Time-uniform and Asymptotic Confidence Sequence of Quantile under Local Differential Privacy

Leheng Cai[1†], **Qirui Hu**[2†*], **Juntao Sun**[2†], **Shuyuan Wu**[2†]
[1] Department of Statistics and Data Science
Tsinghua University, Beijing, China
[2] School of Statistics and Data Science
Shanghai University of Finance and Economics, Shanghai, China
cailh22@mails.tsinghua.edu.cn
huqirui@mail.shufe.edu.cn
sunjuntao1@stu.sufe.edu.cn
wushuyuan@mail.sufe.edu.cn

Abstract

In this paper, we develop a novel algorithm for constructing time-uniform, asymptotic confidence sequences for quantiles under local differential privacy (LDP). The procedure combines dynamically chained parallel stochastic gradient descent (P-SGD) with a randomized response mechanism, thereby guaranteeing privacy protection while simultaneously estimating the target quantile and its variance. A strong Gaussian approximation for the proposed estimator yields asymptotically anytime-valid confidence sequences whose widths obey the law of the iterated logarithm (LIL). Moreover, the method is fully online, offering high computational efficiency and requiring only $\mathcal{O}(\kappa)$ memory, where κ denotes the number of chains and is much smaller than the sample size. Rigorous mathematical proofs and extensive numerical experiments demonstrate the theoretical soundness and practical effectiveness of the algorithm.

1 Introduction

Mobile sensor traces (accelerometer, gyroscope, wireless-charger emissions) can be reverse-engineered to reveal routes, speech, and browsing habits [29, 56, 3, 36]. This shows privacy risks arise whenever fine-grained data are aggregated and mined; Differential Privacy (DP) mitigates this by adding calibrated noise so any individual's presence has negligible effect [21]. But DP assumes a trusted central curator, a model broken by the Netflix deanonymization and the March 2023 ChatGPT exposure [40, 33]. Local Differential Privacy (LDP) removes that single point of failure by randomizing data on-device and is already used in Google's RAPPOR, Apple's iOS telemetry, and Microsoft's Windows diagnostics. [19, 23, 16] As data ecosystems grow, LDP shifts analytics toward provable, user-centric privacy.

Quantile estimation and inference play a critical role in a variety of scientific and practical fields. In finance, quantiles such as value-at-risk and expected shortfall help manage portfolio risks under regulatory requirements and market volatility [9, 4]. Accurate estimation of extreme quantiles is especially important for capturing heavy-tailed financial risks [50]. In healthcare, quantile methods identify clinically significant thresholds, such as safe medication doses and treatment effectiveness [54], and guide resource allocation in treatment prioritization [55]. Reliability engineering also

*Corresponding Author
†The authors are listed in alphabetical order with equal contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

frequently employs quantile estimation to establish conservative safety standards for machinery in harsh operating conditions [17, 28]. In addition, policy evaluation also benefits from quantile approaches to capture intervention effects across diverse population groups, highlighting impacts that mean-based analyses may miss [10, 32, 14]. Unlike traditional methods focused on averages, quantile-based methods are robust when dealing with skewed or heavy-tailed real-world data, thus providing deeper insight into complex data distributions [8]. More discussion can be found in [30].

A substantial body of literature addresses quantile estimation under either CDP or LDP. Early contributions in the CDP setting include [22, 35]. More recent work, such as [47], proposes a rate-optimal sample-quantile estimator that avoids histogram evaluation, and [25] extends this line of research to the simultaneous estimation of multiple quantiles. Quantile estimation under CDP remains an active topic, with applications ranging from bounded-support data [2] to large-scale query systems [5]. In online scenarios such as continual observation [20], algorithms can compress or recompute the added noise at each time step to improve efficiency [49]. Both cases require access to raw data and apply the privacy mechanism iteratively. In the LDP setting, the curator never observes raw data but only privacy-protected reports supplied by users. This constraint makes it considerably more challenging to design algorithms that achieve accurate quantile estimation while supporting rigorous statistical inference; see, for example, [37] proposed an SGD-based estimator, [38] considered the inference for simultaneous quantiles and [1] considered a hierarchical mechanism and noisy binary search.

Inference on quantiles under an LDP constraint is challenging because it requires estimating the asymptotic variance (or other normalizing constants) of the LDP quantile estimator. Classical central limit theorem results show that the efficiency of a quantile estimator hinges on the density value at the true quantile. For SGD-based methods, however, this density is difficult to recover using only the iterates or perturbed gradients. Moreover, estimating the Hessian matrix is non-trivial, owing to the non-smoothness of the quantile loss, even if one is willing to spend additional privacy budget. Pointwise confidence intervals can be built via self-normalization or random-scaling techniques, but asymptotic sequential inference requires an almost surely consistent variance estimator; see [51]. Recently, [57] developed a high-confidence inference framework using P-SGD with identical initial values across chains, thus obtaining an i.i.d. sequence. In their theoretical results, the number of chains is fixed and cannot ensure the consistency of the variance estimator. Inspired by this smart approach, we consider a dynamically chained P-SGD whose number of chains grows with the sample size to ensure variance consistency.

We highlight our contributions as follows:

(i) We develop a novel algorithm based on the dynamically chained P-SGD for constructing time-uniform, asymptotic confidence sequences for quantiles under LDP. The procedure operates fully online, offering high computational efficiency while requiring only $\mathcal{O}(\kappa)$ memory, where κ is the number of chains and diverges to infinity at a rate much slower than T, e.g., at the order of $\log T$.

(ii) We derive an almost surely Gaussian approximation for the Polyak-Ruppert-type estimator of a quantile obtained by P-SGD. This result is non-trivial even in the non-private setting due to the non-smoothness of the loss function and its gradient. Notably, our strong Gaussian approximation is more general than those in [53] and [57], both of which address SGD with smooth loss functions. While the latter only establishes an $\mathcal{L}_2$ approximation for SGD within a fixed chain, which is not applicable to sequential inference. Our approximation rate is $o_{a.s.}\big((T/\log\log T)^{-1/2}\big)$, faster than the LIL rate, yielding asymptotically anytime-valid confidence sequences for quantiles.

(iii) We propose an almost surely consistent estimator of the quantile variance that relies only on the iterates of P-SGD and incurs no additional privacy cost. Unlike [57], which uses a fixed number of chains, we allow the number of chains to grow with the sample size T to ensure the consistency of variance estimation. As a by-product, the true density at the target quantile can also be consistently estimated under LDP. To the best of our knowledge, this is the first result on sequential inference for quantiles within the LDP framework.

The remainder of this paper is organized as follows. We first review the key concepts of DP and LDP. Next, we introduce our methodology, detailing the proposed algorithms together with the theoretical guarantees. Finally, we present experimental results that demonstrate the effectiveness of the approach. All theoretical proofs and additional simulation results are established in Appendix.

2 Methodologies

First, we introduce the mathematical definition of LDP and the asymptotic confidence sequence. Then, we will introduce the problem setting and algorithm details.

Definition 1 (Differential Privacy , see [21]) *A randomized algorithm $\mathcal{A}$, taking a dataset consisting of individuals as its input, is (ϵ, δ)-differentially private if, for any pair of datasets S and S' that differ in the record of a single individual and any event E, satisfies the condition below:*

$$\mathbb{P}[\mathcal{A}(S) \in E] \leq e^{\epsilon}\mathbb{P}\left[\mathcal{A}\left(S'\right) \in E\right] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is called ϵ-differentially private (ϵ-DP).

Definition 2 (Local Differential Privacy,see [31]) *An (ϵ, δ)-randomizer $R : X \to Y$ is an (ϵ, δ)-differentially private function taking a single data point as input.*

CDP regulates the distribution of released data rather than the curator's credibility. A trusted curator can centrally add noise, keeping algorithm design simple and accuracy loss modest [7].

LDP takes a stricter view by removing any trust assumption. The curator merely coordinates users, each holding a private value X_i. In each round it selects a user and specifies a randomized mechanism R_i. Users verify that the stated (ϵ, δ) guarantee suits the study, apply R_i to their data, and return the perturbed result. Interaction may be fully adaptive, sequential, or non-interactive; we adopt the tightest, non-adaptive model, fixing all user-randomizer pairs before data collection (Definitions 2.3 and 2.6 in [11]). Unlike CDP, where the curator adds noise, under LDP the curator must draw inference solely from user-randomized data.

From an inference perspective, the gap between a central-DP (CDP) estimator and its non-private analogue is typically $\mathcal{O}_p(n^{-1})$. Consequently, after $\sqrt{n}$ scaling, both estimators share the same asymptotic distribution, and one can estimate the associated variance from the (slightly perturbed) data by spending a modest additional privacy budget. By contrast, for locally private procedures, the error of an LDP estimator is usually $\mathcal{O}_p(n^{-1/2})$, which alters the limiting distribution and inflates the asymptotic variance. Moreover, in most practical settings, the variance cannot be consistently recovered from locally privatized data that were collected solely for point estimation.

