NEAR-OPTIMAL ESTIMATION AND INFERENCE FOR DISTRIBUTION FUNCTIONS UNDER LOCAL DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

We estimate population cumulative distribution functions under local differential privacy using independent cycles of dyadic optimized unary encoding reports. An exact grid-approximation modulus yields continuous uniform risk of order $(\log n)^{3/2}/\{\tanh(\epsilon/2)\sqrt{n}\}$ under public Hölder regularity. An interactive lower bound establishes optimality up to logarithmic factors on a bounded-density class for $0 < \epsilon \leq 1$. Total-mass-constrained least-squares reconstruction with a symmetric monotone envelope retains this rate. We further establish conditions for studentized pointwise Gaussian limits and construct ordered simultaneous bands with finite-sample or asymptotic coverage guarantees. Ordering band endpoints preserves coverage, whereas recentering pointwise intervals at a monotone estimate can fail. Across four simulated distributions at $\epsilon = 1$ and $n \approx 8.4 \times 10^6$, the least-squares estimator reduces mean uniform error by 47–50% relative to one-bit current-status estimation based on Liu et al. (2024). Fixed-grid inference experiments show a 40–43% reduction in mean band width on identical reports.

1 INTRODUCTION

A population cumulative distribution function (CDF) answers every threshold query at once. Cormode et al. (2019) study locally private range queries on taxi pickup times and show why summing noisy bin frequencies loses accuracy over long ranges. For this variable, a CDF gives the fraction of pickups before any chosen time. Under local differential privacy (LDP), each user randomizes an observation before releasing it. The challenge is to support thresholds chosen after collection while retaining accuracy across the curve and its joint sampling structure for inference. Without privacy, the empirical CDF has worst-threshold error of order $n^{-1/2}$ (Dvoretzky et al., 1956; Massart, 1990; van der Vaart & Wellner, 1996).

A single public threshold admits a simple solution: randomized response (RR) (Warner, 1965) gives an unbiased estimate and Gaussian inference at the locally private root-n scale. A curve requires information about many thresholds. Dividing users among many separate queries reduces the sample available at each threshold. In the current-status approach of Liu et al. (2024), each user instead reports one privatized comparison at a public random threshold, and constrained isotonic estimation pools these observations. At fixed privacy parameter ε, their continuous-threshold uniform bound is $O_{\mathbb{P}}\{n^{-1/3} \log n/\rho_\varepsilon\}$ under positive threshold and population densities, where $\rho_\varepsilon = \tanh(\varepsilon/2)$ measures the signal retained by randomized response. Their fixed-grid and growing-grid Gaussian results concern different targets and resolutions; Appendix B.1 makes this distinction precise.

Hierarchies reuse private node reports across range queries, including prefixes: intervals below a threshold. Consistent histograms and wavelets were developed for centrally private range queries (Hay et al., 2010; Xiao et al., 2011). Under LDP, Cormode et al. (2019) study hierarchical and Haar range estimates, including least-squares reconciliation of redundant node masses. Local frequency reporting developed from randomized response through RAPPOR (Erlingsson et al., 2014) to optimized unary encoding (OUE) and Hadamard response (HR) (Wang et al., 2017; Acharya et al., 2019). Square-wave smoothing (Li et al., 2020), adaptive and piecewise-linear trees (Du et al., 2021; Wang et al., 2024), and the wavelet method of Zhao et al. (2026) offer further distribution reconstructions. The latter provides Wasserstein error bounds and evaluates both Wasserstein and Kolmogorov distances. These

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.

approaches motivate multiscale reporting; continuous-domain Kolmogorov risk and uncertainty after shape enforcement require their own analysis.

Two issues connect the finite hierarchy to a population CDF. Grid approximation must account for every real threshold, including jumps between or exactly on grid points. Inference must also distinguish the error of the curve from the error of an individual prefix. A cumulative maximum produces a monotone estimate without enlarging its worst-grid error, yet it can move an estimate by much more than that target's standard error: neighboring prefixes may combine different numbers of noisy nodes. Thus a shape-valid point estimate and a valid confidence band need not be constructed in the same way.

Our construction, TreeOUE, assigns one user per dyadic level to each independent cycle. Each cycle estimates threshold probabilities by summing at most one unbiased node estimate per level, so each report contributes to many thresholds. This reuse yields continuous uniform risk of order $(\log n)^{3/2}/(\rho_\varepsilon\sqrt{n})$ under a public Hölder bound. On their common bounded-density setting and at fixed privacy, it improves the published continuous-uniform bound of Liu et al. (2024). An interactive LDP lower bound establishes optimality up to logarithmic factors for $0 < \varepsilon \leq 1$; the nonprivate root-n rate supplies a separate benchmark. The gain uses public regularity information and longer messages, whose lengths are given in Section 2.2.

Public reconstruction makes this improvement useful at finite sample sizes. Total-mass-constrained least squares (LS) reconciles redundant node estimates; averaging a running maximum with a reverse running minimum enforces monotonicity symmetrically. We show that the resulting CDF retains the same uniform rate. Independent simulations show an increasing error reduction over current-status estimation as the sample size grows. Independent cycles also supply an estimable joint covariance for inference. Raw pointwise studentization and ordered simultaneous bands preserve their respective targets, while fixed-grid LS reconstruction reduces mean band width by 40–43% on identical reports.

The rest of the paper is organized as follows. Sections 2–4 present the method, theory and experiments; Section 5 concludes. The appendices collect implementation details, further results and proofs.

2 A REUSABLE PRIVATE CDF RELEASE

2.1 POPULATION TARGET AND GRID APPROXIMATION

Let $X_1, \ldots, X_n$ be independent observations with a common law P on the public interval $[0, 1]$ and right-continuous CDF $F(x) = P(X \leq x)$. Each user holds one observation; a known bounded interval can be mapped to $[0, 1]$ by an affine transformation. We measure curve estimation by Kolmogorov loss $\|\widehat{F} - F\|_\infty = \sup_{x\in[0,1]} |\widehat{F}(x) - F(x)|$. A pointwise target $x \in (0, 1)$ is public.

Definition 2.1. *Let Z_i be user i's private report and $Z_{1:i-1} = (Z_1, \ldots, Z_{i-1})$ the reports already released. Conditional on public randomness and this history, user i applies a randomized reporting rule (Markov kernel) $Q_i(\cdot \mid X_i, Z_{1:i-1})$. The protocol is ε-locally differentially private if, almost surely in the past transcript, $Q_i(B \mid x, Z_{1:i-1}) \leq e^\varepsilon Q_i(B \mid x', Z_{1:i-1})$ for every measurable output event B and every $x, x' \in [0, 1]$. A noninteractive protocol is the special case in which Q_i does not depend on previous private reports.*

Our protocol is noninteractive; allowing interaction in the lower bound makes the comparison stronger. Write $p_\varepsilon = e^\varepsilon/(1 + e^\varepsilon), q_\varepsilon = 1/(1 + e^\varepsilon), \rho_\varepsilon = p_\varepsilon - q_\varepsilon = \tanh(\varepsilon/2), a_\varepsilon = \rho_\varepsilon/2$. For $D \geq 2$, let $x_k = k/D$. The ideal left-grid representation is

$$F_D^\circ(x) = \begin{cases} 0, & 0 \leq x < x_1, \\ F(x_k), & x_k \leq x < x_{k+1}, \quad 1 \leq k < D, \\ 1, & x = 1, \end{cases} \tag{1}$$

Its exact deterministic approximation error is $\omega_D(F) = \|F_D^\circ - F\|_\infty$. Equivalently, with $F(t-) = P(X < t)$ denoting the left limit, $\omega_D(F) = \max\left\{F(x_1-), \max_{1\leq k<D}\left[F(x_{k+1}-) - F(x_k)\right]\right\}$.

A jump at a positive grid point is represented exactly, whereas an off-grid atom can prevent uniform approximation regardless of estimation accuracy. In particular, a point mass at a positive grid point has $\omega_D(F) = 0$; a point mass strictly between grid points has $\omega_D(F) = 1$. A mass at zero contributes

through the first cell, whose released value is zero. The largest cell mass

$$\Delta_D(F) = \max\{F(x_1), \max_{2\le k\le D}[F(x_k) - F(x_{k-1})]\}$$

is an upper bound on $\omega_D(F)$, but need not equal it. For $s \in (0, 1]$ and $L \ge 1$, define

$$\mathcal{F}(s, L) = \{F : F(0) = 0, \quad F(y) - F(x) \le L(y - x)^s \text{ for all } 0 \le x < y \le 1\}. \tag{2}$$

Then $\omega_D(F) \le \Delta_D(F) \le LD^{-s}$. For the lower bound, $\mathcal{F}_{\mathrm{bd}}$ denotes CDFs with a density satisfying $1/2 \le f \le 3/2$ almost everywhere. Thus $\mathcal{F}_{\mathrm{bd}} \subset \mathcal{F}(1, 3/2)$. The finite-sample bounds retain $\omega_D(F)$ and apply to arbitrary distributions.

2.2 Independent cycles of hierarchical reports

Let $D = 2^J$ with $1 \le J \le n$. At level $r \in \{1, \dots, J\}$, define $I_{r,j} = ((j-1)2^{-r}, j2^{-r}]$, $j = 1, \dots, 2^r$, with zero included in $I_{r,1}$. Define $C_r(x) = j$ when $x \in I_{r,j}$. Every prefix $[0, k/D]$, $1 \le k < D$, has a canonical disjoint representation

$$[0, k/D] = \dot{\bigcup_{(r,j)\in\mathcal{D}(k)}} I_{r,j}, \qquad |\mathcal{D}(k)| \le J,$$

using at most one cell at each level. Its cardinality is the number of ones in the J-digit binary expansion of k. For example, with $D = 8$, the prefix $[0, 5/8]$ is the union of the level-1 cell $[0, 1/2]$ and the level-3 cell $(1/2, 5/8]$. Assigning distinct users to these levels makes their decoded contributions independent. The cycle construction below provides independent repetitions of this prefix estimate while reusing the same reports for every other prefix.

Put $m = \lfloor n/J \rfloor$ and use mJ users, leaving at most $J - 1$ users unused. Partition these users into m disjoint groups of J users, called cycles. Within each cycle $b = 1, \dots, m$, publicly assign one user to each level. A level-r user computes the cell $C_r(X_i)$ and applies OUE (Wang et al., 2017):

$$\mathbb{P}(Y_{i,j} = 1 \mid C_r(X_i) = c) = \begin{cases} 1/2, & j = c, \\ q_\varepsilon, & j \ne c, \end{cases} \qquad j = 1, \dots, 2^r, \tag{3}$$

independently across users, with conditionally independent coordinates. For node $\nu = (r, j)$, write $I_\nu = I_{r,j}$ and decode

$$W_{i,\nu} = \frac{Y_{i,j} - q_\varepsilon}{a_\varepsilon}. \tag{4}$$

Then $\mathbb{E}(W_{i,\nu} \mid X_i) = \mathbf{1}\{X_i \in I_\nu\}$. Writing $i(b, r)$ for the user at level r in cycle b, set $W_{b,(r,j)} = W_{i(b,r),(r,j)}$.

For cycle b and grid prefix k, define

$$V_{b,k} = \sum_{\nu\in\mathcal{D}(k)} W_{b,\nu}, \qquad \widetilde{F}_{m,D}(x_k) = \frac{1}{m}\sum_{b=1}^{m} V_{b,k}.$$

Set the raw endpoint values to zero and one. The raw grid vector need not be monotone. For a grid vector g, let $\mathrm{clamp}_{[0,1]}(u) = \min\{1, \max\{0, u\}\}$ and define the cumulative-maximum map

$$\mathcal{T}(g)_0 = 0, \qquad \mathcal{T}(g)_k = \mathrm{clamp}_{[0,1]}\left(\max_{1\le j\le k} g_j\right), \quad 1 \le k < D, \qquad \mathcal{T}(g)_D = 1. \tag{5}$$

The shape-valid grid estimator is $\widehat{F}_{m,D}(x_k) = \mathcal{T}(\widetilde{F}_{m,D})_k$, extended by

$$\widehat{F}_{m,D}(x) = \widehat{F}_{m,D}(x_k), \quad x \in [x_k, x_{k+1}), \quad k = 0, \dots, D-1, \qquad \widehat{F}_{m,D}(1) = 1. \tag{6}$$

Each user applies one ε-LDP channel, and all subsequent operations use only the reports. The balanced average message length is $(2D - 2)/J$ bits and the maximum is D bits. With public thresholds, the current-status method uses one bit per user. These lengths count uplink responses; assignments are public. Larger D reduces approximation error and increases communication. The public Hölder pair (s, L) determines a resolution attaining the uniform rate in Corollary 3.2; an off-grid pointwise interval may require a finer resolution or an explicit bias bound. Algorithm 1 in Appendix A collects the complete procedure.

2.3 LEAST-SQUARES RECONSTRUCTION

Least squares reconciles redundant parent and child estimates. Following Cormode et al. (2019, Section 4.5), let $\boldsymbol{W}_b$ collect the $2D-2$ decoded node scores in cycle b. Let $\boldsymbol{A}$ be the $(2D-2)\times D$ node–leaf incidence matrix, with $A_{\nu j}=1$ when leaf j lies in node ν, and zero otherwise. The fixed least-squares (LS) reconstruction is

$$\widehat{\boldsymbol{p}}_b = \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^D:\,\mathbf{1}^\top\boldsymbol{v}=1} \|\boldsymbol{A}\boldsymbol{v}-\boldsymbol{W}_b\|_2^2, \qquad V_{b,k}^{\mathrm{LS}} = \sum_{j=1}^{k}\widehat{p}_{b,j}, \quad 1\le k<D. \tag{7}$$

Here $\mathbf{1}$ is the all-ones D-vector. The public weights and total-mass constraint give an affine reconstruction with possibly signed fitted leaf masses. The bottom-level rows of $\boldsymbol{A}$ form the identity matrix, so the least-squares criterion is strictly convex and has a unique constrained minimizer. Its coefficients depend only on the public tree and can be computed once for all cycles.

For point estimation, put $g_k^{\mathrm{LS}} = m^{-1}\sum_{b=1}^{m} V_{b,k}^{\mathrm{LS}}$ and use the symmetric monotone envelope

$$\widehat{F}_{m,D}^{\mathrm{LS}}(x_k) = \mathrm{clamp}_{[0,1]}\left\{\frac{\max_{1\le j\le k} g_j^{\mathrm{LS}} + \min_{k\le j<D} g_j^{\mathrm{LS}}}{2}\right\}, \quad 1\le k<D. \tag{8}$$

Set the endpoints to zero and one and use the step extension (6). Both envelopes are nondecreasing and stay within the raw vector's maximum error from any monotone grid truth. Corollary 3.3 gives uniform risk control for this LS estimator. For inference, we retain the raw cycle scores and impose order on band endpoints.

2.4 RAW POINTWISE INTERVALS

Fix a public target $x\in(0,1)$ and allow $D_n = 2^{J_n}$ and ε_n to vary, with $m_n = \lfloor n/J_n\rfloor$. Put $k_n = \lfloor D_n x\rfloor$, $x_n = k_n/D_n$ and $\ell_n(x) = |\mathcal{D}(k_n)|$. The prefix complexity $\ell_n(x)$ is the number of selected nodes, equivalently the number of hierarchy levels used by the prefix. For pointwise inference, take a nonempty prefix, $1\le k_n<D_n$, and at least two cycles. For $\nu\in\mathcal{D}(k_n)$, let $\pi_{n,\nu}=P(X\in I_\nu)$ and $\theta_{n,\nu}=q_{\varepsilon_n}+a_{\varepsilon_n}\pi_{n,\nu}$, and define

$$A_n(x) = \sum_{\nu\in\mathcal{D}(k_n)} \theta_{n,\nu}\{1-\theta_{n,\nu}\}, \qquad \sigma_n^2(x) = \frac{A_n(x)}{a_{\varepsilon_n}^2}. \tag{9}$$

Independent cycles estimate this variance without estimating each cell mass:

$$\widehat{\sigma}_n^2(x) = \frac{1}{m_n-1}\sum_{b=1}^{m_n}\left\{V_{b,k_n} - \widetilde{F}_{m_n,D_n}(x_n)\right\}^2, \qquad m_n\ge 2.$$

For $0<\alpha<1$ and a standard-normal quantile $z_{1-\alpha/2}$, the raw interval is

$$\mathcal{I}_{n,\alpha}^{\mathrm{raw}}(x) = \left[\widetilde{F}_{m_n,D_n}(x_n) \pm z_{1-\alpha/2}\frac{\widehat{\sigma}_n(x)}{\sqrt{m_n}}\right]. \tag{10}$$

The raw estimate is unbiased for $F(x_n)$. Thus this interval targets $F(x_n)$ under the studentization conditions below and also targets $F(x)$ when the grid bias is negligible relative to the standard error. If a public bound $0\le F(x)-F(x_n)\le b_n(x)$ is available, the bias-aware interval accounts for the direction of this bias:

$$\mathcal{I}_{n,\alpha}^{\mathrm{bias}}(x) = \left[\widetilde{F}_{m_n,D_n}(x_n) - z_{1-\alpha/2}\frac{\widehat{\sigma}_n(x)}{\sqrt{m_n}},\ \widetilde{F}_{m_n,D_n}(x_n) + z_{1-\alpha/2}\frac{\widehat{\sigma}_n(x)}{\sqrt{m_n}} + b_n(x)\right]. \tag{11}$$

Under (2), one may take $b_n(x) = L(x-x_n)^s \le LD_n^{-s}$; in particular, $b_n(x)=0$ at a grid target. Theorem 3.6 and Corollary B.3 give the conditions for these Gaussian intervals. The monotone CDF $\widehat{F}$ serves uniform estimation, while the intervals retain the raw prefix as their center. All variance estimates and intervals are post-processing and consume no additional privacy budget.

2.5 ORDERED SIMULTANEOUS BANDS

A confidence band can enforce the CDF order on its endpoints while retaining simultaneous coverage. For a public failure probability $0 < \beta < 1$, use each prefix's own node count to set

$$h_k = \frac{1}{a_\varepsilon}\sqrt{\frac{|\mathcal{D}(k)|}{2m}\log\frac{2(D-1)}{\beta}}, \qquad 1 \le k < D. \tag{12}$$

Writing $g_k = \widetilde{F}_{m,D}(x_k)$, form the ordered endpoints

$$L_k = \max\{0, \max_{1\le j\le k}(g_j - h_j)\}, \qquad U_k = \min\{1, \min_{k\le j<D}(g_j + h_j)\}. \tag{13}$$

If any $L_k > U_k$, use $[0,1]$ at all interior grid points. Set $L_0 = 0$ and $L_D = U_D = 1$. A monotone point estimate inside the band is $\overline{F}(x_k) = \min\{U_k, \max\{L_k, \widehat{F}(x_k)\}\}$, with endpoints zero and one and the step extension in (6). Use $[L_k, U_k]$ at positive grid points and $[L_k, U_{k+1}]$ between consecutive grid points; at zero use $[0, U_1]$. This extension uses monotonicity alone and accommodates atoms.

