EMPIRICAL LIKELIHOOD FOR THE DIFFERENTIALLY PRIVATE MEAN

Anonymous authors
Paper under double-blind review

ABSTRACT

We develop a nonparametric framework for confidence intervals for a population mean under differential privacy (DP), without requiring variance estimation or a parametric model for the underlying distribution. We partition clipped observations into disjoint blocks, privatize each block mean with Laplace noise, and construct confidence intervals using empirical likelihood (EL) and adjusted empirical likelihood (AEL). We establish Wilks-type results showing that the EL and AEL likelihood-ratio statistics converge to the standard χ^2_1 distribution under DP. The resulting intervals are calibrated for the clipped population mean μ_c, the estimand targeted by fixed-clipping bounded-sensitivity mechanisms. We explicitly characterize the clipping bias and provide a bias-aware extension for inference on the raw mean. Simulations across various distributions, together with real-data experiments, show that the proposed methods maintain coverage close to the nominal level while typically producing shorter confidence intervals than existing approaches.

1 INTRODUCTION

Confidence intervals under differential privacy (DP) must account for sampling variability and the noise introduced to protect individual records. Existing methods for mean estimation often impose a parametric model or approximate the sampling distribution of a privatized estimator. These requirements can be restrictive when the population distribution is unknown or strongly skewed. Figure 2 shows differences in the empirical age distributions across four states, motivating inference that does not require a prespecified parametric model.

We combine sample partitioning with empirical likelihood (EL) (Owen, 1988). After clipping the observations to a public range, we divide them into disjoint blocks and add Laplace noise to each block mean. Parallel composition makes the collection of privatized block means ϵ-DP. We then construct confidence intervals by profiling an EL or adjusted empirical likelihood (AEL) ratio over these summaries. The method requires neither a parametric population model nor a separate variance estimate. It uses first-order asymptotic chi-squared calibration and requires clipping bounds. Owen (2001); Lazar (2021); Liu & Zhao (2023) review the broader EL literature.

Fixed clipping changes the inferential target. The privatized summaries identify the clipped mean $\mu_c = \mathbb{E}[\mathrm{clip}(X)]$, rather than the raw mean $\mu = \mathbb{E}X$, unless the support lies within the clipping interval or the two tail contributions cancel. We therefore state the formal results for μ_c, quantify the difference $\mu_c - \mu$, and provide a conservative interval for μ under an explicit tail-moment bound.

Our contributions are as follows.

- We construct pure-ϵ-DP EL and AEL confidence intervals from Laplace-privatized block means without specifying a parametric model for the population distribution.
- We prove Wilks-type limits for the resulting triangular array: although the variance of each block summary vanishes, both likelihood-ratio statistics converge to χ^2_1 at μ_c. We also characterize clipping bias and give a bias-aware extension for μ (Proposition 2.1 and Remark 2.4).
- We evaluate finite-sample coverage and interval length across light-tailed, heavy-tailed, and real-data settings, compare against five baselines, and give practical guidance for choosing the clipping bounds and the number of blocks B.

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.

Related literature. Existing work on differentially private confidence intervals for the mean often relies on distributional assumptions or an explicit approximation to the sampling distribution. Karwa & Vadhan (2018); Du et al. (2020) derived confidence intervals under normality, and Ferrando et al. (2022) proposed a parametric bootstrap that depends on correct model specification. Awan & Wang (2025) constructed intervals by repeatedly simulating privatized statistics under a specified generative model. Brawner & Honaker (2018) proposed a bootstrap method that avoids a parametric model for the original data but uses a normal approximation to the sampling distribution. Our approach instead avoids a parametric population model and calibrates the EL and AEL ratios through an asymptotic χ^2_1 limit (Theorems 2.3 and 2.5).

More broadly, several methods construct private confidence sets by estimating or approximating sampling distributions. Covington et al. (2025) combine the bag of little bootstraps (Kleiner et al., 2014) with CoinPress (Biswas et al., 2020). Chadha et al. (2021; 2024) develop privatized bootstrap ensembles. The latter also partitions the data but aggregates with a private median rather than an empirical-likelihood profile. These solve a more general estimation problem with their own hyperparameters and privacy mechanisms, each targeting the scalar mean only as a special case; an apples-to-apples comparison under matched clipping and privacy accounting is left to future work. Other work studies DP inference for differences of means (D'Orazio et al., 2015), medians (Drechsler et al., 2022), multivariate means and covariances (Biswas et al., 2020), empirical risk minimization (Wang et al., 2019), and OLS (Sheffet, 2017). Related work develops DP resampling theory and valid inference from privacy-protected data (Balle et al., 2018; Evans & King, 2023; Evans et al., 2023). Further results cover minimax-optimal private point estimation (Cai et al., 2021), high-dimensional robust-mean intervals under Gaussian DP (Yu et al., 2024), and self-normalized quantile inference under local DP (Liu et al., 2023). Our focus is a variance-estimation-free scalar-mean confidence interval under central, pure ϵ-DP. Liu & Li (2025) also study empirical likelihood under differential privacy (PriDEL/BAPriDEL), but in a distinct setting: K physically separate machines each privatize a local estimate under (ϵ, δ)-DP. We include adapted versions of PriDEL and BAPriDEL as empirical baselines in Table 1. In our experiments, EL and AEL generally yield shorter intervals than the adapted PriDEL and BAPriDEL baselines while providing comparable coverage.

2 Algorithms and Theoretical Results

Throughout, we work with ϵ-DP (Dwork et al., 2006): informally, a randomized algorithm's output distribution changes only mildly when one record of the input dataset is modified, with the privacy budget ϵ controlling the trade-off between privacy and accuracy. Achieving ϵ-DP requires calibrating noise to a function's sensitivity (Definition B.2), typically via the Laplace mechanism (Theorem B.3); Section B reviews these formal definitions. Let $x_1, \ldots, x_n$ be i.i.d. observations from a distribution with mean $\mu = \mathbb{E}X$. We seek a confidence interval based only on an ϵ-DP release.

2.1 EL Confidence Interval for the Private Mean

Algorithm 1 in Section C takes the data, public clipping bounds $x_{\min} < x_{\max}$, the number of blocks B, privacy budget ϵ, and significance level α. We assume for simplicity that the block size $m = n/B$ is an integer. The clipping bounds may come from public domain knowledge; for example, ages may be restricted to a prespecified plausible range. Section F discusses alternatives when only a coarse range is known or the data are unbounded.

After clipping, the algorithm divides the observations into B disjoint blocks of size m. The mean of a block has sensitivity $(x_{\max} - x_{\min})/m$, so adding independent Laplace noise with scale $(x_{\max} - x_{\min})/(m\epsilon)$ makes each block release ϵ-DP. Because the blocks are disjoint, releasing all B noisy means $\tilde{x}_1, \ldots, \tilde{x}_B$ remains ϵ-DP by parallel composition. The subsequent EL calculation is post-processing and consumes no additional privacy budget. Only the privatized block means are used for inference. The choice of B trades asymptotic calibration against privacy noise; Section F gives the data-independent default $B \approx n/\log n$ and Section 3.2 studies sensitivity to this choice.

2.1.1 Theoretical Analysis

The privatized means $\tilde{x}_1, \ldots, \tilde{x}_B$ are computed from disjoint blocks of the *clipped* data. Write $\mathrm{clip}(x) = x_{\min}\mathbb{I}(x < x_{\min}) + x\,\mathbb{I}(x_{\min} \le x \le x_{\max}) + x_{\max}\mathbb{I}(x > x_{\max})$ for the clipping map

applied in the first step of Algorithm 1, and let $\mu_c := \mathbb{E}[\mathrm{clip}(X; x_{\min}, x_{\max})]$ denote the mean of the clipped population, as opposed to $\mu = \mathbb{E}X$. Since $\tilde{x}_j$ is computed from already-clipped observations, the mean of $\tilde{x}_j$ for $j = 1, \ldots, B$ is

$$\mathbb{E}\tilde{x}_j = \mathbb{E}[\mathrm{clip}(x_i)] + \mathbb{E}z_j = \mu_c. \tag{1}$$

Thus the Laplace noise does not change the expectation of a privatized block mean. The pivotal target below is μ_c, rather than the raw mean μ. The two coincide when the population is supported on $[x_{\min}, x_{\max}]$ and may also coincide when the lower- and upper-tail contributions cancel. We first quantify their difference and then give an explicit condition under which an interval for μ_c can be enlarged to cover μ conservatively.

Proposition 2.1 (Clipping bias). *Let $\mathbb{E}|X| < \infty$ and $x_{\min} < x_{\max}$, with $\mu = \mathbb{E}X$ and $\mu_c = \mathbb{E}[\mathrm{clip}(X; x_{\min}, x_{\max})]$. Then:*

(a) $\mu_c - \mu = \mathbb{E}[\max(x_{\min} - X, 0)] - \mathbb{E}[\max(X - x_{\max}, 0)]$, and hence $|\mu_c - \mu| \le \mathbb{E}[\max(x_{\min} - X, 0)] + \mathbb{E}[\max(X - x_{\max}, 0)] = \mathbb{E}|\mathrm{clip}(X) - X|$; this bound is 0 if and only if $\mathbb{P}(x_{\min} \le X \le x_{\max}) = 1$;

(b) if $X \ge x_{\min}$ almost surely, the bias is a pure downward shift, $\mu_c - \mu = -\mathbb{E}[\max(X - x_{\max}, 0)] \le 0$;

(c) fix any reference point $a \in (x_{\min}, x_{\max})$. If $\mathbb{E}|X - a|^p < \infty$ for some $p > 1$, then

$$\mathbb{E}[\max(X - x_{\max}, 0)] \le \frac{\mathbb{E}|X - a|^p}{p\,(x_{\max} - a)^{p-1}}, \qquad \mathbb{E}[\max(x_{\min} - X, 0)] \le \frac{\mathbb{E}|X - a|^p}{p\,(a - x_{\min})^{p-1}}.$$

Part (a) expresses the signed clipping bias as the difference between the lower- and upper-tail contributions. The displayed upper bound is zero exactly when the population is supported on the clipping interval, although the signed bias itself may also vanish through cancellation. Part (b) gives the common one-sided case: when the lower clipping bound is also a support bound, only upper-tail clipping contributes to the bias. Part (c) controls both tail contributions through a moment centered at any reference point inside the clipping interval, so it applies even when the entire interval lies above or below zero. Widening the interval generally reduces these tail contributions but increases the sensitivity $x_{\max} - x_{\min}$ and hence the scale of the Laplace noise. The clipping bounds therefore govern a bias–variance trade-off shared by the bounded-sensitivity methods considered in Section 3.

Proposition 2.2 (Non-identifiability under fixed clipping). *Without an additional support or tail assumption, the raw mean μ is not identifiable from a release that depends on the sample only through observations clipped to a fixed interval.*

We first derive the limiting distribution by empirical likelihood. For a candidate value θ, define

$$\mathcal{R}(\theta) = \max_{w_1, \ldots, w_B} \Pi_{j=1}^{B} B w_j, s.t. \sum_{j=1}^{B} w_j \tilde{x}_j = \theta, \sum_{j=1}^{B} w_j = 1, w_j \ge 0. \tag{2}$$

Here w_j is the probability mass assigned to the jth privatized block mean. The constraints require nonnegative masses that sum to one and whose weighted mean equals θ. The factor B normalizes the constrained likelihood by its unconstrained maximum, attained at $w_j = 1/B$. If the constraint is infeasible, we define $\mathcal{R}(\theta) = 0$ and $-2 \log \mathcal{R}(\theta) = +\infty$.

