DIFFERENTIALLY PRIVATE QUANTILE INFERENCE: PHASE TRANSITIONS AND A GAUSSIANIZATION MECHANISM

Anonymous authors
Paper under double-blind review

ABSTRACT

Private quantile inference requires the distribution of sampling error together with the randomization used for privacy. We characterize this distribution for the rank exponential mechanism under local density regularity and a public bounded output interval. As $\epsilon_n \to 0$ with $n\epsilon_n \to \infty$, averaging its length-weighted order-statistic gaps yields a conditional Laplace limit independent of the sample fluctuation. This identifies the Gaussian, Gaussian–Laplace, and Laplace regimes, with privacy scale $1/(n\epsilon_n)$ and no logarithmic factor. Neighboring private quantiles estimate the common local scale, giving feasible intervals across regimes. We then change the error distribution through a one-step correction with Gaussian noise added to a bounded quantile score. Under explicit pilot and nuisance conditions, this construction under approximate differential privacy has a Gaussian pivot at every relative noise level. Simulations illustrate the phase laws and show how public range and budget allocation determine the width of feasible rank-based intervals, including the tradeoff between center accuracy and scale estimation.

1 INTRODUCTION

For subgroup income medians, Drechsler et al. (2022) study private confidence intervals using samples from the 1940 U.S. Census. Such intervals must account for sampling fluctuation and the privacy mechanism's randomization. When these have the same order, an error bound alone does not specify a critical value, and variance inflation can miss the shape of the inferential distribution. We study this problem under central differential privacy (DP), where a trusted curator observes the sample and releases a private estimate and interval.

Quantiles expose a tension between worst-case and statistical stability. Replacing one record can move a sample quantile across a large gap, whereas near a regular population quantile q_τ at level τ the typical spacing is of order $1/\{nf(q_\tau)\}$, with f the density. Smooth sensitivity, propose–test–release, and rank mechanisms exploit forms of this local stability (Nissim et al., 2007; Dwork & Lei, 2009; Smith, 2011). The rank exponential mechanism, QExp, is particularly simple: it selects an order-statistic gap using its length and an exponential penalty for its distance in rank from the target, then samples uniformly within that gap. Its random gap lengths are also the obstacle to inference. At finite sample sizes, the perturbation is not an independent additive noise with a known distribution.

Existing guarantees illuminate parts of this problem. For the median, smoothed inverse sensitivity is asymptotically equivalent to the empirical median when $\varepsilon_n\sqrt{n}/\log n \to \infty$ (Asi & Duchi, 2020b). For QExp, Lalanne et al. (2023a) derive distribution-level tail bounds; their stated bound uses global density bounds, and their Remark 3.6 discusses localizing the lower-bound assumption. Such bounds control value error but do not identify its first-order distribution when privacy noise persists. Sharp private-median sample complexity (Tzamos et al., 2020) and distribution-free private endpoint intervals (Drechsler et al., 2022) address accuracy and coverage directly. The black-box subsampling theory of Shoham et al. (2026) allows general limiting distributions but requires mechanism error to be asymptotically smaller than sampling error. Here we resolve the local distribution when that requirement fails, including the critical budget $\varepsilon_n \asymp n^{-1/2}$.

Our key observation is that a shrinking budget smooths the random gap geometry. When $\varepsilon_n \to 0$ and $n\varepsilon_n \to \infty$, QExp's effective rank window contains a diverging number of spacings but

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.

remains local to the quantile. Their length-weighted average converges to a Laplace density. We prove this conditionally on the sample, which also establishes asymptotic independence from the classical quantile fluctuation. Consequently, QExp preserves the Gaussian sampling law as soon as $\varepsilon_n\sqrt{n} \to \infty$; at the boundary it has a Gaussian–Laplace convolution; under stronger privacy its error is asymptotically Laplace. This removes the logarithmic factor in the pointwise stochastic privacy scale and identifies the distribution needed for calibration, under local continuity and positivity of the density and without moment assumptions.

The same sparsity, $1/f(q_\tau)$, converts both leading errors to the quantile scale. Two neighboring private quantiles consistently estimate it, and the known center budget determines the convolution critical value. This gives a feasible interval without selecting a regime from the data. We account for every release, using either a same-sample budget split or disjoint blocks. Both constructions can retain the center's first-order law while spending an asymptotically vanishing fraction of the budget or observations on scale estimation.

The phase transition also suggests how to control, rather than merely calibrate, the limiting shape. We release a bounded quantile score at an independent private pilot estimate and take a one-step correction. With Gaussian score noise, the leading sampling and privacy components are both Gaussian. Explicit pilot and nuisance conditions make the remaining terms negligible, yielding a standard-normal pivot at any relative noise level under approximate differential privacy. A pure-DP Laplace score variant halves QExp's critical privacy coefficient under the corresponding conditions. These are distributional benefits of the score construction; their preliminary estimation costs and different privacy guarantees determine which procedure is appropriate.

The rest of the paper is organized as follows. Section 2 defines QExp, Section 3 derives its phase laws, and Section 4 constructs feasible intervals. Section 5 develops score-based inference; Section 6 examines calibration, output range, and budget allocation; and Section 7 concludes. The appendices provide implementation details, additional experiments, joint limits, a regular-class risk lower bound, and proofs.

2 Quantiles and the rank mechanism

Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) with distribution function F. Fix $\tau \in (0,1)$ and define

$$q_\tau = F^{-1}(\tau) := \inf\{x : F(x) \geq \tau\}.$$

Assumption 2.1. *There is a public finite output interval $[A, B]$ such that $q_\tau \in (A, B)$. The quantile q_τ is unique. On an open neighborhood of q_τ, F is absolutely continuous with density f, and*

$$f(x) \longrightarrow f_\tau := f(q_\tau) \in (0, \infty) \qquad \text{as } x \to q_\tau.$$

The observations may be unbounded; $[A, B]$ bounds the output and contains the target, rather than the sample. Only local absolute continuity and density continuity at the target enter the expansion, so atoms elsewhere and heavy tails are allowed. Without a public interval containing the quantile, a private localization step is needed; see Section 7.

Local positivity and continuity make $Q = F^{-1}$ differentiable at τ, with derivative $1/f_\tau$. Write $X_{(1)} \leq \cdots \leq X_{(n)}$, $X_{(0)} = -\infty$, $X_{(n+1)} = \infty$, $k_n = \lfloor n\tau \rfloor$, and $\widehat{q}_{\tau,n} = X_{(k_n)}$.

Datasets are adjacent when one record is replaced. Pure ε-DP requires $\mathbb{P}\{M(D) \in S\} \leq e^\varepsilon \mathbb{P}\{M(D') \in S\}$ for every adjacent pair and measurable S; approximate (ε, δ)-DP permits an additional δ on the right. For $N_D(y) = \sum_i \mathbf{1}\{X_i < y\}$, the exponential mechanism (McSherry & Talwar, 2007) has density

$$p_D(y) = \frac{\exp\{-\varepsilon_n |N_D(y) - k_n|/2\}}{Z_n}, \qquad Z_n = \int_A^B \exp\{-\varepsilon_n |N_D(z) - k_n|/2\}\, dz. \tag{1}$$

The score sensitivity is one, so QExp is pure ε_n-DP. The exact gap probabilities and privacy proof are in Proposition A.3 and Appendix G.2.

The procedure is total whenever $A < B$: zero-length gaps receive zero weight, and at least one clipped gap has positive length. Sorting costs $O(n \log n)$; computing the weights and drawing a release then costs $O(n)$.

Algorithm 1 Exact rank-exponential quantile release

Require: Data $D = (X_1, \dots, X_n)$, level τ, public interval $[A, B]$, budget ε_n.
1: Sort the sample and form the clipped gaps $I_{j,n} = [A, B] \cap [X_{(j)}, X_{(j+1)}]$, $j = 0, \dots, n$.
2: Set $k_n = \lfloor n\tau \rfloor$ and $w_{j,n} = \mathrm{Leb}(I_{j,n}) \exp\{-\varepsilon_n |j - k_n|/2\}$.
3: Draw J_n from the normalized weights $w_{0,n}, \dots, w_{n,n}$.
4: Draw and return the private quantile $\widetilde{q}_{\tau,n}$ uniformly from $I_{J_n,n}$.

The public range enters through the clipped gap lengths: a long gap can retain appreciable probability even when its rank is penalized. The conditional limit below explains how the mechanism localizes around the empirical quantile as $n\varepsilon_n$ grows; the numerical study examines the range's effect at finite sample sizes.

3 From gap randomization to the inferential law

Write $\widetilde{q}_{\tau,n} - q_\tau = (\widehat{q}_{\tau,n} - q_\tau) + (\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n})$. The first term is the empirical quantile fluctuation. The second is QExp's randomization conditional on $D_n = (X_1, \dots, X_n)$. If the budget is bounded away from zero, this perturbation is $O_{\mathbb{P}}(n^{-1})$, so privacy leaves the first-order sampling law unchanged (Theorem G.13). The shrinking-budget case identifies the perturbation itself. Let $L \sim \mathrm{Lap}(0, 1)$ have density $e^{-|t|}/2$ and set $T_n = n\varepsilon_n f_\tau(\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n})/2$.

Theorem 3.1 (Conditional local Laplace limit)**.** *Suppose Assumption 2.1 holds and $\varepsilon_n \to 0, n\varepsilon_n \to \infty$. Then, conditionally on the data,*

$$\mathcal{L}\left(\frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n}) \,\middle|\, D_n \right) \rightsquigarrow \mathrm{Lap}(0, 1) \quad \textit{in probability.}$$

Equivalently, for every bounded continuous h,

$$\mathbb{E}\{h(T_n) \mid D_n\} \longrightarrow_{\mathbb{P}} \frac{1}{2} \int_{\mathbb{R}} h(t) e^{-|t|}\, dt.$$

The convergence is stable with respect to the sample sigma-field.

Representing uniform spacings by normalized exponentials turns the exact gap weights into Riemann sums against $e^{-|t|}\, dt$, with limiting normalizer two. The rank window of order $1/\varepsilon_n$ contains increasingly many gaps but stays local because $n\varepsilon_n \to \infty$. Local differentiability of Q supplies the spatial scaling; the nonrandom conditional limit gives independence from convergent sample statistics. Appendices G.5 and G.6 provide the spacing and stable-convergence arguments.

In particular, the classical expansion gives $\sqrt{n} f_\tau(\widehat{q}_{\tau,n} - q_\tau) \rightsquigarrow G_\tau$, where $G_\tau \sim N\{0, \tau(1-\tau)\}$, independently of L in the joint limit (Theorem G.17). Comparing the sampling scale $n^{-1/2}$ with the privacy scale $(n\varepsilon_n)^{-1}$ yields the following transition.

Theorem 3.2 (Three-regime phase transition)**.** *Assume Assumption 2.1.*

(a) If $\varepsilon_n \sqrt{n} \to \infty$, then $\sqrt{n} f_\tau(\widetilde{q}_{\tau,n} - q_\tau) \rightsquigarrow G_\tau$.

(b) If $\varepsilon_n \sqrt{n} \to \kappa \in (0, \infty)$, then

$$\sqrt{n} f_\tau(\widetilde{q}_{\tau,n} - q_\tau) \rightsquigarrow G_\tau + \frac{2}{\kappa} L. \tag{2}$$

(c) If $\varepsilon_n \to 0$, $n\varepsilon_n \to \infty$, and $\varepsilon_n \sqrt{n} \to 0$, then $n\varepsilon_n f_\tau(\widetilde{q}_{\tau,n} - q_\tau) \rightsquigarrow 2L$.

At the critical budget, the limit in (2) has variance $\tau(1-\tau) + 8/\kappa^2$ and convolution tails. Figure 1 isolates their shape by dividing every limiting variable by its own standard deviation. For the two privacy-heavier curves shown, the Gaussian critical value 1.96 is too small for the standardized mixed law. As sampling dominates, the curves approach the standard-normal tail. Calibrating the convolution uses its non-Gaussian critical value; Gaussianizing the mechanism changes the leading randomization.

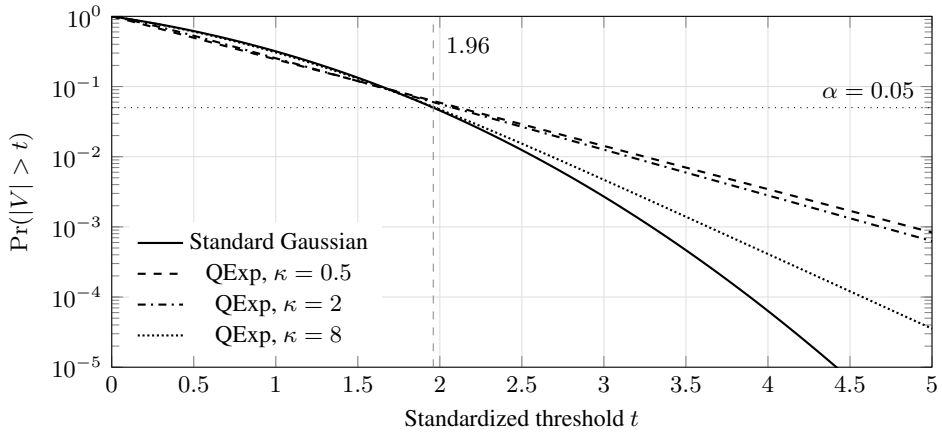


Figure 1: Theoretical tails at $\tau = 1/2$, standardized to variance one. For QExp, $V = (G_{1/2} + 2L/\kappa)/\sqrt{1/4 + 8/\kappa^2}$; the Gaussian curve is the standard-normal pivot after Gaussianization under its remainder conditions. The vertical line marks the Gaussian 95% cutoff 1.96.

The standardized tail probabilities at 1.96 are approximately 0.062, 0.060, and 0.051 for $\kappa = 0.5, 2, 8$. Thus equal variance still leaves a consequential tail difference when the privacy component is large. These probabilities concern the limiting family: finite-sample coverage additionally depends on localization of the released center and estimation of its scale. The numerical study separates those two effects.

The same local analysis gives the following rate, including the neighboring levels needed to estimate the scale.

Corollary 3.3. *Under Assumption 2.1, if $n\varepsilon_n \to \infty$, then*

$$\widetilde{q}_{\tau,n} - q_\tau = O_{\mathbb{P}}\Big(n^{-1/2} + (n\varepsilon_n)^{-1}\Big). \tag{3}$$

The same statement holds for triangular quantile levels $\tau_n \to \tau$, with q_{τ_n} in place of q_τ.

For the median, the Gaussian regime requires $\varepsilon_n\sqrt{n} \to \infty$, compared with the sufficient condition $\varepsilon_n\sqrt{n}/\log n \to \infty$ established for smoothed inverse sensitivity in Asi & Duchi (2020b). The class-uniform expected absolute-risk lower bound in Theorem C.9 contains the corresponding scales $(\ell\sqrt{n})^{-1} \vee (\ell n\varepsilon)^{-1}$ for local density lower bound ℓ under its explicit class and embedding conditions. The two results describe stochastic error at a fixed distribution and unavoidable expected loss over a class, respectively.

4 Feasible intervals from neighboring private quantiles

Fix a confidence level $1 - \alpha$, with $0 < \alpha < 1$. Write Φ for the standard normal cumulative distribution function, $z_p = \Phi^{-1}(p)$, and Π_I for projection onto an interval I. The quantile derivative $Q'(\tau) = 1/f_\tau$ is the common scale in Theorem 3.2. As in classical quantile studentization (Hall & Sheather, 1988), it can be estimated by a local quantile slope, now using private releases:

$$\frac{Q(\tau + h) - Q(\tau - h)}{2h} \longrightarrow \frac{1}{f_\tau}.$$

For $\kappa > 0$, define the public critical value

$$c_{Q,1-\alpha}(\tau, \kappa) = \inf\left\{c : \mathbb{P}\left(\left|G_\tau + \frac{2}{\kappa}L\right| \le c\right) \ge 1 - \alpha\right\}, \tag{4}$$

where G_τ and L are independent. As $\kappa \to \infty$ it tends to $z_{1-\alpha/2}\sqrt{\tau(1-\tau)}$; as $\kappa \downarrow 0$, its product with κ tends to $2\log(1/\alpha)$. Thus the same critical-value family handles all three regimes. Its one-dimensional integral and numerical computation are given in Appendix A.3.

Algorithm 2 QExp-SC: same-sample composition

Require: Data, target τ, public range $[A, B]$, confidence $1-\alpha$, bandwidth $0 < h_n < \min(\tau, 1-\tau)$, and positive budgets $\varepsilon_{0,n} + \varepsilon_{+,n} + \varepsilon_{-,n} \le E_n$.

Require: Public truncation sequences $c_n \downarrow 0$ and $C_n \uparrow \infty$.

1: Run exact QExp on the same sample at τ, $\tau + h_n$, and $\tau - h_n$ with the respective budgets, obtaining $\widetilde{q}_{0,n}, \widetilde{q}_{+,n}, \widetilde{q}_{-,n}$.

2: Estimate sparsity and density:

$$\widetilde{s}_n^{\mathrm{raw}} = \frac{\widetilde{q}_{+,n} - \widetilde{q}_{-,n}}{2h_n}, \qquad \widetilde{s}_n = \Pi_{[c_n, C_n]}(\widetilde{s}_n^{\mathrm{raw}}), \qquad \widetilde{f}_n = \widetilde{s}_n^{-1}. \tag{5}$$

3: Form the interval

$$\mathcal{C}_n^{Q,\mathrm{SC}} = \left[\widetilde{q}_{0,n} \ \pm \ \frac{c_{Q,1-\alpha}(\tau, \varepsilon_{0,n}\sqrt{n})}{\sqrt{n}\,\widetilde{f}_n} \right], \tag{6}$$

4: Release either the full tuple or only the center estimate and interval.

Positive truncation keeps the procedure defined even when the noisy neighboring quantiles cross. A range-based choice is $c_n = (B-A)/n, C_n = (B-A)/(2h_n)$,

$$h_n = \min\left\{ \frac{\min(\tau, 1-\tau)}{2}, [\min\{\sqrt{n}, n\min(\varepsilon_{+,n}, \varepsilon_{-,n})\}]^{-1/3} \right\}.$$

Both sparsity bounds scale with changes of measurement units. The upper bound never shortens a positive raw slope, since both released quantiles lie in $[A, B]$. For a thrice continuously differentiable Q, the bandwidth balances the $O(h_n^2)$ secant bias against a release-error bound divided by h_n; coverage itself only needs the limits below.

Theorem 4.1. *Suppose Assumption 2.1 holds, $n\varepsilon_{0,n} \to \infty$, and*

$$h_n \to 0, \sqrt{n}\, h_n \to \infty, nh_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty.$$

Then $\widetilde{f}_n \to_{\mathbb{P}} f_\tau$ and $\mathbb{P}\{q_\tau \in \mathcal{C}_n^{Q,\mathrm{SC}}\} \longrightarrow 1-\alpha$. This conclusion holds without assuming that $\varepsilon_{0,n}\sqrt{n}$ converges: the critical value in (6) *automatically specializes to the Gaussian, Gaussian–Laplace, or privacy-dominated limit along every subsequence.*

Here $h_n \to 0$ removes slope bias, $\sqrt{n}h_n \to \infty$ controls sampling error relative to the level separation, and $nh_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty$ controls the two privacy errors. Scale consistency is sufficient even though all three releases use the same records; independence between the estimated center and scale is unnecessary. The complete release is pure E_n-DP by sequential composition, with coverage proved in Appendix G.12.

The first-order interval length is $2c_{Q,1-\alpha}(\tau, \varepsilon_{0,n}\sqrt{n})/(\sqrt{n}f_\tau)$: it reduces to $2z_{1-\alpha/2}\sqrt{\tau(1-\tau)}/(\sqrt{n}f_\tau)$ in the Gaussian regime and $4\log(1/\alpha)/(n\varepsilon_{0,n}f_\tau)$ in the privacy-dominated regime. The budget determines the standardized law, while sparsity determines its scale; poor scale estimation can therefore inflate width even when center calibration is accurate.

The auxiliary privacy fraction can vanish. For $nE_n \to \infty$, take $\beta_n = \min\{1/4, (nE_n)^{-1/2}\}$, side budgets $\beta_n E_n$, center budget $(1-2\beta_n)E_n$, and the bandwidth in Corollary G.28. Then $\varepsilon_{0,n}/E_n \to 1$ and coverage remains valid. Alternatively, QExp-PC (parallel composition) uses deterministic disjoint center and nuisance blocks. The total budget is the maximum of the center budget and the sum of the two nuisance budgets. A nuisance block of size $o(n)$ can suffice, leaving the first-order sampling efficiency unchanged (Theorem C.1 and Lemma G.30). These constructions make scale estimation asymptotically inexpensive. The finite-sample choice depends on how accurately the side quantiles localize, as the budget experiment below shows.

5 Gaussianization through a private quantile score

QExp calibration uses the distribution generated by rank randomization. To change that distribution, we instead randomize a bounded score whose noise law can be chosen directly. Gaussian score noise

is the mechanism behind Gaussianization; a Laplace alternative retains pure DP. The construction uses disjoint pilot, nuisance, and score blocks of sizes m_p, m_f, N, with score index set $\mathcal{I}_s$. Release an initial QExp estimate (the pilot) $\widetilde{q}_{p,n}$ with budget $\varepsilon_{p,n}$ and estimate $\widetilde{f}_{f,n}$ from nuisance-block quantiles at $\tau \pm h_n$, each with budget $\varepsilon_{f,\pm,n}$, whose sum is at most $\varepsilon_{f,n}$. With independent score noise W_n, define

$$S_N(y) = N^{-1} \sum_{i \in \mathcal{I}_s} \{\mathbf{1}(X_i \le y) - \tau\}, \qquad \widetilde{q}_n^W = \Pi_{[A,B]} \left\{ \widetilde{q}_{p,n} - \frac{S_N(\widetilde{q}_{p,n}) + W_n}{\widetilde{f}_{f,n}} \right\}.$$

The expectation of the score is $F(y) - \tau$, with local slope f_τ, so the update is a Newton correction. Conditional on the independent pilot, the score has sensitivity $1/N$. For the Gaussian estimator $\widetilde{q}_n^G$, take $W_n = \sigma_n Z$ with $Z \sim N(0,1)$ and analytic $(\varepsilon_{s,n}, \delta_{s,n})$ Gaussian calibration (Balle & Wang, 2018). For the Laplace estimator $\widetilde{q}_n^L$, take $W_n = L_s/(N\varepsilon_{s,n})$ with $L_s \sim \mathrm{Lap}(0,1)$. Draw score noise independently of the data and all other mechanism randomness. The entire release is respectively approximate $(E_n, \delta_{s,n})$-DP or pure E_n-DP, where $E_n = \max\{\varepsilon_{p,n}, \varepsilon_{f,n}, \varepsilon_{s,n}\}$.

Algorithm 3 gives the Gaussian construction; Appendix A.2 gives the Laplace variant and both critical values. The following conditions quantify the accuracy needed from the pilot and nuisance estimates.

Let $\omega_f(t) = \sup_{|x-q_\tau| \le t} |f(x) - f_\tau|$ measure local density variation and $b_Q(h) = |(Q(\tau+h) - Q(\tau-h))/(2h) - 1/f_\tau|$ the quantile-slope bias. Pilot and nuisance errors have scales

$$d_n = m_p^{-1/2} + (m_p \varepsilon_{p,n})^{-1}, \quad \rho_{\pm,n} = m_f^{-1/2} + (m_f \varepsilon_{f,\pm,n})^{-1}, \quad e_{f,n} = b_Q(h_n) + \frac{\rho_{+,n} + \rho_{-,n}}{h_n}.$$

For Gaussian noise put $r_n = N^{-1/2} + \sigma_n$; for Laplace noise put $r_n^L = N^{-1/2} + (N\varepsilon_{s,n})^{-1}$. The two errors that survive a Newton correction are the pilot error times the density-estimation error, of order $d_n e_{f,n}$, and the local linearization error, controlled by $d_n \omega_f(C d_n)$ on neighborhoods of radius $C d_n$. The independent score block also contributes a local empirical-process increment, which is negligible at the sampling scale as the pilot becomes consistent. The condition below requires the first two terms to be smaller than the combined sampling and score-noise scale. Pilot accuracy therefore enters jointly with nuisance accuracy and local density variation.

Theorem 5.1 (One-step pivots)**.** *Under Assumption 2.1, suppose $d_n, e_{f,n}, h_n \to 0$, $N \to \infty$, and $\sigma_n \to 0$. If, for every fixed finite C, $d_n\{e_{f,n} + \omega_f(C d_n)\} = o(r_n)$, then*

$$\frac{\widetilde{f}_{f,n}(\widetilde{q}_n^G - q_\tau)}{\sqrt{\tau(1-\tau)/N + \sigma_n^2}} \rightsquigarrow N(0,1).$$

The interval $[\widetilde{q}_n^G \pm z_{1-\alpha/2}\sqrt{\tau(1-\tau)/N + \sigma_n^2}/\widetilde{f}_{f,n}]$ has coverage tending to $1-\alpha$. With Laplace noise, replace σ_n by $(N\varepsilon_{s,n})^{-1}$ and r_n by r_n^L in these conditions. Then the interval using the $1-\alpha$ quantile of $|G_\tau + L_s/(\varepsilon_{s,n}\sqrt{N})|$, divided by $\sqrt{N}\widetilde{f}_{f,n}$, also has asymptotic coverage $1-\alpha$.

The reason is the asymptotic expansion

$$\widetilde{f}_{f,n}(\widetilde{q}_n^G - q_\tau) = -S_N(q_\tau) - \sigma_n Z + o_{\mathbb{P}}(r_n).$$

The empirical score has variance $\tau(1-\tau)/N$ and is independent of Z. Their sum, divided by its standard deviation, is asymptotically Gaussian at any relative noise size. The remainder condition makes the pilot contribution negligible. Gaussian calibration still controls width: the sufficient rule in (11) includes a $\sqrt{\log(1.25/\delta_{s,n})}$ factor. If $\sqrt{N}\sigma_n \to 0$ and $N/n \to 1$, the classical quantile variance is recovered. When privacy noise is first order, the same standardization remains valid because both leading components are Gaussian.