For covariance calibration with $m \ge 2$, let $\widehat{\sigma}_k^2$ be the cycle sample variance at x_k. Draw independent standard-normal multipliers $\xi_1, \dots, \xi_m$, using the same multiplier for every prefix within a cycle, and compute

$$T^* = \max_{1\le k<D}\frac{|\sum_{b=1}^m \xi_b(V_{b,k} - g_k)|}{\sqrt{m-1}\,\widehat{\sigma}_k}, \qquad h_k^* = c_{1-\beta}^*\widehat{\sigma}_k/\sqrt{m}. \tag{14}$$

Here $c_{1-\beta}^*$ is the conditional $(1-\beta)$-quantile of T^* given the reports. Substitute h_k^* for h_k in (13), using $[0,1]$ if any sample variance vanishes. Proposition 3.8 gives asymptotic coverage for this choice; (12) retains finite-sample coverage.

3 ESTIMATION AND INFERENCE GUARANTEES

3.1 CONTINUOUS UNIFORM RISK AND ITS MINIMAX SCALE

The finite-sample bound separates the noise in the private reports from the exact error of representing a continuous CDF on the chosen grid.

Theorem 3.1 (Finite-sample Kolmogorov bound)**.** *Let $D = 2^J$, $m = \lfloor n/J \rfloor \ge 1$, and $0 < \beta < 1$. Algorithm 1 is noninteractive and user-level ε-LDP. For every distribution F on $[0,1]$, with probability at least $1-\beta$,*

$$\left\|\widehat{F}_{m,D} - F\right\|_\infty \le \frac{1}{a_\varepsilon}\sqrt{\frac{J}{2m}\log\frac{2(D-1)}{\beta}} + \omega_D(F). \tag{15}$$

In addition,

$$\mathbb{E}_F\left\|\widehat{F}_{m,D} - F\right\|_\infty \le \frac{1}{a_\varepsilon}\sqrt{\frac{J}{2m}\log\{2(D-1)\}} + \omega_D(F). \tag{16}$$

Consequently, uniformly over $F \in \mathcal{F}(s,L)$, one may replace $\omega_D(F)$ in both displays by LD^{-s}.

For fixed failure probability and $J = o(n)$, the stochastic term is $J^{3/2}/(\rho_\varepsilon\sqrt{n})$. Its three factors of $\sqrt{J}$ come from the node sum, allocation of about n/J users per level, and simultaneous control of $D-1$ prefixes. Cumulative maxima do not enlarge grid error, and the step extension adds at most $\omega_D(F)$; Appendix D gives the proof.

Corollary 3.2. *Fix $s \in (0,1]$ and $L < \infty$, and allow $\varepsilon = \varepsilon_n$ to vary. Define $r_n = \rho_{\varepsilon_n}\sqrt{n}/\{\log(en)\}^{3/2}$ and suppose $r_n \to \infty$. Let D_n^{bal} be the smallest power of two satisfying*

$$D_n^{\mathrm{bal}} \ge \max\left[2, \{Lr_n\}^{1/s}\right], \qquad J_n = \log_2 D_n^{\mathrm{bal}}, \qquad m_n = \lfloor n/J_n \rfloor. \tag{17}$$

Then, uniformly over $F \in \mathcal{F}(s,L)$,

$$\left\|\widehat{F}_{m_n,D_n^{\mathrm{bal}}} - F\right\|_\infty = O_{\mathbb{P}}(r_n^{-1}), \qquad \sup_{F\in\mathcal{F}(s,L)}\mathbb{E}_F\left\|\widehat{F}_{m_n,D_n^{\mathrm{bal}}} - F\right\|_\infty = O(r_n^{-1}).$$

For fixed $\varepsilon > 0$, the rate is $(\log n)^{3/2}/\sqrt{n}$. The more familiar choice $D_n \asymp n^{1/(2s)}$ has the same stochastic order but generally uses a finer grid than is needed for uniform estimation.

For fixed ε, s, L, the balanced grid has $D_n^{\mathrm{bal}} \asymp n^{1/(2s)}/(\log n)^{3/(2s)}$. It reduces average communication by a factor $(\log n)^{3/(2s)}$ relative to $D_n \asymp n^{1/(2s)}$, with the same uniform risk order. For laws supported on the positive public grid, $\omega_D(F) = 0$ and continuous loss equals maximum grid error (Corollary B.2, Appendix B.2).

Corollary 3.3. *For $D = 2^J$, $m = \lfloor n/J \rfloor \geq 1$ and $0 < \beta < 1$, the OUE-based LS estimator* (8) *satisfies, for every distribution F, with probability at least $1 - \beta$,*

$$\|\widehat{F}^{\mathrm{LS}}_{m,D} - F\|_\infty \leq \sqrt{\left(\frac{1}{2a_\varepsilon^2} + 1\right)\frac{J}{m}\log\frac{2(D-1)}{\beta}} + \omega_D(F). \tag{18}$$

The expected loss obeys the same bound with $\log\{2(D-1)\}$ in place of $\log\{2(D-1)/\beta\}$. Under the assumptions and resolution of Corollary 3.2, both uniform rate conclusions hold for $\widehat{F}^{\mathrm{LS}}$ as well. The public resolution used in Section 4 also attains this rate on $\mathcal{F}(1, 3/2)$.

Constrained LS has threshold coefficient norm no larger than canonical decoding. Combining this fact with concentration of weighted OUE reports and the envelope's error contraction proves the result (Appendix D.1). The bound supports growing-grid point estimation; Proposition 3.9 gives the separate fixed-grid band guarantee.

The following lower bound uses the same continuous loss on $\mathcal{F}_{\mathrm{bd}}$ and allows every sequentially interactive reporting scheme.

Theorem 3.4 (Interactive lower bound for uniform loss)**.** *There are universal constants $c, c_0, N_0 > 0$ such that the following holds. Let $0 < \varepsilon \leq 1$ and put $N_\varepsilon = n\rho_\varepsilon^2$. If $N_\varepsilon \geq N_0$, then*

$$\inf_{\mathcal{M},\widehat{F}} \sup_{F\in\mathcal{F}_{\mathrm{bd}}} \mathbb{P}_F\left\{\left\|\widehat{F} - F\right\|_\infty \geq c\sqrt{\frac{\log N_\varepsilon}{N_\varepsilon}}\right\} \geq c_0, \tag{19}$$

where the infimum ranges over all sequentially interactive ε-LDP protocols and all transcript-measurable function-valued estimators.

Localized triangular CDF deviations of height δ fit in order $1/\delta$ disjoint locations. Distinguishing their locations contributes the logarithm. The proof combines this packing with the sequential privacy contraction of Duchi et al. (2013; 2018) and Fano's inequality (Appendix G).

Corollary 3.5. *Let $R_{n,\varepsilon} = \inf_{\mathcal{M},\widehat{F}} \sup_{F\in\mathcal{F}_{\mathrm{bd}}} \mathbb{E}_F\|\widehat{F} - F\|_\infty$. For $0 < \varepsilon \leq 1$ and $n\rho_\varepsilon^2$ sufficiently large,*

$$c\frac{\sqrt{\log(n\rho_\varepsilon^2)}}{\rho_\varepsilon\sqrt{n}} \leq R_{n,\varepsilon} \leq C\frac{\{\log(en)\}^{3/2}}{\rho_\varepsilon\sqrt{n}}. \tag{20}$$

The bounds share the root-n exponent and privacy scale, with one remaining factor of $\log n$ at fixed privacy. Prefix factorization also yields the upper rate (Edmonds et al., 2020); Appendix B.1 gives the continuous-loss comparison. Here the explicit cycles provide variance estimates and shape-valid inference, while Corollary 3.3 extends the rate to LS reconstruction. The lower bound covers arbitrary messages and sequential interaction.

3.2 Pointwise studentization

For a single target, independent decoded Bernoulli coordinates give an exact variance. Their total Bernoulli variance controls the normal approximation even when the privacy parameter varies.

Theorem 3.6 (Pointwise studentization)**.** *Fix $x \in (0, 1)$, and suppose that $1 \leq k_n < D_n$ for all sufficiently large n.*

(a) *The raw estimator is exactly unbiased for the left-grid target and has exact variance*

$$\mathbb{E}\widetilde{F}_{m_n,D_n}(x_n) = F(x_n), \qquad \mathrm{Var}\{\widetilde{F}_{m_n,D_n}(x_n)\} = \frac{\sigma_n^2(x)}{m_n}. \tag{21}$$

(b) *For every n with $A_n(x) > 0$,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{m_n}\{\widetilde{F}_{m_n,D_n}(x_n) - F(x_n)\}}{\sigma_n(x)} \leq t\right\} - \Phi(t)\right| \leq \frac{C_{\mathrm{BE}}}{\sqrt{m_n A_n(x)}}, \tag{22}$$

where Φ is the standard-normal CDF and C_{BE} is a universal Berry–Esseen constant.

(c) *The conditions*

$$m_n \to \infty, \qquad m_n A_n(x) \to \infty \tag{23}$$

imply

$$\widehat{\sigma}_n^2(x)/\sigma_n^2(x) \xrightarrow{\mathbb{P}} 1, \tag{24}$$

$$\sqrt{m_n}\{\widetilde{F}_{m_n,D_n}(x_n) - F(x_n)\}/\widehat{\sigma}_n(x) \rightsquigarrow N(0,1). \tag{25}$$

(d) *In addition, the following implication holds:*

$$\sqrt{m_n}\{F(x) - F(x_n)\}/\sigma_n(x) \longrightarrow 0 \tag{26}$$

$$\implies \quad \sqrt{m_n}\{\widetilde{F}_{m_n,D_n}(x_n) - F(x)\}/\widehat{\sigma}_n(x) \rightsquigarrow N(0,1). \tag{27}$$

The condition $m_n A_n(x) \to \infty$ controls normal approximation; $m_n \to \infty$ supplies independent cycles for variance estimation. Expanding (9) gives

$$\sigma_n^2(x) = \frac{\ell_n(x) q_{\varepsilon_n}(1 - q_{\varepsilon_n})}{a_{\varepsilon_n}^2} + 2F(x_n) - \sum_{\nu \in \mathcal{D}(k_n)} \pi_{n,\nu}^2. \tag{28}$$

Each selected node contributes background privacy noise. Consequently, at fixed ε and $J_n = o(n)$, a target using finitely many prefix nodes has standard error of order $\sqrt{J_n/n}$, whereas a prefix using order J_n nodes has standard error $J_n/\sqrt{n}$. A dedicated RR query has order $n^{-1/2}$ (Appendix B.2); reusability is purchased by dividing users across levels.

Under (23), (10) covers the grid target asymptotically. It covers $F(x)$ when (26) also holds; alternatively, (11) has limiting coverage at least $1 - \alpha$ with the stated public bias bound. Only the upper endpoint is enlarged because $\overline{F}(x) \geq F(x_n)$. For $F \in \mathcal{F}(s, L)$ and bounded ε_n, the resolution $D_n \asymp n^{1/(2s)}$ makes grid bias negligible on the pointwise standard-error scale (Corollary B.4). This gives a public resolution choice for intervals beyond the grid. Complete interval statements appear in Appendix B.3; Appendix E proves the theorem.

3.3 Shape enforcement and consistent reconstruction

Ordering band endpoints preserves the joint coverage event while producing a monotone curve inside the band. This supplies a shape-valid release without moving a pointwise interval to a new center.

Corollary 3.7. *For every distribution F on $[0,1]$, the band with radii* (12) *covers $F(x)$ simultaneously for all $x \in [0,1]$ with probability at least $1-\beta$. At positive grid points, its endpoints are nondecreasing and contain $\overline{F}$. Whenever the original intervals are compatible, $U_k - L_k \leq 2h_k$.*

The continuous extension uses only the order of the true CDF, so it needs no density assumption or interpolation-bias estimate. Simultaneous coverage allows choosing a threshold after viewing the release. Covariance calibration further uses shared-node dependence and each prefix's own variance.

Proposition 3.8. *Fix $0 < \beta < 1$ and a finite privacy upper bound $\varepsilon_{\max}$. Suppose $0 < \varepsilon_n \leq \varepsilon_{\max}$ and $\{\log(eD_n m_n)\}^7/m_n \to 0$. Uniformly over all distributions on $[0,1]$, the ordered band with radii* (14) *covers F on $[0,1]$ with probability at least $1 - \beta - o(1)$. The same conclusion holds for the levelwise Hadamard reports in Appendix A.3.*

Polynomial resolutions satisfy the stated growth condition. Appendix D.2 proves the result by high-dimensional Gaussian approximation. The Hoeffding radii retain their finite-sample guarantee; Gaussian calibration is asymptotic.

For LS inference, use the reconstructed scores (7) with their sample means and variances.

Proposition 3.9. *Fix $D = 2^J \geq 2$, $\varepsilon > 0$ and $0 < \beta < 1$. Under either the OUE or levelwise HR channel, the scores in* (7) *are unbiased for $F(x_k)$ and have strictly positive variances for every interior prefix. Apply the Gaussian multiplier calibration* (14) *to these LS scores, then order the endpoints by* (13) *and extend as in Section 2.5. As $m \to \infty$, this band covers F on $[0,1]$ with probability at least $1 - \beta - o(1)$ for every distribution F on $[0,1]$.*

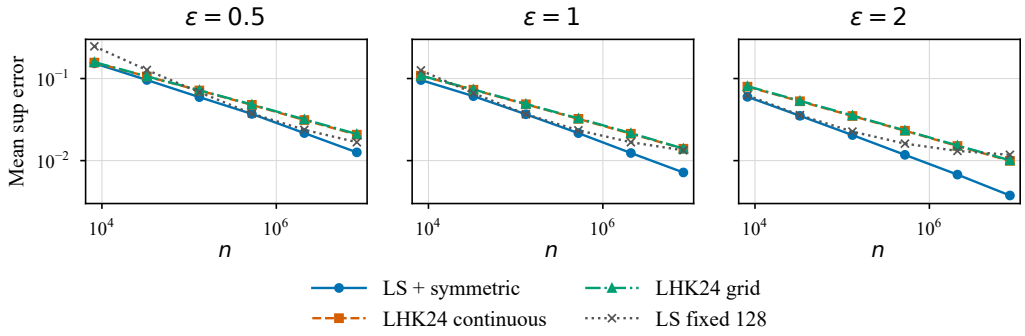

Figure 1: Continuous uniform error versus sample size. Curves average the four law-specific mean errors equally; both axes are logarithmic. "LS + symmetric" is TreeOUE–LS at $D = 4D_0$ ((8)); fixed-128 uses LS reconstruction with cumulative maxima. Figure 2 separates the laws.

Reconstructing every cycle preserves independent repetitions of the whole estimated curve. Because the map is affine, averaging the reconstructed leaf masses gives the same point estimate as reconstructing the average node report. Keeping the cycle vectors also supplies the covariance needed for inference: a prefix now combines information from parent and child reports that canonical decoding leaves unused. Their dependence enters the estimated covariance automatically. Signed intermediate masses preserve unbiasedness, while the ordered endpoints produce the final monotone band and CDF. The proof in Appendix D.3 uses affine unbiasedness, positive variances and a fixed-dimensional Gaussian limit. This guarantee holds for fixed D, ε; the growing-grid guarantee in Proposition 3.8 concerns canonical scores. The reconstruction experiment below evaluates precisely the fixed-grid setting.

The distinction between endpoints and centers is essential: a small uniform error does not make a monotone correction negligible on every pointwise scale. Under the local gap and endpoint separation conditions in (45), the correction is inactive with probability tending to one (Proposition B.5). Proposition F.1 in Appendix F exhibits the opposite regime for a uniform law and a grid of order $\sqrt{m_n J_n}$.

At $1/2$ the canonical prefix uses one node; its predecessor uses $J_n - 1$ nodes and disjoint users. The predecessor's standard deviation is larger by order $\sqrt{J_n}$, while the true gap is only $1/D_n$. This explains why intervals keep the raw center. Ordering the endpoints of a simultaneous band instead preserves the coverage event, consistent with shape-restricted band constructions (Chernozhukov et al., 2009). Pointwise isotonic limits also depend on local shape (Groeneboom & Wellner, 1992). Appendix F proves the pointwise claims.

4 Numerical results

We compare mean continuous sup error with Liu et al. (2024) (LHK24) at common (n, ε) and native message lengths, and compare bands on identical hierarchical reports. TreeOUE–LS combines TreeOUE reports, LS reconstruction and the symmetric envelope; in the band comparison, LS denotes reconstruction alone. Figure 3 illustrates the LS band of Proposition 3.9.

The four densities on $[0, 1]$ are 1, $0.5 + x$, $1.5 - x$ and $1 + 0.5\cos(4\pi x)$, all bounded by $3/2$. We use $D = 4D_0$, where D_0 minimizes (16) with approximation term $3/(2D)$ over public dyadic grids. The multiplier and symmetric envelope were selected under the first two laws and frozen before independent validation; the other laws were held out. Continuous-threshold LHK24 and its independently selected grid version use the published isotonic fit and interpolation (Appendix A.2).

Validation uses $\varepsilon \in \{0.5, 1, 2\}$ and six sample sizes $n \in \{2^{13}, 2^{15}, \ldots, 2^{23}\}$, with 256 replications per condition through $n = 2^{19}$ and 128 at larger sizes (15360 total). Exact aggregate-output simulation generates estimates; both ends of each constant segment determine the supremum.

The error separation grows with n (Figure 1). Table 1 reports reductions relative to continuous-threshold LHK24 and log-ratio delta-method intervals. At the largest size, every lower endpoint exceeds 45%, including both held-out laws. Appendix A.2 gives the calculation and LHK24 grid sensitivity.

Table 1: Mean continuous error reduction (%) of TreeOUE–LS over continuous-threshold LHK24, $\varepsilon = 1$. Brackets: approximate marginal 95% Monte Carlo intervals (256, 256, 128 replications by column). Reverse tilt and wave are held-out laws.