Definition 3 (Asymptotic confidence sequences, see [51]) *Let $\mathcal{T}$ be a totally ordered infinite set (denoting time) that has a minimum value $t_0 \in \mathcal{T}$. We say that the intervals $(\widehat{\theta}_t - L_t, \widehat{\theta}_t + U_t)_{t\in\mathcal{T}}$ centered at the estimators $(\widehat{\theta}_t)_{t\in\mathcal{T}}$ with non-zero bounds $L_t, U_t > 0, \forall t \in \mathcal{T}$, form a $(1-\alpha)$-asymptotic confidence sequence (AsympCS) for a sequence of real parameters $(\theta_t)_{t\in\mathcal{T}}$ if there exists a (typically unknown) non-asymptotic $(1-\alpha)$-CS $(\widehat{\theta}_t - L_t^\star, \widehat{\theta}_t + U_t^\star)_{t\in\mathcal{T}}$ for $(\theta_t)_{t\in\mathcal{T}}$ –, i.e. satisfying*

$$\mathbb{P}\left(\forall t \in \mathcal{T}, \theta_t \in \left[\widehat{\theta}_t - L_t^\star, \widehat{\theta}_t + U_t^\star\right]\right) \geqslant 1 - \alpha,$$

and L_t, U_t become arbitrarily precise almost-sure approximations to $L_t^\star$ and $U_t^\star$:

$$L_t^\star / L_t \xrightarrow{a.s.} 1 \quad \textit{and} \quad U_t^\star / U_t \xrightarrow{a.s.} 1.$$

Compared with classical asymptotic confidence intervals, AsympCS offer several advantages and have therefore attracted considerable research attention; see, for example, [39, 27, 26]. AsympCS quantifies uncertainty uniformly over all sample sizes, rather than at a single, pre-specified size. To guarantee valid coverage across this entire time horizon, the requisite consistency must hold almost surely, rather than merely in probability, as emphasized by [51].

We formulate the problem as follows. Let $\{\xi_t\}_{t=1}^T$ be independent observations drawn sequentially from a distribution F. Our goal is to construct an AsympCS for the τ-quantile of F, denoted by x^*, i.e. $F(x^*) = \tau$, under a LDP framework.

To privatize $\{\xi_t\}_{t=1}^T$ we adopt the interactive, permutation-based binary-response mechanism of [37], which is optimal in certain regimes. Let W_t and V_t be i.i.d. Bernoulli variables, mutually independent and also independent of ξ_t, with

$$\mathbb{P}(W_t = 1) = r, \quad \mathbb{P}(W_t = 0) = 1 - r, \qquad \mathbb{P}(V_t = 1) = \mathbb{P}(V_t = 0) = 1/2.$$

For any $\zeta = (\xi, W, V)^\top$ and scalar x, define

$$G(x,\zeta) = \frac{1+r-2r\tau}{2}\Big[\mathbf{1}\{\xi \le x\}W + (1-W)(1-V)\Big] - \frac{1-r+2r\tau}{2}\Big[\mathbf{1}\{\xi > x\}W + (1-W)V\Big].$$

Given a sequence $\{x_t\}_{t=1}^T$, this yields the privatized sequence $\{G(x_t,\zeta_t)\}_{t=1}^T$, which can be viewed as a permuted stochastic gradient. The parameter r is the truthful-response rate, and [37] shows that the mechanism is ϵ-LDP with $\epsilon = \log(1+r) - \log(1-r)$.

Using the privatized gradients, we run the SGD iteration

$$x_{t+1} = x_t - \eta_t\, G(x_t, \zeta_{t+1}), \qquad t = 0, \dots, T-1.$$

Although this approach yields a consistent LDP estimator of the target quantile, estimating its asymptotic variance from $\{G(x_t,\zeta_t)\}_{i=1}^T$ alone is difficult. To address this, we employ parallel SGD (P-SGD): the data are split into κ disjoint chains, all initialized identically,

$$x_{k,t+1} = x_{k,t} - \eta_t\, G(x_{k,t}, \zeta_{k,t+1}), \qquad t = 0, \dots, T_k - 1,\ k = 1, \dots, \kappa. \tag{1}$$

When each chain has the same length, the trajectories $\{x_{k,t}\}_{t=1}^{T_k}$ are i.i.d. across k, allowing the asymptotic variance to be estimated by the sample variance across chains. Ensuring consistency, however, requires $\kappa \to \infty$. Repartitioning the data would disrupt the SGD structure and consume additional privacy budget, so we adopt a dynamically chained P-SGD in which κ grows with T.

To accommodate a time-varying number of chains, we let $\kappa = h(T)$, where $h : \mathbb{Z}_+ \to \mathbb{Z}_+$ is an increasing, piecewise-constant function. Set $K_0 := h(1)$. For each $k \in \mathbb{N}$ define $m_k := \big|\{T : h(T) = K_0 + k\}\big|$, where $|\cdot|$ denotes cardinality. We require

$$m_0 \ge K_0 \quad \text{and} \quad m_k \ \ge\ \frac{1}{K_0 + k - 1}\sum_{i=0}^{k-1} m_i, \qquad k \in \mathbb{Z}_+.$$

This condition ensures that no new chain will be added before the new chain is aligned in length with the previous ones. For example, $h(T) = \lfloor c \log_a T \rfloor + K_0$ with $a^{1/c} > \max\{K_0^{-1} + 2,\, K_0\}$ satisfies these conditions. Algorithm 1 provides the index of the chain to which each sample from 1 to T is assigned. Figure 1 provides a visual illustration.

When the T-th sample arrives, let T_k denote the number of observations held by the k-th chain. Our online quantile estimator is

$$\widehat{x}_T = \frac{1}{T}\sum_{k=1}^{\kappa}\sum_{t=1}^{T_k} x_{k,t} = \sum_{k=1}^{\kappa}\frac{T_k}{T}\Big(\frac{1}{T_k}\sum_{t=1}^{T_k} x_{k,t}\Big), \tag{2}$$

i.e., a weighted average of the chain-wise means. The asymptotic variance σ^2 of the approximating Gaussian variables Z_i is estimated by the weighted sample variance

$$\widehat{\sigma}_T^2 = \sum_{k=1}^{\kappa}\frac{T_k}{T}\left[\Big(T_k^{-1/2}\sum_{t=1}^{T_k} x_{k,t}\Big) - \sum_{l=1}^{\kappa}\frac{T_l}{T}\Big(T_l^{-1/2}\sum_{t=1}^{T_l} x_{l,t}\Big)\right]^2. \tag{3}$$

Because both the quantile estimator (2) and the corresponding variance estimator (3) are computed directly from the P-SGD iterates in (1), they each satisfy ϵ-LDP with $\epsilon = \log(1+r) - \log(1-r)$.

3 Theoretical results

To investigate the asymptotic properties, some mild assumptions are introduced.

- (A1) The density $f(\cdot)$ is continuous and $f(x^*) > 0$.
- (A2) For some constant $C_{f'} > 0$, $|f'(\cdot)|$ is uniformly bounded by $C_{f'}$.
- (A3) For some constant $a \in (1/2, 1)$, the step size $\eta_t \asymp t^{-a}$.
- (A4) As $T \to \infty$, $\kappa \to \infty$ and $\kappa \ll T^{1-1/(2a)}$.

Algorithm 1 Data allocation for parallel runs

Input T and function $h(\cdot)$.
Initialize array `nums` of length $\kappa_0 = h(0)$ with all zeros
Initialize array `result` of length T with all zeros
for $i = 1$ to T **do**
    if $h(i) > h(i-1)$ **then** Append 0 to the end of `nums`
    end if
    $k \leftarrow$ index of the first minimum in `nums`; $\texttt{result}[i] \leftarrow k$; $\texttt{nums}[k] \leftarrow \texttt{nums}[k] + 1$
end for
    Output `result`

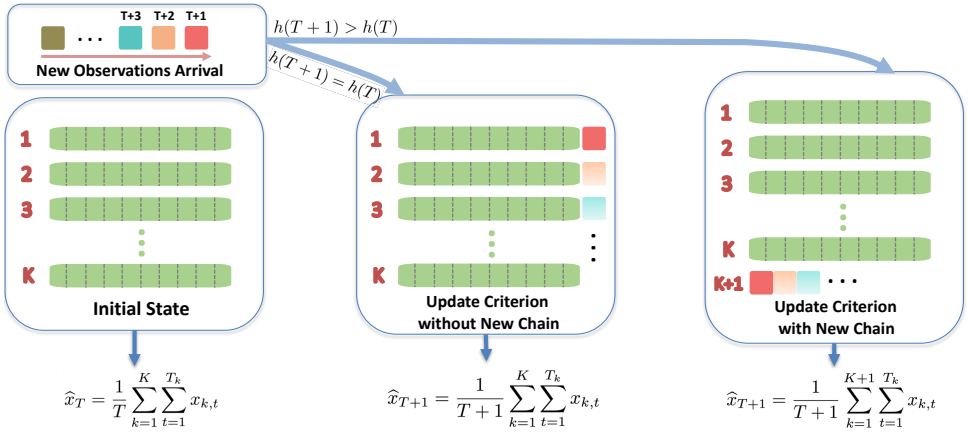

Figure 1: Overview of the Algorithm 1. (1) The left panel illustrates the initial state with T observations partitioned into K chains. (2) When new observations arrive, the algorithm determines whether to introduce a new chain. If not, new observations are sequentially added to existing chains, as shown in the middle panel. (3) If required, a new chain is created, as illustrated in the right panel, which continues receiving observations until it matches the length of existing chains. The update criterion ensures no additional chain is required before alignment.