Suppose θ lies in the interior of the convex hull of the privatized block means. Then the maximizing weights are strictly positive, and the following Lagrange-multiplier representation applies. At a feasible boundary point, $\mathcal{R}(\theta) = 0$ unless every observation equals θ, the exceptional case treated below. Define the Lagrangian function

$$\mathcal{L}(w_1, \ldots, w_B, \lambda, v) = \sum_{j=1}^{B} \log(B w_j) - B\lambda \sum_{j=1}^{B} w_j (\tilde{x}_j - \theta) + v \left(\sum_{j=1}^{B} w_j - 1 \right),$$

where λ and v are Lagrange multipliers. Differentiating $\mathcal{L}$ with respect to w_j and setting the derivative to zero yields $\frac{\partial \mathcal{L}}{\partial w_j} = \frac{1}{w_j} - B\lambda(\tilde{x}_j - \theta) + v = 0$, i.e., $1 - B\lambda w_j (\tilde{x}_j - \theta) + v w_j = 0$. Then, we obtain $v = -B$. Substituting back into the first-order condition gives, for each $j = 1, \ldots, B$,

$$w_j = \frac{1}{B} \frac{1}{1 + \lambda(\tilde{x}_j - \theta)}. \tag{3}$$

Using Equation (2) together with the above expression for w_j, the empirical log-likelihood ratio becomes

$$-2\log\mathcal{R}(\theta) = 2\sum_{j=1}^{B}\log\big[1+\lambda(\tilde{x}_j-\theta)\big], \tag{4}$$

Let $g_j(\theta)=\tilde{x}_j-\theta$ and define the admissible domain $\mathcal{D}_\theta=\{\lambda\in\mathbb{R}: 1+\lambda g_j(\theta)>0,\ j=1,\dots,B\}$. On the interior-feasibility event, λ is the unique solution in $\mathcal{D}_\theta$ of

$$\frac{1}{B}\sum_{j=1}^{B}\frac{g_j(\theta)}{1+\lambda g_j(\theta)}=0. \tag{5}$$

Indeed, the left-hand side is strictly decreasing on $\mathcal{D}_\theta$ unless every residual is zero, because its derivative is $-\frac{1}{B}\sum_{j=1}^{B}\frac{g_j(\theta)^2}{\{1+\lambda g_j(\theta)\}^2}<0$. Interior feasibility supplies both a positive and a negative residual. Consequently, the score tends to $+\infty$ at the left endpoint of $\mathcal{D}_\theta$ and to $-\infty$ at its right endpoint, which proves existence as well as uniqueness of the root. If every residual is zero, we set $\lambda=0$ and $\mathcal{R}(\theta)=1$.

We establish the limiting distribution of $-2\log\mathcal{R}(\mu_c)$ in the following Wilks-type theorem for the clipped mean μ_c.

Theorem 2.3. *Let $x_1,\dots,x_n$ be i.i.d. observations drawn from the distribution of X. Let $\mathrm{clip}(X)$ denote the clipped variable with $\mu_c=\mathbb{E}[\mathrm{clip}(X)]$, and suppose $\mathrm{Var}(\mathrm{clip}(X))<\infty$. Fix the clipping bounds $x_{\min}<x_{\max}$ and the privacy budget $\epsilon>0$. Suppose that the privacy-protected samples $\tilde{x}_1,\dots,\tilde{x}_B$ are generated using Algorithm 1. We assume $B=o(n)$. Then, as $B\to\infty$ and $n\to\infty$, $-2\log\mathcal{R}(\mu_c)\xrightarrow{D}\chi^2_1$, where $\xrightarrow{D}$ represents convergence in distribution and χ^2_1 denotes the chi-squared distribution with 1 degree of freedom.*

The theorem requires $B\to\infty$ and $B=o(n)$, equivalently, $B\to\infty$ and $m=n/B\to\infty$. As both dimensions grow, the variance of each privatized block mean vanishes, so the observations form a triangular array rather than an i.i.d. sample from a fixed distribution. To see why the Wilks limit nevertheless survives privacy, when $\sigma_c^2:=\mathrm{Var}(\mathrm{clip}(X))>0$ the ratio of privacy-noise variance to block-sampling variance is $\frac{2(x_{\max}-x_{\min})^2}{m\epsilon^2\sigma_c^2}$, which vanishes for fixed ϵ as $m\to\infty$. When $\sigma_c^2=0$, the Laplace perturbation is instead the sole source of variation, and the same self-normalized EL limit continues to hold. The proof in Section L covers both cases. Because the χ^2_1 limit contains no nuisance parameter, the resulting interval requires no separate variance estimate: $I_{EL}=\{\theta:-2\log\mathcal{R}(\theta)\le\chi^2_1(\alpha)\}$, where $\chi^2_1(\alpha)$ is the upper α-quantile of χ^2_1.

Remark 2.4 (From μ_c to a bias-aware interval for μ)**.** By Proposition 2.2, μ is not identifiable from I_{EL} without an additional assumption, so any valid interval for μ itself must carry an explicit bias allowance. Fix a public reference point $a\in(x_{\min},x_{\max})$ and suppose a bound $M_{p,a}\ge\mathbb{E}|X-a|^p$ is known for some $p>1$. Such a bound must come from public domain knowledge or an independent source; estimating it from the confidential sample would require an additional privacy mechanism and corresponding privacy accounting. By Proposition 2.1(c),

$$|\mu-\mu_c|\le\frac{M_{p,a}}{p\,(x_{\max}-a)^{p-1}}+\frac{M_{p,a}}{p\,(a-x_{\min})^{p-1}}=:\Delta_{p,a}.$$

On the event that I_{EL} is nonempty, let $L=\inf I_{EL}$ and $U=\sup I_{EL}$. The inflated interval $[L-\Delta_{p,a},\,U+\Delta_{p,a}]$ is asymptotically conservative for μ: whenever $\mu_c\in I_{EL}$, it also contains μ, so its asymptotic coverage is at least $1-\alpha$ under the stated moment bound. This converts the clipping bias from an uncontrolled quantity into an explicit, assumption-indexed correction, at the cost of a wider interval; without such a bound, $\Delta_{p,a}$ is not estimable (Proposition 2.2), and coverage of μ itself cannot be guaranteed. When the support is known to lie in $[x_{\min},x_{\max}]$ (so $\mu=\mu_c$ by Proposition 2.1(a)), no such correction is needed and I_{EL} already targets μ.

2.2 AEL Confidence Interval for the Private Mean

When the number of blocks is small, the EL method may encounter the so-called convex hull constraint problem. Specifically, the population mean may fall outside the convex hull formed by the

sample points with high probability, leading to an ill-defined EL function. To address this issue, we adapt the adjusted empirical likelihood approach introduced by Chen et al. (2008). The AEL method adds a single artificial observation to ensure that the origin lies within the convex hull of the data. We incorporate this idea into the construction of confidence intervals for the private mean.

Let $a_B = \max\{1, (\log B)/2\}$. For each candidate value θ, define the pseudo-point $\tilde{x}_{B+1}(\theta) = \theta - \frac{a_B}{B}\sum_{j=1}^{B}(\tilde{x}_j - \theta)$. This candidate-dependent construction ensures feasibility of the adjusted constraint. The pseudo-point is recomputed for each candidate θ during likelihood profiling and is not an additional privatized observation. Because it is deterministic post-processing of the privatized block means, it incurs no additional privacy cost. In the theorem below, the statistic is evaluated at $\theta = \mu_c$.

To verify feasibility, let $g_j(\theta) = \tilde{x}_j - \theta$ for $j \le B$ and $g_{B+1}(\theta) = -a_B\bar{g}(\theta)$, where $\bar{g}(\theta) = B^{-1}\sum_{j=1}^{B} g_j(\theta)$. The weights $w_j = \frac{a_B}{B(1+a_B)}$ for $j \le B$ and $w_{B+1} = \frac{1}{1+a_B}$ are positive, sum to one, and satisfy $\sum_{j=1}^{B+1} w_j g_j(\theta) = 0$.

The adjusted empirical likelihood ratio at a candidate mean θ is

$$\mathcal{R}^{adj}(\theta) = \max_{w_1,\ldots,w_{B+1}} \prod_{j=1}^{B+1}(B+1)w_j,$$
$$\text{s.t. } \sum_{j=1}^{B+1} w_j(\tilde{x}_j(\theta) - \theta) = 0,\ \sum_{j=1}^{B+1} w_j = 1,\ w_j \ge 0,$$

where $\tilde{x}_j(\theta) = \tilde{x}_j$ for $j \le B$ and the $(B+1)$st value is the pseudo-point above. Following the EL derivation, the adjusted empirical log-likelihood ratio can be expressed as

$$-2\log\mathcal{R}^{adj}(\theta) = 2\sum_{j=1}^{B+1}\log[1 + \lambda(\tilde{x}_j(\theta) - \theta)]. \tag{6}$$

Define $\mathcal{D}_\theta^{adj} = \{\lambda \in \mathbb{R} : 1 + \lambda(\tilde{x}_j(\theta) - \theta) > 0,\ j = 1, \ldots, B+1\}$. The Lagrange multiplier is the unique solution in $\mathcal{D}_\theta^{adj}$ of

$$\frac{1}{B+1}\sum_{j=1}^{B+1}\frac{\tilde{x}_j(\theta) - \theta}{1 + \lambda(\tilde{x}_j(\theta) - \theta)} = 0. \tag{7}$$

The same derivative argument as for EL proves uniqueness unless every adjusted residual is zero; in that case we set $\lambda = 0$ and $\mathcal{R}^{adj}(\theta) = 1$. The strictly positive feasibility weights displayed above imply that the adjusted residuals contain both signs whenever they are not all zero. The adjusted score therefore tends to $+\infty$ and $-\infty$ at the respective endpoints of $\mathcal{D}_\theta^{adj}$, proving existence of the stated root.

Theorem 2.5. *Let $x_1, \ldots, x_n$ be i.i.d. observations drawn from the distribution of X, and fix the clipping bounds and privacy budget as in Theorem 2.3. Suppose that the privacy-protected samples are generated using Algorithm 2 in the Appendix, and that the pseudo-point is defined as above. We assume $B = o(n)$. Then, as $B \to \infty$ and $n \to \infty$, $-2\log\mathcal{R}^{adj}(\mu_c) \xrightarrow{D} \chi_1^2$.*

Using Theorem 2.5, the AEL-based private confidence interval with $(1-\alpha)$ confidence level for μ_c is $I_{AEL} = \{\theta : -2\log\mathcal{R}^{adj}(\theta) \le \chi_1^2(\alpha)\}$. As with I_{EL}, an interval for the raw mean μ is obtained via the bias-aware inflation of Remark 2.4. The algorithm for constructing the private mean confidence interval using AEL is analogous to the EL-based algorithm and is provided as Algorithm 2 in Section D. AEL is a finite-sample adjustment of EL for the same data setting.

3 Simulation Studies

We evaluate coverage probability (CP) and average interval length (AL) under the different distributions in Section E. Section F discusses the practical selection of clipping bounds, and Table 6 lists the bounds used in the simulations. We first compare methods and then vary ϵ, B, and the clipping bounds. Each design uses 2,000 Monte Carlo replications at $\alpha = 0.05$.

