Sigma_r)^{1/2}$. Max critical values instead retain candidate-specific correlations (theorem H.6). The finite-score law thus refines the volume comparison near ties, while the choice between volume and width changes the statistical objective.

2.3 Online explore–then–commit estimation

Algorithm 1 gives the complete procedure. Partition n users into consecutive streams of sizes n_1, n_2, n_3. A randomized-response recursion first localizes the quantiles at its complete average $\widetilde{\boldsymbol{Q}}$. The second stream is split equally between thresholds $\widetilde{\boldsymbol{Q}} \pm \widetilde{h}_n \mathbf{1}_k$; every report estimates all k CDF values. Their decoded means $\overline{a}_+$ and $\overline{a}_-$ give

$$\widehat{f}_j^{\mathrm{raw}} = (\overline{a}_{+,j} - \overline{a}_{-,j})/(2\widetilde{h}_n).$$

Clip these estimates to $[\underline{f}_n, \overline{f}_n]$ and set $\widehat{D} = \mathrm{diag}(\widehat{f}_j)$. The nominal bandwidth h_n is clipped to $\widetilde{h}_n$ near the public bounds and satisfies $h_n \to 0$, $n_2 h_n^2 \to \infty$. The density clipping bounds tend to zero and infinity, respectively. Full bandwidth and step definitions are in appendix A.1.

Fix the estimator branch and its report count before Stage 3: $N = n_3$ for the recursive average, or $N = n_3 - b$ for the fixed-threshold branch below. Minimize $J^W_{r,N}$ or $J^{\max}_{r,N}(\widehat{D})$ accordingly. Starting at $\boldsymbol{q}^m_0 = \widetilde{\boldsymbol{Q}}$, commit to this channel for the final stream and run with step sizes γ^m_s

$$\boldsymbol{q}^m_s = \Pi^{\widehat{B}^{-1}}_{\mathcal{C}}\left[\boldsymbol{q}^m_{s-1} + \gamma^m_s \widehat{B}(\boldsymbol{\tau} - a_{\widehat{r},Y_s})\right], \qquad \overline{\boldsymbol{q}}_{n_3} = n_3^{-1}\sum_{s=1}^{n_3} \boldsymbol{q}^m_{s-1},$$

where $\widehat{B} = \widehat{D}^{-1}$ and $\Pi^W_{\mathcal{C}}$ minimizes $(q-x)^\top W(q-x)$ over $q \in \mathcal{C}$. This is weighted isotonic regression (Barlow et al., 1972). Restarting the step clock at commitment adapts the updates to the newly selected score scale. Each user is queried once. The thresholds and channel assigned to that user depend only on earlier private reports and public randomness, giving user-level ε-LDP for the complete transcript (proposition C.7). The experiments allocate $10\%/20\%/70\%$ to the three streams. The recursive limit is normalized by n_3; sublinear pilot/probe sizes recover the total-sample first-order normalization.

2.4 JOINT CONFIDENCE SETS

For the selected mechanism, write $\Sigma_{\widehat{r}}$ for the public truth-score covariance.

Frozen-score calibrated Wald ellipsoid. Let $(c^W_{\widehat{r},n_3,\alpha}, \eta^W_{\widehat{r},n_3,\alpha})$ be the public rule from (7) and (30). Report

$$\mathcal{C}^W_{1-\alpha} = \left\{\theta \in \mathcal{C} : n_3(\overline{\boldsymbol{q}}_{n_3} - \theta)^\top \widehat{D}\Sigma^{-1}_{\widehat{r}}\widehat{D}(\overline{\boldsymbol{q}}_{n_3} - \theta) \le c^W_{\widehat{r},n_3,\alpha}\right\}, \tag{12}$$

with public boundary randomization when equality occurs.

Frozen-score calibrated simultaneous intervals. Using (10), report the rectangle intersected with the ordered parameter set,

$$\mathcal{C}^{\max}_{1-\alpha} = \mathcal{C} \cap \prod_{j=1}^{k}\left[\overline{q}_{n_3,j} \pm \frac{c^{\max}_{\widehat{r},n_3,\alpha}}{\sqrt{n_3}}\frac{\sqrt{(\Sigma_{\widehat{r}})_{jj}}}{\widehat{f}_j}\right], \tag{13}$$

with public boundary randomization $\eta^{\max}_{\widehat{r},n_3,\alpha}$ when equality occurs. Because multiplication by a positive diagonal matrix does not change correlation, the Gaussian limiting max critical value depends on $\mathrm{Corr}(\Sigma_{\widehat{r}})$ and not on the unknown density vector.

The two geometries serve different reporting tasks. An ellipsoid represents the joint uncertainty in location, spread, and tail summaries, allowing the reader to assess contrasts of quantiles using their dependence. A simultaneous rectangle directly reports an interval for every requested quantile; its maximum half-width measures the least precise coordinate. Minimizing volume may trade precision between directions, whereas minimizing maximum width targets the widest coordinate interval. This distinction explains why a single reporting channel need not be preferred for both summaries.

Localization followed by fixed-threshold inference. To reduce error from the noisy recursive trajectory, average its first b iterates and hold all subsequent thresholds at $\overline{\boldsymbol{q}}_b$. With $N = n_3 - b$ fresh reports, use

$$\check{\boldsymbol{q}}_N = \overline{\boldsymbol{q}}_b + \widehat{D}^{-1}\frac{1}{N}\sum_{s=b+1}^{n_3}(\boldsymbol{\tau} - a_{\widehat{r},Y_s}). \tag{14}$$

Before Stage 3, fix b, select and calibrate at size N, and replace $(\overline{\boldsymbol{q}}_{n_3}, n_3)$ by $(\check{\boldsymbol{q}}_N, N)$ in both confidence sets. This uses the same users and one report per user, allocating b main-stream users to localization. Under the conditions of section 3, $b, N \to \infty$ with N/b bounded gives conditional asymptotic coverage, including conditioning on the localization reports (proposition 3.3).

A score multiplier construction and density-free trajectory normalizers provide alternative confidence sets (equation (34) and appendix E). The trajectory normalizers use the same public score law and permit inference without a density probe.

3 THEORETICAL RESULTS

The results take k and ε as fixed. Assume F is continuous and strictly increasing on $[L, U]$, the quantiles are separated and interior, and f is positive and locally β-Hölder at them for $0 < \beta \leq 1$. The three stream sizes diverge, the library has uniformly bounded cardinality and scores, and its covariance eigenvalues stay bounded away from zero and infinity. Pilot and main steps are proportional to $(t + s_0)^{-a}$ with $\max\{1/2, 1/(1+\beta)\} < a < 1$; density bandwidths satisfy $h_n \to 0$ and $n_2 h_n^2 \to \infty$. These conditions, including bandwidth clipping, are stated precisely in assumptions C.2 to C.4.

The local Hölder condition controls the nonlinear remainder $\sum_t \|\boldsymbol{q}_t - \boldsymbol{Q}\|^{1+\beta}$, and the range in (49) makes it $o_{\mathbb{P}}(\sqrt{n})$. Continuity and positive density at the target are standard conditions for quantile linearization (Bahadur, 1966) and also appear in online private quantile inference (Liu et al., 2023). When the density is locally Lipschitz, $\beta = 1$ and the step-size range becomes $a \in (1/2, 1)$, as in Polyak–Ruppert averaging (Polyak & Juditsky, 1992).

For example, the normal and Gaussian-mixture designs below have positive smooth densities around each of the requested quantiles. A fixed grid with distinct levels gives positive quantile gaps, and public bounds containing the target make it an interior point of $\mathcal{C}$. These properties permit a local linear approximation once the recursion has localized. Small density values still inflate quantile-scale noise through D^{-1}, so the tail and mixture designs test the finite-sample consequences of a condition that holds asymptotically.

The common-shift estimate satisfies

$$\|\operatorname{diag}(\widehat{f}_j^{\mathrm{raw}}) - D\|_{\mathrm{op}} = O_{\mathbb{P}}\Big[(\|\widetilde{\boldsymbol{Q}} - \boldsymbol{Q}\|_\infty + h_n)^\beta + (h_n\sqrt{n_2})^{-1}\Big]$$

and its clipping is asymptotically inactive (theorem C.8). Hence the private probe supplies a consistent density estimate and a bounded positive preconditioner with probability tending to one. For a locally Lipschitz density, $h_n \asymp n_2^{-1/4}$ balances smoothing and probe noise, apart from localization error. The finite-confidence transfer bounds in the appendices use the stronger global conditions in assumptions C.5 and C.6.

3.1 MECHANISM-SELECTION GUARANTEES

The common density factor in (8) makes the Wald selector an exact volume minimizer over the realized library for every positive diagonal D. Including RR gives $J^W_{\widehat{r}^W, n_3} \leq J^W_{r_{\mathrm{RR}}, n_3}$. This is the Wald part of theorem C.9; the comparison is conditional on the realized library and the same report count. For rectangles, density estimation contributes an explicit regret term:

Theorem 3.1 (Plug-in oracle inequality for rectangular selection)**.** *Suppose assumptions C.2 to C.4 hold. Let*

$$r^* \in \arg\min_{r \in \mathfrak{M}_n} J^{\max}_{r,n_3}(D), \qquad \widehat{r}^{\max} \in \arg\min_{r \in \mathfrak{M}_n} J^{\max}_{r,n_3}(\widehat{D}).$$

Then

$$J^{\max}_{\widehat{r}^{\max}, n_3}(D) \leq \min_{r \in \mathfrak{M}_n} J^{\max}_{r,n_3}(D) + 2\Delta_n, \tag{15}$$

where

$$\Delta_n = \max_{r \in \mathfrak{M}_n} \left| J^{\max}_{r,n_3}(\widehat{D}) - J^{\max}_{r,n_3}(D) \right| = O_{\mathbb{P}}(\left\| \widehat{D} - D \right\|_{\mathrm{op}}) = o_{\mathbb{P}}(1). \tag{16}$$

If the oracle minimizer is unique and its gap from the second-best candidate is bounded below by a positive constant, then $\mathbb{P}(\widehat{r}^{\max} = r^*) \to 1$. *If randomized response belongs to the library,*

$$J^{\max}_{\widehat{r}^{\max}, n_3}(D) \leq J^{\max}_{r_{\mathrm{RR}}, n_3}(D) + o_{\mathbb{P}}(1). \tag{17}$$

The oracle inequality controls regret in the inferential criterion and remains informative under ties and near-ties. The uniform error in (16) vanishes with the density-probe error. For the fixed-threshold branch, the same comparison uses $N = n_3 - b$ throughout.

3.2 POST-SELECTION ONLINE ASYMPTOTICS

Let $\mathcal{G}_n$ contain the pilot, probes, and public calibration information. For Stage 3 define

$$h(\boldsymbol{q}) = \boldsymbol{\tau} - \{F(q_1), \dots, F(q_k)\}^\top, \qquad M_s = \{\boldsymbol{\tau} - a_{\widehat{r},Y_s}\} - h(\boldsymbol{q}^m_{s-1}). \tag{18}$$

Then (M_s) is a bounded martingale-difference sequence conditional on $\mathcal{G}_n$.

Theorem 3.2 (Selected online limit and confidence sets)**.** *Under assumptions C.2 to C.4, conditionally on the pilot, probes, and public calibration sigma-field $\mathcal{G}_n$, the complete Stage-3 average has the representation*

$$D\sqrt{n_3}(\overline{\boldsymbol{q}}_{n_3} - \boldsymbol{Q}) = n_3^{-1/2} \sum_{s=1}^{n_3} M_s + o_{\mathbb{P}}(1),$$

uniformly over the finite realized library. The conditional law of $\sqrt{n_3}(\overline{\boldsymbol{q}}_{n_3} - \boldsymbol{Q})$ is approximated, in bounded-Lipschitz distance, by $N_k(0, D^{-1}\Sigma_{\widehat{r}}D^{-1})$. With exact frozen critical rules, the Wald and max sets have conditional coverage converging in probability to $1-\alpha$, and unconditional coverage also converges to $1-\alpha$.

The detailed statements and bootstrap extension are theorems C.10 and C.11. The local Jacobian is $-\widehat{B}D$, so averaging yields influence map $(\widehat{B}D)^{-1}\widehat{B} = D^{-1}$. Thus the public truth-score experiment describes the first-order noise of the online estimator after localization. Conditioning fixes the selected channel and its covariance even if ties make the selected index vary across pilot samples.

Public Monte Carlo calibration introduces an additional, measurable error. For each realized critical rule, let $d_{\mathrm{cal},n}$ be the largest absolute acceptance error over the candidate frozen experiments. Then selected online coverage error is at most $d_{\mathrm{cal},n} + o_{\mathbb{P}}(1)$ (proposition H.7). Empirical calibration with B_{cal} draws controls this component at order $B_{\mathrm{cal}}^{-1/2}$ for a fixed finite library; enumeration gives zero error. This separates public calibration accuracy from the online approximation and motivates the independent frozen holdout reported below.

3.3 FIXED-THRESHOLD INFERENCE

For (14), conditioning also on the localization reports makes the remaining decoded scores independent. Write $e = \overline{\boldsymbol{q}}_b - \boldsymbol{Q}$, $r(e) = h(\boldsymbol{Q} + e) + De$, and $V_s = \boldsymbol{\tau} - a_{\widehat{r},Y_s} - h(\overline{\boldsymbol{q}}_b)$. The exact decomposition is

$$\sqrt{N}\widehat{D}(\breve{\boldsymbol{q}}_N - \boldsymbol{Q}) = \frac{1}{\sqrt{N}} \sum_{s=b+1}^{n_3} V_s + \sqrt{N}(\widehat{D} - D)e + \sqrt{N}r(e). \tag{19}$$

The first term is the centered score average at a single localized threshold vector. Its conditional covariance converges to $\Sigma_{\widehat{r}}$. The remaining terms quantify how density error and local curvature affect the correction. The averaged localization error is $e = O_{\mathbb{P}}(b^{-1/2})$; consistent density estimation and bounded N/b make the product term vanish, and local Hölder smoothness makes the curvature term $O_{\mathbb{P}}(b^{-\beta/2})$. Thus the same Gaussian calibration applies at sample size N (proposition 3.3).

Proposition 3.3 (Conditional inference at fixed thresholds)**.** *Under assumptions C.2 to C.4 with $b, N \to \infty$ and N/b bounded, where b is fixed before Stage 3, the Wald and max sets centered at* (14)*, normalized and calibrated at N, have conditional coverage converging in probability to $1-\alpha$ with exact frozen critical rules. Conditioning includes the localization reports. For a realized Monte Carlo critical rule, the additional calibration error is the corresponding $d_{\mathrm{cal},N}$ from proposition H.7.*

The proof is in appendix H. At fixed n_3, larger b improves the anchor while smaller $N = n_3 - b$ increases the final covariance $D^{-1}\Sigma_{\widehat{r}}D^{-1}/N$. The numerical comparison reports both effects.

4 NUMERICAL EXPERIMENTS

The experiments ask whether selecting for confidence geometry reduces uncertainty, whether fixed thresholds improve the online approximation, and what that correction costs. We report joint coverage of the entire quantile vector at nominal level 0.95, pre-intersection confidence-set size, and the root

Distribution	Final estimator	Coverage: selected / RR		Size: selected / RR	
		Wald	Max	Volume	Half-width
Normal	Recursive average	0.923 / 0.910	0.923 / 0.914	0.570	0.898
	Fixed thresholds	0.938 / 0.949	0.940 / 0.941	**0.584**	**0.907**
Mixture	Recursive average	0.915 / 0.930	0.903 / 0.923	0.548	0.866
	Fixed thresholds	0.930 / 0.935	0.948 / 0.960	**0.567**	**0.863**

Table 1: Joint coverage and confidence-set size at $\varepsilon = 0.5$, $k = 5$, $n = 500000$. Wald (Selected–W) selects by ellipsoid volume; Max (Selected–Max) selects by interval width. Each is compared with fixed RR; rows distinguish recursive averages from fixed-threshold correction. Normal/mixture evaluations use 800/400 replications and 20000-draw calibration. Fixed thresholds use $N = 175000$ final reports versus 350000 recursive-average reports. Sizes precede intersection with $\mathcal{C}$.

mean squared error (RMSE) of an estimator $\widehat{\boldsymbol{q}}$, $[\mathbb{E}\{k^{-1}\|\widehat{\boldsymbol{q}} - \boldsymbol{Q}\|_2^2\}]^{1/2}$. Size ratios compare a selected channel with RR using the *same estimator, density probe, and sample split*. All reported experiments use nonrandomized empirical finite-score critical values; their Monte Carlo error is separate from online coverage error.

4.1 Shared protocol and comparisons

The library contains direct randomized response (RR), optimized unary encoding (OUE), optimized local hashing (OLH), Hadamard response (HR), and PairStair, a trace-optimized channel in a complementary-pair staircase subclass. Every channel uses the decoder in (3). Fixed-channel Wald procedures isolate channel effects. Selected–W and Selected–Max choose among all five channels by Wald volume and maximum interval width, respectively. Both selectors use the $10\%/20\%/70\%$ sample split in (20), reserving users for localization and density probing. Choosing Wald volume does not itself require those density estimates, but constructing its quantile-space ellipsoid does.

The strong-privacy comparison uses $\varepsilon = 0.5$, $n = 500000$, and levels $(0.10, 0.25, 0.50, 0.75, 0.90)$ under $N(0,1)$ and $0.5N(-1, 0.75^2) + 0.5N(1, 0.75^2)$. Fixed-threshold inference uses $b = N = 175000$; the recursive average uses all $n_3 = 350000$ Stage-3 iterates. Both use bandwidth $1.40(U - L)n_2^{-1/4}$, the same clipping and step rules, and paired streams. This sample split and bandwidth were fixed before the independent evaluations: 800 normal and 400 mixture replications. Calibration and independent reference holdouts each use 20000 draws at the relevant report count. Appendix B.1 gives all five channels, paired standard errors, and implementation details.

4.2 Mechanism gains and localization costs

Table 1 shows a clear size reduction within the fixed-threshold procedure: selected/RR volume ratios are 0.584 and 0.567, and maximum half-width ratios are 0.907 and 0.863. Both determinant-only and calibrated Wald selection choose OLH here. Thus, in these two designs, the Wald volume reduction is the gain of fixed OLH over fixed RR under the same protocol; selection supplies no further volume reduction beyond fixed OLH. The 42%–43% volume reduction uses the same privacy and user budgets, with native uplink costs of 5.8 versus 3.0 bits per user. Geometry still matters: max selection chooses OLH for all normal samples but PairStair for 38% of mixture samples, where the coordinatewise objective differs from volume.

Near a tie, calibration uncertainty can exceed the difference between candidates. At $\varepsilon = 1$, $k = 5$, $N = 140000$, the determinant-only OLH/RR volume factor is 0.955. With 20000 calibration draws, the estimated calibrated ratio is 1.011, giving RR, whereas the 500-draw comparison selects OLH. A 95% binomial order-statistic interval for the exact-reference ratio is $[0.931, 1.088]$; the independent 20000-draw holdout gives 0.965 with interval $[0.888, 1.043]$. Both intervals include one, so the observed reversal does not resolve which channel has smaller reference volume. Appendix B.3 gives the calculation and distinguishes reference acceptance from online coverage, including the narrower-bandwidth strong-privacy shortfall.

Fixed thresholds address a different issue. Selected Wald coverage changes from 0.923 to 0.938 for normal data and from 0.915 to 0.930 for the mixture; selected max coverage changes from 0.923 to

Distribution	Channel	RMSE Recursive average	RMSE Fixed thresholds	Native uplink bits / user
Normal	RR	0.0314	0.0449	3.0
	OLH	0.0282	0.0398	5.8
Mixture	RR	0.0341	0.0492	3.0
	OLH	0.0311	0.0450	5.8

Table 2: Estimation error and communication cost in the same strong-privacy evaluation. Uplink averages include pilot/probe reports and the OLH hash identifier. Both procedures use all $n = 500000$ users; fixed-threshold inference allocates 35% of them to final score estimation, versus 70% in the recursive average. All channels fit the common $6.5n$-bit ceiling.

0.940 and from 0.903 to 0.948. The selected fixed-threshold coverages have Monte Carlo standard errors approximately 0.008–0.013. Binomial coverage intervals and paired gain standard errors in appendix B.1 quantify the uncertainty in this comparison. The RR/OLH bias norms fall from $0.020/0.016$ to $0.004/0.006$ for normal data and from $0.016/0.012$ to $0.005/0.002$ for the mixture, consistent with removing accumulated recursive drift in (19).

Fixing thresholds trades precision for bias reduction (table 2). Halving the final report count increases the leading half-width by $\sqrt{2}$ and five-dimensional volume by $2^{5/2} \approx 5.7$ at the same channel and density estimate; calibrated volume factors are 5.3–6.0. RMSE also increases, while OLH retains smaller sets and lower RMSE than RR within each branch.

4.3 Distributional checks and a respondent-data illustration

A complementary 22-configuration study evaluates the recursive average across regular, skewed, heavy-tailed, and mixture distributions, with 500 online replications and 500 calibration draws per configuration. Across the 16 base cells, Selected–W/Selected–Max average coverage is $0.941/0.942$, and Selected–W average RMSE is 0.0118 versus 0.0120 for RR. Sample-size and clustered/nine-level comparisons, together with the full design, are reported in appendix B.2 and tables 4 to 6.

The Behavioral Risk Factor Surveillance System (BRFSS) illustration uses 414633 retained body mass index (BMI) records from its 2024 public file (CDC, 2025). In separate counterfactual deployments, max-width selection chooses OLH at $\varepsilon = 0.5, 1$ and RR at $\varepsilon = 2$ (appendix B.4 and table 7). The target is the empirical distribution of this unweighted finite file. Random ordering samples without replacement; the i.i.d. population coverage theorem therefore provides a calibration model for the illustration. At $\varepsilon = 0.5$, the upper-quantile interval $[35.28, 36.73]$ misses the empirical 0.90 quantile 36.90. The example shows how privacy changes mechanism choice and the precision of upper-tail estimation.

5 Concluding remarks

This paper selects private reporting mechanisms for the geometry of joint quantile uncertainty. The public category law permits exact library comparison for ellipsoid volume, because unknown density factors cancel; independent density probing extends selection to maximum interval width. Conditional online limits justify inference after selection, including near ties between channels. These results match mechanism choice to the uncertainty summary required. Fixed-threshold correction separates final score noise from localization and density errors. In the strong-privacy experiments, it combines coverage near the nominal level with 42%–43% smaller ellipsoid volumes than RR under the same sample split, while quantifying the precision cost of localization.

Reducing the localization fraction would retain more reports for final estimation. For $b = o(n_3)$, this needs uniform rates for the shortened recursion and private density probe that make the remainders in (19) negligible at the $\sqrt{N}$ scale. The allocation balances density–localization error against curvature. Another extension updates the channel from earlier reports while retaining one report per user. Its evolving score covariance and recursive remainders call for martingale Gaussian approximation and multiplier calibration (Belloni and Oliveira, 2018).

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

A. Aamand, F. Boninsegna, A. Gentle, J. Imola, and R. Pagh. Lightweight protocols for distributed private quantile estimation. In *Proceedings of ICML*, volume 267 of PMLR, pages 27–58, 2025.

J. Acharya, Z. Sun, and H. Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *Proceedings of AISTATS*, volume 89 of PMLR, pages 1120–1129, 2019.

R. R. Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966.

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical Inference under Order Restrictions*. Wiley, 1972.

A. Belloni and R. I. Oliveira. A high dimensional central limit theorem for martingales, with applications to context tree models. *arXiv preprint arXiv:1809.02741*, 2018.

V. Bentkus. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113:385–402, 2003.

L. Cai, Q. Hu, J. Sun, and S. Wu. Time-uniform and asymptotic confidence sequence of quantile under local differential privacy. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Centers for Disease Control and Prevention. 2024 BRFSS survey data and documentation. National Center for Chronic Disease Prevention and Health Promotion, 2025. `https://www.cdc.gov/brfss/annual_data/annual_2024.html`.

Centers for Disease Control and Prevention. Adult BMI categories. National Center for Chronic Disease Prevention and Health Promotion, 2024. `https://www.cdc.gov/bmi/adult-calculator/bmi-categories.html`.

X. Chen, J. D. Lee, X. T. Tong, and Y. Zhang. Statistical inference for model parameters in stochastic gradient descent. *Annals of Statistics*, 48(1):251–273, 2020.

K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and Banach spaces. In W. B. Johnson and J. Lindenstrauss, editors, *Handbook of the Geometry of Banach Spaces*, volume 1, pages 317–366. North-Holland, 2001.

H. Dette and C. Graw. Uncertainty quantification by block bootstrap for differentially private stochastic gradient descent. *arXiv preprint arXiv:2405.12553*, 2024.

J. C. Duchi and R. Rogers. Lower bounds for locally private estimation via communication complexity. In *Proceedings of COLT*, volume 99 of PMLR, pages 1161–1191, 2019.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113:182–201, 2018.

Y. Fang, J. Xu, and L. Yang. Online bootstrap confidence intervals for the stochastic gradient descent estimator. *Journal of Machine Learning Research*, 19(78):1–21, 2018.

D. A. Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel Journal of Mathematics*, 50:265–289, 1985.

H. Hotelling. The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2:360–378, 1931.

P. Kairouz, S. Oh, and P. Viswanath. Extremal mechanisms for local differential privacy. *Journal of Machine Learning Research*, 17(17):1–51, 2016.

V. Koltchinskii and S. Mendelson. Bounding the smallest singular value of a random matrix without concentration. *International Mathematics Research Notices*, 2015(23):12991–13008, 2015.

Y. Liu, Q. Hu, and L. Kong. Tuning-free estimation and inference of cumulative distribution function under local differential privacy. In *Proceedings of ICML*, volume 235 of PMLR, pages 31147–31164, 2024.

Y. Liu, Q. Hu, L. Ding, and L. Kong. Online local differential private quantile inference via self-normalization. In *Proceedings of ICML*, volume 202 of PMLR, pages 21698–21714, 2023.

R. McKenna, R. K. Maity, A. Mazumdar, and G. Miklau. A workload-adaptive mechanism for linear queries under local differential privacy. *Proceedings of the VLDB Endowment*, 13(11):1905–1918, 2020.

B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30:838–855, 1992.

M. Raič. A multivariate central limit theorem for Lipschitz and smooth test functions. *arXiv preprint arXiv:1812.08268v2*, 2019.

H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

D. Ruppert. Efficient estimators from a slowly convergent Robbins–Monro process. Technical Report 781, School of Operations Research and Industrial Engineering, Cornell University, 1988.

X. Shao. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society, Series B*, 72:343–366, 2010.

M. Sheshukova, S. Samsonov, D. Belomestny, E. Moulines, Q.-M. Shao, Z.-S. Zhang, and A. Naumov. Gaussian approximation and multiplier bootstrap for stochastic gradient descent. In *Proceedings of AISTATS*, volume 300 of PMLR, pages 1378–1386, 2026.

L. Steinberger. Efficiency in local differential privacy. *Annals of Statistics*, 52(5):2139–2166, 2024.

Y. Sun. A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *The Econometrics Journal*, 16:1–26, 2013.

T. Wang, J. Blocki, N. Li, and S. Jha. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium*, pages 729–745, 2017.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60:63–69, 1965.

I. Waudby-Smith, S. Wu, and A. Ramdas. Nonparametric extensions of randomized response for private confidence sets. In *Proceedings of ICML*, volume 202 of PMLR, pages 36748–36789, 2023.

A. Yu. Zaitsev. Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM: Probability and Statistics*, 2:41–108, 1998.

L. Zhao, S. Park, Z. J. Ward, A. L. Cradock, S. L. Gortmaker, and H. M. Blanck. State-specific prevalence of severe obesity among adults in the US using bias correction of self-reported body mass index. *Preventing Chronic Disease*, 20:230005, 2023. doi:10.5888/pcd20.230005.

Algorithms and implementation appear in appendix A, followed by additional experiments in appendix B. Supporting selection results and alternative inference constructions are in appendices C to F. Proofs are grouped in appendices G to K.

A ALGORITHMS AND IMPLEMENTATION

The main procedure uses the definitions in Section 2. The final estimator and inference size are chosen before Stage 3.

Algorithm 1 Online inference-aware selection for multiple LDP quantiles

Require: Levels $\boldsymbol{\tau}$, privacy ε, bounds $[L, U]$, total n, stage sizes (n_1, n_2, n_3), baseline mechanism r_0, a candidate-library rule, geometry (Wald ellipsoid, simultaneous rectangle, or Haar volume), bandwidth and truncation sequences, step sequences, and confidence level $1-\alpha$. For Wald or max inference, choose the recursive average or fixed-threshold estimator in advance; the latter also requires a localization size $b < n_3$. For Haar volume, fix K, ζ, and any burn-in b before Stage 3.

1: Assign users $1, \dots, n_1$ to (21); compute the full online average $\widetilde{\boldsymbol{Q}}$ in (22).
2: Split the next n_2 users into fixed positive and negative groups; send $\widetilde{\boldsymbol{Q}} \pm \widetilde{h}_n \mathbf{1}_k$ and compute $\widehat{D}$ by (24)–(25). Skip this step only when neither selection nor inference uses D; then set $n_2 = 0$, $n_3 = n - n_1$, and $\widehat{B} = I_k$.
3: Construct the finite candidate library, compute the optimal decoder, verify (27), and calibrate each public frozen score law (5) at the final statistic's report count: $N = n_3 - b$ for fixed thresholds or a retained tail, and $N = n_3$ for the complete recursive average, by the pre-specified public rule in appendix A.2.
4: Select $\widehat{r}$ by (9), (44), or (26), using the same inference size N.
5: Initialize $\boldsymbol{q}_0^m = \widetilde{\boldsymbol{Q}}$ and reset the local step clock.
6: **if** fixed-threshold inference **then**
7: Run (28) for b users and keep its complete average $\overline{\boldsymbol{q}}_b$. Hold the remaining N users' thresholds at $\overline{\boldsymbol{q}}_b$, and accumulate their reports for (14).
8: Report (12) or (13) with center $\breve{\boldsymbol{q}}_N$ and sample size N.
9: **else**
10: Process all n_3 users by (28), updating the complete average, score moments, and optional self-normalizer online. Report (12), (13), (34), or the pre-specified density-free set.
11: **end if**

A.1 COMPLETE THREE-STAGE CONSTRUCTION

Split the user indices deterministically into three consecutive stages,

$$n_1 + n_2 + n_3 = n.$$

The default sample split is

$$(n_1, n_2, n_3) = (\lfloor 0.10n \rfloor, \lfloor 0.20n \rfloor, n - \lfloor 0.10n \rfloor - \lfloor 0.20n \rfloor), \tag{20}$$

with Stage 2 divided as evenly as possible between positive and negative probes. The theory permits any $n_j \to \infty$; choosing $n_1 + n_2 = o(n)$ additionally preserves first-order efficiency relative to the total sample size.

Stage 1: online localization. Use a stable baseline candidate r_0, taken by default to be direct randomized response. Starting from $\boldsymbol{q}_0^p \in \mathcal{C}$, run

$$\boldsymbol{q}_t^p = \Pi_{\mathcal{C}} \left[\boldsymbol{q}_{t-1}^p + \gamma_t^p \{ \boldsymbol{\tau} - a_{r_0, Y_t} \} \right], \qquad t = 1, \dots, n_1, \tag{21}$$

and set

$$\widetilde{\boldsymbol{Q}} = n_1^{-1} \sum_{t=1}^{n_1} \boldsymbol{q}_{t-1}^p. \tag{22}$$

The pilot estimate uses the complete average in (22).

Stage 2: joint density probing. Choose a deterministic nominal bandwidth $h_n \downarrow 0$ and a numerical floor $0 < \ell_n = o(h_n)$, and set

$$\widetilde{h}_n = \max\left[\ell_n, \min\left\{h_n, \frac{\widetilde{Q}_1 - L}{2}, \frac{U - \widetilde{Q}_k}{2}\right\}\right]. \tag{23}$$

Users in the positive group receive $\widetilde{\boldsymbol{Q}} + \widetilde{h}_n \mathbf{1}_k$ and users in the negative group receive $\widetilde{\boldsymbol{Q}} - \widetilde{h}_n \mathbf{1}_k$. Each user sends one report through the baseline channel–decoder pair. Let $\overline{a}_+$ and $\overline{a}_-$ be the two decoded report means and define

$$\widehat{f}_j^{\mathrm{raw}} = \frac{\overline{a}_{+,j} - \overline{a}_{-,j}}{2\widetilde{h}_n}, \qquad j = 1, \ldots, k. \tag{24}$$

To make inversion well defined at every finite sample, use deterministic sequences $\underline{f}_n \downarrow 0$ and $\overline{f}_n \uparrow \infty$ and set

$$\widehat{f}_j = \min\{\overline{f}_n, \max(\underline{f}_n, \widehat{f}_j^{\mathrm{raw}})\}, \qquad \widehat{D} = \mathrm{diag}(\widehat{f}_1, \ldots, \widehat{f}_k). \tag{25}$$

The averages use every probe report. Truncation is applied after averaging to keep the preconditioner positive and numerically stable; theorem C.8 shows that it is asymptotically inactive.