Assumptions (A1) and (A2) are regular conditions for the distribution function. Assumption (A3) is standard in the literature; see [53]. Assumption (A4) restricts the rate at which the number of chains diverges with the sample size. The divergence rate can be arbitrarily slow, which offers great flexibility in practical implementation.

Theorem 1 *Under Assumptions (A1)-(A4), for the quantile estimator (2) there exist i.i.d. normal r.v.'s Z_i's with mean zero and variance $\sigma^2 = (1 - r^2(2\tau-1)^2)/(4r^2 f^2(x^*))$, such that*

$$\left| \widehat{x}_T - \frac{1}{T}\sum_{i=1}^{T} Z_i \right| = o_{a.s.}\left(\sqrt{\frac{\log\log T}{T}} \right).$$

Theorem 1 establishes a strong Gaussian approximation for $\widehat{x}_T - x^*$, providing an almost surely result rather than one in probability. Interestingly, for each fixed k, although $x_{t,k}$ are dependent across t, the deviation of the final weighted sum estimator $\widehat{x}_T$ from the true value x^* can be approximated by the average of T i.i.d. Gaussian random variables. Besides, the rate is significantly faster than the law of iterated logarithm bound, which is crucial for constructing asymptotic confidence sequences. Notably, [57] derived a Gaussian approximation result for a single chain, but their approximation error is measured in terms of mean squared error rather than an almost surely bound, which is not applicable to sequential inference. On the other hand, although [53] provides an almost surely Gaussian approximation, both [53] and [57] consider smooth loss and assume average Lipschitzness of the gradient, which does not hold for the quantile loss. In fact, the gradient of quantile loss only enjoys average $1/2$-Hölder smoothness, since $\left[\mathbb{E}\{\mathbf{1}(\xi \le x) - \mathbf{1}(\xi \le y)\}^2\right]^{1/2} \lesssim |x-y|^{1/2}$ for any random variable ξ with uniformly bounded density functions, which poses challenges for theoretically analyzing the approximation error.

The next theorem shows the almost surely consistency of $\widehat{\sigma}_T^2$.

Theorem 2 *Under Assumptions (A1)-(A4), for the variance estimator (3) as $T \to \infty$,*

$$\left| \widehat{\sigma}_T^2 - \frac{1 - r^2(2\tau - 1)^2}{4r^2 f^2(x^*)} \right| = \mathcal{O}_{a.s.}(1).$$

A byproduct of Theorem 2 is that it enables the estimation of the density at the true quantile x^* under the framework of differential privacy, i.e., $\sqrt{\{1 - r^2(2\tau - 1)^2/(4r^2\widehat{\sigma}_T^2)}$.

Let $[\mu_T - \gamma_{T,m}, \mu_T + \gamma_{T,m}]_{T \geq m}$ be any confidence sequence started from time $m \geq 1$ for the unknown mean of a Gaussian distribution with unit variance.

Theorem 3 *Under Assumptions (A1)-(A4), there exists some nonasymptotic $(1 - \alpha)$-confidence sequence $[\widehat{x}_T - \sigma\gamma^\star_{T,m}, \widehat{x}_T + \sigma\gamma^\star_{T,m}]$, i.e., $\mathbb{P}\left(\forall T \geq m, x^* \in \left[\widehat{x}_T - \sigma\gamma^\star_{T,m}, \widehat{x}_T + \sigma\gamma^\star_{T,m}\right]\right) \geq 1 - \alpha$, such that $(\sigma\gamma^\star_{T,m})/(\widehat{\sigma}_T\gamma_{T,m}) = \mathcal{O}_{a.s.}(1)$ as $T \to \infty$.*

With the help of the strong consistency established in Theorem 2, Theorem 3 provides a general framework for constructing AsympCSs for quantiles under the LDP setting, requiring only a confidence sequence for Gaussian random variables with unit variance. Existing confidence sequences for Gaussian variables in the literature include different types of boundaries, for example, the stitched boundary developed by [27] with a concentration rate of $\mathcal{O}\left(\sqrt{\log\log T/T}\right)$:

$$\gamma_{T,m} = 1.7\sqrt{\frac{\log\log(\max\{2T/m, e\}) + 0.72\log(10.4/\alpha)}{T}}, \tag{4}$$

or Robbins' mixture boundary ([43] and [44]), which achieves a concentration rate of $\mathcal{O}(\sqrt{\log T/T})$:

$$\gamma_{T,m} = \sqrt{\frac{\{g^{-1}(\alpha)\}^2 + \log(T/m)}{T}}, \quad \text{where } g(a) = 2\{1 - \Phi(a) + a\phi(a)\}. \tag{5}$$

Here, $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of a standard Gaussian random variable, respectively. For $m = 1$, the Gaussian mixture bound can be generalized to the following, see [51],

$$\gamma_{T,1} = \sqrt{\frac{2(T\rho^2 + 1)}{T^2\rho^2} \log\left(\frac{\sqrt{T\rho^2 + 1}}{\alpha}\right)}, \quad \forall \rho > 0. \tag{6}$$

By tuning the hyperparameter ρ in (6), one can minimize the width of the confidence interval at a specific time point given a significance level α.

We note that the Robbins' boundary is not inferior due to its slower asymptotic convergence rate. On the contrary, it is often preferable in practice because it tends to be tighter in early stages with finite samples, as also discussed in [51].

There may be some confusion regarding the burn-in strategy used in SGD-based methods versus the construction of AsympCSs starting from index m. The burn-in strategy discards a predetermined number of initial iterates to mitigate the effect of unstable early updates on the final averaged estimator, thereby reducing the effective sample size. In contrast, the coverage probability calculation starting from m retains all iterates from 1 to m and uses them to construct the AsympCSs based on equations (4)-(5). If a burn-in of b iterations is applied and coverage probabilities are reported starting from index m, then the AsympCSs should begin at iteration $b + m + 1$.

Combined with estimators (2), (3) and Theorems 3 ,we summarize the construction of the LDP $(1 - \alpha)$-AsympCS in Algorithm 2. It is worth noting that the entire procedure can be computed sequentially, storing only the most recent updates from each chain, thereby requiring approximately $\mathcal{O}(\kappa)$ memory, where $\kappa \ll T$. As a straightforward derivation, following Theorems 1 and 2, the LDP point-wise confidence interval of quantile is concluded as follows.

Corollary 1 *Under Assumptions (A1)-(A4), the asymptotically correct $(1 - \alpha)$ point-wise confidence interval of quantile x^* is $[\widehat{x}_T - \widehat{\sigma}_T z_{1-\alpha/2}/\sqrt{T}, \widehat{x}_T + \widehat{\sigma}_T z_{1-\alpha/2}/\sqrt{T}]$, i.e., $\mathbb{P}\left(x^* \in \left[\widehat{x}_T - \widehat{\sigma} z_{1-\alpha/2}/\sqrt{T}, \widehat{x}_T + \widehat{\sigma} z_{1-\alpha/2}/\sqrt{T}\right]\right) \geq 1 - \alpha$, as $T \to \infty$, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-quantile of standard normal random variables.*

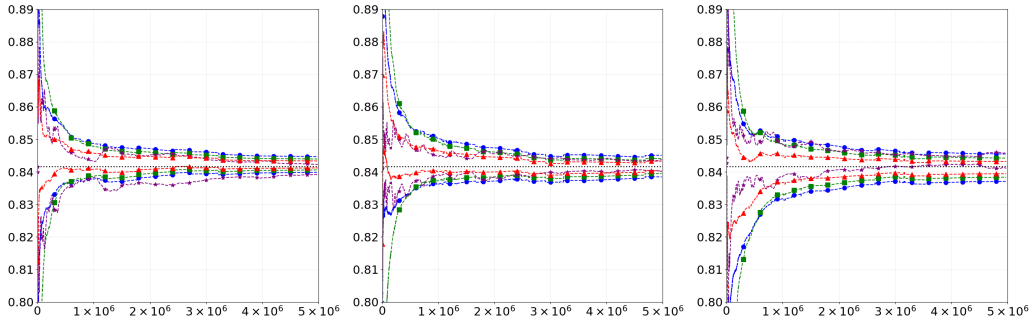

Figure 2: Plots of trajectories when confidential data come from standard normal distribution $\mathcal{N}(0,1)$ for pointwise confidence interval from Corollary 1 (in red with upward-pointing triangles), pointwise confidence interval from [37] (in purple with asterisks), proposed AsympCS based on (4) (in blue with circles) and (6) (in green with squares) with $\tau = 0.8$, $r = 1, 0.9, 0.75$ (left, middle and right panel).