Law	$n = 2^{13}$	$n = 2^{17}$	$n = 2^{23}$
Uniform	11.6 [8.5, 14.6]	22.7 [20.7, 24.6]	46.9 [45.4, 48.4]
Tilted	10.6 [7.6, 13.5]	23.9 [21.8, 25.9]	49.5 [48.1, 50.8]
Reverse tilt	11.6 [8.6, 14.5]	24.6 [22.6, 26.5]	47.6 [45.9, 49.2]
Wave	12.4 [9.4, 15.3]	26.6 [24.6, 28.6]	50.3 [49.0, 51.6]

At $\varepsilon = 1$, mean LS message length grows from 21 to about 205 bits across these sizes; Appendix A.2 reports same-grid ablations.

For inference, $n = 2^{13}$, $D = 128$ and $\varepsilon = 1$, with 600 replications per law and 1999 Gaussian calibration draws. The laws are uniform, $\mathrm{Beta}(2, 1)$ and $0.7\,\mathrm{Unif}(0, 1) + 0.3\delta_{1/2}$. Decoders share reports within each channel. Table 2 averages width over all 127 interior grid points and requires simultaneous coverage; the ordered extension includes the grid atom.

Table 2: Nominal 95% bands from identical reports. Columns C/LS are canonical/least-squares reconstructions (LS: (7)); OUE/HR are TreeOUE/Hadamard reporting channels. Width is in CDF units; Table 7 gives midpoint widths and fallback frequencies.

		Mean grid width		Coverage (%)	
Law	Channel	C	LS	C	LS
Uniform	OUE	0.448	0.261	95.2	96.3
Uniform	HR	0.465	0.277	95.3	94.2
$\mathrm{Beta}(2, 1)$	OUE	0.418	0.243	95.5	95.2
$\mathrm{Beta}(2, 1)$	HR	0.441	0.253	95.2	96.8
Mixture	OUE	0.413	0.244	94.8	95.2
Mixture	HR	0.431	0.257	95.5	96.2

Within each channel, LS reconstruction reduces mean width by 40%–43%, with coverage 94.2%–96.8% including fallback bands. Recentring at the cumulative maximum lowers canonical midpoint coverage under uniformity from 96.2% to 63.8%; LS intervals also undercover (Appendices A.3–A.5).

5 Concluding remarks

Hierarchical reuse and independent cycles connect accurate private CDF release with feasible uncertainty quantification. Under public Hölder regularity and fixed privacy, the continuous uniform rate approaches the root-n scale; the interactive lower bound leaves a logarithmic gap on the bounded-density class. Affine LS reconstruction retains this rate while using redundant reports, and the numerical results show gains in estimation error and fixed-grid band width. The distinction between shaping estimates and ordering band endpoints is essential: uniform error contraction alone does not justify pointwise Gaussian inference. The construction supports thresholds chosen after collection while retaining joint sampling information for simultaneous coverage.

Two extensions are natural. Growing-grid LS bands call for a uniform Gaussian approximation: the coefficient bounds in Appendix D.1 provide a starting point for controlling standardized tails and sample correlations across prefixes, followed by anti-concentration for the studentized maximum (Chernozhukov et al., 2017). Enforcing nonnegative leaf masses is another extension. At zero cell masses, active constraints make the fit nonlinear; its directional derivative and a resampling calibration for the possibly non-Gaussian limit would replace the present affine argument (Fang & Santos, 2019).

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

Algorithm 1 specifies reporting, point estimation and inference. Appendix A gives the public reconstruction rules, simulation designs, aggregate-output sampling and evaluation metrics. Appendix B contains further inference results, followed by complete proofs, including the LS uniform bound. 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

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1120–1129, 2019.

Raef Bassily. Linear queries estimation with local differential privacy. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 721–729, 2019.

Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009. doi: 10.1093/biomet/asp030.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352, 2017. doi: 10.1214/16-AOP1113.

Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. Answering range queries under local differential privacy. *Proceedings of the VLDB Endowment*, 12(10):1126–1138, 2019. doi: 10.14778/3339490.3339496.

Linkang Du, Zhikun Zhang, Shaojie Bai, Changchang Liu, Shouling Ji, Peng Cheng, and Jiming Chen. AHEAD: Adaptive hierarchical decomposition for range query under local differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1266–1288, 2021. doi: 10.1145/3460120.3485668.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *Proceedings of the 54th IEEE Symposium on Foundations of Computer Science*, pp. 429–438, 2013. doi: 10.1109/FOCS.2013.53.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018. doi: 10.1080/01621459.2017.1389735.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956. doi: 10.1214/aoms/1177728174.

Alexander Edmonds, Aleksandar Nikolov, and Jonathan Ullman. The power of factorization mechanisms in local and central differential privacy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 425–438, 2020. doi: 10.1145/3357713.3384297.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1054–1067, 2014. doi: 10.1145/2660267.2660348.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412, 2019. doi: 10.1093/restud/rdy049.

Piet Groeneboom and Jon A. Wellner. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, 1992. doi: 10.1007/978-3-0348-8621-5.

Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the accuracy of differentially private histograms through consistency. *Proceedings of the VLDB Endowment*, 3(1):1021–1032, 2010. URL https://vldb.org/pvldb/vol3/R91.pdf.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

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

Zitao Li, Tianhao Wang, Milan Lopuhaä-Zwakenberg, Ninghui Li, and Boris Škorić. Estimating numerical distributions under local differential privacy. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pp. 621–635, 2020. doi: 10.1145/3318464.3389700.

Yi Liu, Qirui Hu, and Linglong Kong. Tuning-free estimation and inference of cumulative distribution function under local differential privacy. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 31147–31164, 2024.

Pascal Massart. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18 (3):1269–1283, 1990. doi: 10.1214/aop/1176990746.

Valentin V. Petrov. *Sums of Independent Random Variables*, volume 82 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, 1975. doi: 10.1007/978-3-642-65809-9.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009. doi: 10.1007/b13794.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, 1996. doi: 10.1007/978-1-4757-2545-2.

Leixia Wang, Qingqing Ye, Haibo Hu, and Xiaofeng Meng. PriPL-Tree: Accurate range query for arbitrary distribution under local differential privacy. *Proceedings of the VLDB Endowment*, 17(11):3031–3044, 2024. doi: 10.14778/3681954.3681981.

Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium*, pp. 729–745. USENIX Association, 2017.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. doi: 10.1080/01621459.1965.10480775.

Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. Differential privacy via wavelet transforms. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1200–1214, 2011. doi: 10.1109/TKDE.2010.247.

Puning Zhao, Zhikun Zhang, Bo Sun, Li Shen, Liang Zhang, Shaowei Wang, and Zhe Liu. Consistent estimation of numerical distributions under local differential privacy by wavelet expansion. In *2026 IEEE Symposium on Security and Privacy*, pp. 3243–3261, 2026. doi: 10.1109/SP63933.2026.00054.

A Algorithms and experimental details

A.1 Reporting, calibration and evaluation

Algorithm 1 Balanced TreeOUE release, pointwise intervals and simultaneous band

Require: Public n, $\varepsilon > 0$, $D = 2^J$ with $1 \le J \le n$, optional targets $\mathcal{X}_0$, levels $1-\alpha$, $1-\beta$ and public bias bounds $b_n(x)$.

1: Set $m = \lfloor n/J \rfloor$; assign mJ users to m cycles, one user per level.
2: Each level-r user locates $C_r(X_i)$ and releases the full OUE vector (3).
3: Decode the reports by (4); sum the canonical nodes to obtain $V_{b,k}$ and average over cycles for every $1 \le k < D$.
4: Apply cumulative maxima (5) and step extension (6) to release $\widehat{F}$.
5: For LS point estimation at the public resolution D, reconstruct each cycle by (7) and average its prefix scores over cycles.
6: Apply the symmetric envelope (8), fix the endpoints to zero and one, and use (6) to release $\widehat{F}^{\mathrm{LS}}_{m,D}$.
7: **for** each $x \in \mathcal{X}_0$ **do**
8: Require $m \ge 2$ and $1 \le \lfloor Dx \rfloor < D$; compute the cycle sample variance.
9: Report (10) for $F(\lfloor Dx \rfloor / D)$, or for $F(x)$ with negligible grid bias. With a public $b_n(x)$, report (11) for $F(x)$.
10: Apply (13) with radii (12) or (14); form $\overline{F}$ within the band.
11: **return** The canonical and LS CDF estimates, the band, and requested raw estimates, standard errors and pointwise intervals.

For fixed LS bands, replace each canonical cycle score in the final band step by (7), compute its sample covariance, and apply the same multiplier and ordering rules. The public equal weights and total-mass constraint give the affine reconstruction used in Proposition 3.9.

All experiments use right-closed cells: a value exactly at $j2^{-r}$ belongs to cell j, and zero belongs to the first cell. The inference experiments generate every OUE coordinate, including the false-cell bits. User assignments and the tree are public. The inference experiments use $n = 2^{13}$, $D = 128$, $\varepsilon = 1$, $m = 1170$ cycles and 8190 used observations. The comparison designs use the uniform law, $\mathrm{Beta}(2,1)$ and $0.7\,\mathrm{Unif}(0,1) + 0.3\delta_{1/2}$. The true CDF is used only for evaluation.

For each of the channel, calibration and reconstruction comparisons, there are 600 independent replications per law. Each comparison uses an independent simulation stream. Within a comparison, the two channels share observations and public assignments but use independent randomization. Within a channel, alternative calibrations or reconstructions share all reports.

At the target $1/2$, pointwise coverage evaluates the population CDF value and interval length is the upper minus lower endpoint. Simultaneous grid coverage requires all 127 interior CDF values to be covered. Mean grid width averages the ordered width over these same points, including runs using the $[0,1]$ band. The band extension uses $[L_k, U_{k+1}]$ between grid points, so simultaneous grid coverage implies continuous coverage by monotonicity, also for the atomic law. For point estimation, continuous Kolmogorov error is evaluated exactly from both sides of each cell:

Let G be right-continuous and equal to g_k on $[x_k, x_{k+1})$. For a continuous CDF F, monotonicity of F gives

$$\|G - F\|_\infty = \max\left\{ \max_{0 \le k < D} \max\left(|g_k - F(x_k)|, |g_k - F(x_{k+1}-)| \right), |G(1) - F(1)| \right\}. \quad (29)$$

The formula also handles the jump at $1/2$, since the corresponding grid boundary is evaluated with its right value and its left limit.

A.2 Continuous-error comparison and public reconstruction choices

The four simulation CDFs are $F(x) = x$, $x/2 + x^2/2$, $3x/2 - x^2/2$ and $x + \sin(4\pi x)/(8\pi)$ on $[0,1]$, called uniform, tilted, reverse tilt and wave. Their densities lie in $[1/2, 3/2]$. The true CDF is used to generate observations and evaluate loss; the reconstruction uses the public density bound $3/2$.

The initial resolution minimizes the expected canonical upper bound with this density bound. For the sample sizes used here, define

$$\mathcal{G}_n = \{2^j : 1 \le j \le \lceil \tfrac{1}{2} \log_2 n \rceil\}, \qquad D_0 = \operatorname*{arg\,min}_{D \in \mathcal{G}_n} B_{n,\varepsilon}(D),$$
$$B_{n,\varepsilon}(D) = \frac{1}{a_\varepsilon} \sqrt{\frac{\log_2 D}{2 \lfloor n / \log_2 D \rfloor} \log\{2(D-1)\}} + \frac{3}{2D}. \tag{30}$$

Ties use the smaller grid. Development compares the dyadic multipliers $c \in \{1/2, 1, 2, 4, 8\}$ through $D = \max\{2, cD_0\}$, for canonical and LS decoding. Each is followed by cumulative maxima, unweighted isotonic regression on the interior grid, or the symmetric envelope in (8). Clipping to $[0, 1]$ and fixing the endpoints are common to all three shape rules. Both decoders and all shape rules at a common D share the same reports.

Development uses uniform and tilted laws, $n \in \{2^{13}, 2^{15}, 2^{17}, 2^{19}\}$ and $\varepsilon \in \{0.5, 1, 2\}$, with 96 replications in each of the 24 conditions. Each rule is scored by the geometric mean, over these conditions, of its mean error divided by the continuous-threshold LHK24 mean error. Among rules within 2% of the smallest score, smaller average message length, then smaller resolution, then smaller score determine the selection. This chooses the symmetric envelope with $c = 2$ for canonical decoding and $c = 4$ for LS. These two rules remain fixed throughout validation.

For the LHK24 continuous comparator, public thresholds are independent uniform draws on $[0, 1]$. A response at threshold t is Bernoulli with probability $q_\varepsilon + \rho_\varepsilon F(t)$. We sort thresholds, fit isotonic response probabilities, transform by $(\cdot - q_\varepsilon)/\rho_\varepsilon$, clip to $[0, 1]$, and extend by the nearest fitted threshold to the left. The endpoint values are zero and one. This implements Algorithm 1 and the interpolation used in Section 3.3 of Liu et al. (2024). Their preselected-threshold comparator assigns each user uniformly to j/K, $j = 1, \ldots, K$, and uses count-weighted isotonic regression, omitting empty thresholds. We develop $K = \max\{2, \lceil c(n\rho_\varepsilon^2)^{1/3} \rceil\}$ with the same five candidate multipliers, laws, sample sizes and scoring criterion. All such reports have one bit, so the resolution breaks a score tie. The selected multiplier is $c = 8$, the largest value in the specified candidate set. Each LHK24 procedure is retained separately in validation.

Validation uses independent random streams for all four laws, three privacy levels and $n \in \{2^{13}, 2^{15}, 2^{17}, 2^{19}, 2^{21}, 2^{23}\}$. There are 256 replications in each condition through 2^{19} users and 128 at the two larger sizes. Every replication retains eight estimators: selected canonical and LS with their respective symmetric envelopes; each selected decoder with a same-report, same-grid cumulative maximum; canonical and LS cumulative maxima at fixed $D = 128$; and the two LHK24 procedures. The selected canonical and LS estimators use different resolutions, so their comparison includes both reconstruction and grid choice. No choice depends on the validation outcomes. Figure 2 and Table 5 include every condition.

Table 5 includes the small-sample comparisons; same-report ablations isolate the effect of the shape correction. For the uniform law at $\varepsilon = 0.5$, $n = 2^{13}$, the observed mean errors are 0.156 for LS with the symmetric envelope and 0.152 for continuous LHK24. Across all conditions, replacing cumulative maxima by the symmetric envelope on identical LS reports at a common resolution changes mean error from a 0.4% increase to a 13% reduction.

The aggregate simulation preserves the exact point-estimator output distribution. For each level with K cells and $m = \lfloor n / \log_2 D \rfloor$ assigned users, draw the category counts $(N_1, \ldots, N_K)$ from the multinomial law with probabilities $F(j/K) - F((j-1)/K)$. Conditional on these counts, the reported bit totals are independent across coordinates and have distributions

$$T_j \overset{d}{=} \mathrm{Bin}(N_j, 1/2) + \mathrm{Bin}(m - N_j, q_\varepsilon), \qquad y_j = (T_j/m - q_\varepsilon)/a_\varepsilon,$$

with the two binomial terms independent. Conditional coordinate independence follows from the OUE channel: conditional on the entire category assignment, coordinates use independent random bits, and the distribution of each total depends only on the counts. Independent levels therefore give the joint law of all node averages. Affine LS reconstruction commutes with averaging. For continuous LHK24 thresholds, normalized partial sums of $n + 1$ independent exponential variables generate the exact uniform order statistics; responses are then generated conditional on those thresholds. For the grid comparator, multinomial threshold counts and conditional binomial successes give its sufficient

statistics exactly. These simulations do not retain cycle covariances for bands or measure the cost of sending the full messages.

For every estimator, continuous error uses (29) at all fitted segment boundaries. Averaging in Figure 1 gives each of the four law-specific mean errors weight $1/4$ at fixed (n, ε). Table 4 records the grids and LS message lengths used in Figures 1–2; these settings are common to all laws and specify the resolution–communication tradeoff behind the comparisons.

Uncertainty in the error reductions. Table 1 quantifies Monte Carlo uncertainty in the reductions highlighted in Section 4. TreeOUE and LHK24 use independent aggregate-output draws. Same-report pairing applies to decoder and shape comparisons at a common tree resolution. For the ratio of the LS and continuous-LHK24 mean errors, we apply the normal delta method to its logarithm, estimating its variance by the sum of the squared relative Monte Carlo standard errors. Exponentiation and subtraction from one give the reported marginal intervals. At the largest sample size, their lower endpoints exceed 45% for all four laws, including the two held-out laws.

Sensitivity to larger LHK24 grids. The selected LHK24 multiplier $c = 8$ lies at the development search boundary. To assess larger grids, we evaluate $c \in \{8, 16, 32, 64\}$ at $\varepsilon = 1$ and $n \in \{2^{17}, 2^{23}\}$ under all four laws, using 128 fresh independent replications per condition. The grid rule and isotonic fit are otherwise unchanged. Table 3 shows that larger multipliers can reduce this comparator's mean error by about 4%, with the best observed multiplier depending on the law and sample size. At the larger size, all tested LHK24 grids have mean error between 0.0132 and 0.0144; the selected LS estimator has mean error between 0.0070 and 0.0074 in the independent validation. The sensitivity results supplement the fixed-rule validation and retain the separate continuous-threshold comparison.

Table 3: Larger-grid sensitivity for preselected-threshold LHK24 at $\varepsilon = 1$. Entries are mean continuous error (Monte Carlo standard error), both in units of 10^{-3}, over 128 fresh replications per cell. The resolution is $K = \max\{2, \lceil c(n\rho_\varepsilon^2)^{1/3} \rceil\}$.

n	Law	$c = 8$	$c = 16$	$c = 32$	$c = 64$
2^{17}	Uniform	47.21 (0.52)	46.08 (0.51)	47.44 (0.51)	47.46 (0.51)
	Tilted	50.17 (0.71)	48.22 (0.48)	48.25 (0.54)	48.88 (0.60)
	Reverse tilt	49.87 (0.58)	49.49 (0.59)	47.78 (0.59)	49.71 (0.67)
	Wave	51.46 (0.60)	49.29 (0.56)	49.55 (0.58)	49.47 (0.58)
2^{23}	Uniform	13.67 (0.11)	13.60 (0.11)	13.29 (0.12)	13.18 (0.11)
	Tilted	14.31 (0.13)	14.17 (0.12)	14.04 (0.14)	14.18 (0.13)
	Reverse tilt	14.31 (0.13)	13.97 (0.12)	13.98 (0.12)	13.78 (0.12)
	Wave	14.35 (0.12)	14.15 (0.13)	14.16 (0.13)	13.95 (0.11)

Table 5 reports the law-specific errors underlying the central LS–LHK24 comparison in Figure 1 and Table 1. It retains every privacy level and sample size, including the small-sample settings where LS does not improve on LHK24. Entries are rounded to three significant digits.