The expansions and proofs are given in Theorems C.4, C.5 and C.7 and Appendices G.14 and G.15. For locally Lipschitz density and block budgets of order $n^{-1/2}$, $m_p \asymp m_f \asymp n^{7/8}$ and $h_n \asymp n^{-3/16}$ leave $N/n \to 1$ and make the remainder negligible at root-n scale. At matched critical pure-DP budgets $\varepsilon_n\sqrt{n} \to \kappa$, QExp has limit $G_\tau + 2L/\kappa$ and Laplace one-step has $G_\tau + L_s/\kappa$ after scaling by $\sqrt{n} f_\tau$. The halved privacy coefficient comes with the pilot, nuisance, and local remainder requirements. Gaussian one-step instead controls pivot shape under approximate privacy.

In this example, $d_n e_{f,n} = O(n^{-9/16})$ and $d_n^2 = O(n^{-3/4})$ are both negligible relative to $n^{-1/2}$; Corollary C.6 details the preliminary rates. The auxiliary blocks thus absorb their privacy noise while using a vanishing fraction of the records.

Table 1 compares limiting distributions and estimation requirements. Direct QExp requires only local regularity for its exact first-order law. Score randomization controls the limiting shape under the pilot and nuisance remainder conditions; interval width also depends on the score sample size, estimated density, and noise calibration. Proposition C.8 formalizes the matched-budget comparison of leading laws. The experiments next evaluate the complete rank-based intervals.

	QExp	Gaussian one-step	Laplace one-step
Privacy	Pure ε	(ε, δ)	Pure ε
Point-estimation data	Full sample	Pilot, nuisance, score blocks	Pilot, nuisance, score blocks
Local smoothness	Continuity and positivity	Modulus remainder; Hölder is sufficient	Same as Gaussian one-step
Critical privacy law	Laplace component $2L/\kappa$	Gaussian component λZ	Laplace component L/κ
Unified law	Gaussian–Laplace	Standard normal	Gaussian–Laplace
Fixed-privacy first order	Classical if $\varepsilon > 0$	Classical if $\sqrt{N}\sigma_n \to 0$	Classical if $\varepsilon > 0$
Mechanism feature	Direct rank geometry	Gaussian inferential law	Smaller pure-DP Laplace constant

Table 1: Mechanism-level comparison. One-step statements require the pilot and nuisance remainder conditions; critical coefficients use matched budgets and $N/n \to 1$, with $\lambda = \lim \sqrt{N}\sigma_n$ when this limit exists. The privacy guarantees differ for the Gaussian construction.

6 Numerical evidence and budget tradeoffs

The experiments examine how the conditional law translates into population coverage, then how private scale estimation changes interval width. All reported intervals target 0.95 coverage. Coverage is the fraction of independent datasets whose interval contains the population median. At the target coverage, shorter mean length indicates more precise inference; smaller mean absolute center error (MAE) indicates a more accurate point estimate. Exact gap sampling is used throughout. Designs and the finite-sample endpoint method are specified in Appendix B.

6.1 The phase law and finite-sample localization

For the standard normal median with public range $[-8, 8]$, the conditional experiment holds each of 30 datasets fixed and draws QExp 300 times. Along $\varepsilon_n = 2/\sqrt{n}$, the average Kolmogorov distance to the standard Laplace law decreases and the conditional variance approaches two (Table 2). This supports the gap-averaging interpretation of Theorem 3.1.

	$\varepsilon_n = 2/\sqrt{n}$		$\varepsilon = 0.5$	
n	Kolmogorov	Variance	Kolmogorov	Variance
500	0.085	3.18	0.180	2.08
2000	0.067	2.10	0.146	2.32
8000	0.057	1.99	0.169	2.35

Table 2: Conditional QExp perturbation T_n: average Kolmogorov distance to $\mathrm{Lap}(0, 1)$ and conditional variance (target two). Each row averages over 30 datasets and 300 mechanism draws per dataset.

Fixed $\varepsilon = 0.5$ keeps the effective rank window bounded, with larger distributional distances: a few random gaps continue to determine the conditional law at the finer privacy scale. The perturbation is of order n^{-1} by Theorem G.13, so it disappears relative to the sample fluctuation. Classical first-order

efficiency can therefore hold while the conditional privacy law retains the sample's random gap geometry.

To isolate the center law from nuisance estimation, Table 3 uses the true density in two intervals applied to the same private center. One uses the Gaussian–Laplace critical value; the other uses a Gaussian with matched asymptotic variance. The phase parameter $\kappa_0 = \varepsilon_{0,n}\sqrt{n}$ uses the actual center budget. At the three displayed configurations, scaled variances track the limits closely. Both critical-value rules give coverage near nominal; these coverage estimates do not resolve a clear advantage of the convolution critical value over variance matching.

n	κ_0	Convolution coverage	Gaussian coverage	Variance	Limit
8000	0.5	0.965 ± 0.006	0.953 ± 0.008	32.09	32.25
2000	2	0.955 ± 0.006	0.951 ± 0.006	2.159	2.250
2000	8	0.943 ± 0.007	0.942 ± 0.007	0.372	0.375

Table 3: Center calibration for the standard normal median. Variance refers to $\sqrt{n} f_\tau(\widetilde{q}_{0,n} - q_\tau)$; the limit is $1/4 + 8/\kappa_0^2$. Coverage entries include one Monte Carlo standard error, with 1200 replications at $n = 2000$ and 800 at $n = 8000$. The true density is used only in this center-law experiment.

The local approximation requires the center to leave the endpoint gaps. At $n = 2000$, $\kappa_0 = 0.5$, endpoint selection occurs in 7.7% of samples: convolution coverage is 0.827 and the scaled variance is 987.76, far above 32.25. At $n = 8000$ under the same phase parameter, endpoint selection is absent in the reported draws and coverage is 0.965. Thus a nominal phase parameter alone does not guarantee finite-sample localization. Gap-length weighting explains both the eventual local law and its failure when a penalized endpoint gap still carries appreciable mass.

6.2 Feasible intervals at a common total privacy budget

We compare QExp-SC with the widened rank-exponential endpoint construction of Drechsler et al. (2022), using the explicit conservative binomial-tail calibration in Appendix B.2. We call this implementation ExpMech. It targets finite-sample coverage at least 0.95 through endpoint inversion; QExp-SC estimates local sparsity and uses an asymptotic critical law. Each method is pure E_n-DP on the same data and public range. QExp-SC assigns $60/20/20$ percent of $E_n = 2/(0.6\sqrt{n})$ to center and side releases, so $\kappa_0 = 2$. ExpMech assigns $E_n/2$ to each endpoint and uses widening $(B - A)/1000$. QExp-SC uses the public range-based tuning in Section 4.

		QExp-SC		ExpMech	
Model	Range	Coverage	Length	Coverage	Length
$N(0,1)$	$[-8, 8]$	0.969	0.452	1.000	1.102
$N(0,1)$	$[-32, 32]$	0.982	2.265	1.000	1.204
Lognormal	$[0, 16]$	0.963	0.496	1.000	1.167
Lognormal	$[0, 64]$	0.975	2.437	1.000	1.259

Table 4: Pure-DP median intervals at $n = 4000$, with 1000 datasets per setting. Lognormal means $\exp(Z)$ for $Z \sim N(0,1)$. Length is the mean interval length. Coverage Monte Carlo standard errors are at most 0.015.

Table 4 shows that QExp-SC maintains effective empirical coverage while producing shorter intervals in some settings. Coverage is 0.963–0.982 across all four settings, above the nominal 0.95 level. On the standard ranges, mean intervals are 57–59% shorter than ExpMech's, with coverage 0.963–0.969. QExp-SC uses local asymptotic calibration, whereas ExpMech enforces a conservative finite-sample bound. Changing normal measurement units by a factor of three preserves coverage and triples widths. The width advantage depends on the public range: expanding each range by a factor of four makes QExp-SC about 1.9 times as wide. For the standard normal model, at least one side release selects an endpoint gap in 13.7% of standard-range samples and 49.8% of wide-range samples. Long gaps separate the side quantiles, and division by $2h_n$ amplifies the estimated sparsity. Scale-equivariant tuning preserves measurement units but does not remove sensitivity to excess output range.

Larger samples restore short QExp-SC intervals even on wide ranges (Table 6). At $n = 16000$, the wide-range normal and lognormal widths are 0.137 and 0.144, compared with 0.644 and 0.646 for ExpMech: reductions of 79% and 78%, with QExp-SC coverage 0.953 and 0.968. Their standard-range widths are 0.134 and 0.143, so the range penalty has nearly disappeared. Across the full design QExp-SC coverage is 0.945–0.993; ExpMech covers in all 1000 draws per cell, consistent with its conservative calibration.

6.3 Allocating privacy to the center or to its scale

At the standard normal median, $n = 4000$, range $[-32, 32]$, and the same total E_n, we compare three allocations on 2000 paired datasets and mechanism uniforms. All use $h_n = 1/4$; each uses its actual center critical value, with $\kappa_0 = 2, 4/3, 2/3$. Table 5 separates the nuisance benefit from the center cost.

Center/upper/lower	Side endpoint	Center MAE	Coverage	Mean length
$60/20/20$	51.6%	0.043	0.980	2.419 ± 0.058
$40/30/30$	9.4%	0.062	0.967	0.912 ± 0.039
$20/40/40$	0.9%	0.249	0.949	0.870 ± 0.023

Table 5: Budget allocation at fixed range and total privacy. Side endpoint is the frequency that at least one side release selects an endpoint gap. Length includes one Monte Carlo standard error; coverage standard errors are at most 0.005. Center MAE measures absolute error about the population median.

Moving from $60/20/20$ to $40/30/30$ reduces mean length by 62% while increasing center MAE from 0.043 to 0.062; the paired length reduction is 1.507 with standard error 0.056. The $20/40/40$ split gives only a further 0.042 reduction in width, but center MAE rises to 0.249, and the center itself selects an endpoint gap in 0.7% of draws. With the true density, the corresponding interval widths would instead increase from 0.247 to 0.363 to 0.716. The improved private scale estimate more than offsets the larger center error in the interval width. A vanishing nuisance budget is sufficient asymptotically; on a wide range, allocating more privacy budget to scale estimation can be substantially more useful at finite sample sizes.

7 Concluding remarks

The paper connects private quantile inference to the mechanism's local error law. For QExp, conditional gap averaging yields a Laplace limit independent of sampling fluctuation, resolving three regimes and removing a logarithmic factor from the pointwise privacy scale. A common sparsity estimate enables feasible calibration across regimes. Gaussian score randomization instead yields a Gaussian pivot under approximate DP, subject to explicit pilot and nuisance remainder conditions. These alternatives distinguish calibrating an existing law from designing its shape. Empirically, effective coverage can accompany shorter intervals, while the public range and private scale estimation govern attainable precision.

Two extensions require additional theory. With an unknown output range, an AboveThreshold-based localization step (Durfee, 2023) would make the interval random. Extending Theorem 3.1 requires localization failure to vanish and the weighted-gap normalizer and far-tail bound in (37) to be controlled uniformly over the selected intervals. Localization on a separate block preserves independence, but its sample and privacy costs must leave the intended center regime unchanged; reusing observations requires a joint analysis of range selection and local spacings. For a growing collection of quantiles, Theorem C.2 is insufficient: dividing the budget among more releases increases their privacy scales. Joint rank mechanisms or density-based inversion can use the budget more efficiently (Gillenwater et al., 2021; Lalanne et al., 2023a), but couple the privacy errors across levels. Simultaneous bands would require uniform quantile linearization, consistent scale estimation over the index set, and tightness of the joint sampling–privacy process. Density-based inversion additionally introduces smoothing bias. A functional stable limit and inversion analysis (van der Vaart, 1998) would therefore need to account for mechanism-dependent correlated errors, replacing the independent Laplace coordinates proved here.

AI USE STATEMENT

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

REPRODUCIBILITY STATEMENT

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

REFERENCES

Anders Aamand, Fabrizio Boninsegna, Abigail Gentle, Jacob Imola, and Rasmus Pagh. Lightweight protocols for distributed private quantile estimation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 27–58, 2025.

Daniel Alabi, Omri Ben-Eliezer, and Anamay Chaturvedi. Bounded space differentially private quantiles. *Transactions on Machine Learning Research*, 2023. URL `https://openreview.net/forum?id=sixOD8YVvM`.

Hilal Asi and John C. Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. In *Advances in Neural Information Processing Systems*, volume 33, pp. 14106–14117, 2020a.

Hilal Asi and John C. Duchi. Near instance-optimality in differential privacy. *arXiv preprint arXiv:2005.10630*, 2020b.

Marco Avella-Medina. Privacy-preserving parametric inference: A case for robust statistics. *Journal of the American Statistical Association*, 116(534):969–983, 2021. doi: 10.1080/01621459.2019.1700130.

Marco Avella-Medina, Casey Bradshaw, and Po-Ling Loh. Differentially private inference via noisy optimization. *The Annals of Statistics*, 51(5):2067–2092, 2023. doi: 10.1214/23-AOS2321.

Jordan Awan and Aleksandra Slavković. Differentially private uniformly most powerful tests for binomial data. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Jordan Awan and Salil Vadhan. Canonical noise distributions and private hypothesis tests. *The Annals of Statistics*, 51(2):547–572, 2023. doi: 10.1214/23-AOS2259.

R. R. Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966. doi: 10.1214/aoms/1177699450.

Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 394–403, 2018.

Leheng Cai, Qirui Hu, Juntao Sun, and Shuyuan Wu. Time-uniform and asymptotic confidence sequence of quantile under local differential privacy. In *Advances in Neural Information Processing Systems*, volume 38, pp. 114488–114520, 2025a. doi: 10.52202/085713-3818.

Leheng Cai, Qirui Hu, and Shuyuan Wu. Federated learning of quantile inference under local differential privacy. *arXiv preprint arXiv:2509.21800*, 2025b.

Karan Chadha, John C. Duchi, and Rohith Kuditipudi. Resampling methods for private statistical inference. *arXiv preprint arXiv:2402.07131*, 2024.

Du Chen and Geoffrey A. Chua. Differentially private stochastic convex optimization under a quantile loss function. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 4435–4461, 2023.

Herbert A. David and Haikady N. Nagaraja. *Order Statistics*. Wiley, 3 edition, 2003. doi: 10.1002/0471722162.

Jörg Drechsler, Ira Globus-Harris, Audra McMillan, Jayshree Sarathy, and Adam Smith. Nonparametric differentially private confidence intervals for the median. *Journal of Survey Statistics and Methodology*, 10(3):804–829, 2022. doi: 10.1093/jssam/smac021.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual 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.

David Durfee. Unbounded differentially private quantile and maximum estimation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 77691–77712, 2023. doi: 10.52202/075280-3394.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pp. 371–380, 2009. doi: 10.1145/1536414.1536466.

Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*, volume 9 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers, 2014.

Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. Parametric bootstrap for differentially private confidence intervals. In *Proceedings of the Twenty-Fifth International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 1598–1618, 2022.

Marco Gaboardi, Ryan Rogers, and Or Sheffet. Locally private mean estimation: Z-test and tight confidence intervals. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 2545–2554, 2019.

Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3713–3722, 2021.

Peter Hall and Simon J. Sheather. On the distribution of a studentized quantile. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):381–391, 1988. doi: 10.1111/j.2517-6161.1988.tb01735.x.

Jacob Imola, Fabrizio Boninsegna, Hannah Keller, Anders Aamand, Amrita Roy Chowdhury, and Rasmus Pagh. Differentially private quantiles with smaller error. In *Advances in Neural Information Processing Systems*, volume 38, pp. 157193–157218, 2025. doi: 10.52202/085713-4739.

Haim Kaplan, Shachar Schnapp, and Uri Stemmer. Differentially private approximate quantiles. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10751–10761, 2022.

Mikhail Khodak, Kareem Amin, Travis Dick, and Sergei Vassilvitskii. Learning-augmented private algorithms for multiple quantile release. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 16344–16376, 2023.

Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. Private statistical estimation of many quantiles. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 18399–18418, 2023a.

Clément Lalanne, Clément Gastaud, Nicolas Grislain, Aurélien Garivier, and Rémi Gribonval. Private quantiles estimation in the presence of atoms. *Information and Inference: A Journal of the IMA*, 12(3):2197–2223, 2023b. doi: 10.1093/imaiai/iaad030.

Jing Lei. Differentially private M-estimators. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

Yi Liu and Qirui Hu. Locally private online quantile regression: Estimation and inference. *arXiv preprint arXiv:2607.05312*, 2026.

Yi Liu, Qirui Hu, Lei Ding, and Linglong Kong. Online local differential private quantile inference via self-normalization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 21698–21714, 2023.

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.

Qilong Lu, Song Xi Chen, and Yumou Qiu. Versatile differentially private learning for general loss functions. *The Annals of Statistics*, 54(2):692–717, 2026. doi: 10.1214/25-AOS2583.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 94–103, 2007. doi: 10.1109/FOCS.2007.66.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pp. 75–84, 2007. doi: 10.1145/1250790.1250803.

Kelly Ramsay and Jairo Diaz-Rodriguez. Differentially private boxplots. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 51117–51139. PMLR, 2025.

Matthew Reimherr and Jordan Awan. KNG: The K-norm gradient mechanism. In *Advances in Neural Information Processing Systems*, volume 32, pp. 10208–10219, 2019.

Ziliang Shen, Caixing Wang, Shaoli Wang, and Yibo Yan. High-dimensional differentially private quantile regression: Distributed estimation and statistical inference. *arXiv preprint arXiv:2508.05212v2*, 2026.

Tomer Shoham, Moshe Shenfeld, Noa Velner-Harris, and Katrina Ligett. Differentially private nonparametric confidence intervals under minimal distributional assumptions. *arXiv preprint arXiv:2511.01303v2*, 2026.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pp. 813–822, 2011. doi: 10.1145/1993636.1993743.

Lukas Steinberger. Efficiency in local differential privacy. *The Annals of Statistics*, 52(5):2139–2166, 2024. doi: 10.1214/24-AOS2425.

Tran Tran, Matthew Reimherr, and Aleksandra Slavkovic. Differentially private quantile regression. In *Privacy in Statistical Databases*, volume 14915 of *Lecture Notes in Computer Science*, pp. 18–34. Springer, 2024. doi: 10.1007/978-3-031-69651-0_2.

Christos Tzamos, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Ilias Zadik. Optimal private median estimation under minimal distributional assumptions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3301–3311, 2020.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010. doi: 10.1198/jasa.2009.tm08651.

Algorithms and numerical details are collected in Appendices A and B. Additional guarantees and comparisons follow in Appendices C to F; the complete proofs are in Section G.

A Additional algorithms and calibration

The main text gives the exact QExp sampler and feasible intervals. This section collects the Gaussian and Laplace score constructions and their public calibration. All algorithms use the target, range and privacy definitions of Section 2.

A.1 Central differential privacy

Two samples $D, D' \in \mathbb{R}^n$ are adjacent if they differ in at most one coordinate.

Definition A.1. A randomized mechanism M is ε-differentially private if, for every adjacent D, D' and measurable output event S,

$$\mathbb{P}\{M(D) \in S\} \le e^{\varepsilon}\mathbb{P}\{M(D') \in S\}.$$

Definition A.2. A randomized mechanism M is (ε, δ)-differentially private if, for every adjacent D, D' and measurable output event S,

$$\mathbb{P}\{M(D) \in S\} \le e^{\varepsilon}\mathbb{P}\{M(D') \in S\} + \delta.$$

Pure differential privacy is the special case $\delta = 0$.

Proposition A.3. *The mechanism* (1) *is pure ε_n-differentially private under replacement adjacency. For $j = 0, \dots, n$, let*

$$I_{j,n} = [A, B] \cap [X_{(j)}, X_{(j+1)}], \qquad \Delta_{j,n} = \mathrm{Leb}(I_{j,n}),$$

where endpoint conventions are immaterial. Thus the nonempty $I_{j,n}$ partition $[A, B]$ up to endpoints. Conditionally on the data, it can be sampled exactly as follows:

1. *sample $J_n \in \{0, \dots, n\}$ with*

$$\mathbb{P}(J_n = j \mid D_n) = \frac{\Delta_{j,n} \exp\{-\varepsilon_n |j - k_n|/2\}}{\sum_{\ell=0}^{n} \Delta_{\ell,n} \exp\{-\varepsilon_n |\ell - k_n|/2\}}; \tag{7}$$

2. *conditionally on $J_n = j$, sample uniformly from $I_{j,n}$ (zero-length intervals have zero selection probability).*

After sorting, the exact sampler costs $O(n)$ arithmetic operations and no Markov chain approximation.

A.2 Private score releases and a one-step correction

Privatizing a bounded quantile score makes the inferential law depend directly on the chosen noise distribution. Approximate-DP Gaussian score noise yields an asymptotically standard-normal pivot across sampling and privacy regimes; Laplace score noise reduces the critical pure-DP privacy coefficient from QExp's factor two to one, under the matched-budget conditions in Proposition C.8. These gains require private pilot and nuisance estimates on separate blocks, with accuracy sufficient for a one-step correction.

The quantile score has expectation $F(y) - \tau$, which vanishes at q_τ and has local slope f_τ. Subtracting the score at a pilot divided by an estimate of f_τ therefore gives a Newton correction toward the target. We use disjoint pilot, nuisance, and score blocks to obtain parallel privacy composition and independence of the pilot from the score empirical process. Conditional on the private pilot, the average score on N records has sensitivity $1/N$, allowing noise to be calibrated before the correction.

Partition the indices deterministically into $\mathcal{I}_p$, $\mathcal{I}_f$, and $\mathcal{I}_s$, of sizes m_p, m_f, and N, respectively. Let

$$\psi_y(x) = \mathbf{1}\{x \le y\} - \tau, \qquad S_N(y) = \frac{1}{N} \sum_{i \in \mathcal{I}_s} \psi_y(X_i).$$

Algorithm 3 GOS: Gaussian one-step quantile inference

Require: Disjoint blocks $\mathcal{I}_p, \mathcal{I}_f, \mathcal{I}_s$, positive component budgets, $0 < \delta_{s,n} < 1$, and public $0 < h_n < \min(\tau, 1-\tau)$.

1: On $\mathcal{I}_p$, use QExp with budget $\varepsilon_{p,n}$ to release a pilot $\widetilde{q}_{p,n}$ at level τ.

2: On $\mathcal{I}_f$, use QExp at $\tau+h_n$ and $\tau-h_n$ with budgets $\varepsilon_{f,+,n}$ and $\varepsilon_{f,-,n}$ satisfying $\varepsilon_{f,+,n}+\varepsilon_{f,-,n} \le \varepsilon_{f,n}$. Denote the releases by $\widetilde{q}_{f,+,n}$ and $\widetilde{q}_{f,-,n}$ and, for public $c_n \downarrow 0$ and $C_n \uparrow \infty$, form

$$\widetilde{s}^{\mathrm{raw}}_{f,n} = \frac{\widetilde{q}_{f,+,n} - \widetilde{q}_{f,-,n}}{2h_n}, \qquad \widetilde{s}_{f,n} = \Pi_{[c_n, C_n]}(\widetilde{s}^{\mathrm{raw}}_{f,n}), \qquad \widetilde{f}_{f,n} = 1/\widetilde{s}_{f,n}.$$

3: Conditional on $\widetilde{q}_{p,n}$, release

$$\widetilde{S}_n = S_N(\widetilde{q}_{p,n}) + \sigma_n Z, \qquad Z \sim N(0,1), \tag{8}$$

where Z is independent and σ_n is calibrated to sensitivity $\Delta_n = 1/N$. In particular, the analytic Gaussian calibration chooses σ_n so that

$$\Phi\left(\frac{\Delta_n}{2\sigma_n} - \frac{\varepsilon_{s,n}\sigma_n}{\Delta_n}\right) - e^{\varepsilon_{s,n}}\Phi\left(-\frac{\Delta_n}{2\sigma_n} - \frac{\varepsilon_{s,n}\sigma_n}{\Delta_n}\right) \le \delta_{s,n}. \tag{9}$$

4: Form

$$\widetilde{q}^{G}_n = \Pi_{[A,B]}\left\{\widetilde{q}_{p,n} - \frac{\widetilde{S}_n}{\widetilde{f}_{f,n}}\right\}, \qquad \mathcal{C}^{G}_n = \left[\widetilde{q}^{G}_n \pm z_{1-\alpha/2}\frac{\sqrt{\tau(1-\tau)/N + \sigma_n^2}}{\widetilde{f}_{f,n}}\right]. \tag{10}$$

5: Release the final estimate and interval.

Privacy and coverage guarantees are given in Appendix C.5.

The exact condition (9) is the analytic Gaussian mechanism of Balle & Wang (2018). A simpler sufficient calibration, valid for $0 < \varepsilon_{s,n} < 1$ and $0 < \delta_{s,n} < 1$, is

$$\sigma_n = \frac{\sqrt{2\log(1.25/\delta_{s,n})}}{N\varepsilon_{s,n}}, \tag{11}$$

although it is generally more conservative.

A.2.1 Laplace score noise

For a direct pure-DP comparison, replace (8) by

$$\widetilde{S}^{L}_n = S_N(\widetilde{q}_{p,n}) + \frac{L_s}{N\varepsilon_{s,n}}, \qquad L_s \sim \mathrm{Lap}(0,1),$$

where L_s is independent of the data and all other mechanism randomness. Use the update in (10) with $\widetilde{S}^{L}_n$ to obtain $\widetilde{q}^{L}_n$.