It is well known that the tail of a self-normalized distribution is typically heavier than that of the normal distribution, as noted in [46]. As a result, the pointwise confidence interval constructed based on Corollary 1 is more efficient than those proposed in [37]. We further provide a visualization comparing our constructed AsympCSs, the pointwise confidence intervals, and the intervals from [37] in Figures 2 and A.7. One can observe that although the asymptotic widths of both pointwise confidence intervals are similar, our proposed intervals tend to be slightly narrower. Additionally, the AsympCS constructed using equation (6) is numerically tighter than the one based on equation (4), although the latter enjoys a faster asymptotic convergence rate.

Algorithm 2 Algorithm to construct LDP AsympCS of quantile

1: Input data: $\{\xi_t\}_{t=1}^T$, truthful response rate $r \in [0,1]$, significance level α, initial sample size m for sequential inference, initial number of chains κ, learning rate $\{\eta_t\}_{t=1}^T$, initial index $n_k = 0$ and initial values across all chains $\widetilde{x}_k = x_0$.
2: **For** $t = 1, \ldots T$,
3: Computed the current update chain $l_t = \texttt{result}[t]$ from Algorithm 1.
4: **If** $l_t > \kappa$ Set $\kappa = \kappa + 1$, $n_\kappa = 0$, $x_{\kappa, n_\kappa} = 0$.
5: **EndIf**
6: Require perturbed gradient $G(x_{l_t,n_{l_t}}, \zeta_t)$.
7: Update in l_t- chain: $x_{l_t,n_{l_t}+1} = x_{l_t,n_{l_t}} - \eta_{n_{l_t}} G(x_{l_t,n_{l_t}}, \zeta_t)$,
8: $\widetilde{x}_{l_t} = \left\{n_{l_t}\widetilde{x}_{l_t} + x_{l_t,n_{l_t}+1}\right\}/(n_{l_t}+1)$, $n_{l_t} = n_{l_t} + 1$,
9: Update of quantile estimator and corresponding variance estimator:

$$\widehat{x}_t = \sum_{k=1}^{\kappa} \frac{n_k}{t}\widetilde{x}_k, \quad \widehat{\sigma}_t^2 = \sum_{k=1}^{\kappa} \frac{n_k}{t}\left[\left(n_k^{1/2}\widetilde{x}_k\right) - \sum_{l=1}^{\kappa}\frac{n_l}{t}\left(n_l^{1/2}\widetilde{x}_l\right)\right]^2.$$

10: **End For**
11: Output the $(1-\alpha)$-AsympCS $[\widehat{x}_t - \widehat{\sigma}_t\gamma_{t,m}, \widehat{x}_t + \widehat{\sigma}_t\gamma_{t,m}]$ for $t = m, \ldots, T$, where $\gamma_{t,m}$ can be computed by (4), (5) or (6)

4 Experiments

4.1 General setting

In this section, we evaluate the finite-sample performance of the proposed method. The confidential data are generated from two distributions: standard Normal $\mathcal{N}(0,1)$ and standard Cauchy $\mathcal{C}(0,1)$. Target quantiles are set to $\tau = 0.8, 0.5, 0.3$. The truthful response rates are chosen as $r = 1, 0.9, 0.75, 0.5, 0.25$, corresponding to privacy budgets $\varepsilon = \log(1+r) - \log(1-r)$ of $+\infty, 2.94, 1.95, 1.10, 0.51$, respectively. The algorithm uses random initialization with standard

Normal $\mathcal{N}(0,1)$ of all chains and step sizes set to $\eta_{\kappa,t} = 1/t^a$ with $a = 0.6$ for all chains as well, satisfying Assumption (A3). Following [34], we incorporate a burn-in strategy into the algorithm to reduce the impact of initial parameter bias and enhance the stability of statistical inference, with the number of burn-in samples being about $(0.25/r^2)\%$ of the total sample size. Each experiment is replicated 2000 times using 110 Intel® Xeon® Platinum 8352V CPU @ 2.10GHz CPUs with $360GB$ memory and $1200GB$ storage.

4.2 Results

Our first analysis focuses on the time-uniform convergence performance. We consider the number of chains as a function of time via $h(t) = \lfloor 8\log_{10}(T/5) \rfloor$ for $t < T/5$, and $h(t) = \lfloor 8\log_{10}(t) \rfloor$ for $t \geq T/5$, where $T = 5{,}000{,}000$ denotes the total sample size. Time-uniform 95% AsympCSs are constructed using the stitched boundary in (4) and the Gaussian mixture boundary in (6) with $\rho = 0.001$. We report the time-uniform type I error rates and the average lengths of the resulting CSs. As a benchmark, we include the order-statistics-based non-private method proposed in [26].

Results for the standard normal distribution based on equations (4) and (6) are presented in Figures 3 and 4, while additional results for the standard Cauchy distribution are provided in Appendix A. These numerical results are consistent with Theorem 3. Figure 3 shows that all methods maintain the nominal type I error rate (5%) across various values of the parameters r and τ for both AsymCSs based on (4) and (6). Figure 4 indicates that the average length of the constructed AsympCSs decreases as the privacy budget increases. Moreover, From Figure 3, one observes that the AsymCSs based on (4) will be more conservative than (6) in most stages under finite sample sizes, which is also reflected on Figure 4. Therefore, the AsymCSs based on (6) enjoys the better finite sample performance in our setting, even its theoretically asymptotic rate is $\mathcal{O}(\sqrt{\log T/T})$, which is slower than $\mathcal{O}(\sqrt{\log\log T/T})$.

Notably, when $r = 1$, the non-DP AsympCSs based on P-SGD are tighter than the nonasymptotic CSs from [26], while still maintaining valid type I error control. These findings suggest that our proposed confidence sequences can provide improved efficiency for quantile inference, even in the absence of privacy constraints. A similar phenomenon is observed under the Cauchy distribution setting, as illustrated in Figures A.3 and A.4.

Next, we investigate finite-sample variance estimation $\widehat{\sigma}_T^2$. To illustrate consistency with respect to T, we set $\kappa = 20, 40, 80, 100$. Relative absolute errors (RAEs), defined as $|\widehat{\sigma}_T^2 - \sigma^2|/\sigma^2$, are summarized via boxplots in Figures A.5 and A.6 in Appendix A . The results demonstrate that RAEs consistently decrease as T increases, aligning with Theorem 2. Furthermore, for a fixed T, smaller values of r yield lower RAEs.

Finally, to further strengthen our simulation study, we conducted additional experiments, including: (1) sensitivity analysis of tuning parameters, (2) finite-sample performance under a mixture of Beta distributions, and (3) a comparison between our proposed method and [37] under specific settings. Across these settings, the results consistently demonstrate the robustness and effectiveness of our approach; see details in Appendix A.

5 Real data application

In this section, we empirically evaluate the effectiveness of our proposed method on the following two representative real datasets widely used in privacy research:

Law school dataset [52]. This dataset consists of 20,649 examples aiming to predict students' undergraduate GPA based on their personal information and academic abilities. Given that GPA reflects individual educational outcomes and is protected under strict data-use agreements [52], we treat it as sensitive educational information requiring privacy protection.

Government salary dataset [41]. This dataset originates from the 2018 American Community Survey conducted by the U.S. Census Bureau. It includes over 200,000 observations, with annual salary (USD) as the response variable. Annual salary represents typical personal financial information [25]; therefore, we treat it as sensitive data warranting privacy protection.