3.1 COMPARISON

We compare EL and AEL with KV (Karwa & Vadhan, 2018), BH (Brawner & Honaker, 2018), AW (Awan & Wang, 2025), and PriDEL/BAPriDEL (Liu & Li, 2025). Section I describes their assumptions and our implementation choices.

Table 1: Coverage probability of μ_c and average length for the main simulation design.

		Normal(1,1)		Exp(1)		Beta(2,5)		Lognormal(0,1)		Gamma(2,2)		Weibull(2,1)	
	n	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
EL	200	0.935	3.514	0.950	4.159	0.940	0.852	0.941	5.035	0.942	3.372	0.942	1.761
	400	0.937	3.293	0.946	3.608	0.939	0.767	0.939	4.481	0.942	2.997	0.942	1.646
	600	0.938	3.060	0.944	3.194	0.941	0.698	0.940	4.045	0.942	2.700	0.943	1.528
AEL	200	0.944	3.583	0.957	4.248	0.950	0.871	0.949	5.150	0.953	3.448	0.950	1.796
	400	0.947	3.350	0.954	3.680	0.948	0.782	0.948	4.569	0.950	3.054	0.950	1.675
	600	0.945	3.109	0.949	3.253	0.947	0.709	0.946	4.112	0.945	2.746	0.948	1.553
KV	200	0.997	3.952	0.989	4.798	1.000	0.999	0.980	5.696	0.987	3.890	0.999	1.996
	400	0.980	3.828	0.969	4.590	0.999	0.998	0.962	5.457	0.985	3.755	0.998	1.979
	600	0.976	3.733	0.971	4.472	0.998	0.995	0.962	5.328	0.986	3.669	0.988	1.948
BH	200	0.960	3.576	0.968	4.683	0.962	0.943	0.968	5.655	0.959	3.739	0.964	1.901
	400	0.955	2.794	0.960	4.218	0.953	0.872	0.962	5.222	0.954	3.431	0.955	1.820
	600	0.960	2.108	0.955	3.549	0.956	0.779	0.956	4.619	0.959	3.008	0.950	1.706
AW	200	0.980	7.537	0.982	8.350	0.977	1.755	0.980	10.572	0.985	6.919	0.977	3.767
	400	0.980	3.482	0.978	5.615	0.981	0.947	0.978	6.754	0.986	3.998	0.984	1.730
	600	0.979	1.735	0.978	4.509	0.976	0.703	0.973	5.575	0.977	3.119	0.982	0.908
PriDEL	200	0.933	7.120	0.929	8.920	0.937	1.781	0.930	10.644	0.930	7.138	0.940	3.561
	400	0.943	5.392	0.945	6.737	0.948	1.344	0.941	8.040	0.945	5.390	0.944	2.692
	600	0.946	3.749	0.948	4.694	0.945	0.938	0.947	5.608	0.946	3.752	0.941	1.874
BAPriDEL	200	0.959	7.627	0.964	9.546	0.954	1.903	0.957	11.415	0.961	7.629	0.953	3.811
	400	0.953	5.610	0.956	7.021	0.957	1.406	0.957	8.391	0.954	5.610	0.950	2.807
	600	0.951	3.850	0.950	4.809	0.954	0.962	0.951	5.741	0.950	3.847	0.953	1.927

We use $(n, B) = (200, 40), (400, 70), (600, 100)$ and set $\epsilon = 0.1$. Table 1 reports coverage probability (CP) for the clipped mean μ_c, the target of Theorems 2.3 and 2.5, and average interval length (AL). We assess calibration by distance from the nominal level $1 - \alpha = 0.95$ and efficiency by interval length. Section H reports the corresponding coverage comparison for the raw mean μ.

- EL undercovers mildly in most cells (CP 0.935–0.950). AEL moves coverage closer to 0.95 in nearly every cell, with only a small increase in interval length.
- KV generally overcovers (CP as high as 1.000) and is longer than EL and AEL in every reported cell.
- Under normality, where its normal approximation is correctly specified, BH gives shorter intervals than EL and AEL at $n = 400$ and 600. In the non-normal settings, EL and AEL are shorter than BH with comparable calibration.
- In most settings, EL and AEL produce shorter intervals than AW and achieve coverage closer to the nominal 0.95 level; AW generally overcovers.
- PriDEL and BAPriDEL (Liu & Li, 2025) usually attain coverage near 0.95 but are generally among the longest procedures. For example, under Normal$(1, 1)$ at $n = 400$, PriDEL's average length (5.392) is about 1.6 times AEL's (3.350).

Across these settings, AEL offers the most consistent calibration of the two proposed methods, and EL/AEL are usually shorter than KV, BH, PriDEL, and BAPriDEL. The comparison with AW depends more strongly on sample size and distribution.

Heavy-tailed and skewed distributions. The distributions in Table 1 are light-tailed. We therefore add a shifted Student-t_3, Pareto($\alpha = 3$), Pareto($\alpha = 2.5$), and the right-skewed Lognormal(0,1). Section G reports the comparison at $n = 200, 400, 600$. EL and AEL remain close to nominal and are generally shorter than KV, BH, PriDEL, and BAPriDEL; AW is shorter in a few settings but tends to overcover.

Figure 1 directly examines the Wilks approximation underlying the proposed intervals. For Normal$(1, 1)$ data clipped to $[-1, 3]$ at $\epsilon = 0.1$, the empirical distribution functions of both likelihood-ratio statistics closely track the χ^2_1 reference distribution.

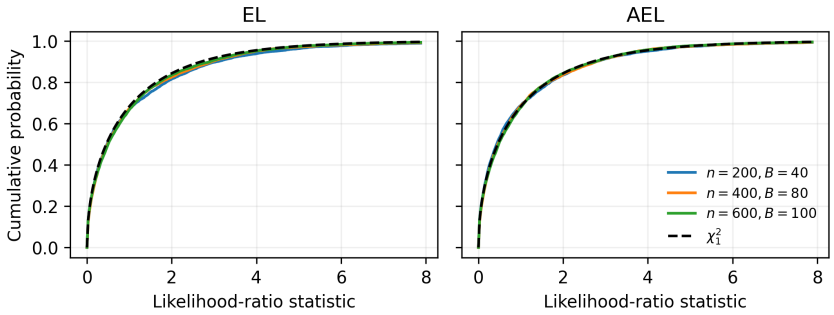


Figure 1: Empirical CDFs of $-2\log\mathcal{R}(\mu_c)$ (EL) and $-2\log\mathcal{R}^{adj}(\mu_c)$ (AEL) against the χ^2_1 reference CDF. Each curve uses 4,000 Monte Carlo replications of Normal$(1, 1)$ data clipped to $[-1, 3]$ with $\epsilon = 0.1$. The AEL approximation is especially accurate at the smallest design.

3.2 Robustness Test

We next vary the privacy budget ϵ, the number of blocks B, and the clipping bounds while holding the sample size at $n = 400$.

First, we examine the sensitivity to ϵ. For these simulations, we set $B = 70$. Table 2 reports coverage probabilities and average interval lengths for EL and AEL across different values of ϵ. The coverage probabilities are close to 0.95 for different privacy budgets. The average length of the confidence intervals decreases as ϵ increases. This is because, as ϵ increases, the variance of the noise added to the private mean becomes smaller.

Table 2: Coverage probability of μ_c and average length under different privacy budgets.

		Normal(1,1)		Exp(1)		Beta(2,5)		Lognormal(0,1)		Gamma(2,2)		Weibull(2,1)	
Method	ϵ	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
EL	0.1	0.941	3.302	0.945	3.608	0.941	0.768	0.942	4.490	0.939	2.995	0.942	1.639
	0.3	0.942	1.581	0.941	1.761	0.941	0.381	0.940	2.235	0.934	1.488	0.939	0.790
	0.5	0.945	0.962	0.944	1.177	0.940	0.238	0.940	1.434	0.932	0.951	0.939	0.481
AEL	0.1	0.949	3.359	0.953	3.677	0.950	0.783	0.949	4.579	0.949	3.056	0.949	1.667
	0.3	0.949	1.634	0.950	1.809	0.948	0.393	0.950	2.301	0.942	1.531	0.948	0.816
	0.5	0.952	0.995	0.948	1.214	0.950	0.246	0.951	1.481	0.940	0.982	0.945	0.497

Next, we evaluate sensitivity to the number of blocks B. For these simulations, we set $\epsilon = 0.1$. Table 3 reports coverage probabilities and average interval lengths under different values of B. Coverage remains close to 0.95 across the block counts considered, while interval length increases with B because smaller blocks require more Laplace noise.

Finally, we evaluate the sensitivity of the clipping bounds $x_{\min}$ and $x_{\max}$. For these simulations, we set $B = 70$ and $\epsilon = 0.1$. We introduce a clipping perturbation factor δ to scale the original bounds. The perturbed clipping range is defined as $[(1 - \delta)x_{\min}, (1 + \delta)x_{\max}]$. We consider $\delta \in \{-0.3, -0.2, -0.1, 0.1, 0.2, 0.3\}$, which either shrinks or expands the clipping range. This allows us to measure how the performance of the methods changes when the sensitivity is increased or decreased. Table 4 reports coverage probabilities and average interval lengths for EL and AEL across different values of δ. The coverage probabilities are close to the nominal level (0.95) for different clipping perturbation factors.

Table 3: Coverage probability of μ_c and average length for different numbers of blocks.

Method	B	Normal(1,1) CP	Normal(1,1) AL	Exp(1) CP	Exp(1) AL	Beta(2,5) CP	Beta(2,5) AL	Lognormal(0,1) CP	Lognormal(0,1) AL	Gamma(2,2) CP	Gamma(2,2) AL	Weibull(2,1) CP	Weibull(2,1) AL
EL	20	0.911	2.299	0.912	2.370	0.913	0.524	0.911	3.042	0.910	2.023	0.909	1.140
	40	0.933	2.935	0.937	3.049	0.935	0.665	0.935	3.851	0.934	2.568	0.933	1.455
	50	0.938	3.098	0.943	3.276	0.938	0.709	0.940	4.120	0.936	2.750	0.935	1.535
	80	0.939	3.372	0.944	3.719	0.943	0.789	0.945	4.620	0.941	3.085	0.940	1.674
	100	0.940	3.452	0.947	3.924	0.942	0.821	0.944	4.838	0.943	3.230	0.943	1.717
AEL	20	0.936	2.485	0.936	2.553	0.938	0.565	0.938	3.270	0.936	2.177	0.935	1.232
	40	0.944	3.036	0.948	3.163	0.943	0.690	0.945	3.997	0.943	2.667	0.945	1.506
	50	0.949	3.178	0.952	3.378	0.947	0.729	0.950	4.243	0.950	2.831	0.948	1.576
	80	0.946	3.420	0.950	3.785	0.948	0.802	0.949	4.702	0.949	3.140	0.947	1.698
	100	0.945	3.489	0.953	3.981	0.948	0.831	0.950	4.902	0.950	3.269	0.947	1.736

Table 4: Coverage probability of μ_c and average length under different clipping perturbations.