Let

$$c_{L,1-\alpha}(\tau,\kappa) = \inf\{c : \mathbb{P}(|G_\tau + L_s/\kappa| \le c) \ge 1-\alpha\}.$$

Define

$$\mathcal{C}^{L}_n = \left[\widetilde{q}^{L}_n \pm \frac{c_{L,1-\alpha}(\tau, \varepsilon_{s,n}\sqrt{N})}{\sqrt{N}\,\widetilde{f}_{f,n}}\right].$$

The Gaussian and Laplace constructions use the same pilot, nuisance estimate, and Newton update; the score noise determines the critical-value family.

A.3 Computing the QExp critical value

For fixed public τ, κ, α, the critical value in (4) is the root in c of the following probability equal to $1-\alpha$:

$$\mathbb{P}\left(\left|G_\tau + \frac{2L}{\kappa}\right| \le c\right)$$

$$= \frac{1}{2}\int_{\mathbb{R}} \left[\Phi\left\{\frac{c-2\ell/\kappa}{\sqrt{\tau(1-\tau)}}\right\} - \Phi\left\{\frac{-c-2\ell/\kappa}{\sqrt{\tau(1-\tau)}}\right\}\right] e^{-|\ell|}\, d\ell, \tag{12}$$

Only public quantities enter this calculation. The Gaussian and Laplace boundary values are obtained by taking $\kappa \to \infty$ and $\kappa \downarrow 0$, respectively.

A.4 Parallel-composition QExp interval

QExp-PC (parallel composition) partitions indices before observing data into a center block of size n_0 and a nuisance block of size $m = n - n_0$ for scale estimation. The center uses budget $\varepsilon_{0,n}$; the nuisance quantiles use $\varepsilon_{+,n}, \varepsilon_{-,n}$ summing to at most $\varepsilon_{f,n}$. Apply the slope estimate in (5) to the nuisance block, obtaining $\widetilde{f}_{f,n}$, and use

$$\mathcal{C}_n^{Q,\mathrm{PC}} = \left[\widetilde{q}_{0,n} \;\pm\; \frac{c_{Q,1-\alpha}(\tau, \varepsilon_{0,n}\sqrt{n_0})}{\sqrt{n_0}\,\widetilde{f}_{f,n}}\right].$$

Parallel composition gives total privacy $\max\{\varepsilon_{0,n}, \varepsilon_{f,n}\}$; the two nuisance releases still compose sequentially within their block. Thus SC allocates privacy budget to scale estimation, while PC allocates observations.

B Numerical design and additional comparisons

The center-law and conditional-law experiments use seed 20260728, with independent random streams across design cells. The public-tuning comparison uses seed 20260912 and 1000 independent datasets per model, range, and sample size. The budget comparison uses seed 20260913 and 2000 paired datasets. Sample generation and mechanism randomization use NumPy. Seed and replication counts describe the reported experiments; simulation-only density and endpoint measurements are not released by the private procedure.

B.1 Sampling and calibration

For sorted data and target rank $k = \lfloor n\tau \rfloor$, the implementation computes all clipped gaps, including the two endpoint gaps, and assigns log-weight

$$\log \Delta_{j,n} - \frac{\varepsilon_n}{2}|j-k|.$$

Zero-length gaps have weight zero. Log-sum-exp normalization gives the selection probabilities in (7). One uniform draw selects the gap by its cumulative probability, and a second independent uniform draw selects the value within that gap. Thus the implementation samples the continuous mechanism directly, up to floating-point arithmetic.

The integral in (12) is evaluated by adaptive quadrature, followed by Brent root finding for the target probability $1-\alpha$. The bracket is enlarged until it contains the root. At $\tau = 1/2$ and $\alpha = 0.05$, the critical values for $\kappa_0 = 0.5, 2, 8$ are approximately 12.014, 3.121, and 1.208. This computation uses only public quantities.

In the conditional experiment, the empirical cumulative distribution function of $M = 300$ draws is compared with the standard Laplace cumulative distribution function. The Kolmogorov distance is evaluated at each ordered draw and its left limit. Distances and conditional moments are computed separately for 30 independent datasets and then averaged; variances use denominator $M-1$. A fixed Monte Carlo sample size introduces approximation error in each conditional law.

Population coverage uses independent datasets and one center release per dataset. The convolution and variance-matched Gaussian intervals share that center and use the true density to isolate center calibration. The QExp-SC comparisons instead use three conditionally independent private releases and estimate the density. Models are standard normal, its rescaling by three, and $\exp(Z)$ for $Z \sim N(0,1)$. The rescaling shares data and mechanism uniforms, so changes of units can be compared directly. Public-tuning QExp-SC uses $60/20/20$ percent of $E_n = 2/(0.6\sqrt{n})$ and the

range-based rule in Section 4. The ExpMech endpoints use independent randomization on the same datasets.

The wide-range budget experiment uses the same total E_n for $60/20/20$, $40/30/30$, and $20/40/40$ allocations. The resulting center parameters are $\kappa_0 = 2, 4/3, 2/3$ and the common bandwidth is $h_n = 1/4$. Pairing uses common datasets and mechanism uniforms; differences in mean length therefore use paired standard errors. Table 5 gives the complete comparison needed to interpret the budget tradeoff.

B.2 Finite-sample endpoint comparison

We use the widened rank-exponential endpoints of Drechsler et al. (2022) with an explicit conservative calibration. Let $\theta = (B-A)/1000$. After clipping observations to $[A, B]$, put $N_D(y) = \sum_i \mathbf{1}\{X_i < y\}$. For public target rank k, the widened utility is

$$u_k(y) = -\max\{N_D(y-\theta) - k,\ k - N_D(y+\theta),\ 0\}.$$

An endpoint Y_k has density proportional to $\exp\{E_n u_k(y)/4\}$ on $[A, B]$. Its utility has sensitivity one, so the endpoint is $E_n/2$-DP. To sample it, shift the first k clipped order statistics down by θ and the remaining ones up by θ, clip again to $[A, B]$, and sample a gap with weights equal to length times $\exp\{-E_n|j-k|/4\}$.

For $M \sim \mathrm{Bin}(n, 1/2)$, choose the largest $k \in \{1, \ldots, \lfloor n/2 \rfloor\}$ satisfying

$$\mathbb{E}\left[\min\left\{1, \frac{B-A}{\theta}\exp\left(-\frac{E_n}{4}(M-k)_+\right)\right\}\right] \le \frac{\alpha}{2}.$$

Set $k_L = k$ and $k_U = n - k$, draw the two endpoints independently, and return the convex hull of $Y_{k_L} - \theta$ and $Y_{k_U} + \theta$. If no such rank exists, return $[A, B]$. Every tuning quantity and the fallback decision depend only on public inputs. The complete interval is pure E_n-DP.

The coverage bound accounts for both sampling and mechanism randomness. The zero-penalty region has length at least θ, including at a clipped endpoint. Conditional on $M = m$ observations below the true median, the event $Y_{k_L} - \theta > q_{1/2}$ has penalty at least $(m - k_L)_+$. Its probability is bounded by the summand above, using total output length $B - A$. Averaging over the binomial rank bounds lower-tail noncoverage by $\alpha/2$; symmetry gives the same upper-tail bound for $k_U = n - k_L$. Thus the interval covers with probability at least $1 - \alpha$ for a continuous distribution with median in (A, B). This calibration uses a uniform normalizing bound and can be conservative.

B.3 Sample size, measurement units, and output range

Table 6 retains both methods across all model, range, and sample-size settings. QExp-SC coverage ranges from 0.945 to 0.993, and from 0.953 to 0.968 at $n = 16000$. Its upper sparsity bound is never active, as implied by the public construction. ExpMech covers in all 1000 draws in each cell; this is an empirical outcome rather than a population coverage of exactly one. Long gaps inflate QExp-SC widths at smaller sample sizes, especially on wide ranges. The narrowing at $n = 16000$ is consistent with localization of the side releases.

C Additional results and mechanism comparisons

The results below develop the joint laws, nuisance allocations and one-step comparisons used in the main text. Their proofs follow the local spacing analysis later in the supplement.

C.1 Nonvanishing budgets and auxiliary allocations

The following statements complement the conditional shrinking-budget limit. The fixed-budget perturbation result establishes first-order efficiency without requiring a Laplace approximation at the finer n scale.

Under Assumption 2.1, budgets bounded away from zero give $\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n} = O_{\mathbb{P}}(n^{-1})$ by Theorem G.13. Combining this with Theorem G.16 gives the classical limit $\sqrt{n}(\widetilde{q}_{\tau,n} - q_\tau) \rightsquigarrow N(0, \tau(1-\tau)/f_\tau^2)$.

Model	Range	n	QExp-SC		ExpMech	
			Coverage	Length	Coverage	Length
$N(0,1)$	$[-8,8]$	1000	0.971	1.997	1.000	6.514
$N(0,1)$	$[-8,8]$	4000	0.969	0.452	1.000	1.102
$N(0,1)$	$[-8,8]$	16000	0.959	0.134	1.000	0.549
$N(0,1)$	$[-32,32]$	1000	0.993	10.614	1.000	28.844
$N(0,1)$	$[-32,32]$	4000	0.982	2.265	1.000	1.204
$N(0,1)$	$[-32,32]$	16000	0.953	0.137	1.000	0.644
$N(0,9)$	$[-24,24]$	1000	0.971	5.991	1.000	19.543
$N(0,9)$	$[-24,24]$	4000	0.969	1.356	1.000	3.307
$N(0,9)$	$[-24,24]$	16000	0.959	0.401	1.000	1.646
$N(0,9)$	$[-96,96]$	1000	0.993	31.843	1.000	86.531
$N(0,9)$	$[-96,96]$	4000	0.982	6.794	1.000	3.613
$N(0,9)$	$[-96,96]$	16000	0.953	0.410	1.000	1.932
Lognormal	$[0,16]$	1000	0.960	1.898	1.000	6.433
Lognormal	$[0,16]$	4000	0.963	0.496	1.000	1.167
Lognormal	$[0,16]$	16000	0.963	0.143	1.000	0.552
Lognormal	$[0,64]$	1000	0.945	9.736	1.000	25.662
Lognormal	$[0,64]$	4000	0.975	2.437	1.000	1.259
Lognormal	$[0,64]$	16000	0.968	0.144	1.000	0.646

Table 6: Additional pure-DP median interval comparisons with public range-based tuning. Each cell has 1000 independent datasets; coverage standard errors are at most 0.016. The ExpMech calibration is given in Appendix B.2.

For a vanishing auxiliary budget fraction, Corollary G.28 supplies the explicit allocation and capped bandwidth. Its construction satisfies all coverage conditions while $\varepsilon_{0,n}/E_n \to 1$.

C.2 QExp joint limits and privacy accounting

Define

$$G_{\tau,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\tau - \mathbf{1}(X_i \le q_\tau)\}.$$

Then $G_{\tau,n} \rightsquigarrow G_\tau$, where $G_\tau \sim N\{0, \tau(1-\tau)\}$, and the usual quantile expansion is

$$\sqrt{n} f_\tau(\widehat{q}_{\tau,n} - q_\tau) = G_{\tau,n} + o_{\mathbb{P}}(1). \tag{13}$$

Under the assumptions used here, (13) follows directly from the binomial central limit theorem and local differentiability of F; no second derivative of F is required.

Under the conditions of Theorem 3.1, the empirical fluctuation and scaled privacy perturbation converge jointly to the independent pair (G_τ, L); Theorem G.17 gives the statement and proof.

By Proposition G.24, Algorithm 2 is pure $(\varepsilon_{0,n}+\varepsilon_{+,n}+\varepsilon_{-,n})$-DP and hence pure E_n-DP. Computing and releasing $\widetilde{s}_n$, $\widetilde{f}_n$, the critical value, and the interval consumes no further privacy.

The symmetric allocation $\varepsilon_{0,n} = (1-2\beta_n)E_n$ and $\varepsilon_{+,n} = \varepsilon_{-,n} = \beta_n E_n$ is a special case. It does not add privacy beyond E_n, but the center mechanism receives only $(1-2\beta_n)E_n$. The explicit choice in Corollary G.28 makes this division asymptotically costless for the center release.

C.3 Disjoint-sample parallel-composition inference

The construction is given in Appendix A.4.

Theorem C.1. *The complete release is pure*

$$\max\{\varepsilon_{0,n}, \varepsilon_{f,n}\}\text{-}DP.$$

Suppose Assumption 2.1 holds, $n_0 \to \infty$, $m \to \infty$, $n_0\varepsilon_{0,n} \to \infty$, *and*

$$h_n \to 0, \qquad \sqrt{m}\, h_n \to \infty, \qquad m h_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty. \tag{14}$$

Then $\widetilde{f}_{f,n} \to_{\mathbb{P}} f_\tau$ *and*

$$\mathbb{P}\{q_\tau \in \mathcal{C}_n^{Q,\mathrm{PC}}\} \to 1-\alpha.$$

If $n_0/n \to 1$*, the center estimator and interval have the same first-order sampling efficiency as their full-sample counterparts.*

When $nE_n \to \infty$, Lemma G.30 constructs deterministic $m_n = o(n)$ and $h_n \downarrow 0$ satisfying (14), with budget E_n on the center block and total budget E_n on the nuisance block. The complete release is pure E_n-DP and $n_0/n \to 1$.

An explicit construction and the coverage proof are given in Appendix G.13. This construction shows that parallel QExp inference can estimate its nuisance scale privately at the same end-to-end privacy parameter, using an asymptotically vanishing sample fraction. Together with Corollary G.28, it gives two ways to preserve the first-order center law: allocate a vanishing fraction of observations under parallel composition or a vanishing fraction of privacy budget under same-sample composition.

C.4 A FINITE COLLECTION OF QUANTILES

Let $0 < \tau_1 < \cdots < \tau_m < 1$ be fixed and suppose Assumption 2.1 holds at every $q_j = F^{-1}(\tau_j)$ with $f_j = f(q_j) > 0$. Allocate privacy fractions $\alpha_j > 0$, $\sum_{j=1}^m \alpha_j \le 1$, and independently run QEXP at level τ_j with budget $\alpha_j \varepsilon_n$. The joint release is pure ε_n-DP.

Let $G = (G_1, \ldots, G_m)^\top$ be Gaussian with

$$\mathrm{Cov}(G_j, G_k) = \min(\tau_j, \tau_k) - \tau_j \tau_k,$$

and let $L_1, \ldots, L_m$ be independent standard Laplace variables, independent of G.

Theorem C.2. *If* $\varepsilon_n \to 0$*,* $n\varepsilon_n \to \infty$*, and* $\varepsilon_n\sqrt{n} \to \kappa \in (0,\infty)$*, then*

$$\left(\sqrt{n} f_j \{\widetilde{q}_{\tau_j,n} - q_j\}\right)_{j=1}^m \rightsquigarrow \left(G_j + \frac{2L_j}{\alpha_j \kappa}\right)_{j=1}^m. \tag{15}$$

If $\varepsilon_n\sqrt{n} \to \infty$*, the Laplace terms vanish and the classical joint quantile Gaussian law is recovered.*

The independent Laplace terms in (15) arise because the mechanism randomizations are conditionally independent. JointExp and recursive mechanisms may generate dependent privacy errors and can use the budget more efficiently when m grows. Deriving their exact joint local limits is a separate problem. For fixed m, Theorem C.2 provides a transparent benchmark for simultaneous inference.

The classical and single-quantile special cases, together with the corresponding reductions for parallel QExp and Gaussian one-step inference, are collected in Proposition G.44.

C.5 GUARANTEES FOR THE GAUSSIAN AND LAPLACE ONE-STEP ESTIMATORS

The score constructions in Appendix A.2 separate two tasks. Bounded score sensitivity gives the privacy guarantee; accuracy of the pilot and estimated density determines whether the Newton correction leaves only sampling error and the chosen score noise. The latter is the additional condition needed to replace QExp's direct rank randomization by a controlled score release.

Theorem C.3. *Suppose* $\varepsilon_{s,n} > 0$ *and* $0 < \delta_{s,n} < 1$*. Under replacement adjacency, Algorithm 3 is*

$$(\max\{\varepsilon_{p,n}, \varepsilon_{f,n}, \varepsilon_{s,n}\}, \delta_{s,n})\text{-}DP.$$

If the same records were used for all three stages, the standard generic guarantee would instead be $(\varepsilon_{p,n} + \varepsilon_{f,n} + \varepsilon_{s,n}, \delta_{s,n})$*-DP. The interval in* (10) *is post-processing and adds no privacy loss.*

For a target end-to-end guarantee (E_n, δ_n), one may set

$$\varepsilon_{p,n} = \varepsilon_{f,n} = \varepsilon_{s,n} = E_n, \qquad \varepsilon_{f,+,n} = \varepsilon_{f,-,n} = E_n/2, \qquad \delta_{s,n} = \delta_n.$$

Parallel composition then gives exactly (E_n, δ_n) rather than $(3E_n, \delta_n)$.

The rate comparison with direct QExp is therefore explicit. Once the one-step remainder is negligible, GOS has

$$\widetilde{q}_n^G - q_\tau = O_{\mathbb{P}}\left(N^{-1/2} + \sigma_n\right).$$

Under the sufficient calibration (11), this becomes

$$O_{\mathbb{P}}\left\{N^{-1/2}+\frac{\sqrt{\log(1.25/\delta_{s,n})}}{N\varepsilon_{s,n}}\right\},$$

whereas direct QExp has the pointwise regular-model rate $O_{\mathbb{P}}\{n^{-1/2}+(n\varepsilon_{0,n})^{-1}\}$. GOS gives a single Gaussian pivot, with a privacy contribution determined by the released standard deviation σ_n. Under the conventional sufficient calibration, that contribution includes the factor $\sqrt{\log(1.25/\delta_{s,n})}$; analytic calibration can reduce the noise. The corresponding phase parameter is $\lambda_n=\sqrt{N}\sigma_n$, which incorporates both privacy parameters. Rate comparisons with QExp use $N \asymp n$ and comparable center and score budgets.

C.5.1 Low-order conditions and the one-step expansion

Define the local density modulus

$$\omega_f(t)=\sup_{|x-q_\tau|\le t}|f(x)-f_\tau|,$$

and the deterministic symmetric quantile-slope bias

$$b_Q(h)=\left|\frac{Q(\tau+h)-Q(\tau-h)}{2h}-\frac{1}{f_\tau}\right|.$$

Under Assumption 2.1, both tend to zero. The relevant error scales are the pilot error d_n, the neighboring-quantile errors $\rho_{\nu,n}$, the density error $e_{f,n}$, and the final sampling and score-noise scale r_n:

$$\begin{aligned} d_n &= m_p^{-1/2}+(m_p\varepsilon_{p,n})^{-1}, \\ \rho_{\nu,n} &= m_f^{-1/2}+(m_f\varepsilon_{f,\nu,n})^{-1}, \qquad \nu\in\{+,-\}, \\ e_{f,n} &= b_Q(h_n)+\frac{\rho_{+,n}+\rho_{-,n}}{h_n}, \qquad r_n = N^{-1/2}+\sigma_n. \end{aligned}$$

The Newton remainder has two sources. A pilot error of order d_n interacts with density-estimation error through the product $d_n e_{f,n}$. Linearizing F between the pilot and q_τ contributes $d_n\omega_f(Cd_n)$ for a fixed C. The next theorem requires both terms to be negligible relative to the final sampling and score-noise scale r_n.

Theorem C.4. *Suppose Assumption 2.1 holds, $m_p\varepsilon_{p,n}\to\infty$, $m_f\varepsilon_{f,+,n}\to\infty$, $m_f\varepsilon_{f,-,n}\to\infty$, $h_n\to 0$, $e_{f,n}\to 0$, $d_n\to 0$, $N\to\infty$, and $\sigma_n\to 0$. Assume, for every fixed $C<\infty$,*

$$d_n\{e_{f,n}+\omega_f(Cd_n)\}=o(r_n). \tag{16}$$

Then

$$\widetilde{q}_n^G-q_\tau=\frac{1}{Nf_\tau}\sum_{i\in\mathcal{I}_s}\{\tau-\mathbf{1}(X_i\le q_\tau)\}-\frac{\sigma_n Z}{f_\tau}+o_{\mathbb{P}}(r_n).$$

Moreover $\widetilde{f}_{f,n}-f_\tau=O_{\mathbb{P}}(e_{f,n})$, and the projection onto $[A,B]$ is inactive with probability tending to one.

The expansion separates the usual quantile influence-function average from the independent Gaussian score noise. The proof, including the empirical-process remainder, is in Appendix G.14.

Theorem C.5. *Under Theorem C.4,*

$$\frac{\widetilde{f}_{f,n}(\widetilde{q}_n^G-q_\tau)}{\sqrt{\tau(1-\tau)/N+\sigma_n^2}}\rightsquigarrow N(0,1).$$

Consequently

$$\mathbb{P}\{q_\tau\in\mathcal{C}_n^G\}\to 1-\alpha.$$

Writing $\lambda_n=\sqrt{N}\,\sigma_n$, the corresponding root-N limits are

$$\sqrt{N}(\widetilde{q}_n^G-q_\tau)\rightsquigarrow N\left(0,\frac{\tau(1-\tau)+\lambda^2}{f_\tau^2}\right) \quad \text{if } \lambda_n\to\lambda<\infty.$$

If $\lambda_n\to 0$, the classical efficient variance is recovered. If $\lambda_n\to\infty$ while $\sigma_n\to 0$, then $\widetilde{q}_n^G$ remains consistent and $(\widetilde{q}_n^G-q_\tau)/\sigma_n\rightsquigarrow N(0,f_\tau^{-2})$.

Corollary C.6. *Suppose, for some $\gamma \in (0, 1]$ and finite H,*

$$|f(x) - f_\tau| \le H|x - q_\tau|^\gamma$$

near q_τ. Then $\omega_f(t) = O(t^\gamma)$ and $b_Q(h) = O(h^\gamma)$, so a sufficient version of (16) *is*

$$d_n^{1+\gamma} + d_n e_{f,n} = o(r_n).$$

For local Lipschitz continuity ($\gamma = 1$), fixed positive block budgets, and $m_p \asymp m_f \asymp n^{3/4}$, the choice $h_n \asymp n^{-3/16}$ satisfies the condition when $N/n \to 1$ and $\sqrt{N}\sigma_n = O(1)$. If all block privacy parameters are of order $n^{-1/2}$, the choice $m_p \asymp m_f \asymp n^{7/8}$ and $h_n \asymp n^{-3/16}$ is sufficient.

These choices leave $N/n \to 1$ while making the nuisance error small enough for the pilot correction. In both examples $d_n = O(n^{-3/8})$ and $e_{f,n} = O(n^{-3/16})$, so the product is $O(n^{-9/16}) = o(n^{-1/2})$. The larger auxiliary blocks in the shrinking-budget example compensate for the increased error of each private quantile release; the bandwidth balances the Lipschitz slope bias with the amplified side-release error.

C.5.2 Pure-DP Laplace one-step variant and comparison with QExp

Theorem C.7. *The full pilot–nuisance–score release is pure $\max\{\varepsilon_{p,n}, \varepsilon_{f,n}, \varepsilon_{s,n}\}$-DP. Put $r_n^L = N^{-1/2} + (N\varepsilon_{s,n})^{-1}$. Under the assumptions of Theorem C.4, with σ_n replaced by $(N\varepsilon_{s,n})^{-1}$ and* (16) *measured against r_n^L,*

$$\widetilde{q}_n^L - q_\tau = \frac{1}{N f_\tau} \sum_{i \in \mathcal{I}_s} \{\tau - \mathbf{1}(X_i \le q_\tau)\} - \frac{L_s}{N\varepsilon_{s,n} f_\tau} + o_{\mathbb{P}}(r_n^L).$$

If $N\varepsilon_{s,n} \to \infty$, the interval $\mathcal{C}_n^L$ in Appendix A.2 has asymptotic coverage $1 - \alpha$ without requiring convergence of $\varepsilon_{s,n}\sqrt{N}$.

The proof in Appendix G.15 uses the same score expansion with Laplace noise. At critical privacy budgets, the direct and one-step constructions can therefore be compared through the coefficients of their Laplace components.

Proposition C.8. *Assume $N/n \to 1$, $\varepsilon_n\sqrt{n} \to \kappa \in (0, \infty)$, and that all pilot and nuisance remainders are negligible at root-n scale. Under the same pure-DP center or score budget ε_n,*

$$\sqrt{n} f_\tau(\widetilde{q}_n^Q - q_\tau) \rightsquigarrow G_\tau + 2L_Q/\kappa, \qquad \mathrm{Var}(2L_Q/\kappa) = 8/\kappa^2,$$
$$\sqrt{n} f_\tau(\widetilde{q}_n^L - q_\tau) \rightsquigarrow G_\tau + L_s/\kappa, \qquad \mathrm{Var}(L_s/\kappa) = 2/\kappa^2.$$

If instead $\sqrt{N}\sigma_n \to \lambda$, then GOS has the Gaussian limit $N\{0, [\tau(1-\tau) + \lambda^2]/f_\tau^2\}$ after root-n scaling. The factor-four privacy-variance comparison between QExp and the Laplace one-step release is mechanism specific and does not order the complete procedures: the latter uses pilots, nuisance estimation, sample splitting, and the stronger one-step remainder condition, while direct QExp is valid under continuity alone.

GOS yields a Gaussian standardized limit for every relative magnitude of sampling and score noise. Its approximate-privacy calibration determines the noise scale through both $\varepsilon_{s,n}$ and $\delta_{s,n}$. Under the sufficient rule (11),

$$\lambda_n = \frac{\sqrt{2\log(1.25/\delta_{s,n})}}{\varepsilon_{s,n}\sqrt{N}},$$

so the relevant boundary contains the $\delta_{s,n}$ calibration as well as $\varepsilon_{s,n}\sqrt{N}$.