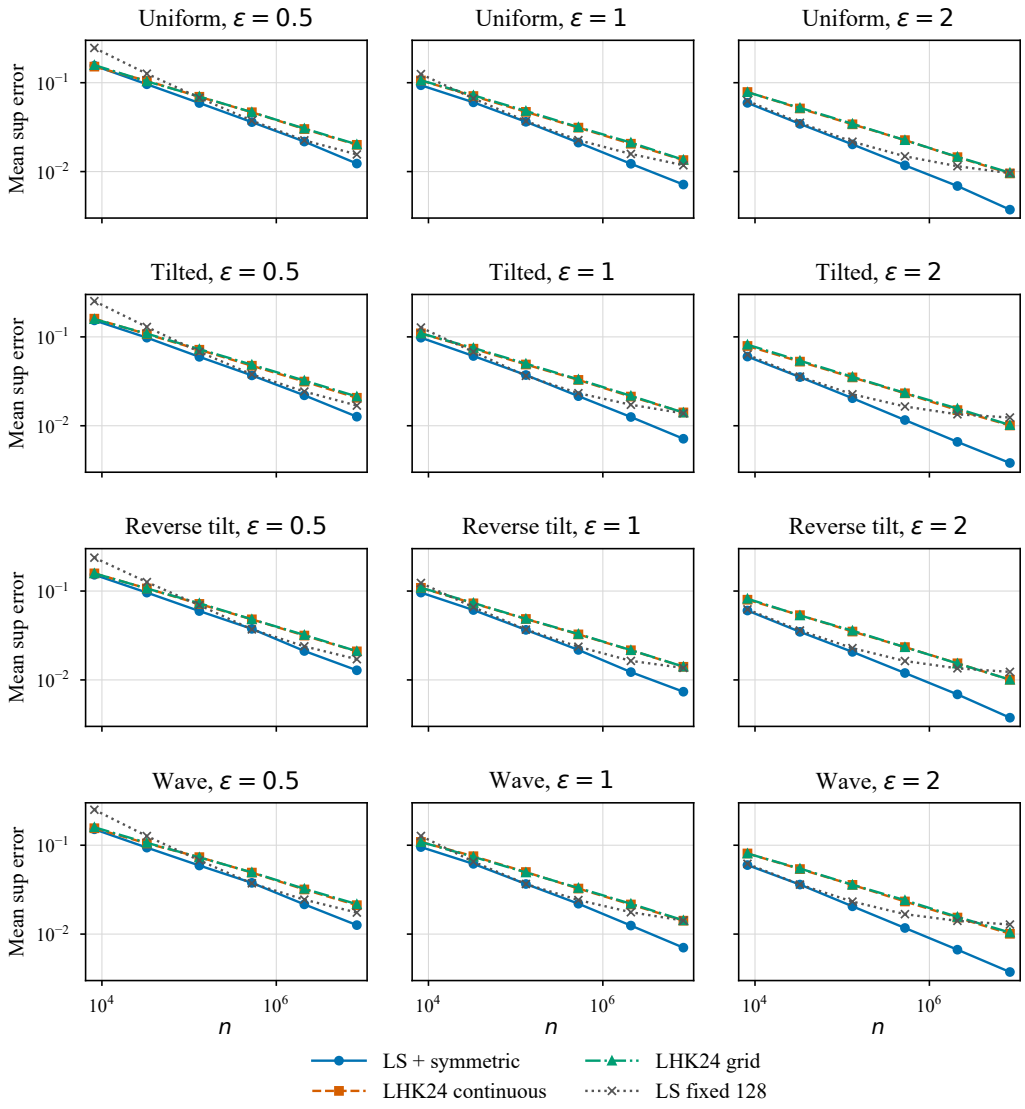

Figure 2: Mean continuous sup error for every validation law and privacy parameter. "LS + symmetric" denotes TreeOUE–LS; the fixed-128 LS curve uses cumulative maxima instead of the symmetric envelope. The density is 1 (uniform), $0.5 + x$ (tilted), $1.5 - x$ (reverse tilt) or $1 + 0.5\cos(4\pi x)$ (wave). Reverse tilt and wave were excluded from development. Each point is the mean over the specified 256 or 128 independent replications. Both axes are logarithmic.

Table 4: Public resolutions in validation. Canonical uses $2D_0$, LS uses $4D_0$, and LHK24 grid uses K. Bits are the average uplink length for selected LS. The tree uses $\lfloor n/\log_2 D \rfloor \log_2 D$ users and the LHK24 estimators use all n users.

ε	n	D_0	Canonical D	LS D	LHK24 K	LS bits
0.5	2^{13}	8	16	32	64	12.4
0.5	2^{15}	16	32	64	101	21.0
0.5	2^{17}	32	64	128	160	36.3
0.5	2^{19}	32	64	128	253	36.3
0.5	2^{21}	64	128	256	401	63.8
0.5	2^{23}	128	256	512	637	113.6
1	2^{13}	16	32	64	97	21.0
1	2^{15}	16	32	64	154	21.0
1	2^{17}	32	64	128	243	36.3
1	2^{19}	64	128	256	386	63.8
1	2^{21}	128	256	512	613	113.6
1	2^{23}	256	512	1024	972	204.6
2	2^{13}	16	32	64	135	21.0
2	2^{15}	32	64	128	214	36.3
2	2^{17}	64	128	256	339	63.8
2	2^{19}	128	256	512	538	113.6
2	2^{21}	256	512	1024	854	204.6
2	2^{23}	512	1024	2048	1356	372.2

Table 5: Mean continuous sup error for selected TreeOUE–LS (LS) and continuous-threshold LHK24 (LHK24$_c$), for all 72 validation conditions. There are 256 replications per condition through $n = 2^{19}$ and 128 at the two larger sizes. Figure 2 also displays preselected-grid LHK24 and fixed-128 LS reconstruction with cumulative maxima.

		Uniform		Tilted		Reverse tilt		Wave	
ε	n	LS	LHK24$_c$	LS	LHK24$_c$	LS	LHK24$_c$	LS	LHK24$_c$
0.5	2^{13}	0.156	0.152	0.153	0.16	0.153	0.158	0.151	0.155
	2^{15}	0.0961	0.105	0.098	0.107	0.096	0.107	0.0939	0.105
	2^{17}	0.0591	0.0695	0.0596	0.0716	0.0596	0.0716	0.0591	0.0734
	2^{19}	0.036	0.0465	0.0369	0.0473	0.0376	0.048	0.0377	0.0492
	2^{21}	0.0218	0.0301	0.022	0.0314	0.0212	0.0316	0.0216	0.0318
	2^{23}	0.0123	0.02	0.0126	0.0206	0.0128	0.0209	0.0126	0.0211
1	2^{13}	0.0937	0.106	0.0979	0.11	0.0961	0.109	0.0953	0.109
	2^{15}	0.0601	0.0708	0.0609	0.0736	0.0611	0.0727	0.0617	0.0752
	2^{17}	0.0363	0.047	0.0372	0.0489	0.0366	0.0485	0.0366	0.0499
	2^{19}	0.0212	0.0312	0.0215	0.0328	0.0217	0.0325	0.022	0.0325
	2^{21}	0.0122	0.0206	0.0125	0.0213	0.0122	0.0216	0.0125	0.0215
	2^{23}	0.00715	0.0135	0.00714	0.0141	0.00737	0.0141	0.00704	0.0142
2	2^{13}	0.0593	0.0786	0.06	0.079	0.0604	0.0799	0.0598	0.0808
	2^{15}	0.0344	0.0513	0.0353	0.0527	0.0348	0.0535	0.0361	0.0539
	2^{17}	0.0202	0.034	0.0204	0.035	0.0207	0.035	0.0205	0.0358
	2^{19}	0.0118	0.0226	0.0116	0.0232	0.012	0.0234	0.0117	0.0233
	2^{21}	0.00688	0.0145	0.00658	0.015	0.00687	0.0154	0.00667	0.0153
	2^{23}	0.00373	0.00953	0.00381	0.0101	0.00376	0.01	0.00374	0.0101

CDF paths and LS Gaussian bands. Figure 3 illustrates the band of Proposition 3.9 under the uniform law, using one independently simulated OUE release per displayed condition at $D = 128$. Each release retains $m = \lfloor n/7 \rfloor$ independent cycles and reconstructs the LS scores in (7). We estimate their full sample covariance and use 1999 Gaussian draws to calibrate (14); drawing from the empirical correlation matrix is conditionally equivalent to the Gaussian multiplier construction. Endpoints are ordered by (13) and extended as in Section 2.5. The point estimate $\overline{F}$ clips the cumulative maximum of the raw LS means to this band. The simulations use NumPy SeedSequence with entropy $(2026092201, n, 100\varepsilon)$ and three child streams for observations, reports and calibration; no replication is selected by its outcome. These panels illustrate individual releases, while Table 2 evaluates coverage over repeated samples.

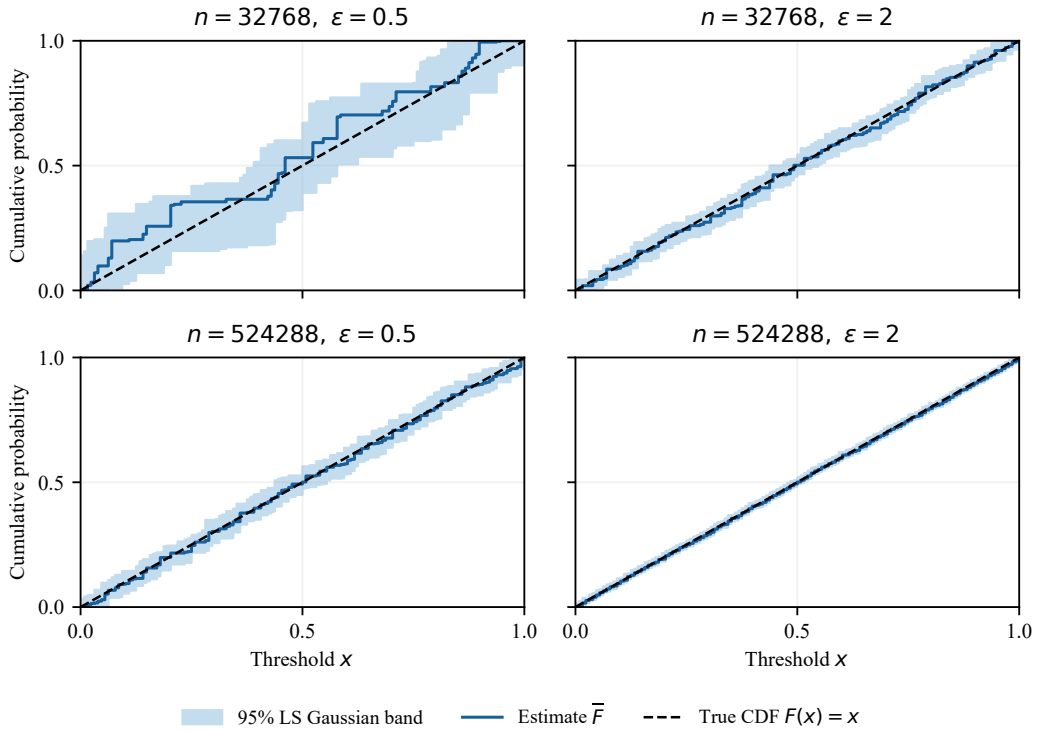

Figure 3: The nominal 95% LS Gaussian band of Proposition 3.9, using scores (7), calibration (14) and ordered endpoints (13), at $D = 128$. The dashed line is $F(x) = x$, the solid step curve is the band-constrained estimate $\overline{F}$, and shading shows the continuous extension $[L_k, U_{k+1}]$ between grid points. The coverage guarantee is asymptotic at fixed D and ε; this is distinct from the finite-sample Hoeffding band with radii (12).

A.3 Hadamard response and channel comparison

We apply the high-privacy Hadamard response of Acharya et al. (2019) separately at each assigned level. For $K = 2^r$ categories, let H be the Sylvester Hadamard matrix of size $2K$, with rows and columns indexed from zero. Category $c \in \{1, \dots, K\}$ uses nonconstant row c and releases a column $Z \in \{0, \dots, 2K - 1\}$ according to

$$\mathbb{P}(Z = z \mid C_r(X) = c) = \frac{1 + \rho_\varepsilon H_{cz}}{2K}, \qquad W^{\mathrm{HR}}_{i,(r,j)} = \frac{H_{jZ_i}}{\rho_\varepsilon}. \tag{31}$$

Every nonconstant row is balanced, so these probabilities sum to one. The largest likelihood ratio is $(1 + \rho_\varepsilon)/(1 - \rho_\varepsilon) = e^\varepsilon$. Orthogonality gives $\mathbb{E}(W^{\mathrm{HR}}_{i,(r,j)} \mid X_i) = \mathbf{1}\{C_r(X_i) = j\}$. Each report therefore supports all node estimates at its level, using $\log_2(2K) = r + 1$ bits. Define $V^{\mathrm{HR}}_{b,k}$ by summing the decoded HR scores over $\mathcal{D}(k)$, using independent users across levels as for OUE.

For a node with mass π_ν, the Hadamard score has variance $\rho_\varepsilon^{-2} - \pi_\nu^2$. Since $\rho_\varepsilon^{-2} = 1 + q_\varepsilon(1-q_\varepsilon)/a_\varepsilon^2$, subtracting the OUE cycle variance in (28) gives

$$\mathrm{Var}(V_{b,k}^{\mathrm{HR}}) - \mathrm{Var}(V_{b,k}) = |\mathcal{D}(k)| - 2F(x_k). \tag{32}$$

Thus compact reports can have either larger or smaller pointwise variance, depending on prefix complexity and mass. For fixed D and ε, both methods have independent bounded cycle scores with positive variance, so the ordinary central limit theorem and sample-variance consistency justify their pointwise studentized intervals as $m \to \infty$. This comparison uses fixed $D = 128$ and $\varepsilon = 1$. The Hadamard scores have range length $2/\rho_\varepsilon = 1/a_\varepsilon$; hence the concentration argument for (12) and Corollary 3.7 applies unchanged.

The following comparison reports both CDF error and pointwise inference. It uses the finite-sample Hoeffding band to define the CDF $\overline{F}$ inside the band; thus it is distinct from the LS Gaussian-band experiment. The pointwise intervals use raw canonical scores. The continuous errors and interval lengths are in CDF units; lower values indicate better precision provided the corresponding coverage remains adequate.

Table 6: Continuous Kolmogorov errors (KS; smaller is better) and raw 95% pointwise inference at $x = 1/2$, with $n = 2^{13}$, $D = 128$, $\varepsilon = 1$ and 600 replications per law. The estimates $\widehat{F}$ and $\overline{F}$ use cumulative maxima and clamping within the Hoeffding band, respectively. OUE denotes TreeOUE and HR denotes Hadamard response. Entries are means or coverage percentages.

Law	Channel	KS $\widehat{F}$	KS $\overline{F}$	Raw cov. (%)	Raw width
Uniform	OUE	0.236	0.233	95.7	0.241
Uniform	HR	0.259	0.252	95.7	0.241
Beta$(2,1)$	OUE	0.242	0.239	94.5	0.233
Beta$(2,1)$	HR	0.267	0.259	93.0	0.246
Mixture	OUE	0.223	0.221	96.2	0.245
Mixture	HR	0.246	0.241	94.3	0.237

The paired HR-minus-OUE increases in mean error for the CDF inside the band are 0.0195, 0.0206 and 0.0206 across the three laws. HR uses fewer bits and has the shorter raw interval on the mixture. Equation (32) explains this target-dependent variance ordering.

A.4 Covariance calibration

For each law, a separate 600-replication experiment compares Hoeffding radii, Gaussian Bonferroni radii and joint Gaussian radii on identical canonical cycle scores. Bonferroni uses each observed standard error and $z_{1-0.05/[2(D-1)]}$. Joint calibration uses 1999 Gaussian draws and the 1900th ordered maximum. Generating the Gaussian vector through a factorization of the cycle correlation matrix is distributionally equivalent to the multiplier construction in (14). Every method applies the same ordering and $[0,1]$ replacement rule.

Joint midpoint widths are 16–22% shorter than Hoeffding widths, with ordered simultaneous coverage 94.2–95.8%. Estimated marginal variances account for most of the improvement; the joint critical value is about 3.41–3.42, compared with the Gaussian Bonferroni value 3.54, giving roughly a further 4% reduction. Proposition 3.8 supplies the asymptotic justification for this Gaussian procedure; its finite-sample performance is evaluated here.

A.5 Same-report consistency reconstruction

The experiment in Table 7 fixes the grid, equal reconstruction weights and nominal level across all laws. Canonical and LS decoding share every report within a channel, and both estimate their actual cycle covariance. Calibration uses 1999 Gaussian draws in each of 600 independent replications per law. The reconstruction uses the public equal weights and total-mass constraint in (7).

Mean grid widths decrease from 0.413–0.465 to 0.243–0.277, with paired absolute reductions 0.169–0.188. Widths at $1/2$ decrease by 46–49%. The $[0,1]$ band is used in 0–0.5% of LS runs and is included in every summary. The mean continuous error of the CDF inside the band decreases from

0.218–0.252 to 0.120–0.138. The fixed public grid and weights place this experiment under the coverage guarantee of Proposition 3.9.

Table 7: Gaussian-calibrated bands from canonical (C) and fixed least-squares (LS) cycle scores on identical reports. Each cell uses 600 replications.

Law	Channel	Decode	Width at $1/2$	Mean grid width	Coverage (%)	$[0,1]$ band (%)
Uniform	OUE	C	0.407	0.448	95.2	0.0
Uniform	OUE	LS	0.211	0.261	96.3	0.0
Uniform	HR	C	0.409	0.465	95.3	0.2
Uniform	HR	LS	0.217	0.277	94.2	0.2
Beta$(2,1)$	OUE	C	0.392	0.418	95.5	0.2
Beta$(2,1)$	OUE	LS	0.213	0.243	95.2	0.5
Beta$(2,1)$	HR	C	0.405	0.441	95.2	0.0
Beta$(2,1)$	HR	LS	0.214	0.253	96.8	0.0
Mixture	OUE	C	0.417	0.413	94.8	0.0
Mixture	OUE	LS	0.211	0.244	95.2	0.2
Mixture	HR	C	0.408	0.431	95.5	0.3
Mixture	HR	LS	0.212	0.257	96.2	0.0

The band-constrained point estimate differs from the symmetric-envelope point estimate in (8). In the pointwise experiment, moving raw midpoint intervals to the cumulative maximum reduces uniform-law coverage from 96.2% to 63.8%; the paired decrease is 32.3 percentage points before rounding. The predecessor has 2.28 times the standard deviation across a true gap of $1/128$. Even after LS reconstruction, recentering gives 67.0–78.7% coverage under the two continuous laws. These results motivate using the raw scores for pointwise inference and ordering the endpoints for simultaneous inference.