Selection and commitment. After Stage 2, construct any pilot-measurable candidates, compute their optimal decoders and frozen score laws, and select one mechanism. For joint ellipsoids or Haar self-normalization use (9) or (44). For simultaneous rectangles use

$$\widehat{r}^{\max} = \arg\min_{r \in \mathfrak{M}_n} J_{r,n_3}^{\max}(\widehat{D}). \tag{26}$$

The selected channel remains fixed for every Stage-3 user.

Before admission to the library, each candidate is checked by

$$\max_{y,c,c'} \frac{\mathsf{Q}_{r,yc}}{\mathsf{Q}_{r,yc'}} \le e^{\varepsilon}, \qquad \|A_r \mathsf{Q}_r - P\|_{\max} \le \mathrm{tol}, \qquad \lambda_{\min}(\Sigma_r) > \mathrm{tol}\,. \tag{27}$$

Stage 3: restarted preconditioned online recursion. Let $\widehat{r}$ denote the selected mechanism and define $\widehat{B} = \widehat{D}^{-1}$. For positive definite W, let

$$\Pi_{\mathcal{C}}^{W}(x) = \arg\min_{q \in \mathcal{C}} (q - x)^{\top} W (q - x).$$

Starting from $\boldsymbol{q}_0^m = \widetilde{\boldsymbol{Q}}$, run the local-clock recursion

$$\boldsymbol{q}_s^m = \Pi_{\mathcal{C}}^{\widehat{B}^{-1}}\left[\boldsymbol{q}_{s-1}^m + \gamma_s^m \widehat{B}\{\boldsymbol{\tau} - a_{\widehat{r},Y_s}\}\right], \qquad s = 1, \ldots, n_3. \tag{28}$$

The step index restarts at $s = 1$ after commitment. This avoids carrying a very small global step size into a recursion whose mechanism and scale have just changed. The primary estimator is the complete Stage-3 average

$$\overline{\boldsymbol{q}}_{n_3} = n_3^{-1} \sum_{s=1}^{n_3} \boldsymbol{q}_{s-1}^m. \tag{29}$$

The quantitative Haar and integral self-normalized variants may instead use a post-commit burn-in. The weighted projection is diagonal weighted isotonic regression and can be computed by the pool-adjacent-violators algorithm (PAVA) (Barlow et al., 1972). The server stores the current iterate, the average in (29), score first and second moments, and optional bootstrap accumulators.

Communication. For $k = 5$, RR and HR require 3 uplink bits per main-stage report, OUE and paired staircase require 6, and OLH requires 7 at $\varepsilon = 0.5$ and 8 at $\varepsilon \in \{1, 2\}$, including its hash identifier. Pilot and probe reports use 3 bits. Thus every procedure fits a common budget of $6.5n$ uplink bits with 8-bit main-stage slots; native totals are $3n$ for RR/HR, $5.1n$ for OUE/paired staircase, and $5.8n$ or $6.5n$ for OLH. These counts concern user reports; downlink threshold vectors have the same dimension across methods.

Fixed-threshold implementation. Choose b before processing Stage 3, with $b \to \infty$, $N = n_3 - b \to \infty$, and N/b bounded. Select the channel using $J^W_{r,N}$ or $J^{\max}_{r,N}(\widehat{D})$. Run (28) for b users, retain the average $\overline{\boldsymbol{q}}_b$, and use it as the threshold vector for every remaining user. Accumulating their decoded reports gives (14). The recursion and the final accumulation both use $O(k)$ storage; the latter needs no projection or step-size update. The estimator is left unprojected, and the reported confidence sets are intersected with $\mathcal{C}$ as before.

A.2 Public critical values

For a discrete statistic T formed from (6), an exact public critical rule at confidence level $1-\alpha$ uses

$$\mathbb{P}(T < c) + \eta \mathbb{P}(T = c) = 1 - \alpha, \qquad c \in [0, \infty], \quad \eta \in [0, 1]. \tag{30}$$

Boundary randomization gives exact acceptance for the discrete frozen law and is public post-processing.

A.3 Candidate construction and numerical implementation

The experimental library uses direct d-ary RR and enumerates all 2^d binary outputs for OUE, with independent bits having success probability $1/2$ at the input category and $1/(1+e^\varepsilon)$ elsewhere. OLH draws uniformly from the affine hashes with nonzero slope over the smallest prime field containing the d categories, reduces the hash value modulo $g = \max\{2, \min(d, \mathrm{round}(e^\varepsilon)+1)\}$, and reports the hash together with a g-ary randomized response. HR uses the first d nonconstant rows of the Sylvester Hadamard matrix of order $2^{\lceil \log_2(d+1) \rceil}$; output probabilities are proportional to e^ε on positive entries and to one on negative entries. All channels use (3).

PairStair numerically minimizes the trace objective in (38) with $W = I_k$ and equal weights on each complementary pair: $\theta_s = \theta_{(1+e^\varepsilon)\mathbf{1}_d - s}$. The pair probabilities sum to one, with each pair contributing its two complementary rows divided by $1 + e^\varepsilon$. The implementation uses a Frank–Wolfe solver with gap tolerance 10^{-9} times the larger of one and the absolute objective value, and a maximum of 400 iterations. Column-sum and LDP residuals must be at most 2×10^{-8}, the maximum decoder residual at most 10^{-7}, and the smallest covariance eigenvalue greater than 2×10^{-8}.

The simulations initialize each coordinate at the standard normal quantile for its requested level, clipped to $[L + 0.02(U-L), U - 0.02(U-L)]$. They use $\ell_n = 10^{-8}(U-L)/n_2$, $\underline{f}_n = 0.25/\sqrt{n_2}$, and $\overline{f}_n = \max\{2, \log n_2\}$ in the bandwidth and density truncations. The 22-configuration benchmark uses the shared step and bandwidth constants in appendix B.2; the independent fixed-threshold comparison uses the wider bandwidth in appendix B.1.

For moderate k, enumerate the 2^{k+1} staircase vectors and solve (38) by a conic or disciplined convex optimizer. At a nonsingular iterate, the gradient is

$$\frac{\partial}{\partial \theta_s} \operatorname{tr}\{W P H(\theta)^{-1} P^\top\} = -\frac{s^\top H(\theta)^{-1} P^\top W P H(\theta)^{-1} s}{(\boldsymbol{\pi}^*)^\top s}. \tag{31}$$

A candidate solution is accepted only when primal feasibility, dual feasibility, complementarity, and the decoder/privacy checks in (27) all pass. For larger k, use column generation: solve the restricted program, search for a staircase ray with the most negative reduced cost using (31), add it, and repeat. This is a computational device; statistical validity still follows from the final finite channel and exact decoder.

A.4 Score multiplier bootstrap

The multiplier bootstrap (MB) uses centered decoded scores. For Stage-3 users put

$$U_s = \boldsymbol{\tau} - a_{\widehat{r}, Y_s}, \qquad \overline{U} = n_3^{-1} \sum_{s=1}^{n_3} U_s, \qquad \widehat{\Sigma}_U = n_3^{-1} \sum_{s=1}^{n_3} (U_s - \overline{U})(U_s - \overline{U})^\top. \tag{32}$$

Given the transcript, draw $e_1, \ldots, e_{n_3} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and form

$$Z^* = n_3^{-1/2} \sum_{s=1}^{n_3} e_s \widehat{D}^{-1} (U_s - \overline{U}). \tag{33}$$

Let $\widehat{s}_j^2 = (\widehat{D}^{-1}\widehat{\Sigma}_U\widehat{D}^{-1})_{jj}$ and let $c^*_{1-\alpha}$ be the conditional $(1-\alpha)$ quantile of $\max_j |Z_j^*|/\widehat{s}_j$. The bootstrap rectangle is

$$\mathcal{C}^{\mathrm{MB}}_{1-\alpha} = \mathcal{C} \cap \prod_{j=1}^{k} \left[\overline{q}_{n_3,j} \pm \frac{c^*_{1-\alpha}\widehat{s}_j}{\sqrt{n_3}} \right]. \tag{34}$$

The covariance in (32) is updated recursively. At the end, one may draw Z^* directly from $N_k(0, \widehat{D}^{-1}\widehat{\Sigma}_U\widehat{D}^{-1})$, or maintain multiplier sums online. Validity follows from the influence-function representation proved below. Related online and block-bootstrap constructions for stochastic approximation are developed by Fang et al. (2018); Dette & Graw (2024); Sheshukova et al. (2026).

A.5 Algorithms for density-free inference

The fixed-mechanism recursion below supplies the retained trajectory for the self-normalized procedures in Appendix E. Algorithm 3 uses the block-Haar weights and normalizer defined there and in Appendix D.

Algorithm 2 Fixed-mechanism one-report online multiple-quantile estimation

Require: A fixed unbiased channel–decoder pair (Q, A), initial $\boldsymbol{q}_0 \in \mathcal{C}$, steps $\gamma_t = \gamma_0 t^{-a}$, and sample size n.

1: **for** $t = 1, \ldots, n$ **do**
2: Send $\boldsymbol{q}_{t-1}$ to user t; the user forms $C_t = C(\boldsymbol{q}_{t-1}; X_t)$ and sends one draw $Y_t \sim \mathsf{Q}(\cdot \mid C_t)$.
3: Update $\boldsymbol{q}_t = \Pi_{\mathcal{C}}[\boldsymbol{q}_{t-1} + \gamma_t\{\boldsymbol{\tau} - a_{Y_t}\}]$.
4: **end for**
5: Return the Polyak–Ruppert average and the pre-specified retained path used by the selected self-normalized statistic.

Algorithm 3 Haar score-calibrated joint quantile inference

Require: Outputs of algorithm 2, levels $\boldsymbol{\tau}$, channel–decoder pair (Q, A), burn-in $0 \le b < n$, number of contrasts $K \ge k$, confidence level $1-\alpha$, smoothing scale $h_T \ge 0$, calibration size B_{cal}, and calibration mode in $\{\text{exact finite}, \text{Monte Carlo finite}, \text{exchangeable rank}\}$.

1: Set $m = n - b$, require $m \ge 4K$, let $s = \lfloor m/(2K) \rfloor$, and construct the block-Haar weights in (69).
2: From the public vector $\boldsymbol{\pi}^*$, generate frozen experiments with $C_i^* \sim \boldsymbol{\pi}^*$, $Y_i^* \sim \mathsf{Q}(\cdot \mid C_i^*)$, and $\xi_i^* = \boldsymbol{\tau} - a_{Y_i^*}$.
3: In each frozen experiment, form the endpoint coefficient, the K block-Haar covariance coefficients, and the frozen statistic; record $+\infty$ when the normalizer is singular.
4: Calibrate the h_T-smoothed frozen statistic: enumerate its law when feasible; otherwise use an empirical finite-critical rule with reported d_{cal} and a one-sided singular-mass certificate, or use the exchangeable randomized-rank rule. In finite-critical mode, increase (m, K) unless $p_{\infty,m} \le \alpha$ is certified.
5: From the retained online path, compute $\bar{\boldsymbol{q}}_{b,m}$ and $S^{\mathrm{SER}}_{K,m}$ in (89).
6: **if** $S^{\mathrm{SER}}_{K,m}$ is singular **then**
7: **if** exchangeable-rank mode **then**
8: Apply the rank decision at $+\infty$; return $\mathcal{C}$ on acceptance and $\varnothing$ on rejection.
9: **else**
10: Return $\varnothing$ and a rank-deficiency diagnostic; do not use an uncalibrated generalized inverse.
11: **end if**
12: **else**
13: Let c_{eff} be the calibrated cutoff minus the public smoothing draw $h_T U$; apply the calibrated boundary or rank randomization and report

$$\mathcal{C}^{\mathrm{HSC}}_{1-\alpha} = \left\{\theta \in \mathcal{C} : m(\bar{\boldsymbol{q}}_{b,m} - \theta)^\top (S^{\mathrm{SER}}_{K,m})^{-1}(\bar{\boldsymbol{q}}_{b,m} - \theta) \le c_{\mathrm{eff}}\right\}.$$

14: **end if**
15: Report d_{cal}, $\lambda_{\min}(S^{\mathrm{SER}}_{K,m})$, its condition number, the frozen singular probability, and the exact-score and analytic F critical values.

Distribution	Channel	Recursive average / fixed thresholds Wald coverage	RMSE	Fixed-threshold Volume / RR
Normal	RR	0.910 / 0.949	0.0314 / 0.0449	1.000
Normal	OUE	0.932 / 0.943	0.0293 / 0.0417	0.789
Normal	OLH	0.923 / 0.938	0.0282 / 0.0398	0.584
Normal	HR	0.915 / 0.946	0.0305 / 0.0428	0.841
Normal	PairStair	0.904 / 0.951	0.0285 / 0.0390	0.790
Mixture	RR	0.930 / 0.935	0.0341 / 0.0492	1.000
Mixture	OUE	0.925 / 0.945	0.0336 / 0.0471	0.791
Mixture	OLH	0.915 / 0.930	0.0311 / 0.0450	0.567
Mixture	HR	0.882 / 0.938	0.0340 / 0.0490	0.834
Mixture	PairStair	0.912 / 0.930	0.0298 / 0.0423	0.757

Table 3: All five fixed channels in the independent strong-privacy comparison. Normal and mixture designs have 800 and 400 paired replications, respectively. Both procedures use the wider probe bandwidth; the confidence levels are unchanged.

B Additional numerical results

The results below supplement the main comparison: all-channel fixed-threshold performance, the broader recursive benchmark, reference calibration, and the respondent-data illustration.

B.1 Fixed-threshold comparisons under strong privacy

The two independent evaluations in table 1 use $n_1 = 50000$, $n_2 = 100000$, $n_3 = 350000$, $b = N = 175000$, and $\varepsilon = 0.5$. The first half of Stage 3 runs the original recursion. The remaining reports are generated at its average and accumulated for (14). Each candidate's two runs share raw observations, pilot and probe reports, and report-randomization streams. The probe bandwidth is $1.40(U - L)n_2^{-1/4}$, with the original boundary clipping, density truncation, and step rules. Both procedures use estimated densities throughout. Public calibration and independent holdout samples each contain 20000 draws at the actual inference size N; the normal recursive average reuses its reference at n_3.

For the normal distribution, the RR/OLH Euclidean bias norms fall from $0.020/0.016$ to $0.004/0.006$; for the mixture they fall from $0.016/0.012$ to $0.005/0.002$. The paired Wald coverage changes for RR/OLH are $0.039/0.015$ with standard errors $0.011/0.010$ in the normal design, and $0.005/0.015$ with standard errors $0.016/0.017$ in the mixture. Selected max coverage improves by 0.018 and 0.045, with paired standard errors 0.011 and 0.017. The RMSE increase in table 3 reflects the smaller number of final inference reports. The max selector chooses OLH in every normal replication and in 62% of mixture replications, choosing PairStair in the remaining 38%. Across the two public holdouts and five channels, acceptance ranges from 0.946 to 0.954 for Wald and from 0.945 to 0.953 for max statistics.

Conditional on the realized public calibration, 95% exact binomial intervals for selected fixed-threshold coverage are $[0.918, 0.953]$ (Wald) and $[0.921, 0.955]$ (max) in the normal design, and $[0.900, 0.953]$ and $[0.921, 0.967]$ in the mixture. These intervals use 800 and 400 independent online replications, respectively. They quantify the precision of the coverage estimates; the point estimates and intervals retain the possibility of a practically relevant shortfall from 0.95.

Each user still sends one ε-LDP report. RR uses $3n$ native uplink bits and OLH uses $5.8n$, including pilot and probe reports and the hash identifier; both fit the common $6.5n$-bit ceiling in appendix A.1. Fixed-threshold inference preserves these totals and the dimension of the downlink thresholds. It replaces 175000 recursive updates by score additions, with $O(k)$ state throughout. At the same $\widehat{D}$ and channel, the leading volume and width inflation factors are $2^{k/2}$ and $\sqrt{2}$; the public finite-score critical values give realized fixed-threshold/full-average volume ratios of 5.3–6.0 across the five channels and two distributions. The comparison includes all 500000 users, including those used only for localization and density probing.

Distribution	ϵ	RR–W	OUE–W	PairStair–W	Selected–W	Selected–Max
Normal	1	0.910 (0.0216)	0.938 (0.0225)	0.882 (0.0220)	0.912 (0.0209)	0.938 (0.0204)
Normal	2	0.932 (0.0085)	0.910 (0.0111)	0.954 (0.0127)	0.946 (0.0083)	0.950 (0.0084)
Gamma	1	0.922 (0.0219)	0.946 (0.0222)	0.916 (0.0225)	0.950 (0.0207)	0.934 (0.0215)
Gamma	2	0.940 (0.0090)	0.954 (0.0108)	0.958 (0.0131)	0.938 (0.0090)	0.926 (0.0089)
t_3	1	0.954 (0.0183)	0.934 (0.0207)	0.914 (0.0209)	0.948 (0.0191)	0.936 (0.0186)
t_3	2	0.928 (0.0077)	0.942 (0.0093)	0.934 (0.0115)	0.946 (0.0073)	0.940 (0.0076)
Mixture	1	0.932 (0.0228)	0.914 (0.0246)	0.896 (0.0244)	0.942 (0.0230)	0.924 (0.0236)
Mixture	2	0.926 (0.0096)	0.926 (0.0122)	0.926 (0.0137)	0.922 (0.0095)	0.926 (0.0100)

Table 4: Joint coverage with RMSE in parentheses for $k = 5$ and $n = 200{,}000$. RR–W, OUE–W and PairStair–W fix the channel and use Wald calibration; Selected–W/Selected–Max choose a channel by volume/width. Each entry uses 500 replications.

B.2 Distributional, sample-size, and grid comparisons

Four benchmark distributions represent regular, skewed, heavy-tailed, and low-density settings: $N(0,1)$; $(G-4)/2$ for $G \sim \mathrm{Gamma}(4,1)$; $T/\sqrt{3}$ for $T \sim t_3$; and $0.5N(-1,0.75^2) + 0.5N(1,0.75^2)$. The main design crosses these distributions with $\varepsilon \in \{1,2\}$ and the quantile grids

$$(0.25, 0.50, 0.75) \qquad \text{and} \qquad (0.10, 0.25, 0.50, 0.75, 0.90)$$

at $n = 200{,}000$, giving 16 configurations. Four sample-size configurations use the normal and mixture distributions, the five-level grid, $\varepsilon = 1$, and $n \in \{100{,}000, 400{,}000\}$. Two stress configurations use the normal distribution with nine equally spaced levels from 0.10 to 0.90 and the mixture distribution with the clustered grid

$$(0.45, 0.475, 0.50, 0.525, 0.55),$$

respectively; both set $\varepsilon = 1$ and $n = 500{,}000$. Thus the study contains 22 configurations.

The 16 base configurations and six extensions use 500 online replications and 500 public calibration draws each. Public bounds are each distribution's 0.0025 and 0.9975 quantiles, supplied as public design information. The probe bandwidth here is $0.70(U-L)n_2^{-1/4}$, distinct from the wider bandwidth in the independent fixed-threshold comparison. The bounds and all remaining tuning are common to the compared channels. Increasing the sample size from 100000 to 400000 approximately halves selected-method RMSE; nine-level and clustered grids retain coverage between 0.930 and 0.942 with RMSE below 0.015.

Distribution	ϵ	RR–W	OUE–W	PairStair–W	Selected–W	Selected–Max
Normal	1	0.938 (0.0120)	0.904 (0.0152)	0.950 (0.0131)	0.944 (0.0120)	0.938 (0.0123)
Normal	2	0.926 (0.0059)	0.938 (0.0079)	0.928 (0.0081)	0.942 (0.0059)	0.950 (0.0059)
Gamma	1	0.936 (0.0117)	0.944 (0.0138)	0.928 (0.0131)	0.932 (0.0117)	0.964 (0.0117)
Gamma	2	0.938 (0.0059)	0.956 (0.0075)	0.942 (0.0078)	0.938 (0.0055)	0.930 (0.0059)
t_3	1	0.950 (0.0082)	0.954 (0.0100)	0.956 (0.0095)	0.954 (0.0080)	0.952 (0.0084)
t_3	2	0.942 (0.0040)	0.950 (0.0054)	0.956 (0.0055)	0.948 (0.0039)	0.972 (0.0039)
Mixture	1	0.916 (0.0167)	0.920 (0.0199)	0.950 (0.0178)	0.950 (0.0161)	0.958 (0.0165)
Mixture	2	0.936 (0.0079)	0.936 (0.0108)	0.932 (0.0108)	0.936 (0.0082)	0.926 (0.0083)

Table 5: Joint coverage with RMSE in parentheses for $k = 3$ and $n = 200{,}000$. Method labels follow Table 4; each entry uses 500 replications.

The tables display fixed RR, fixed OUE, and PairStair, each paired with a frozen-score Wald ellipsoid; Selected–W pairs volume-based channel selection with that ellipsoid, while Selected–Max pairs width-based selection with simultaneous coordinate intervals. Both selectors search the full five-mechanism library. The step exponent is $2/3$, with pilot and main constants 0.50 and 0.30. The common-shift bandwidth is $0.70(U-L)n_2^{-1/4}$. For each mechanism the local clock offset is the smallest integer for which

$$\gamma_{1,r} \left\| \widehat{B} \right\|_{\mathrm{op}} G_r \le 0.03\widehat{\Delta}, \qquad \widehat{\Delta} = \min\{\widetilde{Q}_1 - L, U - \widetilde{Q}_k, \widetilde{Q}_{j+1} - \widetilde{Q}_j\}.$$

Configuration	RR–W	OUE–W	PairStair–W	Selected–W	Selected–Max
Normal, $k = 5$, $\epsilon = 1$, $n = 100{,}000$	0.928 (0.0285)	0.934 (0.0309)	0.926 (0.0305)	0.902 (0.0289)	0.932 (0.0289)
Mixture, $k = 5$, $\epsilon = 1$, $n = 100{,}000$	0.874 (0.0334)	0.928 (0.0348)	0.912 (0.0329)	0.910 (0.0316)	0.900 (0.0332)
Normal, $k = 5$, $\epsilon = 1$, $n = 400{,}000$	0.918 (0.0149)	0.934 (0.0158)	0.910 (0.0159)	0.920 (0.0148)	0.960 (0.0142)
Mixture, $k = 5$, $\epsilon = 1$, $n = 400{,}000$	0.924 (0.0170)	0.916 (0.0180)	0.938 (0.0165)	0.924 (0.0165)	0.930 (0.0163)
Normal, nine-level grid, $\epsilon = 1$, $n = 500{,}000$	0.946 (0.0159)	0.936 (0.0141)	0.946 (0.0139)	0.940 (0.0135)	0.940 (0.0137)
Mixture, clustered grid, $\epsilon = 1$, $n = 500{,}000$	0.936 (0.0155)	0.950 (0.0158)	0.930 (0.0141)	0.930 (0.0148)	0.942 (0.0144)

Table 6: Sample-size and stress designs. Method labels follow Table 4; entries are joint coverage with RMSE in parentheses, based on 500 replications.

The factor 0.03 sets the permitted initial update to three percent of the smallest pilot separation or boundary margin, so the step scale adapts to both the decoded-score magnitude and the spacing of the target quantiles. All procedures use the complete Stage-3 Polyak–Ruppert average.

B.3 Reference calibration and the online approximation

The reference law and the online estimator answer different coverage questions. A separate paired comparison uses 20000 calibration draws and 400 online replications per cell, the five-level grid, and the narrower probe bandwidth $0.70(U - L)n_2^{-1/4}$. At $\varepsilon = 1$ and $n = 200000$ ($n_3 = 140000$), the determinant-only OLH/RR volume factor is 0.955, but the calibrated level-0.95 OLH/RR volume ratio is 1.011, so calibrated Wald selection chooses RR. For normal data, RR and OLH coverage is 0.940 and 0.923; for the mixture, it is 0.905 and 0.935. In contrast, the separate 500-draw, 100-replication comparison selects OLH with a volume ratio 0.895. The calibration sizes and independent evaluation samples distinguish these comparisons.

To quantify the near-tie uncertainty at $\varepsilon = 1$, sort each candidate's $B = 20000$ calibration statistics. For its population 0.95 quantile, use order statistics with ranks ℓ and u, where ℓ is the 0.0125 quantile of $\mathrm{Bin}(B, 0.95)$ and u is one plus its 0.9875 quantile. Each interval has coverage at least 97.5%, including for discrete score laws. Bonferroni adjustment across RR and OLH gives simultaneous coverage at least 95%. Substituting the lower OLH and upper RR endpoints, and then the upper OLH and lower RR endpoints, into $(c_{\mathrm{OLH}}/c_{\mathrm{RR}})^{5/2}(\det \Sigma_{\mathrm{OLH}}/\det \Sigma_{\mathrm{RR}})^{1/2}$ yields $[0.931, 1.088]$. Applying the same calculation separately to the independent holdout gives ratio 0.965 and interval $[0.888, 1.043]$. The holdout is used for this uncertainty assessment, leaving the reported selector and online results unchanged. Within this pair, the calibration interval permits excess reference volume of up to 7.4% for choosing RR or 8.8% for choosing OLH, relative to the better candidate.

For the originally calibrated critical values at $\varepsilon = 1$, holdout acceptance is 0.947 for RR and 0.952 for OLH, with respective 95% exact binomial intervals $[0.944, 0.951]$ and $[0.949, 0.955]$. Accurate reference acceptance can thus coexist with an unresolved volume ranking: the volume raises the critical value to power $k/2$, amplifying relative critical-value error.

At $\varepsilon = 2$, the 20000-draw OLH/RR volume ratio is 1.331 and calibrated selection again retains RR. At $\varepsilon = 0.5$, $n = 500000$, and reference level 0.95, RR/OLH coverage under the narrower bandwidth is $0.865/0.850$, despite a selected/RR volume ratio 0.570. Substituting the true density only in the final standardization raises both coverages to 0.908; the estimator remains feasible but the substituted inference is an oracle comparison. Across the reference-level-0.95 cells, independent frozen holdout acceptance lies in $[0.946, 0.955]$. The gap between reference acceptance and online coverage reflects density-standardization and online approximation errors. The wider-bandwidth, fixed-threshold results in tables 1 and 3 use their own independently evaluated protocol.

τ	Nonprivate	$\epsilon = 0.5$	$\epsilon = 1$	$\epsilon = 2$
0.10	21.52	21.20 [20.87, 21.52]	21.37 [21.19, 21.54]	21.51 [21.43, 21.59]
0.25	24.14	24.00 [23.75, 24.24]	24.04 [23.91, 24.18]	24.13 [24.07, 24.19]
0.50	27.44	27.32 [27.00, 27.64]	27.32 [27.17, 27.48]	27.43 [27.37, 27.49]
0.75	31.75	31.66 [31.23, 32.09]	31.68 [31.47, 31.89]	31.73 [31.64, 31.82]
0.90	36.90	36.01 [35.28, 36.73]	36.77 [36.41, 37.14]	36.94 [36.79, 37.09]

Table 7: BRFSS respondent-BMI quantiles. Private entries are the estimate [method-calibrated simultaneous 95% interval]; the nonprivate column is the empirical benchmark.

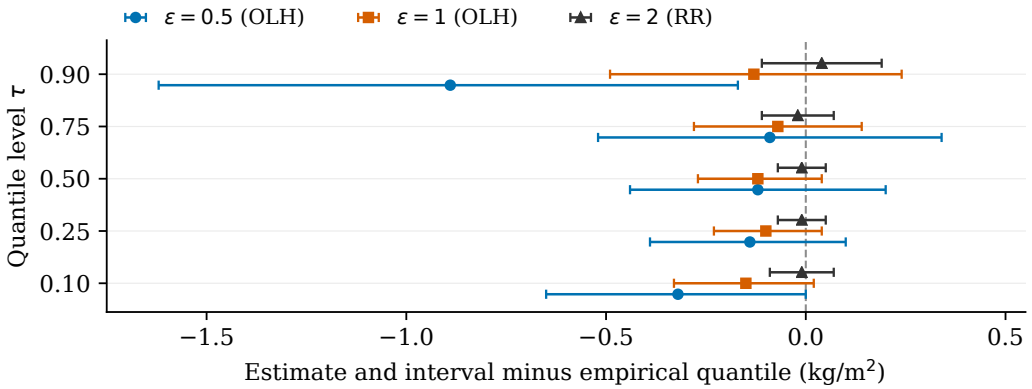

Figure 1: Private BMI estimates and method-calibrated simultaneous 95% intervals from table 7, each translated by its empirical quantile. The vertical line denotes the empirical benchmark. The three privacy levels are separate counterfactual deployments. At $\varepsilon = 0.5$, the upper-tail interval is widest and misses its benchmark; increasing ε contracts the intervals.

B.4 BRFSS RESPONDENT-BMI ILLUSTRATION

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey monitoring health-related behaviors and chronic conditions in the United States (CDC, 2025). Its scale and streaming data-collection structure provide a natural illustration of one-report local privacy. We use the 2024 combined landline and cell-phone public file, which contains 457,670 records, and extract the calculated BMI variable specified in the official layout. Retaining nonmissing values in the pre-specified range $12 \leq \text{BMI} < 100$ leaves 414,633 respondents. We treat their empirical BMI distribution as a randomly ordered stream and display intervals produced by the method for this finite-file illustration. Survey weights are not used.

BMI is especially useful here because distributional tails carry public-health information beyond the mean. CDC adult categories use 25 and 30 as the thresholds for overweight and obesity, while 35 marks class 2 obesity (CDC, 2024). Moreover, BRFSS height and weight are self-reported, and quantile-specific bias can be more pronounced in the upper tail (Zhao et al., 2023). We estimate the five quantiles $(0.10, 0.25, 0.50, 0.75, 0.90)$ to describe both central and upper-tail BMI.

Because the released BMI variable is recorded to two decimal places, each user first adds independent $\text{Unif}(-0.005, 0.005)$ dequantization noise before categorization. This local preprocessing removes rounding atoms while retaining the reported BMI scale. A fixed random order and fixed dequantization draws are used within each counterfactual deployment. For each of $\varepsilon \in \{0.5, 1, 2\}$, treated as a separate counterfactual deployment, users are split 10%/20%/70% and the pilot max-width selector searches the same five-mechanism library as in the simulation. Public bounds are $[12, 60]$, initial values are $(20, 24, 27.5, 30, 35)$, and 500 public draws calibrate each simultaneous interval. Table 7 compares private estimates with the empirical quantiles; figure 1 displays their differences on a common BMI scale.

The nonprivate quartiles place the median respondent in the overweight range and the upper quartile above the obesity threshold; the 0.90 quantile lies in the class 2 obesity range. At $\varepsilon = 0.5$ and $\varepsilon = 1$, the selector chooses OLH, whereas at $\varepsilon = 2$ it chooses RR. The maximum absolute difference from the empirical benchmark decreases from 0.891 at $\varepsilon = 0.5$ to 0.151 at $\varepsilon = 1$ and 0.042 at $\varepsilon = 2$. The simultaneous intervals contract in the same order. Thus the application exhibits both the privacy–precision tradeoff and the purpose of library selection: the preferred channel can change with the privacy regime even when the estimand and stream are held fixed.

The interval calibration uses the i.i.d. population model, while this illustration samples without replacement from the finite file. Survey-design coverage would require accounting for the sampling weights and dependence. In particular, the $\varepsilon = 0.5$ interval for $\tau = 0.90$ excludes the empirical target 36.90.

C Supporting results for mechanism selection and inference

C.1 Score geometry and staircase designs

For the quantitative bounds, f_+ denotes the global density upper bound in assumption C.5. At the truth, let $C^* \sim \boldsymbol{\pi}^*$, $Y^* \sim \mathsf{Q}(\cdot \mid C^*)$, and define

$$\xi^* = \boldsymbol{\tau} - a_{Y^*}, \qquad \Sigma = \mathrm{Var}(\xi^*), \qquad \sigma_- = \lambda_{\min}(\Sigma)^{1/2}, \tag{35}$$

$$G = \max_y \|\boldsymbol{\tau} - a_y\|_2, \qquad A_0 = \max_y \|a_y\|_2, \qquad H_0 = 2A_0 + f_+\sqrt{k}(U - L),$$

$$\beta_3 = \mathbb{E}\left\|\Sigma^{-1/2}\xi^*\right\|_2^3, \qquad \ell = \max_y \left\|\Sigma^{-1/2}(a_y - \boldsymbol{\tau})\right\|_2. \tag{36}$$

Candidate-specific quantities in (35)–(36) carry a subscript r. The stability and approximation bounds use G_r and $\beta_{3,r}$, while the selection criteria are defined in section 2.2 from the frozen confidence sets.