C.6 A class-matched lower bound for the two error scales

The privacy-noise constants above describe particular mechanisms. The next result identifies the error scales imposed by sampling and pure privacy over a regular class. We measure expected absolute error and write $\mathsf{DP}(\varepsilon)$ for all pure ε-DP estimators under replacement adjacency.

For $\ell, r > 0$ and a public interval $[A, B]$, let $\mathcal{F}(\tau, \ell, r; A, B)$ be the class of continuous distributions supported on $[A, B]$ with a unique τ-quantile $q_F \in [A + 2r, B - 2r]$, absolutely continuous on $[q_F - r, q_F + r]$, and with density satisfying

$$f(x) \geq \ell, \qquad |x - q_F| \leq r.$$

The support restriction is imposed only for this lower-bound subclass; the upper theory allows unbounded observations.

Theorem C.9. *There are positive constants c_0, c_1, c_2, c_3, depending only on τ and on the fixed smooth bump density used in the proof, with the following property. Suppose*

$$\ell r \leq c_0, \qquad B - A \geq \frac{c_1}{\ell} + 6r, \tag{17}$$

and

$$\frac{1}{\ell\sqrt{n}} \vee \frac{1}{\ell n\varepsilon} \leq c_2 r.$$

Then, for all sufficiently large n and every $0 < \varepsilon \leq 1$,

$$\inf_{M \in \mathsf{DP}(\varepsilon)} \sup_{F \in \mathcal{F}(\tau,\ell,r;A,B)} \mathbb{E}_F |M(X_1, \ldots, X_n) - q_F| \geq c_3 \left\{ \frac{1}{\ell\sqrt{n}} \vee \frac{1}{\ell n\varepsilon} \right\}.$$

The range conditions in (17) place an explicit compactly supported translated subfamily inside $[A, B]$. The proof then uses two members of that subfamily. For the sampling term, the Hellinger distance between translations separated by Δ is $O(\ell\Delta)$, so choosing $\Delta \asymp 1/(\ell\sqrt{n})$ keeps the product experiments indistinguishable. For the privacy term, their one-observation total variation distance is $O(\ell\Delta)$. A maximal coupling makes the two n-samples differ in a binomial number of coordinates. Group privacy on the event that this Hamming distance is at most m yields an approximate likelihood-ratio bound $e^{m\varepsilon}$ between the output laws. Taking $\Delta \asymp 1/(\ell n\varepsilon)$ keeps $m\varepsilon = O(1)$ and therefore prevents reliable testing. Le Cam's two-point inequality converts testing difficulty into absolute-error risk. A complete proof, including the coupling inequality, is given in Appendix G.17.

Remark C.10. The lower bound has the same two scales as (3), agreeing with the private-median theory of Tzamos et al. (2020) and the quantile lower bounds of Ramsay & Diaz-Rodriguez (2025). Here the translated alternatives verify these scales for the stated class and absolute-error loss. The upper result is a pointwise stochastic-order statement; a uniform expected-risk guarantee would additionally require control of tails over the class. For the local distributional problem, Theorem 3.1 identifies the exact privacy fluctuation without a logarithmic factor.

C.7 Central versus local quantile inference

The contrast with local privacy can be made explicit. Under the randomized-response stochastic approximation of Liu et al. (2023), the fixed-ε LDP variance is (20). At the median,

$$\mathrm{AVar}_{\mathrm{LDP}} = \frac{1}{4r^2 f(q_{1/2})^2}, \qquad r = \tanh(\varepsilon/2),$$

so the variance inflation relative to the nonprivate median is

$$\frac{1}{r^2} = \coth^2(\varepsilon/2) > 1.$$

For small ε, $r \sim \varepsilon/2$, and the standard error is of order $1/(\varepsilon\sqrt{n}f)$.

In the central model, the fixed-ε standard error remains $n^{-1/2}/f$ and the privacy perturbation is of order n^{-1}; when $\varepsilon_n \to 0$ it has the sharper scale $1/(n\varepsilon_n f)$. QExp therefore recovers the classical root-n law when $\varepsilon_n\sqrt{n} \to \infty$, while randomized response retains positive first-order variance inflation at every fixed finite budget.

This difference reflects where randomization occurs. A trusted central curator aggregates the raw records into one rank score before releasing QExp. The local procedure randomizes each binary response before it reaches the curator, protecting the raw record from that curator and supporting one-pass, low-storage inference. Its persistent variance inflation pays for this stronger trust guarantee.

D BOUNDARY CASES AND SCOPE

This section records boundary cases that materially affect the theorem statements.

When $\varepsilon_n \equiv \varepsilon > 0$, the weights $e^{-\varepsilon|j-k|/2}$ place appreciable mass on only a tight number of gaps. Each normalized central spacing has a nondegenerate exponential limit. Therefore the conditional n-scale law remains a random mixture determined by finitely many limiting spacings; it does not average to a deterministic Laplace density. Theorem G.13 gives the perturbation bound $\widetilde{q}_n - \widehat{q}_n = O_{\mathbb{P}}(n^{-1})$ for all budgets bounded away from zero, including diverging budgets. The deterministic Laplace limit requires both $\varepsilon_n \to 0$, so the rank window diverges, and $n\varepsilon_n \to \infty$, so that window remains local in value space.

Although the observations may be unbounded, on all of $\mathbb{R}$ the rank score in (21) is constant to the left of the sample minimum and to the right of the sample maximum. The Lebesgue integral of the unnormalized density is then infinite. A public bounded interval, clipping rule, discrete reference measure, or separately private localization algorithm is necessary.

At an atom, Q is not differentiable in the form used in Lemma G.3; the sample quantile may converge faster than $n^{-1/2}$ or have a non-Gaussian limit, and the width of an order-statistic gap may be zero. The local continuity and positivity in Assumption 2.1 provide the differentiable quantile geometry used by the normal–Laplace expansion.

All proofs use $k_n/n \to \tau \in (0,1)$. If $\tau_n \to 0$ or 1, the local exponential representation must be combined with intermediate- or extreme-order-statistic theory. In particular, the privacy rank scale $1/\varepsilon_n$ must be negligible relative to the effective tail count before a classical tail limit can survive.

The Laplace scaling is

$$\frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n),$$

because the replacement-adjacency score sensitivity is one and the standard exponential mechanism uses $\exp\{\varepsilon u/(2\Delta u)\}$. A different adjacency convention or a differently calibrated score changes this constant. The rate and proof structure remain the same.

The one-dimensional gap sampler is exact after sorting. No Markov-chain approximation is used. If an approximate sampler were substituted, its approximation error would have to be included in the final privacy and distributional guarantees; it cannot be ignored as post-processing.

E CONNECTIONS TO EXISTING QUANTITATIVE GUARANTEES

High-probability error bounds, sample-complexity guarantees, and local distributional limits describe different aspects of private quantile accuracy. This section expresses the first two in terms of the sampling and privacy scales used in the main text, derives the central–local variance comparison, and relates pointwise stochastic order to expected-risk lower bounds.

E.1 FROM A QEXP TAIL INEQUALITY TO A VALUE-ERROR BOUND

The statistical analysis of Lalanne et al. (2023a) yields, under global density upper and lower bounds on a compact interval, an inequality of the generic form

$$\mathbb{P}_F\{|\widetilde{q}_{\tau,n} - q_\tau| > t\} \le C_1 n \exp(-c_1 n\varepsilon\ell t) + C_2 \exp(-c_2 n\ell^2 t^2), \tag{18}$$

where ℓ is a lower density bound and the positive constants may depend on the fixed quantile level and on the global regularity constants. Their Remark 3.6 discusses localizing the density lower bound. The following inversion records the consequence of the stated inequality; the logarithmic term belongs to this bound, not a demonstrated obstruction for QExp.

Proposition E.1. *Suppose* (18) *holds. For every* $\alpha \in (0,1)$*,*

$$|\widetilde{q}_{\tau,n} - q_\tau| \le \frac{1}{c_1 n\varepsilon\ell}\log\left(\frac{2C_1 n}{\alpha}\right) \vee \frac{1}{\ell}\sqrt{\frac{1}{c_2 n}\log\left(\frac{2C_2}{\alpha}\right)} \tag{19}$$

with probability at least $1-\alpha$*.*

For fixed confidence level, (19) has order

$$\frac{1}{\ell\sqrt{n}} + \frac{\log n}{\ell n\varepsilon}.$$

This is a finite-sample high-probability guarantee under global assumptions. The conditional local limit in Theorem G.8 identifies the first-order privacy fluctuation at a fixed regular distribution under shrinking privacy: its scale is $(n\varepsilon f_\tau)^{-1}$ without a logarithm.

E.2 Inverting sample-complexity expressions

The first two terms in the private-median sample complexity of Tzamos et al. (2020) have the form

$$n \gtrsim \frac{\log(1/\beta)}{L^2a^2} + \frac{\log(1/\beta)}{\varepsilon La},$$

where L is the local density lower bound, a is value accuracy and $1-\beta$ is success probability. To satisfy this sample-size condition, each summand must be no larger than a constant multiple of n, which requires

$$a \gtrsim \frac{\sqrt{\log(1/\beta)}}{L\sqrt{n}}, \qquad a \gtrsim \frac{\log(1/\beta)}{Ln\varepsilon}.$$

At fixed confidence this is the same pair of scales as in the present local theory. Their additional term, involving the public range and the radius of the density lower bound, is a genuine localization cost and cannot be removed by this algebraic inversion.

Likewise, a lower-bound statement of the form reported for regular quantiles by Ramsay & Diaz-Rodriguez (2025),

$$n \gtrsim \frac{1}{L^2a^2} \vee \frac{1}{La\varepsilon},$$

is algebraically equivalent, in its nontrivial local range, to

$$a \gtrsim \frac{1}{L\sqrt{n}} \vee \frac{1}{Ln\varepsilon}.$$

The translated bump family and bounded-Hamming coupling in Theorem G.51 establish the lower bound within the regular class used in Theorem C.9, for expected absolute error.

E.3 Derivation of the fixed-LDP variance inflation

We now derive the variance of the randomized-response stochastic-approximation estimator of Liu et al. (2023), used in the central-versus-local comparison. Let

$$B = \mathbf{1}\{X \le q_\tau\}, \qquad \mathbb{P}(B=1) = \tau,$$

and apply symmetric randomized response. Write

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

The released bit Y satisfies

$$\mathbb{P}(Y=1 \mid B) = \frac{1-r}{2} + rB,$$

so the unbiased reconstruction of B is

$$\widetilde{B} = \frac{Y - (1-r)/2}{r}.$$

At the true quantile,

$$p_Y := \mathbb{P}(Y=1) = \frac{1-r}{2} + r\tau = \frac{1 + r(2\tau-1)}{2}.$$

Consequently,

$$\mathrm{Var}(\widetilde{B}) = \frac{p_Y(1-p_Y)}{r^2}$$

$$= \frac{1 - r^2(2\tau - 1)^2}{4r^2}$$
$$= \tau(1 - \tau) + \frac{1 - r^2}{4r^2}.$$

The derivative of the population quantile score $F(x) - \tau$ at q_τ is f_τ. The stochastic approximation and averaging result of Liu et al. (2023, Theorem 3.3) applies this score variance through the inverse slope $1/f_\tau$. It therefore gives

$$\mathrm{AVar}_{\mathrm{LDP}} = \frac{\tau(1 - \tau)}{f_\tau^2} + \frac{1 - r^2}{4r^2 f_\tau^2}.$$

For every finite fixed ε, $0 < r < 1$ and the second term is strictly positive. As $\varepsilon \downarrow 0$, $r \sim \varepsilon/2$, so the private contribution to the standard error is of order $(\varepsilon\sqrt{n} f_\tau)^{-1}$. This calculation identifies the statistical source of the central–local contrast: the central mechanism perturbs one aggregate rank, whereas local randomized response perturbs every Bernoulli contribution before aggregation.

E.4 Pointwise limits and uniform expected risk

The present upper statement

$$\widetilde{q}_{\tau,n} - q_\tau = O_{\mathbb{P}}\{n^{-1/2} + (n\varepsilon_n)^{-1}\}$$

is pointwise in a fixed regular distribution and is an assertion about stochastic order. The lower statement in Theorem G.51 is uniform over a class and concerns expected absolute loss. Their matching algebraic scales show that both sampling and privacy components are unavoidable and that the local limit has the correct regular order. Establishing finite-sample minimax optimality in expected risk over the entire lower-bound class would additionally require a uniform upper bound and uniform-integrability control.

F Further related work

Classical quantile inference combines a linear representation of sampling error with estimation of the local density for studentization (Bahadur, 1966; Hall & Sheather, 1988). Bahadur's original representation assumes two local derivatives and gives an almost-sure remainder; Appendix G.8 proves the first-order expansion needed here under local continuity. Hall and Sheather show that smoothing for confidence accuracy can differ from smoothing for density mean squared error. Private side releases add another source of error to this scale-estimation problem.

A statistical framework for differential privacy compares the accuracy of distributions and densities reconstructed from randomized releases (Wasserman & Zhou, 2010); private estimation and subsample-and-aggregate extend this perspective to asymptotic inference (Lei, 2011; Smith, 2011). The comparisons below connect these routes to the rank mechanism's local law. Appendix E translates existing tail and sample-complexity bounds into the sampling and privacy scales.

F.1 Central private medians and quantiles

The global sensitivity of an order statistic is the diameter of a bounded data domain and is infinite on an unbounded domain. Smooth sensitivity replaces this worst-case quantity by a smoothed envelope of local sensitivities and was illustrated using the median (Nissim et al., 2007). Propose–test–release similarly exploits stability but releases only after a private test succeeds (Dwork & Lei, 2009). These approaches established that robust statistics can be substantially more accurate than global-sensitivity perturbation. Their noise calibration is nevertheless data dependent, and precise population-level limit laws require a separate analysis.

Smith (2011) developed a general subsample-and-aggregate construction that preserves optimal convergence rates and can recover the nonprivate asymptotic distribution for broad classes of regular statistics. This is an important existence and generality result. For a single scalar quantile, however, rank mechanisms are more direct and can exploit the exact one-dimensional order structure.

The inverse-sensitivity program of Asi & Duchi (2020a;b) relates privacy accuracy to the minimum number of sample modifications needed to change the target. For the median, their smoothed

continuous mechanism uses the record distance to a neighborhood of the output value; the underlying unsmoothed distance is constant between order statistics. Proposition 5.1 and its asymptotic consequence in Section 5.1 give equivalence to the sample median for bounded observations, with smoothing radius $1/n$ and $\varepsilon_n\sqrt{n}/\log n \to \infty$. That result does not identify the limit when $\varepsilon_n\sqrt{n}$ is bounded and is stated directly for the median rather than an arbitrary fixed quantile.

The mechanism studied here is the one-quantile exponential mechanism used in Smith (2011) and later denoted QEXP or PRIVATEQUANTILE. Lalanne et al. (2023a) prove, on $[0,1]$ under global density upper and lower bounds, a bound of the form

$$\mathbb{P}(|\widetilde{q}_\tau - q_\tau| > \gamma) \le C_1 n \exp(-c_1 n\varepsilon\gamma) + C_2 \exp(-c_2 n\gamma^2).$$

Their Remark 3.6 also discusses localizing the density lower bound. Inverting this inequality produces the upper scale $n^{-1/2} + \log n/(n\varepsilon)$. The polynomial prefactor comes from a generic minimum-gap argument and does not by itself show that $\log n$ is intrinsic. Our local limit shows that it is not intrinsic for stochastic first-order accuracy at a regular fixed quantile.

For several empirical quantiles, JointExp scores all inter-quantile bin counts simultaneously (Gillenwater et al., 2021); recursive methods improve privacy-budget dependence for approximate quantiles (Kaplan et al., 2022); bounded-space methods address streaming and memory constraints (Alabi et al., 2023); learning-augmented mechanisms use public or predicted structure (Khodak et al., 2023); and correlated continual-counting mechanisms improve the maximum rank error under spacing or approximate-DP conditions (Imola et al., 2025). Distribution-level analyses compare repeated quantile releases with private density or histogram estimation (Lalanne et al., 2023a). These works focus on rank error, high-probability value error, or computation rather than the joint weak limit of sampling and privacy randomness.

The behavior at atoms is qualitatively different. JointExp and unsmoothed inverse sensitivity can be inconsistent at isolated atoms; data jittering supplies a tractable repair under the conditions of Lalanne et al. (2023b). Our regular theory deliberately excludes this nonregular case through uniqueness and local absolute continuity.

F.2 FINITE-SAMPLE OPTIMALITY AND CONFIDENCE INTERVALS

For the median, Tzamos et al. (2020) assume that the median lies in a bounded public range of radius R and that the density is bounded below by L on a neighborhood of radius r. They obtain pure-DP sample complexity, up to constants,

$$\frac{\log(1/\beta)}{L^2\alpha^2} + \frac{\log(1/\beta)}{\varepsilon L\alpha} + \frac{\log(R/\alpha+1)}{\varepsilon L r},$$

where α is value accuracy and $1-\beta$ is success probability. The first two terms correspond to the sampling and privacy scales studied here, while the third is a global localization cost on an unbounded data domain. Ramsay & Diaz-Rodriguez (2025) derive related lower bounds for general quantile levels and use private quantiles to construct boxplots.

Thresholding observations at a fixed candidate quantile gives a binomial count. Exact or optimal tests for this count under pure or approximate privacy (Awan & Slavković, 2018; Awan & Vadhan, 2023) can be inverted to obtain confidence sets.

Drechsler et al. (2022) construct nonparametric differentially private confidence intervals for the median under pure or concentrated privacy by privately estimating appropriate order-statistic endpoints. Their intervals have finite-sample coverage and directly combine sampling and privacy randomness. Their objective is distribution-free confidence validity, whereas ours is sharp local asymptotic inference and first-order interval length under $f(q_\tau) > 0$. The two perspectives are complementary: exact distribution-free validity is available without regularity, while efficient local intervals exploit regularity. The black-box subsampling construction of Shoham et al. (2026) applies to private estimators satisfying its limit-law and rate conditions, with privacy perturbation negligible at the sampling scale; the present analysis instead derives the mechanism's exact first-order law and an analytic critical value without repeated private subsampling.

General private inference has advanced through robust parametric procedures, private M-estimation, noisy optimization, parametric and resampling-based bootstrap methods, and general-loss data

releases under local f-DP (Lei, 2011; Avella-Medina, 2021; Avella-Medina et al., 2023; Ferrando et al., 2022; Chadha et al., 2024; Lu et al., 2026). Quantile-loss optimization under approximate DP has been studied through convolution smoothing (Chen & Chua, 2023); pure-, approximate-, and concentrated-DP quantile regression has also been investigated through smoothed loss and gradient mechanisms (Tran et al., 2024); and recent work develops distributed high-dimensional private quantile regression with debiasing and simultaneous inference (Shen et al., 2026). These frameworks cover models much broader than a scalar quantile, but the corresponding conditions or privacy variance are not tailored to the exact rank geometry of QExp. The smooth-objective theorem for the K-norm gradient mechanism (Reimherr & Awan, 2019), for example, needs an additional argument to cover the nonsmooth empirical check loss.

F.3 Local privacy

The local model protects each record before it reaches the analyst. A locally private binary-search quantile routine is used as a subroutine in the mean-estimation construction of Gaboardi et al. (2019). General minimax theory shows that local privacy contracts statistical information and can alter classical rates and efficiency bounds (Duchi et al., 2013; 2018; Steinberger, 2024). For online quantile inference, Liu et al. (2023) use randomized binary comparisons and stochastic approximation. With randomized-response contrast

$$r = \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1} = \tanh(\varepsilon/2),$$

their averaged estimator has asymptotic variance

$$\frac{1 - r^2(2\tau - 1)^2}{4r^2 f(q_\tau)^2} = \frac{\tau(1-\tau)}{f(q_\tau)^2} + \frac{1 - r^2}{4r^2 f(q_\tau)^2}. \tag{20}$$

They also develop self-normalized inference. Subsequent work treats LDP CDF inference, lightweight distributed protocols, time-uniform quantile confidence sequences, federated quantile inference, and locally private online quantile regression (Liu et al., 2024; Aamand et al., 2025; Cai et al., 2025a;b; Liu & Hu, 2026). For every fixed finite ε, the second term in (20) is positive. By contrast, the central mechanism studied here has no first-order privacy term when ε is fixed.

G Proofs

G.1 Notation and preliminary facts

The proof proceeds from normalized exponential spacings to the conditional mechanism law, then to feasible calibration and one-step inference.

Throughout, $X_1, \dots, X_n$ are i.i.d. with distribution function F on $\mathbb{R}$. A public finite interval $[A, B]$ is used only as the output range of the private mechanism and contains the target quantile. Fix $\tau \in (0, 1)$ and write

$$q = q_\tau = F^{-1}(\tau), \qquad f_\tau = f(q) > 0, \qquad k = k_n = \lfloor n\tau \rfloor.$$

The order statistics are $X_{(1)} \le \cdots \le X_{(n)}$, with $X_{(0)} = -\infty$ and $X_{(n+1)} = +\infty$. The empirical anchor is $\widehat{q}_n = X_{(k)}$. For $j = 0, \dots, n$, define the clipped gap

$$I_{j,n} = [A, B] \cap [X_{(j)}, X_{(j+1)}], \qquad \Delta_{j,n} = \mathrm{Leb}(I_{j,n}), \qquad a_n = \frac{\varepsilon_n}{2}.$$

The nonempty clipped gaps partition $[A, B]$ up to endpoints. Conditionally on the sample, the private estimator $\widetilde{q}_n$ has density

$$p_n(y \mid D_n) = \frac{\exp\{-a_n |N_n(y) - k|\}}{\int_A^B \exp\{-a_n |N_n(z) - k|\}\, dz}, \qquad N_n(y) = \sum_{i=1}^{n} \mathbf{1}\{X_i < y\}. \tag{21}$$

Equivalently, it first selects a gap J_n with probabilities proportional to $\Delta_{j,n} e^{-a_n |j-k|}$ and then draws uniformly inside the corresponding clipped gap $I_{j,n}$.

The regularity condition used throughout the limit theory is the following.

Assumption G.1. *The quantile q is unique and belongs to the public output interval (A, B), and there is an open interval $\mathcal{N}$ containing q on which F is absolutely continuous with a density f satisfying*

$$f(x) \longrightarrow f_\tau \in (0, \infty) \qquad \text{as } x \to q.$$

The observations need not lie in $[A, B]$. Atoms and other irregularities are permitted outside $\mathcal{N}$. No derivative of f, moment or tail condition, or global density lower bound is imposed. All random variables used to generate the privacy mechanism are taken on an extension of the sample probability space and are conditionally independent of any other mechanism randomization given the data.

For a sequence of random conditional laws $K_n(\omega, \cdot)$, the notation $K_n \rightsquigarrow K$ in probability means

$$\int h \, dK_n \longrightarrow_{\mathbb{P}} \int h \, dK$$

for every bounded Lipschitz function h. Since the limiting laws below are tight, this is equivalent to testing against all bounded continuous h.

G.2 Privacy and the exact gap sampler

Proposition G.2. *Under replacement adjacency, the mechanism* (21) *is pure ε_n-differentially private. Conditionally on D_n, it is exactly generated by*

$$\mathbb{P}(J_n = j \mid D_n) = \frac{\Delta_{j,n} e^{-a_n|j-k|}}{\sum_{\ell=0}^{n} \Delta_{\ell,n} e^{-a_n|\ell-k|}}, \qquad j = 0, \dots, n, \tag{22}$$

$$\widetilde{q}_n \mid (J_n = j, D_n) \sim \mathrm{Unif}(I_{j,n}) \quad \textit{whenever } \Delta_{j,n} > 0. \tag{23}$$

Proof. Let D and D' differ in one coordinate. For every y,

$$|N_D(y) - N_{D'}(y)| \le 1.$$

The reverse triangle inequality therefore gives

$$\big||N_D(y) - k| - |N_{D'}(y) - k|\big| \le 1.$$

Thus the score $u_D(y) = -|N_D(y) - k|$ has global sensitivity one. The exponential-mechanism density proportional to $\exp\{\varepsilon_n u_D(y)/2\}$ is consequently ε_n-differentially private. More explicitly, if Z_D denotes the normalizing constant, then

$$e^{-\varepsilon_n/2} \le \frac{\exp\{\varepsilon_n u_D(y)/2\}}{\exp\{\varepsilon_n u_{D'}(y)/2\}} \le e^{\varepsilon_n/2}$$

and integration implies $e^{-\varepsilon_n/2} \le Z_D/Z_{D'} \le e^{\varepsilon_n/2}$. Hence

$$\frac{p_D(y)}{p_{D'}(y)} \le e^{\varepsilon_n}$$

for almost every y, which yields the required event inequality after integration.

Whenever the clipped gap $I_{j,n}$ has positive length, $N_n(y) = j$ for Lebesgue-almost every y in that gap; gaps created inside a tied block have zero length and zero selection probability. The conditional density is therefore constant on that gap. Integrating it over the gap gives a mass proportional to $\Delta_{j,n} e^{-a_n|j-k|}$, proving (22); division by the clipped-gap length gives the uniform conditional law (23). The endpoint conventions affect only sets of Lebesgue measure zero. □

G.3 Local inverse-CDF geometry

Let $Q(u) = F^{-1}(u) = \inf\{x : F(x) \ge u\}$.

Lemma G.3. *Under Assumption G.1, there is $\eta_0 > 0$ such that Q is strictly increasing and Lipschitz on $[\tau - \eta_0, \tau + \eta_0]$. It is differentiable at τ with*

$$Q'(\tau) = \frac{1}{f_\tau}.$$

More strongly, for every deterministic $\eta_n \downarrow 0$,

$$\sup_{\substack{u,v\in[\tau-\eta_n,\tau+\eta_n]\\ u\neq v}} \left| \frac{f_\tau\{Q(u)-Q(v)\}}{u-v} - 1 \right| \longrightarrow 0. \tag{24}$$

In particular, for all sufficiently small fixed $\eta > 0$, there are constants $0 < c_Q < C_Q < \infty$ such that

$$c_Q|u-v| \le |Q(u)-Q(v)| \le C_Q|u-v| \tag{25}$$

whenever $u, v \in [\tau-\eta, \tau+\eta]$.