Method	δ	Normal(1,1) CP	Normal(1,1) AL	Exp(1) CP	Exp(1) AL	Beta(2,5) CP	Beta(2,5) AL	Lognorm(0,1) CP	Lognorm(0,1) AL	Gamma(2,2) CP	Gamma(2,2) AL	Weibull(2,1) CP	Weibull(2,1) AL
EL	-0.3	0.938	2.715	0.941	2.674	0.940	0.569	0.940	3.307	0.936	2.227	0.942	1.140
	-0.2	0.939	2.932	0.942	2.988	0.940	0.638	0.941	3.714	0.937	2.493	0.941	1.317
	-0.1	0.940	3.129	0.943	3.298	0.941	0.704	0.941	4.107	0.938	2.748	0.941	1.484
	0.1	0.941	3.453	0.948	3.919	0.942	0.830	0.943	4.870	0.941	3.243	0.942	1.784
	0.2	0.941	3.582	0.948	4.215	0.943	0.892	0.944	5.241	0.942	3.485	0.942	1.923
	0.3	0.940	3.695	0.950	4.519	0.943	0.953	0.944	5.613	0.943	3.730	0.943	2.057
AEL	-0.3	0.949	2.765	0.953	2.725	0.949	0.579	0.950	3.368	0.947	2.269	0.949	1.160
	-0.2	0.948	2.983	0.951	3.048	0.949	0.650	0.950	3.785	0.948	2.541	0.950	1.341
	-0.1	0.948	3.182	0.952	3.366	0.950	0.718	0.950	4.185	0.948	2.801	0.950	1.510
	0.1	0.949	3.514	0.954	3.994	0.951	0.847	0.950	4.968	0.949	3.307	0.949	1.816
	0.2	0.949	3.646	0.955	4.310	0.951	0.909	0.951	5.350	0.951	3.558	0.948	1.958
	0.3	0.949	3.762	0.955	4.616	0.952	0.972	0.952	5.728	0.951	3.805	0.948	2.095

4 Data Analysis

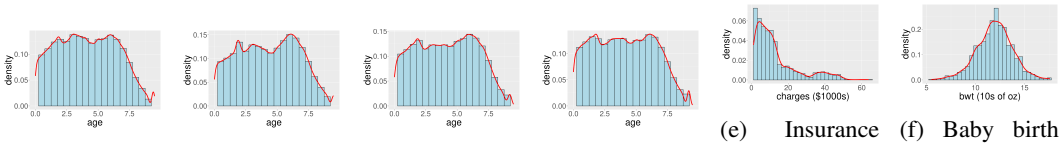
We evaluate the methods using the 2023 American Community Survey Public Use Microdata Sample[1], accessed through the R package `tidycensus` (Walker, 2023).

We focus on the variable `age` and compare its distribution across four states: California (CA), New York (NY), Georgia (GA), and Texas (TX). The total population sizes for each state are: CA: 1,853,429; NY: 973,112; GA: 483,479; TX: 1,306,799. We scaled the `age` variable by dividing by 10. Following the evaluation framework of Brawner & Honaker (2018), we treat the full dataset for each state as the population and compute the mean as the true parameter. For each state, we restrict the variable `age` to its minimum and maximum. Then, for each experimental run, we draw a random sample of size $n = 400$ from the population and construct confidence intervals for the mean using each method. We repeat this process 2,000 times to estimate coverage probability and average interval length. The significance level is set at $\alpha = 0.05$, the privacy parameter is set at $\epsilon = 0.1$, and the number of blocks is fixed at $B = 70$.

Figure 2 shows the empirical age distributions for the four states. It is difficult to identify a single parametric distribution that adequately describes all four. The visibly different distributional shapes in Figure 2 illustrate why inference that does not require a common parametric model is useful here. For completeness, Section J reports pairwise Wilcoxon rank-sum tests. These tests provide evidence of differences between the state distributions.

To broaden the real-data evidence beyond the ACS age variable, we analyze two additional public datasets: the *insurance* dataset (1,338 individuals; we take the mean of *charges*, medical insurance cost) and the classic Nolan Stat Labs *babies* dataset (1,174 individuals after removing missing-value

[1] www.census.gov/programs-surveys/acs/microdata.html



(a) CA (b) NY (c) GA (d) TX (e) Insurance charges (f) Baby birth weight

Figure 2: Age distributions by state (CA, NY, GA, and TX), insurance charges, and baby birth weight.

sentinel codes, following standard practice for this dataset; we take the mean of *bwt*, birth weight). These populations contain 1,174–1,338 records. We use $n = 200$, $B = 40$, $\epsilon = 0.1$, and $N = 2000$. We resample with replacement from each population and report values in \$1000s for *charges* and units of 10 ounces for *bwt*, matching the age/10 convention used for the states. Figure 2 also shows the two population distributions. The *charges* distribution is heavily right-skewed with a secondary mode (consistent with a smoker/non-smoker mixture in this dataset), whereas the *bwt* distribution is approximately symmetric and unimodal.

Table 5: Coverage probability and average length for the ACS `age` variable across four states, insurance charges, and baby birth weight. Clipping bounds are the population minimum and maximum, so $\mu = \mu_c$ and coverage is identical under either target.

	CA		NY		GA		TX		charges		bwt	
Method	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
EL	0.941	7.706	0.940	7.803	0.940	7.558	0.939	7.547	0.948	52.019	0.939	10.659
AEL	0.948	7.840	0.949	7.939	0.950	7.690	0.946	7.678	0.952	53.044	0.947	10.856
KV	0.930	8.447	0.931	8.552	0.927	8.274	0.928	8.270	0.857	49.930	0.946	11.263
BH	0.956	8.587	0.957	8.690	0.953	8.386	0.953	8.360	0.966	58.703	0.960	11.558
AW	0.982	8.578	0.978	8.628	0.980	8.402	0.977	8.504	0.983	102.724	0.974	20.705
PriDEL	0.945	12.761	0.940	12.866	0.939	12.488	0.947	12.464	0.940	111.805	0.934	21.578
BAPriDEL	0.950	13.186	0.960	13.404	0.955	12.941	0.959	12.863	0.958	119.453	0.960	23.151

Table 5 compares EL and AEL with KV, BH, AW, and adapted versions of PriDEL and BAPriDEL. Across the four ACS states, EL and AEL have CP between 0.939 and 0.950 and are shorter than every baseline. Their average lengths are about 7–14% below those of KV, BH, and AW. PriDEL and BAPriDEL also have coverage near the nominal level but are more than 1.6 times as long as AEL in each state.

For *charges*, EL and AEL have CP 0.948 and 0.952 and are the shortest methods among those with CP in the 0.95 ± 0.01 band; KV is slightly shorter but undercovers substantially (CP $= 0.857$). For *bwt*, AEL has CP 0.947 and the second-shortest interval, after EL. AW overcovers on both datasets and has substantially longer intervals. These results show a favorable coverage–length trade-off for EL and especially AEL, while also illustrating that the ordering of methods depends on the calibration criterion.

5 Conclusion

We proposed EL and AEL confidence intervals for a population mean under ϵ-DP, built from Laplace-privatized block means. Although these block means form a triangular array with vanishing variance, both likelihood-ratio statistics have χ^2_1 limits at the clipped mean μ_c, so the intervals require neither a parametric model nor a separate variance estimate. For the raw mean μ, Proposition 2.1 quantifies the clipping bias and Remark 2.4 gives a conservative correction under a tail-moment bound. In simulations and real-data analyses, AEL is the better-calibrated variant, and both methods typically yield shorter intervals than existing approaches while maintaining near-nominal coverage. Coverage remains close to the nominal level across privacy budgets, numbers of blocks, and perturbed clipping bounds, as well as under heavy-tailed and skewed distributions.

AI USE STATEMENT

In this work, we used generative AI tools to provide critical ingredients for proving mathematical claims, design or provide feedback on research methodology or experiments, implement methods, and interpret results. We have not used generative AI tools to assist in the writing of proofs, generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, propose or refine hypotheses, assist with translation, clean and reformat dataset, or support qualitative and thematic data analysis, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools to create or modify scientific figures or images, create or edit software code, edit a research paper to improve readability, and suggest a structure for a research paper. We have reviewed all AI-assisted work. The authors checked all proof revisions against the underlying mathematical arguments, executed and verified AI-assisted code, inspected the resulting figure, and reviewed all textual edits for accuracy and consistency with the research. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Galen Andrew, Om Thakkar, H. Brendan McMahan, and Swaroop Ramaswamy. Differentially private learning with adaptive clipping. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Jordan Awan and Zhanyu Wang. Simulation-based, finite-sample inference for privatized data. *Journal of the American Statistical Association*, 120(551):1669–1682, 2025.

Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 14475–14485. Curran Associates, Inc., 2020.

Thomas Brawner and James Honaker. Bootstrap inference and differential privacy: Standard errors for free. In *Summer Meetings of the Society for Political Methodology*, Provo, UT, 2018.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography*, pp. 635–658, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.

T Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

Karan Chadha, John Duchi, and Rohith Kuditipudi. Private confidence sets. In *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

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

Jiahua Chen, Asokan Mulayath Variyath, and Bovas Abraham. Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443, 2008.

Christian Covington, Xi He, James Honaker, and Gautam Kamath. Unbiased statistical estimation and valid confidence intervals under differential privacy. *Statistica Sinica*, 35:651–670, 2025.

Vito D'Orazio, James Honaker, and Gary King. Differential privacy for social science inference. *Sloan Foundation Economics Research Paper*, 2015.

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.

Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.

David Durfee. Unbounded differentially private quantile and maximum estimation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pp. 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

Georgina Evans and Gary King. Statistically valid inferences from differentially private data releases, with application to the facebook URLs dataset. *Political Analysis*, 31(1):1–21, 2023.

Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. Statistically valid inferences from privacy-protected data. *American Political Science Review*, 117(4):1275–1290, 2023.

Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. Parametric bootstrap for differentially private confidence intervals. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 1598–1618. PMLR, 28–30 Mar 2022.

Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. In *Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR 139*, 2021.

Haim Kaplan, Shachar Schnapp, and Uri Stemmer. Differentially private approximate quantiles. In *Proceedings of the 39th International Conference on Machine Learning (ICML), PMLR 162*, 2022.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 44:1–44:9, 2018.

Ariel Kleiner, Ameet Talwalkar, Purnamrita Sarkar, and Michael I Jordan. A scalable bootstrap for massive data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(4): 795–816, 2014.

Nicole A Lazar. A review of empirical likelihood. *Annual Review of Statistics and Its Application*, 8 (1):329–344, 2021.

Pangpang Liu and Yichuan Zhao. A review of recent advances in empirical likelihood. *WIREs Computational Statistics*, 15(3):e1599, 2023.

Qianqian Liu and Zhouping Li. Distributed empirical likelihood inference with privacy guarantees. *Statistics and Computing*, 35(4):88, 2025.

Yi Liu, Qirui Hu, Lei Ding, and Linglong Kong. Online local differential private quantile inference via self-normalization. In *International Conference on Machine Learning*, pp. 21698–21714. PMLR, 2023.

Art B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75 (2):237–249, 1988.

Art B. Owen. *Empirical likelihood*. Chapman and Hall/CRC, Boca Raton, 2001.

Or Sheffet. Differentially private ordinary least squares. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3105–3114. PMLR, 2017.

Kyle Walker. *Analyzing US Census Data: Methods, Maps, and Models in R*. Chapman & Hall/CRC The R Series. CRC Press, New York, 2023.