B Comparisons and Further Inference Results

B.1 Reporting experiments

In the noninteractive design of Liu et al. (2024), each user returns one randomized comparison bit at an independently generated public threshold. Their truthful-response parameter is $\rho_\varepsilon = \tanh(\varepsilon/2)$. Theorem 4.1 gives $L^2[0,1]$ error $O_{\mathbb{P}}(\rho_\varepsilon^{-1} n^{-1/3})$ when the threshold density is positive; its additional positive population-density condition gives full-interval uniform error $O_{\mathbb{P}}(\rho_\varepsilon^{-1} n^{-1/3} \log n)$. Theorem 4.2 gives a Chernoff limit at an interior target under the density assumptions. For grid spacing $n^{-\gamma}$, $0 < \gamma < 1/3$, Theorem 4.3 instead gives Gaussian error of order $n^{-(1-\gamma)/2}$ at the two grid points bracketing a fixed interior target, at fixed privacy and with continuous positive population density. Its equation (5) gives joint root-n normality on a fixed finite grid.

The introduction compares bounds for the same full-interval uniform loss at fixed privacy on a common bounded-density class. The LS estimator satisfies this improved bound by Corollary 3.3. The bound of Liu et al. (2024) is an upper guarantee for their continuous-threshold estimator, rather than a lower bound on its attainable error. Our expected-risk bound additionally permits the minimax comparison in Corollary 3.5. Their grid-valued Gaussian limits concern designated points; continuous interpolation adds a resolution-dependent error. Section 2.2 gives the communication comparison.

Hierarchical and factorization bounds. Finite-workload methods control simultaneous error over a prescribed query family (Bassily, 2019; Edmonds et al., 2020). For a binary hierarchy of depth J, Theorem 1 of Cormode et al. (2019) bounds the variance of a range estimate by $O(J^2/(n\varepsilon^2))$ for $0 < \varepsilon \le 1$ with uniform level allocation. Their Section 4.7 specializes the decomposition to prefixes, and Section 4.5 uses least squares to reconcile node masses. This is a single-query variance guarantee.

For simultaneous continuous loss, specialize the exact factorization mechanism of Edmonds et al. (2020) to the prefix matrix $\boldsymbol{P}$, where $P_{kj} = \mathbf{1}\{j \le k\}$, $1 \le k < D$, $1 \le j \le D$. With the node–leaf matrix $\boldsymbol{A}$ from Section 2.3, let row k of $\boldsymbol{R}$ select the nodes in $\mathcal{D}(k)$. Then $\boldsymbol{P} = \boldsymbol{R}\boldsymbol{A}$. Every column

of $\boldsymbol{A}$ has J ones, and every row of $\boldsymbol{R}$ has at most J ones, so the product of the maximum column and row Euclidean norms is at most J. Their mechanism therefore gives expected maximum empirical-prefix error $O(J\sqrt{\log(2D)}/(\varepsilon\sqrt{n}))$. For i.i.d. observations, the empirical-CDF fluctuation adds $O(n^{-1/2})$ (Dvoretzky et al., 1956; Massart, 1990). Applying (5) and (6) gives a continuous estimator $\widehat{F}^{\mathrm{fac}}$ satisfying

$$\mathbb{E}\|\widehat{F}^{\mathrm{fac}} - F\|_\infty \le \frac{CJ\sqrt{\log(2D)}}{\varepsilon\sqrt{n}} + \frac{C}{\sqrt{n}} + \omega_D(F), \qquad 0 < \varepsilon \le 1.$$

Balancing the grid on $\mathcal{F}(s, L)$ yields the upper rate in Corollary 3.2, since $\varepsilon \asymp \rho_\varepsilon$ in this privacy range. The cycle construction attains that rate with explicit report lengths and supplies the variance and covariance estimates used for inference. Corollary 3.3 also controls the reconstructed estimator, and Theorem 3.4 gives a lower bound on the bounded-density continuous class allowing sequential interaction.

B.2 Fixed-threshold and finite-grid benchmarks

Report and estimator. At a public threshold x, user i forms $B_i(x) = \mathbf{1}\{X_i \le x\}$ and releases $Z_i(x) \in \{0, 1\}$ by Warner randomized response (RR),

$$\mathbb{P}(Z_i = 1 \mid B_i = 1) = p_\varepsilon, \qquad \mathbb{P}(Z_i = 1 \mid B_i = 0) = q_\varepsilon. \tag{33}$$

The standard unbiased estimator is

$$\widehat{F}_n^{\mathrm{RR}}(x) = \frac{\overline{Z}_n(x) - q_\varepsilon}{\rho_\varepsilon}, \qquad \overline{Z}_n(x) = n^{-1}\sum_{i=1}^n Z_i(x). \tag{34}$$

This benchmark quantifies the variance attainable when every user reports on the same threshold. TreeOUE instead uses the OUE binary contrast and allocates users across hierarchy levels to release a reusable CDF.

Variance and minimax benchmark.

Proposition B.1. *Fix $x \in (0, 1)$. The channel in* (33) *is ε-LDP, and the estimator in* (34) *is exactly unbiased. For every distribution F,*

$$\mathrm{Var}_F\{\widehat{F}_n^{\mathrm{RR}}(x)\} = \frac{1}{n}\left[F(x)\{1 - F(x)\} + \frac{e^\varepsilon}{(e^\varepsilon - 1)^2}\right]. \tag{35}$$

For fixed $\varepsilon > 0$,

$$\sqrt{n}\{\widehat{F}_n^{\mathrm{RR}}(x) - F(x)\} \rightsquigarrow N\left(0, F(x)\{1 - F(x)\} + \frac{e^\varepsilon}{(e^\varepsilon - 1)^2}\right). \tag{36}$$

Moreover, for each fixed $x \in (0, 1)$ there are constants $c_x, c_0, C_x > 0$ such that, whenever $n\rho_\varepsilon^2 \ge C_x$,

$$\inf_{\mathcal{M},T}\sup_{F\in\mathcal{F}_{\mathrm{bd}}} \mathbb{P}_F\left\{|T - F(x)| \ge \frac{c_x}{\rho_\varepsilon\sqrt{n}}\right\} \ge c_0. \tag{37}$$

Here the infimum ranges over all sequentially interactive ε-LDP protocols and all transcript-measurable estimators T.

The privacy term in (35) is of order ε^{-2} for small ε and vanishes as $\varepsilon \to \infty$, leaving the ordinary Bernoulli variance. Proposition B.1 is the standard one-query benchmark supplied by randomized response and classical locally private minimax theory (Warner, 1965; Duchi et al., 2013; 2018; Kairouz et al., 2016). Its exact variance calibrates the cost of a reusable CDF, for which TreeOUE divides users among several public levels.

Corollary B.2. *Suppose that P is supported on the public set $\{x_1, \ldots, x_D\}$, with no mass at zero. Then $\omega_D(F) = 0$, and*

$$\left\|\widehat{F}_{m,D} - F\right\|_\infty = \max_{0\le k\le D}\left|\widehat{F}_{m,D}(x_k) - F(x_k)\right|. \tag{38}$$

Thus Algorithm 1 reduces exactly to release of the finite prefix workload through a binary hierarchy and OUE.

The fixed-threshold proof is in Appendix E.1; the finite-grid identity is proved in Appendix D.

B.3 POINTWISE INFERENCE REFINEMENTS

The following consequences give bias bounds and sufficient conditions for the raw and monotone pointwise limits. Their proofs are in Appendices E and F.

Corollary B.3. *Under* (23) *and* (26)*, the interval* $\mathcal{I}^{\mathrm{raw}}_{n,\alpha}(x)$ *in* (10) *has asymptotic coverage* $1-\alpha$ *for* $F(x)$*. More generally, suppose that a public deterministic bound satisfies*

$$0 \le F(x) - F(x_n) \le b_n(x). \tag{39}$$

Under (23)*, the asymmetric interval* $\mathcal{I}^{\mathrm{bias}}_{n,\alpha}(x)$ *in* (11) *obeys*

$$\liminf_{n\to\infty} \mathbb{P}_F\{F(x) \in \mathcal{I}^{\mathrm{bias}}_{n,\alpha}(x)\} \ge 1-\alpha. \tag{40}$$

Neither interval uses additional privacy budget.

Corollary B.4. *For every* n*,*

$$\frac{q_{\varepsilon_n}}{2}\ell_n(x) \le A_n(x) \le \frac{1}{4}\ell_n(x), \tag{41}$$

and hence

$$\frac{\sqrt{2q_{\varepsilon_n}\ell_n(x)}}{\rho_{\varepsilon_n}\sqrt{m_n}} \le \frac{\sigma_n(x)}{\sqrt{m_n}} \le \frac{\sqrt{\ell_n(x)}}{\rho_{\varepsilon_n}\sqrt{m_n}}. \tag{42}$$

Therefore the public conditions

$$m_n \to \infty, \qquad m_n\ell_n(x)q_{\varepsilon_n} \to \infty$$

are sufficient for (24) *and* (25)*.*

If $F \in \mathcal{F}(s,L)$*,* $\varepsilon_n \le \bar{\varepsilon} < \infty$*, and* $J_n = o(n)$*, then the additional public undersmoothing condition*

$$D_n^{-s}\sqrt{\frac{n}{J_n\ell_n(x)}} \longrightarrow 0 \tag{43}$$

implies (26)*. In this bounded-privacy regime,*

$$\frac{\sigma_n(x)}{\sqrt{m_n}} \asymp_{\bar{\varepsilon}} \frac{\sqrt{\ell_n(x)J_n}}{\rho_{\varepsilon_n}\sqrt{n}}. \tag{44}$$

In particular, $D_n \asymp n^{1/(2s)}$ *satisfies* (43) *whenever* $J_n \to \infty$*.*

Proposition B.5. *For a target grid index* k_n*, write* $g_{n,j} = \widetilde{F}_{m_n,D_n}(x_j)$ *and* $f_{n,j} = F(x_j)$*. Suppose there is a deterministic sequence* $\eta_n \downarrow 0$ *such that*

$$\mathbb{P}\left\{\max_{0\le j\le D_n} |g_{n,j} - f_{n,j}| \le \eta_n\right\} \longrightarrow 1.$$

If, eventually,

$$\eta_n < f_{n,k_n} < 1-\eta_n, \qquad f_{n,k_n} - f_{n,k_n-1} > 2\eta_n, \tag{45}$$

then

$$\mathbb{P}\{\widehat{F}_{m_n,D_n}(x_{k_n}) = \widetilde{F}_{m_n,D_n}(x_{k_n})\} \longrightarrow 1. \tag{46}$$

Every raw pointwise limit therefore transfers to the released monotone estimate under (45)*.*

This is a sufficient local separation condition, not a regularity assumption required by Algorithm 1. For a density bounded below near x, the condition follows from $D_n\eta_n \to 0$. The uniform rate in Corollary 3.2 does not imply that relation at the usual Lipschitz resolutions.

Since $\sum_\nu \pi^2_{n,\nu} \le F(x_n)$, the decomposition implies $A_n(x) \ge a^2_{\varepsilon_n}F(x_n)$. Thus, if $\varepsilon_n \to \infty$ and $\liminf_n F(x_n) > 0$, the condition $m_n \to \infty$ already ensures $m_nA_n(x) \to \infty$. For example, at an eventually represented fixed dyadic target with $F(x) > 0$, this remains true when $q_{\varepsilon_n} = o(m_n^{-1})$. The public sufficient condition in Corollary B.4 then fails because $\ell_n(x)$ is fixed, although the theorem still gives studentized inference. The exact variance therefore permits regimes that a bound based only on the privacy level and prefix size would miss.

C Auxiliary Results

C.1 Grid approximation

Lemma C.1. *For every CDF F on $[0,1]$, the function F_D° in (1) satisfies*

$$\|F_D^\circ - F\|_\infty = \max\left\{F(x_1-), \max_{1\le k<D}\{F(x_{k+1}-) - F(x_k)\}\right\} = \omega_D(F). \tag{47}$$

Moreover, $\omega_D(F) \le \Delta_D(F)$, and $\omega_D(F) \le LD^{-s}$ for $F \in \mathcal{F}(s,L)$.

Proof. On $[0,x_1)$, $F_D^\circ(x) = 0$. Since F is nondecreasing,

$$\sup_{0\le x<x_1} |F_D^\circ(x) - F(x)| = \sup_{0\le x<x_1} F(x) = F(x_1-). \tag{48}$$

For $1 \le k < D$ and $x \in [x_k, x_{k+1})$, $F_D^\circ(x) = F(x_k)$ and $F(x) \ge F(x_k)$. Hence

$$\sup_{x_k\le x<x_{k+1}} |F_D^\circ(x) - F(x)| = F(x_{k+1}-) - F(x_k). \tag{49}$$

At $x = 1$ the error is zero. Taking the maximum in (48)–(49) proves the equality.

Because $F(0-) = 0$,

$$F(x_1-) \le F(x_1) - F(0-) = F(x_1).$$

For $1 \le k < D$,

$$F(x_{k+1}-) - F(x_k) = P(x_k < X < x_{k+1}) \le F(x_{k+1}) - F(x_k).$$

Thus every term in (47) is at most $\Delta_D(F)$. If $F \in \mathcal{F}(s,L)$, then $F(0) = 0$ and

$$F(x_1-) \le F(x_1) - F(0) \le LD^{-s},$$

while

$$F(x_{k+1}-) - F(x_k) \le F(x_{k+1}) - F(x_k) \le LD^{-s}.$$

This proves the remaining assertions. □

Lemma C.2. *Let $\widehat{f}_0 = 0 \le \widehat{f}_1 \le \cdots \le \widehat{f}_D = 1$ and let $\widehat{f}$ be its left-grid step extension, as in (6). Then*

$$\left\|\widehat{f} - F\right\|_\infty \le \max_{1\le k<D}\left|\widehat{f}_k - F(x_k)\right| + \omega_D(F). \tag{50}$$

If P is supported on $\{x_1,\dots,x_D\}$, then equality in (38) holds.

Proof. Both $\widehat{f}$ and F_D° equal zero on the first cell and one at the upper endpoint. On the remaining cells their values differ by $\widehat{f}_k - F(x_k)$, so

$$\left\|\widehat{f} - F_D^\circ\right\|_\infty = \max_{1\le k<D}\left|\widehat{f}_k - F(x_k)\right|. \tag{51}$$

The triangle inequality and Lemma C.1 give (50).

If P is supported on $\{x_1,\dots,x_D\}$, then $F(x) = 0$ on $[0,x_1)$ and $F(x) = F(x_k)$ on $[x_k,x_{k+1})$. Thus $F = F_D^\circ$ and $\omega_D(F) = 0$. Equation (51) then becomes the claimed equality. □

C.2 Dyadic decomposition and OUE

Lemma C.3. *For each $1 \le k < D = 2^J$, there is a unique canonical set $\mathcal{D}(k)$ containing at most one dyadic cell from each level such that*

$$[0, k/D] = \dot{\bigcup_{\nu\in\mathcal{D}(k)}} I_\nu. \tag{52}$$

Its cardinality is the number of ones in the J-digit binary expansion of k.

Proof. Write

$$k = \sum_{h=0}^{J-1} b_h 2^h, \qquad b_h \in \{0, 1\}.$$

Read the nonzero bits from $h = J - 1$ down to $h = 0$. Whenever $b_h = 1$, append immediately to the right of the previously selected cells a dyadic interval of length

$$\frac{2^h}{2^J} = 2^{-(J-h)}.$$

This interval belongs to level $r = J - h$. Distinct nonzero bits give distinct levels. The selected intervals are adjacent and disjoint, start at zero, and have total length

$$\sum_{h=0}^{J-1} b_h 2^{-(J-h)} = \frac{k}{D}.$$

They therefore form the disjoint union in (52). Their number is $\sum_h b_h$. Uniqueness of the binary expansion determines the selected lengths; adjacency then determines their locations. This proves uniqueness of the canonical set. □

Lemma C.4. *Let a category $c \in \{1, \dots, K\}$ generate conditionally independent coordinates satisfying*

$$\mathbb{P}(Y_j = 1 \mid c) = \begin{cases} 1/2, & j = c, \\ q_\varepsilon, & j \neq c. \end{cases}$$

Then $c \mapsto Y$ is ε-LDP. For $W_j = (Y_j - q_\varepsilon)/a_\varepsilon$,

$$\mathbb{E}(W_j \mid c) = \mathbf{1}\{j = c\},$$

and the two possible values of W_j differ by $1/a_\varepsilon$.

Proof. Consider two inputs $c \neq c'$. Coordinates other than c and c' have the same distribution under both inputs. For the two differing coordinates, the largest output likelihood ratio occurs when $(Y_c, Y_{c'}) = (1, 0)$ and equals

$$\frac{(1/2)(1 - q_\varepsilon)}{q_\varepsilon(1 - 1/2)} = \frac{1 - q_\varepsilon}{q_\varepsilon} = e^\varepsilon.$$

The reciprocal configuration has ratio $e^{-\varepsilon}$, and the configurations $(0, 0)$ and $(1, 1)$ have ratio one. Conditional independence therefore bounds every vector likelihood ratio in either direction by e^ε.