Staircase channels are useful candidate generators. Let

$$\mathcal{S}_\varepsilon = \{s \in \mathbb{R}^d : s_c \in \{1, e^\varepsilon\},\ 0 \le c \le k\}.$$

For nonnegative weights satisfying $\sum_{s \in \mathcal{S}_\varepsilon} \theta_s s = \mathbf{1}_d$, form a channel with output row $\theta_s s^\top$ for each s with $\theta_s > 0$. The constraint normalizes each input column, and the two entry levels ensure ε-LDP. Its information matrix is

$$H(\theta) = \sum_{s \in \mathcal{S}_\varepsilon} \theta_s \frac{ss^\top}{(\boldsymbol{\pi}^*)^\top s}. \tag{37}$$

Theorem C.1 (Staircase reduction and convex candidate generation)**.** *Fix $W \succ 0$. Among finite-output ε-LDP channels admitting an unbiased CDF decoder, the minimum of $\mathrm{tr}\{W\Sigma(\mathsf{Q})\}$ is attained by a staircase channel. It is obtained from*

$$\begin{aligned} \textit{minimize}_{\theta \ge 0} \quad & \mathrm{tr}\{WPH(\theta)^{-1}P^\top\}, \\ \textit{subject to} \quad & \sum_{s \in \mathcal{S}_\varepsilon} \theta_s s = \mathbf{1}_d, \end{aligned} \tag{38}$$

with objective $+\infty$ when $H(\theta)$ is singular. For a nonsingular feasible design,

$$a_s = \frac{PH(\theta)^{-1}s}{(\boldsymbol{\pi}^*)^\top s}, \qquad \Sigma_\theta = PH(\theta)^{-1}P^\top - \boldsymbol{\tau}\boldsymbol{\tau}^\top. \tag{39}$$

The feasible set is compact, the objective is convex, and a minimizer exists.

The theorem builds on staircase extremality (Kairouz et al., 2016), with an objective and decoder specialized to the CDF workload. Each solution of (38) enters the candidate library. For example, $W = I_k$ minimizes decoded-score trace, whereas the final quantile covariance is $D^{-1}\Sigma D^{-1}$ with

$$D = \mathrm{diag}\{f(Q_1), \ldots, f(Q_k)\}.$$

The final selection compares the candidates through the geometry of the chosen confidence set.

For direct d-ary randomized response (kRR, with $d = k + 1$), let p and q be the probabilities of the true and each other category:

$$p = \frac{e^\varepsilon}{e^\varepsilon + d - 1}, \qquad q = \frac{1}{e^\varepsilon + d - 1}, \qquad \delta_\varepsilon = p - q = \frac{e^\varepsilon - 1}{e^\varepsilon + k}. \tag{40}$$

Its truth output probabilities are

$$\mu = q\mathbf{1}_d + \delta_\varepsilon \boldsymbol{\pi}^*, \qquad \mu_c = q + \delta_\varepsilon \pi_c^*, \qquad \mu_\varepsilon = \min_c \mu_c. \tag{41}$$

The probabilities in (41) and the explicit decoder constants in proposition G.1 quantify the privacy and dimension dependence of randomized response.

C.2 Density-free Haar selection

For density-free self-normalized inference, use $K \geq k$ orthonormal, zero-sum contrasts together with the constant weight $w_{0,N} = N^{-1/2}\mathbf{1}_N$. The block-Haar construction contrasts two adjacent groups within each of K disjoint blocks; its weights satisfy (67) and are specified in (69). The resulting score coefficients provide a public reference for the self-normalizer used in appendix E. In the frozen score experiment, define the coefficients

$$Z^\circ_{r,0} = \sum_{i=1}^N w_{0,N}(i)\xi^\circ_{r,i}, \qquad Z^\circ_{r,\ell} = \sum_{i=1}^N w_{\ell,N}(i)\xi^\circ_{r,i}, \qquad S^{\xi,\circ}_{r,K,N} = K^{-1}\sum_{\ell=1}^K Z^\circ_{r,\ell} Z^{\circ\top}_{r,\ell}. \tag{42}$$

The coefficients in (42) also give a public self-normalized experiment. Let $c^H_{r,N,K,\alpha}$ be the exact frozen critical value for $Z^{\circ\top}_{r,0}(S^{\xi,\circ}_{r,K,N})^{-1}Z^\circ_{r,0}$, with $+\infty$ on singular normalizers. For fixed $\zeta \in (0,1)$, define the robust frozen size

$$J^H_{r,N,K,\zeta} = \mathrm{Quantile}_{1-\zeta}\left[\{c^H_{r,N,K,\alpha}\}^{k/2}\det(S^{\xi,\circ}_{r,K,N})^{1/2}\right], \tag{43}$$

where the bracketed variable is $+\infty$ when the normalizer is singular. The parameter ζ determines which quantile of this random volume is minimized. Under the local quantile linearization, the size in (43) acquires the common quantile-space factor $v_k N^{-k/2}\det(D^{-1})$. Thus

$$\widehat{r}^H = \arg\min_{r\in\mathfrak{M}_n} J^H_{r,N,K,\zeta} \tag{44}$$

is density-free and incorporates the actual calibrated critical value, score distribution, and singularity behavior of the chosen self-normalizer.

C.3 Main assumptions

The principal post-selection results require smoothness near the target quantiles and uniform control of the finite mechanism library.

Assumption C.2 (Identification and local smoothness)**.** *The distribution function F is continuous and strictly increasing on $[L, U]$, with $F(L) < \tau_1 < \tau_k < F(U)$. The target satisfies*

$$\Delta_Q = \min\{Q_1 - L, U - Q_k, Q_{j+1} - Q_j : 1 \leq j < k\} > 0. \tag{45}$$

For some $\beta \in (0, 1]$ and neighborhoods $\mathcal{N}_j$ of Q_j, F has a density f on $\cup_j \mathcal{N}_j$, every $f(Q_j) > 0$, and

$$|f(x) - f(Q_j)| \leq L_j|x - Q_j|^\beta, \qquad x \in \mathcal{N}_j. \tag{46}$$

Assumption C.3 (Stages, library, and privacy scores)**.** *The three user groups are disjoint and $n_1, n_2, n_3 \to \infty$. Stage-3 users are independent of the sigma-field $\mathcal{G}_n$ generated by the pilot, probes, public simulation, and selected library. Conditional on $\mathcal{G}_n$, the library has at most $R_{\max} < \infty$ candidates. The baseline r_0 is retained in the library. Every retained candidate is an ε-LDP unbiased channel–decoder pair and, for deterministic constants $0 < \underline{\sigma} < \overline{\sigma} < \infty$ and $G_{\max} < \infty$,*

$$\underline{\sigma}^2 I_k \preceq \Sigma_r \preceq \overline{\sigma}^2 I_k, \qquad \max_y \|\boldsymbol{\tau} - a_{r,y}\|_2 \leq G_{\max} \tag{47}$$

for every retained r, almost surely for all sufficiently large n.

Assumption C.4 (Bandwidth, truncation, and steps)**.** *The nominal density bandwidth and numerical floor satisfy*

$$h_n \to 0, \qquad n_2 h_n^2 \to \infty, \qquad \ell_n = o(h_n). \tag{48}$$

The truncation sequences satisfy $\underline{f}_n \downarrow 0$ *and* $\overline{f}_n \uparrow \infty$*. The pilot and main steps have the form* $\gamma_t = c(t + s_0)^{-a}$ *with constants bounded away from zero and infinity, where*

$$\max\left\{\frac{1}{2}, \frac{1}{1+\beta}\right\} < a < 1. \tag{49}$$

At every finite n*, the density truncation makes* $\widehat{B} = \widehat{D}^{-1}$ *positive definite and finite.*

The separation condition can be assessed through the pilot gaps and the frequency of active isotonic projection. Density regularity can be examined by comparing common-shift estimates at $h_n/2, h_n$, and $2h_n$. Privacy, unbiasedness, and the eigenvalue bounds in (47) are computable from the public channel, decoder, and $\boldsymbol{\pi}^*$. Candidate generation may use a pilot-weighted staircase optimizer followed by the same deterministic checks. The pivotal limits take k and ε as fixed; the appendices give their explicit contribution to finite-confidence bounds.

The quantitative post-burn-in transfer results use the following stronger conditions.

Assumption C.5 (Global regularity for quantitative transfer)**.** *The target is strictly interior and separated as in* (45)*. The distribution has a continuously differentiable density on* $[L, U]$ *satisfying*

$$0 < f_- \le f(x) \le f_+ < \infty, \qquad |f'(x)| \le L_f. \tag{50}$$

Assumption C.6 (Decoder and steps for quantitative transfer)**.** *The output alphabet is finite,* $\Sigma \succ 0$*, and* $G, A_0 < \infty$*. The unpreconditioned fixed-mechanism recursion uses* $\gamma_t = \gamma_0 t^{-a}$ *with* $a \in (1/2, 1)$ *and* $\gamma_0 f_+ \le 1/4$*.*

The density bounds in (50) supply the global contraction and exponential entrance estimates used by the finite-confidence online-to-frozen results in the appendices. The principal post-selection central limit theorem uses assumption C.2.

C.4 Privacy and density probing

Proposition C.7 (Interactive privacy and communication)**.** *Under assumption C.3, algorithm 1 is user-level* ε*-LDP for every user conditional on the preceding transcript. Every user sends one symbol from the output alphabet of the mechanism assigned to that stage. Pilot averaging, common-shift density estimation, mechanism construction and selection, weighted projection, frozen-score simulation, multiplier calibration, and confidence-set construction are post-processing and consume no additional privacy budget.*

Separate user groups make the selected mechanism and preconditioner fixed when the main stream begins. Each user participates in one stage and uses the full privacy budget ε.

Theorem C.8 (Consistency and rate of the common-shift density probe)**.** *Suppose assumptions C.2 to C.4 hold. Then the Stage-1 average satisfies* $\widetilde{\boldsymbol{Q}} \to \boldsymbol{Q}$ *in probability. Write* $\widehat{D}^{\mathrm{raw}} = \mathrm{diag}(\widehat{f}_j^{\mathrm{raw}})$*. Let* $n_{2,+}$ *and* $n_{2,-}$ *be the two probe sizes, with* $n_{2,+}/n_2 \to \rho_+ \in (0,1)$ *and* $n_{2,-}/n_2 \to \rho_- \in (0,1)$*. Then*

$$\left\|\widehat{D}^{\mathrm{raw}} - D\right\|_{\mathrm{op}} = O_{\mathbb{P}}\left(\{\left\|\widetilde{\boldsymbol{Q}} - \boldsymbol{Q}\right\|_{\infty} + h_n\}^{\beta} + \frac{1}{h_n\sqrt{n_2}}\right). \tag{51}$$

Consequently, $\widehat{D} \to D$ *in probability and*

$$\mathbb{P}(\widehat{D} = \widehat{D}^{\mathrm{raw}}) \to 1. \tag{52}$$

For $\beta = 1$*, balancing smoothing and stochastic terms gives the conventional choice* $h_n \asymp n_2^{-1/4}$*, apart from pilot error.*

Let $d_- = \min_j f(Q_j) > 0$, $d_+ = \max_j f(Q_j) < \infty$, $b_- = (2d_+)^{-1}$, and $b_+ = 2d_-^{-1}$, and define

$$\mathcal{E}_n^B = \left\{b_- I_k \preceq \widehat{B} \preceq b_+ I_k\right\}. \tag{53}$$

Then theorem C.8 gives $\mathbb{P}(\mathcal{E}_n^B) \to 1$. On the event in (53), the preconditioner has uniform spectral bounds, and its complement contributes to the outer $o_{\mathbb{P}}(1)$ term.

Each shifted report contains all k CDF indicators, so every quantile uses the full positive- or negative-shift probe sample.

C.5 Exact density cancellation and library oracle

Theorem C.9 (Finite-score library oracle and density cancellation)**.** *Fix n_3, α, and a realized finite candidate library.*

(i) The Wald selector in (9) *exactly minimizes the volume in* (8) *over the library for every positive diagonal D.*

(ii) For the Haar construction, the selector in (44) *minimizes the $(1-\zeta)$ quantile of the frozen quantile-space volume*

$$v_k n_3^{-k/2} \det(D^{-1}) \{c^H_{r,n_3,K,\alpha}\}^{k/2} \det(S^{\xi,\circ}_{r,K,n_3})^{1/2}$$

for every positive diagonal D, with infinite volume on a singular normalizer.

(iii) If direct randomized response belongs to the library, then, for $g \in \{W, H\}$,

$$J^g_{\widehat{r}^g} \le J^g_{r_{\mathrm{RR}}}. \tag{54}$$

These conclusions are deterministic identities conditional on the realized library. In particular, (54) *compares the selected criterion directly with randomized response.*

C.6 Detailed post-selection limits and inference

Theorem C.10 (Conditional Polyak–Ruppert representation after selection)**.** *Let $\mathcal{F}_s$ contain the first s Stage-3 reports, and write $\mathbb{P}^*$ for probability over the bootstrap draws conditional on the transcript. Under assumptions C.2 to C.4, the recursive-average branch of Algorithm 1 has the following properties conditionally on $\mathcal{G}_n$, uniformly over the finite realized library on $\mathcal{E}_n^B$, whose probability tends to one.*

(i) The Stage-3 iterates satisfy $\boldsymbol{q}_s^m \to \boldsymbol{Q}$ almost surely with respect to the main-stream probability, and the weighted order projection is inactive after a finite random time.

(ii) The full Stage-3 average obeys

$$D\sqrt{n_3}(\overline{\boldsymbol{q}}_{n_3} - \boldsymbol{Q}) = \frac{1}{\sqrt{n_3}} \sum_{s=1}^{n_3} M_s + o_{\mathbb{P}}(1), \tag{55}$$

where the conditional remainder converges to zero in probability.

(iii) The conditional covariance average satisfies

$$\left\| \frac{1}{n_3} \sum_{s=1}^{n_3} \mathbb{E}(M_s M_s^\top \mid \mathcal{F}_{s-1}, \mathcal{G}_n) - \Sigma_{\widehat{r}} \right\|_{\mathrm{op}} = o_{\mathbb{P}}(1) \tag{56}$$

conditionally on $\mathcal{G}_n$. Let $Z_{\widehat{r},n}$, conditionally on $\mathcal{G}_n$, have distribution $N_k(0, D^{-1}\Sigma_{\widehat{r}}D^{-1})$. If BL_1 denotes the functions bounded by one and Lipschitz with constant one, then

$$\sup_{\varphi \in \mathrm{BL}_1} \left| \mathbb{E}\big[\varphi\{\sqrt{n_3}(\overline{\boldsymbol{q}}_{n_3} - \boldsymbol{Q})\} \mid \mathcal{G}_n\big] - \mathbb{E}\{\varphi(Z_{\widehat{r},n}) \mid \mathcal{G}_n\} \right| \longrightarrow 0 \tag{57}$$

in probability.

(iv) The empirical Stage-3 score covariance is consistent:

$$\widehat{\Sigma}_U \longrightarrow \Sigma_{\widehat{r}} \tag{58}$$

in probability, conditionally on $\mathcal{G}_n$.

The first-order covariance is independent of the positive diagonal preconditioner $\widehat{B}$. If $\Sigma_{\widehat{\tau}} \to \Sigma_\star$ in probability for a deterministic positive-definite matrix $\Sigma_\star$, (57) *gives the conditional weak limit* $N_k(0, D^{-1}\Sigma_\star D^{-1})$.

The cancellation follows from the local Jacobian $-\widehat{B}D$: the averaged influence map is $(\widehat{B}D)^{-1}\widehat{B} = D^{-1}$. Preconditioning improves local scaling while preserving the first-order influence function. The functional version and the fixed-mechanism specialization are stated in corollary H.5.

Theorem C.11 (Conditional validity of selected confidence sets)**.** *Suppose assumptions C.2 to C.4 hold and $\widehat{D} \to D$ in probability.*

(i) The frozen-score calibrated Wald set satisfies

$$\mathbb{P}\left\{\boldsymbol{Q} \in \mathcal{C}^W_{1-\alpha} \mid \mathcal{G}_n\right\} \longrightarrow 1-\alpha \tag{59}$$

in probability.

(ii) The frozen-score calibrated simultaneous rectangle satisfies

$$\mathbb{P}\left\{\boldsymbol{Q} \in \mathcal{C}^{\max}_{1-\alpha} \mid \mathcal{G}_n\right\} \longrightarrow 1-\alpha \tag{60}$$

in probability.

(iii) Let $c^*_{1-\alpha}$ be the exact conditional quantile of the Gaussian multiplier statistic in (33)*. Then*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\max_j \frac{\sqrt{n_3}|\overline{q}_{n_3,j} - Q_j|}{\widehat{s}_j} \le x \,\middle|\, \mathcal{G}_n\right\} - \mathbb{P}^*\left\{\max_j \frac{|Z^*_j|}{\widehat{s}_j} \le x\right\}\right| \longrightarrow 0 \tag{61}$$

in probability, and $\mathbb{P}\{\boldsymbol{Q} \in \mathcal{C}^{\mathrm{MB}}_{1-\alpha} \mid \mathcal{G}_n\} \to 1-\alpha$.

(iv) The corresponding unconditional coverages also converge to $1-\alpha$.

Sample splitting makes the selected mechanism and nuisance estimates fixed for the main stream, so conditional calibration transfers directly to the selected confidence set.

The coverage guarantee applies even when several mechanisms have the same optimal criterion and the selected index varies across pilot samples. Conditional inference uses the covariance of the realized channel. Consistent selection additionally yields the oracle covariance approximation in corollary H.8.

The appendices give the functional process representation, the oracle limit under consistent selection, uniform frozen critical-value limits, and finite public-calibration bounds. They also develop density-free self-normalized alternatives based on orthonormal series, block-Haar contrasts, integral normalization, and independent blocks.

D Post-burn-in localization and linearization

This appendix develops the fixed-mechanism recursion and the quantitative online-to-frozen bounds used by the density-free procedures.

Using the drift h in (18), define

$$M_t = (\boldsymbol{\tau} - a_{Y_t}) - h(\boldsymbol{q}_{t-1}).$$

Let $\mathcal{F}_t$ denote the report history through time t. Then $(M_t, \mathcal{F}_t)$ is a bounded martingale-difference sequence and, whenever the projection is inactive,

$$\boldsymbol{q}_t = \boldsymbol{q}_{t-1} + \gamma_t\{h(\boldsymbol{q}_{t-1}) + M_t\}. \tag{62}$$

The decomposition in (62) separates the restoring drift from the martingale noise that determines the averaged estimator's limiting covariance.

D.1 Consistency, eventual inactivity, and functional linearization

Theorem D.1 (Online asymptotics)**.** *Under assumptions C.5 and C.6:*

(i) $\boldsymbol{q}_t \to \boldsymbol{Q}$ *almost surely;*

(ii) with probability one, the projection is inactive after a finite random time;

(iii) for $\bar{\boldsymbol{q}}_n = n^{-1}\sum_{t=1}^{n} \boldsymbol{q}_{t-1}$,

$$D\sqrt{n}(\bar{\boldsymbol{q}}_n - \boldsymbol{Q}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} M_t + o_{\mathbb{P}}(1); \tag{63}$$

(iv)

$$\sqrt{n}(\bar{\boldsymbol{q}}_n - \boldsymbol{Q}) \Rightarrow N_k(0,\Omega), \qquad \Omega = D^{-1}\Sigma D^{-1}; \tag{64}$$

(v) with the process set equal to zero at $r = 0$*,*

$$\sup_{0\le r\le 1}\left\| \frac{\lfloor nr\rfloor}{\sqrt{n}}(\bar{\boldsymbol{q}}_{\lfloor nr\rfloor} - \boldsymbol{Q}) - D^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} M_t \right\|_2 = o_{\mathbb{P}}(1). \tag{65}$$

D.2 Tail localization

Fix a burn-in b, set $m = n - b$, and let $\mathcal{I}_{b,n}$ be the event that the projection is inactive for $t = b+1,\dots,n$. We write $\pi_{b,n} = \mathbb{P}(\mathcal{I}_{b,n}^c)$. The exponential-stability result in lemma I.1 and corollary I.2 gives, whenever $\gamma_{b+1}G \le \Delta_Q/8$,

$$\pi_{b,n} \le C_0(m+1)\exp\left\{-\frac{c_0\Delta_Q^2}{64G^2\gamma_b}\right\}. \tag{66}$$

The dependence on privacy enters through G: a tighter privacy budget enlarges the decoded update and therefore delays the time at which the order projection is guaranteed to remain inactive.

Assumption D.2 (Balanced burn-in and retained segment)**.** *For a fixed constant* $c_b \in (0, 1/2)$*, the burn-in satisfies* $c_b n \le b \le (1-c_b)n$*. Hence* b*,* $m = n - b$*, and all tail times are comparable. Constants in the quantitative tail bounds may depend on* c_b*.*

This condition is used only to express every remainder in terms of the retained length m. More general splits can be handled by retaining separate powers of b and m.

D.3 Orthonormal weights and basis roughness

Let $w_{0,m} = m^{-1/2}\mathbf{1}_m$. A K-contrast system $\mathcal{W}$ is a collection $w_{1,m},\dots,w_{K,m} \in \mathbb{R}^m$ satisfying

$$w_{r,m}^{\top} w_{s,m} = \mathbf{1}\{r = s\}, \qquad 0 \le r, s \le K. \tag{67}$$

Thus every contrast $w_{\ell,m}$, $\ell \ge 1$, sums to zero. Define

$$\mu(\mathcal{W}) = \sqrt{m}\max_{0\le\ell\le K}\max_{1\le i\le m}|w_{\ell,m}(i)|,$$

$$\mathfrak{v}(\mathcal{W}) = \sqrt{m}\max_{0\le\ell\le K}\left\{2\max_i|w_{\ell,m}(i)| + \sum_{i=1}^{m-1}|w_{\ell,m}(i+1) - w_{\ell,m}(i)|\right\}.$$

The first quantity controls martingale increments; the second controls the Abel boundary term.

Two systems will be used.

1. The discrete cosine transform (DCT) system is

$$w_{\ell,m}(i) = \sqrt{\frac{2}{m}}\cos\left\{\frac{\pi\ell(i-1/2)}{m}\right\}, \qquad 1 \le \ell \le K. \tag{68}$$

It satisfies $\mu(\mathcal{W}) \le \sqrt{2}$ and $\mathfrak{v}(\mathcal{W}) \le C(1+K)$.

2. For the block-Haar system, assume $m \geq 4K$, set $s = \lfloor m/(2K) \rfloor$, and define

$$w_{\ell,m}(i) = \frac{1}{\sqrt{2s}} \begin{cases} 1, & 2s(\ell-1) < i \leq 2s(\ell-1)+s, \\ -1, & 2s(\ell-1)+s < i \leq 2s\ell, \\ 0, & \text{otherwise.} \end{cases} \tag{69}$$

Then $\mu(\mathcal{W}) \leq \sqrt{2K}$ and $\mathfrak{v}(\mathcal{W}) \leq 6\sqrt{2K}$. The K frozen-score contrast coefficients use disjoint observations and are therefore independent even when the score distribution is non-Gaussian.

D.4 Coupling and weighted linearization

For each tail user, define a truth category $C_t^* = C(\boldsymbol{Q}; X_t)$ and a truth report Y_t^* generated from the same channel. The coupling uses the same raw X_t and a maximal coupling of $\mathsf{Q}(\cdot \mid C_t)$ and $\mathsf{Q}(\cdot \mid C_t^*)$. Then $\xi_t^* = \boldsymbol{\tau} - a_{Y_t^*}$ are i.i.d. copies of ξ^*.

For $i = 1, \ldots, m$, write $t_i = b + i$ and $e_i = \boldsymbol{q}_{t_i-1} - \boldsymbol{Q}$; also set $e_{m+1} = \boldsymbol{q}_n - \boldsymbol{Q}$. Define the standardized observed coefficients

$$\widehat{X}_0(\theta) = \Sigma^{-1/2} D \sum_{i=1}^{m} w_{0,m}(i)(\boldsymbol{q}_{t_i-1} - \theta),$$

$$\widehat{X}_\ell = \Sigma^{-1/2} D \sum_{i=1}^{m} w_{\ell,m}(i) \boldsymbol{q}_{t_i-1}, \qquad 1 \leq \ell \leq K,$$

and the frozen-score coefficients

$$X_\ell^\circ = \Sigma^{-1/2} \sum_{i=1}^{m} w_{\ell,m}(i) \xi_{t_i}^*, \qquad 0 \leq \ell \leq K. \tag{70}$$

The unknown common transformation $\Sigma^{-1/2} D$ is used only in the proof; all quadratic-form statistics below are invariant to it.

Let

$$\Lambda_{m,K}(\delta) = k + \log \frac{8(K+1)m}{\delta}.$$

Theorem D.3 (Weighted online-to-frozen linearization)**.** *Under assumptions C.5, C.6 and D.2, let $m = n - b$, and let $\mathcal{W}$ satisfy* (67)*. For every $0 < \delta < 1/4$, with probability at least $1 - \pi_{b,n} - 3\delta$,*

$$\max_{0 \leq \ell \leq K} \left\| \widehat{X}_\ell(\boldsymbol{Q}) - X_\ell^\circ \right\|_2 \leq r_{m,K}(\delta), \tag{71}$$

where

$$\begin{aligned} r_{m,K}(\delta) \leq C\sigma_-^{-1} \Big[& G\mathfrak{v}(\mathcal{W}) m^{-(1-a)/2} \Lambda_{m,K}(\delta)^{1/2} + G^2 m^{1/2-a} \Lambda_{m,K}(\delta) \\ & + A_0 (f_+ \sqrt{k} G)^{1/2} m^{-a/4} \Lambda_{m,K}(\delta) + f_+ G m^{-a/2} \Lambda_{m,K}(\delta) \\ & + H_0 \mu(\mathcal{W}) m^{-1/2} \Lambda_{m,K}(\delta) \Big]. \end{aligned} \tag{72}$$

At the balanced exponent $a = 2/3$,

$$r_{m,K}(\delta) \leq C \mathfrak{R}_{k,\varepsilon,K}(\mathcal{W}) m^{-1/6} \Lambda_{m,K}(\delta), \tag{73}$$

where

$$\mathfrak{R}_{k,\varepsilon,K}(\mathcal{W}) = \sigma_-^{-1} \left[G\mathfrak{v}(\mathcal{W}) + G^2 + A_0 (f_+ \sqrt{k} G)^{1/2} + f_+ G + H_0 \mu(\mathcal{W}) \right]. \tag{74}$$

The bound is deliberately conservative in logarithms; all powers of m, k, privacy-dependent mechanism constants, and basis roughness are explicit.

D.5 A QUANTITATIVE TAIL-PATH COUPLING

The integral self-normalizer depends on the complete post-burn-in trajectory, so an endpoint bound alone is insufficient. For $0 \le s \le m$, define

$$\widehat{\mathbb{P}}_m(s) = \Sigma^{-1/2} D\, m^{-1/2} \sum_{i=1}^{s} e_i, \qquad \mathbb{P}^\circ_m(s) = \Sigma^{-1/2} m^{-1/2} \sum_{i=1}^{s} \xi^*_{t_i},$$
$$\widehat{\mathbb{U}}_m(s) = \widehat{\mathbb{P}}_m(s) - \frac{s}{m}\widehat{\mathbb{P}}_m(m), \qquad \mathbb{U}^\circ_m(s) = \mathbb{P}^\circ_m(s) - \frac{s}{m}\mathbb{P}^\circ_m(m).$$

Empty sums are zero. Put

$$\Lambda^{\mathrm{path}}_m(\delta) = k + \log\frac{8m}{\delta}.$$

Theorem D.4 (Uniform tail-path linearization)**.** *Under assumptions C.5, C.6 and D.2, let $m = n - b$. For every $0 < \delta < 1/4$, with probability at least $1 - \pi_{b,n} - 3\delta$,*

$$\max_{0\le s\le m} \left\| \widehat{\mathbb{P}}_m(s) - \mathbb{P}^\circ_m(s) \right\|_2 \le r^{\mathrm{path}}_m(\delta), \tag{75}$$

where

$$r^{\mathrm{path}}_m(\delta) \le C\sigma_-^{-1}\Big[G m^{-(1-a)/2}\Lambda^{\mathrm{path}}_m(\delta)^{1/2} + G^2 m^{1/2-a}\Lambda^{\mathrm{path}}_m(\delta) + A_0(f_+\sqrt{k}G)^{1/2} m^{-a/4}\Lambda^{\mathrm{path}}_m(\delta) + f_+ G m^{-a/2}\Lambda^{\mathrm{path}}_m(\delta) + H_0 m^{-1/2}\Lambda^{\mathrm{path}}_m(\delta)\Big]. \tag{76}$$

Consequently,

$$\max_{0\le s\le m} \left\| \widehat{\mathbb{U}}_m(s) - \mathbb{U}^\circ_m(s) \right\|_2 \le 2 r^{\mathrm{path}}_m(\delta). \tag{77}$$

At $a = 2/3$,

$$r^{\mathrm{path}}_m(\delta) \le C\mathfrak{R}^{\mathrm{path}}_{k,\varepsilon} m^{-1/6}\Lambda^{\mathrm{path}}_m(\delta), \tag{78}$$

where

$$\mathfrak{R}^{\mathrm{path}}_{k,\varepsilon} = \sigma_-^{-1}\{G + G^2 + A_0(f_+\sqrt{k}G)^{1/2} + f_+ G + H_0\}.$$

For balanced kRR, meaning $\pi^*_c \asymp 1/k$, and $0 < \varepsilon \le 1$,

$$r^{\mathrm{path}}_m(\delta) \le C_f \frac{k^{5/2}}{\varepsilon} m^{-1/6}\Lambda^{\mathrm{path}}_m(\delta). \tag{79}$$

Corollary D.5 (Explicit privacy and dimension order for kRR)**.** *For the block-Haar basis and kRR, let $K \le C_K k$. Then*

$$\mathfrak{R}^{\mathrm{kRR}}_{k,\varepsilon,K} \le C_f\left[\frac{\sqrt{k}(1+\sqrt{K})}{\sqrt{\mu_\varepsilon}} + \frac{k}{\delta_\varepsilon\sqrt{\mu_\varepsilon}} + \frac{k}{\sqrt{\delta_\varepsilon\mu_\varepsilon}} + \frac{\sqrt{k}\mu(\mathcal{W})}{\sqrt{\mu_\varepsilon}}\right].$$

If $0 < \varepsilon \le 1$ and $\pi^*_c \asymp 1/k$, then

$$\mathfrak{R}^{\mathrm{kRR}}_{k,\varepsilon,K} \le C_f \frac{k^{5/2}}{\varepsilon}. \tag{80}$$

Consequently,

$$r_{m,K}(\delta) \le C_f \frac{k^{5/2}}{\varepsilon} m^{-1/6}\Lambda_{m,K}(\delta)$$

outside probability $\pi_{b,n} + 3\delta$. This upper bound gives the retained-path order used below.

E INFERENCE PROCEDURES AND SCORE-CALIBRATED HAAR CONTRASTS

All procedures in this section use the same segment $\boldsymbol{q}_b, \boldsymbol{q}_{b+1}, \ldots, \boldsymbol{q}_{n-1}$ of length $m = n - b$, matching the tail segment covered by the quantitative linearization.

If both selection and inference are density-free, omit the probe stream and set $\widehat{B} = I_k$ in (28). For $k = 1$, binary randomized response, $\widehat{B} = 1$, and a singleton library, this architecture reduces to the online self-normalized quantile framework of Liu et al. (2023). For a singleton library and arbitrary k, it reduces to the fixed-mechanism procedure below.

E.1 Oracle and plug-in Wald procedures

For smooth strongly convex losses, Chen et al. (2020) estimate the covariance of averaged stochastic-gradient iterates by online plug-in and batch-means methods. Here the selected channel supplies the score covariance Σ publicly; Wald inference estimates the density transformation, while the self-normalizers below use the retained quantile path. Define

$$\bar{\boldsymbol{q}}_{b,m} = m^{-1} \sum_{i=1}^{m} \boldsymbol{q}_{b+i-1}.$$

The oracle statistic is

$$T_m^{\mathrm{or}}(\theta) = m(\bar{\boldsymbol{q}}_{b,m} - \theta)^\top \Omega^{-1} (\bar{\boldsymbol{q}}_{b,m} - \theta).$$

Its analytic critical value is $\chi^2_{k,1-\alpha}$. A plug-in statistic replaces D by an independent estimator $\widehat{D}$ and uses $\widehat{\Omega} = \widehat{D}^{-1} \Sigma \widehat{D}^{-1}$.