To facilitate analysis, we applied a logarithmic transformation for two datasets and then back-transformed the confidence sequence bounds after prediction. We apply our proposed method:

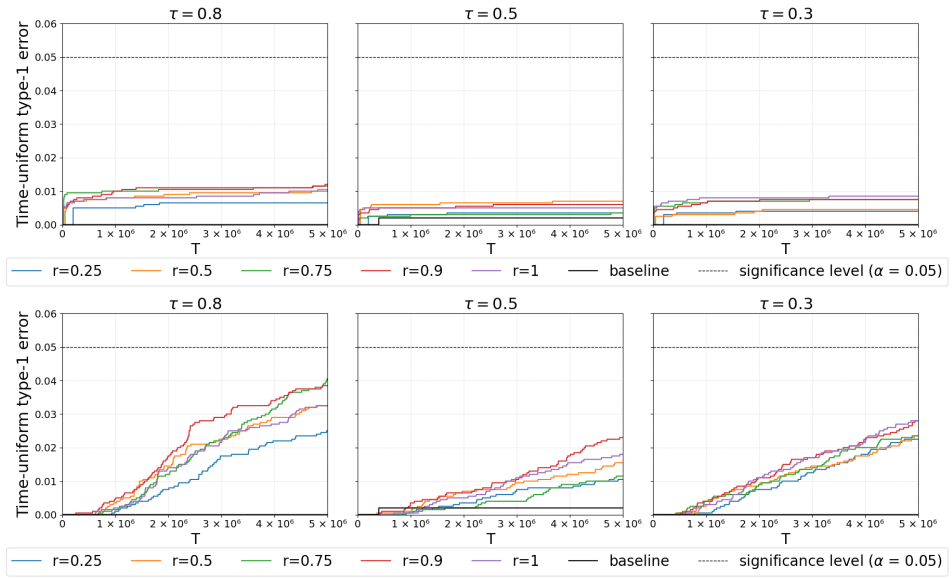

Figure 3: Time-uniform type I error for AsypmCS constructed by (4) (on top panel) and (6) (on bottom panel) and non-DP non asymptotic CS in [26],when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

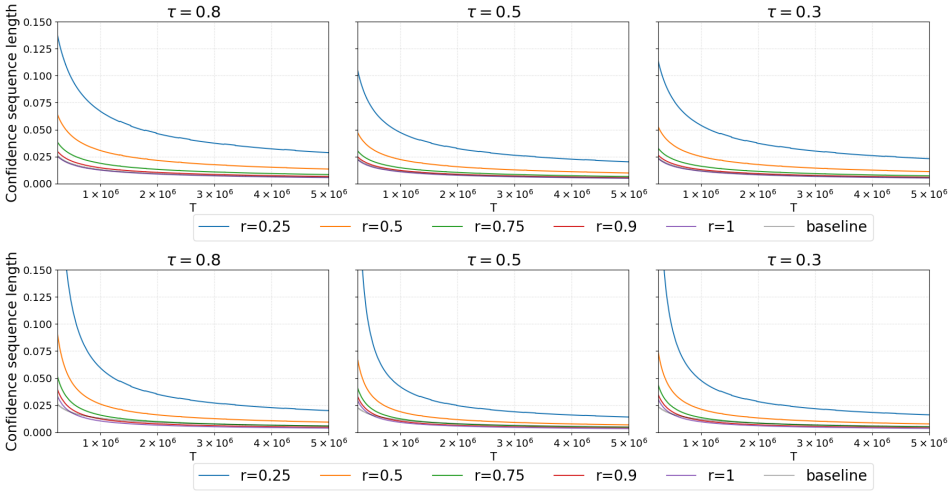

Figure 4: Average length for AsypmCS constructed by (4) (on top panel) and (6) (on bottom panel) and non-DP non asymptotic CS in [26],when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

AsympCS based on equation (4) and equation (6) to conduct privacy-preserving inference on the median ($\tau = 0.5$), targeting GPA in the first dataset and annual salary in the second. Specifically, we construct time-uniform CS for the respective quantiles under truthful response rates $r = 0.75$ and 0.9, and set the hyperparameter ρ in equation (6) to $\rho = 0.01$, while keeping all other tuning parameters consistent with those used in Section 4. The upper and lower bounds of the confidence sequences are presented in Figure 5. From the results in Figure 5, we observe that the CSs produced by our two methods under different response rates r covering similar central values. In addition, in both datasets, the length of the constructed CS decreases as r increases. As t grows, the sequence based on (4) becomes more conservative than that based on (6). These observations align with our simulation findings and further demonstrate the methods' adaptability.

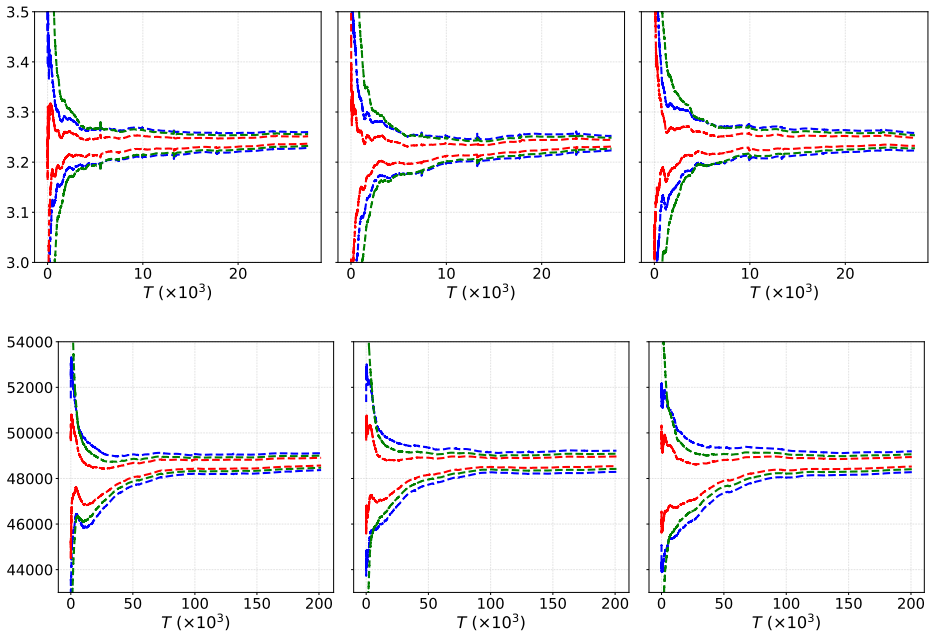

Figure 5: Confidence sequence boundaries for GPA in the Law dataset (on top panel) and annual salary in the government salary dataset (on bottom panel). The pointwise confidence interval from Corollary 1 (red), and the proposed AsympCS based on equation (4) (blue) and equation (6) (green), with target quantile $\tau = 0.5$ and truthful response rates $r = 0.9$, 0.8, and 0.75 (left, middle, and right panels, respectively).

6 Concluding remark

In this paper, we introduce an online, $\mathcal{O}(\kappa)$-memory algorithm that provides time-uniform, asymptotic confidence sequences for quantiles under LDP. We establish an almost-sure Gaussian approximation for the Polyak-Ruppert quantile estimator obtained via parallel SGD, which is non-trivial even in the non-DP case at rate $o_{a.s.}\big((T/\log\log T)^{-1/2}\big)$, thereby sharpening the $\mathcal{L}_2$ and smooth-loss based results of [57] and [53]. In addition, we devise an almost-surely consistent estimator of the quantile variance (and density) using only the SGD iterates, thus providing the first sequential quantile-inference procedure in the LDP setting.

Nonetheless, our methodology has some limitations. First, the SGD-based procedure depends on tuning parameters, such as the learning rate and the initial values, whose optimal calibration can be delicate. Second, the rate of variance consistency hinges on the number of parallel chains, κ, and the dynamically increasing-chain scheme requires relatively sharp assumptions on the relationship between κ and T. A fixed-κ variant could leverage a t-distribution with $(\kappa - 1)$ degrees of freedom to form confidence sequences, but deriving non-asymptotic t-based bounds is far from straightforward. Finally, although non-asymptotic error bounds for SGD estimators have been extensively studied (e.g., 8), extending these results to obtain fully non-asymptotic confidence sequences for SGD iterates under LDP remains an attractive yet challenging avenue for future research.

7 Acknowledgments

The authors sincerely thank the anonymous reviewers, AC, and PCs for their valuable suggestions that have greatly improved the quality of our work. Leheng Cai would thank to the funding supported by China Association for Science and Technology and National Natural Science Foundation of China (No. 12171269). Shuyuan Wu's research is partially supported by National Natural Science Foundation of China (No. 12401392).

References

[1] Anders Aamand, Fabrizio Boninsegna, Abigail Gentle, Jacob Imola, and Rasmus Pagh. Lightweight protocols for distributed private quantile estimation. *arXiv preprint arXiv:2502.02990*, 2025.

[2] Daniel Alabi, Omri Ben-Eliezer, and Anamay Chaturvedi. Bounded space differentially private quantiles. *arXiv preprint arXiv:2201.03380*, 2022.