Yue Wang, Daniel Kifer, and Jaewoo Lee. Differentially private confidence intervals for empirical risk minimization. *Journal of Privacy and Confidentiality*, 9(1), 2019.

Myeonghun Yu, Zhao Ren, and Wen-xin Zhou. Gaussian differentially private robust mean estimation and inference. *Bernoulli*, 30(4):3059–3088, 2024.

A Proof of Proposition 2.2

Fix $x_{\min} < x_0 < x_{\max}$ and $q \in (0,1)$, and let $P_M(X = x_0) = 1-q$ and $P_M(X = x_{\max}+M) = q$. The clipped distribution is identical for every $M > 0$, whereas its raw mean $(1-q)x_0 + q(x_{\max}+M)$ is unbounded in M. Thus every mechanism based only on the clipped observations has the same output distribution under raw distributions with arbitrarily different means.

B Differential Privacy Background

This section reviews the differential privacy background needed for our results; see Dwork et al. (2006) for a comprehensive treatment. Under differential privacy, the output of an algorithm should remain nearly unchanged when a single individual record of the data is modified. We say that two data sets are neighbors if they differ in exactly one data record.

Definition B.1. (Differential privacy (Dwork et al., 2006)) A randomized algorithm M is ϵ-differentially private (ϵ-DP) if for neighboring data sets X and X' and any subset $O \in Range(M)$,

$$\mathbb{P}(M(X) \in O) \leq e^{\epsilon}\mathbb{P}(M(X') \in O).$$

The core idea behind differential privacy is to ensure that the inclusion or exclusion of one data record in a dataset does not significantly impact the output of a computation. The goal is to protect the privacy of individuals contributing data while still allowing for meaningful and accurate analyses on the aggregated dataset.

In the context of differential privacy, the parameter ϵ serves as a crucial control for privacy leakage and is termed the privacy budget. A smaller value of ϵ signifies more robust privacy protection, albeit potentially at the expense of accuracy. The pursuit of differential privacy necessitates the introduction of some level of noise into the data, with the magnitude of this noise typically determined by the sensitivity, as defined below.

Definition B.2. (Sensitivity (Dwork et al., 2006)) The L_1 sensitivity of a function f is

$$\Delta f = \max_{X,X'} \|f(X) - f(X')\|_1,$$

where X and X' are any two neighboring data sets.

To achieve ϵ-DP, the following Laplace mechanism is frequently considered.

Theorem B.3. *(Laplace mechanism (Dwork et al., 2006)) Given a function $f : \mathcal{X} \to \mathbb{R}$, the Laplace mechanism outputs*

$$\tilde{f}(X) = f(X) + Z \text{ with } Z \sim Lap\left(\frac{\Delta f}{\epsilon}\right),$$

where $Lap(b)$ is the Laplace distribution with the mean 0, the scale b, and the probability density function $e^{-|z|/b}/(2b)$. $\tilde{f}$ is ϵ-differentially private.

The Laplace mechanism introduces additional randomness as a privacy safeguard, a measure that often incurs a trade-off by affecting the accuracy and efficiency of estimation or inference. Statistical inference must therefore account for this additional randomness.

C Confidence Interval for Private Mean via EL

We present the construction of the confidence interval for the private mean using EL in Algorithm 1.

D Confidence Interval for Private Mean via AEL

Algorithm 2 uses the same private block releases as Algorithm 1 and adds the candidate-dependent pseudo-point defined in Section 2 during post-processing.

E Data Generation

We generate data from six baseline distributions (Tables 1 to 4) using the corresponding R functions listed below.

- Normal(1,1): the normal distribution with mean 1 and standard deviation 1, using the R function `rnorm`.
- Exp(1): the exponential distribution with the rate parameter 1, using the R function `rexp`.
- Beta(2, 5): the beta distribution with shape parameters 2 and 5, using the R function `rbeta`.
- Gamma(2, 2): the gamma distribution with shape 2 and rate 2, using the R function `rgamma`.
- Weibull(2,1): the Weibull distribution with shape 2 and scale 1, using the R function `rweibull`.
- Lognormal(0, 1): the lognormal distribution whose logarithm has mean 0 and standard deviation 1, using the R function `rlnorm`.

More details about these distributions can be found in the documentation for the R functions.

For Section G, we additionally generate data from three heavy-tailed or skewed distributions, together with the already-listed Lognormal(0,1) (used there with different, tighter clipping bounds, Table 6):

- t_3: a Student-t distribution with 3 degrees of freedom, shifted to have mean 5, generated as `5 + rt(n, df=3)`.
- Pareto($\alpha = 3$): a Pareto distribution with scale 1 and shape 3 (mean 1.5), generated via inverse-CDF sampling as `1/(1-runif(n))^(1/3)`.
- Pareto($\alpha = 2.5$): a Pareto distribution with scale 1 and shape 2.5 (mean 1.667), generated as `1/(1-runif(n))^(1/2.5)`.

F Practical Guidance for Clipping Bounds and the Number of Blocks

Choosing the clipping bounds in practice. We distinguish three practical scenarios. (i) *Public or domain bounds.* When a physical or regulatory bound is known a priori (e.g., age, a bounded survey scale), $x_{\min}, x_{\max}$ can be set directly from this domain knowledge, without spending any privacy budget; this is the case for the age data analyzed in Section 4. (ii) *A coarse public range is available but too wide.* One can split the budget $\epsilon = \epsilon_q + \epsilon_m$, spend ϵ_q on differentially private quantile estimation (e.g., the $1\%/99\%$ quantiles, via Gillenwater et al. (2021); Kaplan et al. (2022)) to tighten a loose public range into $[x_{\min}, x_{\max}]$, and then run Algorithm 1 (or 2) with the remaining budget ϵ_m; sequential composition gives overall ϵ-DP. This mirrors the private-quantile-based clipping used in adaptive DP gradient clipping (Andrew et al., 2021). Because this still requires a coarse public range to search over, it tightens the loose bound rather than removing the need for one. (iii) *Genuinely unbounded data, with no public range.* One can instead spend ϵ_q on an *unbounded* DP quantile estimator (Durfee, 2023), which locates extreme quantiles via the Sparse Vector Technique without requiring a prior search range, applied once in each direction (i.e., once to X and once to $-X$) to obtain $x_{\max}$ and $x_{\min}$ respectively; the remaining budget ϵ_m is then used exactly as in (ii). The clipping bounds themselves are a genuine bias–variance choice regardless of how they are obtained: widening them reduces the clipping bias of Proposition 2.1 but increases the sensitivity $x_{\max} - x_{\min}$ and hence the privacy noise, a trade-off we probe empirically via the clipping-perturbation sensitivity analysis in Section 3.2.

For each distribution, the clipping bounds used in our simulation studies are selected using distributional properties to cover most of the probability mass. Table 6 shows the clipping bounds $(x_{\min}, x_{\max})$ used in the simulation studies for each data-generating distribution.

Practical choices for B and the EL variant. B controls a two-sided trade-off: it must grow with n for the χ^2_1 calibration of Theorem 2.3 to take hold, while larger B means smaller blocks and more privacy noise per block. A convenient public-data default is

$$B \approx n/\log n,$$

which satisfies the asymptotic conditions and gives $B \approx 38, 67, 94$ for $n = 200, 400, 600$, close to the values used in Table 1; coverage is essentially flat over this range (Table 3). When B is small, the

Table 6: Clipping bounds $(x_{\min}, x_{\max})$ and the resulting clipping bias $\mu - \mu_c$ (computed by numerical integration) used in the simulation studies for each data-generating distribution.

Distribution	$x_{\min}$	$x_{\max}$	$\mu - \mu_c$
Normal(1, 1)	-1	3	0.000
Exp(1)	0	5	0.007
Beta(2,5)	0	1	0.000
Lognormal(0,1) (Table 1)	0	5.97	0.135
Gamma(2,2)	0	4	0.002
Weibull(2,1)	0	2	0.004
t_3 (Table 7)	1.82	8.18	0.000
Pareto(3) (Table 7)	1.01	3.42	0.043
Pareto(2.5) (Table 7)	1.01	4.38	0.073
Lognormal(0,1) (Table 7)	0.14	7.10	0.099

EL ratio can be undefined because the candidate mean falls outside the convex hull of the privatized block means. AEL's pseudo-point guarantees feasibility and restores near-nominal coverage in this regime (Table 3), at the cost of a marginally longer interval. We therefore recommend AEL as the default, with EL as a lightweight alternative when B is known to be large.

G Heavy-Tailed and Skewed Distributions

We evaluate our methods on three heavy-tailed or skewed generating distributions where a normal working model is badly misspecified: a shifted Student-t_3, Pareto($\alpha = 3$), and Pareto($\alpha = 2.5$), together with the already-considered Lognormal(0,1), which is itself right-skewed. We use $n = 200, 400, 600$ with $B = 40, 70, 100$ respectively, $\epsilon = 0.1$, $N = 2000$, and tight central-mass clipping bounds for each distribution (Table 6 reports the bounds and the resulting clipping bias $\mu - \mu_c$). We

Table 7: Coverage probability of μ_c and average length under heavy-tailed and skewed distributions.

n	Method	t_3		Pareto(3)		Pareto(2.5)		Lognormal(0,1)	
		CP	AL	CP	AL	CP	AL	CP	AL
200	EL	0.938	5.604	0.947	1.992	0.949	2.777	0.945	5.785
	AEL	0.948	5.710	0.954	2.035	0.957	2.835	0.953	5.912
	KV	0.972	6.121	0.995	2.370	0.994	3.281	0.969	6.567
	BH	0.968	6.122	0.970	2.320	0.968	3.233	0.974	6.591
	AW	0.967	11.555	0.981	3.956	0.982	5.462	0.976	11.561
	PriDEL	0.939	11.378	0.941	4.310	0.941	6.027	0.935	12.430
	BAPriDEL	0.953	12.085	0.962	4.568	0.963	6.388	0.957	13.320
400	EL	0.941	5.246	0.942	1.718	0.945	2.382	0.944	5.032
	AEL	0.948	5.337	0.951	1.754	0.952	2.428	0.950	5.134
	KV	0.949	5.900	0.989	2.309	0.986	3.167	0.961	6.258
	BH	0.951	5.959	0.963	2.268	0.960	3.133	0.962	5.994
	AW	0.981	5.211	0.974	2.913	0.975	4.128	0.981	8.071
	PriDEL	0.945	8.599	0.942	3.262	0.943	4.561	0.936	9.406
	BAPriDEL	0.946	8.884	0.949	3.377	0.949	4.722	0.954	9.742
600	EL	0.939	4.875	0.944	1.518	0.945	2.093	0.941	4.469
	AEL	0.944	4.953	0.947	1.545	0.948	2.129	0.948	4.547
	KV	0.952	5.825	0.995	2.269	0.981	3.094	0.963	6.106
	BH	0.956	5.919	0.956	2.188	0.955	2.954	0.955	5.165
	AW	0.978	2.492	0.972	2.468	0.970	3.450	0.971	6.883
	PriDEL	0.957	5.990	0.939	2.269	0.939	3.173	0.941	6.545
	BAPriDEL	0.959	6.123	0.951	2.310	0.951	3.231	0.959	6.693

report CP for the clipped mean μ_c here, as in Table 1; Table 9 repeats the comparison for the raw mean μ. EL and AEL remain near 0.95 across the four distributions and are shorter than KV and BH in every cell. AEL is, for example, 8–16% shorter than BH for the two Pareto distributions. PriDEL

and BAPriDEL also have coverage near the nominal level but are usually longer than EL and AEL. AW generally overcovers (CP 0.967–0.982); its relative length improves with n and it is the shortest method for shifted t_3 at $n = 400$ and 600. Thus no method dominates in interval length alone, while EL and AEL provide the most stable combination of near-nominal coverage and comparatively short intervals in these designs.