E.2 Integral self-normalization

This construction uses the recursive-estimate self-normalization (SN) of Shao (2010). For $1 \le s \le m-1$, define the tail bridge

$$B^q_{s,m} = \frac{1}{\sqrt{m}} \left\{ \sum_{i=1}^{s} \boldsymbol{q}_{b+i-1} - \frac{s}{m} \sum_{i=1}^{m} \boldsymbol{q}_{b+i-1} \right\}.$$

Set

$$S_m^{\mathrm{SN}} = \frac{1}{m} \sum_{s=1}^{m-1} B^q_{s,m} B^{q\top}_{s,m}, \qquad T_m^{\mathrm{SN}}(\theta) = m(\bar{\boldsymbol{q}}_{b,m} - \theta)^\top (S_m^{\mathrm{SN}})^{-1} (\bar{\boldsymbol{q}}_{b,m} - \theta). \tag{81}$$

For fixed k, let B be standard k-dimensional Brownian motion. Its Brownian reference is

$$U_k = B(1)^\top V_k^{-1} B(1), \qquad V_k = \int_0^1 \{B(r) - rB(1)\}\{B(r) - rB(1)\}^\top dr. \tag{82}$$

Proposition E.1 (Brownian normalizer spectrum)**.** *Let $G_1, G_2, \ldots$ be i.i.d. $N_k(0, I_k)$. Then*

$$V_k \overset{d}{=} \sum_{\ell=1}^{\infty} \lambda_\ell G_\ell G_\ell^\top, \qquad \lambda_\ell = \frac{1}{\pi^2 \ell^2}. \tag{83}$$

Its temporal effective rank is

$$\frac{(\sum_\ell \lambda_\ell)^2}{\sum_\ell \lambda_\ell^2} = \frac{5}{2}. \tag{84}$$

It is positive definite almost surely and, for $k \ge 2$,

$$\mathbb{E}\lambda_{\min}(V_k) \le \sum_{\ell=k}^{\infty} \lambda_\ell \le \frac{1}{\pi^2 (k-1)}. \tag{85}$$

For every integer $J \ge k$,

$$V_k \succeq \lambda_J \sum_{\ell=1}^{J} G_\ell G_\ell^\top. \tag{86}$$

Hence increasing $J - k$ gives polynomial lower-tail bounds of increasing order. Taking $J \asymp k$ in (86) *produces the dimension-relevant lower spectral scale $1/k$.*

For the DCT weights (68), the integral normalizer has the exact finite-m decomposition

$$S_m^{\mathrm{SN}} = \sum_{\ell=1}^{m-1} \bar{\lambda}_{\ell,m} Y_{\ell,m} Y_{\ell,m}^\top, \qquad \bar{\lambda}_{\ell,m} = \{4m^2 \sin^2(\pi\ell/(2m))\}^{-1}, \tag{87}$$

where $Y_{\ell,m} = \sum_i w_{\ell,m}(i) \boldsymbol{q}_{b+i-1}$. In particular,

$$\bar{\lambda}_{J,m} \ge \frac{1}{\pi^2 J^2}, \qquad 1 \le J < m. \tag{88}$$

The proof is given in appendix J.1.

E.3 EQUAL-WEIGHT ORTHONORMAL-CONTRAST PIVOT

Orthonormal-series (SER) variance estimation also underlies the fixed-smoothing F tests of Sun (2013). Let $\mathcal{W} = (w_{0,m}, \dots, w_{K,m})$ satisfy (67). Define observable coefficients

$$Z_{0,m}(\theta) = \sum_{i=1}^{m} w_{0,m}(i)(\boldsymbol{q}_{b+i-1} - \theta) = \sqrt{m}(\bar{\boldsymbol{q}}_{b,m} - \theta),$$

$$Z_{\ell,m} = \sum_{i=1}^{m} w_{\ell,m}(i)\boldsymbol{q}_{b+i-1}, \qquad 1 \le \ell \le K,$$

$$S_{K,m}^{\mathrm{SER}} = \frac{1}{K}\sum_{\ell=1}^{K} Z_{\ell,m} Z_{\ell,m}^{\top}. \tag{89}$$

The series statistic is

$$T_{K,m}^{\mathrm{SER}}(\theta) = Z_{0,m}(\theta)^{\top}(S_{K,m}^{\mathrm{SER}})^{-1} Z_{0,m}(\theta).$$

A common nonsingular transformation of all $Z_{\ell,m}$ leaves this statistic unchanged, avoiding estimation of D and Σ.

Theorem E.2 (Exact Gaussian F law for every orthonormal basis)**.** *Let $Z_1, \dots, Z_m$ be i.i.d. $N_k(0, I_k)$ and let $w_{0,m}, \dots, w_{K,m}$ satisfy* (67)*. Put*

$$X_\ell = \sum_{i=1}^{m} w_{\ell,m}(i) Z_i, \qquad 0 \le \ell \le K, \qquad \widehat{S}_K = K^{-1}\sum_{\ell=1}^{K} X_\ell X_\ell^{\top}.$$

Then $X_0, \dots, X_K$ are independent $N_k(0, I_k)$ vectors and, for $K \ge k$,

$$\frac{K-k+1}{kK} X_0^{\top} \widehat{S}_K^{-1} X_0 \sim F_{k,K-k+1}. \tag{90}$$

Under assumptions C.5 and C.6, the same relation holds asymptotically for fixed K provided the triangular contrast array also satisfies the Lindeberg leverage condition

$$\ell \max_{1 \le i \le m} \left\{ \sum_{r=0}^{K} w_{r,m}(i)^2 \right\}^{1/2} \longrightarrow 0 \tag{91}$$

and the online radius $r_{m,K}(\delta_m)$ tends to zero for some $\delta_m \downarrow 0$. Both the fixed-K DCT and block-Haar systems satisfy (91)*.*

The analytic critical value is

$$\frac{kK}{K-k+1} F_{k,K-k+1}^{-1}(1-\alpha). \tag{92}$$

The default finite-sample method combines the block-Haar basis with finite-score calibration; (92) provides the analytic reference value.

E.4 INDEPENDENT-BLOCK HOTELLING PROCEDURE

For the Hotelling (HOT) comparison, partition users into B_0 independent groups. Restart the recursion in each group, discard a within-group burn-in, and let $\widehat{\boldsymbol{Q}}_1, \dots, \widehat{\boldsymbol{Q}}_{B_0}$ be the group tail averages. Put

$$\widetilde{\boldsymbol{Q}} = B_0^{-1}\sum_{r=1}^{B_0} \widehat{\boldsymbol{Q}}_r, \qquad S_{B_0} = \frac{1}{B_0 - 1}\sum_{r=1}^{B_0} (\widehat{\boldsymbol{Q}}_r - \widetilde{\boldsymbol{Q}})(\widehat{\boldsymbol{Q}}_r - \widetilde{\boldsymbol{Q}})^{\top},$$

and

$$T_{B_0}^{\mathrm{HOT}}(\theta) = B_0(\widetilde{\boldsymbol{Q}} - \theta)^{\top} S_{B_0}^{-1}(\widetilde{\boldsymbol{Q}} - \theta).$$

For independent Gaussian block estimators (Hotelling, 1931),

$$\frac{B_0 - k}{k(B_0 - 1)} T_{B_0}^{\mathrm{HOT}}(\boldsymbol{Q}) \sim F_{k,B_0-k}.$$

This method provides genuinely independent covariance replications, but it restarts stochastic approximation B_0 times and replaces a length-n recursion by recursions of length approximately n/B_0.

E.5 Finite-score calibration and singular normalizers

The truth-score distribution is fully known from $(\boldsymbol{\tau}, \varepsilon, \mathsf{Q}, A)$ because (1) is known. For a selected sample size and weight system, one can therefore generate the frozen statistic T_m° without using private data; set $T_m^\circ = +\infty$ whenever its normalizer is singular. For any public calibration rule, let

$$d_{\mathrm{cal}} = |\mathbb{P}\{\text{the frozen rule accepts}\} - (1-\alpha)|$$

denote its unconditional frozen-experiment calibration error, including any public simulation and boundary randomization.

When the finite support is manageable, enumerate the frozen law and choose a critical value $c \in [0, \infty]$ and boundary probability $\eta \in [0, 1]$ satisfying

$$\mathbb{P}(T_m^\circ < c) + \eta \mathbb{P}(T_m^\circ = c) = 1 - \alpha. \tag{93}$$

Acceptance at equality with probability η yields $d_{\mathrm{cal}} = 0$, including for a lattice-valued statistic. A nonrandomized upper quantile is conservative.

Let $p_{\infty,m} = \mathbb{P}(T_m^\circ = +\infty)$. If $p_{\infty,m} \le \alpha$, the exact rule in (93) can be chosen to reject every singular realization and to use a finite critical value on the nonsingular support. If $p_{\infty,m} > \alpha$, no finite-critical rule that always rejects a singular normalizer can attain acceptance probability $1 - \alpha$; the sample size, block length, or number of contrasts must be increased.

Finite Monte Carlo admits two implementations. First, an empirical finite-critical rule estimates the frozen CDF. Conditional on the simulated reference library $\mathcal{R}$, write $d_{\mathrm{MC}}(\mathcal{R})$ for the resulting frozen calibration error. With B_{cal} reference draws, the Dvoretzky–Kiefer–Wolfowitz event gives

$$d_{\mathrm{MC}}(\mathcal{R}) \le 2\sqrt{\frac{\log(2/\delta_{\mathrm{cal}})}{2B_{\mathrm{cal}}}} \tag{94}$$

with probability at least $1 - \delta_{\mathrm{cal}}$ over $\mathcal{R}$, after empirical boundary randomization. Consequently, the unconditional error is bounded by the right-hand side of (94) plus δ_{cal}. A one-sided binomial upper confidence bound separately certifies $p_{\infty,m} \le \alpha$. Second, one may rank the observed statistic together with B_{cal} independent frozen statistics after attaching continuous tie breakers. The exchangeable randomized-rank rule has $d_{\mathrm{cal}} = 0$ jointly over the reference simulation. At $+\infty$, acceptance returns the full parameter cone and rejection returns the empty set, preserving exchangeability at the singular atom. The numerical study evaluates the independent-holdout probability $\mathbb{P}\{T_m^\circ \le c\}$ for the empirical finite critical value.

For the online-to-frozen transfer, smooth the discrete statistic with an independent $U \sim \mathrm{Unif}[-1, 1]$: for $h > 0$, define

$$\widetilde{T}_{m,h} = \begin{cases} T_m + hU, & T_m < \infty, \\ +\infty, & T_m = +\infty. \end{cases} \tag{95}$$

The finite component of the frozen law then has density at most $(2h)^{-1}$, which controls the local-concentration term. The singular mass remains at $+\infty$.

E.6 Haar contrasts: spectral control and inference

The block-Haar basis provides a lower-tail certificate under the exact non-Gaussian score law. Its proof is a finite-dimensional small-ball argument in the spirit of Koltchinskii & Mendelson (2015). For every unit vector v,

$$\mathbb{E}\left|v^\top \Sigma^{-1/2} \xi^*\right|^3 \le \beta_3.$$

With Φ the standard normal CDF, let

$$p_G = 2\{1 - \Phi(1/2)\}.$$

Theorem E.3 (Haar finite-score spectral certificate)**.** *Use the block-Haar basis* (69) *with half-block length $s = \lfloor m/(2K) \rfloor$. In the frozen standardized experiment, the K covariance coefficients $X_1^\circ, \dots, X_K^\circ$ are independent and isotropic. Put $S_{K,m}^\circ = K^{-1} \sum_{\ell=1}^K X_\ell^\circ X_\ell^{\circ\top}$. For every unit v,*

$$\mathbb{P}\{|v^\top X_1^\circ| \ge 1/2\} \ge p_s := p_G - \frac{2C_{\mathrm{BE}}\beta_3}{\sqrt{2s}}, \tag{96}$$

where C_{BE} is any valid scalar Berry–Esseen constant. If $p_s > 0$ and

$$16\sqrt{\frac{k}{K}} + \sqrt{\frac{2\log(1/\delta)}{K}} \le \frac{p_s}{2}, \tag{97}$$

then

$$\mathbb{P}\left\{\lambda_{\min}(S^{\circ}_{K,m}) < \frac{p_s}{32}\right\} \le \delta. \tag{98}$$

In particular, if

$$s \ge \frac{8C^2_{\mathrm{BE}}\beta_3^2}{p_G^2} \quad \textit{and} \quad K \ge C\{k + \log(1/\delta)\},$$

then $p_s \ge p_G/2$ and the lower threshold in (98) *is the numerical constant $p_G/64$.*

Remark E.4 (Meaning of the sample-size condition)**.** *For balanced kRR, $\beta_3^2 \lesssim k^3$. With $K \asymp k$, the sufficient condition $m \gtrsim K\beta_3^2$ is therefore of order k^4. This certificate follows from a uniform one-dimensional Berry–Esseen bound. In computation, the exact frozen-score lower eigenvalue distribution provides a sharper offline diagnostic.*

The Haar score-calibrated (HSC) construction leads to a complete inferential algorithm. The privacy mechanism is fixed from public inputs, the online estimator is run once, and the reference law is generated offline from the known target interval probabilities $\boldsymbol{\pi}^*$. No density or covariance estimate is required.

Algorithm 3 gives the implementation.

The theorem-level certificate uses theorem E.3. In computation, K should be at least a constant multiple of k and leave sufficiently long half-blocks. The exactly simulable frozen spectral distribution is reported alongside the analytic certificate.

For $c_{\mathrm{eff}} > 0$, the untruncated confidence set is an ellipsoid centered at $\bar{\boldsymbol{q}}_{b,m}$ with volume

$$\frac{\pi^{k/2}}{\Gamma(k/2+1)}\left(\frac{c_{\mathrm{eff}}}{m}\right)^{k/2}\det(S^{\mathrm{SER}}_{K,m})^{1/2}.$$

Its projection onto coordinate j has half-width $\{c_{\mathrm{eff}}(S^{\mathrm{SER}}_{K,m})_{jj}/m\}^{1/2}$ before intersection with the ordered parameter set. These are projections of a joint ellipsoid; rectangular simultaneous intervals use max-type critical values.

F Fixed-confidence coverage errors

Superscripts identify the procedure whose coverage is evaluated. For a confidence set $\mathcal{C}_{1-\alpha}$, write $\mathrm{CE}_\alpha(\mathcal{C}) = |\mathbb{P}\{\boldsymbol{Q} \in \mathcal{C}\} - 1 + \alpha|$. Throughout this section, $\alpha \in [\alpha_0, 1-\alpha_0]$ for a fixed $\alpha_0 \in (0, 1/2)$.

F.1 Coverage transfer and spectral thresholds

For a real random variable T, let $\omega_T(c,t) = \mathbb{P}\{|T-c| \le t\}$ denote its local concentration function. Also let

$$L_K(\delta) = k + \log\frac{4(K+1)}{\delta}, \qquad M_{m,K}(\delta) = C\left\{\sqrt{L_K(\delta)} + \frac{\ell\mu(\mathcal{W})L_K(\delta)}{\sqrt{m}}\right\}. \tag{99}$$

The technical bounds in lemmas K.1 and K.2 imply that, outside probability at most $\pi_{b,n} + 4\delta$, the online endpoint and series normalizer differ from their frozen-score counterparts by at most $(1 + 2M_{m,K})r_{m,K} + r^2_{m,K}$. The quadratic-form perturbation lemma in lemma K.3 shows why every random-normalizer result contains a lower spectral threshold u: inversion amplifies an object error ρ by u^{-2}. The bounds display reference calibration, coupling failure, the spectral lower tail, and local anti-concentration as separate terms.

Gaussian spectral thresholds. Write $S^G_K = K^{-1}\sum_{\ell=1}^K G_\ell G_\ell^\top$ for independent $G_\ell \sim N_k(0, I_k)$, and $S^{\mathrm{SN,G}}_m = \sum_{\ell=1}^{m-1}\bar{\lambda}_{\ell,m}G_\ell G_\ell^\top$, using the weights in (87).

Proposition F.1 (Dimension-explicit Gaussian normalizer thresholds)**.** *Let $x \ge 0$.*

(i) For the K-mode equal-weight series normalizer,

$$u_{\mathrm{SER}}(K,k,x)=\left(1-\sqrt{\frac{k}{K}}-\frac{x}{\sqrt{K}}\right)_+^2$$

satisfies

$$\mathbb{P}\{\lambda_{\min}(S_K^G)<u_{\mathrm{SER}}(K,k,x)\}\le e^{-x^2/2}.$$

(ii) For the finite-m Gaussian integral normalizer and any $k\le J<m$,

$$u_{\mathrm{SN}}(J,k,x)=\frac{(\sqrt{J}-\sqrt{k}-x)_+^2}{\pi^2J^2}$$

satisfies

$$\mathbb{P}\{\lambda_{\min}(S_m^{\mathrm{SN,G}})<u_{\mathrm{SN}}(J,k,x)\}\le e^{-x^2/2}.$$

The same bound holds for the Brownian normalizer.

(iii) For a Gaussian sample covariance from B_0 observations, the threshold

$$u_{\mathrm{HOT}}(B_0,k,x)=\left(1-\sqrt{\frac{k}{B_0-1}}-\frac{x}{\sqrt{B_0-1}}\right)_+^2$$

has failure probability at most $e^{-x^2/2}$.

If $K=\lceil ck\rceil$ with $c>1$ and $x=(\sqrt{K}-\sqrt{k})/2$, then $u_{\mathrm{SER}}\ge(1-c^{-1/2})^2/4$, a positive constant, and the failure probability is at most $\exp\{-k(\sqrt{c}-1)^2/8+o(k)\}$. If $J=\lceil ck\rceil$ and the same choice of x is made, then

$$u_{\mathrm{SN}}(J,k,x)\ge\frac{(\sqrt{c}-1)^2}{4\pi^2c^2k}+o(k^{-1}).$$

The required reference-density bounds are stated and proved in lemma K.4. For K/k bounded above and away from one, the scaled series-F density near a fixed-confidence critical value is $O(k^{-1/2})$; for the integral Gaussian pivot it is $O(k)$.

F.2 Oracle Wald coverage

Theorem F.2 (Oracle Wald fixed-confidence error)**.** *Let $r_{m,0}(\delta)$ be* (72) *for the endpoint weight only. If $r_{m,0}(\delta)$ is smaller than a fixed neighborhood of the chi-square critical radius, then*

$$\mathrm{CE}_\alpha^{\mathrm{or}}\le 1200\frac{k^{1/4}\beta_3}{\sqrt{m}}+\pi_{b,n}+3\delta+C_{\alpha_0}r_{m,0}(\delta). \tag{100}$$

In the exact frozen-score experiment, replacing the chi-square critical value by an exact randomized critical rule removes the reference-calibration Berry–Esseen term. Transferring that discrete rule to the online estimator still requires its local concentration function, or the smoothing device in (95)*.*

Corollary F.3 (Oracle kRR dependence on m,ε,k)**.** *For kRR,*

$$\mathrm{CE}_\alpha^{\mathrm{or}}\le C\left[\frac{k^{1/4}(k+1)}{\sqrt{\mu_\varepsilon m}}+\mathfrak{R}_{k,\varepsilon,0}^{\mathrm{kRR}}m^{-1/6}\Lambda_{m,0}(\delta)\right]+\pi_{b,n}+3\delta.$$

For $0<\varepsilon\le 1$ and balanced interval probabilities,

$$\mathrm{CE}_\alpha^{\mathrm{or}}\le C_f\left[\frac{k^{7/4}}{\sqrt{m}}+\frac{k^{5/2}}{\varepsilon}m^{-1/6}\{k+\log(m/\delta)\}\right]+\pi_{b,n}+3\delta. \tag{101}$$

The first term is essentially privacy invariant after score standardization; the privacy loss enters the displayed online nonlinear remainder through G^2/σ_-. The bound is truncated at one.

F.3 PLUG-IN WALD AND AN EXPLICIT PRIVATE DENSITY CALIBRATION

The plug-in procedure uses disjoint users for the main recursion, pilot centers, and density calibration. Disjointness permits every user to use the full privacy budget ε.

Assumption F.4 (Density calibration smoothness)**.** *On a neighborhood of all target quantiles, f is twice continuously differentiable and $\sup_x |f''(x)| \le L_{f,2}$.*

Let $\widetilde{Q}_j$ be independent pilot centers satisfying

$$\mathbb{P}\left\{\max_j \left|\widetilde{Q}_j - Q_j\right| > r_p(\delta)\right\} \le \delta. \tag{102}$$

A concrete pilot radius follows from an independent LDP recursion. Let m_p be its retained length, let π_p be its localization-failure bound, and define

$$\omega_\infty = \max_{1\le j\le k} \sqrt{\Omega_{jj}}.$$

Proposition F.5 (Finite-confidence common-shift density probe)**.** *Let $G_0 = \max_y \|\boldsymbol{\tau} - a_{r_0,y}\|_2 \le G_{\max}$. Under assumptions C.5 and F.4, suppose the pilot guarantee* (102) *holds and $h + r_p(\delta)$ is small enough that every interval $[\widetilde{Q}_j - h, \widetilde{Q}_j + h]$ lies in the stated smoothness neighborhood. Use the two joint categorical groups in* (24)*, with sizes $n_{2,+}$ and $n_{2,-}$. Then, with probability at least $1 - 2\delta$,*

$$\max_{1\le j\le k} |\widehat{f}_j^{\mathrm{raw}} - f(Q_j)| \le \frac{L_{f,2}}{6}h^2 + L_f r_p(\delta) + \frac{G_0\sqrt{2\log(4k/\delta)}}{2h}\left(n_{2,+}^{-1/2} + n_{2,-}^{-1/2}\right). \tag{103}$$

If the deterministic truncation interval contains $[f_-, f_+]$, the same bound holds for $\max_j |\widehat{f}_j - f(Q_j)|$. When both group proportions are bounded away from zero, balancing the first and third terms gives

$$\max_j |\widehat{f}_j - f(Q_j)| = O_{\mathbb{P}}\left[\left\{\frac{G_0^2\log(k/\delta)}{n_2}\right\}^{1/3} + r_p(\delta)\right]. \tag{104}$$

Consequently, theorem F.7 applies to the implemented common-shift probe by using (103) *in lemma K.6; the remaining online-transfer conditions are unchanged.*

For comparison, an alternative coordinate-split calibration assigns each of m_c independent users publicly to one of the $2k$ cells $(j,+)$ and $(j,-)$, with equal cell size $N_c = \lfloor m_c/(2k)\rfloor$. In cell $(j,\pm)$, the user forms $B = \mathbf{1}\{X \le \widetilde{Q}_j \pm h\}$ and sends binary randomized response

$$\mathbb{P}(Z = 1 \mid B) = c_\varepsilon + r_\varepsilon B, \qquad c_\varepsilon = \frac{1}{1+e^\varepsilon}, \qquad r_\varepsilon = \frac{e^\varepsilon - 1}{e^\varepsilon + 1} = \tanh(\varepsilon/2).$$

The decoded bit $(Z - c_\varepsilon)/r_\varepsilon$ is unbiased. Let $\widehat{F}_{j,\pm}$ be its cell average and set

$$\widehat{f}_j = \frac{\widehat{F}_{j,+} - \widehat{F}_{j,-}}{2h}.$$

Proposition F.6 (Private density-estimation error)**.** *Under assumptions C.5 and F.4, assume $m_c \ge 4k$. If $h + r_p(\delta) < \Delta_Q/4$, then with probability at least $1 - 2\delta$,*

$$\max_{1\le j\le k}\left|\widehat{f}_j - f(Q_j)\right| \le \frac{L_{f,2}}{6}h^2 + L_f r_p(\delta) + \frac{C}{r_\varepsilon h}\sqrt{\frac{k\log(4k/\delta)}{m_c}}. \tag{105}$$

The bandwidth

$$h \asymp \left\{\frac{k\log(4k/\delta)}{m_c r_\varepsilon^2}\right\}^{1/6} \tag{106}$$

yields

$$\max_j\left|\widehat{f}_j - f(Q_j)\right| \le C_f\left[\left\{\frac{k\log(4k/\delta)}{m_c r_\varepsilon^2}\right\}^{1/3} + r_p(\delta)\right] \tag{107}$$

with probability at least $1 - 2\delta$. For $0 < \varepsilon \le 1$, the calibration term is of order $\{k\log(k/\delta)/(m_c\varepsilon^2)\}^{1/3}$.

Let $\kappa_\Sigma = \lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ and $\delta_\Omega = \|\Omega^{-1/2}(\widehat{\Omega} - \Omega)\Omega^{-1/2}\|_{\mathrm{op}}$, the relative covariance error.

Theorem F.7 (Plug-in Wald coverage)**.** *For any deterministic $0 < d < 1/2$,*

$$\mathrm{CE}_\alpha^{\mathrm{plug}} \le 1200\frac{k^{1/4}\beta_3}{\sqrt{m}} + \pi_{b,n} + 3\delta + C_{\alpha_0}\{r_{m,0}(\delta) + \sqrt{k}\,d\} + \mathbb{P}(\delta_\Omega > d). \tag{108}$$

Combining either proposition F.5 or proposition F.6 with lemma K.6 gives a fully explicit choice of d. For the coordinate-split binary calibration, suppressing condition-number and density constants,

$$\textit{additional plug-in error} = O\left(\sqrt{k}\left[\left\{\frac{k\log(k/\delta)}{m_c\tanh^2(\varepsilon/2)}\right\}^{1/3} + r_p(\delta)\right]\right).$$

F.4 General random-normalizer coverage transfer

Let (X_m°, S_m°) be a chosen reference object and let $T_m^\circ = X_m^{\circ\top} S_m^{\circ -1} X_m^\circ$. Suppose the critical value c_m has reference calibration error

$$d_m^\circ = \left|\mathbb{P}(T_m^\circ \le c_m) - (1-\alpha)\right|.$$

Theorem F.8 (Coverage transfer)**.** *Suppose the online object (X_m, S_m) and its reference counterpart (X_m°, S_m°) can be coupled so that $\|X_m - X_m^\circ\|_2 + \|S_m - S_m^\circ\|_{\mathrm{op}} \le \rho$ outside an event of probability at most η. For every $M > 0$ and $2\rho \le u \le 1$,*

$$\begin{aligned}\mathrm{CE}_\alpha \le d_m^\circ + \eta + \mathbb{P}(\|X_m^\circ\| > M) + \mathbb{P}\{\lambda_{\min}(S_m^\circ) < u\} \\ + \omega_{T_m^\circ}\left(c_m, \frac{C(M+1)^2\rho}{u^2}\right).\end{aligned} \tag{109}$$

If the coupling good event is defined jointly so that it also includes $\|X_m^\circ\|_2 \le M$, the separate term $\mathbb{P}(\|X_m^\circ\| > M)$ is omitted and its failure probability is already included in η. The same statement holds for a critical rule generated by public randomness independent of the private data: d_m° is then defined jointly over the frozen object and that randomness, and the local-concentration bound must hold uniformly conditional on the public randomization. For an exactly calibrated h-smoothed frozen-score rule with finite critical value, $d_m^\circ = 0$ and the local-concentration term is at most $C(M+1)^2\rho/(hu^2)$.

F.5 Haar score-calibrated coverage

Theorem F.9 (Haar finite-score coverage)**.** *Use block-Haar contrasts with half-block length s, a public finite-critical calibration rule for the h-smoothed frozen statistic, and smoothing parameter $h > 0$. Let d_{cal} be the frozen-experiment calibration error of that rule. Let p_s be* (96)*. Suppose $p_s > 0$,* (97) *holds, and put*

$$u_H = \frac{p_s}{32}.$$

Let $M = M_{m,K}(\delta)$, $r = r_{m,K}(\delta)$, and $\rho = (1+2M)r + r^2$. If $2\rho \le u_H$ and the frozen critical value is finite, then

$$\mathrm{CE}_\alpha^{\mathrm{HSC}} \le d_{\mathrm{cal}} + \pi_{b,n} + 5\delta + \frac{C(M+1)^2}{hu_H^2}\{(1+2M)r + r^2\}. \tag{110}$$

At step exponent $a = 2/3$, if the leverage correction in (99) *is no larger than its square-root term and the quadratic term is smaller than the linear term,*

$$\mathrm{CE}_\alpha^{\mathrm{HSC}} \le d_{\mathrm{cal}} + \pi_{b,n} + 5\delta + \frac{C\mathfrak{R}_{k,\varepsilon,K}(\mathcal{W})}{hu_H^2} m^{-1/6}\Lambda_{m,K}(\delta)^{5/2}. \tag{111}$$

For balanced kRR, $0 < \varepsilon \le 1$, and $K \asymp k$,

$$\mathrm{CE}_\alpha^{\mathrm{HSC}} \le d_{\mathrm{cal}} + \pi_{b,n} + 5\delta + \frac{C_f}{hu_H^2}\frac{k^{5/2}}{\varepsilon} m^{-1/6}\{k + \log(m/\delta)\}^{5/2}. \tag{112}$$

The powers of k and the logarithm in (112) *are conservative consequences of uniform nets and worst-case decoded-score bounds; the theorem's sharper, mechanism-specific form is* (110)*. Exact enumeration has $d_{\mathrm{cal}} = 0$, whereas an empirical finite-critical rule carries its reported d_{cal}. The exchangeable-rank variant also has exact frozen level, but a singular accepted realization returns the full cone and is therefore outside the finite-ellipsoid statement of this theorem.*

Remark F.10 (Unsmoothing). *Without smoothing, exact randomized calibration remains exact in the frozen experiment, but* (110) *must replace the final term by the local concentration function in* (109). *No universal density bound exists for every finite-alphabet score distribution. The theory permits both unsmoothed randomized calibration and small relative smoothing; the numerical study evaluates the unsmoothed event* $T_m^\circ \le c$.

F.6 Analytic series F coverage

To compare the exact finite-score calibration with analytic F calibration, let $\mathbf{X}^\circ = (X_0^{\circ\top}, \dots, X_K^{\circ\top})^\top$ and let $\mathbf{G}$ contain $K+1$ independent $N_k(0, I_k)$ vectors. Write W_1 for Wasserstein distance with Euclidean cost and $\mathcal{L}$ for a probability law. Denote by

$$\Gamma_{m,K} = W_1(\mathcal{L}(\mathbf{X}^\circ), \mathcal{L}(\mathbf{G}))$$

a valid Wasserstein–one normal-approximation bound. The independent-summand Wasserstein bound of Raič (2019, Section 3, equation (26)) gives, in dimension $p = (K+1)k$,

$$\Gamma_{m,K} \le C_p \beta_3 \sum_{i=1}^{m} \|(w_{0,m}(i), \dots, w_{K,m}(i))\|_2^3 . \tag{113}$$

For both the DCT and block-Haar systems,

$$\sum_{i=1}^{m} \|(w_{0,m}(i), \dots, w_{K,m}(i))\|_2^3 \le \frac{C(1+K)^{3/2}}{\sqrt{m}}. \tag{114}$$

Here C_p is the dimension-dependent constant from the selected quantitative multivariate central limit theorem.

By the coupling characterization of W_1, for any $0 < \eta_G < 1$ there is a coupling with

$$\mathbb{P}\{\|\mathbf{X}^\circ - \mathbf{G}\|_2 > \Gamma_{m,K}/\eta_G\} \le \eta_G. \tag{115}$$

Put

$$\begin{aligned} r_G &= \Gamma_{m,K}/\eta_G, & r_{\mathrm{tot}} &= r_{m,K}(\delta) + r_G, \\ M_G &= \sqrt{k} + \sqrt{2\log\{(K+1)/\delta\}}, & \rho_G &= (1+2M_G) r_{\mathrm{tot}} + r_{\mathrm{tot}}^2. \end{aligned} \tag{116}$$

Theorem F.11 (Analytic series transfer in fixed dimension). *Let $K \ge k$, choose $x \ge 0$, and put $u = u_{\mathrm{SER}}(K, k, x)$. Let $A^{\mathrm{SER}}_{k,K,\alpha}$ bound the Gaussian pivot density near its critical value, as in lemma K.4. Use M_G and ρ_G from* (116). *If $2\rho_G \le u$, then the analytic F procedure satisfies*

$$\mathrm{CE}_\alpha^{\mathrm{SER,F}} \le \pi_{b,n} + 4\delta + \eta_G + e^{-x^2/2} + \frac{C A^{\mathrm{SER}}_{k,K,\alpha} (M_G+1)^2 \rho_G}{u^2}. \tag{117}$$

If $K = \lceil ck \rceil$ for fixed $c > 1$ and $x = (\sqrt{K} - \sqrt{k})/2$, then

$$\mathrm{CE}_\alpha^{\mathrm{SER,F}} \le \pi_{b,n} + 4\delta + \eta_G + e^{-c_c k} + C_{c,\alpha_0} \frac{k + \log(1/\delta)}{\sqrt{k}} \rho_G. \tag{118}$$

For fixed (k, K), taking $\eta_G = \sqrt{\Gamma_{m,K}}$ makes the Gaussian replacement failure and coefficient radius both of order $\Gamma_{m,K}^{1/2}$, which is $O(m^{-1/4})$ when (113) *is $O(m^{-1/2})$.*

F.7 Integral self-normalization coverage

The path theorem gives a fully explicit online-to-frozen comparison without any Gaussian approximation. Put $V_m^\circ = m^{-1} \sum_{s=1}^{m-1} \mathbb{U}_m^\circ(s) \mathbb{U}_m^\circ(s)^\top$. Define

$$T_m^{\mathrm{SN},\circ} = \mathbb{P}_m^\circ(m)^\top (V_m^\circ)^{-1} \mathbb{P}_m^\circ(m), \qquad s_{m,\mathrm{SN}}^\circ(u) = \mathbb{P}\{\lambda_{\min}(V_m^\circ) < u\}.$$

Exact draws from both $T_m^{\mathrm{SN},\circ}$ and the lower-eigenvalue event defining $s_{m,\mathrm{SN}}^\circ$ can be generated from the known truth-score law. Their probabilities are evaluated by enumeration when feasible or by a separately quantified Monte Carlo calculation. For later use, set

$$M_m^{\mathrm{path}}(\delta) = C \left\{ \sqrt{k + \log \frac{4m}{\delta}} + \frac{\ell\{k + \log(4m/\delta)\}}{\sqrt{m}} \right\}. \tag{119}$$

Theorem F.12 (Exact-score integral-SN coverage)**.** *Use a public finite-critical frozen-score calibration rule with error d_{cal} and the smoothed statistic* (95)*. Let $M = M_m^{\mathrm{path}}(\delta)$, $r = r_m^{\mathrm{path}}(\delta)$, and*

$$\rho_{\mathrm{SN}}^{\mathrm{score}} = C\{(1+M)r + r^2\}. \tag{120}$$

For every $u \in (0,1]$ satisfying $2\rho_{\mathrm{SN}}^{\mathrm{score}} \le u$ and every finite frozen critical value,

$$\mathrm{CE}_\alpha^{\mathrm{SN,score}} \le d_{\mathrm{cal}} + \pi_{b,n} + 4\delta + s^\circ_{m,\mathrm{SN}}(u) + \frac{C(M+1)^2\rho_{\mathrm{SN}}^{\mathrm{score}}}{hu^2}. \tag{121}$$

Exact enumeration has $d_{\mathrm{cal}} = 0$, whereas an empirical finite-critical rule contributes its certified d_{cal}. The exact-rank variant uses the full/empty decision on singular paths described in appendix E. At $a = 2/3$, if the leverage correction in (119) *and the quadratic term in* (120) *are smaller-order,*

$$\mathrm{CE}_\alpha^{\mathrm{SN,score}} \le d_{\mathrm{cal}} + \pi_{b,n} + 4\delta + s^\circ_{m,\mathrm{SN}}(u) + \frac{C\mathfrak{R}_{k,\varepsilon}^{\mathrm{path}}}{hu^2} m^{-1/6}\{k + \log(4m/\delta)\}^{5/2}. \tag{122}$$

For balanced kRR and $0 < \varepsilon \le 1$, the last term is at most

$$\frac{C_f}{hu^2}\frac{k^{5/2}}{\varepsilon} m^{-1/6}\{k + \log(4m/\delta)\}^{5/2}. \tag{123}$$

Equation (121) *retains the actual integral-normalizer tail $s^\circ_{m,\mathrm{SN}}(u)$, which can be evaluated from the known frozen-score law.*

For analytic Brownian or finite-m Gaussian calibration, a second comparison is needed. Let a coupling of the frozen score path and a Gaussian partial-sum path satisfy

$$\mathbb{P}\left\{\max_{0\le s\le m} \|\mathbb{P}_m^\circ(s) - \mathbb{G}_m(s)\|_2 > r_m^{\mathrm{proc}}\right\} \le \eta_m^{\mathrm{proc}}, \tag{124}$$

where $\mathbb{G}_m$ is formed from i.i.d. $N_k(0, I_k)$ increments. Because the frozen score is bounded, multidimensional Komlós–Major–Tusnády (KMT)-type results yield $r_m^{\mathrm{proc}} = O_{k,\ell}\{\log(m/\eta_m^{\mathrm{proc}})/\sqrt{m}\}$ for fixed k (Zaitsev, 1998). The theorem keeps the process approximation pair $(r_m^{\mathrm{proc}}, \eta_m^{\mathrm{proc}})$ explicit.