[3] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. Spearphone: A speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers. *arXiv preprint arXiv:1907.05972*, 2019.

[4] Luca Barbaglia, Sergio Consoli, and Sebastiano Manzan. Forecasting with economic news. *Journal of Business & Economic Statistics*, 41(3):708–719, 2023.

[5] Omri Ben-Eliezer, Dan Mikulincer, and Ilias Zadik. Archimedes meets privacy: On privately estimating quantiles in high dimensions under minimal assumptions. *arXiv preprint arXiv:2208.07438*, 2022.

[6] Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

[7] 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.

[8] Likai Chen, Georg Keilbar, and Wei Biao Wu. Recursive quantile estimation: Non-asymptotic confidence bounds. *Journal of Machine Learning Research*, 24(91):1–25, 2023.

[9] Song Xi Chen. Nonparametric estimation of expected shortfall. *Journal of financial econometrics*, 6(1):87–107, 2008.

[10] Victor Chernozhukov and Iván Fernández-Val. Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies*, 78(2):559–589, 2011.

[11] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*, pages 375–403. Springer, 2019.

[12] Jeremy Cohen, Simran Kaur, Yuanzhi Li, J Zico Kolter, and Ameet Talwalkar. Gradient descent on neural networks typically occurs at the edge of stability. In *International Conference on Learning Representations*, 2021.

[13] Miklos Csörgo and Pál Révész. *Strong approximations in probability and statistics*. Academic press, 1981.

[14] David Deuber, Jinzhou Li, Sebastian Engelke, and Marloes H Maathuis. Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. *Journal of the American Statistical Association*, 119(547):2206–2216, 2024.

[15] Aymeric Dieuleveut and Francis Bach. Nonparametric stochastic approximation with large step-sizes. 2016.

[16] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

[17] Dana Draghicescu, Serge Guillas, and Wei Biao Wu. Quantile curve estimation and visualization for nonstationary time series. *Journal of Computational and Graphical Statistics*, 18(1):1–20, 2009.

[18] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[19] 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*, pages 429–438. IEEE, 2013.

[20] Cynthia Dwork. Differential privacy in new settings. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 174–183. SIAM, 2010.

[21] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006.

[22] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[23] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

[24] Sébastien Gadat and Fabien Panloup. Optimal non-asymptotic analysis of the ruppert–polyak averaging stochastic algorithm. *Stochastic Processes and their Applications*, 156:312–348, 2023.

[25] Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3713–3722. PMLR, 18–24 Jul 2021.

[26] Steven R Howard and Aaditya Ramdas. Sequential estimation of quantiles with applications to a/b testing and best-arm identification. *Bernoulli*, 28(3):1704–1728, 2022.

[27] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

[28] Jiaqiao Hu, Yijie Peng, Gongbo Zhang, and Qi Zhang. A stochastic approximation method for simulation-based quantile optimization. *INFORMS Journal on Computing*, 34(6):2889–2907, 2022.

[29] Jingyu Hua, Zhenyu Shen, and Sheng Zhong. We can track you if you take the metro: Tracking metro riders using accelerometers on smartphones. *IEEE Transactions on Information Forensics and Security*, 12(2):286–297, 2016.

[30] Qi Huang, Hanze Zhang, Jiaqing Chen, and MJJBB He. Quantile regression models and their applications: A review. *Journal of Biometrics & Biostatistics*, 8(3):1–6, 2017.

[31] Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 94–105, 2019.

[32] Nathan Kallus, Xiaojie Mao, and Masatoshi Uehara. Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. *Journal of Machine Learning Research*, 25(16):1–59, 2024.

[33] In Lee. An analysis of data breaches in the us healthcare industry: diversity, trends, and risk profiling. *Information Security Journal: A Global Perspective*, 31(3):346–358, 2022.

[34] Sokbae Lee, Yuan Liao, Myung Hwan Seo, and Youngki Shin. Fast and robust online inference with stochastic gradient descent via random scaling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7):7381–7389, June 2022.

[35] Jing Lei. Differentially private m-estimators. *Advances in Neural Information Processing Systems*, 24, 2011.

[36] Jianwei Liu, Xiang Zou, Leqi Zhao, Yusheng Tao, Sideng Hu, Jinsong Han, and Kui Ren. Privacy leakage in wireless charging. *IEEE Transactions on Dependable and Secure Computing*, 2022.

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

[38] 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*, pages 31147–31164, 2024.

[39] Tudor Manole and Aaditya Ramdas. Martingale methods for sequential estimation of convex functionals and divergences. *IEEE Transactions on Information Theory*, 69(7):4641–4658, 2023.

[40] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the Netflix prize dataset. *arXiv preprint cs/0610105*, 2006.

[41] Drago Plečko, Nicolas Bennett, and Nicolai Meinshausen. fairadapt: Causal reasoning for fair data preprocessing. *Journal of Statistical Software*, 110:1–35, 2024.

[42] Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855, 1992.

[43] Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, 1970.

[44] Herbert Robbins and David Siegmund. Boundary crossing probabilities for the wiener process and sample sums. *The Annals of Mathematical Statistics*, pages 1410–1429, 1970.

[45] Robert J Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.

[46] Xiaofeng Shao. Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817, 2015.

[47] Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

[48] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. pmlr, 2013.

[49] Jay Tenenbaum, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Concurrent shuffle differential privacy under continual observation. In *International Conference on Machine Learning*, pages 33961–33982. PMLR, 2023.

[50] Huixia Judy Wang, Deyuan Li, and Xuming He. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464, 2012.

[51] Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *The Annals of Statistics*, 52(6):2613–2640, 2024.

[52] Linda F Wightman. Lsac national longitudinal bar passage study. lsac research report series. 1998.

[53] Chuhan Xie, Kaicheng Jin, Jiadong Liang, and Zhihua Zhang. Asymptotic time-uniform inference for parameters in averaged stochastic approximation. *arXiv preprint arXiv:2410.15057*, 2024.

[54] Dandan Xu, Michael J Daniels, and Almut G Winterstein. A bayesian nonparametric approach to causal inference on quantiles. *Biometrics*, 74(3):986–996, 2018.

[55] Steve Yadlowsky, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. Evaluating treatment prioritization rules via rank-weighted average treatment effects. *Journal of the American Statistical Association*, 120(549):38–51, 2025.

[56] Li Zhang, Parth H Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. Accelword: Energy efficient hotword detection through accelerometer. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 301–315, 2015.

[57] Wanrong Zhu, Zhipeng Lou, Ziyang Wei, and Wei Biao Wu. High confidence level inference is almost free using parallel stochastic optimization. *arXiv preprint arXiv:2401.09346*, 2024.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We have accurately summarized the paper's contributions and scope in the abstract and introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussed the limitations in Conclusion section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We have provided detailed theoretical assumptions and corresponding comments in Section 3, and complete proofs in the Appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have described the experimental setting in Section 4 in detail.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We have uploaded the code that reproduces the experimental results in the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We have established these details in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: This information has been defined correctly in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have reported the computer resources in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have checked that our paper satisfies the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification:We discuss our contributions and potential future directions in the last section. No apparent negative societal impacts are foreseen.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our paper doesn't involve this.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: No existing assets were used.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: We only use the LLM for writing and editing.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Additional simulation results

A.1 Other results in Section 4

In this section, we provide additional figures not shown in the main text (see Figures A.1–A.7). Notably, in the left panel of Figure A.1, the time-uniform type I error slightly exceeds the nominal level of 0.05 (approximately 0.08), with most errors occurring during the start of the algorithm. This primarily occurs because the initial values ($\mathcal{N}(0,1)$) are relatively far from the true values, resulting in poor estimation at the initial moments. Additionally, although (4) is generally more conservative than (6) during the early stage, calculations indicate that the boundary given by (4) is narrower at the very beginning of the algorithm, causing slightly poorer coverage during these initial moments. Increasing the burn-in period could further reduce this error rate.

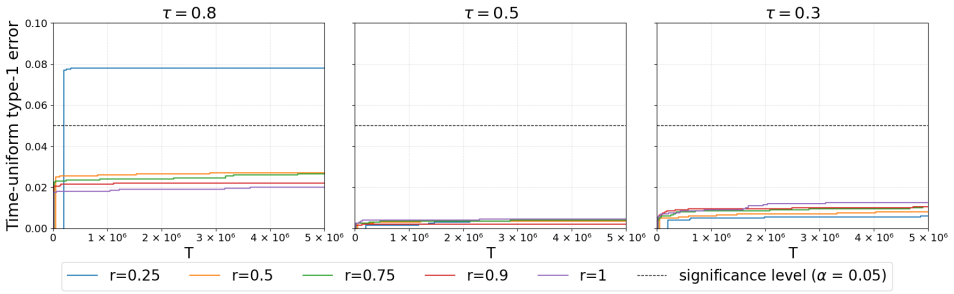

Figure A.1: Time-uniform type I error for AsypmCS constructed by (4) and non-DP non asymptotic CS in [26],when confidential data come from standard Cauchy $\mathcal{C}(0,1)$ with $r = 0.25, 0.5, 0.75, 1$ and $\tau = 0.3, 0.5, 0.8$.