H Coverage of the Raw Mean μ versus the Clipped Mean μ_c

Table 1 evaluates coverage of $\mu_c = \mathbb{E}[\mathrm{clip}(X)]$, the target of Theorems 2.3 and 2.5. Table 8 repeats the comparison with coverage evaluated at the raw mean μ. Average length does not depend on which target is used to score coverage; the small AL differences between Tables 1 and 8 reflect Monte Carlo variation from separate simulation runs. Table 7 similarly reports CP for μ_c; Table 9 repeats that

Table 8: Coverage probability (CP) and average length (AL) of confidence intervals at $\alpha = 0.05$.

		Normal(1,1)		Exp(1)		Beta(2,5)		Lognormal(0,1)		Gamma(2,2)		Weibull(2,1)	
	n	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
EL	200	0.939	3.525	0.949	4.157	0.943	0.853	0.945	5.045	0.944	3.376	0.939	1.754
	400	0.942	3.301	0.946	3.609	0.942	0.768	0.944	4.488	0.944	2.997	0.940	1.639
	600	0.939	3.067	0.941	3.191	0.940	0.698	0.938	4.050	0.944	2.701	0.942	1.522
AEL	200	0.948	3.588	0.952	4.241	0.953	0.871	0.951	5.159	0.954	3.452	0.947	1.788
	400	0.949	3.359	0.950	3.680	0.950	0.783	0.951	4.580	0.949	3.057	0.948	1.668
	600	0.943	3.117	0.947	3.250	0.944	0.710	0.945	4.118	0.948	2.746	0.946	1.547
KV	200	0.997	3.951	0.989	4.803	1.000	0.999	0.981	5.695	0.986	3.882	1.000	1.997
	400	0.979	3.825	0.968	4.589	1.000	0.998	0.963	5.463	0.986	3.762	0.998	1.979
	600	0.977	3.731	0.971	4.463	0.998	0.995	0.960	5.320	0.986	3.666	0.987	1.947
BH	200	0.961	3.579	0.967	4.683	0.960	0.940	0.969	5.651	0.959	3.746	0.963	1.901
	400	0.956	2.803	0.963	4.213	0.955	0.873	0.960	5.221	0.956	3.431	0.954	1.818
	600	0.959	2.114	0.954	3.548	0.957	0.778	0.955	4.616	0.959	3.008	0.948	1.707
AW	200	0.970	7.505	0.980	8.317	0.978	1.713	0.974	10.418	0.980	6.859	0.976	3.758
	400	0.979	3.485	0.975	5.693	0.982	0.955	0.978	6.751	0.980	4.062	0.978	1.758
	600	0.978	1.747	0.975	4.543	0.981	0.702	0.958	5.554	0.977	3.099	0.976	0.898
PriDEL	200	0.937	7.144	0.934	8.991	0.943	1.793	0.939	10.663	0.933	7.177	0.941	3.577
	400	0.937	5.407	0.945	6.770	0.943	1.354	0.937	8.069	0.955	5.389	0.942	2.708
	600	0.942	3.762	0.950	4.715	0.953	0.937	0.947	5.615	0.941	3.764	0.937	1.883
BAPriDEL	200	0.956	7.656	0.960	9.454	0.957	1.910	0.958	11.427	0.961	7.618	0.964	3.792
	400	0.953	5.600	0.948	6.973	0.951	1.402	0.955	8.357	0.949	5.593	0.949	2.803
	600	0.959	3.847	0.951	4.810	0.951	0.960	0.961	5.742	0.945	3.856	0.951	1.917

heavy-tailed comparison against the raw mean μ.

Throughout, the primary tables (Tables 1 to 4 and 7) report coverage of μ_c, and Tables 8 and 9 give the corresponding raw-mean coverage for the main and heavy-tailed designs. Lognormal(0,1) drives essentially all of the gap (Normal and Beta have zero clipping bias under the chosen bounds, and the bias is small for Exponential, Gamma, and Weibull; Lognormal has the largest bias in Table 6, $\mu - \mu_c \approx 0.135$). These results show that the empirical ranking in the reported main design is not driven solely by the choice between μ and μ_c, while the formal guarantee remains specific to μ_c unless the bias correction applies.

I Description of Baseline Methods

Karwa & Vadhan (2018) proposed a histogram-based interval for the mean of a normal population. We also evaluate this method, denoted KV, under non-normal distributions, where its normal-model calibration is misspecified. Brawner & Honaker (2018) proposed a bootstrap method that avoids a parametric model for the original data but uses a normal approximation to the sampling distribution. The method, denoted BH, is based on zero-concentrated differential privacy (zCDP), which can be converted to (ϵ, δ)-DP (Bun & Steinke, 2016). Awan & Wang (2025) constructed intervals by

Table 9: Coverage probability of the raw mean μ and average length for the same heavy-tailed comparison as Table 7.

n	Method	t_3		Pareto(3)		Pareto(2.5)		Lognormal(0,1)	
		CP	AL	CP	AL	CP	AL	CP	AL
200	EL	0.938	5.604	0.948	1.992	0.951	2.777	0.948	5.785
	AEL	0.948	5.710	0.955	2.035	0.957	2.835	0.954	5.912
	KV	0.972	6.121	0.994	2.370	0.994	3.281	0.968	6.567
	BH	0.968	6.122	0.970	2.320	0.968	3.233	0.974	6.591
	AW	0.967	11.555	0.981	3.956	0.980	5.462	0.976	11.561
	PriDEL	0.939	11.378	0.939	4.310	0.939	6.027	0.938	12.430
	BAPriDEL	0.953	12.085	0.964	4.568	0.965	6.388	0.958	13.320
400	EL	0.941	5.246	0.945	1.718	0.949	2.382	0.945	5.032
	AEL	0.948	5.337	0.953	1.754	0.952	2.428	0.952	5.134
	KV	0.949	5.900	0.989	2.309	0.984	3.167	0.961	6.258
	BH	0.951	5.959	0.963	2.268	0.960	3.133	0.961	5.994
	AW	0.981	5.211	0.970	2.913	0.968	4.128	0.979	8.071
	PriDEL	0.945	8.599	0.944	3.262	0.945	4.561	0.936	9.406
	BAPriDEL	0.946	8.884	0.948	3.377	0.947	4.722	0.955	9.742
600	EL	0.939	4.875	0.942	1.518	0.943	2.093	0.944	4.469
	AEL	0.944	4.953	0.947	1.545	0.948	2.129	0.946	4.547
	KV	0.952	5.825	0.994	2.269	0.978	3.094	0.963	6.106
	BH	0.956	5.919	0.957	2.188	0.954	2.954	0.955	5.165
	AW	0.978	2.492	0.961	2.468	0.959	3.450	0.960	6.883
	PriDEL	0.957	5.990	0.935	2.269	0.935	3.173	0.947	6.545
	BAPriDEL	0.959	6.123	0.948	2.310	0.947	3.231	0.960	6.693

simulating privatized statistics under a working generative model. We implement this AW procedure with a location-scale normal working model, including when the generating distribution is non-normal; its results therefore incorporate the effect of model misspecification. Finally, Liu & Li (2025) proposed PriDEL and BAPriDEL for K physically separate machines under (ϵ, δ)-DP. We include adapted versions of their scalar-mean procedures as empirical baselines.

J Pairwise Wilcoxon rank-sum tests

For completeness, Table 10 reports pairwise Wilcoxon rank-sum tests, which provide evidence of distributional differences across states.

Table 10: P-values from pairwise Wilcoxon rank-sum tests

	CA	GA	NY
GA	2×10^{-16}		
NY	2×10^{-16}	2×10^{-16}	
TX	2×10^{-16}	2×10^{-16}	2×10^{-16}

K Proof of Proposition 2.1

(a) By definition, $\mathrm{clip}(X) - X = \max(x_{\min} - X, 0) - \max(X - x_{\max}, 0)$: if $X < x_{\min}$ this reads $x_{\min} - X = (x_{\min} - X) - 0$; if $X > x_{\max}$ it reads $x_{\max} - X = 0 - (X - x_{\max})$; and if $x_{\min} \le X \le x_{\max}$ both sides are 0. Taking expectations gives $\mu_c - \mu = \mathbb{E}[\max(x_{\min} - X, 0)] - \mathbb{E}[\max(X - x_{\max}, 0)]$. The triangle inequality gives $|\mu_c - \mu| \le \mathbb{E}[\max(x_{\min} - X, 0)] + \mathbb{E}[\max(X - x_{\max}, 0)]$. Because the two terms on the right are nonnegative, their sum is zero if and only if both vanish almost surely. This condition is equivalent to $X \in [x_{\min}, x_{\max}]$ almost surely, or $\mathbb{P}(x_{\min} \le X \le x_{\max}) = 1$.

(b) If $X \ge x_{\min}$ almost surely, then $\max(x_{\min} - X, 0) = 0$ almost surely, so by part (a), $\mu_c - \mu = -\mathbb{E}[\max(X - x_{\max}, 0)] \le 0$.

(c) Fix $a \in (x_{\min}, x_{\max})$ and $p > 1$. Put $d_+ = x_{\max} - a > 0$. For $y \geq d_+$, define $h_+(y) = y^p/(p\, d_+^{p-1}) - (y - d_+)$. Then $h_+(d_+) = d_+/p > 0$ and $h'_+(y) = (y/d_+)^{p-1} - 1 \geq 0$, so $h_+ > 0$ on $[d_+, \infty)$; for $y < d_+$ the left side below is 0. Hence

$$(y - d_+)_+ \leq \frac{|y|^p}{p\, d_+^{p-1}} \qquad \text{for every } y \in \mathbb{R}.$$

Taking $y = X - a$ gives

$$\mathbb{E}[\max(X - x_{\max}, 0)] \leq \frac{\mathbb{E}|X - a|^p}{p\,(x_{\max} - a)^{p-1}}.$$

For the lower tail, put $d_- = a - x_{\min} > 0$ and apply the same inequality to $y = a - X$. Because $\max(x_{\min} - X, 0) = ((a - X) - d_-)_+$ and $|a - X| = |X - a|$, we obtain

$$\mathbb{E}[\max(x_{\min} - X, 0)] \leq \frac{\mathbb{E}|X - a|^p}{p\,(a - x_{\min})^{p-1}}.$$

□

L Proof of Theorem 2.3

The main difficulty is that the effective sample $\tilde{x}_1, \ldots, \tilde{x}_B$ is not a fixed-distribution i.i.d. sample: it is a triangular array whose row variance σ_B^2 vanishes as $n \to \infty$, since each $\tilde{x}_j$ is itself an average of $m = n/B \to \infty$ clipped observations. Steps 1–3 below establish the uniform moment and maximal-control conditions this triangular array satisfies, which is what allows the expansion in Step 4 to go through.