A Brownian critical value also introduces a purely Gaussian discretization error. Let $T_m^{\mathrm{SN},G}$ be the exact finite-grid Gaussian statistic and let $c_{k,\alpha}^B$ be the $(1-\alpha)$ quantile of the Brownian statistic U_k in (82). The finite-grid comparison in proposition K.9 supplies an explicit bound; it is zero when the exact finite-m Gaussian critical value is used.

If $c_{m,\alpha}^G$ denotes the critical value actually used, define

$$d_{m,k}^{\mathrm{ref}} = \left|\mathbb{P}(T_m^{\mathrm{SN},G} \le c_{m,\alpha}^G) - (1-\alpha)\right|.$$

This term is zero for the exact finite-m Gaussian critical value. For a Brownian critical value it is the finite-grid-to-Brownian calibration error and can be simulated independently. Let

$$\rho_{\mathrm{SN}}^G = \rho_{\mathrm{SN}}^{\mathrm{score}} + C\left[\{1 + M_m^{\mathrm{path}}(\delta) + M_m^G(\delta)\}r_m^{\mathrm{proc}} + (r_m^{\mathrm{proc}})^2\right], \tag{125}$$

where $M_m^G(\delta) = C\sqrt{k + \log(4m/\delta)}$ controls the Gaussian path maximum.

Theorem F.13 (Analytic integral-SN transfer)**.** *Choose $k \le J < m$, $x \ge 0$, and $u = u_{\mathrm{SN}}(J,k,x)$. If $2\rho_{\mathrm{SN}}^G \le u$, then for $k \ge 2$,*

$$\mathrm{CE}_\alpha^{\mathrm{SN}} \le d_{m,k}^{\mathrm{ref}} + \pi_{b,n} + 5\delta + \eta_m^{\mathrm{proc}} + e^{-x^2/2} + C\frac{k\{k + \log(4m/\delta)\}\rho_{\mathrm{SN}}^G}{u^2}. \tag{126}$$

If $J = \lceil ck\rceil$, $c > 1$, and $x = (\sqrt{J} - \sqrt{k})/2$, then

$$\mathrm{CE}_\alpha^{\mathrm{SN}} \le d_{m,k}^{\mathrm{ref}} + \pi_{b,n} + 5\delta + \eta_m^{\mathrm{proc}} + e^{-c_c k} + C_c k^3\{k + \log(4m/\delta)\}\rho_{\mathrm{SN}}^G. \tag{127}$$

Thus, at proportional J/k, inverse perturbation contributes a factor of order k^2 because the certified spectral scale is $u \asymp k^{-1}$, and the analytic pivot density contributes another factor of order k. The online component of ρ_{SN}^G is given by (76)–(120)*; only $(r_m^{\mathrm{proc}}, \eta_m^{\mathrm{proc}})$ belongs to the additional non-Gaussian process approximation.*

Without smoothing, the exact-score theorem replaces its last term by the local concentration function of $T_m^{\mathrm{SN},\circ}$.

F.8 Independent-block Hotelling coverage

Assume for simplicity that every restarted block has the same usable tail length m_B. Under the common transformation $\Sigma^{-1/2}D\sqrt{m_B}$, write the rth block estimator as $\widehat{X}_r$, its frozen endpoint as X_r°, and let $G_r \sim N_k(0, I_k)$ be independent. Use

$$r_B^{\mathrm{on}} = r_{m_B,0}(\delta/B_0),$$

so a union bound over the restarted recursions gives online-to-frozen failure at most $B_0\pi_B + 3\delta$, where π_B is the block analogue of (155).

Let

$$\Gamma_{B_0,m_B} = W_1\left\{\mathcal{L}(X_1^\circ,\dots,X_{B_0}^\circ), \mathcal{L}(G_1,\dots,G_{B_0})\right\}.$$

Applying Raič (2019, Section 3, equation (26)) to the stacked independent block scores gives

$$\Gamma_{B_0,m_B} \le C_{kB_0}\frac{\beta_3}{\sqrt{m_B}}. \tag{128}$$

For $0 < \eta_G < 1$, put $r_B^G = \Gamma_{B_0,m_B}/\eta_G$ and $r_B = r_B^{\mathrm{on}} + r_B^G$. Let

$$M_B^G(\delta) = C\left\{\sqrt{k} + \sqrt{\log(4B_0/\delta)}\right\}, \qquad \rho_B = C\{(1+M_B^G)r_B + r_B^2\}.$$

The centered sample covariance and sample mean are both perturbed by at most ρ_B on the resulting coupling event.

Theorem F.14 (Block Hotelling transfer)**.** *Choose $x \ge 0$ and $u = u_{\mathrm{HOT}}(B_0, k, x)$. Let $A^{\mathrm{HOT}}_{k,B_0-1,\alpha}$ bound the Gaussian Hotelling-pivot density near its critical value. If $2\rho_B \le u$, then*

$$\mathrm{CE}_\alpha^{\mathrm{HOT}} \le B_0\pi_B + 4\delta + \eta_G + e^{-x^2/2} + \frac{CA^{\mathrm{HOT}}_{k,B_0-1,\alpha}\{k+\log(4B_0/\delta)\}\rho_B}{u^2}. \tag{129}$$

If $B_0 - 1 = \lceil ck\rceil$, $c > 1$, and $x = (\sqrt{B_0-1} - \sqrt{k})/2$, then

$$\mathrm{CE}_\alpha^{\mathrm{HOT}} \le B_0\pi_B + 4\delta + \eta_G + e^{-c_c k} + C_{c,\alpha_0}\frac{k+\log(4B_0/\delta)}{\sqrt{k}}\rho_B. \tag{130}$$

At $a = 2/3$, balanced kRR gives

$$r_B^{\mathrm{on}} \le C_f\frac{k^{5/2}}{\varepsilon}m_B^{-1/6}\left\{k + \log\frac{m_B B_0}{\delta}\right\}. \tag{131}$$

If the usable block length satisfies $m_B \asymp n/B_0$ and $B_0 \asymp k$, its leading sample-size/privacy factor is therefore $k^{8/3}\varepsilon^{-1}n^{-1/6}$ before norm and anti-concentration multipliers. In addition, localization is paid B_0 times and the Gaussian replacement in (128) *is based on only m_B observations per block.*

An exact-score block calibration may replace the Gaussian F critical value. Its valid finite-sample bound is obtained from Theorem F.8 using the exactly simulable block spectral tail and local concentration function; it removes $(\Gamma_{B_0,m_B}, \eta_G)$, while the block construction still restarts the recursion.

F.9 Privacy and dimension dependence

Coverage error and confidence-set diameter have different privacy dependence. Under kRR, the jth decoded CDF coordinate is

$$a_{Y,j} = \delta_\varepsilon^{-1}\{\mathbf{1}(Y \le j-1) - qj\},$$

so

$$\mathrm{Var}(a_{Y,j}) = \delta_\varepsilon^{-2}\nu_j(1-\nu_j), \qquad \nu_j = jq + \delta_\varepsilon\tau_j. \tag{132}$$

Consequently, the exact first-order marginal standard error is

$$\mathrm{se}(\widehat{Q}_j) \asymp \frac{\sqrt{\nu_j(1-\nu_j)}}{\delta_\varepsilon f(Q_j)\sqrt{m}}.$$

For $0 < \varepsilon \le 1$, $\delta_\varepsilon \asymp \varepsilon/k$ and $\nu_j \asymp j/k$ away from the right boundary. Hence

$$\mathrm{se}(\widehat{Q}_j) \asymp \frac{\sqrt{j(k-j)}}{\varepsilon f(Q_j)\sqrt{m}}$$

for interior indices, up to endpoint constants. The largest coordinate width is of order $k/(\varepsilon\sqrt{m})$, while a tail coordinate with fixed j has order $\sqrt{k}/(\varepsilon\sqrt{m})$. Thus privacy inflates central and tail-coordinate widths at different rates.

After standardization by $\Sigma^{-1/2}$, the frozen-score Berry–Esseen term is governed by β_3. Privacy re-enters finite-sample coverage through the nonlinear online path because the unstandardized decoded update has radius $G \asymp \sqrt{k}/\delta_\varepsilon$. For kRR this leads to the conservative $\varepsilon^{-1}m^{-1/6}$ term in (101) and (112). Plug-in density estimation has exponent $\tanh(\varepsilon/2)^{-2/3} \asymp \varepsilon^{-2/3}$ because it balances second-order finite-difference bias with binary-RR variance.

Table 8: Leading fixed-confidence coverage-error terms under a common benchmark regime. Assume balanced target intervals, direct $(k+1)$-ary randomized response, $0 < \varepsilon \le 1$, step exponent $a = 2/3$, fixed confidence level, bounded density and conditioning constants, and $K \asymp B_0 \asymp k$. Let $L_m = k + \log(m/\delta)$. The rate column gives conservative sufficient upper bounds and omits localization and failure probabilities. The Haar row also assumes h, u_H bounded away from zero; $L_B = k + \log(n/\delta)$.

Procedure	Reference and spectral requirement	Leading explicit coverage-error terms
Oracle Wald	Deterministic oracle covariance	$k^{7/4}m^{-1/2} + \varepsilon^{-1}k^{5/2}L_m m^{-1/6}$.
Plug-in Wald	Independent private pilot and density sample	Oracle terms $+\sqrt{k}[\{k\log(k/\delta)/(m_c\varepsilon^2)\}^{1/3} + r_p]$.
Haar score-calibrated	Finite-score calibration; $u_H \asymp 1$ when theorem E.3 is certified	$d_{\mathrm{cal}} + \varepsilon^{-1}k^{5/2}L_m^{5/2}m^{-1/6}$.
Analytic series F	Gaussian F reference; constant spectral threshold for $K/k > 1$	$\varepsilon^{-1}k^2L_m^{5/2}m^{-1/6} + O_{k,K}(m^{-1/4})$.
Integral SN, exact score	Exact frozen-score law; actual tail $s^\circ_{m,\mathrm{SN}}(u)$	$d_{\mathrm{cal}} + s^\circ_{m,\mathrm{SN}}(u) + \varepsilon^{-1}k^{5/2}L_m^{5/2}m^{-1/6}/(hu^2)$.
Integral SN, analytic	Finite-grid Gaussian or Brownian reference; certified $u \asymp k^{-1}$	Process/calibration error $+\varepsilon^{-1}k^{11/2}L_m^{5/2}m^{-1/6}$.
Block Hotelling	B_0 restarted recursions; constant Gaussian spectral threshold	$\varepsilon^{-1}k^{13/6}n^{-1/6}L_B^{5/2} + O_{k,B_0}\{(n/k)^{-1/4}\}$.

Coverage comparison. The reductions in table 8 use $M \asymp L_m^{1/2}$, $\rho \asymp Mr$, the fixed-level reference-density orders $k^{-1/2}$ for series and Hotelling pivots and k for integral self-normalization, and the spectral scales $u_{\mathrm{SER}} \asymp u_{\mathrm{HOT}} \asymp 1$ and $u_{\mathrm{SN}} \asymp k^{-1}$. They follow from the preceding nonasymptotic theorems after suppressing constants that depend on the fixed confidence level and the stated density and conditioning bounds. Under balanced kRR, $a = 2/3$, and a balanced burn-in, the localization term in (66) is $\pi_{b,n} \le Cm\exp\{-c\Delta_Q^2\varepsilon^2 m^{2/3}/k^3\}$, up to fixed mechanism and step-size constants. This expression retains any dimension dependence of the quantile separation Δ_Q.

Frozen calibration and online error. The probability $\mathbb{P}\{T^\circ_m \le c\}$ and the coverage of the online confidence set are different estimands. In the frozen experiment the standardized score law is public. To make the privacy dependence precise, write direct d-ary randomized response as $\mathsf{Q} = q\mathbf{1}_d\mathbf{1}_d^\top + \delta_\varepsilon I_d$. For the report law μ in (41), the decoded frozen score at report y is

$$-\delta_\varepsilon^{-1}P(e_y - \mu).$$

Thus every frozen quadratic pivot invariant under a common nonsingular transformation depends on ε through μ rather than the scalar δ_ε^{-1}. On a balanced grid $\boldsymbol{\pi}^* = d^{-1}\mathbf{1}_d$, hence $\mu = d^{-1}\mathbf{1}_d$ for every $\varepsilon > 0$ and the standardized frozen laws of oracle Wald and the four self-normalized procedures are exactly ε-invariant. For plug-in Wald, the private pilot and density estimate also enter the plug-in law; on a nonbalanced grid, μ can depend on ε. An enumerated or randomized finite-score rule has acceptance probability $1 - \alpha$ with its calibrated singular branch. A finite critical value rejecting all singular normalizers exists when $p_{\infty,m} \le \alpha$. Online transfer introduces the decoded-update radius and the privacy dependence shown in table 8.

The analytic references have different approximation errors. For oracle Wald, the balanced-benchmark χ^2 error is $O(k^{7/4}m^{-1/2})$, so $k = o(m^{2/7})$ suffices. For integral SN, finite-grid Gaussian or Brownian calibration uses the fixed-dimensional process approximation in (124). The analytic F replacement for DCT series and block-Haar series is $O_{k,K}(m^{-1/4})$, and the Gaussian replacement for block Hotelling is $O_{k,B_0}(m_B^{-1/4})$. The oracle Wald, series, and Hotelling approximation terms have no explicit standardized ε^{-1} factor. Ideal score calibration is exact with the stated singular branch; for plug-in Wald, fixing the pilot at the truth gives the oracle Wald score endpoint, while plug-in calibration also includes the independent pilot and density-calibration law.

Table 9 collects the corresponding leading online conditions. For exact-score integral SN, the displayed condition uses $u \asymp k^{-1}$ and also requires $s^\circ_{m,\mathrm{SN}}(u) = o(1)$. Exact-score DCT series retains the actual DCT spectral tail and local concentration. For block-Haar series with $K \asymp k$, $m \gtrsim k^4$ is a conservative sufficient certificate for a constant spectral threshold. Analytic series and Hotelling conditions retain their fixed-dimensional Gaussian-replacement errors; block Hotelling also retains blockwise localization failures. The plug-in Wald conditions are subject to the density and covariance conditioning factors.

Table 9: Leading online small-error conditions under $L_m \asymp k$, $K \asymp B_0 \asymp k$, balanced kRR, and $0 < \varepsilon \le 1$. Reference-approximation, spectral, and localization requirements are stated above.

Method	Leading online conditions
Oracle Wald	$k = o\{(\varepsilon m^{1/6})^{2/7}\}$; analytic error: $k = o(m^{2/7})$ suffices.
Plug-in Wald	Oracle Wald plus $k^{5/2}\log k = o(m_c\varepsilon^2)$ and $\sqrt{k}\, r_p = o(1)$.
Integral SN	Exact score: $k = o\{(\varepsilon m^{1/6})^{1/7}\}$. Analytic diagnostic: $k = o\{(\varepsilon m^{1/6})^{1/8}\}$ plus process error.
DCT series	Analytic reference: $k = o\{(\varepsilon m^{1/6})^{2/9}\}$.
Block-Haar series	Exact-score diagnostic: $k = o\{(\varepsilon m^{1/6})^{1/5}\}$. Analytic reference: $k = o\{(\varepsilon m^{1/6})^{2/9}\}$.
Block Hotelling	$k = o\{(\varepsilon n^{1/6})^{3/14}\}$.

These conditions separate frozen calibration from online coverage. Privacy can cancel after standardization in the frozen pivot while enlarging the nonlinear online remainder. For plug-in Wald, the plug-in frozen probability conditions on the pilot or integrates over its law; the score-only calculation equals the oracle Wald probability.

The comparison identifies conditions under which the score-calibrated Haar series pivot is attractive. It removes the Gaussian-reference calibration error, produces independent frozen covariance replications, has a constant certified spectral threshold under explicit conditions, and runs one online recursion. When D can be estimated accurately, plug-in Wald may be shorter; when the sample is too small for $K \gtrsim k$ stable replications, ellipsoidal random-normalizer methods can be unstable.

G Mechanism-design lemmas and proofs

This appendix contains the linear-algebra and convex-design arguments used in section 2.1.

Proposition G.1 (Direct randomized-response constants)**.** *For direct randomized response in* (40)*:*

(i) The unique unbiased decoder is

$$A = P\mathsf{Q}^{-1} = \delta_\varepsilon^{-1}\{P - q(P\mathbf{1}_d)\mathbf{1}_d^\top\}, \tag{133}$$

and

$$A_0 \le \frac{2\sqrt{k}}{\delta_\varepsilon}, \qquad G \le \frac{3\sqrt{k}}{\delta_\varepsilon}. \tag{134}$$

(ii) The standardized leverage and third moment satisfy

$$\ell^2 = \max_c(\mu_c^{-1} - 1) \le \mu_\varepsilon^{-1}, \qquad \beta_3 = \sum_{c=0}^{k} \mu_c(\mu_c^{-1} - 1)^{3/2} \le \frac{k+1}{\sqrt{\mu_\varepsilon}}. \tag{135}$$

(iii) The covariance satisfies

$$\lambda_{\min}(\Sigma) \ge \frac{\mu_\varepsilon}{4\delta_\varepsilon^2}, \qquad \sigma_-^{-1} \le \frac{2\delta_\varepsilon}{\sqrt{\mu_\varepsilon}}. \tag{136}$$

Lemma G.2 (Automatic full column rank)**.** *If (Q, A) is an unbiased channel–decoder pair, then Q has column rank $k+1$.*

Proof. The first k columns of P form a nonsingular lower-triangular matrix and the last column is zero, so $\ker(P) = \operatorname{span}(e_d)$, where e_d is the last coordinate vector in $\mathbb{R}^d$. If $\mathsf{Q}v = 0$, then $Pv = A\mathsf{Q}v = 0$, hence $v = ce_d$. Since every channel column sums to one, $0 = \mathbf{1}_{|\mathcal{Y}|}^\top \mathsf{Q}v = \mathbf{1}_d^\top v = c$. Thus $v = 0$. □

For every feasible decoder, the variance comparison in theorem 2.1 follows from

$$ARA^\top = A^* R A^{*\top} + (A - A^*)R(A - A^*)^\top, \tag{137}$$

Proof of theorem 2.1. First, $A^*\mathsf{Q} = PM^{-1}\mathsf{Q}^\top R^{-1}\mathsf{Q} = P$. For any other feasible A, put $H = A - A^*$. Then $H\mathsf{Q} = 0$ and

$$A^* R H^\top = PM^{-1}\mathsf{Q}^\top H^\top = PM^{-1}(H\mathsf{Q})^\top = 0.$$

Expanding $(A^* + H)R(A^* + H)^\top$ proves (137). Direct multiplication gives $A^* R A^{*\top} = PM^{-1}P^\top$. Finally, $\mathbb{E}a_Y = A^*\mathsf{Q}\boldsymbol{\pi}^* = P\boldsymbol{\pi}^* = \boldsymbol{\tau}$, so subtracting $\boldsymbol{\tau}\boldsymbol{\tau}^\top$ yields the covariance formula. □

Lemma G.3 (Information increases under output refinement)**.** *If a nonzero channel row vector q is split into nonnegative vectors $q_1, \dots, q_m$ with $q = \sum_{r=1}^m q_r$, then*

$$\sum_{r=1}^{m} \frac{q_r q_r^\top}{(\boldsymbol{\pi}^*)^\top q_r} \succeq \frac{qq^\top}{(\boldsymbol{\pi}^*)^\top q}.$$

Proof. For arbitrary v, weighted Cauchy–Schwarz gives

$$\frac{(v^\top q)^2}{(\boldsymbol{\pi}^*)^\top q} = \frac{\{\sum_r v^\top q_r\}^2}{\sum_r (\boldsymbol{\pi}^*)^\top q_r} \le \sum_r \frac{(v^\top q_r)^2}{(\boldsymbol{\pi}^*)^\top q_r}.$$

The quadratic-form inequality for every v proves the result. □

Proof of theorem C.1. Take a nonzero output row q of an arbitrary ε-LDP channel and put $m_q = \min_c q_c$. If $m_q = 0$, local privacy forces $q = 0$, contrary to the choice of row. Hence $q/m_q \in [1, e^\varepsilon]^d$. Every point in this box is a convex combination of its 2^d vertices, so $q = \sum_{s \in \mathcal{S}_\varepsilon} \theta_{q,s} s$ with $\theta_{q,s} \ge 0$. Replace the output by the associated suboutputs. By lemma G.3, the information matrix cannot decrease. Matrix inversion reverses Loewner order, and theorem 2.1 shows that the optimal decoded-score covariance cannot increase. Merge proportional staircase rows to obtain (37).

For a staircase row s, the marginal output probability is $\theta_s(\boldsymbol{\pi}^*)^\top s$. Therefore

$$M(\mathsf{Q}_\theta) = \sum_s \frac{\theta_s^2 s s^\top}{\theta_s(\boldsymbol{\pi}^*)^\top s} = H(\theta),$$

and theorem 2.1 gives (39). The map $H \mapsto \operatorname{tr}(CH^{-1})$ is convex for $C \succeq 0$; here $C = P^\top W P$. Since $H(\theta)$ is affine, the objective is convex. Feasibility implies $\sum_s \theta_s \le 1$, so the feasible set is bounded and closed.

It remains to justify the singular boundary. For any feasible design, $\ker H(\theta) = \ker \mathsf{Q}_\theta$. If v lies in this kernel, channel normalization gives $\mathbf{1}_d^\top v = 0$. Since $\ker P = \operatorname{span}(e_d)$ and $\mathbf{1}_d^\top e_d = 1$, every nonzero such v satisfies $Pv \neq 0$. Thus $v^\top P^\top W P v > 0$. Along any positive-definite sequence converging to a singular feasible H, the inverse objective diverges in a null direction. Extension by $+\infty$ is lower semicontinuous, and compactness yields existence. □

Proof of proposition G.1. The channel matrix is $\mathsf{Q} = \delta_\varepsilon I_d + q\mathbf{1}_d\mathbf{1}_d^\top$. Since $\delta_\varepsilon + qd = 1$, the Sherman–Morrison formula gives $\mathsf{Q}^{-1} = \delta_\varepsilon^{-1}(I_d - q\mathbf{1}_d\mathbf{1}_d^\top)$, proving (133). Each column of P has norm at most $\sqrt{k}$, and $\|qP\mathbf{1}_d\|_2 \le q(\sum_{j=1}^k j^2)^{1/2} \le \sqrt{k}$. Thus $A_0 \le 2\sqrt{k}/\delta_\varepsilon$. Since $\|\boldsymbol{\tau}\| \le \sqrt{k}$ and $\delta_\varepsilon \le 1$, the bound for G follows.

The centered categorical vector $e_Y - \mu$ has covariance $\operatorname{diag}(\mu) - \mu\mu^\top$. On the simplex tangent space, its inverse quadratic form is $v^\top \operatorname{diag}(\mu)^{-1} v$. The decoded CDF score is a nonsingular linear image of this tangent vector. Mahalanobis distances are invariant under nonsingular linear maps, so the squared distance at output c equals $\mu_c^{-1} - 1$. Averaging the $3/2$ power proves the identity in (135); the bounds follow from $\mu_c \ge \mu_\varepsilon$.

For the covariance lower bound, centered categorical vectors satisfy $A(e_Y - \mu) = \delta_\varepsilon^{-1} P(e_Y - \mu)$ because $\mathbf{1}_d^\top(e_Y - \mu) = 0$. Hence

$$\Sigma = \delta_\varepsilon^{-2} P\{\operatorname{diag}(\mu) - \mu\mu^\top\}P^\top.$$

For $x \in \mathbb{R}^k$, put $v = P^\top x$. Then

$$x^\top \Sigma x = \delta_\varepsilon^{-2} \min_{a\in\mathbb{R}} \sum_{c=0}^{k} \mu_c (v_c - a)^2 \ge \delta_\varepsilon^{-2}\mu_\varepsilon \min_a \|v - a\mathbf{1}_d\|_2^2 .$$

The discrete differences of v equal the coordinates of x. The first-difference operator has norm at most two on $\mathbf{1}_d^\perp$, so $\|x\|_2 \le 2\min_a \|v - a\mathbf{1}_d\|_2$. This gives (136). The balanced orders follow by substitution. □

H PROOFS FOR ONLINE SELECTION

Throughout this appendix, constants may change from line to line. Conditional statements are understood with respect to the pilot sigma-field $\mathcal{G}_n$. For a candidate r, let $\mathbb{P}_r$ and $\mathbb{E}_r$ denote the conditional law and expectation of the hypothetical Stage-3 recursion that uses r and the realized preconditioner. Uniformity means that the maximum of the relevant conditional errors over $r \in \mathfrak{M}_n$ converges to zero. The common deterministic bounds in (47), the bounded cardinality $R_{\max}$, and the bounds on $\mathcal{E}_n^B$ establish this uniformity for the triangular library.

Proof of proposition 3.3. Let $e = \overline{\boldsymbol{q}}_b - \boldsymbol{Q}$, let $h(q)$ have coordinates $\tau_j - F(q_j)$, and put $r(e) = h(\boldsymbol{Q} + e) + De$. For the fixed-threshold reports define $V_s = \boldsymbol{\tau} - a_{\widehat{r},Y_s} - h(\overline{\boldsymbol{q}}_b)$. Conditional on the transcript through localization, the V_s are independent, bounded, and mean zero. Direct substitution gives (19). The uniform representation in theorem C.10, applied at b, gives $e = O_{\mathbb{P}}(b^{-1/2})$. By theorem C.8, $\widehat{D} - D = o_{\mathbb{P}}(1)$, so the second term is $o_{\mathbb{P}}(1)$. Local Hölder regularity gives $\|r(e)\| \le C\|e\|^{1+\beta}$ with probability tending to one; its scaled contribution is $O_{\mathbb{P}}(b^{-\beta/2})$. The conditional covariance of V_s converges to $\Sigma_{\widehat{r}}$ because $\overline{\boldsymbol{q}}_b \to \boldsymbol{Q}$ and the score alphabet is finite. The conditional triangular-array central limit theorem and the Gaussian continuity argument used in proposition H.7 now give both conclusions, uniformly over the finite library. □

H.1 PRIVACY OF THE FIXED-MECHANISM VARIANTS

At time t, condition on the complete past and the public server state $\boldsymbol{q}_{t-1}$. For any two possible raw values x, x' and every report y,

$$\mathbb{P}(Y_t = y \mid X_t = x, \mathcal{F}_{t-1}) = \mathsf{Q}_{y,C(\boldsymbol{q}_{t-1};x)} \le e^\varepsilon \mathsf{Q}_{y,C(\boldsymbol{q}_{t-1};x')} = e^\varepsilon \mathbb{P}(Y_t = y \mid X_t = x', \mathcal{F}_{t-1}).$$

Thus each user receives interactive user-level ε-LDP. Decoding all k CDF indicators, projecting, averaging, constructing Haar coefficients, optimizing a mechanism from public $(\boldsymbol{\tau}, \varepsilon)$, and simulating frozen critical values are post-processing. They consume no additional privacy budget. In the plug-in procedure, main, pilot, and calibration users are disjoint; every user therefore uses the full budget ε without composition across the three samples.

H.2 Privacy and density-probe proofs

Proof of proposition C.7. Fix a user t and condition on the complete transcript before that user is queried. This transcript determines the user's stage, the threshold vector sent to the user, and a single channel Q_t. Let x, x' be two possible raw values and write $c = C(q; x)$ and $c' = C(q; x')$ for the corresponding interval categories. Since Q_t is ε-LDP,

$$\mathbb{P}(Y_t = y \mid X_t = x, \mathcal{H}_{t-1}) = (\mathsf{Q}_t)_{y,c} \le e^{\varepsilon}(\mathsf{Q}_t)_{y,c'} = e^{\varepsilon}\mathbb{P}(Y_t = y \mid X_t = x', \mathcal{H}_{t-1})$$

for every output y and every possible preceding transcript $\mathcal{H}_{t-1}$. Hence the interactive protocol is user-level ε-LDP for user t. The same argument applies separately to pilot, positive-probe, negative-probe, and main-stage users. The groups are disjoint, so no user contributes reports in two stages and no within-user composition is incurred.

All remaining operations—decoding, averaging, projection, density truncation, construction and comparison of public channels, frozen-score simulation, multiplier generation, and formation of confidence sets—are measurable functions of public quantities and already privatized reports. The post-processing property completes the proof. □

Proof of theorem C.8. Apply lemmas H.2 and H.3 to the Stage-1 recursion with $B = I_k$ and $r = r_0$. Their proofs use only the drift, score, projection, and step conditions and hence do not depend on the density probe. Thus $\boldsymbol{q}_t^p \to \boldsymbol{Q}$ conditionally almost surely, and Cesàro averaging gives $\widetilde{\boldsymbol{Q}} \to \boldsymbol{Q}$ in probability. Because the target has the positive box margins in (45),

$$\mathbb{P}(\widetilde{h}_n = h_n) \longrightarrow 1.$$

It therefore suffices to work on this event. Condition on $\widetilde{\boldsymbol{Q}}$. Unbiasedness of the baseline CDF decoder yields, coordinatewise,

$$\mathbb{E}(\overline{a}_{+,j} \mid \widetilde{\boldsymbol{Q}}) = F(\widetilde{Q}_j + h_n), \qquad \mathbb{E}(\overline{a}_{-,j} \mid \widetilde{\boldsymbol{Q}}) = F(\widetilde{Q}_j - h_n).$$

Consequently,

$$\begin{aligned}\mathbb{E}(\widehat{f}_j^{\mathrm{raw}} \mid \widetilde{\boldsymbol{Q}}) &= \frac{F(\widetilde{Q}_j + h_n) - F(\widetilde{Q}_j - h_n)}{2h_n} \\ &= \frac{1}{2h_n}\int_{\widetilde{Q}_j - h_n}^{\widetilde{Q}_j + h_n} f(u)\, du.\end{aligned}$$

For all sufficiently large n, on an event whose probability tends to one, the interval of integration is contained in $\mathcal{N}_j$. By (46),

$$\begin{aligned}\left|\mathbb{E}(\widehat{f}_j^{\mathrm{raw}} \mid \widetilde{\boldsymbol{Q}}) - f(Q_j)\right| &\le \frac{L_j}{2h_n}\int_{\widetilde{Q}_j - h_n}^{\widetilde{Q}_j + h_n} |u - Q_j|^{\beta}\, du \\ &\le L_j\big(|\widetilde{Q}_j - Q_j| + h_n\big)^{\beta}.\end{aligned}$$

This is the combined pilot-location and smoothing bias.