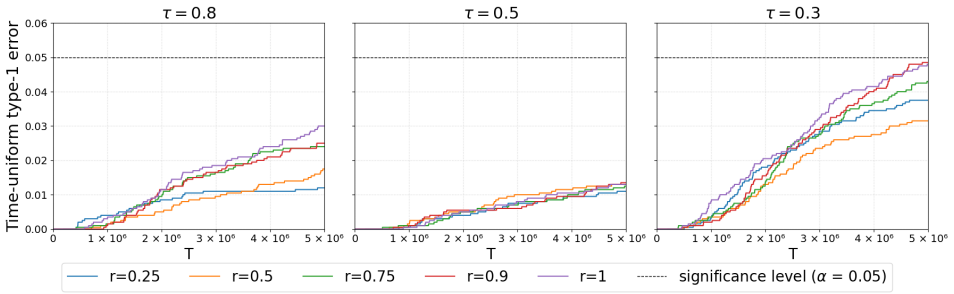

Figure A.2: Time-uniform type I error for AsypmCS constructed by (6) and non-DP non asymptotic CS in [26],when confidential data come from standard Cauchy $\mathcal{C}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

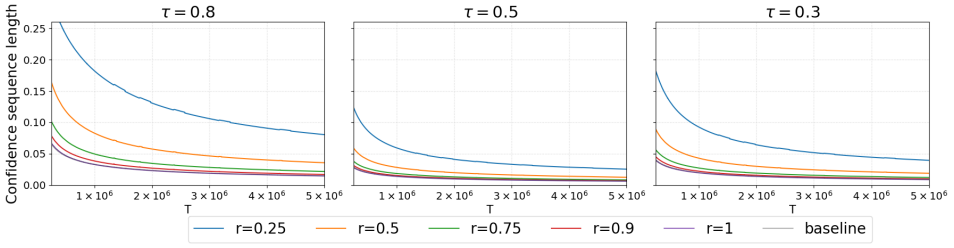

Figure A.3: Average length for AsypmCS constructed by (4) and non-DP non asymptotic CS in [26],when confidential data come from standard Cauchy $\mathcal{C}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

Next, we evaluate the finite-sample performance under a mixture of Beta distributions and make some discussions about our Assumptions. To be specific, for our Assumption (A1), according to

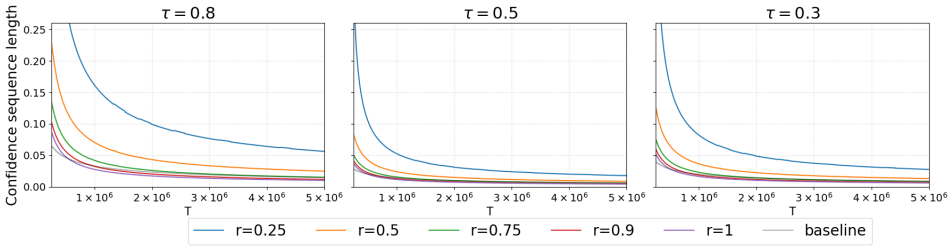

Figure A.4: Average length for AsypmCS constructed by (6) and non-DP non asymptotic CS in [26],when confidential data come from standard Cauchy $\mathcal{C}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

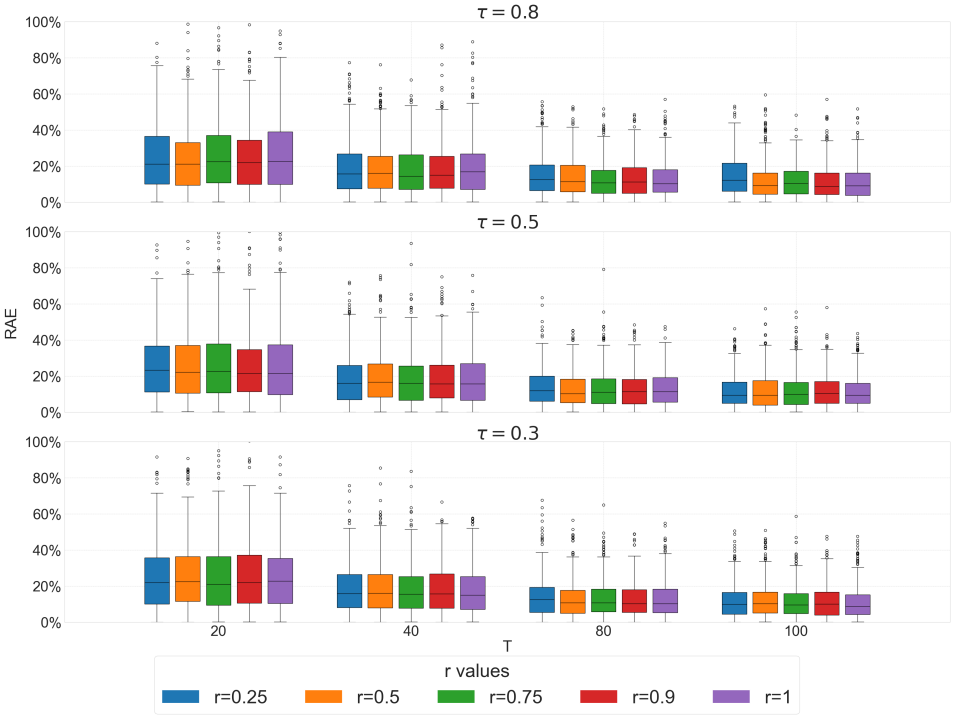

Figure A.5: Relative error of the variance estimator (3) when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

Corollary 2.3.3.A in [45], the asymptotic normality of the sample quantile relies on the assumption that the distribution function $F(\cdot)$ is differentiable at the true quantile value x^*, with a strictly positive derivative. While it's not strictly necessary that a density function $f(\cdot)$ equals the derivative of the distribution function, $f(\cdot) = F'(\cdot)$, this relationship holds if the density $f(\cdot)$ is continuous at x^*, in which case $F'(x^*) = f(x^*) > 0$. In addition, Assumption (A2) is a technical requirement crucial for controlling a negligible term in the Gaussian approximation, where a second-order Taylor expansion is applied (refer to equation 7). This is a mild assumption that holds for many common distributions, including heavy-tailed ones. For example, the derivative of the density function for a standard Cauchy random variable is:

$$f'(x) = \frac{d}{dx} f(x) = \frac{d}{dx}\left(\frac{1}{\pi(1+x^2)}\right) = -\frac{2x}{\pi(1+x^2)^2},$$

This derivative is both continuous and bounded, thereby satisfying Assumption (A2).

While Assumptions (A1) and (A2) hold for a wide range of distributions, our theoretical results require the underlying density to have continuous and bounded derivatives. This condition is not met by all irregular or "spiky" distributions. To investigate our method's practical performance under

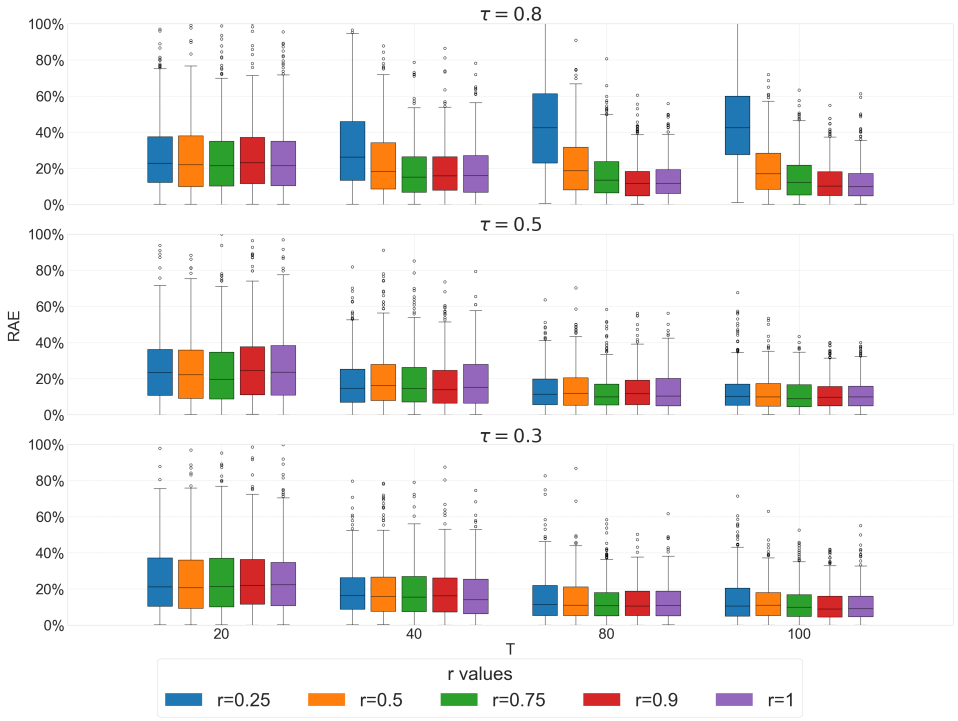

Figure A.6: Relative error of the variance estimator (3) when confidential data come from standard Cauchy $\mathcal{C}(0,1)$ with $r = 0.25, 0.5, 0.75, 0.9, 1$ and $\tau = 0.3, 0.5, 0.8$.