If $j = c$, then $\mathbb{E}Y_j = 1/2 = q_\varepsilon + a_\varepsilon$, so $\mathbb{E}W_j = 1$. If $j \neq c$, then $\mathbb{E}Y_j = q_\varepsilon$, so $\mathbb{E}W_j = 0$. Finally, the two decoded values are $-q_\varepsilon/a_\varepsilon$ and $(1 - q_\varepsilon)/a_\varepsilon$, whose difference is $1/a_\varepsilon$. □

C.3 Monotone stability and standard probability inequalities

Lemma C.5. *Let $0 = f_0 \leq f_1 \leq \cdots \leq f_D = 1$. If $\max_k |g_k - f_k| \leq \delta$, then*

$$\max_{0 \leq k \leq D} |\mathcal{T}(g)_k - f_k| \leq \delta. \tag{53}$$

Proof. Fix $1 \leq k < D$. Since $g_k \geq f_k - \delta$,

$$\max_{1 \leq j \leq k} g_j \geq f_k - \delta.$$

For every $j \leq k$, monotonicity of the true vector gives

$$g_j \leq f_j + \delta \leq f_k + \delta,$$

so the same maximum is at most $f_k + \delta$. Projection of a real number onto $[0, 1]$ cannot increase its distance from $f_k \in [0, 1]$. The endpoint errors are zero. Taking the maximum over k proves (53). □

We use Hoeffding's lemma and its Chernoff consequence in their standard forms (Hoeffding, 1963). Specifically, if a centered random variable Y has range length c, then

$$\mathbb{E}e^{\lambda Y} \le e^{\lambda^2 c^2/8}, \qquad \lambda \in \mathbb{R}. \tag{54}$$

We also use the independent, non-identically distributed Berry–Esseen inequality: for centered independent U_i, with $B^2 = \sum_i \mathbb{E}U_i^2 > 0$,

$$\sup_t \left| \mathbb{P}\left(\frac{\sum_i U_i}{B} \le t \right) - \Phi(t) \right| \le C_{\mathrm{BE}} \frac{\sum_i \mathbb{E}\,|U_i|^3}{B^3}. \tag{55}$$

See Petrov (1975) for this form of the Berry–Esseen inequality.

Lemma C.6. *Let P, Q be two input distributions. Under any sequentially interactive ε-LDP protocol applied to n independent users, let $M_P^{(n)}$ and $M_Q^{(n)}$ denote the transcript laws. Write* KL *for Kullback–Leibler divergence and* $\mathrm{TV}(P,Q) = \sup_A |P(A) - Q(A)|$ *for total variation distance. Then*

$$\mathrm{KL}(M_P^{(n)} \| M_Q^{(n)}) + \mathrm{KL}(M_Q^{(n)} \| M_P^{(n)}) \le 8n(e^\varepsilon - 1)^2\, \mathrm{TV}(P,Q)^2. \tag{56}$$

For $0 < \varepsilon \le 1$, the right side is at most $C_{\mathrm{LDP}} n \rho_\varepsilon^2\, \mathrm{TV}(P,Q)^2$ for a universal constant C_{LDP}.

Proof. The sequential contraction of Duchi et al. (2018) bounds each directed transcript KL divergence by $4n(e^\varepsilon - 1)^2\, \mathrm{TV}(P,Q)^2$. Applying it in both directions and adding gives (56). The conditional channels remain ε-LDP after every realized past transcript, as required by that result.

For $0 < \varepsilon \le 1$,

$$e^\varepsilon - 1 = \frac{2\rho_\varepsilon}{1 - \rho_\varepsilon} \le \frac{2}{1 - \tanh(1/2)} \rho_\varepsilon.$$

Substitution into (56) gives the second assertion. □

D Proofs for Uniform Estimation

Proof of Theorem 3.1. Privacy follows user by user from Lemma C.4. Public level assignment does not depend on private data, and decoding, prefix summation, monotone post-processing, and step extension are post-processing.

Fix a grid prefix $1 \le k < D$. By Lemma C.3, its selected cells are disjoint and have union $[0, x_k]$. Lemma C.4 therefore gives

$$\mathbb{E}V_{b,k} = \sum_{\nu \in \mathcal{D}(k)} P(X \in I_\nu) = F(x_k).$$

Within cycle b, the selected coordinates in $V_{b,k}$ come from distinct level users, because $\mathcal{D}(k)$ contains at most one cell at each level. They are independent. Each decoded coordinate has range length $1/a_\varepsilon$. Applying Hoeffding's lemma (54) to the centered coordinates and multiplying their moment generating functions gives, for every $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\exp[\lambda\{V_{b,k} - F(x_k)\}] &\le \exp\left\{ \frac{\lambda^2\, |\mathcal{D}(k)|}{8a_\varepsilon^2} \right\} \\ &\le \exp\left(\frac{\lambda^2 J}{8a_\varepsilon^2} \right). \end{aligned} \tag{57}$$

The m cycles are independent. Applying (57) to their average yields

$$\mathbb{E}\exp\left[\lambda\{\widetilde{F}_{m,D}(x_k) - F(x_k)\}\right] \le \exp\left(\frac{\lambda^2 J}{8ma_\varepsilon^2} \right). \tag{58}$$

Chernoff optimization for each sign gives

$$\mathbb{P}\left\{ \left| \widetilde{F}_{m,D}(x_k) - F(x_k) \right| > t \right\} \le 2\exp\left(-\frac{2ma_\varepsilon^2 t^2}{J} \right).$$

A union bound over $k = 1, \dots, D-1$ shows that, with probability at least $1-\beta$,

$$\max_{1\le k<D} \left|\widetilde{F}_{m,D}(x_k) - F(x_k)\right| \le \frac{1}{a_\varepsilon}\sqrt{\frac{J}{2m}\log\frac{2(D-1)}{\beta}}. \tag{59}$$

Apply Lemma C.5 to the nondecreasing comparison vector $(0, F(x_1), \dots, F(x_{D-1}), 1)$, whose interior entries are the true prefix probabilities. The lemma transfers (59) to the post-processed grid vector without increasing its right side. Lemma C.2 then adds exactly the deterministic approximation error $\omega_D(F)$. This proves (15).

For the expectation, put $Z_k = \widetilde{F}_{m,D}(x_k) - F(x_k)$. For every $\lambda > 0$, the elementary inequality $e^{\lambda \max_j u_j} \le \sum_j e^{\lambda u_j}$ and (58) give

$$\begin{aligned}\mathbb{E}\max_{1\le k<D}|Z_k| &\le \frac{1}{\lambda}\log\sum_{k=1}^{D-1}\left(\mathbb{E}e^{\lambda Z_k} + \mathbb{E}e^{-\lambda Z_k}\right)\\ &\le \frac{\log\{2(D-1)\}}{\lambda} + \frac{\lambda J}{8ma_\varepsilon^2}.\end{aligned}$$

The right side is minimized at

$$\lambda = \sqrt{\frac{8ma_\varepsilon^2\log\{2(D-1)\}}{J}},$$

which yields

$$\mathbb{E}\max_{1\le k<D}|Z_k| \le \frac{1}{a_\varepsilon}\sqrt{\frac{J}{2m}\log\{2(D-1)\}}.$$

The same two deterministic stability steps prove (16). Lemma C.1 supplies the Hölder-class substitutions. □

Proof of Corollary 3.7. Retain $|\mathcal{D}(k)|$ in the first line of (57). Chernoff optimization for the average of m cycles gives

$$\mathbb{P}\{|g_k - F(x_k)| > h_k\} \le 2\exp\{-2ma_\varepsilon^2 h_k^2/|\mathcal{D}(k)|\} = \frac{\beta}{D-1}.$$

By the union bound, all positive interior grid values belong to their original intervals with probability at least $1-\beta$. On this event, for $j \le k$, $g_j - h_j \le F(x_j) \le F(x_k)$; for $j \ge k$, $g_j + h_j \ge F(x_j) \ge F(x_k)$. Hence $L_k \le F(x_k) \le U_k$, so incompatible endpoints cannot trigger replacement by $[0,1]$ on the coverage event. The replacement produces ordered, nonempty intervals. Both endpoint sequences are nondecreasing, and minima and maxima of nondecreasing sequences remain nondecreasing, proving the same property for $\overline{F}$. In the compatible case, $L_k \ge g_k - h_k$ and $U_k \le g_k + h_k$, which gives the width bound. For $x \in (x_k, x_{k+1})$ with $k \ge 1$, $L_k \le F(x_k) \le F(x) \le F(x_{k+1}) \le U_{k+1}$. For the first cell, $0 \le F(x) \le F(x_1) \le U_1$, including mass at zero. The value at one is known. These inequalities prove continuous simultaneous coverage and show that the step CDF $\overline{F}$ lies in its continuous band. □

Proof of Corollary 3.2. By (17),

$$L(D_n^{\mathrm{bal}})^{-s} \le r_n^{-1}. \tag{60}$$

Because $\rho_{\varepsilon_n} \le 1$ and s, L are fixed,

$$J_n = \log_2 D_n^{\mathrm{bal}} = O\{\log(en)\}.$$

The assumption $r_n \to \infty$ implies $n \to \infty$, so $J_n = o(n)$ and $m_n \asymp n/J_n$. Also $\log\{2(D_n^{\mathrm{bal}}-1)\} = O(J_n)$. The stochastic term in (15) is therefore

$$\begin{aligned}\frac{1}{a_{\varepsilon_n}}\sqrt{\frac{J_n}{m_n}\log\{2(D_n^{\mathrm{bal}}-1)\}} &\le C\frac{J_n^{3/2}}{\rho_{\varepsilon_n}\sqrt{n}}\\ &\le C\frac{\{\log(en)\}^{3/2}}{\rho_{\varepsilon_n}\sqrt{n}} = Cr_n^{-1}.\end{aligned} \tag{61}$$

For any fixed failure probability, Theorem 3.1, (60), and (61) prove the uniform $O_{\mathbb{P}}(r_n^{-1})$ assertion. The expectation assertion follows from the same calculation applied to (16).

If $D_n \asymp n^{1/(2s)}$, then $J_n \asymp \log n$ and the stochastic calculation is unchanged, while $LD_n^{-s} = O(n^{-1/2})$. This proves the final comparison. □

Proof of Corollary B.2. Under the stated support condition, F is zero on $[0, x_1)$ and is constant on every half-open interval $[x_k, x_{k+1})$. Hence $F = F_D^\circ$ and $\omega_D(F) = 0$ by Lemma C.1. Lemma C.2 gives (38). The reported cell at each level is precisely a finite-domain category, and each prefix is reconstructed by its canonical binary hierarchy. This proves the reduction. □

D.1 Uniform error after least-squares reconstruction

Lemma D.1. *Let $\boldsymbol{y} = m^{-1}\sum_{b=1}^m \boldsymbol{W}_b$ be the vector of average OUE node scores on the complete balanced tree. Let $\boldsymbol{h}_k$ indicate the first k leaves, and let $\boldsymbol{Q}$ have orthonormal columns spanning $\mathbf{1}^\perp$. Put*

$$\boldsymbol{G} = \boldsymbol{Q}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{Q}, \qquad \boldsymbol{w}_k = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{G}^{-1}\boldsymbol{Q}^\top \boldsymbol{h}_k.$$

Then $\boldsymbol{G}$ is invertible and the reconstructed threshold estimate has the form

$$g_k^{\mathrm{LS}} = \frac{k}{D} + \boldsymbol{w}_k^\top \boldsymbol{y}, \qquad \mathbb{E}_F g_k^{\mathrm{LS}} = F(x_k), \qquad \|\boldsymbol{w}_k\|_2^2 \le |\mathcal{D}(k)| \le J. \tag{62}$$

For every deterministic node coefficient vector $\boldsymbol{w}$ and $t \in \mathbb{R}$,

$$\mathbb{E}_F \exp\{t\boldsymbol{w}^\top(\boldsymbol{y} - \mathbb{E}_F\boldsymbol{y})\} \le \exp\left\{\frac{t^2}{m}\left(\frac{1}{8a_\varepsilon^2} + \frac{1}{4}\right)\|\boldsymbol{w}\|_2^2\right\}. \tag{63}$$

Proof. The leaf rows of $\boldsymbol{A}$ form the identity matrix. Hence $\boldsymbol{A}$ has full column rank and $\boldsymbol{G}$ is positive definite. Every feasible leaf vector can be written uniquely as $\mathbf{1}/D + \boldsymbol{Q}\boldsymbol{z}$. Each leaf belongs to one node at every level, and all level-r nodes contain $D/2^r$ leaves. Consequently

$$\boldsymbol{A}^\top \boldsymbol{A}\mathbf{1} = \left(\sum_{r=1}^J \frac{D}{2^r}\right)\mathbf{1} = (D-1)\mathbf{1}, \qquad \boldsymbol{Q}^\top \boldsymbol{A}^\top \boldsymbol{A}\mathbf{1} = 0.$$

The constrained least-squares normal equation at input $\boldsymbol{y}$ therefore gives $\widehat{\boldsymbol{z}} = \boldsymbol{G}^{-1}\boldsymbol{Q}^\top \boldsymbol{A}^\top \boldsymbol{y}$. Taking its first-$k$ leaf sum proves the affine formula in (62). Since the map is affine, it also equals the mean of the reconstructed cycle scores.

The vector $\boldsymbol{w}_k$ is the minimum-Euclidean-norm solution of $\boldsymbol{Q}^\top \boldsymbol{A}^\top \boldsymbol{w} = \boldsymbol{Q}^\top \boldsymbol{h}_k$: for $\boldsymbol{H} = \boldsymbol{Q}^\top \boldsymbol{A}^\top$, the full-row-rank formula for this solution is $\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{Q}^\top \boldsymbol{h}_k$. The canonical coefficient vector $\boldsymbol{v}_k$, with ones on $\mathcal{D}(k)$ and zeros elsewhere, satisfies $\boldsymbol{A}^\top \boldsymbol{v}_k = \boldsymbol{h}_k$. It is feasible, and hence $\|\boldsymbol{w}_k\|_2^2 \le \|\boldsymbol{v}_k\|_2^2 = |\mathcal{D}(k)|$. Also $\boldsymbol{w}_k^\top \boldsymbol{A}\mathbf{1} = 0$ by the displayed identity. If $\boldsymbol{p} = \mathbf{1}/D + \boldsymbol{Q}\boldsymbol{z}$ is the true leaf-mass vector, then $\mathbb{E}_F\boldsymbol{y} = \boldsymbol{A}\boldsymbol{p}$ and

$$\frac{k}{D} + \boldsymbol{w}_k^\top \boldsymbol{A}\boldsymbol{p} = \frac{k}{D} + \boldsymbol{h}_k^\top \boldsymbol{Q}\boldsymbol{z} = \boldsymbol{h}_k^\top \boldsymbol{p} = F(x_k).$$

This includes the first leaf's mass at zero.

For the concentration claim, consider one user at one level, with category C and an arbitrary coefficient vector $\boldsymbol{u}$ at that level. Conditional on C, the OUE coordinates are independent and $\mathbb{E}(\boldsymbol{u}^\top \boldsymbol{W} \mid C) = u_C$. The range length of $u_j W_j$ is $|u_j|/a_\varepsilon$. Hoeffding's lemma yields

$$\mathbb{E}\{\exp[t(\boldsymbol{u}^\top \boldsymbol{W} - u_C)] \mid C\} \le \exp\{t^2\|\boldsymbol{u}\|_2^2/(8a_\varepsilon^2)\}.$$

The range of u_C is at most $\sqrt{2}\|\boldsymbol{u}\|_2$, because $(u_i - u_j)^2 \le 2(u_i^2 + u_j^2) \le 2\|\boldsymbol{u}\|_2^2$. A second application of Hoeffding's lemma and conditional expectation give

$$\mathbb{E}\exp\{t(\boldsymbol{u}^\top \boldsymbol{W} - \mathbb{E}\boldsymbol{u}^\top \boldsymbol{W})\} \le \exp\left\{t^2\left(\frac{1}{8a_\varepsilon^2} + \frac{1}{4}\right)\|\boldsymbol{u}\|_2^2\right\}.$$

There are m independent users at each level, with distinct users across levels. Apply this bound with each level's coefficient vector divided by m, multiply over those users and levels, and sum the squared coefficient norms. This proves (63). □

Proof of Corollary 3.3. Write $Z_k = g_k^{\mathrm{LS}} - F(x_k)$. Lemma D.1 gives $\mathbb{E}e^{tZ_k} \le e^{t^2v^2/2}$, where

$$v^2 = \left(\frac{1}{4a_\varepsilon^2} + \frac{1}{2}\right)\frac{J}{m}.$$

Chernoff's inequality and a union bound over both signs of the $D-1$ interior errors give $\mathbb{P}\{\max_k |Z_k| > u\} \le 2(D-1)e^{-u^2/(2v^2)}$. For the expectation, the exponential maximum bound gives, for $t > 0$,

$$\mathbb{E}\max_k |Z_k| \le \frac{\log\{2(D-1)\}}{t} + \frac{tv^2}{2}.$$

Minimizing in t yields $\sqrt{2v^2\log\{2(D-1)\}}$. Dependence between threshold estimates is allowed in both arguments.

For the deterministic postprocessing step, let $f_k = F(x_k)$ and suppose that $\max_{1\le k<D} |g_k - f_k| \le e$. Define $M_k = \max_{1\le j\le k} g_j$ and $m_k^- = \min_{k\le j<D} g_j$. For $j \le k$, $g_j \le f_j + e \le f_k + e$, while $M_k \ge g_k \ge f_k - e$. For $j \ge k$, $g_j \ge f_j - e \ge f_k - e$, while $m_k^- \le g_k \le f_k + e$. Thus both M_k and m_k^- lie in $[f_k - e, f_k + e]$. Both are nondecreasing in k, so their average is nondecreasing and has the same error bound. Clipping to $[0,1]$ preserves these properties. With the fixed endpoints, Lemma C.2 adds precisely the approximation error $\omega_D(F)$. Substitution into the probability and expectation bounds above proves (18) and its expected-loss version for arbitrary F.

The stochastic term is $\sqrt{1+2a_\varepsilon^2}$ times its canonical counterpart. Since $0 < a_\varepsilon < 1/2$, this factor is at most $\sqrt{3/2}$. At D_n^{bal}, the proof of Corollary 3.2 therefore gives the same uniform probability and expected-risk orders for every fixed Hölder pair (s, L).