Proof. Since $f(x) \to f_\tau > 0$, choose an open interval $I = (q-\delta, q+\delta) \subset \mathcal{N}$ on which $f_\tau/2 \le f(x) \le 2f_\tau$. For $x < y$ in I,

$$F(y) - F(x) = \int_x^y f(t)\,dt \ge (f_\tau/2)(y-x) > 0.$$

Thus F is strictly increasing on I. By continuity and uniqueness of q, there is $\eta_0 > 0$ such that $Q([\tau-\eta_0, \tau+\eta_0]) \subset I$.

Take $u < v$ in a shrinking probability neighborhood of τ, and put $x = Q(u)$ and $y = Q(v)$. Then $x, y \to q$ uniformly over that neighborhood. Absolute continuity of F on I gives

$$v - u = F(y) - F(x) = \int_x^y f(t)\,dt.$$

Therefore

$$\left| \frac{v-u}{y-x} - f_\tau \right| \le \sup_{t \text{ between } x \text{ and } y} |f(t) - f_\tau|,$$

which tends to zero uniformly because every such t lies in a shrinking neighborhood of q. Since the average on the left is bounded away from zero, inversion yields (24). Taking one endpoint equal to τ proves differentiability at τ. Taking a fixed sufficiently small neighborhood and using $f_\tau/2 \le f \le 2f_\tau$ gives the two-sided Lipschitz bounds. □

G.4 Uniform order statistics and local exponential grids

We may construct $X_i = Q(U_i)$ with U_i i.i.d. $\mathrm{Unif}(0,1)$. Since Q is nondecreasing, $X_{(j)} = Q(U_{(j)})$ almost surely. Let $U_{(0)} = 0$ and $U_{(n+1)} = 1$.

The representation below is classical; see, for example, David & Nagaraja (2003, Chapter 6).

Lemma G.4. *Let $E_1, \dots, E_{n+1}$ be independent standard exponential variables, $S_j = \sum_{i=1}^{j} E_i$, and $S = S_{n+1}$. Then*

$$(U_{(1)}, \dots, U_{(n)}) \overset{d}{=} (S_1/S, \dots, S_n/S), \tag{26}$$

and the uniform spacings satisfy

$$U_{(j+1)} - U_{(j)} \overset{d}{=} E_{j+1}/S, \qquad j = 0, \dots, n. \tag{27}$$

Proof. The vector of $n+1$ uniform spacings has the Dirichlet$(1, \dots, 1)$ distribution. Independent gamma variables normalized by their sum have the corresponding Dirichlet law, and a standard exponential is Gamma$(1,1)$. Cumulative sums of the normalized spacings give (26); subtraction gives (27). □

The next lemma is the local-grid statement used repeatedly below.

Lemma G.5. *Suppose*

$$a_n \to 0, \qquad na_n \to \infty, \qquad k_n/n \to \tau \in (0,1).$$

For every fixed $M < \infty$,

$$\max_{|r|\le M/a_n} \left| a_n n\{U_{(k_n+r)} - U_{(k_n)}\} - a_n r \right| \longrightarrow_{\mathbb{P}} 0, \tag{28}$$

$$a_n n \max_{|r|\le M/a_n} \{U_{(k_n+r+1)} - U_{(k_n+r)}\} \longrightarrow_{\mathbb{P}} 0. \tag{29}$$

Indices outside $\{0, \dots, n+1\}$ may be omitted; for large n none occur in the displayed window.

Proof. Use the representation in Lemma G.4. Since $S/n \to 1$ almost surely, $n/S \to 1$ in probability. Put $m_n = \lfloor M/a_n \rfloor$. For $r \geq 1$,

$$a_n n(U_{(k+r)} - U_{(k)}) = a_n \frac{n}{S} \sum_{i=k+1}^{k+r} E_i.$$

Hence

$$\begin{aligned} &\max_{1\leq r\leq m_n} \left| a_n n(U_{(k+r)} - U_{(k)}) - a_n r \right| \\ &\quad \leq a_n \frac{n}{S} \max_{1\leq r\leq m_n} \left| \sum_{i=k+1}^{k+r} (E_i - 1) \right| + a_n m_n \left| \frac{n}{S} - 1 \right|. \end{aligned}$$

The second term is at most $M|n/S - 1| + a_n|n/S - 1|$ and therefore converges to zero in probability. For the first term, Kolmogorov's maximal inequality gives, for every $\eta > 0$,

$$\mathbb{P}\left(a_n \max_{1\leq r\leq m_n} \left| \sum_{i=k+1}^{k+r} (E_i - 1) \right| > \eta \right) \leq \frac{a_n^2 m_n}{\eta^2} \leq \frac{M a_n}{\eta^2}.$$

This tends to zero. The negative ranks are handled by the same argument applied to the reversed block of exponentials. This proves (28).

For the mesh, fix $\eta > 0$ and work on the event $n/S \leq 2$, whose probability tends to one. A union bound and the exponential tail yield

$$\begin{aligned} \mathbb{P}\left(a_n \frac{n}{S} \max_{|r|\leq m_n} E_{k+r+1} > \eta, \frac{n}{S} \leq 2 \right) &\leq (2m_n + 2) \exp\{-\eta/(2a_n)\} \\ &\leq (2M/a_n + 4) e^{-\eta/(2a_n)} \longrightarrow 0. \end{aligned}$$

Together with (27), this proves (29). □

Lemma G.6. *Under Assumption G.1 and the conditions of Lemma G.5, for every fixed $M < \infty$,*

$$\max_{|r|\leq M/a_n} \left| a_n n f_\tau \{X_{(k_n+r)} - X_{(k_n)}\} - a_n r \right| \longrightarrow_{\mathbb{P}} 0, \tag{30}$$

$$a_n n f_\tau \max_{|r|\leq M/a_n} \Delta_{k_n+r,n} \longrightarrow_{\mathbb{P}} 0. \tag{31}$$

Proof. Because $k_n/n \to \tau$, $U_{(k_n)} \to \tau$ in probability. Also, $M/(na_n) \to 0$. By (28),

$$\max_{|r|\leq M/a_n} |U_{(k_n+r)} - U_{(k_n)}| = O_{\mathbb{P}}\{(na_n)^{-1}\} = o_{\mathbb{P}}(1).$$

Thus all uniform order statistics in the local window lie in a shrinking neighborhood of τ with probability tending to one. Because $q \in (A, B)$, the corresponding data order statistics lie in (A, B) with probability tending to one, so every clipped gap in this window equals the full adjacent spacing. Apply the uniform inverse-CDF expansion (24) with $u = U_{(k_n+r)}$ and $v = U_{(k_n)}$ to obtain

$$a_n n f_\tau \{X_{(k_n+r)} - X_{(k_n)}\} = a_n n \{U_{(k_n+r)} - U_{(k_n)}\} + o_{\mathbb{P}}(1)$$

uniformly over the window. Combining this with (28) proves (30). Applying the same expansion to adjacent pairs and using (29) proves (31). □

G.5 A WEIGHTED-SPACING RIEMANN THEOREM

For a bounded measurable function $h : \mathbb{R} \to \mathbb{R}$, define

$$R_n(h) = a_n n f_\tau \sum_{j=0}^{n} e^{-a_n|j-k|} \int_{I_{j,n}} h\{a_n n f_\tau (y - X_{(k)})\}\, dy. \tag{32}$$

The denominator of the conditional mechanism, after multiplication by $a_n n f_\tau$, is $R_n(1)$.

Theorem G.7. *Under Assumption G.1, if $a_n \to 0$ and $na_n \to \infty$, then for every bounded continuous h,*

$$R_n(h) \longrightarrow_{\mathbb{P}} \int_{\mathbb{R}} h(t)e^{-|t|}\,dt. \tag{33}$$

In particular,

$$a_n n f_\tau \sum_{j=0}^{n} \Delta_{j,n} e^{-a_n|j-k|} \longrightarrow_{\mathbb{P}} 2. \tag{34}$$

Proof. Fix h and write $H = \|h\|_\infty$. The argument has three parts.

Step 1: compact rank windows. For $r = j - k$, set

$$t_{r,n} = a_n n f_\tau \{X_{(k+r)} - X_{(k)}\}.$$

Because $q \in (A, B)$, for every fixed M all clipped gaps with $|r| \le M/a_n$ coincide with the full order-statistic gaps with probability tending to one. On that event, after the change of variable $t = a_n n f_\tau (y - X_{(k)})$, the contribution of the gap $j = k + r$ equals

$$e^{-a_n|r|} \int_{t_{r,n}}^{t_{r+1,n}} h(t)\,dt.$$

For fixed M, let $R_{n,M}(h)$ be the sum over $|r| \le M/a_n$, omitting at most one boundary gap on each side. By Lemma G.6,

$$\max_{|r|\le M/a_n} |t_{r,n} - a_n r| \to_{\mathbb{P}} 0$$

and the largest mesh length in this window converges to zero in probability. Therefore, on every event on which these two quantities are small, the step function that equals $e^{-a_n|r|}$ on $[t_{r,n}, t_{r+1,n})$ converges uniformly on compact subsets of $[-M, M]$ to $e^{-|t|}$. Since h is uniformly continuous on compact sets,

$$R_{n,M}(h) \longrightarrow_{\mathbb{P}} \int_{-M}^{M} h(t)e^{-|t|}\,dt \tag{35}$$

up to endpoint intervals whose total length is $o_{\mathbb{P}}(1)$. To see the Riemann step explicitly, write

$$\begin{aligned}\left| R_{n,M}(h) - \sum_{|r|\le M/a_n} \int_{t_{r,n}}^{t_{r+1,n}} h(t)e^{-|t|}\,dt \right| \\ \le H \sum_{|r|\le M/a_n} \int_{t_{r,n}}^{t_{r+1,n}} \left| e^{-a_n|r|} - e^{-|t|} \right| dt,\end{aligned}$$

and the integrand supremum tends to zero because $\sup |t - a_n r| \to_{\mathbb{P}} 0$ over the relevant intervals. The union of those intervals converges to $[-M, M]$ in Hausdorff distance, which proves (35).

Step 2: moderate tails. Choose a fixed $\delta > 0$ small enough that $Q([\tau - 3\delta, \tau + 3\delta]) \subset (A, B)$ and Q is C_Q-Lipschitz on $[\tau - 3\delta, \tau + 3\delta]$, as in (25). Let $\mathcal{E}_n$ be the event that

$$U_{(k-\lfloor \delta n \rfloor)} \ge \tau - 2\delta, \qquad U_{(k+\lfloor \delta n \rfloor + 1)} \le \tau + 2\delta.$$

The law of large numbers for order statistics implies $\mathbb{P}(\mathcal{E}_n) \to 1$. On $\mathcal{E}_n$, every clipped gap with $|j - k| \le \delta n$ is the full data gap and satisfies

$$\Delta_{j,n} \le C_Q \{U_{(j+1)} - U_{(j)}\}.$$

Let $T_{n,M}^{\mathrm{mod}}$ be the absolute contribution to (32) from $M/a_n < |j - k| \le \delta n$. On $\mathcal{E}_n$,

$$T_{n,M}^{\mathrm{mod}} \le H a_n n f_\tau C_Q \sum_{M/a_n < |j-k| \le \delta n} e^{-a_n|j-k|} \{U_{(j+1)} - U_{(j)}\}.$$

Every uniform spacing has expectation $1/(n+1)$. Therefore

$$\mathbb{E}\big[T_{n,M}^{\mathrm{mod}}\mathbf{1}(\mathcal{E}_n)\big] \le HC_Q f_\tau \frac{a_n n}{n+1} \sum_{|r|>M/a_n} e^{-a_n|r|}$$
$$\le CH\frac{a_n e^{-M}}{1-e^{-a_n}} \le C'He^{-M} \tag{36}$$

for all sufficiently large n, because $a_n/(1-e^{-a_n}) \to 1$. Markov's inequality shows that the moderate-tail contribution is $O_{\mathbb{P}}(e^{-M})$ uniformly in large n.

Step 3: far tails. For $|j-k| > \delta n$, the exponential weight is at most $e^{-a_n\delta n}$. Since the clipped gaps partition $[A,B]$ and hence sum to $B-A$, the absolute far-tail contribution is bounded by

$$Ha_n n f_\tau (B-A)e^{-a_n\delta n} \longrightarrow 0, \tag{37}$$

where the last step follows from $a_n n \to \infty$ and $xe^{-\delta x} \to 0$.

Combining (35), (36), and (37), first let $n \to \infty$ and then $M \to \infty$. Since $\int_{|t|>M} e^{-|t|}\,dt = 2e^{-M}$, we obtain (33). Taking $h \equiv 1$ gives (34). □

G.6 Conditional Laplace convergence and stable independence

Theorem G.8. *Under Assumption G.1, suppose*

$$\varepsilon_n \to 0, \qquad n\varepsilon_n \to \infty.$$

Then

$$\mathcal{L}\left(\frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n)\,\middle|\, D_n\right) \rightsquigarrow \mathrm{Lap}(0,1) \quad \textit{in probability.}$$

Proof. Let

$$T_n = a_n n f_\tau(\widetilde{q}_n - \widehat{q}_n), \qquad a_n = \varepsilon_n/2.$$

For bounded continuous h, the exact conditional density (21) and the gap decomposition give

$$\mathbb{E}\{h(T_n) \mid D_n\} = \frac{\sum_{j=0}^n e^{-a_n|j-k|}\int_{I_{j,n}} h\{a_n n f_\tau(y - X_{(k)})\}\,dy}{\sum_{j=0}^n \Delta_{j,n} e^{-a_n|j-k|}}$$
$$= \frac{R_n(h)}{R_n(1)}.$$

By Theorem G.7,

$$R_n(h) \to_{\mathbb{P}} \int_{\mathbb{R}} h(t)e^{-|t|}\,dt, \qquad R_n(1) \to_{\mathbb{P}} 2.$$

The denominator limit is strictly positive, so Slutsky's theorem applied to the ratio yields

$$\mathbb{E}\{h(T_n) \mid D_n\} \to_{\mathbb{P}} \frac{1}{2}\int_{\mathbb{R}} h(t)e^{-|t|}\,dt.$$

The right-hand side is the expectation of $h(L)$ for a standard Laplace variable L. This proves conditional weak convergence in probability. □

Lemma G.9. *Let $\mathcal{G}_n$ be sigma-fields and let T_n be random variables such that*

$$\mathbb{E}\{h(T_n) \mid \mathcal{G}_n\} \to_{\mathbb{P}} \mathbb{E}h(T)$$

for every bounded continuous h, where T has a fixed, nonrandom law. If Y_n is $\mathcal{G}_n$-measurable and $Y_n \rightsquigarrow Y$, then

$$(Y_n, T_n) \rightsquigarrow (Y, T^*),$$

where T^* has the law of T and is independent of Y.

Proof. It is enough first to consider product test functions $g(y)h(t)$ with g, h bounded and continuous. Since the conditional expectations are bounded and converge in probability to a constant, they also converge in L^1. Therefore

$$\begin{aligned}\mathbb{E}\{g(Y_n)h(T_n)\} &= \mathbb{E}\left[g(Y_n)\mathbb{E}\{h(T_n) \mid \mathcal{G}_n\}\right] \\ &= \mathbb{E}g(Y_n)\,\mathbb{E}h(T) + o(1) \\ &\longrightarrow \mathbb{E}g(Y)\,\mathbb{E}h(T).\end{aligned}$$

The marginal sequences are tight. Finite linear combinations of product functions form a convergence-determining algebra on compact sets; tightness extends the conclusion to all bounded continuous functions. The limiting joint law is the product of the laws of Y and T. □

Corollary G.10. *Under the assumptions of Theorem G.8, the variable*

$$\frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n)$$

converges stably, relative to the sample sigma-field, to a standard Laplace variable independent of every weak limit of a sample-measurable statistic.

Proof. Apply Lemma G.9 with $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$ and use Theorem G.8. □

G.7 The fixed-privacy perturbation is of order n^{-1}

For privacy budgets bounded away from zero, the selected rank stays within a tight number of gaps around the target. Each local gap has width of order n^{-1}, which gives the perturbation bound below. The first lemma quantifies the central spacing used to control the mechanism's denominator.

Lemma G.11. *Under Assumption G.1,*

$$nf_\tau\Delta_{k,n} \rightsquigarrow E, \qquad E \sim \mathrm{Exp}(1). \tag{38}$$

Consequently, for every $\eta > 0$ there are $0 < c < C < \infty$ such that, for all sufficiently large n,

$$\mathbb{P}\{c/n \le \Delta_{k,n} \le C/n\} \ge 1 - \eta.$$

Proof. In the normalized-exponential representation,

$$n\{U_{(k+1)} - U_{(k)}\} = \frac{n}{S}E_{k+1}.$$

Although the index $k+1$ changes with n, E_{k+1} has the standard exponential distribution for every n. Also $n/S \to 1$ in probability. Since $q \in (A, B)$ and $X_{(k)}, X_{(k+1)} \to_{\mathbb{P}} q$, the clipped target gap equals the full spacing $X_{(k+1)} - X_{(k)}$ with probability tending to one. The local inverse-CDF expansion in Lemma G.3 therefore gives

$$f_\tau\Delta_{k,n} = \{U_{(k+1)} - U_{(k)}\}\{1 + o_{\mathbb{P}}(1)\}.$$

Slutsky's theorem proves (38); the probability statement follows from tightness and the continuity of the exponential law at zero. □

Lemma G.12. *Under Assumption G.1, if $\inf_n \varepsilon_n > 0$, then*

$$J_n - k_n = O_{\mathbb{P}}(1).$$

Proof. Write $a_0 = \inf_n a_n > 0$ and

$$Z_n = \sum_{j=0}^{n} \Delta_{j,n} e^{-a_n|j-k|}, \qquad N_{n,M} = \sum_{|j-k|>M} \Delta_{j,n} e^{-a_n|j-k|}.$$

Then

$$\mathbb{P}(|J_n - k| > M \mid D_n) = N_{n,M}/Z_n.$$

Choose a small fixed $\delta > 0$ such that Q is locally Lipschitz on the corresponding probability interval. As in the moderate-tail argument of Theorem G.7,

$$\mathbb{E}\left[n \sum_{M<|j-k|\le \delta n} \Delta_{j,n} e^{-a_n|j-k|}; \mathcal{E}_n\right] \le C \sum_{|r|>M} e^{-a_0|r|} \le C' e^{-a_0 M}. \tag{39}$$

The far part is deterministically bounded by

$$n(B-A)e^{-a_0\delta n} \to 0.$$

Therefore $nN_{n,M}$ can be made arbitrarily small in probability by first taking M large and then n large.

The denominator obeys $Z_n \ge \Delta_{k,n}$. By Lemma G.11, for every $\eta > 0$ one can choose $c > 0$ so that

$$\limsup_n \mathbb{P}(nZ_n < c) \le \eta.$$

On the event $nZ_n \ge c$,

$$\mathbb{P}(|J_n - k| > M \mid D_n) \le nN_{n,M}/c.$$

To pass from the conditional ratio to the joint probability, fix $c > 0$ and use

$$\begin{aligned}\mathbb{P}(|J_n - k| > M) &= \mathbb{E}\left(\frac{N_{n,M}}{Z_n}\right) \\ &\le \mathbb{P}(\mathcal{E}_n^c) + \mathbb{P}(nZ_n < c) + \mathbb{E}\left[\min\left\{1, \frac{nN_{n,M}}{c}\right\}; \mathcal{E}_n\right].\end{aligned}$$

The first term tends to zero. The second can be made arbitrarily small by choosing c small, uniformly in large n, using Lemma G.11. For fixed c, the last term is bounded by $c^{-1}\mathbb{E}[nN_{n,M}; \mathcal{E}_n]$ and hence tends to zero uniformly in large n as $M \to \infty$ by (39) and the deterministic far-tail bound. Therefore

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(|J_n - k| > M) = 0,$$

which is the claimed tightness. □

Theorem G.13. *Under Assumption G.1, if* $\inf_n \varepsilon_n > 0$*, then*

$$\widetilde{q}_n - \widehat{q}_n = O_{\mathbb{P}}(n^{-1}). \tag{40}$$

Proof. Fix $\eta > 0$. By Lemma G.12, choose M such that $\mathbb{P}(|J_n - k| > M) < \eta$ for all sufficiently large n. On the complementary event, $\widetilde{q}_n$ lies between two order statistics whose ranks differ from k by at most $M + 1$. Hence

$$|\widetilde{q}_n - \widehat{q}_n| \le \max_{|r|\le M+1} |X_{(k+r)} - X_{(k)}|.$$

Because $q \in (A, B)$, the relevant fixed collection of clipped gaps equals the corresponding full spacings with probability tending to one. For each fixed M, the normalized-exponential representation and the local inverse-CDF expansion imply

$$n \max_{|r|\le M+1} |X_{(k+r)} - X_{(k)}| = O_{\mathbb{P}}(1),$$

because it is bounded by a finite sum of $O_{\mathbb{P}}(1)$ exponential spacings divided by f_τ. This proves (40). □

G.8 The classical sample-quantile expansion under minimal smoothness

Let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \le x\}, \qquad G_{\tau,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\tau - \mathbf{1}(X_i \le q)\}.$$

Lemma G.14. *Under Assumption G.1,*

$$\widehat{q}_n - q = O_{\mathbb{P}}(n^{-1/2}).$$

Proof. Fix $C > 0$ and put $x_{n,+} = q + C/\sqrt{n}$. By local differentiability,

$$F(x_{n,+}) = \tau + f_\tau C/\sqrt{n} + o(n^{-1/2}).$$

The event $\widehat{q}_n > x_{n,+}$ implies that fewer than k observations are at most $x_{n,+}$. The corresponding binomial count has mean $nF(x_{n,+}) = n\tau + f_\tau C\sqrt{n} + o(\sqrt{n})$. The binomial central limit theorem therefore shows

$$\limsup_n \mathbb{P}(\widehat{q}_n > x_{n,+}) \le \Phi(-f_\tau C/\sqrt{\tau(1-\tau)}) + o(1).$$

This bound tends to zero as $C \to \infty$. The lower tail follows identically from $x_{n,-} = q - C/\sqrt{n}$. □

Lemma G.15. *For every fixed $C < \infty$,*

$$\sup_{|x-q| \le C/\sqrt{n}} \sqrt{n}\,|(F_n - F)(x) - (F_n - F)(q)| \longrightarrow_{\mathbb{P}} 0. \tag{41}$$

Proof. Write P_n for the empirical measure and P for the population law. For $x \ge q$, the increment equals

$$(F_n - F)(x) - (F_n - F)(q) = (P_n - P)\mathbf{1}\{q < X \le x\},$$

and for $x < q$ it is the negative of the same expression for an interval $(x, q]$. By local continuity and positivity of f, there is $C_1 < \infty$ such that all these intervals have probability at most $p_n = C_1/\sqrt{n}$ whenever $|x - q| \le C/\sqrt{n}$ and n is large.

We prove the upper-side bound; the lower side is identical. Let

$$\mathcal{A}_n^+ = \{(q, x] : q \le x \le q + C/\sqrt{n}\}.$$

Symmetrization with independent Rademacher variables $\xi_1, \dots, \xi_n$ gives

$$\mathbb{E} \sup_{A \in \mathcal{A}_n^+} \left|\sqrt{n}(P_n - P)\mathbf{1}_A\right| \le 2\mathbb{E} \sup_{A \in \mathcal{A}_n^+} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \mathbf{1}\{X_i \in A\}\right|. \tag{42}$$

Condition on the observations and let M_n be the number falling in the largest interval $(q, q + C/\sqrt{n}]$. Order those M_n observations. Because the class is nested, the conditional supremum in (42) is the maximum of absolute partial sums of M_n Rademacher variables. Doob's L^2 maximal inequality and Cauchy–Schwarz imply that its conditional expectation is at most $2\sqrt{M_n}$. Consequently,

$$\mathbb{E} \sup_{A \in \mathcal{A}_n^+} \left|\sqrt{n}(P_n - P)\mathbf{1}_A\right| \le \frac{4}{\sqrt{n}} \mathbb{E}\sqrt{M_n} \le 4\sqrt{\frac{\mathbb{E}M_n}{n}} \le 4\sqrt{p_n} = O(n^{-1/4}).$$

The same bound holds for the lower-side nested class. Their union covers all $|x - q| \le C/\sqrt{n}$. Markov's inequality proves (41). □

The following is the standard first-order quantile expansion under local regularity (van der Vaart, 1998, Chapter 21). The proof is included to make the triangular interaction with the private perturbation fully explicit.