Conditional on the pilot, the positive reports are i.i.d., the negative reports are i.i.d., and the two groups are independent. Under assumption C.3, decoded outputs have uniformly bounded norm. Thus, for fixed k,

$$\max_{1 \le j \le k}\left|\overline{a}_{+,j} - \mathbb{E}(\overline{a}_{+,j} \mid \widetilde{\boldsymbol{Q}})\right| = O_{\mathbb{P}}(n_{2,+}^{-1/2}),$$

and the same bound holds for the negative group. Since both group proportions have positive limits,

$$\max_j\left|\widehat{f}_j^{\mathrm{raw}} - \mathbb{E}(\widehat{f}_j^{\mathrm{raw}} \mid \widetilde{\boldsymbol{Q}})\right| = O_{\mathbb{P}}\{(h_n\sqrt{n_2})^{-1}\}.$$

By pilot consistency and (45), the boundary cap and numerical floor in (23) are inactive with probability tending to one, so $\widetilde{h}_n = h_n$ on that event. Combining the last two displays and using that the operator norm of a diagonal matrix equals the largest absolute diagonal entry proves (51). The right-hand side tends to zero by (48) and pilot consistency.

Every $f(Q_j)$ is finite and strictly positive. Since $\underline{f}_n \downarrow 0$ and $\overline{f}_n \uparrow \infty$, for all sufficiently large n each $f(Q_j)$ lies strictly between the truncation boundaries. Uniform consistency over the fixed set of coordinates then gives

$$\mathbb{P}\{\underline{f}_n < \widehat{f}_j^{\mathrm{raw}} < \overline{f}_n \text{ for every } j\} \longrightarrow 1,$$

which is exactly (52). □

H.3 Selection proofs

Proof of theorem C.9. For candidate r, the local Wald ellipsoid in displacement coordinates is

$$\left\{u \in \mathbb{R}^k : N u^\top D \Sigma_r^{-1} D u \le c_{r,N,\alpha}^W\right\}.$$

Apply the linear change of variables

$$v = \sqrt{N/c_{r,N,\alpha}^W}\,\Sigma_r^{-1/2} D u.$$

Its inverse Jacobian determinant is

$$N^{-k/2}\{c_{r,N,\alpha}^W\}^{k/2}\det(D^{-1})\det(\Sigma_r)^{1/2}.$$

Multiplication by the unit-ball volume v_k proves (8). The factors $v_k N^{-k/2}\det(D^{-1})$ do not depend on r, so minimizing the volume is equivalent to minimizing $J_{r,N}^W$.

For the Haar construction, the local quantile coefficient corresponding to the score coefficient $Z_{r,\ell}^\circ$ is $D^{-1}Z_{r,\ell}^\circ$. Hence the quantile-space normalizer is

$$D^{-1} S_{r,K,N}^{\xi,\circ} D^{-1}.$$

On a nonsingular realization, the same change-of-variables calculation gives volume

$$v_k N^{-k/2}\det(D^{-1})\{c_{r,N,K,\alpha}^H\}^{k/2}\det(S_{r,K,N}^{\xi,\circ})^{1/2}.$$

A singular normalizer is assigned infinite volume. Multiplication by the common positive constant $v_k N^{-k/2}\det(D^{-1})$ commutes with quantiles, proving part (ii). Part (iii) follows immediately because randomized response is one of the candidates over which the minimum is taken. □

Proof of theorem 3.1. By definition of $\widehat{r}^{\max}$,

$$J^{\max}_{\widehat{r}^{\max},n_3}(\widehat{D}) \le J^{\max}_{r^*,n_3}(\widehat{D}).$$

Adding and subtracting the corresponding true-D criteria gives

$$\begin{aligned} J^{\max}_{\widehat{r}^{\max},n_3}(D) &\le J^{\max}_{\widehat{r}^{\max},n_3}(\widehat{D}) + \Delta_n \\ &\le J^{\max}_{r^*,n_3}(\widehat{D}) + \Delta_n \\ &\le J^{\max}_{r^*,n_3}(D) + 2\Delta_n, \end{aligned}$$

which proves (15).

Write $d_j = f(Q_j)$ and $\widehat{d}_j = \widehat{f}_j$. On an event whose probability tends to one, all d_j and $\widehat{d}_j$ lie in a fixed compact interval $[d_-/2, 2d_+]$ with $d_- > 0$. Therefore

$$\left|\frac{1}{\widehat{d}_j} - \frac{1}{d_j}\right| = \frac{|\widehat{d}_j - d_j|}{\widehat{d}_j d_j} \le C|\widehat{d}_j - d_j|.$$

The diagonal score standard deviations are uniformly bounded by (47). Moreover, a union bound and Chebyshev's inequality give a common upper bound for the $(1-\alpha)$ frozen max critical values. Using $|\max_j x_j - \max_j y_j| \le \max_j |x_j - y_j|$ therefore yields

$$\Delta_n \le C\left\|\widehat{D} - D\right\|_{\mathrm{op}} = o_{\mathbb{P}}(1).$$

If the oracle gap is at least $\delta_0 > 0$, the event $2\Delta_n < \delta_0$ forces the empirical minimizer to equal the unique oracle minimizer; hence selection is consistent. Taking randomized response as the competitor in (15) proves (17). □

H.4 POST-SELECTION STOCHASTIC-APPROXIMATION PROOFS

Condition on $\mathcal{G}_n$ and let B be any $\mathcal{G}_n$-measurable matrix satisfying $b_- I_k \preceq B \preceq b_+ I_k$. For a retained candidate r, write $q_s = q_{r,s}$, $e_s = q_s - \boldsymbol{Q}$, $\gamma_s = \gamma_s^m$, $N = n_3$, and

$$U_s = \boldsymbol{\tau} - a_{r,Y_s} = h(q_{s-1}) + M_s.$$

The selected recursion is obtained by taking $B = \widehat{B}$ and $r = \widehat{r}$ on $\mathcal{E}_n^B$. All constants below depend only on the deterministic bounds in the assumptions and are therefore uniform in (B, r). The complement of $\mathcal{E}_n^B$ has probability tending to zero.

Lemma H.1 (Target-centered global drift)**.** *Under assumption C.2, there is a constant $\lambda_F > 0$ such that, for every $q \in \mathcal{C}$,*

$$-(q - \boldsymbol{Q})^\top h(q) = \sum_{j=1}^{k} (q_j - Q_j)\{F(q_j) - F(Q_j)\} \geq \lambda_F \|q - \boldsymbol{Q}\|_2^2 . \tag{138}$$

Proof. For coordinate j, define, with the continuous extension at Q_j,

$$g_j(x) = \begin{cases} \{F(x) - F(Q_j)\}/(x - Q_j), & x \neq Q_j, \\ f(Q_j), & x = Q_j. \end{cases}$$

The local Hölder condition implies differentiability of F at Q_j, so g_j is continuous at Q_j and is positive on a neighborhood of Q_j. On the compact complement of that neighborhood, continuity and strict monotonicity of F give a positive minimum. Taking $\lambda_F = \min_j \inf_{x \in [L,U]} g_j(x) > 0$ and summing $(q_j - Q_j)^2 g_j(q_j)$ proves (138). □

Lemma H.2 (Uniform conditional moment order)**.** *Under assumptions C.2 to C.4, for every fixed $p \geq 2$ there is a deterministic $C_p < \infty$ such that*

$$\max_{r \in \mathfrak{M}_n} \mathbb{E}_r(\|q_{r,s} - \boldsymbol{Q}\|_2^p \mid \mathcal{G}_n) \leq C_p \gamma_s^{p/2}, \qquad s \geq 1, \tag{139}$$

whenever B lies between fixed positive spectral bounds. In particular, the bound holds with the constants (b_-, b_+) on $\mathcal{E}_n^B$ for Stage 3 and with fixed bounds containing I_k for the pilot, and $e_s = O_{\mathbb{P}}(\gamma_s^{1/2})$ conditionally and uniformly over the realized library.

Proof. Step 1: the quadratic recursion. Let $V_s = \|e_s\|_{B^{-1}}^2$. Metric projection is nonexpansive in the B^{-1} norm and fixes $\boldsymbol{Q}$. Because $\|U_s\|_2 \leq G_{\max}$ and $\mathbb{E}_{s-1} U_s = h(q_{s-1})$, lemma H.1 gives

$$\begin{aligned} \mathbb{E}_{s-1} V_s &\leq V_{s-1} + 2\gamma_s e_{s-1}^\top h(q_{s-1}) + \gamma_s^2 \mathbb{E}_{s-1}(U_s^\top B U_s) \\ &\leq (1 - c_2 \gamma_s) V_{s-1} + C_2 \gamma_s^2, \end{aligned}$$

where one may take $c_2 = 2\lambda_F b_-$ after increasing the deterministic starting index so that $c_2 \gamma_s \leq 1$.

Step 2: higher moments. Put $w = B^{-1/2} e_{s-1}$ and $z = \gamma_s B^{1/2} U_s$. For $p \geq 2$, Taylor's inequality for $x \mapsto \|x\|_2^p$ gives

$$\|w + z\|_2^p \leq \|w\|_2^p + p \|w\|_2^{p-2} w^\top z + C_p\{\|w\|_2^{p-2} \|z\|_2^2 + \|z\|_2^p\}.$$

Conditional expectation, (138), the bounds on B, and $\|z\| \leq C\gamma_s$ yield

$$\mathbb{E}_{s-1} V_s^{p/2} \leq (1 - c_p \gamma_s) V_{s-1}^{p/2} + C_p \gamma_s^2 V_{s-1}^{(p-2)/2} + C_p \gamma_s^p.$$

For any sufficiently small fixed $\eta > 0$, Young's inequality in the form

$$\gamma_s^2 x^{p-2} \leq \eta \gamma_s x^p + C_{p,\eta} \gamma_s^{1+p/2}$$

absorbs the middle term. Since $\gamma_s^p \leq C \gamma_s^{1+p/2}$ for $p \geq 2$, for all sufficiently large s,

$$\mathbb{E}_{s-1} V_s^{p/2} \leq (1 - c_p' \gamma_s) V_{s-1}^{p/2} + C_p' \gamma_s^{1+p/2}. \tag{140}$$

Step 3: comparison with the step size. Regular variation gives

$$\frac{\gamma_{s-1}^{p/2}}{\gamma_s^{p/2}} = 1 + O(s^{-1}) = 1 + o(\gamma_s),$$

because $a < 1$. Choose $K_p \geq 2C_p'/c_p'$. If $\mathbb{E}V_{s-1}^{p/2} \leq K_p\gamma_{s-1}^{p/2}$, then, for all sufficiently large s, (140) is at most $K_p\gamma_s^{p/2}$. The finitely many earlier indices are absorbed by increasing K_p, since projection keeps q_s in the compact set $\mathcal{C}$. Norm equivalence on $b_-I_k \preceq B \preceq b_+I_k$ proves (139), with the same constants for every retained candidate. □

Lemma H.3 (Conditional consistency and projection inactivity)**.** *Under assumptions C.2 to C.4, conditionally on $\mathcal{G}_n$, $q_s \to \boldsymbol{Q}$ almost surely and the weighted projection in* (28) *is inactive after a finite random time. More quantitatively, choose $\delta_0 > 0$ so that $\|q - \boldsymbol{Q}\|_\infty \leq \delta_0$ lies inside the density neighborhoods and the strict interior of $\mathcal{C}$. For every fixed $p > 2/a$ and all sufficiently large b,*

$$\max_{r\in\mathfrak{M}_n} \mathbb{P}_r\left\{\sup_{s\geq b-1} \|q_{r,s} - \boldsymbol{Q}\|_\infty > \delta_0 \,\middle|\, \mathcal{G}_n\right\} \leq C_p \sum_{s=b-1}^{\infty} \gamma_s^{p/2} \leq C_p b^{1-ap/2}. \tag{141}$$

On the complementary event every update after time b is unprojected.

Proof. Fix $p > 2/a$. By Markov's inequality and (139), for every $\eta > 0$,

$$\sum_{s=1}^{\infty} \max_r \mathbb{P}_r(\|e_s\|_2 > \eta \mid \mathcal{G}_n) \leq C_{p,\eta} \sum_{s=1}^{\infty} \gamma_s^{p/2} < \infty.$$

Conditional Borel–Cantelli, first on rational $\eta > 0$ and then by intersection, gives $e_s \to 0$ conditionally almost surely. The same Markov bound and a union bound over $s \geq b-1$ prove (141).

Take δ_0 no larger than $\Delta_Q/8$ and small enough that every coordinate remains in its density neighborhood. For large b, $\gamma_s b_+ G_{\max} \leq \Delta_Q/8$ for all $s \geq b$. If $\|e_{s-1}\|_\infty \leq \delta_0$, the current adjacent gaps are at least $3\Delta_Q/4$ and both box margins are positive. Each coordinate of the proposed update changes by at most $\Delta_Q/8$, so every adjacent gap and box margin remains strictly positive. The proposed point is therefore in the interior of $\mathcal{C}$ and the projection is inactive. This also shows explicitly that re-exit after time b is included in the event on the left of (141). □

Lemma H.4 (Averaged linear representation)**.** *Under assumptions C.2 to C.4,*

$$\frac{D}{\sqrt{N}} \sum_{s=1}^{N} e_{s-1} = \frac{1}{\sqrt{N}} \sum_{s=1}^{N} M_s + o_{\mathbb{P}}(1) \tag{142}$$

conditionally on $\mathcal{G}_n$.

Proof. Fix $0 < \rho < 1/2$ and put $b_N = \lfloor N^\rho \rfloor$. Choose $p > 2/a$ in lemma H.3, and let $\mathcal{A}_N$ be the event that $\|e_s\|_\infty \leq \delta_0$ for every $s \geq b_N - 1$. Uniformly over the library, (141) gives $\mathbb{P}_r(\mathcal{A}_N^c \mid \mathcal{G}_n) \to 0$. On $\mathcal{A}_N$, every update with $s > b_N$ is unprojected and

$$e_s = e_{s-1} + \gamma_s B\{h(\boldsymbol{Q} + e_{s-1}) + M_s\}.$$

The local Hölder density condition implies the componentwise Taylor expansion

$$h(\boldsymbol{Q} + e) = -De + r(e), \qquad \|r(e)\|_2 \leq C\,\|e\|_2^{1+\beta}.$$

Rearranging gives

$$De_{s-1} = M_s + r(e_{s-1}) - B^{-1}\gamma_s^{-1}(e_s - e_{s-1}). \tag{143}$$

We control the initial-segment, nonlinear, and Abel terms separately.

Step 1: the initial projected segment. Both e_s and M_s are uniformly bounded. Hence

$$\frac{1}{\sqrt{N}} \left\| D \sum_{s=1}^{b_N} e_{s-1} - \sum_{s=1}^{b_N} M_s \right\|_2 \leq C b_N N^{-1/2} = o(1).$$

Step 2: the nonlinear remainder. Since $1+\beta \le 2$, Lyapunov's inequality and (139) give

$$\max_r \mathbb{E}_r\{\|e_s\|_2^{1+\beta} \mid \mathcal{G}_n\} \le C\gamma_s^{(1+\beta)/2}.$$

Therefore Markov's inequality yields

$$\begin{aligned}\frac{1}{\sqrt{N}} \sum_{s=b_N+1}^{N} \|r(e_{s-1})\|_2 &= O_{\mathbb{P}}\left(N^{-1/2}\sum_{s=1}^{N} s^{-a(1+\beta)/2}\right)\\ &= o_{\mathbb{P}}(1),\end{aligned}$$

uniformly in r, because $a(1+\beta) > 1$.

Step 3: the Abel term. For $l = b_N + 1$, summation by parts gives

$$\sum_{s=l}^{N} \gamma_s^{-1}(e_s - e_{s-1}) = \gamma_N^{-1} e_N - \gamma_l^{-1} e_{l-1} - \sum_{s=l}^{N-1} (\gamma_{s+1}^{-1} - \gamma_s^{-1}) e_s.$$

By (139),

$$\max_r \mathbb{E}_r(\|e_s\|_2 \mid \mathcal{G}_n) \le C\gamma_s^{1/2}.$$

Since $\gamma_s^{-1} \asymp s^a$ and $\gamma_{s+1}^{-1} - \gamma_s^{-1} \le Cs^{a-1}$,

$$\begin{aligned}\max_r \mathbb{E}_r\left[\left\|\sum_{s=l}^{N} \gamma_s^{-1}(e_s - e_{s-1})\right\|_2 \,\middle|\, \mathcal{G}_n\right] &\le C\left\{N^{a/2} + l^{a/2} + \sum_{s=l}^{N-1} s^{a/2-1}\right\}\\ &\le CN^{a/2}.\end{aligned}$$

Division by $\sqrt{N}$ and Markov's inequality make this $o_{\mathbb{P}}(1)$ because $a < 1$. Summing (143) on $\mathcal{A}_N$, adding the initial segment, and then adding $\mathbb{P}_r(\mathcal{A}_N^c \mid \mathcal{G}_n) = o(1)$ proves (142) uniformly over the realized library. □

Proof of theorem C.10. Part (i) is lemma H.3. Since $\sqrt{N}(\overline{\boldsymbol{q}}_N - \boldsymbol{Q}) = N^{-1/2}\sum_{s=1}^{N} e_{s-1}$, part (ii) follows from lemma H.4.

For part (iii), let

$$V_r(q) = \mathrm{Var}\{\boldsymbol{\tau} - a_{r,Y} \mid q\},$$

where the report is generated using thresholds q. The interval-probability vector is $\boldsymbol{\pi}(q)$ from section 2.1. A Markov channel is an ℓ_1 contraction and

$$\|\boldsymbol{\pi}(q) - \boldsymbol{\pi}^*\|_1 \le 2\sum_{j=1}^{k} |F(q_j) - F(Q_j)|.$$

Because every score has norm at most $G_{\max}$,

$$\max_{r\in\mathfrak{M}_n} \|V_r(q) - \Sigma_r\|_{\mathrm{op}} \le C\left\{\sum_{j=1}^{k} |F(q_j) - F(Q_j)| + \|h(q)\|_2^2\right\}. \tag{144}$$

On $\|q - \boldsymbol{Q}\|_\infty \le \delta_0$, local boundedness of the density makes the right-hand side at most $C\|q - \boldsymbol{Q}\|_2$. Outside that neighborhood it is bounded, while (139) and Markov's inequality give probability $O(\gamma_s)$. Consequently,

$$\max_r \mathbb{E}_r\{\|V_r(q_{r,s-1}) - \Sigma_r\|_{\mathrm{op}} \mid \mathcal{G}_n\} \le C\gamma_s^{1/2}.$$

Averaging and applying Markov's inequality proves (56) uniformly.

Moreover, $\|M_s\|_2 \le 2G_{\max}$. Whiten the array conditionally on $\mathcal{G}_n$ by setting

$$\widetilde{M}_{r,s} = \Sigma_r^{-1/2} M_s.$$

The lower eigenvalue bound makes $\left\|\widetilde{M}_{r,s}\right\|_2 \le 2G_{\max}/\underline{\sigma}$, so for every $\eta > 0$ its conditional Lindeberg sum is identically zero once $2G_{\max}/\underline{\sigma} < \eta\sqrt{N}$. Equation (56) and the two uniform eigenvalue bounds give

$$\max_r \left\| \frac{1}{N}\sum_{s=1}^{N} \mathbb{E}_{s-1}(\widetilde{M}_{r,s}\widetilde{M}_{r,s}^\top) - I_k \right\|_{\mathrm{op}} = o_{\mathbb{P}}(1).$$

The conditional martingale triangular-array central limit theorem now has the same fixed target $N_k(0, I_k)$ for every candidate. Uniformity follows by contradiction: any sequence of candidates violating the uniform conclusion still satisfies the displayed variance convergence and the common Lindeberg bound, and hence satisfies that fixed-target theorem. Multiplication by $\Sigma_r^{1/2}$ changes bounded-Lipschitz constants by at most the common factor $\overline{\sigma}$, so the unwhitened martingale sum is uniformly approximated by $N_k(0, \Sigma_r)$. Combining this result with (55) proves (57).

For part (iv), write $U_s = h(q_{s-1}) + M_s$. Consistency and boundedness imply

$$N^{-1}\sum_s \|h(q_{s-1})\|^2 \to 0, \qquad \overline{U} \to 0$$

in probability. The matrix differences $M_s M_s^\top - V_r(q_{s-1})$ are bounded martingale differences, and orthogonality of their increments gives

$$\mathbb{E}_r \left\| \frac{1}{N}\sum_{s=1}^{N} \{M_s M_s^\top - V_r(q_{s-1})\} \right\|_F^2 \le \frac{C}{N}.$$

The cross terms obey

$$\left\| \frac{1}{N}\sum_s M_s h(q_{s-1})^\top \right\|_{\mathrm{op}} \le \left(\frac{1}{N}\sum_s \|M_s\|_2^2\right)^{1/2} \left(\frac{1}{N}\sum_s \|h(q_{s-1})\|_2^2\right)^{1/2} = o_{\mathbb{P}}(1),$$

and the transpose has the same bound. Together with (56), this gives $N^{-1}\sum_s U_s U_s^\top = \Sigma_r + o_{\mathbb{P}}(1)$ uniformly. Finally, $\overline{U} = N^{-1}\sum_s M_s + N^{-1}\sum_s h(q_{s-1}) = o_{\mathbb{P}}(1)$ by the martingale variance bound and Cauchy–Schwarz. Subtracting $\overline{U}\,\overline{U}^\top$ proves (58). □

Corollary H.5 (Functional representation and the fixed-mechanism recursion)**.** *Under the conditions of theorem C.10, the partial-average process satisfies, with value zero at $r = 0$,*

$$\sup_{0\le r\le 1} \left\| \frac{\lfloor n_3 r\rfloor}{\sqrt{n_3}}(\overline{\boldsymbol{q}}_{\lfloor n_3 r\rfloor} - \boldsymbol{Q}) - D^{-1}\frac{1}{\sqrt{n_3}}\sum_{s=1}^{\lfloor n_3 r\rfloor} M_s \right\|_2 = o_{\mathbb{P}}(1). \tag{145}$$

With a singleton library, no probe stage, $\widehat{B} = I_k$, and $n_3 = n$, this gives the fixed-mechanism online asymptotics. Under assumptions C.5 and C.6, the post-burn-in version is theorem D.1.

Proof of corollary H.5. Use the same $b_N = N^\rho$, $0 < \rho < 1/2$, and localization event $\mathcal{A}_N$ as in the proof of lemma H.4. The initial-segment contribution is bounded by $Cb_N/\sqrt{N} = o(1)$ uniformly over all upper limits $m \le b_N$. For $m > b_N$, sum (143) from $b_N + 1$ to m.

The nonlinear term is controlled uniformly by its full absolute sum:

$$\max_{m\le N} \sum_{s=b_N+1}^{m} \|r(e_{s-1})\|_2 \le \sum_{s=b_N+1}^{N} \|r(e_{s-1})\|_2 = o_{\mathbb{P}}(\sqrt{N})$$

by Step 2 of the preceding proof. The cumulative part of the Abel term satisfies

$$\max_{m\le N} \left\| \sum_{s=b_N+1}^{m-1} (\gamma_{s+1}^{-1} - \gamma_s^{-1}) e_s \right\|_2 \le \sum_{s=b_N+1}^{N-1} (\gamma_{s+1}^{-1} - \gamma_s^{-1}) \|e_s\|_2 ,$$

whose conditional expectation is $O(N^{a/2}) = o(\sqrt{N})$ by (139).

It remains to control all Abel endpoints at once. Choose a fixed $p > 2/(1-a)$. Markov's inequality and a union bound give, for every $\eta > 0$,

$$\begin{aligned}
&\max_r \mathbb{P}_r \left\{ \max_{b_N \le m \le N} \gamma_m^{-1} \|e_m\|_2 > \eta\sqrt{N} \,\middle|\, \mathcal{G}_n \right\} \\
&\quad \le C_{p,\eta} N^{-p/2} \sum_{m=b_N}^{N} \gamma_m^{-p} \mathbb{E}_r(\|e_m\|_2^p \mid \mathcal{G}_n) \\
&\quad \le C_{p,\eta} N^{-p/2} \sum_{m=1}^{N} m^{ap/2} = O\{N^{1-p(1-a)/2}\} = o(1).
\end{aligned}$$

Thus the endpoint, cumulative Abel, nonlinear, and initial-segment terms are all $o_{\mathbb{P}}(\sqrt{N})$ uniformly in the upper limit. Adding $\mathbb{P}_r(\mathcal{A}_N^c \mid \mathcal{G}_n) = o(1)$ and dividing by $\sqrt{N}$ proves (145). The singleton-library and unpreconditioned cases follow by setting r fixed and $B = I_k$ in the same bounds. □

H.5 Inference proofs

Let $\chi^2_{k,1-\alpha}$ be the usual chi-square critical value and let $c^{\max}_{r,\infty,\alpha}$ denote the $(1-\alpha)$ quantile of $\max_j |Z_j|/\sqrt{(\Sigma_r)_{jj}}$ for $Z \sim N_k(0, \Sigma_r)$.

Theorem H.6 (Frozen-score critical limits)**.** *Under assumption C.3, uniformly over the finite candidate library,*

$$c^W_{r,n_3,\alpha} \to \chi^2_{k,1-\alpha}, \qquad c^{\max}_{r,n_3,\alpha} \to c^{\max}_{r,\infty,\alpha}.$$

The boundary contribution vanishes as the frozen laws approach their continuous Gaussian limits.

Proposition H.7 (Transfer of public finite calibration)**.** *Let $\mathcal{R}_n$ be public reference randomness, included in $\mathcal{G}_n$, and let $(c^g_{r,n_3}(\mathcal{R}_n), \eta^g_{r,n_3}(\mathcal{R}_n))$ be any measurable critical rule for $g \in \{W, \max\}$. With an independent $U \sim \mathrm{Unif}(0,1)$, define*

$$d^g_{\mathrm{cal},n} = \max_{r \in \mathfrak{M}_n} \left| \mathbb{P}\left\{ T^{g,\circ}_{r,n_3} < c^g_{r,n_3} \text{ or } (T^{g,\circ}_{r,n_3} = c^g_{r,n_3}, U \le \eta^g_{r,n_3}) \,\middle|\, \mathcal{R}_n, \mathcal{G}_n \right\} - (1-\alpha) \right|. \tag{146}$$

The probability is over a fresh frozen statistic and U, with the realized public rule held fixed. The selected online confidence set satisfies

$$\left| \mathbb{P}\{\boldsymbol{Q} \in \mathcal{C}^g_{1-\alpha}(\mathcal{R}_n) \mid \mathcal{G}_n\} - (1-\alpha) \right| \le d^g_{\mathrm{cal},n} + o_{\mathbb{P}}(1). \tag{147}$$

For B_{cal} frozen draws, empirical boundary randomization and the Dvoretzky–Kiefer–Wolfowitz inequality give, simultaneously over the library,

$$d^g_{\mathrm{cal},n} \le \sqrt{\frac{\log(2R_{\max}/\delta_{\mathrm{cal}})}{2B_{\mathrm{cal}}}}$$

with probability at least $1-\delta_{\mathrm{cal}}$. Exact enumeration has $d^g_{\mathrm{cal},n} = 0$. An independent randomized-rank calibration has exact acceptance probability $1-\alpha$ in the frozen experiment, jointly over its new reference sample and independent test scores. The online procedure inherits asymptotic coverage through the transfer proved below.

Corollary H.8 (Oracle limit after consistent selection)**.** *If the plug-in selector is consistent for a unique oracle sequence r_n^*, then the selected estimator is approximated, in bounded-Lipschitz distance, by $N_k(0, D^{-1}\Sigma_{r_n^*}D^{-1})$. If $\Sigma_{r_n^*} \to \Sigma_\star$ for a deterministic positive-definite matrix $\Sigma_\star$, then*

$$\sqrt{n_3}(\overline{\boldsymbol{q}}_{n_3} - \boldsymbol{Q}) \Rightarrow N_k(0, D^{-1}\Sigma_\star D^{-1}). \tag{148}$$

If the selected covariance converges in distribution, the unconditional limit is the corresponding Gaussian mixture.

Proof of theorem H.6. Put $X_{r,i} = \Sigma_r^{-1/2}\xi^\circ_{r,i}$. Then $\mathbb{E}X_{r,i} = 0$, $\mathbb{E}X_{r,i}X_{r,i}^\top = I_k$, and

$$\sup_{n,r} \mathbb{E}\,\|X_{r,i}\|_2^3 \le G^3_{\max}\underline{\sigma}^{-3} < \infty.$$

Bentkus's convex-set Berry–Esseen theorem (Bentkus, 2003) therefore gives

$$\sup_{r\in\mathfrak{M}_n}\sup_{A\text{ convex}}\left|\mathbb{P}\left(N^{-1/2}\sum_{i=1}^{N}X_{r,i}\in A\right)-\mathbb{P}(Z\in A)\right|\le\frac{C}{\sqrt{N}},\qquad Z\sim N_k(0,I_k).\tag{149}$$

Wald statistic. Taking A to be a Euclidean ball in (149) yields

$$\sup_r\sup_{x\ge0}\left|\mathbb{P}(T_{r,N}^{W,\circ}\le x)-\mathbb{P}(\chi_k^2\le x)\right|\le CN^{-1/2}.$$

The chi-square distribution is continuous and strictly increasing at its $(1-\alpha)$ quantile. The defining inequalities for a generalized quantile, together with the last uniform cdf bound, imply $\sup_r|c_{r,N,\alpha}^{W}-\chi^2_{k,1-\alpha}|\to0$.

Maximum statistic. For fixed x, the event in (10) is a rectangle and hence convex, so (149) also gives a uniform cdf approximation to the candidate-specific Gaussian maximum. The standardized covariance matrices

$$R_r=\operatorname{diag}(\Sigma_r)^{-1/2}\Sigma_r\operatorname{diag}(\Sigma_r)^{-1/2}$$

belong to a compact set of correlation matrices whose smallest eigenvalue is bounded away from zero by a constant depending only on $\underline{\sigma}$ and $\overline{\sigma}$. Gaussian rectangle probabilities are jointly continuous in (x,R) on this set, and the law of $\max_j|Z_j|$ has no atoms. If uniform quantile convergence failed, compactness would produce a subsequence $R_r\to R_0$ and critical values separated from the $(1-\alpha)$ quantile for R_0, contradicting cdf convergence and strict increase of the limiting Gaussian-max cdf at that quantile. Hence $\sup_r|c_{r,N,\alpha}^{\max}-c_{r,\infty,\alpha}^{\max}|\to0$.

Finally, the largest point mass of either frozen statistic tends to zero uniformly by the same cdf approximation and continuity of the Gaussian limits. Thus any boundary probability satisfying (30) has an asymptotically vanishing contribution, which proves the final assertion. □

Proof of theorem C.11. Condition on $\mathcal{G}_n$ and write $r=\widehat{r}$, $N=n_3$. By (57), $\widehat{D}\to D$, and the uniform eigenvalue bounds, the conditional law of $\sqrt{N}\,\widehat{D}(\overline{\boldsymbol{q}}_N-\boldsymbol{Q})$ is uniformly approximated by $N_k(0,\Sigma_r)$ in bounded-Lipschitz distance. The compact covariance class and Gaussian anti-concentration upgrade this to uniform cdf approximation for ellipsoids and rectangles. The independent frozen sum has the same candidate-indexed Gaussian approximation by (149). Consequently, for $g\in\{W,\max\}$,

$$\max_{r\in\mathfrak{M}_n}\sup_x\left|\mathbb{P}_r(T_{r,N}^{g,\mathrm{on}}\le x\mid\mathcal{G}_n)-\mathbb{P}(T_{r,N}^{g,\circ}\le x\mid\mathcal{G}_n)\right|=o_{\mathbb{P}}(1),\tag{150}$$

where the online statistics use $\widehat{D}$ and the frozen statistics are those in (7) and (10). The same bound holds for strict inequalities because the largest frozen and Gaussian point masses vanish uniformly.

For the exact randomized critical pair, the acceptance probability is $F(c-)+\eta\{F(c)-F(c-)\}$. Applying (150) to both $F(c-)$ and $F(c)$ shows that the online acceptance probability differs by $o_{\mathbb{P}}(1)$ from the exactly calibrated frozen probability $1-\alpha$. This proves (59) and (60), including ties and atoms.

Conditional on the privatized Stage-3 transcript, Z^* in (33) is exactly Gaussian with covariance

$$\widehat{\Omega}=\widehat{D}^{-1}\widehat{\Sigma}_U\widehat{D}^{-1}.$$

By (58) and $\widehat{D}\to D$, $\left\|\widehat{\Omega}-\Omega_r\right\|_{\mathrm{op}}=o_{\mathbb{P}}(1)$, where $\Omega_r=D^{-1}\Sigma_rD^{-1}$. Its diagonal entries are bounded away from zero with probability tending to one, and the corresponding correlation matrix differs from $\operatorname{Corr}(\Sigma_r)$ by $o_{\mathbb{P}}(1)$. Conditional on the transcript, the multiplier maximum is exactly the Gaussian maximum for this estimated correlation matrix. Compactness of the admissible correlation matrices makes its cdf uniformly continuous in the matrix argument. The actual studentized maximum has the candidate-indexed Gaussian approximation established above. The triangle inequality proves (61); continuity and strict increase of the Gaussian-max cdf at its fixed-level quantile give the asserted bootstrap coverage.