Throughout, $m = n/B$ is an integer, as assumed in Section 2, so every block has exactly m observations. Write $D_j = \frac{1}{m}\sum_{i=(j-1)m+1}^{jm} \mathrm{clip}(x_i)$ for the j-th clipped block mean and $L_j = z_j \sim \mathrm{Lap}(w/(m\epsilon))$, $w := x_{\max} - x_{\min}$, for the corresponding Laplace noise, so $\tilde{x}_j = D_j + L_j$, $j = 1, \ldots, B$, with D_j and L_j independent.

Step 0 (moments of $\tilde{x}_j$; Equation (1) revisited). Since $\mathrm{clip}(x_i) \in [x_{\min}, x_{\max}]$ a.s., $|D_j - \mu_c| \leq w$ a.s., and $\mathrm{Var}(D_j) = \sigma_c^2/m$ with $\sigma_c^2 := \mathrm{Var}(\mathrm{clip}(X)) < \infty$; L_j has mean 0 and variance $\sigma_L^2 := 2w^2/(m^2\epsilon^2)$. Hence

$$\mathbb{E}\tilde{x}_j = \mu_c, \qquad \sigma_B^2 := \mathrm{Var}(\tilde{x}_j) = \frac{\sigma_c^2}{m} + \frac{2w^2}{m^2\epsilon^2}. \tag{8}$$

Because $w > 0$, $\epsilon > 0$, and $m < \infty$ in every row, the Laplace term ensures $\sigma_B^2 > 0$, even when $\sigma_c^2 = 0$. As $x_{\min}, x_{\max}, \epsilon$ are fixed while $m = n/B \to \infty$ (since $B = o(n)$), $\sigma_B^2 \to 0$: the row law of $\{\tilde{x}_j\}_{j=1}^B$ changes with n, so $\{\tilde{x}_j\}$ is a triangular array rather than a classical fixed-distribution sample; for each fixed n, $\tilde{x}_1, \ldots, \tilde{x}_B$ are (row-wise) i.i.d., being computed from disjoint blocks. Let $u_j := (\tilde{x}_j - \mu_c)/\sigma_B$, so $\mathbb{E}u_j = 0$, $\mathbb{E}u_j^2 = 1$.

Lemma L.1. *Fix the law of X, $\epsilon > 0$, $x_{\min} < x_{\max}$, and $\delta \in (0, 2]$. Along any sequence (n_k, B_k) satisfying $B_k \to \infty$, $m_k = n_k/B_k \to \infty$, and $m_k \in \mathbb{N}$, the standardized row variables $u_{k,j}$ satisfy $\sup_k \mathbb{E}|u_{k,1}|^{2+\delta} < \infty$; in particular, $\sup_k \mathbb{E}[u_{k,1}^4] < \infty$. The bound may depend on the fixed law of X, ϵ, the clipping bounds, and δ.*

Proof. Suppress the row index k and write $\tilde{x}_1 - \mu_c = D_1' + L_1$, where $D_1' := D_1 - \mu_c$ and $L_1 \sim \mathrm{Lap}(w/(m\epsilon))$. By the c_r-inequality,

$$\mathbb{E}|u_1|^{2+\delta} \leq 2^{1+\delta}\sigma_B^{-(2+\delta)}\{\mathbb{E}|D_1'|^{2+\delta} + \mathbb{E}|L_1|^{2+\delta}\}.$$

The Laplace moment equals $c'_\delta(w/(m\epsilon))^{2+\delta}$. Since $w/(m\epsilon) = \sigma_L/\sqrt{2} \leq \sigma_B/\sqrt{2}$, its standardized contribution is bounded uniformly over the rows.

If $\sigma_c^2 = 0$, then the clipped variable is constant almost surely, so $D_1' \equiv 0$ and the Laplace bound alone proves the result. Suppose instead that $\sigma_c^2 > 0$. The variable D_1' is a mean-zero average of m independent terms bounded in absolute value by w. Rosenthal's inequality gives

$$\mathbb{E}|D_1'|^{2+\delta} \leq C_\delta \left\{(\sigma_c^2/m)^{(2+\delta)/2} + w^{2+\delta} m^{-(1+\delta)}\right\}.$$

Because $\sigma_B^2 \geq \sigma_c^2/m$, division by $\sigma_B^{2+\delta}$ bounds the first term by a constant and the second by $Cm^{-\delta/2}$, where C may depend on the fixed clipped distribution. Thus the standardized moment is uniformly bounded for all sufficiently large rows; the finite initial set of rows is also bounded because $\sigma_B^2 > 0$ and both summands have finite moments. This proves the lemma. □

Step 1 (a triangular-array CLT for $\bar{\tilde{x}}$). By Lemma L.1, the Lyapunov ratio $B^{-\delta/2}\mathbb{E}|u_1|^{2+\delta} \to 0$ as $B \to \infty$ for any fixed $\delta \in (0, 2]$. Since $u_1, \ldots, u_B$ are row-wise i.i.d. with mean 0 and variance 1, the Lyapunov CLT for triangular arrays gives

$$\sqrt{B}\,\bar{u} = \sqrt{B}\,\frac{\bar{\tilde{x}} - \mu_c}{\sigma_B} \xrightarrow{D} N(0,1), \qquad \bar{\tilde{x}} = \frac{1}{B}\sum_{j=1}^{B} \tilde{x}_j. \tag{9}$$

Step 2 (maximal control in probability). By a union bound and Markov's inequality on the $(2+\delta)$-th moment,

$$\begin{aligned}\mathbb{P}\Big(\max_{j\leq B}|u_j| > \eta B^{1/2}\Big) &\leq B\,\mathbb{P}(|u_1| > \eta B^{1/2}) \\ &\leq \eta^{-(2+\delta)} B^{-\delta/2}\mathbb{E}|u_1|^{2+\delta} \longrightarrow 0.\end{aligned}$$

so, writing $c_B := \max_{1\leq j\leq B}|\tilde{x}_j - \mu_c| = \sigma_B \max_j |u_j|$,

$$c_B = o_p(B^{1/2}\sigma_B). \tag{10}$$

Step 3 (consistency of S). Let $S = \frac{1}{B}\sum_{j=1}^{B}(\tilde{x}_j - \mu_c)^2 = \sigma_B^2 \cdot \frac{1}{B}\sum_{j=1}^{B} u_j^2$, analogous to the (unprivatized) empirical-likelihood sample variance in Section 2, but now for the clipped mean μ_c. Since $u_1^2, \ldots, u_B^2$ are row-wise i.i.d. with $\mathbb{E}[u_1^2] = 1$ and, by Lemma L.1 with $\delta = 2$, their fourth moments are uniformly bounded across rows, Chebyshev's inequality gives $\mathrm{Var}(\frac{1}{B}\sum_j u_j^2) \leq \mathbb{E}[u_1^4]/B \to 0$, so

$$\frac{1}{B}\sum_{j=1}^{B} u_j^2 \xrightarrow{p} 1, \qquad \text{i.e.} \qquad S/\sigma_B^2 \xrightarrow{p} 1. \tag{11}$$

Lemma L.2 (Asymptotic EL feasibility)**.** *Under the conditions of Theorem 2.3,*

$$\mathbb{P}\left(\min_{1\leq j\leq B}(\tilde{x}_j - \mu_c) < 0 < \max_{1\leq j\leq B}(\tilde{x}_j - \mu_c)\right) \longrightarrow 1.$$

Proof. Write $Y_j = \tilde{x}_j - \mu_c$. If $\sigma_c^2 > 0$, the classical CLT for a block mean gives $\sqrt{m}(D_j - \mu_c)/\sigma_c \xrightarrow{D} N(0,1)$. Moreover, $\sqrt{m}L_j \sim \mathrm{Lap}(w/(\epsilon\sqrt{m})) \xrightarrow{p} 0$. Slutsky's theorem therefore gives $\sqrt{m}Y_j/\sigma_c \xrightarrow{D} N(0,1)$, so $q_B^+ := \mathbb{P}(Y_j > 0) \to 1/2$ and $q_B^- := \mathbb{P}(Y_j < 0) \to 1/2$. If $\sigma_c^2 = 0$, then $D_j = \mu_c$ almost surely and $Y_j = L_j$ is symmetric and continuous, so $q_B^+ = q_B^- = 1/2$. Row-wise independence now yields

$$\mathbb{P}\left(0 \notin \mathrm{int\,conv}\{Y_1, \ldots, Y_B\}\right) \leq (1 - q_B^+)^B + (1 - q_B^-)^B \longrightarrow 0.$$

□

Step 4 (the EL expansion, at μ_c). On the event in Lemma L.2, whose probability tends to one, the EL constraint has an interior solution and the unique admissible multiplier is well defined. Write $Y_j = \tilde{x}_j - \mu_c$, $\bar{Y} = \bar{\tilde{x}} - \mu_c$, and $y_j = \lambda Y_j$. Rewriting equation 5 at $\theta = \mu_c$ as $\frac{1}{B}\sum_j [1 - \frac{y_j}{1+y_j}]Y_j = 0$ gives

$$\lambda\tilde{S} = \bar{Y}, \qquad \tilde{S} := \frac{1}{B}\sum_{j=1}^{B}\frac{Y_j^2}{1+y_j}.$$

Since $w_j > 0$ in equation 3 forces $1 + y_j > 0$ for every j, $\lambda^2 S \leq \lambda^2\tilde{S}(1 + \max_j y_j) \leq \lambda^2\tilde{S}(1 + |\lambda|c_B) = \lambda\bar{Y}(1 + |\lambda|c_B)$, so

$$|\lambda|\big(|\lambda|S - |\lambda||\bar{Y}|c_B\big) \leq |\lambda||\bar{Y}|. \tag{12}$$

By equation 9, $\bar{Y} = O_p(\sigma_B B^{-1/2})$; combined with $S/\sigma_B^2 \to_p 1$ (Step 3) and $c_B = o_p(B^{1/2}\sigma_B)$ (Step 2), this gives

$$\frac{|\bar{Y}|c_B}{S} = O_p(\sigma_B B^{-1/2})\frac{o_p(B^{1/2}\sigma_B)}{\sigma_B^2\{1+o_p(1)\}} = o_p(1).$$

Consequently, $S - |\bar{Y}|c_B \geq S/2$ with probability tending to one. On this event, equation 12 (with the case $\lambda = 0$ immediate) implies

$$|\lambda| \leq \frac{|\bar{Y}|}{S - |\bar{Y}|c_B} \leq \frac{2|\bar{Y}|}{S} = O_p\big((\sqrt{B}\,\sigma_B)^{-1}\big),$$

and hence

$$\max_{1\leq j\leq B}|y_j| \leq |\lambda|c_B = O_p(B^{-1/2}\sigma_B^{-1})\cdot o_p(B^{1/2}\sigma_B) = o_p(1). \tag{13}$$

Expanding equation 5 to second order, $0 = \frac{1}{B}\sum_j Y_j(1 - y_j + \frac{y_j^2}{1+y_j}) = \bar{Y} - \lambda S + R$, where $R := \frac{1}{B}\sum_j Y_j y_j^2/(1+y_j)$ satisfies, using equation 13 (so $1+y_j$ is bounded away from 0 with probability tending to 1),

$$\begin{aligned}|R| &\leq \Big(\max_{j\leq B}|Y_j|\Big)\lambda^2\Big(\frac{1}{B}\sum_{j=1}^{B}Y_j^2\Big)O_p(1)\\ &= o_p(B^{1/2}\sigma_B)\,O_p(B^{-1}\sigma_B^{-2}), O_p(\sigma_B^2) = o_p(B^{-1/2}\sigma_B).\end{aligned}$$

so that

$$\lambda = S^{-1}\bar{Y} + \beta, \qquad \beta := S^{-1}R = o_p\big((\sqrt{B}\,\sigma_B)^{-1}\big).$$