Theorem G.16. *Under Assumption G.1,*

$$\sqrt{n} f_\tau(\widehat{q}_n - q) = G_{\tau,n} + o_{\mathbb{P}}(1),$$

and hence

$$\sqrt{n}(\widehat{q}_n - q) \rightsquigarrow N\left(0, \frac{\tau(1-\tau)}{f_\tau^2}\right). \tag{43}$$

Proof. By root-n localization, $\widehat{q}_n$ belongs to the locally absolutely continuous neighborhood with probability tending to one. On that event there are almost surely no ties at $\widehat{q}_n$, so $F_n(\widehat{q}_n) = k/n$. All following identities are used on this event; its complement has probability $o(1)$. Decompose

$$\frac{k}{n} - \tau = \{F_n(q) - \tau\} + \{F(\widehat{q}_n) - F(q)\} + \{(F_n - F)(\widehat{q}_n) - (F_n - F)(q)\}. \tag{44}$$

The left side is $O(n^{-1})$. By Lemma G.14 and continuity of f at q,

$$F(\widehat{q}_n) - F(q) = f_\tau(\widehat{q}_n - q) + o_{\mathbb{P}}(n^{-1/2}).$$

The last term in (44) is $o_{\mathbb{P}}(n^{-1/2})$ by combining root-n localization with Lemma G.15: first restrict to $|\widehat{q}_n - q| \le C/\sqrt{n}$, then let $C \to \infty$. Multiplication by $\sqrt{n}$ gives

$$\sqrt{n} f_\tau(\widehat{q}_n - q) = -\sqrt{n}\{F_n(q) - \tau\} + o_{\mathbb{P}}(1) = G_{\tau,n} + o_{\mathbb{P}}(1).$$

Finally, $G_{\tau,n}$ is a normalized sum of centered Bernoulli variables with variance $\tau(1-\tau)$, so the ordinary central limit theorem proves (43). □

The same argument permits any integer rank with $k_n - n\tau = o(\sqrt{n})$: the left side of (44) is then $o(n^{-1/2})$. More generally, if $(k_n - n\tau)/\sqrt{n} \to b$, the root-$n$ expansion acquires the deterministic shift b. The conditional privacy limit remains centered at $X_{(k_n)}$, since the spacing argument itself needs only $k_n/n \to \tau$.

G.9 Joint limit, phase transition, and stochastic rate

Theorem G.17. *Under Assumption G.1, if $\varepsilon_n \to 0$ and $n\varepsilon_n \to \infty$, then*

$$\left(\sqrt{n} f_\tau(\widehat{q}_n - q), \frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n)\right) \rightsquigarrow (G_\tau, L),$$

where $G_\tau \sim N\{0, \tau(1-\tau)\}$, $L \sim \mathrm{Lap}(0,1)$, and the two limits are independent.

Proof. The first coordinate is sample measurable and converges to G_τ by Theorem G.16. The conditional law of the second coordinate converges in probability to the fixed Laplace law by Theorem G.8. Apply Lemma G.9. □

Theorem G.18. *Under Assumption G.1:*

(a) if $\varepsilon_n\sqrt{n} \to \infty$, then

$$\sqrt{n} f_\tau(\widetilde{q}_n - q) \rightsquigarrow G_\tau;$$

(b) if $\varepsilon_n\sqrt{n} \to \kappa \in (0,\infty)$, then

$$\sqrt{n} f_\tau(\widetilde{q}_n - q) \rightsquigarrow G_\tau + \frac{2}{\kappa} L;$$

(c) if $\varepsilon_n \to 0$, $n\varepsilon_n \to \infty$, and $\varepsilon_n\sqrt{n} \to 0$, then

$$n\varepsilon_n f_\tau(\widetilde{q}_n - q) \rightsquigarrow 2L.$$

Proof. Always write

$$\widetilde{q}_n - q = (\widehat{q}_n - q) + (\widetilde{q}_n - \widehat{q}_n). \tag{45}$$

To prove part (a) without imposing a limit on ε_n, use the subsequence principle. Take any subsequence. If its privacy budgets have a further subsequence bounded below by a positive constant, then Theorem G.13 gives $\sqrt{n}(\widetilde{q}_n - \widehat{q}_n) = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(1)$ on that further subsequence. Otherwise it has a further subsequence on which $\varepsilon_n \to 0$. Since $\varepsilon_n\sqrt{n} \to \infty$, that further subsequence also satisfies $n\varepsilon_n \to \infty$, and the conditional Laplace theorem gives

$$\sqrt{n} f_\tau(\widetilde{q}_n - \widehat{q}_n) = \frac{2}{\varepsilon_n\sqrt{n}} \left\{ \frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n) \right\} = o_{\mathbb{P}}(1).$$

Thus every subsequence has a further subsequence on which the privacy perturbation is $o_{\mathbb{P}}(n^{-1/2})$; hence this holds for the full sequence. Combining with Theorem G.16 proves part (a).

For parts (b) and (c), $\varepsilon_n\sqrt{n}$ has a finite limit and therefore $\varepsilon_n \to 0$; the assumption $n\varepsilon_n \to \infty$ is automatic in part (b) and imposed in part (c). Define

$$V_n = \frac{n\varepsilon_n f_\tau}{2}(\widetilde{q}_n - \widehat{q}_n).$$

By Theorem G.17, $(\sqrt{n}f_\tau(\widehat{q}_n - q), V_n) \rightsquigarrow (G_\tau, L)$. Multiplying the second coordinate by $2/(\varepsilon_n\sqrt{n})$ in (45) proves (a) when this factor tends to zero and proves (b) when it tends to $2/\kappa$.

For part (c), multiply (45) by $n\varepsilon_n f_\tau$:

$$n\varepsilon_n f_\tau(\widetilde{q}_n - q) = \varepsilon_n\sqrt{n}\{\sqrt{n}f_\tau(\widehat{q}_n - q)\} + 2V_n.$$

The first term tends to zero in probability and the second tends to $2L$. □

Corollary G.19. *Under Assumption G.1, if $n\varepsilon_n \to \infty$, then*

$$\widetilde{q}_n - q = O_{\mathbb{P}}\left(n^{-1/2} + (n\varepsilon_n)^{-1}\right). \tag{46}$$

Proof. The sample error is always $O_{\mathbb{P}}(n^{-1/2})$ by Theorem G.16. It remains to control the privacy perturbation. Take any subsequence. It has a further subsequence on which either the privacy budgets are bounded below by a positive constant or $\varepsilon_n \to 0$. In the first case, Theorem G.13 gives $\widetilde{q}_n - \widehat{q}_n = O_{\mathbb{P}}(n^{-1})$, which is absorbed by the $n^{-1/2}$ term in (46). In the second case, the assumed relation $n\varepsilon_n \to \infty$ and Theorem G.8 give $\widetilde{q}_n - \widehat{q}_n = O_{\mathbb{P}}\{(n\varepsilon_n)^{-1}\}$. Therefore (46) holds on a further subsequence of every subsequence, which proves it for the full sequence. □

G.10 Triangular quantile levels

The feasible sparsity estimator uses levels $\tau \pm h_n$. We therefore record a uniform local version of the preceding rate.

Lemma G.20. *Under Assumption G.1, there is $\eta > 0$ such that, for every $u \in [\tau - \eta, \tau + \eta]$, the quantile $Q(u)$ is unique, $f\{Q(u)\}$ is bounded above and bounded away from zero, and the constants in the local inverse-CDF and order-spacing arguments can be chosen uniformly in u.*

Proof. By the assumed limit $f(x) \to f_\tau > 0$ as $x \to q$, choose $r_0 > 0$ such that

$$\frac{f_\tau}{2} \le f(x) \le \frac{3f_\tau}{2} \qquad \text{for almost every } x \in (q - r_0, q + r_0).$$

Local absolute continuity then implies, whenever $x < y$ lie in this interval,

$$\frac{f_\tau}{2}(y - x) \le F(y) - F(x) \le \frac{3f_\tau}{2}(y - x).$$

Hence F is strictly increasing there. Let

$$\eta = \frac{1}{2}\min\{F(q + r_0) - \tau, \tau - F(q - r_0)\} > 0.$$

For every $u \in [\tau - \eta, \tau + \eta]$, the generalized inverse $Q(u)$ lies in $(q - r_0, q + r_0)$ and is the unique solution of $F(x) = u$ within that interval. The preceding two-sided inequality gives, uniformly in such u, v,

$$\frac{2}{3f_\tau}|u - v| \le |Q(u) - Q(v)| \le \frac{2}{f_\tau}|u - v|.$$

Finally, since $Q(u) \to q$ uniformly as the probability neighborhood shrinks and $f(x) \to f_\tau$ at q, decreasing η if necessary gives uniform upper and lower bounds for $f\{Q(u)\}$. Every local inverse-CDF and spacing estimate used later depends only on these displayed deterministic bounds, so its constants can be chosen uniformly in u. □

Proposition G.21. *Let $\tau_n \to \tau$, $q_n = Q(\tau_n)$, and let $\widetilde{q}_{\tau_n,n}$ be the rank exponential mechanism with integer target k_n satisfying $|k_n/n - \tau_n| \le 1/n$ and privacy budget $\eta_n > 0$. If $n\eta_n \to \infty$, then*

$$\widetilde{q}_{\tau_n,n} - q_n = O_{\mathbb{P}}\left(n^{-1/2} + (n\eta_n)^{-1}\right). \tag{47}$$

Proof. Take an arbitrary subsequence. It has a further subsequence on which either η_n is bounded below by a positive constant or $\eta_n \to 0$.

In the first case, the selected rank is tight uniformly in τ_n because the denominator can again be bounded below by the target gap and the geometric rank tail is independent of the precise center,

while Lemma G.20 supplies uniform local Lipschitz constants. A fixed number of spacings around k_n is therefore $O_{\mathbb{P}}(n^{-1})$ uniformly. The ordinary triangular sample quantile satisfies

$$\widehat{q}_{\tau_n,n} - q_n = O_{\mathbb{P}}(n^{-1/2})$$

by the same binomial localization argument as Lemma G.14, because $\tau_n \to \tau \in (0,1)$ and $f(q_n) \to f_\tau > 0$.

In the second case, put $a_n = \eta_n/2$. Then $a_n \to 0$ and $na_n \to \infty$. The local-grid proof uses only $k_n/n \to \tau$, $f(q_n) \to f_\tau$, and uniform inverse-CDF linearity on the neighborhood supplied by Lemma G.20. Consequently,

$$\frac{n\eta_n f(q_n)}{2}\{\widetilde{q}_{\tau_n,n} - \widehat{q}_{\tau_n,n}\} = O_{\mathbb{P}}(1),$$

while the sample component remains $O_{\mathbb{P}}(n^{-1/2})$. Since $f(q_n)$ is uniformly bounded above and away from zero, this yields (47). Every subsequence has a further subsequence with the required bound, so the subsequence principle proves the full assertion. □

G.11 Privacy accounting for feasible inference

This section separates three notions that are often conflated. A confidence interval computed from already private outputs is post-processing and has zero additional privacy cost. A private estimator of the sparsity $1/f(q)$ is itself a data release and must be included in the mechanism. Whether that nuisance release adds privacy parameters or instead uses parallel composition depends on whether it reuses the center records.

Definition G.22. A mechanism M is (ε, δ)-differentially private if, for every pair of replacement-adjacent datasets D, D' and measurable S,

$$\mathbb{P}\{M(D) \in S\} \le e^{\varepsilon}\mathbb{P}\{M(D') \in S\} + \delta.$$

Pure privacy is the case $\delta = 0$.

Post-processing and adaptive sequential composition are standard; see Dwork & Roth (2014, Proposition 2.1 and Appendix B). The lemma below also states and proves the disjoint-block form used here.

Lemma G.23. *The following statements hold.*

(a) If M_j is pure ε_j-DP on the same records, possibly chosen adaptively from previous private outputs, then their joint release is pure $(\sum_j \varepsilon_j)$-DP.

(b) Let $D = (D_1, \dots, D_K)$ be a data-independent partition into disjoint record blocks. Suppose that, when the changed record belongs to block k, the joint output is $(\varepsilon_k, \delta_k)$-DP as a function of D_k after conditioning on all randomness and outputs generated from the other blocks in the causal order of the algorithm. Then the complete release is

$$\left(\max_k \varepsilon_k, \max_k \delta_k\right)\text{-DP}.$$

(c) Any randomized or deterministic function of a private output has the same privacy guarantee.

Proof. For (a), first consider two mechanisms. Write their conditional kernels as $K_1(D, dy_1)$ and $K_2(D, y_1, dy_2)$. For every nonnegative measurable g and adjacent D, D',

$$\int g(y_1)K_1(D, dy_1) \le e^{\varepsilon_1} \int g(y_1)K_1(D', dy_1).$$

Apply this inequality with $g(y_1) = K_2(D, y_1, A_{y_1})$, where $A_{y_1} = \{y_2 : (y_1, y_2) \in A\}$. Conditional ε_2-privacy gives $K_2(D, y_1, A_{y_1}) \le e^{\varepsilon_2} K_2(D', y_1, A_{y_1})$ for every y_1. Combining the two inequalities bounds the probability of the joint event A by $e^{\varepsilon_1+\varepsilon_2}$ times its probability under D'. Induction proves the general statement.

For (b), adjacent full datasets differ in exactly one block, say block k. All other block datasets are identical. Conditional on the outputs generated before the block-k private operation, the required likelihood inequality is therefore the one for $(\varepsilon_k, \delta_k)$. Integrating over the identical law of all quantities generated from unchanged blocks preserves the same inequality. Outputs generated later from unchanged block data and the block-k private output form a Markov kernel and are covered by post-processing. Hence the full mechanism is $(\varepsilon_k, \delta_k)$-DP for this neighboring pair. Taking the maximum over the possible changed blocks proves the claim. Part (c) follows by applying the defining inequality to the inverse image of each output event. □

G.12 QExp inference on one sample

Let the center and side budgets be $\varepsilon_{0,n}, \varepsilon_{+,n}, \varepsilon_{-,n} > 0$. On the same sample, independently conditional on the data, release QExp estimates at τ, $\tau + h_n$, and $\tau - h_n$. Define

$$\widetilde{s}_n^{\mathrm{raw}} = \frac{\widetilde{q}_{+,n} - \widetilde{q}_{-,n}}{2h_n}, \qquad \widetilde{s}_n = \Pi_{[c_n, C_n]}(\widetilde{s}_n^{\mathrm{raw}}), \qquad \widetilde{f}_n = 1/\widetilde{s}_n,$$

where $c_n \downarrow 0$ and $C_n \uparrow \infty$.

Proposition G.24. *The joint tuple consisting of all three QExp releases, $\widetilde{s}_n$, $\widetilde{f}_n$, and any confidence interval formed from them is pure $(\varepsilon_{0,n} + \varepsilon_{+,n} + \varepsilon_{-,n})$-DP.*

Proof. Each QExp release is pure DP by Proposition G.2. Because the three mechanisms access the same records, Lemma G.23(a) adds their privacy parameters. The sparsity estimate, reciprocal, critical value, and interval are functions of the private tuple and public constants, so Lemma G.23(c) adds no privacy loss. □

Theorem G.25. *Under Assumption G.1, suppose*

$$h_n \to 0, \qquad \sqrt{n}h_n \to \infty, \qquad nh_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty.$$

Then

$$\widetilde{s}_n \to_{\mathbb{P}} 1/f_\tau, \qquad \widetilde{f}_n \to_{\mathbb{P}} f_\tau.$$

Proof. By Lemma G.3, the quantile function Q is differentiable at τ and $Q'(\tau) = 1/f_\tau$. Hence

$$s_n^0 := \frac{Q(\tau + h_n) - Q(\tau - h_n)}{2h_n} \longrightarrow 1/f_\tau. \tag{48}$$

For $\nu \in \{+, -\}$, Proposition G.21, with privacy parameter $\varepsilon_{\nu,n}$, gives

$$\widetilde{q}_{\nu,n} - Q(\tau + \nu h_n) = O_{\mathbb{P}}\left(n^{-1/2} + (n\varepsilon_{\nu,n})^{-1}\right).$$

The conditions imply $n\varepsilon_{\nu,n} \to \infty$, because $n\varepsilon_{\nu,n} \geq nh_n\varepsilon_{\nu,n}$ for all sufficiently large n with $h_n < 1$. Therefore

$$\begin{aligned}\widetilde{s}_n^{\mathrm{raw}} - s_n^0 &= \frac{\{\widetilde{q}_{+,n} - Q(\tau + h_n)\} - \{\widetilde{q}_{-,n} - Q(\tau - h_n)\}}{2h_n} \\ &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}h_n} + \frac{1}{nh_n\varepsilon_{+,n}} + \frac{1}{nh_n\varepsilon_{-,n}}\right) = o_{\mathbb{P}}(1).\end{aligned}$$

Together with (48), this proves $\widetilde{s}_n^{\mathrm{raw}} \to_{\mathbb{P}} 1/f_\tau > 0$. Since the truncation interval eventually contains every fixed neighborhood of $1/f_\tau$, the projection is inactive with probability tending to one. Continuity of $x \mapsto 1/x$ at $1/f_\tau$ proves the result for $\widetilde{f}_n$. □

For $\kappa > 0$, let

$$c_{Q,1-\alpha}(\tau, \kappa) = \inf\left\{c : \mathbb{P}\left(|G_\tau + 2L/\kappa| \leq c\right) \geq 1 - \alpha\right\}.$$

Lemma G.26. *For fixed $\tau \in (0,1)$ and $\alpha \in (0,1)$, the map $\kappa \mapsto c_{Q,1-\alpha}(\tau, \kappa)$ is continuous on $(0, \infty)$ and*

$$\begin{aligned} c_{Q,1-\alpha}(\tau, \kappa) &\longrightarrow z_{1-\alpha/2}\sqrt{\tau(1-\tau)} && \text{as } \kappa \to \infty, \\ \kappa c_{Q,1-\alpha}(\tau, \kappa) &\longrightarrow 2\log(1/\alpha) && \text{as } \kappa \downarrow 0. \end{aligned} \tag{49}$$

Proof. Construct all variables on one space with fixed independent G_τ and L. If $\kappa_j \to \kappa \in (0,\infty)$, then $G_\tau + 2L/\kappa_j \to G_\tau + 2L/\kappa$ almost surely. Every such convolution has a continuous strictly positive density, so the distribution of its absolute value is continuous and strictly increasing on $[0,\infty)$. Weak convergence therefore implies convergence of its unique $(1-\alpha)$ quantile, proving continuity.

As $\kappa \to \infty$, the variable converges almost surely to G_τ. The $(1-\alpha)$ quantile of $|G_\tau|$ is $z_{1-\alpha/2}\sqrt{\tau(1-\tau)}$, which proves (49). As $\kappa \downarrow 0$,

$$\kappa\left(G_\tau + \frac{2L}{\kappa}\right) = \kappa G_\tau + 2L \longrightarrow 2L$$

almost surely. Scaling a positive absolute-value quantile by κ shows that $\kappa c_{Q,1-\alpha}(\tau,\kappa)$ is the $(1-\alpha)$ quantile of $|\kappa G_\tau + 2L|$. The latter converges to the corresponding quantile of $|2L|$. Since $\mathbb{P}(|L| \le x) = 1 - e^{-x}$ for $x \ge 0$, that quantile equals $2\log(1/\alpha)$. □

Theorem G.27. *Assume the conditions of Theorem G.25 and $n\varepsilon_{0,n} \to \infty$. Put $\kappa_{0,n} = \varepsilon_{0,n}\sqrt{n}$ and define*

$$\mathcal{C}_n^{Q,\mathrm{SC}} = \left[\widetilde{q}_{0,n} \pm \frac{c_{Q,1-\alpha}(\tau,\kappa_{0,n})}{\sqrt{n}\,\widetilde{f}_n}\right].$$

Then $\mathbb{P}(q \in \mathcal{C}_n^{Q,\mathrm{SC}}) \to 1-\alpha$. No convergence assumption on $\kappa_{0,n}$ is required.

Proof. It suffices to prove that every subsequence has a further subsequence along which the coverage tends to $1-\alpha$. Given any subsequence, compactness of the extended half-line yields a further subsequence on which $\kappa_{0,n} \to \kappa \in [0,\infty]$.

If $\kappa = \infty$, Theorem G.18(a), including the nonvanishing-privacy case, gives

$$\sqrt{n} f_\tau(\widetilde{q}_{0,n} - q) \rightsquigarrow G_\tau.$$

By Lemma G.26, the critical value converges to the absolute Gaussian quantile. Since $\widetilde{f}_n / f_\tau \to_{\mathbb{P}} 1$, Slutsky's theorem gives coverage $1-\alpha$.

If $\kappa \in (0,\infty)$, Theorem G.18(b) gives

$$\sqrt{n} f_\tau(\widetilde{q}_{0,n} - q) \rightsquigarrow G_\tau + 2L/\kappa.$$

Continuity of the critical value and the limiting absolute-value distribution again gives the result.

If $\kappa = 0$, multiply the coverage event by $\kappa_{0,n}$. Because $\kappa_{0,n}\sqrt{n} = n\varepsilon_{0,n}$,

$$\left|\sqrt{n}\widetilde{f}_n(\widetilde{q}_{0,n} - q)\right| \le c_{Q,1-\alpha}(\tau,\kappa_{0,n})$$

is equivalent to

$$\left|n\varepsilon_{0,n}\widetilde{f}_n(\widetilde{q}_{0,n} - q)\right| \le \kappa_{0,n} c_{Q,1-\alpha}(\tau,\kappa_{0,n}).$$

The left side converges in distribution to $|2L|$ by Theorem G.18(c) and consistent studentization, while the right side converges to $2\log(1/\alpha)$ by Lemma G.26. The limiting distribution is continuous, so the coverage tends to $1-\alpha$. The subsequence principle completes the proof. □

Corollary G.28. *Let $E_n > 0$ satisfy $nE_n \to \infty$. Define*

$$\beta_n = \min\left\{\frac{1}{4}, (nE_n)^{-1/2}\right\}, \quad \varepsilon_{+,n} = \varepsilon_{-,n} = \beta_n E_n, \quad \varepsilon_{0,n} = (1-2\beta_n)E_n,$$

and

$$A_n = \min\{\sqrt{n}, n\beta_n E_n\}, \qquad h_n = \min\left\{\frac{\min(\tau,1-\tau)}{2}, A_n^{-1/2}\right\}.$$

Then the joint center estimate, side estimates, sparsity estimate, and confidence interval are pure E_n-DP,

$$\beta_n \to 0, \qquad \frac{\varepsilon_{0,n}}{E_n} \to 1, \qquad n\varepsilon_{0,n} \to \infty,$$

and

$$h_n \to 0, \qquad \sqrt{n}h_n \to \infty, \qquad nh_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty.$$

Consequently Theorem G.27 applies.

Proof. Because $nE_n \to \infty$, eventually $\beta_n = (nE_n)^{-1/2}$, so

$$\beta_n \to 0, \qquad n\beta_n E_n = \sqrt{nE_n} \to \infty.$$

It follows that $\varepsilon_{0,n}/E_n = 1 - 2\beta_n \to 1$ and $n\varepsilon_{0,n} = (1-2\beta_n)nE_n \to \infty$. Moreover, both arguments of the minimum defining A_n diverge, hence $A_n \to \infty$; the bandwidth cap is eventually inactive, so $h_n = A_n^{-1/2} \to 0$. Since $A_n \le \sqrt{n}$,

$$\sqrt{n}h_n = \frac{\sqrt{n}}{\sqrt{A_n}} \ge \sqrt{A_n} \to \infty.$$

Since $A_n \le n\beta_n E_n$,

$$nh_n\varepsilon_{+,n} = \frac{n\beta_n E_n}{\sqrt{A_n}} \ge \sqrt{A_n} \to \infty,$$

and the same calculation holds for $\varepsilon_{-,n}$. The three same-sample budgets add exactly to $\{1-2\beta_n + \beta_n + \beta_n\}E_n = E_n$, so Proposition G.24 gives pure E_n-DP. All hypotheses of Theorem G.27 therefore hold. □

G.13 Parallel-composition QExp inference

Partition the data deterministically into a center block D_0 of size n_0 and a nuisance block D_f of size m, with $n_0 + m = n$. Release the center QExp estimate from D_0 with budget $\varepsilon_{0,n}$. On D_f, release QExp estimates at $\tau \pm h_n$ with budgets $\varepsilon_{+,n}, \varepsilon_{-,n}$ satisfying $\varepsilon_{+,n} + \varepsilon_{-,n} \le \varepsilon_{f,n}$. Construct $\widetilde{f}_{f,n}$ from the side releases exactly as above.

Theorem G.29. *The complete release is pure* $\max(\varepsilon_{0,n}, \varepsilon_{f,n})$*-DP. Under Assumption G.1, if*

$$n_0 \to \infty, \quad m \to \infty, \quad n_0\varepsilon_{0,n} \to \infty, \quad h_n \to 0, \quad \sqrt{m}h_n \to \infty, \quad mh_n \min(\varepsilon_{+,n}, \varepsilon_{-,n}) \to \infty,$$

then $\widetilde{f}_{f,n} \to_{\mathbb{P}} f_\tau$*, and*

$$\mathcal{C}_n^{Q,\mathrm{PC}} = \left[\widetilde{q}_{0,n} \pm \frac{c_{Q,1-\alpha}(\tau, \varepsilon_{0,n}\sqrt{n_0})}{\sqrt{n_0}\,\widetilde{f}_{f,n}}\right]$$

has coverage tending to $1-\alpha$*. If* $n_0/n \to 1$*, replacing* $\sqrt{n_0}$ *by the full-sample first-order scale causes no asymptotic efficiency loss.*

Proof. The two side releases act on the same nuisance block and are pure $(\varepsilon_{+,n} + \varepsilon_{-,n})$-DP by sequential composition. The center and nuisance blocks are disjoint, so Lemma G.23(b) gives the maximum rather than the sum. All subsequent calculations are post-processing.