All approximation constants are common over the realized library. Finally, the absolute conditional coverage error is bounded by one; convergence in probability to zero and uniform integrability imply convergence of its expectation to zero. This proves the unconditional claims. □

Proof of proposition H.7. Condition on $\mathcal{G}_n$, including $\mathcal{R}_n$, and put $r = \widehat{r}$. For a cdf F, acceptance with boundary probability η at c is $(1-\eta)F(c-) + \eta F(c)$. Hence the difference between the online and frozen acceptance probabilities is at most the sum of their cdf differences at $c-$ and c. The uniform comparison (150) makes this difference $o_{\mathbb{P}}(1)$, while (146) bounds the difference between frozen acceptance and $1-\alpha$. This proves (147), including when the selector depends on the same public reference randomness because the maximum in (146) is taken before substituting $r = \widehat{r}$.

For the empirical boundary rule, apply the Dvoretzky–Kiefer–Wolfowitz inequality to each candidate and use a union bound over at most $R_{\max}$ candidates. With probability at least $1-\delta_{\mathrm{cal}}$, every empirical cdf differs uniformly from its population cdf by at most the displayed square root; the same is true for the left limits. Because randomized acceptance is the convex combination $(1-\eta)F(c-) + \eta F(c)$, its population value at the empirical critical value differs from its empirical value $1-\alpha$ by no more than that same square root.

For the rank construction, condition only on the pilot, probes, selection references, and selected mechanism. In the frozen experiment, the observed statistic and the B_{cal} new references are exchangeable after continuous tie breaking, so the observed rank is exactly uniform. Randomization at one rank realizes any acceptance probability in $[0,1]$ and in particular $1-\alpha$. Replacing the frozen observed statistic by the online statistic changes the acceptance probability by $o(1)$ through (150); integrating over the new reference sample proves the claim. □

Proof of corollary H.8. Selection consistency gives $\mathbb{P}(\widehat{r} = r_n^*) \to 1$. For every $\varphi \in \mathrm{BL}_1$, split its expectation according to this event. The complementary contribution is at most $2\mathbb{P}(\widehat{r} \neq r_n^*) = o(1)$, while (57) approximates the contribution on the selection event by the Gaussian law with covariance $D^{-1}\Sigma_{r_n^*}D^{-1}$. This proves the candidate-indexed approximation. If $\Sigma_{r_n^*} \to \Sigma_\star$, continuity of Gaussian laws in their covariance matrices and Slutsky's theorem give (148). The coverage statements follow from theorem C.11 and proposition H.7. □

I Stochastic-approximation tools and proofs

The proofs use the same post-burn-in segment as the statistics in appendix E. Constants may depend on the fixed regularity parameters in assumptions C.5 and C.6.

Proof of theorem D.1. Under assumption C.5, the local assumptions hold with $\beta = 1$, and the singleton fixed-mechanism recursion is the specialization $B = I_k$, r fixed of the construction in lemmas H.1 to H.3. Those lemmas give parts (i)–(ii), including the quantitative last-projection argument.

For later reference, Taylor's theorem under $\sup|f'| \le L_f$ gives

$$h(\boldsymbol{q}_{t-1}) = -D(\boldsymbol{q}_{t-1} - \boldsymbol{Q}) + r_{t-1}, \qquad \|r_{t-1}\|_2 \le \frac{L_f}{2}\,\|\boldsymbol{q}_{t-1} - \boldsymbol{Q}\|_2^2\,, \tag{151}$$

and every unprojected update obeys

$$De_{t-1} = M_t + r_{t-1} - \gamma_t^{-1}(e_t - e_{t-1}). \tag{152}$$

The three-step proof of lemma H.4, with $N = n$, now applies verbatim: an initial segment of length $b_n = n^\rho$, $0 < \rho < 1/2$, is deterministically negligible; the full nonlinear sum has expectation $O(\sum_t \gamma_t) = O(n^{1-a}) = o(\sqrt{n})$; and the expectation of the complete Abel term is $O(n^{a/2}) = o(\sqrt{n})$. This proves (63) without summing pointwise stochastic orders.

The covariance-continuity calculation (144) gives $n^{-1}\sum_t \mathbb{E}_{t-1}(M_t M_t^\top) \to \Sigma$. Bounded increments make the conditional Lindeberg sum eventually zero. The martingale triangular-array central limit theorem and (63) therefore prove (64).

Finally, the proof of corollary H.5 specializes directly to $B = I_k$ and a singleton library. In particular, it uses the pth-moment bound with $p > 2/(1-a)$ to control all Abel endpoints and the expectation of the full weighted Abel sum to control all cumulative terms. This proves (65) and completes every part of the theorem. □

Lemma I.1 (Exponential stability)**.** *Under assumptions C.5 and C.6, there are constants $c_0, C_0 > 0$ such that*

$$\sup_{t\geq 1} \mathbb{E} \exp\left\{ \frac{c_0 \left\| \boldsymbol{q}_t - \boldsymbol{Q} \right\|_2^2}{G^2 \gamma_t} \right\} \leq C_0. \tag{153}$$

Consequently, for every $1 \leq b < n$ and $r > 0$,

$$\mathbb{P}\left\{ \max_{b\leq t\leq n} \left\| \boldsymbol{q}_t - \boldsymbol{Q} \right\|_2 > r \right\} \leq C_0(n-b+1) \exp\left\{ -\frac{c_0 r^2}{G^2 \gamma_b} \right\}. \tag{154}$$

Proof. Put $V_t = \left\| \boldsymbol{q}_t - \boldsymbol{Q} \right\|_2^2$, write $f_0 = f_-$, and let $e = \boldsymbol{q}_{t-1} - \boldsymbol{Q}$. Projection nonexpansiveness, the strong monotonicity in assumption C.5, and $\left\| \boldsymbol{\tau} - a_{Y_t} \right\|_2 \leq G$ give

$$V_t \leq (1 - 2f_0\gamma_t) V_{t-1} + 2\gamma_t \left\langle e, M_t \right\rangle + G^2 \gamma_t^2.$$

Conditionally on $\mathcal{F}_{t-1}$, $\langle e, M_t \rangle$ has mean zero and absolute value at most $2G\sqrt{V_{t-1}}$. Hoeffding's lemma therefore yields, for every $\lambda > 0$,

$$\mathbb{E}_{t-1} \exp\{2\lambda\gamma_t \left\langle e, M_t \right\rangle\} \leq \exp\{8\lambda^2 \gamma_t^2 G^2 V_{t-1}\}.$$

Fix $0 < \eta \leq 1/16$ and set $\lambda_t = \eta f_0/(G^2\gamma_t)$. Combining the preceding two displays gives

$$\begin{aligned} \mathbb{E}_{t-1} e^{\lambda_t V_t} &\leq \exp(\eta f_0 \gamma_t) \exp[\lambda_t \{1 - 2f_0\gamma_t + 8\eta f_0 \gamma_t\} V_{t-1}] \\ &\leq \exp(\eta f_0 \gamma_t) \exp[\lambda_t (1 - 3f_0\gamma_t/2) V_{t-1}]. \end{aligned}$$

Because $\gamma_t = \gamma_0 t^{-a}$ with $a < 1$, $\lambda_t/\lambda_{t-1} = \gamma_{t-1}/\gamma_t = 1 + O(t^{-1})$, whereas $\gamma_t \asymp t^{-a}$. Hence there is a deterministic t_0 such that, for every $t \geq t_0$,

$$\lambda_t (1 - 3f_0\gamma_t/2) \leq (1 - f_0\gamma_t)\lambda_{t-1}.$$

The step-size condition implies $0 < f_0\gamma_t \leq 1/4$. With $L_t = \log \mathbb{E} e^{\lambda_t V_t}$, concavity of $x \mapsto x^{1-f_0\gamma_t}$ now gives

$$L_t \leq \eta f_0 \gamma_t + (1 - f_0\gamma_t) L_{t-1}, \qquad t \geq t_0.$$

Consequently, $L_t \leq \max\{L_{t_0}, \eta\}$ for all $t \geq t_0$. The finitely many preceding indices are harmless because $\mathcal{C}$ is compact. Thus (153) holds with $c_0 = \eta f_0$ and a finite C_0.

For $t \geq b$, Markov's inequality and $\gamma_t \leq \gamma_b$ imply

$$\mathbb{P}\{\left\| \boldsymbol{q}_t - \boldsymbol{Q} \right\|_2 > r\} \leq C_0 \exp\left\{ -\frac{c_0 r^2}{G^2 \gamma_b} \right\}.$$

A union bound over $t = b, \ldots, n$ proves (154). □

Corollary I.2 (Explicit tail-interior failure)**.** *If $\gamma_{b+1} G \leq \Delta_Q/8$, then*

$$\pi_{b,n} \leq C_0(m+1) \exp\left\{ -\frac{c_0 \Delta_Q^2}{64 G^2 \gamma_b} \right\}. \tag{155}$$

Proof. On the event $\max_{b\leq t\leq n} \left\| \boldsymbol{q}_t - \boldsymbol{Q} \right\|_\infty \leq \Delta_Q/8$, the gap calculation in the proof of theorem D.1 makes every tail update interior. Since the Euclidean norm dominates the sup norm, lemma I.1 with $r = \Delta_Q/8$ gives the result. □

Lemma I.3 (One-step adaptive-to-frozen bound)**.** *There is a coupling for which, conditionally on $\mathcal{F}_{t-1}$,*

$$\mathbb{P}(Y_t \neq Y_t^* \mid \mathcal{F}_{t-1}) \leq f_+ \sqrt{k} \left\| \boldsymbol{q}_{t-1} - \boldsymbol{Q} \right\|_2, \tag{156}$$

$$\mathbb{E}\{M_t - \xi_t^* \mid \mathcal{F}_{t-1}\} = 0, \tag{157}$$

$$\mathbb{E}\{\left\| M_t - \xi_t^* \right\|_2^2 \mid \mathcal{F}_{t-1}\} \leq 8A_0^2 f_+ \sqrt{k} \left\| \boldsymbol{q}_{t-1} - \boldsymbol{Q} \right\|_2 + 2f_+^2 \left\| \boldsymbol{q}_{t-1} - \boldsymbol{Q} \right\|_2^2. \tag{158}$$

Proof. The categories differ only if X_t lies between $q_{t-1,j}$ and Q_j for at least one j. Hence

$$\mathbb{P}(C_t \neq C_t^* \mid \mathcal{F}_{t-1}) \leq \sum_{j=1}^{k} |F(q_{t-1,j}) - F(Q_j)| \leq f_+ \sqrt{k} \, \|\boldsymbol{q}_{t-1} - \boldsymbol{Q}\|_2 .$$

A maximal channel coupling makes the report mismatch probability no larger than the category mismatch probability, proving (156).

Both M_t and ξ_t^* have conditional mean zero, proving (157). Moreover,

$$M_t - \xi_t^* = a_{Y_t^*} - a_{Y_t} - h(\boldsymbol{q}_{t-1}).$$

The first difference is zero when the reports agree and has norm at most $2A_0$ otherwise. Also $\|h(\boldsymbol{q}_{t-1})\|_2 \leq f_+ \|\boldsymbol{q}_{t-1} - \boldsymbol{Q}\|_2$. Using $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$ gives (158). □

Proof of theorem D.3. Work first on $\mathcal{I}_{b,n}$. Multiplying (152) by a deterministic weight $w_i = w_{\ell,m}(i)$ and summing gives

$$D \sum_{i=1}^{m} w_i e_i = \sum_{i=1}^{m} w_i M_{t_i} + \sum_{i=1}^{m} w_i r_{t_i - 1} - \sum_{i=1}^{m} w_i \gamma_{t_i}^{-1} (e_{i+1} - e_i), \tag{159}$$

By lemma I.1 and a union bound, outside an event of probability at most δ,

$$\max_{b \leq t \leq n} \|\boldsymbol{q}_t - \boldsymbol{Q}\|_2 \leq C G \gamma_b^{1/2} \Lambda_{m,K}(\delta)^{1/2}. \tag{160}$$

By assumption D.2, all tail step sizes are comparable to m^{-a}. By (151), Cauchy–Schwarz, and $\|w\|_1 \leq \sqrt{m}$,

$$\begin{aligned} \left\| \sum_{i=1}^{m} w_i r_{t_i - 1} \right\|_2 &\leq \frac{L_f}{2} \sum_i |w_i| \, \|e_i\|_2^2 \\ &\leq C G^2 m^{1/2-a} \Lambda_{m,K}(\delta). \end{aligned}$$

For the Abel term, put $v_i = w_i \gamma_{t_i}^{-1}$. Then

$$\sum_{i=1}^{m} v_i (e_{i+1} - e_i) = v_m e_{m+1} - v_1 e_1 - \sum_{i=1}^{m-1} (v_{i+1} - v_i) e_{i+1}.$$

Regular variation gives

$$\sum_{i=1}^{m-1} |v_{i+1} - v_i| \leq C \gamma_b^{-1} \left\{ \sum_{i=1}^{m-1} |w_{i+1} - w_i| + m^{-1} \sum_{i=1}^{m} |w_i| \right\}.$$

Combining this with (160), $\sum_i |w_i| \leq \sqrt{m}$, and the definition of $\mathfrak{v}(\mathcal{W})$ yields

$$\left\| \sum_{i=1}^{m} w_i \gamma_{t_i}^{-1} (e_{i+1} - e_i) \right\|_2 \leq C G \mathfrak{v}(\mathcal{W}) m^{-(1-a)/2} \Lambda_{m,K}(\delta)^{1/2}. \tag{161}$$

Equations (159)–(161), followed by multiplication by $\Sigma^{-1/2}$, control the online-to-martingale error simultaneously after a union bound over $K+1$ weights.

It remains to replace M_{t_i} by $\xi_{t_i}^*$. Put $\Delta_i = M_{t_i} - \xi_{t_i}^*$. By lemma I.3, $(\Delta_i, \mathcal{F}_{t_i})$ is a martingale-difference sequence. On (160), its weighted conditional quadratic variation satisfies

$$\begin{aligned} \sum_i w_i^2 \mathbb{E}_{t_i - 1} \|\Delta_i\|_2^2 \leq C \big\{ & A_0^2 f_+ \sqrt{k} G m^{-a/2} \Lambda_{m,K}(\delta)^{1/2} \\ & + f_+^2 G^2 m^{-a} \Lambda_{m,K}(\delta) \big\}. \end{aligned}$$

Also, $\|w_i \Delta_i\| \leq H_0 \max_i |w_i|$. Apply scalar Freedman's inequality (Freedman, 1975) on a $1/2$-net of the unit sphere, whose cardinality is at most 5^k, and take a union bound over $K+1$ weights. The

standard net inequality $\|z\| \le 2\max_{u\in\mathcal{N}} \left|u^\top z\right|$ gives, outside an additional event of probability at most 2δ,

$$\max_{\ell}\left\|\sum_i w_{\ell,m}(i)\Delta_i\right\|_2 \le C\big[A_0(f_+\sqrt{k}G)^{1/2}m^{-a/4}\Lambda_{m,K}(\delta) + f_+Gm^{-a/2}\Lambda_{m,K}(\delta) + H_0\mu(\mathcal{W})m^{-1/2}\Lambda_{m,K}(\delta)\big].$$

Multiplication by σ_-^{-1} and addition of the three errors prove (72). Setting $a = 2/3$ and upper-bounding smaller powers by $m^{-1/6}$ proves (73). □

Proof of theorem D.4. Work on $\mathcal{I}_{b,n}$ and sum (152) from $i = 1$ to an arbitrary $s \le m$. After division by $\sqrt{m}$,

$$\Sigma^{-1/2}D\,m^{-1/2}\sum_{i=1}^{s} e_i = \Sigma^{-1/2}m^{-1/2}\sum_{i=1}^{s} M_{t_i} + \Sigma^{-1/2}m^{-1/2}\sum_{i=1}^{s} r_{t_i-1} - \Sigma^{-1/2}m^{-1/2}\sum_{i=1}^{s}\gamma_{t_i}^{-1}(e_{i+1}-e_i),$$

where $e_{m+1} = \boldsymbol{q}_n - \boldsymbol{Q}$. On the event (160),

$$\frac{1}{\sqrt{m}}\max_{s\le m}\left\|\sum_{i=1}^{s} r_{t_i-1}\right\|_2 \le CG^2m^{1/2-a}\Lambda_m^{\mathrm{path}}(\delta).$$

For the boundary term, Abel summation gives

$$\sum_{i=1}^{s}\gamma_{t_i}^{-1}(e_{i+1}-e_i) = \gamma_{t_s}^{-1}e_{s+1} - \gamma_{t_1}^{-1}e_1 - \sum_{i=1}^{s-1}(\gamma_{t_{i+1}}^{-1} - \gamma_{t_i}^{-1})e_{i+1}.$$

Since the inverse step sizes increase and all tail times are comparable to m,

$$\frac{1}{\sqrt{m}}\max_{s\le m}\left\|\sum_{i=1}^{s}\gamma_{t_i}^{-1}(e_{i+1}-e_i)\right\|_2 \le CGm^{-(1-a)/2}\Lambda_m^{\mathrm{path}}(\delta)^{1/2}.$$

It remains to replace M_{t_i} by $\xi_{t_i}^*$. Put $\Delta_i = M_{t_i} - \xi_{t_i}^*$. By lemma I.3, the partial sums of Δ_i form a vector martingale. On (160), their total conditional quadratic variation is at most

$$C\left\{A_0^2 f_+\sqrt{k}Gm^{1-a/2}\Lambda_m^{\mathrm{path}}(\delta)^{1/2} + f_+^2G^2m^{1-a}\Lambda_m^{\mathrm{path}}(\delta)\right\}.$$

Apply the maximal scalar Freedman inequality (Freedman, 1975) on a $1/2$-net of the unit sphere, use $\|\Delta_i\| \le H_0$, and take a union bound over the net. After division by $\sqrt{m}$, this gives, outside probability 2δ,

$$\max_{s\le m}\left\|\frac{1}{\sqrt{m}}\sum_{i=1}^{s}\Delta_i\right\|_2 \le C\big[A_0(f_+\sqrt{k}G)^{1/2}m^{-a/4}\Lambda_m^{\mathrm{path}}(\delta) + f_+Gm^{-a/2}\Lambda_m^{\mathrm{path}}(\delta) + H_0m^{-1/2}\Lambda_m^{\mathrm{path}}(\delta)\big].$$

Multiplication by σ_-^{-1} and addition of the preceding terms prove (76). Centering a partial-sum path by its endpoint can at most double the sup-norm error, proving (77). Setting $a = 2/3$ gives (78); (79) follows from the calculation in corollary D.5 with the basis-roughness terms removed. □

Lemma I.4 (Frozen path norm and integral-object conversion)**.** *Let* $M_m^{\mathrm{path}}(\delta)$ *be defined by* (119)*. Then*

$$\mathbb{P}\left\{\max_{0\le s\le m}\|\mathbb{P}_m^\circ(s)\|_2 > M_m^{\mathrm{path}}(\delta)\right\} \le \delta. \tag{162}$$

Define the standardized integral normalizers

$$\widehat{V}_m = \frac{1}{m}\sum_{s=1}^{m-1}\widehat{\mathbb{U}}_m(s)\widehat{\mathbb{U}}_m(s)^\top, \qquad V_m^\circ = \frac{1}{m}\sum_{s=1}^{m-1}\mathbb{U}_m^\circ(s)\mathbb{U}_m^\circ(s)^\top.$$

On the events in (75) *and* (162)*,*

$$\left\|\widehat{\mathbb{P}}_m(m) - \mathbb{P}^\circ_m(m)\right\|_2 + \left\|\widehat{V}_m - V^\circ_m\right\|_{\mathrm{op}} \le C\{(1 + M^{\mathrm{path}}_m) r^{\mathrm{path}}_m + (r^{\mathrm{path}}_m)^2\}. \tag{163}$$

Thus (163) *fails with probability at most* $\pi_{b,n} + 4\delta$*.*

Proof. For each fixed s and unit vector v, $v^\top \mathbb{P}^\circ_m(s)$ is a sum of independent centered variables with variance at most one and increments bounded by $\ell/\sqrt{m}$. Bernstein's inequality, a $1/2$-net of $\mathbb{S}^{k-1}$, and a union bound over $s = 1, \dots, m$ prove (162).

The bridge error is at most $2r^{\mathrm{path}}_m$ by (77), while every frozen bridge has norm at most $2M^{\mathrm{path}}_m$. Hence each summand in $\widehat{V}_m - V^\circ_m$ has operator norm at most $C\{M^{\mathrm{path}}_m r^{\mathrm{path}}_m + (r^{\mathrm{path}}_m)^2\}$. Averaging and adding the endpoint error proves (163). □

Proof of corollary D.5. Insert (134) and (136) into (74). The terms are bounded respectively by $C\sqrt{k}(1+\sqrt{K})/\sqrt{\mu_\varepsilon}$, $Ck/(\delta_\varepsilon\sqrt{\mu_\varepsilon})$, $Ck/(\sqrt{\delta_\varepsilon\mu_\varepsilon})$, and $C\sqrt{k}\mu(\mathcal{W})/\sqrt{\mu_\varepsilon}$. For $\varepsilon \le 1$, $\delta_\varepsilon \asymp \varepsilon/k$ and $\mu_\varepsilon \asymp 1/k$ on a balanced grid. The second term is then of order $k^{5/2}/\varepsilon$ and dominates the displayed bound. □

I.1 Exponential-stability comparison

This appendix fills in the scalar comparison step in lemma I.1. Let $V_t = \|\boldsymbol{q}_t - \boldsymbol{Q}\|^2$ and $f_0 = f_-$. The projected update gives

$$V_t \le (1 - 2f_0\gamma_t)V_{t-1} + 2\gamma_t \langle e_{t-1}, M_t\rangle + G^2\gamma_t^2. \tag{164}$$

Because $\|\boldsymbol{\tau} - a_{Y_t}\| \le G$ and its conditional mean is $h(\boldsymbol{q}_{t-1})$, we have $\|M_t\| \le 2G$. Thus $|\langle e_{t-1}, M_t\rangle| \le 2G\sqrt{V_{t-1}}$, and conditional Hoeffding gives

$$\mathbb{E}_{t-1}\exp\{2\lambda\gamma_t \langle e_{t-1}, M_t\rangle\} \le \exp\{8\lambda^2\gamma_t^2 G^2 V_{t-1}\}. \tag{165}$$

Choose $\lambda_t = c/(G^2\gamma_t)$ with $0 < c \le f_0/16$. Combining (164)–(165),

$$\begin{aligned}\mathbb{E}_{t-1}e^{\lambda_t V_t} &\le \exp(c\gamma_t)\exp\left(\lambda_t[1 - 2f_0\gamma_t + 8c\gamma_t]V_{t-1}\right)\\ &\le \exp(c\gamma_t)\exp\left(\lambda_t[1 - \tfrac{3}{2}f_0\gamma_t]V_{t-1}\right).\end{aligned}$$

Since $\gamma_t = \gamma_0 t^{-a}$ with $a < 1$,

$$\frac{\lambda_t}{\lambda_{t-1}} = \left(\frac{t}{t-1}\right)^a \le 1 + \frac{2a}{t} \le 1 + \frac{f_0\gamma_t}{2}$$

for all $t \ge t_0$. Hence

$$\lambda_t(1 - \tfrac{3}{2}f_0\gamma_t) \le (1 - f_0\gamma_t)\lambda_{t-1}$$

for $t \ge t_0$. Put $Z_t = \mathbb{E}e^{\lambda_t V_t}$ and $L_t = \log Z_t$. By concavity of $x \mapsto x^{1-f_0\gamma_t}$,

$$\begin{aligned}Z_t &\le e^{c\gamma_t}\mathbb{E}\left[e^{\lambda_{t-1}V_{t-1}}\right]^{1-f_0\gamma_t},\\ L_t &\le c\gamma_t + (1 - f_0\gamma_t)L_{t-1}.\end{aligned}$$

Therefore $L_t \le \max\{L_{t_0}, c/f_0\}$ by induction. The finitely many indices below t_0 are finite because $\mathcal{C}$ is bounded. This proves (153) after adjusting constants. Markov's inequality gives

$$\mathbb{P}(V_t > r^2) \le C_0\exp\{-c_0 r^2/(G^2\gamma_t)\},$$

and summing this bound over $t = b, \dots, n$ proves (154).

J SPECTRAL AND PIVOTAL LEMMAS AND PROOFS

Proof of proposition E.1. The coordinatewise Karhunen–Loève expansion of a Brownian bridge is

$$B(r) - rB(1) = \sum_{\ell=1}^{\infty} \frac{\sqrt{2}\sin(\pi\ell r)}{\pi\ell} G_\ell.$$

Orthogonality of the sine functions gives (83). The sums $\sum \ell^{-2} = \pi^2/6$ and $\sum \ell^{-4} = \pi^4/90$ yield (84). The first k Gaussian vectors span $\mathbb{R}^k$ almost surely, which proves positive definiteness. Choose a measurable unit vector u orthogonal to $G_1, \ldots, G_{k-1}$. Conditional on these vectors, u is independent of the remaining Gaussian vectors and $\mathbb{E}(u^\top G_\ell)^2 = 1$. Thus

$$\lambda_{\min}(V_k) \le u^\top V_k u = \sum_{\ell=k}^{\infty} \lambda_\ell (u^\top G_\ell)^2.$$

Expectation and the integral test prove (85). The matrix inequality (86) is immediate because the weights decrease. A J-degree Wishart lower tail has polynomial exponent $(J-k+1)/2$, which can be made arbitrarily large at fixed k; however, $\lambda_J J \asymp 1/J$, giving scale $1/k$ when $J \asymp k$. □

Proof of theorem E.2. The vector $(X_0^\top, \ldots, X_K^\top)^\top$ is jointly Gaussian. Its block covariance is

$$\mathrm{Cov}(X_r, X_s) = \sum_{i=1}^{m} w_{r,m}(i) w_{s,m}(i) I_k = \mathbf{1}\{r = s\} I_k.$$

Thus the blocks are independent standard normals and $K\widehat{S}_K \sim W_k(I_k, K)$ independently of X_0. The Hotelling–Wishart identity (Hotelling, 1931) proves (90). For the online conclusion, stack the $K+1$ frozen coefficients. Its ith independent summand is $(w_{0,m}(i), \ldots, w_{K,m}(i)) \otimes \Sigma^{-1/2}\xi_i^*$. Condition (91) makes the maximal summand norm vanish and orthonormality makes the covariance the identity. The triangular-array multivariate Lindeberg theorem therefore gives independent Gaussian limits. Combine this with theorem D.3 and the continuous mapping theorem on the event of a nonsingular limiting Wishart matrix. □

Lemma J.1 (Uniform small-ball covariance bound)**.** *Let $Y_1, \ldots, Y_K$ be independent and identically distributed isotropic random vectors in $\mathbb{R}^k$. Suppose that for constants $a_0, p_0 > 0$,*

$$\inf_{\|v\|_2=1} \mathbb{P}\{|v^\top Y_1| \ge a_0\} \ge p_0.$$

For $0 < \delta < 1$, if

$$\frac{8}{a_0}\sqrt{\frac{k}{K}} + \sqrt{\frac{2\log(1/\delta)}{K}} \le \frac{p_0}{2}, \tag{166}$$

then, with probability at least $1-\delta$,

$$\lambda_{\min}\left(\frac{1}{K}\sum_{\ell=1}^{K} Y_\ell Y_\ell^\top\right) \ge \frac{a_0^2 p_0}{8}. \tag{167}$$

Proof. Let $\psi : \mathbb{R} \to [0,1]$ equal zero on $[-a_0/2, a_0/2]$, equal one outside $[-a_0, a_0]$, and be linear in between. It is $2/a_0$-Lipschitz and $t^2 \ge (a_0^2/4)\psi(t)$. For every unit v, $\mathbb{E}\psi(v^\top Y_1) \ge p_0$.

By symmetrization and the contraction principle,

$$\begin{aligned}\mathbb{E}\sup_{\|v\|=1}\left|\frac{1}{K}\sum_{\ell=1}^{K}\{\psi(v^\top Y_\ell) - \mathbb{E}\psi(v^\top Y_\ell)\}\right| &\le \frac{8}{a_0}\mathbb{E}\left\|\frac{1}{K}\sum_{\ell=1}^{K}\varepsilon_\ell Y_\ell\right\|_2 \\ &\le \frac{8}{a_0}\sqrt{\frac{k}{K}},\end{aligned}$$

where ε_ℓ are independent Rademacher variables and isotropy gives the last inequality. The supremum changes by at most $1/K$ when one observation changes, so bounded differences add at most

$\sqrt{2\log(1/\delta)/K}$ with probability $1-\delta$. Under (166), every empirical average of $\psi(v^\top Y_\ell)$ is at least $p_0/2$. Therefore

$$\frac{1}{K}\sum_{\ell=1}^{K}(v^\top Y_\ell)^2 \geq \frac{a_0^2}{4}\frac{1}{K}\sum_{\ell=1}^{K}\psi(v^\top Y_\ell) \geq \frac{a_0^2 p_0}{8}$$

uniformly over the unit sphere. The Rayleigh quotient proves (167). □

Proof of theorem E.3. Each Haar coefficient is

$$X_\ell^\circ = \frac{1}{\sqrt{2s}}\left(\sum_{i=1}^{s} Z_{\ell,i} - \sum_{i=s+1}^{2s} Z_{\ell,i}\right), \qquad Z_{\ell,i} = \Sigma^{-1/2}\xi_{\ell,i}^*.$$

Disjoint supports give independence across ℓ; orthonormality gives isotropy. For fixed unit v, the scalar summands have mean zero, total variance one, and sum of third absolute moments at most $\beta_3/\sqrt{2s}$. The scalar Berry–Esseen theorem gives

$$\sup_x \left|\mathbb{P}(v^\top X_\ell^\circ \leq x) - \Phi(x)\right| \leq \frac{C_{\mathrm{BE}}\beta_3}{\sqrt{2s}}.$$

Applying this at $x = 1/2$ and $x = -1/2$ proves (96). Invoke lemma J.1 with $a_0 = 1/2$ and $p_0 = p_s$ to obtain (98). The final statement follows by elementary substitution. □

Lemma J.2 (Gaussian smallest singular value)**.** *Let G be a $J \times k$ matrix with independent $N(0,1)$ entries, $J \geq k$. For every $x \geq 0$,*

$$\mathbb{P}\{s_{\min}(G) \leq \sqrt{J} - \sqrt{k} - x\} \leq e^{-x^2/2}. \tag{168}$$

Proof. Gordon's Gaussian comparison inequality (Gordon, 1985; Davidson & Szarek, 2001) gives $\mathbb{E}s_{\min}(G) \geq \sqrt{J} - \sqrt{k}$. The map $G \mapsto s_{\min}(G)$ is one-Lipschitz in Frobenius norm. Gaussian concentration therefore gives the lower-tail deviation (168). □

Proof of proposition F.1. For the series normalizer, write $S_K^G = G^\top G/K$ and apply lemma J.2. For integral self-normalization, retain the first J terms in (87) and use $\bar{\lambda}_{J,m} \geq 1/(\pi^2 J^2)$; the Brownian proof is identical using (86). A centered Gaussian sample covariance with B_0 observations is Wishart with $B_0 - 1$ degrees of freedom, giving the third statement. The final scales follow by substitution. □

J.1 Discrete bridge transform

We prove (87). It is enough to work with one scalar coordinate. Let $z = (z_1, \ldots, z_m)^\top$ and

$$b_s = \frac{1}{\sqrt{m}}\left\{\sum_{i=1}^{s} z_i - \frac{s}{m}\sum_{i=1}^{m} z_i\right\}, \qquad 1 \leq s \leq m-1.$$

If z_i are independent standard normals, the covariance matrix C_m of $(b_1, \ldots, b_{m-1})$ is

$$(C_m)_{rs} = \frac{1}{m}\left\{\min(r,s) - \frac{rs}{m}\right\}.$$

Let L_m be the $(m-1)\times(m-1)$ tridiagonal matrix with diagonal entries two and adjacent entries -1. Direct multiplication gives $C_m = (mL_m)^{-1}$. The vectors

$$u_{\ell,m}(s) = \sqrt{\frac{2}{m}}\sin\left(\frac{\pi \ell s}{m}\right), \qquad 1 \leq \ell, s \leq m-1, \tag{169}$$

form an orthonormal basis and satisfy

$$L_m u_{\ell,m} = 4\sin^2\left(\frac{\pi\ell}{2m}\right) u_{\ell,m}.$$

Therefore the squared singular values of the map $z \mapsto (b_s)_{s<m}$ are $\{4m\sin^2(\pi\ell/(2m))\}^{-1}$.