Writing $\log(1+y_j) = y_j - \frac{1}{2}y_j^2 + \eta_j$ with $\mathbb{P}(|\eta_j| \leq C|y_j|^3,\ 1 \leq j \leq B) \to 1$ for some finite $C > 0$ (by equation 13), and using $\sum_j y_j = B\lambda\bar{Y}$, $\sum_j y_j^2 = B\lambda^2 S$,

$$-2\log\mathcal{R}(\mu_c) = 2\sum_{j=1}^{B}\log(1+y_j) = 2B\lambda\bar{Y} - B\lambda^2 S + 2\sum_{j=1}^{B}\eta_j.$$

Substituting $\lambda = S^{-1}\bar{Y} + \beta$ into $2B\lambda\bar{Y} - B\lambda^2 S$, the terms linear in β cancel exactly, leaving $2B\lambda\bar{Y} - B\lambda^2 S = B\bar{Y}^2/S - BS\beta^2$. Hence

$$-2\log\mathcal{R}(\mu_c) = \frac{B\bar{Y}^2}{S} - BS\beta^2 + 2\sum_{j=1}^{B}\eta_j = \Big(\frac{\sqrt{B}\,\bar{Y}}{\sigma_B}\Big)^2\frac{\sigma_B^2}{S} - BS\beta^2 + 2\sum_{j=1}^{B}\eta_j.$$

By equation 9 and Step 3, $(\sqrt{B}\bar{Y}/\sigma_B)^2(\sigma_B^2/S) \xrightarrow{D} \chi_1^2\cdot 1 = \chi_1^2$ (Slutsky's theorem). The remainder terms vanish in probability: $BS\beta^2 = B\cdot O_p(\sigma_B^2)\cdot o_p(B^{-1}\sigma_B^{-2}) = o_p(1)$ (using $\beta = o_p((\sqrt{B}\sigma_B)^{-1})$ and $S = O_p(\sigma_B^2)$), and, using equation 13, $|2\sum_j \eta_j| \leq 2C\sum_j |y_j|^3 \leq 2C\max_j|y_j|\sum_j y_j^2 = 2C\,o_p(1)\cdot O_p(1) = o_p(1)$ (since $\sum_j y_j^2 = B\lambda^2 S = O_p(1)\cdot O_p(1) = O_p(1)$, as $\lambda^2\sigma_B^2 = O_p(B^{-1})$ and $S/\sigma_B^2 = O_p(1)$). Therefore

$$-2\log\mathcal{R}(\mu_c) \xrightarrow{D} \chi_1^2,$$

on the feasibility event. Since the complement of this event has probability tending to zero by Lemma L.2, the same convergence holds for the extended statistic (defined as $+\infty$ off the convex hull), which proves Theorem 2.3.

M Proof of Theorem 2.5

We build on the notation and results of Section L, evaluated at μ_c: $Y_j = \tilde{x}_j - \mu_c$ for $j = 1,\ldots,B$, $\sigma_B^2 = \mathrm{Var}(\tilde{x}_j)$, $\bar{Y} = \bar{\tilde{x}} - \mu_c$, and $S = \frac{1}{B}\sum_{j=1}^{B}Y_j^2$, for which equation 9, equation 10, and equation 11 give $\sqrt{B}\bar{Y}/\sigma_B \xrightarrow{D} N(0,1)$, $c_B := \max_{1\leq j\leq B}|Y_j| = o_p(B^{1/2}\sigma_B)$, and $S/\sigma_B^2 \to_p 1$, respectively.

Put $N = B + 1$. The AEL pseudo-point is $Y_{B+1} = \tilde{x}_{B+1} - \mu_c = -a_B\bar{Y}$, where $a_B = \max\{1, (\log B)/2\} = O(\log B) = o(B)$ (Section 2). Although the augmented observations are dependent, their empirical first and second moments have the exact forms

$$\bar{Y}_N := \frac{1}{N}\sum_{j=1}^{N} Y_j = \frac{B - a_B}{B+1}\bar{Y}, \qquad Q_N := \frac{1}{N}\sum_{j=1}^{N} Y_j^2 = \frac{BS + a_B^2\bar{Y}^2}{B+1}. \tag{14}$$

Thus $\bar{Y}_N = O_p(\sigma_B B^{-1/2})$ and $Q_N/\sigma_B^2 \to_p 1$, because $S/\sigma_B^2 \to_p 1$, $\bar{Y}^2/\sigma_B^2 = O_p(B^{-1})$, and $a_B^2/B = O((\log B)^2/B)$. Moreover,

$$\max_{j \le B+1} |Y_j| = \max\{c_B, a_B|\bar{Y}|\} = \max\{o_p(B^{1/2}\sigma_B), O(\log B)O_p(\sigma_B B^{-1/2})\} = o_p(B^{1/2}\sigma_B), \tag{15}$$

since $a_B = o(B)$ makes the second term $o_p(B^{1/2}\sigma_B)$ as well.

Let λ_{adj} denote the unique admissible multiplier solving equation 7 at $\theta = \mu_c$, and set $y_j = \lambda_{\mathrm{adj}} Y_j$ for $j = 1, \ldots, N$. The explicit weights preceding Theorem 2.5 guarantee feasibility. If $c_N := \max_{j \le N} |Y_j|$, then equation 14–equation 15 imply

$$\frac{|\bar{Y}_N| c_N}{Q_N} = o_p(1).$$

Applying the inequality used in equation 12 to the augmented sample therefore gives $Q_N - |\bar{Y}_N| c_N \ge Q_N/2$ with probability tending to one and

$$|\lambda_{\mathrm{adj}}| \le \frac{2|\bar{Y}_N|}{Q_N} = O_p\big((\sqrt{B}\,\sigma_B)^{-1}\big).$$

Consequently,

$$\lambda_{\mathrm{adj}}\sigma_B = O_p(B^{-1/2}), \qquad \max_{1 \le j \le N} |y_j| = o_p(1), \qquad \lambda_{\mathrm{adj}} = Q_N^{-1}\bar{Y}_N + o_p\big((\sqrt{B}\sigma_B)^{-1}\big). \tag{16}$$

Indeed, the last relation follows by expanding the augmented score exactly as in Step 4: its remainder is bounded by $c_N \lambda_{\mathrm{adj}}^2 Q_N O_p(1) = o_p(\sigma_B B^{-1/2})$.

The logarithmic Taylor expansion from Step 4, now summed over N terms, yields

$$-2\log \mathcal{R}^{adj}(\mu_c) = \frac{N\bar{Y}_N^2}{Q_N} + o_p(1).$$

Here the quadratic multiplier remainder is $NQ_N o_p((\sqrt{B}\sigma_B)^{-2}) = o_p(1)$, and the cubic logarithmic remainder is bounded by $O_p(1)\max_{j \le N}|y_j| = o_p(1)$. The leading term reduces exactly to

$$\frac{N\bar{Y}_N^2}{Q_N} = \frac{B\bar{Y}^2}{S}\,\frac{(1 - a_B/B)^2}{1 + (a_B^2/B)(\bar{Y}^2/S)}. \tag{17}$$

Because $B\bar{Y}^2/S = O_p(1)$, $\bar{Y}^2/S = O_p(B^{-1})$, and $a_B = O(\log B)$, the second factor in equation 17 converges in probability to one. By equation 9, equation 11, and Slutsky's theorem,

$$-2\log \mathcal{R}^{adj}(\mu_c) \xrightarrow{D} \chi_1^2,$$

which proves Theorem 2.5. Because $\mathrm{clip}(X)$ is bounded, the χ_1^2 limits for inference on μ_c require no moment condition on the raw X. The assumption $\mathbb{E}|X| < \infty$ is needed only to define the raw mean μ and its clipping bias, while the stronger centered-moment condition $\mathbb{E}|X - a|^p < \infty$ enters only through the bias bound in Proposition 2.1(c) used in Remark 2.4.

Algorithm 1 Confidence Interval for the Private Mean via EL

Input: $\{x_i\}_{i=1}^n, x_{\min}, x_{\max}, B, \epsilon, \alpha$
Output: Confidence interval hull $(\bar{x}_{min}, \bar{x}_{max})$, or $\varnothing$
for $i = 1$ to n **do**
    $x_i = x_i \mathbb{I}(x_{\min} \le x_i \le x_{\max}) + x_{\min} \mathbb{I}(x_i < x_{\min}) + x_{\max} \mathbb{I}(x_i > x_{\max})$.
end for
Set $m = \frac{n}{B}$.
for $j = 1$ to B **do**
    Sample $z_j \sim \text{Lap}\big(\frac{x_{\max} - x_{\min}}{m\epsilon}\big)$.
    Calculate $\tilde{x}_j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} x_i + z_j$.
end for
Define $A_\alpha^{EL} = \{x \in [x_{\min}, x_{\max}] : -2 \log \mathcal{R}(x) \le \chi_1^2(\alpha)\}$, where $-2 \log \mathcal{R}(x)$ is defined in Equation (4) and $\chi_1^2(\alpha)$ is the upper α-quantile of the χ_1^2 distribution.
if $A_\alpha^{EL} = \varnothing$ **then**
    return $\varnothing$
else
    Set $\bar{x}_{min} = \inf A_\alpha^{EL}$ and $\bar{x}_{max} = \sup A_\alpha^{EL}$.
    return $(\bar{x}_{min}, \bar{x}_{max})$
end if

Algorithm 2 Confidence Interval for the Private Mean via AEL

Input: $\{x_i\}_{i=1}^n, x_{\min}, x_{\max}, B, \epsilon, \alpha$
Output: Confidence interval hull $(\bar{x}_{min}, \bar{x}_{max})$, or $\varnothing$
for $i = 1$ to n **do**
    $x_i = x_i \mathbb{I}(x_{\min} \le x_i \le x_{\max}) + x_{\min} \mathbb{I}(x_i < x_{\min}) + x_{\max} \mathbb{I}(x_i > x_{\max})$.
end for
Set $m = \frac{n}{B}$.
for $j = 1$ to B **do**
    Sample $z_j \sim \text{Lap}\big(\frac{x_{\max} - x_{\min}}{m\epsilon}\big)$.
    Calculate $\tilde{x}_j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} x_i + z_j$.
end for
Define $A_\alpha^{AEL} = \{x \in [x_{\min}, x_{\max}] : -2 \log \mathcal{R}^{adj}(x) \le \chi_1^2(\alpha)\}$, where $-2 \log \mathcal{R}^{adj}(x)$ is defined in Equation (6), including the pseudo-point $\tilde{x}_{B+1}(x)$, and $\chi_1^2(\alpha)$ is the upper α-quantile of the χ_1^2 distribution.
if $A_\alpha^{AEL} = \varnothing$ **then**
    return $\varnothing$
else
    Set $\bar{x}_{min} = \inf A_\alpha^{AEL}$ and $\bar{x}_{max} = \sup A_\alpha^{AEL}$.
    return $(\bar{x}_{min}, \bar{x}_{max})$
end if