Apply Theorem G.25 with sample size m to obtain $\widetilde{f}_{f,n} \to_{\mathbb{P}} f_\tau$. Apply the QExp phase theorem and the proof of Theorem G.27 with sample size n_0 to the center release. The nuisance estimate is computed from an independent block, although only consistency is required for Slutsky's theorem. This proves coverage. Finally, $n_0/n \to 1$ implies $\sqrt{n}/\sqrt{n_0} \to 1$, so the ordinary quantile variance based on n_0 equals the full-sample variance to first order. □

Lemma G.30. *Let* $E_n > 0$ *satisfy* $nE_n \to \infty$*. Put*

$$r_n = \min(n, nE_n), \qquad m_n = \left\lfloor \frac{n}{r_n^{1/4}} \right\rfloor, \qquad n_{0,n} = n - m_n,$$

and, for sufficiently large n*,*

$$A_n = \min\{\sqrt{m_n}, m_n E_n\}, \qquad h_n = \min\left\{\frac{\min(\tau, 1-\tau)}{2}, A_n^{-1/2}\right\}.$$

Use budget E_n *for the center block and budgets* $E_n/2, E_n/2$ *for the two nuisance releases. Then*

$$m_n \to \infty, \quad m_n/n \to 0, \quad h_n \to 0, \quad \sqrt{m_n}h_n \to \infty, \quad m_n E_n h_n \to \infty.$$

The complete release is pure E_n*-DP and* $n_{0,n}/n \to 1$*.*

Proof. Because $r_n \to \infty$ and $r_n \le n$, $m_n/n \sim r_n^{-1/4} \to 0$, while $m_n \ge n^{3/4}/2 \to \infty$ for all sufficiently large n. If $E_n \ge 1$, then $r_n = n$ and $m_n E_n \ge m_n \to \infty$. If $E_n < 1$, then $r_n = nE_n$ and

$$m_n E_n \sim nE_n(nE_n)^{-1/4} = r_n^{3/4} \to \infty.$$

Thus both $\sqrt{m_n}$ and $m_n E_n$ diverge, so $A_n \to \infty$ and $h_n \to 0$. The bandwidth cap is eventually inactive. Since $A_n \le \sqrt{m_n}$, for all sufficiently large n,

$$\sqrt{m_n} h_n = \frac{\sqrt{m_n}}{\sqrt{A_n}} \ge \sqrt{A_n} \to \infty.$$

Similarly, $A_n \le m_n E_n$ gives $m_n E_n h_n \ge \sqrt{A_n} \to \infty$. Constants of $1/2$ in the side budgets do not change the limit. Privacy follows from Theorem G.29, and $n_{0,n}/n = 1 - m_n/n \to 1$. □

G.14 Gaussian one-step (GOS) quantile inference

The score release uses the analytic Gaussian mechanism of Balle & Wang (2018); the new work in this section is the end-to-end block accounting and the quantile-specific stochastic expansion.

Let the data be partitioned deterministically into a pilot block D_p of size m_p, a nuisance block D_f of size m_f, and a score block D_s of size N. The pilot is a QExp release at level τ with budget $\varepsilon_{p,n}$. On D_f, two QExp releases at $\tau \pm h_n$ have budgets $\varepsilon_{f,+,n}$ and $\varepsilon_{f,-,n}$, whose sum is at most $\varepsilon_{f,n}$. Writing the two releases as $\widetilde{q}_{f,+,n}$ and $\widetilde{q}_{f,-,n}$, define, for public $c_n \downarrow 0$ and $C_n \uparrow \infty$,

$$\widetilde{s}_{f,n}^{\mathrm{raw}} = \frac{\widetilde{q}_{f,+,n} - \widetilde{q}_{f,-,n}}{2h_n}, \qquad \widetilde{s}_{f,n} = \Pi_{[c_n, C_n]}(\widetilde{s}_{f,n}^{\mathrm{raw}}), \qquad \widetilde{f}_{f,n} = 1/\widetilde{s}_{f,n}.$$

On D_s, with index set $\mathcal{I}_s$, define

$$\psi_y(x) = \mathbf{1}\{x \le y\} - \tau, \qquad S_N(y) = \frac{1}{N} \sum_{i \in \mathcal{I}_s} \psi_y(X_i).$$

The Gaussian score release is

$$\widetilde{S}_n = S_N(\widetilde{q}_{p,n}) + \sigma_n Z, \qquad Z \sim N(0,1).$$

The unprojected and projected one-step estimators are

$$\bar{q}_n^G = \widetilde{q}_{p,n} - \frac{\widetilde{S}_n}{\widetilde{f}_{f,n}}, \qquad \widetilde{q}_n^G = \Pi_{[A,B]}(\bar{q}_n^G).$$

G.14.1 End-to-end privacy

Theorem G.31. *Suppose $\varepsilon_{s,n} > 0$ and $0 < \delta_{s,n} < 1$. Conditional on the pilot, the scalar query $S_N(\widetilde{q}_{p,n})$ has replacement sensitivity $1/N$. If σ_n satisfies*

$$\Phi\left(\frac{1}{2N\sigma_n} - \varepsilon_{s,n} N\sigma_n\right) - e^{\varepsilon_{s,n}} \Phi\left(-\frac{1}{2N\sigma_n} - \varepsilon_{s,n} N\sigma_n\right) \le \delta_{s,n}, \tag{50}$$

then the complete estimator, nuisance estimate, and confidence interval are

$$(\max\{\varepsilon_{p,n}, \varepsilon_{f,n}, \varepsilon_{s,n}\}, \delta_{s,n})\text{-}\textit{DP}.$$

Proof. Fix the pilot value y. Replacing one score-block record changes one summand $\mathbf{1}(X_i \le y) - \tau$ to another value in the same two-point set $\{-\tau, 1-\tau\}$, so the absolute change in the average is at most $1/N$. The exact privacy characterization for a scalar Gaussian mechanism with sensitivity Δ is

$$\Phi\left(\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) - e^{\varepsilon} \Phi\left(-\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) \le \delta;$$

substitution of $\Delta = 1/N$ yields (50) (Balle & Wang, 2018).

We now examine the location of the changed record. If it lies in D_p, the pilot is $\varepsilon_{p,n}$-DP. Conditional on its value, the score block and nuisance block are unchanged, and all subsequent operations form a Markov kernel; post-processing therefore preserves $\varepsilon_{p,n}$-privacy. If the changed record lies in D_f, the two side releases have combined pure privacy at most $\varepsilon_{f,n}$, and all uses of their output are post-processing. If it lies in D_s, the pilot and nuisance outputs have the same distribution under both datasets and, conditional on them, the Gaussian score release is $(\varepsilon_{s,n}, \delta_{s,n})$-DP. Taking the maximum over these three cases proves the claim by Lemma G.23(b). □

Corollary G.32. *For a target (E_n, δ_n) guarantee, set*

$$\varepsilon_{p,n} = \varepsilon_{f,n} = \varepsilon_{s,n} = E_n, \qquad \varepsilon_{f,+,n} = \varepsilon_{f,-,n} = E_n/2, \qquad \delta_{s,n} = \delta_n.$$

Then the complete GOS release is (E_n, δ_n)-DP.

Proof. The two nuisance releases compose to pure E_n-DP on their block. The pilot block is pure E_n-DP and the score block is (E_n, δ_n)-DP. Apply the maximum rule in Theorem G.31. □

G.14.2 Pilot and nuisance rates

Define

$$\begin{aligned} d_n &= m_p^{-1/2} + (m_p \varepsilon_{p,n})^{-1}, \\ \rho_{\nu,n} &= m_f^{-1/2} + (m_f \varepsilon_{f,\nu,n})^{-1}, \qquad \nu \in \{+,-\}, \\ b_Q(h) &= \left| \frac{Q(\tau+h) - Q(\tau-h)}{2h} - \frac{1}{f_\tau} \right|, \\ e_{f,n} &= b_Q(h_n) + \frac{\rho_{+,n} + \rho_{-,n}}{h_n}. \end{aligned}$$

Lemma G.33. *Under Assumption G.1, if $m_p \varepsilon_{p,n} \to \infty$, then*

$$\widetilde{q}_{p,n} - q = O_{\mathbb{P}}(d_n).$$

If additionally $h_n \to 0$ and $m_f \varepsilon_{f,+,n}, m_f \varepsilon_{f,-,n} \to \infty$, then

$$\widetilde{s}^{\mathrm{raw}}_{f,n} - 1/f_\tau = O_{\mathbb{P}}(e_{f,n}).$$

If $e_{f,n} \to 0$, projection is inactive with probability tending to one and

$$\widetilde{f}_{f,n} - f_\tau = O_{\mathbb{P}}(e_{f,n}).$$

Proof. The pilot assertion is Corollary G.19, applied with sample size m_p and privacy budget $\varepsilon_{p,n}$. For the side releases, the triangular-level rate gives

$$\widetilde{q}_{f,\nu,n} - Q(\tau + \nu h_n) = O_{\mathbb{P}}(\rho_{\nu,n}), \qquad \nu \in \{+,-\}.$$

Subtracting the two releases and dividing by $2h_n$ yields

$$\widetilde{s}^{\mathrm{raw}}_{f,n} - \frac{Q(\tau+h_n) - Q(\tau-h_n)}{2h_n} = O_{\mathbb{P}}\left(\frac{\rho_{+,n} + \rho_{-,n}}{h_n} \right).$$

Adding the deterministic bias $b_Q(h_n)$ proves the sparsity rate. Under $e_{f,n} \to 0$, the raw estimator converges to the positive constant $1/f_\tau$, so projection onto $[c_n, C_n]$ is inactive with probability tending to one. On a fixed neighborhood of $1/f_\tau$, the reciprocal map has bounded derivative, and the mean value theorem gives the stated density rate. □

G.14.3 Local score expansion

Let

$$\omega_f(t) = \sup_{|x-q| \le t} |f(x) - f_\tau|.$$

It is finite and tends to zero for sufficiently small t under Assumption G.1.

Lemma G.34. *Let $\Delta_{p,n}$ be independent of the score block and satisfy $\Delta_{p,n} = O_{\mathbb{P}}(d_n)$ with $d_n \to 0$. Define*

$$\begin{aligned} R_{F,n} &= F(q + \Delta_{p,n}) - F(q) - f_\tau \Delta_{p,n}, \\ R_{N,n} &= (P_N - P)\{\mathbf{1}(X \le q + \Delta_{p,n}) - \mathbf{1}(X \le q)\}, \end{aligned}$$

where P_N is the score-block empirical measure. Then, for every fixed $C < \infty$,

$$|R_{F,n}| \le |\Delta_{p,n}| \omega_f(|\Delta_{p,n}|), \qquad R_{F,n} = O_{\mathbb{P}}\{d_n \omega_f(C d_n)\} \text{ on } \{|\Delta_{p,n}| \le C d_n\}, \tag{51}$$

and

$$R_{N,n} = O_{\mathbb{P}}\left(\sqrt{d_n / N} \right) = o_{\mathbb{P}}(N^{-1/2}). \tag{52}$$

Proof. For either sign of $\Delta_{p,n}$, local absolute continuity gives the oriented integral

$$F(q+\Delta_{p,n}) - F(q) = \int_0^{\Delta_{p,n}} f(q+t)\,dt.$$

Subtracting $f_\tau \Delta_{p,n}$ and taking absolute values proves the first bound in (51); the second follows on the displayed event.

Conditional on $\Delta_{p,n}$, the empirical increment has mean zero. The squared increment indicator equals the indicator that X lies between q and $q+\Delta_{p,n}$. Consequently

$$\mathrm{Var}(R_{N,n} \mid \Delta_{p,n}) \le \frac{1}{N}|F(q+\Delta_{p,n}) - F(q)|.$$

On a sufficiently small deterministic neighborhood, continuity and positivity of f give $|F(q+t) - F(q)| \le C_0|t|$. Because $\Delta_{p,n} = O_{\mathbb{P}}(d_n)$ and $d_n \to 0$, conditional Chebyshev's inequality, first on $|\Delta_{p,n}| \le C d_n$ and then letting C increase, yields $R_{N,n} = O_{\mathbb{P}}\{(d_n/N)^{1/2}\}$. Multiplication by $\sqrt{N}$ gives $O_{\mathbb{P}}(d_n^{1/2}) = o_{\mathbb{P}}(1)$. □

Theorem G.35. *Assume the conditions of Lemma G.33, with $d_n \to 0$, $e_{f,n} \to 0$, $N \to \infty$, and $\sigma_n \to 0$. Put $r_n = N^{-1/2} + \sigma_n$. Suppose that, for every fixed $C < \infty$,*

$$d_n\{e_{f,n} + \omega_f(C d_n)\} = o(r_n). \tag{53}$$

Then

$$\widetilde{q}_n^G - q = \frac{1}{N f_\tau} \sum_{i \in \mathcal{I}_s} \{\tau - \mathbf{1}(X_i \le q)\} - \frac{\sigma_n Z}{f_\tau} + o_{\mathbb{P}}(r_n). \tag{54}$$

The projection onto $[A, B]$ is inactive with probability tending to one.

Proof. Write

$$\Delta_{p,n} = \widetilde{q}_{p,n} - q, \qquad \Delta_{f,n} = \widetilde{f}_{f,n} - f_\tau, \qquad B_N = P_N \psi_q = \frac{1}{N}\sum_{i \in \mathcal{I}_s} \{\mathbf{1}(X_i \le q) - \tau\}.$$

Because $F(q) = \tau$, Lemma G.34 gives the exact decomposition

$$S_N(q+\Delta_{p,n}) = B_N + f_\tau \Delta_{p,n} + R_{F,n} + R_{N,n}.$$

Substitution into the unprojected update gives

$$\begin{aligned} \bar{q}_n^G - q &= \Delta_{p,n} - \frac{B_N + f_\tau \Delta_{p,n} + R_{F,n} + R_{N,n} + \sigma_n Z}{f_\tau + \Delta_{f,n}} \\ &= -\frac{B_N + \sigma_n Z}{f_\tau} + \mathcal{R}_{1n} + \mathcal{R}_{2n} + \mathcal{R}_{3n}, \end{aligned} \tag{55}$$

where

$$\begin{aligned} \mathcal{R}_{1n} &= -\{B_N + \sigma_n Z\}\left\{\frac{1}{f_\tau + \Delta_{f,n}} - \frac{1}{f_\tau}\right\}, \\ \mathcal{R}_{2n} &= \frac{\Delta_{p,n}\Delta_{f,n}}{f_\tau + \Delta_{f,n}}, \\ \mathcal{R}_{3n} &= -\frac{R_{F,n} + R_{N,n}}{f_\tau + \Delta_{f,n}}. \end{aligned}$$

We bound the three terms separately.

First, $B_N = O_{\mathbb{P}}(N^{-1/2})$ by its bounded variance and $\sigma_n Z = O_{\mathbb{P}}(\sigma_n)$, so their sum is $O_{\mathbb{P}}(r_n)$. Since $\Delta_{f,n} = O_{\mathbb{P}}(e_{f,n}) = o_{\mathbb{P}}(1)$ and $f_\tau > 0$, the reciprocal difference is $O_{\mathbb{P}}(e_{f,n})$. Hence $\mathcal{R}_{1n} = O_{\mathbb{P}}(r_n e_{f,n}) = o_{\mathbb{P}}(r_n)$.

Second, $\Delta_{p,n}\Delta_{f,n} = O_{\mathbb{P}}(d_n e_{f,n}) = o_{\mathbb{P}}(r_n)$ by (53); the denominator is bounded away from zero with probability tending to one. Thus $\mathcal{R}_{2n} = o_{\mathbb{P}}(r_n)$.

Third, fix $\eta > 0$. Choose C so that $\mathbb{P}(|\Delta_{p,n}| > Cd_n) < \eta$ for all sufficiently large n. On the complementary event, (51) and (53) give $R_{F,n} = o_{\mathbb{P}}(r_n)$. Since C can be chosen for arbitrary η, the same holds unconditionally. Also $R_{N,n} = o_{\mathbb{P}}(N^{-1/2}) = o_{\mathbb{P}}(r_n)$ by (52). Therefore $\mathcal{R}_{3n} = o_{\mathbb{P}}(r_n)$.

Combining these bounds in (55) proves the expansion for $\bar{q}_n^G$. Its right side is $o_{\mathbb{P}}(1)$ because $r_n \to 0$. Since $q \in (A, B)$, choose $\eta_0 > 0$ with $[q - \eta_0, q + \eta_0] \subset (A, B)$. Then $\mathbb{P}(|\bar{q}_n^G - q| < \eta_0) \to 1$, on which projection is inactive. Hence the same expansion holds for $\widetilde{q}_n^G$. □

Corollary G.36. *Under Theorem G.35,*

$$\widetilde{q}_n^G - q = O_{\mathbb{P}}(N^{-1/2} + \sigma_n). \tag{56}$$

If $0 < \varepsilon_{s,n} < 1$, $0 < \delta_{s,n} < 1$, and the classical sufficient Gaussian calibration

$$\sigma_n = \frac{\sqrt{2\log(1.25/\delta_{s,n})}}{N\varepsilon_{s,n}}$$

is used, then

$$\widetilde{q}_n^G - q = O_{\mathbb{P}}\left\{N^{-1/2} + \frac{\sqrt{\log(1.25/\delta_{s,n})}}{N\varepsilon_{s,n}}\right\}. \tag{57}$$

If additionally $N/n \to \rho \in (0, 1]$ and $\varepsilon_{s,n} \asymp \varepsilon_{0,n}$, direct QExp at center budget $\varepsilon_{0,n}$ has

$$\widetilde{q}_n^Q - q = O_{\mathbb{P}}\{n^{-1/2} + (n\varepsilon_{0,n})^{-1}\}.$$

Consequently the conventional sufficient GOS calibration is not an algebraic-rate improvement over QExp and may contain a $\sqrt{\log(1.25/\delta_{s,n})}$ privacy factor. Under analytic Gaussian calibration the exact first-order comparison is instead determined by σ_n itself. The pure-DP Laplace one-step estimator has $O_{\mathbb{P}}\{N^{-1/2} + (N\varepsilon_{s,n})^{-1}\}$ under its corresponding remainder condition.

Proof. In (54), the bounded empirical-score average is $O_{\mathbb{P}}(N^{-1/2})$, the standard normal variable is $O_{\mathbb{P}}(1)$, and the remainder is $o_{\mathbb{P}}(N^{-1/2} + \sigma_n)$. Since $f_\tau > 0$, their sum proves (56). Substitution of the classical sufficient calibration, with fixed numerical factors absorbed by stochastic-order notation, gives (57). The QExp statement is Corollary G.19, and the Laplace one-step statement follows from (61) by the same argument. No ordering of the complete procedures follows from these rate displays because their privacy definitions, block sizes, nuisance requirements, and leading constants differ. □

Theorem G.37. *Under Theorem G.35, let $v_\tau = \tau(1 - \tau)$ and $d_{G,n} = \{v_\tau/N + \sigma_n^2\}^{1/2}$. Then*

$$\frac{\widetilde{f}_{f,n}(\widetilde{q}_n^G - q)}{d_{G,n}} \rightsquigarrow N(0, 1). \tag{58}$$

Therefore

$$\left[\widetilde{q}_n^G \pm z_{1-\alpha/2}\frac{d_{G,n}}{\widetilde{f}_{f,n}}\right]$$

has asymptotic coverage $1 - \alpha$.

Proof. Define

$$V_N = \frac{1}{\sqrt{Nv_\tau}}\sum_{i\in\mathcal{I}_s}\{\tau - \mathbf{1}(X_i \le q)\}.$$

The bounded i.i.d. central limit theorem gives $V_N \rightsquigarrow N(0, 1)$, and V_N is independent of Z. Since $v_\tau > 0$,

$$d_{G,n} \le \sqrt{v_\tau}\, N^{-1/2} + \sigma_n \le \max(\sqrt{v_\tau}, 1)r_n, \qquad r_n \le \{v_\tau^{-1/2} + 1\}d_{G,n}.$$

Thus $d_{G,n}$ and $r_n = N^{-1/2} + \sigma_n$ are equivalent up to fixed constants, and the $o_{\mathbb{P}}(r_n)$ term in Theorem G.35 is $o_{\mathbb{P}}(d_{G,n})$. Therefore

$$\frac{f_\tau(\widetilde{q}_n^G - q)}{d_{G,n}} = a_n^* V_N - b_n^* Z + o_{\mathbb{P}}(1),$$

where

$$a_n^* = \frac{\sqrt{v_\tau/N}}{d_{G,n}}, \qquad b_n^* = \frac{\sigma_n}{d_{G,n}}, \qquad (a_n^*)^2 + (b_n^*)^2 = 1.$$

Take an arbitrary subsequence. Compactness of $[0,1]^2$ gives a further subsequence on which $(a_n^*, b_n^*) \to (a, b)$ with $a^2 + b^2 = 1$. Joint convergence and independence imply $a_n^* V_N - b_n^* Z \rightsquigarrow aZ_1 - bZ_2$, where Z_1, Z_2 are independent standard normals. This limit is standard normal because its variance is $a^2 + b^2 = 1$. Every subsequence has a further subsequence with the same limiting law, proving full convergence. Finally, $\widetilde{f}_{f,n}/f_\tau \to_{\mathbb{P}} 1$ and Slutsky's theorem prove (58); inversion gives coverage. □

Corollary G.38. *Let* $\lambda_n = \sqrt{N}\sigma_n$.

(a) If $\lambda_n \to \lambda < \infty$*, then*

$$\sqrt{N}(\widetilde{q}_n^G - q) \rightsquigarrow N\left(0, \frac{v_\tau + \lambda^2}{f_\tau^2}\right).$$

(b) If $\lambda_n \to 0$*, the nonprivate quantile variance is recovered.*

(c) If $\lambda_n \to \infty$ *and* $\sigma_n \to 0$*, then*

$$\frac{\widetilde{q}_n^G - q}{\sigma_n} \rightsquigarrow N(0, f_\tau^{-2}).$$

Proof. Multiply (54) by $\sqrt{N}$ for (a) and use the ordinary CLT jointly with Z. Part (b) is the case $\lambda = 0$. For (c), divide by σ_n; the empirical term has standard deviation of order $(\sqrt{N}\sigma_n)^{-1} \to 0$, while the normalized Gaussian term is $-Z/f_\tau$ and the remainder is $o_{\mathbb{P}}(1)$. □

G.14.4 Hölder sufficient conditions and explicit tuning

Lemma G.39. *Suppose that for some* $\gamma \in (0, 1]$*,* $H < \infty$*, and all sufficiently small* $|x - q|$*,*

$$|f(x) - f_\tau| \le H|x - q|^\gamma.$$

Then

$$\omega_f(t) = O(t^\gamma), \qquad Q(\tau + t) - q = \frac{t}{f_\tau} + O(|t|^{1+\gamma}), \qquad b_Q(h) = O(h^\gamma).$$

Proof. The first assertion is immediate. Local positivity gives $f(x) \ge f_\tau/2$ for x near q, so $|Q(\tau + t) - q| \le 2|t|/f_\tau$ for small t. Write $x_t = Q(\tau + t) - q$. Local absolute continuity and the Hölder bound give

$$t = F(q + x_t) - F(q) = f_\tau x_t + R_t, \qquad |R_t| \le \int_0^{|x_t|} H u^\gamma du = \frac{H}{1+\gamma}|x_t|^{1+\gamma}.$$

Since $|x_t| = O(|t|)$, $x_t = t/f_\tau + O(|t|^{1+\gamma})$. Apply this expansion at $t = h$ and $t = -h$, subtract, and divide by $2h$ to obtain $b_Q(h) = O(h^\gamma)$. □

Corollary G.40. *Under Lemma G.39, a sufficient condition for* (53) *is*

$$d_n^{1+\gamma} + d_n e_{f,n} = o(N^{-1/2} + \sigma_n).$$

Assume $N/n \to 1$ *and* $\sqrt{N}\sigma_n = O(1)$.

(a) If the pilot and nuisance privacy parameters are bounded below, take $m_p \asymp m_f \asymp n^a$ *and* $h_n \asymp n^{-a/\{2(1+\gamma)\}}$*, where*

$$a > \max\left\{\frac{1}{1+\gamma}, \frac{1+\gamma}{1+2\gamma}\right\}, \qquad a < 1. \tag{59}$$

(b) *If the pilot and nuisance privacy parameters are of order $n^{-1/2}$, take $m_p \asymp m_f \asymp n^a$ and $h_n \asymp n^{-(a-1/2)/(1+\gamma)}$, where*

$$a > \max\left\{\frac{1}{2} + \frac{1}{2(1+\gamma)}, \frac{1}{2} + \frac{1+\gamma}{2(1+2\gamma)}\right\}, \qquad a < 1. \tag{60}$$

For $\gamma = 1$, the concrete choices are $a = 3/4$, $h_n = n^{-3/16}$ in case (a) and $a = 7/8$, $h_n = n^{-3/16}$ in case (b). For $\gamma = 1/2$, valid choices are $a = 7/8$, $h_n = n^{-7/24}$ in case (a) and $a = 15/16$, $h_n = n^{-7/24}$ in case (b). If the auxiliary budgets are $n^{-3/4}$, the locally Lipschitz choice $a = 15/16$, $h_n = n^{-3/32}$ makes the one-step remainder $o\{n^{-1/4}\}$, the pure-score privacy-dominated scale. Under local Hölder-$1/2$ continuity at the same auxiliary privacy order, the concrete choice $a = 31/32$, $h_n = n^{-7/48}$ has the same property.

Proof. The Hölder bounds give $d_n\omega_f(Cd_n) = O(d_n^{1+\gamma})$ for each fixed C, proving the first statement.

In case (a), d_n and both side-release errors are $O(n^{-a/2})$. If $b = a/\{2(1+\gamma)\}$ and $h_n = n^{-b}$, then

$$h_n^\gamma = n^{-a\gamma/\{2(1+\gamma)\}}, \qquad n^{-a/2}h_n^{-1} = n^{-a\gamma/\{2(1+\gamma)\}}.$$

Thus $e_{f,n} = O\{n^{-a\gamma/(2+2\gamma)}\}$. The two remainders have orders

$$d_n^{1+\gamma} = O\{n^{-a(1+\gamma)/2}\}, \qquad d_n e_{f,n} = O\left\{n^{-a(1+2\gamma)/(2+2\gamma)}\right\}.$$

Both are $o(n^{-1/2})$ exactly under the two strict inequalities in (59).

In case (b), because $a < 1$, the privacy component dominates the auxiliary QExp rate and $d_n, \rho_{+,n}, \rho_{-,n} = O\{n^{-(a-1/2)}\}$. Put $c = a - 1/2$ and $b = c/(1+\gamma)$. Then

$$e_{f,n} = O\{n^{-c\gamma/(1+\gamma)}\}, \quad d_n^{1+\gamma} = O\{n^{-c(1+\gamma)}\}, \quad d_n e_{f,n} = O\{n^{-c(1+2\gamma)/(1+\gamma)}\}.$$

Requiring the last two exponents to exceed $1/2$ gives (60).