It remains to check the publicly selected resolution in (30). Let $r_n = \rho_{\varepsilon_n}\sqrt{n}/\{\log(en)\}^{3/2} \to \infty$ and take $(s, L) = (1, 3/2)$. The balanced resolution obeys $D_n^{\mathrm{bal}} \le 3r_n$ eventually. Since $\rho_{\varepsilon_n} \le 1$, it eventually belongs to the candidate set $\mathcal{G}_n$ in (30). If $B_{n,\varepsilon}(D)$ denotes that equation's minimized bound, then

$$B_{n,\varepsilon_n}(D_0) \le B_{n,\varepsilon_n}(D_n^{\mathrm{bal}}) = O(r_n^{-1}).$$

Put $J_0 = \log_2 D_0$ and $J' = \log_2(4D_0) = J_0 + 2 \le 3J_0$. Both are $O(\log n)$, so $\lfloor n/J_0 \rfloor \asymp n/J_0$ and $\lfloor n/J' \rfloor \asymp n/J'$ uniformly over these choices. Also $\log\{2(4D_0-1)\} \le 4\log\{2(D_0-1)\}$ for $D_0 \ge 2$. The stochastic term at $4D_0$ is therefore at most a fixed multiple of that at D_0, and its approximation term is four times smaller. The LS factor remains bounded as above. This proves the expected-risk order at the actual selected grid, uniformly over $\mathcal{F}(1, 3/2)$; Markov's inequality also gives the uniform $O_{\mathbb{P}}(r_n^{-1})$ assertion. All resolutions and reconstruction coefficients depend only on public choices, n and ε. □

D.2 Covariance calibration from independent cycles

Proof of Proposition 3.8. Write $\sigma_k^2 = \mathrm{Var}(V_{b,k})$ and standardize each cycle by $Z_{b,k} = (V_{b,k} - F(x_k))/\sigma_k$. We suppress the sequence index n throughout this proof. The vectors Z_b are independent and identically distributed across cycles; their coordinates can be dependent. For OUE, independence of the selected levels and (9) give

$$\sigma_k^2 \ge \frac{|\mathcal{D}(k)|q_\varepsilon(1-q_\varepsilon)}{a_\varepsilon^2}, \qquad \mathbb{E}e^{tZ_{b,k}} \le \exp\left\{\frac{t^2|\mathcal{D}(k)|}{8a_\varepsilon^2\sigma_k^2}\right\} \le \exp\left\{\frac{t^2}{8q_\varepsilon(1-q_\varepsilon)}\right\}.$$

The last bound is uniform in k, D, F and $0 < \varepsilon \le \varepsilon_{\max}$. Thus the standardized coordinates have uniformly bounded sub-Gaussian tails, unit variances, and uniformly bounded third and fourth moments. For HR, (31) gives a node score with range length $2/\rho_\varepsilon$ and variance $\rho_\varepsilon^{-2} - \pi_\nu^2 \ge \rho_\varepsilon^{-2} - 1$. Consequently,

$$\sigma_k^2 \ge |\mathcal{D}(k)|(\rho_\varepsilon^{-2} - 1), \qquad \mathbb{E}e^{tZ_{b,k}} \le \exp\left\{\frac{t^2}{2(1-\rho_\varepsilon^2)}\right\} = \exp\left\{\frac{t^2}{8q_\varepsilon(1-q_\varepsilon)}\right\}.$$

Both channels therefore satisfy the same moment and tail conditions. No lower bound on ε is needed after standardization.

Let G be a centered Gaussian vector with covariance $\mathbb{E}Z_b Z_b^\top$, and put $M_G = \max_k |G_k|$. Proposition 2.1 of Chernozhukov et al. (2017), applied to these standardized vectors, gives

$$\sup_t \left| \mathbb{P}\left\{ \max_k \left| m^{-1/2} \sum_b Z_{b,k} \right| \le t \right\} - \mathbb{P}(M_G \le t) \right| \le C_{\varepsilon_{\max}} \left\{ \frac{\log^7(eDm)}{m} \right\}^{1/6} = o(1).$$

For one or two coordinates the ordinary fixed-dimensional Gaussian approximation gives the same convergence. The theorem for rectangles allows arbitrary correlations, including singular covariance matrices. Corollary 4.2 of the same reference gives the corresponding conditional Gaussian multiplier approximation in probability, first with true standard deviations in the denominators. The factor $\sqrt{m/(m-1)}$ introduced by (14) tends to one and does not change the approximation.

It remains to justify the observed standard deviations simultaneously. Products of the uniformly sub-Gaussian coordinates have uniformly sub-exponential tails. Bernstein's inequality over all coordinates, together with the bound on their sample means, gives

$$\max_k \left| \frac{\widehat{\sigma}_k^2}{\sigma_k^2} - 1 \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log(eD)}{m}} + \frac{\log(eD)}{m} \right).$$

In particular, the probability of a zero sample variance tends to zero uniformly. The maximum of the standardized averages is $O_{\mathbb{P}}(\sqrt{\log(eD)})$ by the sub-Gaussian bound; the conditional multiplier maximum has the same order in probability. Replacing true by sample standard deviations changes either maximum by $O_{\mathbb{P}}\{\log(eD)/\sqrt{m}\}$. Gaussian anti-concentration (Lemma A.1 of the cited reference, applied to the coordinates and their negatives) bounds the probability of an interval of length u for M_G by $Cu\sqrt{\log(eD)}$. The studentization error thus vanishes in distribution under the stated growth condition. We obtain

$$\sup_t |\mathbb{P}(T \le t) - \mathbb{P}(M_G \le t)| = o(1), \qquad \sup_t |\mathbb{P}(T^* \le t \mid V) - \mathbb{P}(M_G \le t)| = o_{\mathbb{P}}(1),$$

where $T = \max_k \sqrt{m}\,|g_k - F(x_k)|/\widehat{\sigma}_k$ and V denotes the observed cycle vectors. All bounds are uniform over the asserted distributions and privacy levels. Bracketing the random conditional quantile $c^*_{1-\beta}$ between the Gaussian quantiles at $1-\beta \pm o(1)$ now yields

$$\mathbb{P}\{T \le c^*_{1-\beta}\} = 1 - \beta + o(1).$$

On this event every raw grid interval covers its target. The deterministic ordering and continuous extension in the proof of Corollary 3.7 apply to these data-dependent radii as well. Using the $[0,1]$ band only adds coverage, proving the asserted lower bound for the full CDF. All calculations use the existing private reports, so privacy and communication are unchanged. □

Conditionally on the reports, T^* is the maximum absolute coordinate of a Gaussian vector whose correlation matrix is the sample correlation matrix of the cycle scores. It can be simulated either by cycle multipliers or by factoring this matrix. If the conditional quantile is approximated using an increasing number of independent Gaussian draws, uniform convergence of their empirical distribution preserves the coverage limit. Our numerical evaluation uses 1999 draws and the 1900th ordered maximum; this Monte Carlo approximation is included in the reported coverage results.

D.3 Consistency reconstruction

Proof. The last level of $\boldsymbol{A}$ is the identity, so $\boldsymbol{A}$ has full column rank and the constrained least-squares solution is a unique affine function of $\boldsymbol{W}_b$. Write it as $\boldsymbol{d} + \boldsymbol{B}\boldsymbol{W}_b$, with fixed $\boldsymbol{d}$ and $\boldsymbol{B}$. If $\boldsymbol{p}$ is the vector of true leaf masses, then $\mathbb{E}\boldsymbol{W}_b = \boldsymbol{A}\boldsymbol{p}$ and $\mathbf{1}^\top \boldsymbol{p} = 1$. At input $\boldsymbol{A}\boldsymbol{p}$ the unique minimizer is $\boldsymbol{p}$. Affinity therefore gives $\mathbb{E}\widehat{\boldsymbol{p}}_b = \boldsymbol{p}$ and $\mathbb{E}V^{\mathrm{LS}}_{b,k} = F(x_k)$. It also shows that reconstruction of the averaged node scores equals the average of the cycle reconstructions.

The full node-score covariance is positive definite. For OUE, conditional coordinate independence and the channel probabilities give

$$\mathrm{Cov}(\boldsymbol{W}_b) \succeq \frac{q_\varepsilon(1-q_\varepsilon)}{a_\varepsilon^2} \boldsymbol{I}_{2D-2}.$$

For HR at a level with K cells, every output column has probability at least $(1-\rho_\varepsilon)/(2K)$. Under a uniform column, the nonconstant Hadamard rows have zero means and orthogonal unit second moments. For any coefficient vector $\boldsymbol{t}$, minimizing the expected squared deviation over its center therefore gives $\mathrm{Var}\{\boldsymbol{t}^\top \boldsymbol{W}_{b,r}\} \geq (1-\rho_\varepsilon)\|\boldsymbol{t}\|_2^2/\rho_\varepsilon^2$, where $\boldsymbol{W}_{b,r}$ collects that level's scores. Independent levels give a strictly positive lower bound for the full covariance. Here $\boldsymbol{I}$ denotes the identity of the indicated dimension and $\succeq$ means that the difference is positive semidefinite.

For every vector $\boldsymbol{u}$ with $\mathbf{1}^\top \boldsymbol{u} = 0$, exact recovery on the affine constraint gives $\boldsymbol{B}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{u}$. The linear part of the kth prefix functional is nonzero: taking $\boldsymbol{u}$ to have first coordinate one, last coordinate minus one, and all other coordinates zero gives prefix sum one for every $1 \leq k < D$. Its variance is therefore strictly positive by the covariance bounds above.

With D and ε fixed, the transformed cycle vectors are independent, identically distributed and bounded. The multivariate central limit theorem and consistency of their sample covariance imply convergence of the studentized error vector and of its conditional Gaussian multiplier analogue to the same Gaussian law. Each marginal has unit variance, so the absolute maximum has a continuous distribution. Conditional quantile convergence gives simultaneous raw-grid coverage $1-\beta+o(1)$. The ordering, continuous extension and $[0,1]$ replacement argument from Corollary 3.7 give the asserted full-CDF coverage. All operations use the original private reports. □

E Proofs for Pointwise Inference

Proof of Theorem 3.6. Write $Y_{b,(r,j)} = Y_{i(b,r),j}$ for the bit from level r in cycle b. For a selected node $\nu \in \mathcal{D}(k_n)$, the corresponding report is Bernoulli with parameter $\theta_{n,\nu} = q_{\varepsilon_n} + a_{\varepsilon_n}\pi_{n,\nu}$. Therefore

$$\mathbb{E}\frac{Y_{b,\nu} - q_{\varepsilon_n}}{a_{\varepsilon_n}} = \pi_{n,\nu}. \tag{64}$$

The cells in $\mathcal{D}(k_n)$ are disjoint and have union $[0, x_n]$. Summing (64) gives

$$\mathbb{E}V_{b,k_n} = \sum_{\nu\in\mathcal{D}(k_n)} \pi_{n,\nu} = F(x_n).$$

The selected coordinates come from distinct level users and are independent. Thus

$$\mathrm{Var}(V_{b,k_n}) = \frac{1}{a_{\varepsilon_n}^2} \sum_{\nu\in\mathcal{D}(k_n)} \theta_{n,\nu}(1-\theta_{n,\nu}) = \sigma_n^2(x).$$

Independence across cycles proves (21).

For the Berry–Esseen bound, abbreviate $a_n = a_{\varepsilon_n}$ and define

$$U_{b,\nu} = \frac{Y_{b,\nu} - \theta_{n,\nu}}{a_n}, \qquad b = 1,\dots,m_n, \quad \nu \in \mathcal{D}(k_n).$$

The total variance of the sum is

$$B_n^2 := \sum_{b,\nu} \mathbb{E}U_{b,\nu}^2 = m_n a_n^{-2} A_n(x) = m_n \sigma_n^2(x). \tag{65}$$

For $Y \sim \mathrm{Bernoulli}(\theta)$,

$$\begin{aligned}\mathbb{E}\,|Y-\theta|^3 &= \theta(1-\theta)^3 + (1-\theta)\theta^3 \\ &= \theta(1-\theta)\{(1-\theta)^2 + \theta^2\} \\ &\leq \theta(1-\theta).\end{aligned}$$

Consequently,

$$\sum_{b,\nu} \mathbb{E}\,|U_{b,\nu}|^3 \leq m_n a_n^{-3} A_n(x). \tag{66}$$

Substitution of (65)–(66) into the classical inequality (55) gives

$$\sup_t \left| \mathbb{P}\left\{ \frac{\sum_{b,\nu} U_{b,\nu}}{B_n} \leq t \right\} - \Phi(t) \right| \leq C_{\mathrm{BE}} \frac{m_n a_n^{-3} A_n(x)}{\{m_n a_n^{-2} A_n(x)\}^{3/2}}$$

$$= \frac{C_{\mathrm{BE}}}{\sqrt{m_n A_n(x)}}. \tag{67}$$

The standardized sum in (67) is exactly the statistic in (22). This proves part (b), without any boundedness assumption on ε_n.

It remains to prove studentization. Define

$$X_{b,n} = \frac{V_{b,k_n} - F(x_n)}{\sigma_n(x)}.$$

For each n, these variables are i.i.d. across b, with mean zero and variance one. Let $S_{b,n} = V_{b,k_n} - F(x_n) = \sum_\nu U_{b,\nu}$. Independence and centering give the exact fourth-moment expansion

$$\mathbb{E}S_{b,n}^4 = \sum_\nu \mathbb{E}U_{b,\nu}^4 + 6\sum_{\nu<\mu} \mathbb{E}U_{b,\nu}^2 \mathbb{E}U_{b,\mu}^2. \tag{68}$$

The second sum runs over unordered pairs of distinct selected nodes. For a Bernoulli variable,

$$\begin{aligned}\mathbb{E}(Y-\theta)^4 &= \theta(1-\theta)^4 + (1-\theta)\theta^4 \\ &= \theta(1-\theta)\{(1-\theta)^3 + \theta^3\} \\ &\le \theta(1-\theta).\end{aligned} \tag{69}$$

Using (69) in (68), and bounding the cross term by three times the square of the total variance, gives

$$\mathbb{E}S_{b,n}^4 \le a_n^{-4} A_n(x) + 3\{a_n^{-2} A_n(x)\}^2.$$

Since $\sigma_n^2(x) = a_n^{-2} A_n(x)$,

$$\mathbb{E}X_{b,n}^4 \le 3 + \frac{1}{A_n(x)}.$$

Therefore

$$\begin{aligned}\mathrm{Var}\left\{\frac{1}{m_n}\sum_{b=1}^{m_n} X_{b,n}^2\right\} &\le \frac{\mathbb{E}X_{1,n}^4}{m_n} \\ &\le \frac{3}{m_n} + \frac{1}{m_n A_n(x)} \longrightarrow 0\end{aligned}$$

under (23). Hence $m_n^{-1}\sum_b X_{b,n}^2 \xrightarrow{\mathbb{P}} 1$. Also

$$\mathbb{E}\left(\frac{1}{m_n}\sum_{b=1}^{m_n} X_{b,n}\right)^2 = \frac{1}{m_n} \longrightarrow 0,$$

so $\overline{X}_n = m_n^{-1}\sum_b X_{b,n} \xrightarrow{\mathbb{P}} 0$. The exact sample-variance identity now yields

$$\begin{aligned}\frac{\widehat{\sigma}_n^2(x)}{\sigma_n^2(x)} &= \frac{1}{m_n - 1}\sum_{b=1}^{m_n}(X_{b,n} - \overline{X}_n)^2 \\ &= \frac{m_n}{m_n - 1}\left\{\frac{1}{m_n}\sum_{b=1}^{m_n} X_{b,n}^2 - \overline{X}_n^2\right\} \xrightarrow{\mathbb{P}} 1.\end{aligned} \tag{70}$$

Because $m_n A_n(x) \to \infty$, (22) gives the standardized normal limit. Slutsky's theorem and (70) prove (25).

Finally,

$$\begin{aligned}&\frac{\sqrt{m_n}\{\widetilde{F}_{m_n,D_n}(x_n) - F(x)\}}{\widehat{\sigma}_n(x)} \\ &= \frac{\sqrt{m_n}\{\widetilde{F}_{m_n,D_n}(x_n) - F(x_n)\}}{\widehat{\sigma}_n(x)} - \frac{\sqrt{m_n}\{F(x) - F(x_n)\}}{\sigma_n(x)}\frac{\sigma_n(x)}{\widehat{\sigma}_n(x)}.\end{aligned}$$

The first term converges to $N(0,1)$, and the second is $o_{\mathbb{P}}(1)$ under (26) and (24). This proves (27). □

Proof of Corollary B.4. For every selected node, $0 \le \pi_{n,\nu} \le 1$ and $a_{\varepsilon_n} = 1/2 - q_{\varepsilon_n}$, so

$$q_{\varepsilon_n} \le \theta_{n,\nu} \le 1/2.$$

The function $u(1-u)$ is increasing on $[0, 1/2]$. Therefore

$$\frac{q_{\varepsilon_n}}{2} \le q_{\varepsilon_n}(1 - q_{\varepsilon_n}) \le \theta_{n,\nu}(1 - \theta_{n,\nu}) \le \frac{1}{4}.$$

Summing over the $\ell_n(x)$ nodes proves (41). Since $a_{\varepsilon_n} = \rho_{\varepsilon_n}/2$, taking square roots after division by $a_{\varepsilon_n}^2 m_n$ proves (42). The lower bound in (41) shows that $m_n \ell_n(x) q_{\varepsilon_n} \to \infty$ implies $m_n A_n(x) \to \infty$, proving the public sufficient condition together with $m_n \to \infty$.

If $F \in \mathcal{F}(s, L)$, then $x_n \le x < x_n + D_n^{-1}$ and

$$0 \le F(x) - F(x_n) \le L D_n^{-s}.$$

Under $\varepsilon_n \le \overline{\varepsilon}$, $q_{\varepsilon_n} \ge (1 + e^{\overline{\varepsilon}})^{-1} > 0$. The lower standard-deviation bound gives

$$\begin{aligned} \frac{\sqrt{m_n}\{F(x) - F(x_n)\}}{\sigma_n(x)} &\le C_{L,\overline{\varepsilon}} D_n^{-s} \sqrt{\frac{m_n}{\ell_n(x)}} \\ &\le C_{L,\overline{\varepsilon}} D_n^{-s} \sqrt{\frac{n}{J_n \ell_n(x)}}, \end{aligned}$$

where the final inequality uses $m_n \le n/J_n$. Condition (43) proves (26).

In the same bounded-privacy regime, q_{ε_n} is bounded below and $m_n \asymp n/J_n$ whenever $J_n = o(n)$. The two inequalities in (42) then give (44). Finally, if $D_n \asymp n^{1/(2s)}$, then

$$D_n^{-s} \sqrt{\frac{n}{J_n \ell_n(x)}} \asymp \frac{1}{\sqrt{J_n \ell_n(x)}} \to 0$$

when $J_n \to \infty$, since $\ell_n(x) \ge 1$ eventually. □

Proof of Corollary B.3. Under the bias condition, (27) implies

$$\mathbb{P}\left\{ \left| \widetilde{F}_{m_n, D_n}(x_n) - F(x) \right| \le z_{1-\alpha/2} \frac{\widehat{\sigma}_n(x)}{\sqrt{m_n}} \right\} \longrightarrow 1 - \alpha,$$

which is exactly the coverage statement for the raw interval.