The corresponding right singular vectors in increment space are the DCT vectors

$$w_{\ell,m}(i) = \sqrt{\frac{2}{m}} \cos\left\{\frac{\pi\ell(i-1/2)}{m}\right\}, \qquad 1 \le \ell \le m-1.$$

This follows either by taking discrete differences of (169), or by direct multiplication. Since $S_m^{\mathrm{SN}} = m^{-1}\sum_{s=1}^{m-1} b_s^2$, one additional factor m^{-1} gives

$$S_m^{\mathrm{SN}} = \sum_{\ell=1}^{m-1} \frac{Y_{\ell,m}^2}{4m^2 \sin^2\{\pi\ell/(2m)\}}, \qquad Y_{\ell,m} = \sum_{i=1}^{m} w_{\ell,m}(i) z_i.$$

Applying the identity coordinatewise proves (87). Finally, $\sin x \le x$ gives (88).

K Coverage-transfer tools and proofs

Lemma K.1 (Frozen coefficient norm)**.** *Let $X_0^\circ, \dots, X_K^\circ$ be the frozen coefficients in* (70)*, and let $M_{m,K}(\delta)$ be defined by* (99)*. Then*

$$\mathbb{P}\left\{\max_{0\le\ell_0\le K} \left\|X_{\ell_0}^\circ\right\|_2 > M_{m,K}(\delta)\right\} \le \delta. \tag{170}$$

Proof. For a fixed unit vector u and coefficient index ℓ_0, $u^\top X_{\ell_0}^\circ$ is a sum of independent mean-zero variables with total variance one and individual absolute values bounded by $\ell\mu(\mathcal{W})/\sqrt{m}$. Bernstein's inequality gives

$$\mathbb{P}\{\left|u^\top X_{\ell_0}^\circ\right| > t\} \le 2\exp\left\{-c\min\left(t^2, \frac{t\sqrt{m}}{\ell\mu(\mathcal{W})}\right)\right\}.$$

Take a $1/2$-net of the unit sphere with at most 5^k points, use $\|x\| \le 2\max_{u\in\mathcal{N}} \left|u^\top x\right|$, and take a union bound over $K+1$ coefficients. Solving the Bernstein threshold with $M_{m,K}(\delta)$ in (99) proves (170). □

Lemma K.2 (Coefficient-to-object conversion)**.** *Suppose*

$$\max_{0\le\ell\le K} \left\|\widehat{X}_\ell - X_\ell^\circ\right\|_2 \le r, \qquad \max_{0\le\ell\le K} \|X_\ell^\circ\|_2 \le M.$$

Let $\widehat{S} = K^{-1}\sum_{\ell=1}^K \widehat{X}_\ell \widehat{X}_\ell^\top$ and $S^\circ = K^{-1}\sum_{\ell=1}^K X_\ell^\circ X_\ell^{\circ\top}$. Then

$$\left\|\widehat{X}_0 - X_0^\circ\right\|_2 + \left\|\widehat{S} - S^\circ\right\|_{\mathrm{op}} \le (1+2M)r + r^2.$$

Proof. For every ℓ,

$$\left\|\widehat{X}_\ell\widehat{X}_\ell^\top - X_\ell^\circ X_\ell^{\circ\top}\right\|_{\mathrm{op}} \le \left\|\widehat{X}_\ell - X_\ell^\circ\right\| \left\{\left\|\widehat{X}_\ell\right\| + \|X_\ell^\circ\|\right\} \le r(2M+r).$$

Average over ℓ and add the endpoint error. □

Lemma K.3 (Quadratic-form transfer)**.** *Let (X, S) and (X°, S°) be coupled. Suppose that, outside an event of probability η,*

$$\|X - X^\circ\|_2 + \|S - S^\circ\|_{\mathrm{op}} \le \rho.$$

Let $T = X^\top S^{-1} X$ and $T^\circ = X^{\circ\top} S^{\circ -1} X^\circ$, with value $+\infty$ for a singular normalizer. If $0 < 2\rho \le u \le 1$, then for every $M, c > 0$,

$$|\mathbb{P}(T \le c) - \mathbb{P}(T^\circ \le c)| \le \eta + \mathbb{P}(\|X^\circ\| > M) + \mathbb{P}\{\lambda_{\min}(S^\circ) < u\} + \omega_{T^\circ}\left(c, \frac{C(M+1)^2\rho}{u^2}\right).$$

If T° has density bounded by A near c, the last term is at most $2CA(M+1)^2\rho/u^2$.

Proof. On the coupling event and $\{\lambda_{\min}(S^\circ) \geq u\}$, Weyl's inequality gives $\lambda_{\min}(S) \geq u - \rho \geq u/2$. Thus

$$\left\|S^{-1}\right\|_{\mathrm{op}} \leq 2u^{-1}, \qquad \left\|S^{-1} - S^{\circ -1}\right\|_{\mathrm{op}} \leq 2\rho u^{-2}.$$

If $\|X^\circ\| \leq M$, then $\|X\| \leq M + \rho$ and expansion of the two quadratic forms gives $|T - T^\circ| \leq C(M+1)^2 \rho u^{-2}$. The two indicators can differ only on a failed good event or when T° lies in the stated interval around c. Taking probabilities proves the result. □

Lemma K.4 (Reference densities). *(i) Let $T^G_{k,K} = X^\top S_K^{-1} X$ in the Gaussian series experiment and put $\nu = K - k + 1$. Its density is*

$$\frac{\nu}{kK} f_{F_{k,\nu}}\left(\frac{\nu t}{kK}\right). \tag{171}$$

For every $C_0 > 1$, $\eta_0 > 0$, and $\alpha_0 > 0$, if $1 + \eta_0 \leq K/k \leq C_0$, then the density in a fixed relative neighborhood of its $(1-\alpha)$ quantile is bounded by

$$A^{\mathrm{SER}}_{k,K,\alpha} \leq \frac{C_{C_0,\eta_0,\alpha_0}}{\sqrt{k}}. \tag{172}$$

The same order holds for the block-Hotelling reference when $(B_0 - 1)/k$ is bounded above and away from one.

(ii) For $k \geq 2$, the finite-m Gaussian integral pivot and its Brownian limit have densities bounded by

$$A^{\mathrm{SN}}_k \leq \frac{k}{12}.$$

For $k = 1$, the density is bounded in a neighborhood of every positive critical value.

Proof. The exact F law gives (171). Alternatively, if $U \sim \mathrm{Beta}(k/2, \nu/2)$, then $T^G_{k,K} = KU/(1-U)$ and

$$f_U\left(\frac{t}{K+t}\right)\frac{K}{(K+t)^2}$$

is its density at t. When K/k is bounded above and bounded away from one and α stays away from zero and one, the relevant beta quantile remains in a compact subinterval of $(0,1)$. Stirling's bounds for the beta normalizing constant give $\sup f_U \leq C\sqrt{k}$, while $K/(K+t)^2 \leq C/k$. This proves (172).

For integral self-normalization, the Gaussian endpoint is independent of the Gaussian bridge and hence of its normalizer, both on the finite grid and in the Brownian limit. Condition on a positive definite normalizer S and diagonalize it. The pivot is $\sum_{j=1}^k a_j Z_j^2$ with $a_j = \lambda_j(S)^{-1}$. The convolution of two scaled χ^2_1 densities is bounded by $(2\sqrt{a_1 a_2})^{-1}$. Selecting the two largest a_j makes this at most $\mathrm{tr}(S)/2$; convolution with the remaining nonnegative terms cannot increase the sup norm. For the finite-m integral normalizer, $\mathbb{E}\,\mathrm{tr}(S) = k\sum_{\ell=1}^{m-1} \bar{\lambda}_{\ell,m} < k/6$; for Brownian motion the sum equals $k/6$. Taking expectation gives $k/12$. The one-dimensional conditional density is locally bounded away from zero by the finite moment of $\sqrt{S}$. □

Proof of theorem F.2. For $Z_i = \Sigma^{-1/2}\xi_i^*$, the frozen endpoint is $m^{-1/2}\sum_i Z_i$. The acceptance region $\{x : \|x\|^2 \leq \chi^2_{k,1-\alpha}\}$ is convex. Bentkus's convex-set theorem (Bentkus, 2003) gives

$$\sup_{C \text{ convex}} \left|\mathbb{P}(m^{-1/2}\textstyle\sum_i Z_i \in C) - \mathbb{P}(N_k(0, I_k) \in C)\right| \leq 400 k^{1/4}\beta_3/\sqrt{m}.$$

On the coupling event from theorem D.3, the online and frozen radii differ by at most $r_{m,0}$. The shell probability is the difference of two ball probabilities. Applying Bentkus to both balls contributes at most twice the preceding error, while the standard Gaussian radial density is uniformly bounded. Adding the frozen calibration error and the coupling failure proves (100). □

Proof of corollary F.3. Apply proposition G.1 and corollary D.5 to theorem F.2. □

Proposition K.5 (Explicit independent-pilot radius)**.** *If $\widetilde{\boldsymbol{Q}}$ is the pilot tail average, then with probability at least $1-\pi_p-4\delta$,*

$$\max_j \left|\widetilde{Q}_j - Q_j\right| \le r_p^{\mathrm{SA}}(\delta) := \frac{\omega_\infty}{\sqrt{m_p}}\{M_{m_p,0}(\delta) + r_{m_p,0}(\delta)\}. \tag{173}$$

For kRR,

$$\Omega_{jj} = \frac{\nu_j(1-\nu_j)}{\delta_\varepsilon^2 f(Q_j)^2}, \qquad \omega_\infty = \max_j \frac{\sqrt{\nu_j(1-\nu_j)}}{\delta_\varepsilon f(Q_j)}. \tag{174}$$

Thus the pilot contribution has a fully explicit (m_p, ε, k) dependence; for small ε, its largest coordinate scale is at most of order $k/(\varepsilon f_-\sqrt{m_p})$ before the frozen-norm and online-remainder factors.

Proof. The standardized pilot endpoint obeys $\left\|\Sigma^{-1/2} D\sqrt{m_p}(\widetilde{\boldsymbol{Q}} - \boldsymbol{Q})\right\| \le M_{m_p,0} + r_{m_p,0}$ on the intersection of the frozen norm and online coupling events. The Euclidean norm of the jth row of $D^{-1}\Sigma^{1/2}$ is $\sqrt{\Omega_{jj}}$, which proves (173). Equation (174) follows from (132). □

Proof of proposition F.5. Work first on the pilot event in (102), whose complement has probability at most δ, and condition on $\widetilde{\boldsymbol{Q}}$. For coordinate j, any two decoded values satisfy

$$|a_{r_0,y,j} - a_{r_0,y',j}| \le |a_{r_0,y,j} - \tau_j| + |a_{r_0,y',j} - \tau_j| \le 2G_0.$$

Hoeffding's inequality therefore gives, for each sign,

$$\mathbb{P}\left\{|\overline{a}_{\pm,j} - F(\widetilde{Q}_j \pm h)| > G_0\sqrt{\frac{2\log(4k/\delta)}{n_{2,\pm}}} \,\middle|\, \widetilde{\boldsymbol{Q}}\right\} \le \frac{\delta}{2k}.$$

A union bound over the $2k$ coordinate–sign pairs shows that, outside conditional probability δ, the stochastic part of every central difference is at most the third term of (103).

Taylor's theorem with integral remainder and $\sup|f''| \le L_{f,2}$ gives

$$\left|\frac{F(x+h) - F(x-h)}{2h} - f(x)\right| \le \frac{L_{f,2}}{6}h^2.$$

The bound $\sup|f'| \le L_f$ from assumption C.5 also gives $|f(\widetilde{Q}_j) - f(Q_j)| \le L_f r_p(\delta)$. Adding the stochastic and two bias terms proves (103) on the intersection of the pilot and report events, whose probability is at least $1-2\delta$.

Clipping to an interval that contains $f(Q_j)$ is the Euclidean projection onto that interval and cannot increase absolute error, proving the truncated claim. If $n_{2,+}$ and $n_{2,-}$ are proportional to n_2, the stochastic term is A/h with $A \asymp G_0\{\log(k/\delta)/n_2\}^{1/2}$. Balancing h^2 and A/h gives $h \asymp A^{1/3}$ and (104). Finally, on the displayed event set e_f in lemma K.6 equal to the right-hand side of (103); substituting its resulting deterministic bound for d in theorem F.7 proves the last assertion. □

Proof of proposition F.6. Conditional on the pilot, every decoded bit is unbiased and has range at most $1/r_\varepsilon$. Since $m_c \ge 4k$, the common cell size obeys $N_c \ge m_c/(4k)$. Hoeffding's inequality and a union bound over $2k$ cells give

$$\max_{j,\pm}\left|\widehat{F}_{j,\pm} - F(\widetilde{Q}_j \pm h)\right| \le \frac{C}{r_\varepsilon}\sqrt{\frac{k\log(4k/\delta)}{m_c}}$$

with probability $1-\delta$. Division of the difference by $2h$ gives the stochastic term in (105). Taylor's theorem for the symmetric difference gives bias at most $L_{f,2}h^2/6$ at $\widetilde{Q}_j$, while the displacement to Q_j contributes at most $L_f r_p(\delta)$. Balancing h^2 with the stochastic term A/h gives (106) and (107). □

Lemma K.6 (Relative covariance perturbation)**.** *Put $e_f = \max_j\left|\widehat{f}_j - f(Q_j)\right|$ and assume $e_f \le f_-/2$. Then*

$$\delta_\Omega := \left\|\Omega^{-1/2}(\widehat{\Omega} - \Omega)\Omega^{-1/2}\right\|_{\mathrm{op}} \le 2\sqrt{\kappa_\Sigma}\eta_f + \kappa_\Sigma\eta_f^2, \qquad \eta_f = \frac{2e_f}{f_-}.$$

Proof. Let $R = D\widehat{D}^{-1} = I+E$. Since $\widehat{\Omega} = D^{-1}R\Sigma RD^{-1}$, the relative perturbation is orthogonally similar to $\Sigma^{-1/2}(R\Sigma R-\Sigma)\Sigma^{-1/2}$. Expand $R\Sigma R-\Sigma = E\Sigma+\Sigma E+E\Sigma E$. The first two normalized terms have norm at most $\sqrt{\kappa_\Sigma}\,\|E\|$, and the third has norm at most $\kappa_\Sigma\,\|E\|^2$. Finally, $\|E\| \le 2e_f/f_-$ when $e_f \le f_-/2$. □

Lemma K.7 (Fixed-level chi-square shell)**.** *Let $q_{k,\alpha} = \chi^2_{k,1-\alpha}$. Uniformly over $\alpha \in [\alpha_0, 1-\alpha_0]$ and $0 < d < 1/2$,*

$$\mathbb{P}\{q_{k,\alpha}(1-d) < \chi^2_k \le q_{k,\alpha}(1+d)\} \le C_{\alpha_0}\sqrt{k}\,d. \tag{175}$$

Proof. For k larger than a constant depending only on α_0, Chebyshev's inequality applied to χ^2_k, whose mean and variance are k and $2k$, shows that $c_{\alpha_0}k \le q_{k,\alpha} \le C_{\alpha_0}k$. The finitely many smaller k are absorbed into the same constants. For $k \ge 2$, Stirling's bounds applied to the gamma density give $\sup_x f_{\chi^2_k}(x) \le C/\sqrt{k}$; for $k = 1$, the interval in (175) stays inside a compact subset of $(0,\infty)$ uniformly over the stated (α, d) range because $d < 1/2$ and the fixed-level quantile is bounded away from zero. The shell length is $2q_{k,\alpha}d \le C_{\alpha_0}kd$, proving (175). □

Proof of theorem F.7. Put $q = q_{k,\alpha}$, $x = \sqrt{m}\,\Omega^{-1/2}(\bar{\boldsymbol{q}}_{b,m} - \boldsymbol{Q})$, and $E_\Omega = \Omega^{-1/2}(\widehat{\Omega} - \Omega)\Omega^{-1/2}$. On $\{\delta_\Omega \le d\}$, the eigenvalues of $I + E_\Omega$ lie in $[1-d, 1+d]$, so

$$\frac{T^{\mathrm{or}}_m}{1+d} \le T^{\mathrm{plug}}_m = x^\top(I+E_\Omega)^{-1}x \le \frac{T^{\mathrm{or}}_m}{1-d}.$$

Consequently,

$$\{T^{\mathrm{or}}_m \le q(1-d)\} \subseteq \{T^{\mathrm{plug}}_m \le q\} \subseteq \{T^{\mathrm{or}}_m \le q(1+d)\}$$

on that event. Repeat the online-to-frozen coupling and Bentkus comparison in the proof of theorem F.2, now uniformly for the two balls with squared radii $q(1-d)$ and $q(1+d)$. This contributes at most

$$1200\frac{k^{1/4}\beta_3}{\sqrt{m}} + \pi_{b,n} + 3\delta + C_{\alpha_0}r_{m,0}(\delta).$$

The remaining Gaussian probability is the multiplicative chi-square shell in lemma K.7, bounded by $C_{\alpha_0}\sqrt{k}\,d$. Adding $\mathbb{P}(\delta_\Omega > d)$ proves (108). □

Corollary K.8 (Fully allocated plug-in bound)**.** *Use three disjoint user groups: a main online group of retained size m, a pilot online group of retained size m_p, and a density-calibration group of size m_c. For a common $\delta \in (0, 1/8)$, define*

$$e^{\mathrm{tot}}_f(h,\delta) = \frac{L_{f,2}}{6}h^2 + L_f r^{\mathrm{SA}}_p(\delta) + \frac{C}{r_\varepsilon h}\sqrt{\frac{k\log(4k/\delta)}{m_c}}, \tag{176}$$

$$\eta^{\mathrm{tot}}_f(h,\delta) = \frac{2e^{\mathrm{tot}}_f(h,\delta)}{f_-}, \qquad d^{\mathrm{tot}}(h,\delta) = 2\sqrt{\kappa_\Sigma}\eta^{\mathrm{tot}}_f + \kappa_\Sigma(\eta^{\mathrm{tot}}_f)^2.$$

If $e^{\mathrm{tot}}_f(h,\delta) \le f_-/2$ and $d^{\mathrm{tot}}(h,\delta) < 1/2$, then

$$\begin{aligned}\mathrm{CE}^{\mathrm{plug}}_\alpha \le\ & 1200\frac{k^{1/4}\beta_3}{\sqrt{m}} + \pi_{b,n} + \pi_p + 8\delta \\ & + C_{\alpha_0}\{r_{m,0}(\delta) + \sqrt{k}\,d^{\mathrm{tot}}(h,\delta)\}.\end{aligned} \tag{177}$$

With h chosen according to (106)*, the additional covariance-perturbation order is*

$$O_f\left(\sqrt{k}\left[\left\{\frac{k\log(k/\delta)}{m_c\tanh^2(\varepsilon/2)}\right\}^{1/3} + \frac{\omega_\infty}{\sqrt{m_p}}\{M_{m_p,0}(\delta) + r_{m_p,0}(\delta)\}\right]\right), \tag{178}$$

up to the displayed condition-number multiplier and a quadratic term of the same bracket. For kRR, ω_∞ is given exactly by (174)*, so* (177) *is explicit in $(m, m_p, m_c, \varepsilon, k)$.*

Proof. Intersect the pilot event in proposition K.5, the calibration event in proposition F.6, and the main-recursion events in theorem F.2. Their failure probabilities sum to at most $\pi_p + 5\delta$ for the pilot and density stages and $\pi_{b,n} + 3\delta$ for the main stage. On the intersection, (176) bounds e_f; lemma K.6 then gives $\delta_\Omega \le d^{\mathrm{tot}}$. Apply theorem F.7 with this deterministic d. Balancing the first and third terms of (176) yields (178). □

Proof of theorem F.8. Apply lemma K.3 and add the reference calibration error. If the coupling event has already been intersected with $\{\|X_m^\circ\|_2 \le M\}$, repeat the proof of lemma K.3 on that joint event; the norm-tail event is then part of the single failure probability η and is not counted again. For an independent public random critical rule, condition on its randomization, apply the same deterministic-threshold argument, and integrate; the assumed uniform conditional local-concentration bound is preserved. For smoothing, use the same auxiliary uniform variable in the online and reference pivots. Their difference is unchanged, while the frozen finite component has density at most $(2h)^{-1}$. □

Proof of theorem F.9. Let E_1 be the intersection of the coefficient-coupling event in (71) and the frozen coefficient-norm event in (170). By theorem D.3 and lemma K.1,

$$\mathbb{P}(E_1^c) \le \pi_{b,n} + 4\delta.$$

On E_1, lemma K.2 gives

$$\left\|\widehat{X}_0 - X_0^\circ\right\|_2 + \left\|\widehat{S} - S^\circ\right\|_{\mathrm{op}} \le \rho = (1+2M)r + r^2.$$

Let $E_2 = \{\lambda_{\min}(S^\circ) \ge u_H\}$. The Haar spectral theorem gives $\mathbb{P}(E_2^c) \le \delta$. Since $2\rho \le u_H$, Weyl's inequality guarantees that the online normalizer is nonsingular on $E_1 \cap E_2$.

By definition, the public finite-critical calibration rule has frozen error d_{cal}. Couple the online and frozen smoothing variables identically. Conditional on every realization of the public calibration randomness, the finite component of the smoothed frozen statistic has density at most $(2h)^{-1}$; hence the same bound holds uniformly for a random calibrated threshold. Applying the public-randomness version of theorem F.8 therefore gives

$$\mathrm{CE}_\alpha^{\mathrm{HSC}} \le d_{\mathrm{cal}} + \mathbb{P}(E_1^c) + \mathbb{P}(E_2^c) + \frac{C(M+1)^2\rho}{hu_H^2},$$

which is (110).

For the simplified rate, the stated leverage condition gives $M \le C\Lambda_{m,K}^{1/2}$, while (73) gives $r \le C\mathfrak{R}_{k,\varepsilon,K}(\mathcal{W})m^{-1/6}\Lambda_{m,K}$. When r^2 is dominated by $(1+2M)r$, multiplication by $(M+1)^2$ produces the factor $\Lambda_{m,K}^{5/2}$ in (111). Finally, substitute (80) and use $K \asymp k$ to obtain (112). □

Proof of theorem F.11. The online-to-frozen coefficient event in theorem D.3 fails with probability at most $\pi_{b,n} + 3\delta$. By (115), couple the stacked frozen coefficient vector $\mathbf{X}^\circ$ and the Gaussian vector $\mathbf{G}$ so that $\|\mathbf{X}^\circ - \mathbf{G}\|_2 \le r_G$ outside probability η_G. A Gaussian norm bound and a union bound over $K+1$ blocks give $\max_{0\le\ell\le K}\|G_\ell\|_2 \le M_G$ outside probability δ. On the intersection of these events, every online coefficient differs from its Gaussian counterpart by at most $r_{\mathrm{tot}} = r_{m,K} + r_G$. Expanding the K outer products as in lemma K.2 gives the object radius $\rho_G = (1+2M_G)r_{\mathrm{tot}} + r_{\mathrm{tot}}^2$.

In the Gaussian reference experiment the critical value in (92) is exact, so the reference-calibration error is zero. By proposition F.1, the Gaussian normalizer has $\lambda_{\min} \ge u$ outside probability $e^{-x^2/2}$. By lemma K.4, the reference statistic has density at most $A_{k,K,\alpha}^{\mathrm{SER}}$ in the required neighborhood. Apply theorem F.8 to the online and Gaussian objects. The failure probabilities are $\pi_{b,n} + 3\delta$, η_G, the additional Gaussian norm probability δ, and the spectral probability $e^{-x^2/2}$, which proves (117).

When $K = \lceil ck \rceil$ and $x = (\sqrt{K} - \sqrt{k})/2$, proposition F.1 gives a constant lower spectral threshold and an exponentially small spectral failure, while lemma K.4 gives $A_{k,K,\alpha}^{\mathrm{SER}} \le C_{c,\alpha_0}k^{-1/2}$ and $(M_G+1)^2 \le C\{k + \log(1/\delta)\}$. Substitution proves (118). Finally, for fixed (k, K), (113)–(114) give $\Gamma_{m,K} = O_{k,K}(m^{-1/2})$. Choosing $\eta_G = \Gamma_{m,K}^{1/2}$ makes both η_G and $r_G = \Gamma_{m,K}/\eta_G$ of order $m^{-1/4}$. □

Proof of theorem F.12. The raw statistic (81) is unchanged by applying the same nonsingular transformation $\Sigma^{-1/2}D$ to its endpoint and every bridge vector. Its standardized online object is therefore $(\widehat{\mathbb{P}}_m(m), \widehat{V}_m)$, and the corresponding frozen object is $(\mathbb{P}^\circ_m(m), V^\circ_m)$.

By lemma I.4, outside probability $\pi_{b,n} + 4\delta$ these objects differ by at most $\rho^{\mathrm{score}}_{\mathrm{SN}}$. For an arbitrary threshold u satisfying $2\rho^{\mathrm{score}}_{\mathrm{SN}} \le u$, the frozen small-eigenvalue probability is exactly $s^\circ_{m,\mathrm{SN}}(u)$; no Gaussian tail is inserted. The public finite-critical calibration rule has frozen error d_{cal}. With the same public smoothing variable in the online and frozen statistics, the finite frozen component has density at most $(2h)^{-1}$. The public-randomness version of theorem F.8 now gives

$$\mathrm{CE}^{\mathrm{SN,score}}_\alpha \le d_{\mathrm{cal}} + \pi_{b,n} + 4\delta + s^\circ_{m,\mathrm{SN}}(u) + \frac{C(M+1)^2\rho^{\mathrm{score}}_{\mathrm{SN}}}{hu^2},$$

which is (121).

At $a = 2/3$, (78) gives $r \le C\mathfrak{R}^{\mathrm{path}}_{k,\varepsilon} m^{-1/6} L$, where $L = k + \log(4m/\delta)$. Under the stated smaller-order conditions, $M \lesssim L^{1/2}$ and $\rho^{\mathrm{score}}_{\mathrm{SN}} \lesssim Mr$, so $(M+1)^2\rho^{\mathrm{score}}_{\mathrm{SN}} \lesssim \mathfrak{R}^{\mathrm{path}}_{k,\varepsilon} m^{-1/6} L^{5/2}$. This proves (122); inserting (79) proves (123). □

Proposition K.9 (Finite-grid-to-Brownian calibration)**.** *Let*

$$L^B_m(\delta) = k + \log\frac{8m}{\delta}.$$

For $J = \lceil ck \rceil$ with fixed $c > 1$, suppose $J < m$ and $CL^B_m(\delta)/\sqrt{m} \le u_{\mathrm{SN}}(J, k, (\sqrt{J} - \sqrt{k})/2)/2$. Then, for $k \ge 2$,

$$\left|\mathbb{P}\{T^{\mathrm{SN},G}_m \le c^B_{k,\alpha}\} - (1-\alpha)\right| \le 3\delta + e^{-c_c k} + C_{c,\alpha_0}\frac{k^3\{L^B_m(\delta)\}^2}{\sqrt{m}}. \tag{179}$$

For $k = 1$, the same conclusion holds with a constant depending on a positive neighborhood of the Brownian critical value.

Proof. Construct a k-dimensional Brownian motion W whose increments $\sqrt{m}\{W(i/m) - W((i-1)/m)\}$ are the Gaussian score increments. The finite Gaussian endpoint is then exactly $W(1)$, and its discrete bridge is $U(s/m) = W(s/m) - (s/m)W(1)$. The Brownian reference uses the same endpoint and $V = \int_0^1 U(r)U(r)^\top dr$. Its finite-grid counterpart is $V^G_m = m^{-1}\sum_{s=1}^{m-1} U(s/m)U(s/m)^\top$.

A $1/2$-net argument, the scalar reflection principle on each mesh interval, and a union bound give, outside probability δ,

$$\max_{0\le s<m}\sup_{s/m\le r\le(s+1)/m}\|U(r) - U(s/m)\|_2 \le C\sqrt{\frac{L^B_m(\delta)}{m}}.$$

The same argument for the full interval gives $\sup_{0\le r\le 1}\|U(r)\|_2 \le C\sqrt{L^B_m(\delta)}$ outside another event of probability δ. Comparing the integral on each mesh interval with its left endpoint and using $\|xx^\top - yy^\top\|_{\mathrm{op}} \le \|x-y\|(\|x\| + \|y\|)$ yields

$$\|V^G_m - V\|_{\mathrm{op}} \le C\frac{L^B_m(\delta)}{\sqrt{m}}$$

outside probability 2δ; the omitted left endpoint at zero is identically zero. The endpoints are identical. Apply lemmas K.3 and K.4 and proposition F.1 with the Brownian object, its critical value, and $u = u_{\mathrm{SN}}(J, k, (\sqrt{J} - \sqrt{k})/2) \asymp k^{-1}$. The Brownian endpoint norm contributes one further δ, the spectral failure is $e^{-c_c k}$, and $A^{\mathrm{SN}}_k \lesssim k$ together with $u^{-2} \asymp k^2$ and $\|W(1)\|^2 = O\{L^B_m(\delta)\}$ gives the final term in (179). □

Proof of theorem F.13. Start from the online-to-frozen path event in lemma I.4, whose failure probability is at most $\pi_{b,n} + 4\delta$. On the additional process-coupling event (124), the frozen partial-sum path differs from the Gaussian path by at most r^{proc}_m. Centering each path by its endpoint doubles this

additional error. Intersect also with the Gaussian path-maximal event $\max_{s\le m}\|\mathbb{G}_m(s)\|_2 \le M_m^G(\delta)$, whose failure probability is at most δ. Expanding the endpoint and bridge outer products exactly as in lemma I.4 gives the total object radius ρ_{SN}^G in (125). Hence the combined coupling and norm event fails with probability at most $\pi_{b,n} + 5\delta + \eta_m^{\mathrm{proc}}$.

The Gaussian reference critical value has calibration error $d_{m,k}^{\mathrm{ref}}$ by definition. By proposition F.1, its normalizer has smallest eigenvalue at least $u = u_{\mathrm{SN}}(J,k,x)$ outside probability $e^{-x^2/2}$. The endpoint norm is already part of the joint event, and lemma K.4 bounds the reference density by Ck for $k \ge 2$. Applying theorem F.8 therefore yields

$$\mathrm{CE}_\alpha^{\mathrm{SN}} \le d_{m,k}^{\mathrm{ref}} + \pi_{b,n} + 5\delta + \eta_m^{\mathrm{proc}} + e^{-x^2/2} + C\frac{k\{k+\log(4m/\delta)\}\rho_{\mathrm{SN}}^G}{u^2},$$

which is (126).

For $J = \lceil ck \rceil$ and $x = (\sqrt{J} - \sqrt{k})/2$, proposition F.1 gives $u \ge c_c/k$ and spectral failure at most $e^{-c_c k}$. Thus $k/u^2 \le C_c k^3$, proving (127). This calculation also isolates the two sources of dimension amplification: $u^{-2} \asymp k^2$ from inversion and the factor k from Gaussian-reference anti-concentration. □

Proof of theorem F.14. Apply theorem D.3 to each restarted block with failure allocation δ/B_0. A union bound gives

$$\max_{1\le r\le B_0}\left\|\widehat{X}_r - X_r^\circ\right\|_2 \le r_B^{\mathrm{on}}$$

outside probability $B_0\pi_B + 3\delta$. By the coupling characterization of the Wasserstein distance and Markov's inequality, couple the stacked frozen vector with $(G_1,\dots,G_{B_0})$ so that $\max_r \|X_r^\circ - G_r\|_2 \le r_B^G$ outside probability η_G. A Gaussian norm bound and a union bound give $\max_r \|G_r\|_2 \le M_B^G(\delta)$ outside an additional probability δ. On the joint event, every online endpoint is within $r_B = r_B^{\mathrm{on}} + r_B^G$ of its Gaussian counterpart.

The sample mean is Lipschitz with constant one under the maximum endpoint perturbation. After subtracting the two sample means, every centered observation changes by at most $2r_B$. Expanding the centered outer products and using the Gaussian norm event yields

$$\left\|\widehat{X}_{\mathrm{mean}} - G_{\mathrm{mean}}\right\|_2 + \left\|\widehat{S} - S_G\right\|_{\mathrm{op}} \le \rho_B = C\{(1+M_B^G)r_B + r_B^2\}.$$

The Gaussian Hotelling critical value is exact. By proposition F.1, the Gaussian sample covariance has smallest eigenvalue at least $u = u_{\mathrm{HOT}}(B_0,k,x)$ outside probability $e^{-x^2/2}$, and lemma K.4 supplies the density bound $A_{k,B_0-1,\alpha}^{\mathrm{HOT}}$. The master coverage theorem gives (129) after adding $B_0\pi_B + 3\delta$, η_G, the Gaussian norm probability δ, and the spectral failure.

If $B_0 - 1 = \lceil ck \rceil$, the spectral threshold is constant and the reference density is $O(k^{-1/2})$, which proves (130). Substitution of the endpoint kRR bound with block length m_B gives (131). When $m_B \asymp n/B_0$ and $B_0 \asymp k$, the online radius contributes $k^{5/2}(n/k)^{-1/6}/\varepsilon = k^{8/3}n^{-1/6}/\varepsilon$ before the remaining norm and anti-concentration multipliers. □