Substitution verifies the displayed concrete choices. Finally, with auxiliary privacy $n^{-3/4}$, $a = 15/16$ gives $d_n, \rho_{\nu,n} = O(n^{-3/16})$. Taking $h_n = n^{-3/32}$ yields $e_{f,n} = O(n^{-3/32})$ for $\gamma = 1$, so $d_n^2 = O(n^{-3/8})$ and $d_n e_{f,n} = O(n^{-9/32})$, both $o(n^{-1/4})$.

For $\gamma = 1/2$ at the same auxiliary privacy order, take $a = 31/32$. The auxiliary QExp rate is then

$$d_n, \rho_{\nu,n} = O\{n^{-(a-3/4)}\} = O(n^{-7/32}).$$

With $h_n = n^{-7/48}$,

$$h_n^{1/2} = n^{-7/96}, \qquad \rho_{\nu,n}h_n^{-1} = O(n^{-7/32+7/48}) = O(n^{-7/96}),$$

so $e_{f,n} = O(n^{-7/96})$. Therefore

$$d_n^{3/2} = O(n^{-21/64}), \qquad d_n e_{f,n} = O(n^{-7/32-7/96}) = O(n^{-7/24}),$$

and both exponents exceed $1/4$. Hence both remainder terms are $o(n^{-1/4})$, as claimed. □

Corollary G.41. *Suppose the local Hölder condition in Lemma G.39 holds, and suppose m_p/n, m_f/n, and N/n converge to strictly positive constants. Let all pilot, side, and score pure privacy parameters be bounded above and below by positive constant multiples of a deterministic E_n satisfying $nE_n \to \infty$. Define*

$$\xi_n = n^{-1/2} + (nE_n)^{-1}, \qquad h_n \asymp \xi_n^{1/(1+\gamma)}.$$

Then

$$d_n = O(\xi_n), \qquad e_{f,n} = O\{\xi_n^{\gamma/(1+\gamma)}\},$$

and

$$d_n^{1+\gamma} + d_n e_{f,n} = o(\xi_n).$$

Consequently the Laplace one-step remainder condition holds because $r_n^L \asymp \xi_n$. It also holds for the Gaussian one-step estimator whenever $N^{-1/2} + \sigma_n \gtrsim \xi_n$; in particular this is ensured by the public overcalibration $\sigma_n \geq (N\varepsilon_{s,n})^{-1}$.

Proof. Because every block size is proportional to n and every relevant privacy parameter is proportional to E_n,

$$m_p^{-1/2} + (m_p \varepsilon_{p,n})^{-1} = O(\xi_n)$$

and the same bound holds for both side-release errors. The Hölder quantile-slope lemma and the definition of h_n give

$$b_Q(h_n) = O(h_n^\gamma) = O\{\xi_n^{\gamma/(1+\gamma)}\},$$
$$\frac{\rho_{+,n} + \rho_{-,n}}{h_n} = O\{\xi_n^{1-1/(1+\gamma)}\} = O\{\xi_n^{\gamma/(1+\gamma)}\}.$$

This proves the displayed rates. Since $\xi_n \to 0$,

$$\frac{d_n^{1+\gamma}}{\xi_n} = O(\xi_n^\gamma) \to 0, \qquad \frac{d_n e_{f,n}}{\xi_n} = O\{\xi_n^{\gamma/(1+\gamma)}\} \to 0.$$

For the Laplace score, $N^{-1/2} + (N\varepsilon_{s,n})^{-1} \asymp \xi_n$. For the Gaussian score the same remainder conclusion follows whenever its leading scale is bounded below by a constant multiple of ξ_n. The stated public lower bound on σ_n is sufficient, and adding independent Gaussian noise beyond the analytic minimum is post-processing of the noised query, or equivalently a more dispersed Gaussian mechanism, so it cannot weaken the advertised privacy guarantee. □

G.15 Pure-DP Laplace one-step inference and mechanism comparison

Replace the Gaussian score noise by

$$W_n^L = \frac{L_s}{N\varepsilon_{s,n}}, \qquad L_s \sim \mathrm{Lap}(0,1),$$

and define

$$\bar{q}_n^L = \widetilde{q}_{p,n} - \frac{S_N(\widetilde{q}_{p,n}) + W_n^L}{\widetilde{f}_{f,n}}, \qquad \widetilde{q}_n^L = \Pi_{[A,B]}(\bar{q}_n^L).$$

Theorem G.42. *The full mechanism is pure* $\max\{\varepsilon_{p,n}, \varepsilon_{f,n}, \varepsilon_{s,n}\}$*-DP. Put* $r_n^L = N^{-1/2} + (N\varepsilon_{s,n})^{-1}$*. Under the assumptions of Theorem G.35, with* σ_n *replaced by* $(N\varepsilon_{s,n})^{-1}$ *and the remainder condition measured against* r_n^L*,*

$$\widetilde{q}_n^L - q = \frac{1}{N f_\tau} \sum_{i \in \mathcal{I}_s} \{\tau - \mathbf{1}(X_i \le q)\} - \frac{L_s}{N\varepsilon_{s,n} f_\tau} + o_{\mathbb{P}}(r_n^L). \tag{61}$$

Let

$$c_{L,1-\alpha}(\tau, \kappa) = \inf\{c : \mathbb{P}(|G_\tau + L_s/\kappa| \le c) \ge 1 - \alpha\}.$$

If $N\varepsilon_{s,n} \to \infty$*, then*

$$\left[\widetilde{q}_n^L \pm \frac{c_{L,1-\alpha}(\tau, \varepsilon_{s,n}\sqrt{N})}{\sqrt{N}\,\widetilde{f}_{f,n}}\right]$$

has asymptotic coverage $1-\alpha$*, without requiring convergence of* $\varepsilon_{s,n}\sqrt{N}$*.*

Proof. The sensitivity of the score is $1/N$, and adding a standard Laplace variable multiplied by $(N\varepsilon_{s,n})^{-1}$ is pure $\varepsilon_{s,n}$-DP. The same blockwise case analysis as in Theorem G.31 proves the maximum guarantee. The algebra and remainder bounds in Theorem G.35 apply verbatim after replacing $\sigma_n Z$ by $L_s/(N\varepsilon_{s,n})$, proving (61).

For coverage, take subsequences according to $\kappa_n = \varepsilon_{s,n}\sqrt{N}$. If $\kappa_n \to \infty$, the privacy term is negligible and the classical Gaussian limit holds. If $\kappa_n \to \kappa \in (0, \infty)$, multiplication of (61) by $\sqrt{N} f_\tau$ gives $G_\tau - L_s/\kappa$, which has the same distribution as $G_\tau + L_s/\kappa$. If $\kappa_n \to 0$, multiplication by $N\varepsilon_{s,n} f_\tau$ gives $-L_s$. The critical values satisfy

$$c_{L,1-\alpha}(\tau, \kappa) \to z_{1-\alpha/2}\sqrt{\tau(1-\tau)} \quad (\kappa \to \infty), \qquad \kappa c_{L,1-\alpha}(\tau, \kappa) \to \log(1/\alpha) \quad (\kappa \downarrow 0),$$

by the same coupling proof as Lemma G.26. Consistent studentization and the subsequence principle complete the proof. □

Proposition G.43. *Assume $N/n \to 1$, $\varepsilon_n\sqrt{n} \to \kappa \in (0,\infty)$, and all pilot and nuisance remainders are negligible at root-n scale.*

(a) Full-sample QExp with center budget ε_n satisfies

$$\sqrt{n} f_\tau(\widetilde{q}_n^Q - q) \rightsquigarrow G_\tau + 2L_Q/\kappa, \qquad \mathrm{Var}(2L_Q/\kappa) = 8/\kappa^2.$$

(b) The pure-DP Laplace one-step estimator with score budget ε_n satisfies

$$\sqrt{n} f_\tau(\widetilde{q}_n^L - q) \rightsquigarrow G_\tau + L_s/\kappa, \qquad \mathrm{Var}(L_s/\kappa) = 2/\kappa^2.$$

(c) If $\sqrt{N}\sigma_n \to \lambda$, GOS satisfies

$$\sqrt{n} f_\tau(\widetilde{q}_n^G - q) \rightsquigarrow N\{0, \tau(1-\tau) + \lambda^2\}.$$

The variance ratio in (a)–(b) is four, but it is not a minimax ordering of the complete procedures.

Proof. Parts (a), (b), and (c) are direct applications of Theorem G.18, Theorem G.42, and Corollary G.38, respectively, together with $N/n \to 1$. A standard Laplace variable has variance two, so the privacy variances are $4 \times 2/\kappa^2 = 8/\kappa^2$ and $2/\kappa^2$. The final statement follows because the two one-step methods require extra blocks and a local linearization condition that the direct QExp theorem does not require; GOS also uses approximate rather than pure privacy. □

Proposition G.44. *Under the assumptions of the corresponding results:*

(a) if $\varepsilon_n\sqrt{n} \to \infty$, then $\sqrt{n}(\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n}) \to_{\mathbb{P}} 0$;

(b) setting $m = 1$ and $\alpha_1 = 1$ in Theorem G.46 yields the critical-regime law in Theorem G.18;

(c) under Theorem G.29 and Lemma G.30, a nuisance block of size $o(n)$ leaves the center's first-order distribution and interval length unchanged relative to a full-sample QExp center with the same budget; and

(d) under Corollary G.38, $\sqrt{N}\sigma_n \to 0$ recovers the classical quantile limit.

Proof. For (a), Theorem G.13 applies when the privacy budget is bounded below. Otherwise, along any subsequence on which $\varepsilon_n \to 0$, Theorem G.8 gives

$$\sqrt{n} f_\tau(\widetilde{q}_{\tau,n} - \widehat{q}_{\tau,n}) = \frac{2}{\varepsilon_n\sqrt{n}} O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

The subsequence principle proves the full assertion. Part (b) follows by direct substitution in Theorem G.46. For (c), consistency of the nuisance estimator follows from Theorem G.29; $n_0/n \to 1$ makes both the sampling normalization and the center phase parameter asymptotically equal to their full-sample counterparts, and Slutsky's theorem gives the interval-length statement. Part (d) is part (b) of Corollary G.38. □

G.16 A FINITE VECTOR OF QUANTILES

Let $0 < \tau_1 < \cdots < \tau_m < 1$ be fixed. At each $q_j = Q(\tau_j)$ assume the local conditions in Assumption G.1, with $f_j = f(q_j) > 0$. Let $\alpha_j > 0$ and $\sum_j \alpha_j \le 1$. Conditional on the sample, run the m mechanisms independently with budgets $\alpha_j\varepsilon_n$.

Lemma G.45. *Write $\widehat{q}_{j,n} = X_{(\lfloor n\tau_j \rfloor)}$ for the sample quantile at level τ_j. Let*

$$G_{j,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\tau_j - \mathbf{1}(X_i \le q_j)\}.$$

Then

$$\big(\sqrt{n} f_j(\widehat{q}_{j,n} - q_j)\big)_{j=1}^m = (G_{j,n})_{j=1}^m + o_{\mathbb{P}}(1),$$

and the vector converges to a centered Gaussian vector G with

$$\mathrm{Cov}(G_j, G_\ell) = \min(\tau_j, \tau_\ell) - \tau_j\tau_\ell.$$

Proof. Apply the scalar Bahadur argument in Theorem G.16 at each of the finitely many quantiles. The vector of Bernoulli scores satisfies the multivariate central limit theorem. For $j \le \ell$,

$$\begin{aligned}\mathrm{Cov}\{\mathbf{1}(X \le q_j), \mathbf{1}(X \le q_\ell)\} &= \mathbb{P}(X \le q_j) - \tau_j\tau_\ell \\ &= \tau_j - \tau_j\tau_\ell,\end{aligned}$$

which is the stated formula. □

Theorem G.46. *If $\varepsilon_n \to 0$, $n\varepsilon_n \to \infty$, and $\varepsilon_n\sqrt{n} \to \kappa \in (0,\infty)$, then*

$$\left(\sqrt{n} f_j\{\widetilde{q}_{j,n} - q_j\}\right)_{j=1}^m \rightsquigarrow \left(G_j + \frac{2L_j}{\alpha_j\kappa}\right)_{j=1}^m,$$

where $L_1, \dots, L_m$ are independent standard Laplace variables and are independent of G. If $\varepsilon_n\sqrt{n} \to \infty$, only the Gaussian vector remains.

Proof. For each j, the conditional law of

$$\frac{n\alpha_j\varepsilon_n f_j}{2}(\widetilde{q}_{j,n} - \widehat{q}_{j,n})$$

converges in probability to the standard Laplace law. Conditional independence of the mechanism randomizations implies that the joint conditional characteristic function is the product of the marginal conditional characteristic functions. Hence the conditional vector converges to the product Laplace law. Apply the vector version of Lemma G.9 together with Lemma G.45. The scaling identity

$$\sqrt{n} f_j(\widetilde{q}_{j,n} - \widehat{q}_{j,n}) = \frac{2}{\alpha_j\varepsilon_n\sqrt{n}}\left\{\frac{n\alpha_j\varepsilon_n f_j}{2}(\widetilde{q}_{j,n} - \widehat{q}_{j,n})\right\}$$

gives the critical limit. The high-privacy conclusions follow exactly as in Theorem G.18. □

G.17 A CLASS-MATCHED PURE-DP MINIMAX LOWER BOUND

This section proves the two lower scales using an explicit smooth location subfamily and a private version of Le Cam's two-point method. For $\ell, r > 0$, let $\mathcal{F}(\tau, \ell, r; A, B)$ be the class of continuous distributions supported on $[A, B]$ with a unique τ-quantile $q_F \in [A + 2r, B - 2r]$, absolutely continuous on $[q_F - r, q_F + r]$ with density at least ℓ. The supremum below is over this class. Write $\mathsf{DP}(\varepsilon)$ for all pure ε-DP estimators under replacement adjacency.

The lower scales agree with existing private-median and private-quantile results (Tzamos et al., 2020; Ramsay & Diaz-Rodriguez, 2025). The argument below uses the stated local class, replacement adjacency, and absolute-error loss.

G.17.1 A REGULAR TRANSLATED SUBFAMILY

Fix $\tau \in (0,1)$. Let

$$g_0(u) = C_0 \exp\{-1/(1-u^2)\}\mathbf{1}\{|u| < 1\},$$

where C_0 normalizes the density. Let z_τ be its unique τ-quantile. Choose $d_0 > 0$ such that $[z_\tau - d_0, z_\tau + d_0] \subset (-1, 1)$, and define

$$b_0 = \min_{|u - z_\tau| \le d_0} g_0(u) > 0.$$

The function g_0 is smooth on $\mathbb{R}$, and $(\sqrt{g_0})' \in L^2(\mathbb{R})$ and $g_0' \in L^1(\mathbb{R})$.

Given $\ell > 0$, set $s = b_0/(2\ell)$ and define the translated location family

$$f_\theta(x) = \frac{1}{s} g_0\left(\frac{x - \theta}{s}\right), \qquad q_\theta = \theta + sz_\tau.$$

Let P_θ denote the law with density f_θ. If $\ell r \le b_0 d_0/2$, then for $|x - q_\theta| \le r$,

$$\left|\frac{x - \theta}{s} - z_\tau\right| \le r/s \le d_0,$$

and therefore

$$f_\theta(x) \geq b_0/s = 2\ell \geq \ell.$$

Let $c = (A+B)/2$ and choose two quantiles $q_0 = c - \Delta/2$ and $q_1 = c + \Delta/2$, with corresponding location parameters $\theta_i = q_i - sz_\tau$. Write $P_i = P_{\theta_i}$ for their laws. The union of their supports has width at most

$$2s(1+|z_\tau|) + \Delta = \frac{b_0(1+|z_\tau|)}{\ell} + \Delta.$$

If $\Delta \leq 2r$ and

$$B - A \geq \frac{b_0(1+|z_\tau|)}{\ell} + 6r,$$

then both supports lie in $[A, B]$ and both quantiles lie in $[A+2r, B-2r]$. Combined with $\ell r \leq b_0 d_0/2$, this embeds the two translated alternatives in the class used in the main theorem.

Lemma G.47. *There are finite constants C_H, C_T, depending only on g_0, such that for every Δ,*

$$H(P_\theta, P_{\theta+\Delta}) \leq C_H \frac{|\Delta|}{s}, \tag{62}$$

$$\mathrm{TV}(P_\theta, P_{\theta+\Delta}) \leq C_T \frac{|\Delta|}{s}. \tag{63}$$

Here $H(P,Q) = \{\int(\sqrt{p} - \sqrt{q})^2\}^{1/2}$.

Proof. Let $f_s(x) = s^{-1}g_0(x/s)$ and $r_s(x) = \sqrt{f_s(x)}$. Translation by Δ and the fundamental theorem of calculus give

$$\|r_s(\cdot - \Delta) - r_s\|_2 \leq |\Delta| \|r_s'\|_2.$$

A change of variables shows

$$\|r_s'\|_2 = s^{-1}\|(\sqrt{g_0})'\|_2.$$

This proves (62). Similarly,

$$\|f_s(\cdot - \Delta) - f_s\|_1 \leq |\Delta| \|f_s'\|_1 = \frac{|\Delta|}{s}\|g_0'\|_1.$$

Since total variation is one half of the L^1 distance, (63) follows. □

G.17.2 A COUPLING CONTRACTION LEMMA FOR PURE DIFFERENTIAL PRIVACY

Lemma G.48. *Let M be an ε-DP mechanism under replacement adjacency. Suppose random datasets (D, D') form a coupling of input laws P and Q, and let $d_H(D, D')$ be their Hamming distance. If*

$$\mathbb{P}\{d_H(D, D') > m\} \leq \beta,$$

then the output laws P_M and Q_M satisfy, for every event S,

$$\begin{aligned} P_M(S) &\leq e^{m\varepsilon} Q_M(S) + \beta, \\ Q_M(S) &\leq e^{m\varepsilon} P_M(S) + \beta. \end{aligned} \tag{64}$$

Consequently,

$$\mathrm{TV}(P_M, Q_M) \leq \frac{e^{m\varepsilon} - 1 + 2\beta}{e^{m\varepsilon} + 1}. \tag{65}$$

Proof. Group privacy gives $M(D, S) \leq e^{m\varepsilon} M(D', S)$ whenever $d_H(D, D') \leq m$. Therefore

$$\begin{aligned} P_M(S) &= \mathbb{E}M(D, S) \\ &\leq \mathbb{E}\{M(D, S)\mathbf{1}(d_H \leq m)\} + \mathbb{P}(d_H > m) \\ &\leq e^{m\varepsilon}\mathbb{E}M(D', S) + \beta = e^{m\varepsilon} Q_M(S) + \beta. \end{aligned}$$

Interchanging D, D' proves (64).

Put $a = e^{m\varepsilon}$. For an event with $p = P_M(S) \geq q = Q_M(S)$, the first inequality gives $d := p - q \leq (a-1)q + \beta$. Applying the second inequality to S^c gives

$$1 - q \leq a(1-p) + \beta,$$

or $(a-1)q + ad \le a - 1 + \beta$. Maximizing d subject to the two linear inequalities yields

$$d \le \frac{a - 1 + 2\beta}{a + 1}.$$

Taking the supremum over events proves (65). □

Lemma G.49. *Let P and Q be one-observation laws with $\delta = \mathrm{TV}(P, Q)$. There is a coordinatewise coupling of P^n and Q^n for which the Hamming distance $H_n \sim \mathrm{Bin}(n, \delta)$. If $\mu = n\delta$ and $m = \lceil 2\mu + 6 \rceil$, then*

$$\mathbb{P}(H_n > m) \le 2^{-6}.$$

Proof. A maximal coupling of one coordinate disagrees with probability exactly δ. Taking independent maximal couplings over coordinates makes the number of disagreements binomial. For $t = \log 2$, Markov's inequality gives

$$\begin{aligned}
\mathbb{P}(H_n \ge 2\mu + 6) &\le e^{-t(2\mu+6)} \mathbb{E} e^{tH_n} \\
&= e^{-t(2\mu+6)} (1 - \delta + \delta e^t)^n \\
&\le \exp\{-t(2\mu + 6) + \mu(e^t - 1)\} \\
&= \exp\{-(2\log 2 - 1)\mu - 6\log 2\} \le 2^{-6}.
\end{aligned}$$

Since $m \ge 2\mu + 6$, the same bound holds for $\mathbb{P}(H_n > m)$. □

G.17.3 Le Cam conversion and the two scales

Lemma G.50. *Let output laws R_0, R_1 correspond to parameters q_0, q_1 with $|q_1 - q_0| = \Delta$. Every estimator T based on the output satisfies*

$$\max_{i=0,1} \mathbb{E}_i |T - q_i| \ge \frac{\Delta}{4} \{1 - \mathrm{TV}(R_0, R_1)\}.$$

Proof. Let $S = \{|T - q_0| < \Delta/2\}$. On S^c, the error under parameter zero is at least $\Delta/2$; on S, the triangle inequality gives $|T - q_1| \ge \Delta/2$. Hence

$$\begin{aligned}
\mathbb{E}_0 |T - q_0| + \mathbb{E}_1 |T - q_1| &\ge \frac{\Delta}{2} \{R_0(S^c) + R_1(S)\} \\
&\ge \frac{\Delta}{2} \{1 - \mathrm{TV}(R_0, R_1)\}.
\end{aligned}$$

The maximum is at least half the sum. □

Theorem G.51. *Let b_0, d_0, z_τ be the fixed constants in the translated family above. There are constants $c_2, c_3 > 0$, depending only on g_0 and τ, such that, if*

$$\ell r \le b_0 d_0 / 2, \qquad B - A \ge \frac{b_0(1 + |z_\tau|)}{\ell} + 6r, \tag{66}$$

and

$$\frac{1}{\ell\sqrt{n}} \vee \frac{1}{\ell n \varepsilon} \le c_2 r, \tag{67}$$

then, for all sufficiently large n and $0 < \varepsilon \le 1$,

$$\inf_{M \in \mathsf{DP}(\varepsilon)} \sup_F \mathbb{E}_F |M(D_n) - q_F| \ge c_3 \left\{ \frac{1}{\ell\sqrt{n}} \vee \frac{1}{\ell n \varepsilon} \right\}. \tag{68}$$

Proof. We prove the two terms separately.

Step 1: the ordinary sampling term. Place two translated alternatives symmetrically about $(A+B)/2$ as described above, with quantile separation

$$\Delta_s = \frac{c_s}{\ell\sqrt{n}}.$$

Because $s = b_0/(2\ell)$, Lemma G.47 gives

$$H(P_0, P_1) \le C\ell\Delta_s = Cc_s/\sqrt{n}.$$

For product measures,

$$H(P_0^n, P_1^n) \le \sqrt{n}\, H(P_0, P_1) \le Cc_s.$$

Also $\mathrm{TV} \le H$. Choose $c_s > 0$ sufficiently small that $\mathrm{TV}(P_0^n, P_1^n) \le 1/2$. Data processing cannot increase total variation, so the output laws under any mechanism also have total variation at most $1/2$. Their quantiles are separated by exactly Δ_s. By Lemma G.50, the maximal risk is at least $\Delta_s/8$, proving a constant multiple of $1/(\ell\sqrt{n})$.

Step 2: the privacy term. For the privacy term, place another symmetric pair in the same way and take

$$\Delta_p = \frac{c_p}{\ell n\varepsilon}.$$

By (63), the one-observation total variation satisfies

$$\delta \le C\ell\Delta_p = \frac{Cc_p}{n\varepsilon}.$$

Under the binomial maximal coupling in Lemma G.49, the mean Hamming distance is $\mu = n\delta \le Cc_p/\varepsilon$. Set $m = \lceil 2\mu + 6 \rceil$ and $\beta = 2^{-6}$. Then

$$m\varepsilon \le (2\mu + 7)\varepsilon \le 2Cc_p + 7,$$

because $\varepsilon \le 1$. By Lemma G.48, the output total variation is at most

$$\rho_p := \frac{\exp(2Cc_p + 7) - 1 + 2^{-5}}{\exp(2Cc_p + 7) + 1} < 1.$$

This is a numerical constant strictly below one, uniformly in n and ε. Applying Lemma G.50 gives

$$\max_{i=0,1} \mathbb{E}_i|T - q_i| \ge \frac{1-\rho_p}{4}\Delta_p,$$

which is a constant multiple of $1/(\ell n\varepsilon)$.

Step 3: combination. The minimax risk is at least each of the two lower bounds, hence at least the smaller of their constants times the maximum of the two scales. Choose the numerical constants in the two separations small enough that (67) implies $\Delta_s, \Delta_p \le 2r$. Then (66) ensures, by the explicit support calculation above, that both pairs of translated alternatives remain in the asserted class. This proves (68). □

Remark G.52. The lower bound is local. If $1/(\ell\sqrt{n})$ or $1/(\ell n\varepsilon)$ exceeds the radius on which the density is controlled, a two-point shift of that size need not remain in the class. Likewise, a bounded output range caps every absolute-error risk. The theorem therefore states the sharp regular scale only in the nontrivial local regime.

G.18 Proof of the tail-bound inversion

Proof. Let t_1 and t_2 denote the first and second quantities on the right of (19). If $t \ge t_1$, then

$$C_1 n \exp(-c_1 n\varepsilon\ell t) \le C_1 n \exp(-c_1 n\varepsilon\ell t_1) = \alpha/2.$$

If $t \ge t_2$, then

$$C_2 \exp(-c_2 n\ell^2 t^2) \le C_2 \exp(-c_2 n\ell^2 t_2^2) = \alpha/2.$$

Taking $t = t_1 \vee t_2$ in (18) proves the claim. □