For the bias-aware interval, define

$$E_n = \left\{ \left| \widetilde{F}_{m_n, D_n}(x_n) - F(x_n) \right| \le z_{1-\alpha/2} \frac{\widehat{\sigma}_n(x)}{\sqrt{m_n}} \right\}.$$

By (25), $\mathbb{P}(E_n) \to 1 - \alpha$. On E_n, the lower endpoint in (11) is no larger than $F(x_n)$, which is no larger than $F(x)$. The upper endpoint is at least $F(x_n) + b_n(x)$, which is no smaller than $F(x)$ by (39). Hence $E_n \subseteq \{F(x) \in \mathcal{I}_{n,\alpha}^{\mathrm{bias}}(x)\}$. Taking probabilities and lower limits proves (40). Both intervals are deterministic functions of the same private transcript and public constants, so post-processing gives the privacy statement. □

E.1 Fixed-threshold benchmark proof

Proof of Proposition B.1. For output one, the likelihood ratio between inputs $B = 1$ and $B = 0$ is $p_\varepsilon / q_\varepsilon = e^\varepsilon$. For output zero, it is $(1 - p_\varepsilon)/(1 - q_\varepsilon) = e^{-\varepsilon}$. Hence the channel is ε-LDP.

Put $f_x = F(x)$. Marginally, $Z_i(x)$ is Bernoulli with parameter

$$\theta_\varepsilon(x) = p_\varepsilon f_x + q_\varepsilon(1 - f_x) = q_\varepsilon + \rho_\varepsilon f_x.$$

Therefore

$$\mathbb{E}\widehat{F}_n^{\mathrm{RR}}(x) = \frac{\theta_\varepsilon(x) - q_\varepsilon}{\rho_\varepsilon} = f_x.$$

Moreover,

$$\mathrm{Var}\{\widehat{F}_n^{\mathrm{RR}}(x)\} = \frac{\theta_\varepsilon(x)\{1-\theta_\varepsilon(x)\}}{n\rho_\varepsilon^2}. \tag{71}$$

Since $p_\varepsilon = 1 - q_\varepsilon$ and $p_\varepsilon - q_\varepsilon = \rho_\varepsilon$,

$$\begin{aligned}\theta_\varepsilon(x)\{1-\theta_\varepsilon(x)\} &= (q_\varepsilon + \rho_\varepsilon f_x)(p_\varepsilon - \rho_\varepsilon f_x) \\ &= p_\varepsilon q_\varepsilon + \rho_\varepsilon^2 f_x(1-f_x).\end{aligned} \tag{72}$$

Finally,

$$\frac{p_\varepsilon q_\varepsilon}{\rho_\varepsilon^2} = \frac{e^\varepsilon}{(e^\varepsilon-1)^2}.$$

Combining this identity with (71)–(72) proves (35). For fixed ε, the ordinary central limit theorem for bounded Bernoulli variables, followed by the affine decoding in (34), proves (36).

We now prove the lower bound. Let

$$w = \min(x, 1-x)/2,$$

and define disjoint intervals

$$A = [x-w, x], \qquad B = (x, x+w].$$

Set $h = \mathbf{1}_A - \mathbf{1}_B$. For $0 < \delta \le w$, let $\eta = \delta/(2w) \le 1/2$ and define

$$f_+(u) = 1 + \eta h(u), \qquad f_-(u) = 1 - \eta h(u).$$

Both densities integrate to one and take values in $[1/2, 3/2]$. Their CDFs satisfy

$$F_+(x) - F_-(x) = 2\eta w = \delta, \tag{73}$$

and

$$\mathrm{TV}(P_+, P_-) = \frac{1}{2}\int |f_+ - f_-| = \eta \int |h| = 2\eta w = \delta. \tag{74}$$

First suppose $0 < \varepsilon \le 1$. Choose

$$\delta = \frac{c_1}{\rho_\varepsilon \sqrt{n}}$$

for a sufficiently small universal $c_1 > 0$. The condition $n\rho_\varepsilon^2 \ge (c_1/w)^2$ ensures $\delta \le w$. By Lemma C.6 and (74), every sequentially interactive transcript pair obeys

$$\mathrm{KL}(M_+^{(n)} \| M_-^{(n)}) \le C_{\mathrm{LDP}} n \rho_\varepsilon^2 \delta^2 = C_{\mathrm{LDP}} c_1^2.$$

Pinsker's inequality therefore gives

$$\mathrm{TV}(M_+^{(n)}, M_-^{(n)}) \le \sqrt{C_{\mathrm{LDP}}/2}\, c_1.$$

Now suppose $\varepsilon > 1$. The complete transcript is a Markov kernel of the raw sample, so data processing gives

$$\mathrm{KL}(M_+^{(n)} \| M_-^{(n)}) \le n\, \mathrm{KL}(P_+ \| P_-). \tag{75}$$

Using $\mathrm{KL}(P_+\|P_-) \le \chi^2(P_+, P_-)$, where χ^2 is Pearson divergence, and $f_- \ge 1/2$,

$$\begin{aligned}\mathrm{KL}(P_+\|P_-) &\le \int \frac{(f_+ - f_-)^2}{f_-} \\ &\le 8\eta^2 \int h^2 = 16\eta^2 w = \frac{4\delta^2}{w}.\end{aligned} \tag{76}$$

Take $\delta = c_2\sqrt{w}/\sqrt{n}$ with a sufficiently small universal c_2. Then (75)–(76) are bounded by $4c_2^2$, and $\delta \le w$ once $n \ge c_2^2/w$. Pinsker again makes the transcript total variation uniformly smaller than one.

For either regime, any estimator T induces a test that selects the closer of $F_+(x)$ and $F_-(x)$. By (73), that test is correct whenever $|T - F_\pm(x)| < \delta/2$. Le Cam's two-point inequality therefore yields

$$\sup_{\nu\in\{+,-\}} \mathbb{P}_\nu\{|T - F_\nu(x)| \ge \delta/2\} \ge \frac{1 - \mathrm{TV}(M_+^{(n)}, M_-^{(n)})}{2}.$$

Choose c_1, c_2 so that the right side is at least a universal $c_0 > 0$. In the small-privacy regime, $\delta/2 = c_1/(2\rho_\varepsilon\sqrt{n})$. In the large-privacy regime, $\rho_\varepsilon \geq \rho_1 := \tanh(1/2)$, and therefore

$$\frac{c_2\rho_1\sqrt{w}}{2\rho_\varepsilon\sqrt{n}} \leq \frac{c_2\sqrt{w}}{2\sqrt{n}} = \frac{\delta}{2}.$$

Taking

$$c_x = \min\{c_1/2, c_2\rho_1\sqrt{w}/2\}$$

and choosing C_x to imply both feasibility conditions proves (37) after taking the infimum over protocols and estimators. □

F Proofs for Monotone Post-Processing

Proof of Proposition B.5. Let

$$E_n = \left\{ \max_{0\leq j\leq D_n} |g_{n,j} - f_{n,j}| \leq \eta_n \right\}.$$

By assumption, $\mathbb{P}(E_n) \to 1$. On E_n, every $j \leq k_n - 1$ satisfies

$$\begin{aligned} g_{n,j} &\leq f_{n,j} + \eta_n \leq f_{n,k_n-1} + \eta_n \\ &< f_{n,k_n} - \eta_n \leq g_{n,k_n}, \end{aligned}$$

where the strict inequality uses the second condition in (45). Thus the running maximum through index k_n is attained at k_n. The first condition in (45) gives, again on E_n,

$$0 < f_{n,k_n} - \eta_n \leq g_{n,k_n} \leq f_{n,k_n} + \eta_n < 1.$$

The clamp in (5) is therefore inactive at the target. Hence $\widehat{F}_{m_n,D_n}(x_{k_n}) = g_{n,k_n}$ on E_n. Taking probabilities proves (46). Equality with probability tending to one transfers any weak limit of the raw target by Slutsky's theorem. □

Proposition F.1. *Fix $\varepsilon > 0$ and let $F(u) = u$ on $[0,1]$. Let $D_n = 2^{J_n}$ and $N_n = m_n J_n$, and suppose*

$$J_n \to \infty, \qquad 0 < c_D \leq \frac{D_n}{\sqrt{N_n}} \leq C_D < \infty. \tag{77}$$

Set

$$s_{A,n}^2 = \mathrm{Var}\{\widetilde{F}_{m_n,D_n}(1/2)\}.$$

Then, for every fixed $M > 0$,

$$\liminf_{n\to\infty} \mathbb{P}\left\{ \frac{\widehat{F}_{m_n,D_n}(1/2) - \widetilde{F}_{m_n,D_n}(1/2)}{s_{A,n}} > M \right\} \geq \frac{1}{2}. \tag{78}$$

Proof of Proposition F.1. Let $k_n = D_n/2$. Its J_n-digit binary representation is $10\cdots0$, so the canonical decomposition of $[0, 1/2]$ contains only the level-1 cell. The predecessor $k_n - 1$ has binary representation $01\cdots1$, so its canonical decomposition contains one cell at each level $r = 2, \ldots, J_n$. Define

$$A_n = \widetilde{F}_{m_n,D_n}(1/2), \qquad B_n = \widetilde{F}_{m_n,D_n}(1/2 - 1/D_n).$$

The two estimators use disjoint level assignments and are independent.

Under the uniform distribution, a selected level-r cell has probability 2^{-r}. Its OUE success probability and decoded variance are

$$\theta_r = q_\varepsilon + a_\varepsilon 2^{-r}, \qquad v_r = a_\varepsilon^{-2}\theta_r(1-\theta_r).$$

Because ε is fixed and $q_\varepsilon \leq \theta_r \leq 1/2$, there are finite positive constants $c_\varepsilon, C_\varepsilon$ such that

$$c_\varepsilon \leq v_r \leq C_\varepsilon \qquad \text{for every } r.$$

Exact independence across cycles gives

$$s_{A,n}^2 = \mathrm{Var}(A_n) = \frac{v_1}{m_n}, \qquad s_{B,n}^2 := \mathrm{Var}(B_n) = \frac{1}{m_n}\sum_{r=2}^{J_n} v_r. \tag{79}$$

Consequently,

$$\frac{s_{B,n}}{s_{A,n}} \asymp \sqrt{J_n}, \qquad \tau_n^2 := \mathrm{Var}(B_n - A_n) = s_{A,n}^2 + s_{B,n}^2 \asymp \frac{J_n}{m_n}. \tag{80}$$

Unbiasedness gives

$$\mathbb{E}A_n = \frac{1}{2}, \qquad \mathbb{E}B_n = \frac{1}{2} - \frac{1}{D_n}. \tag{81}$$

The centered difference $B_n - A_n + D_n^{-1}$ is a sum of $m_n J_n = N_n$ independent centered Bernoulli coordinates after deterministic scaling and signs. Under fixed ε, every coordinate variance is bounded above and below by positive constants before division by m_n^2, and the third absolute moment is bounded by a constant multiple of its variance. The same calculation as in (67) therefore gives the Berry–Esseen bound

$$\sup_t \left| \mathbb{P}\left\{ \frac{B_n - A_n + D_n^{-1}}{\tau_n} \le t \right\} - \Phi(t) \right| \le \frac{C_\varepsilon}{\sqrt{m_n J_n}} = \frac{C_\varepsilon}{\sqrt{N_n}} \to 0.$$

Thus

$$\frac{B_n - A_n + D_n^{-1}}{\tau_n} \rightsquigarrow N(0,1). \tag{82}$$

The scaling condition (77) gives $N_n \asymp D_n^2$ and $m_n = N_n/J_n$. Using (80),

$$\frac{D_n^{-1}}{\tau_n} = O(J_n^{-1}), \qquad \frac{M s_{A,n}}{\tau_n} = O(M J_n^{-1/2}). \tag{83}$$

Both terms vanish. Equations (82)–(83) imply

$$\mathbb{P}(B_n - A_n > M s_{A,n}) \longrightarrow \frac{1}{2}. \tag{84}$$

It remains to control the upper clamp. By (79) and $N_n \asymp D_n^2$,

$$\mathrm{Var}(B_n) \asymp \frac{J_n}{m_n} = \frac{J_n^2}{N_n} \asymp \frac{J_n^2}{D_n^2} \longrightarrow 0.$$

Together with (81), this gives $B_n \xrightarrow{\mathbb{P}} 1/2$ and hence $\mathbb{P}(B_n \le 1) \to 1$. At index k_n, the cumulative maximum includes the predecessor value B_n. On $\{B_n \le 1\}$, the clamped running maximum is at least B_n; this remains true when $B_n < 0$ because the lower clamp is zero. Therefore

$$\widehat{F}_{m_n,D_n}(1/2) - A_n \ge B_n - A_n \quad \text{on } \{B_n \le 1\}.$$

Using (84) and $\mathbb{P}(B_n \le 1) \to 1$ proves (78). □

G Proofs for the Interactive Lower Bound

Proof of Theorem 3.4. Fix $0 < \delta < 1/16$ and let

$$M = \lfloor 1/(4\delta) \rfloor .$$

Choose M disjoint intervals $A_1, \dots, A_M$ of length 4δ in $[0,1]$. Split each interval into its left and right halves $A_{j,L}$ and $A_{j,R}$, each of length 2δ, and define

$$f_0(u) = 1, \qquad f_j(u) = 1 + \frac{1}{2}\mathbf{1}_{A_{j,L}}(u) - \frac{1}{2}\mathbf{1}_{A_{j,R}}(u), \quad j = 1, \dots, M.$$

The positive and negative perturbations have equal integral. Every f_j therefore integrates to one and takes values in $[1/2, 3/2]$; the corresponding CDFs belong to $\mathcal{F}_{\mathrm{bd}}$.

Within A_j, $F_j - F_0$ rises linearly from zero to $(1/2)(2\delta) = \delta$ at the midpoint, and then returns linearly to zero. It vanishes outside A_j. Consequently,

$$\|F_j - F_0\|_\infty = \delta, \qquad \|F_j - F_k\|_\infty \ge \delta \quad (j \ne k).$$

The total variation distance from the baseline is

$$\mathrm{TV}(P_j, P_0) = \frac{1}{2}\int |f_j - 1| = \frac{1}{2}\left(\frac{1}{2}\right)(4\delta) = \delta.$$

Fix an arbitrary sequentially interactive ε-LDP protocol and let $M_j^{(n)}$ be its transcript law under $P_j^{\otimes n}$. By Lemma C.6, for $0 < \varepsilon \le 1$,

$$\mathrm{KL}(M_j^{(n)} \| M_0^{(n)}) \le C_{\mathrm{LDP}} n \rho_\varepsilon^2 \delta^2 = C_{\mathrm{LDP}} N_\varepsilon \delta^2.$$

Let J^* be uniform on $\{1, \dots, M\}$ and let Z denote the transcript. Their mutual information $I(J^*; Z)$ satisfies the information-radius inequality, using $M_0^{(n)}$ as the reference measure,

$$I(J^*; Z) \le \frac{1}{M}\sum_{j=1}^{M} \mathrm{KL}(M_j^{(n)} \| M_0^{(n)}) \le C_{\mathrm{LDP}} N_\varepsilon \delta^2. \tag{85}$$

Given an arbitrary CDF estimator $\widehat{F}$, define the minimum-distance decoder

$$\widehat{J} \in \arg\min_{1 \le j \le M} \left\| \widehat{F} - F_j \right\|_\infty,$$

with deterministic tie breaking. If the true index is j and $\left\| \widehat{F} - F_j \right\|_\infty < \delta/2$, then for every $k \ne j$,

$$\begin{aligned} \left\| \widehat{F} - F_k \right\|_\infty &\ge \|F_j - F_k\|_\infty - \left\| \widehat{F} - F_j \right\|_\infty \\ &> \delta/2 > \left\| \widehat{F} - F_j \right\|_\infty. \end{aligned}$$

Thus $\widehat{J} = j$, and

$$\mathbb{P}_{F_j}\{ \left\| \widehat{F} - F_j \right\|_\infty \ge \delta/2 \} \ge \mathbb{P}_{F_j}(\widehat{J} \ne j). \tag{86}$$

Fano's inequality (Tsybakov, 2009), combined with (85), gives

$$\frac{1}{M}\sum_{j=1}^{M} \mathbb{P}_{F_j}(\widehat{J} \ne j) \ge 1 - \frac{C_{\mathrm{LDP}} N_\varepsilon \delta^2 + \log 2}{\log M}. \tag{87}$$

Choose

$$\delta = c_* \sqrt{\frac{\log N_\varepsilon}{N_\varepsilon}} \tag{88}$$

with a sufficiently small universal $c_* > 0$. For sufficiently large N_ε, this choice is smaller than $1/16$ and

$$\log M \ge \log \frac{1}{8\delta} \ge \frac{1}{4}\log N_\varepsilon. \tag{89}$$

The information term in (87) is $C_{\mathrm{LDP}} c_*^2 \log N_\varepsilon$. Choose c_* so that its ratio to (89) is at most $1/4$, and then increase N_0 so that $\log 2 / \log M \le 1/4$. The right side of (87) is then bounded below by a universal positive constant. Equations (86) and (88), followed by a supremum over the packing and an infimum over all protocols and estimators, prove (19) after renaming constants. □

Proof of Corollary 3.5. The probability lower bound in Theorem 3.4 implies the expected lower bound by

$$\mathbb{E}_F \left\| \widehat{F} - F \right\|_\infty \ge c_0 c \sqrt{\frac{\log(n\rho_\varepsilon^2)}{n\rho_\varepsilon^2}}$$

for at least one $F \in \mathcal{F}_{\mathrm{bd}}$. Taking the infimum gives the left side of (20).

For the upper bound, note that $\mathcal{F}_{\mathrm{bd}} \subset \mathcal{F}(1, 3/2)$. Take D to be the smallest power of two not smaller than $\sqrt{n}$. Then $J \asymp \log(en)$, $m \asymp n/J$, and $\omega_D(F) \le (3/2)D^{-1} = O(n^{-1/2})$. The expectation bound (16), together with $a_\varepsilon = \rho_\varepsilon/2$, gives

$$\begin{aligned}\sup_{F \in \mathcal{F}_{\mathrm{bd}}} \mathbb{E}_F \left\| \widehat{F}_{m,D} - F \right\|_\infty &\le C \frac{\{\log(en)\}^{3/2}}{\rho_\varepsilon \sqrt{n}} + C n^{-1/2} \\ &\le C' \frac{\{\log(en)\}^{3/2}}{\rho_\varepsilon \sqrt{n}}.\end{aligned}$$

TreeOUE is noninteractive and therefore belongs to the larger class of sequentially interactive protocols over which the minimax infimum is taken. This proves the right side. □