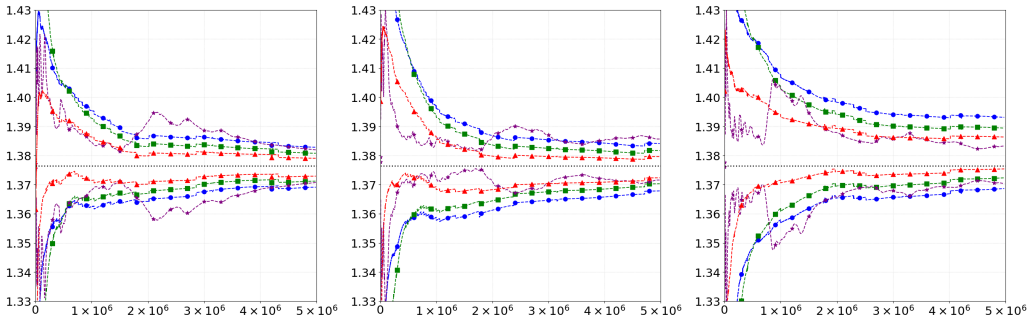

Figure A.7: Plots of trajectories when confidential data come from standard Cauchy distribution $\mathcal{C}(0,1)$, for pointwise confidence interval from Corollary 1 (in red with upward-pointing triangles), pointwise confidence interval from [37] (in purple with asterisks), proposed AsympCS based on (4) (in blue with circles) and (6) (in green with squares) with $\tau = 0.8$, $r = 1, 0.9, 0.75$ (left, middle and right panel).

these challenging conditions, we conducted further experiments. To be specific, we test our method using a mixture of Beta distributions with the following density:

$$f(x) = \{\beta_{10,100}(x) + \beta_{100,100}(x) + \beta_{100,10}(x)\}/3.$$

where $\beta_{\alpha,\beta}(x)$ is the density of a Beta distribution with parameters α and β. This specific mixture creates a sharp spike at $\tau = 0.5$, resulting in a large derivative of the density at that point, which slightly violates Assumption (A2). Despite this, our numerical simulations under $r = 0.75$ and $\tau = 0.5$ confirm that our method remains valid. The results are shown in Figure A.8.

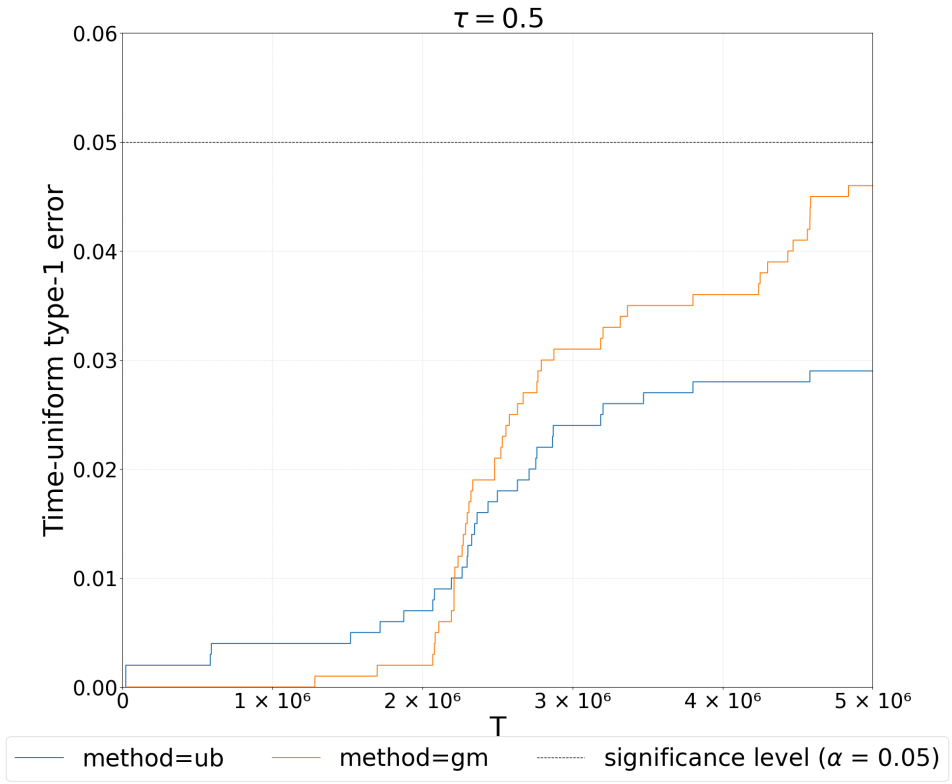

Figure A.8: Time-uniform type I error for AsypmCS constructed by by (4) and (6), when confidential data come from a mixture of Beta distributions with $\tau = 0.5$ and $r = 0.75$.

A.2 Comparison between the proposed method and [37]

We conduct additional experiments to compare our proposed quantile estimation and confidence interval in Corollary 1 with [37]. Recall that [37] adopts a pointwise estimation approach and employs self-normalization for inference. Specifically, we considered the same simulation setting as in Section 4, with total sample size $T = 5,000,000$, quantile level $\tau = 0.3, 0.5, 0.8$, the truthful response rates $r = 0.25, 0.5, 0.75, 0.9, 1$ and distribution type set to normal. The results are in figure A.9. While both methods achieve empirical coverage rates close to the nominal confidence level, the average length of our confidence interval is more narrow across various settings of r and τ, indicating higher efficiency of our approach.

For point estimation accuracy of quantiles, we use the same simulation settings as in the confidence-interval study and evaluate performance by the mean squared error (MSE) of the estimated quantiles. The detailed comparison results are provided in fig A.10. We find that when τ is close to 0.5, our method achieves comparable MSE to that in [37]. However, as τ deviates from 0.5, the MSE of our method becomes slightly worse. This can be attributed to the dynamically chained parallel procedure used procedure employed in our quantile inference.

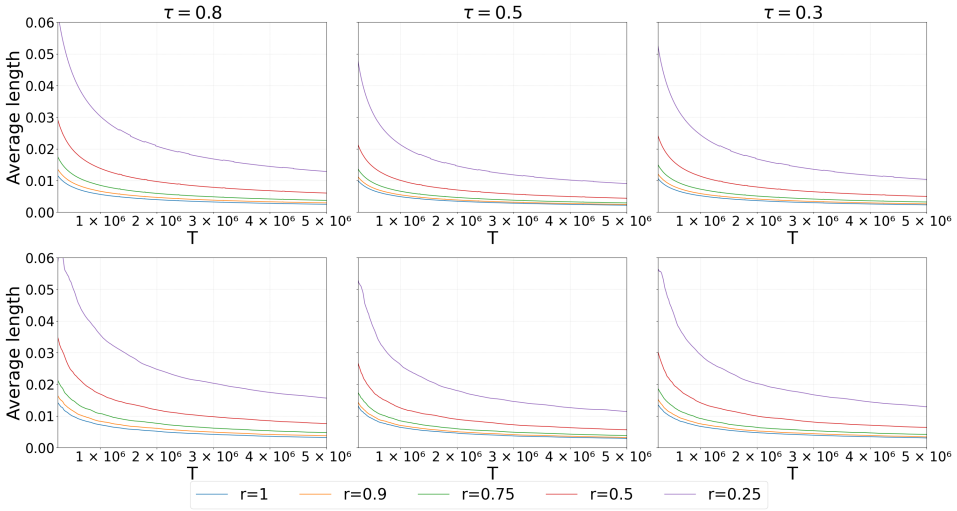

Figure A.9: Average length onstructed by the proposed method in Corollary 1 (on top panel) and [37]) (on bottom panel), when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $\tau = 0.3, 0.5, 0.8$ and $r = 0.25, 0.5, 0.75, 0.9, 1$.

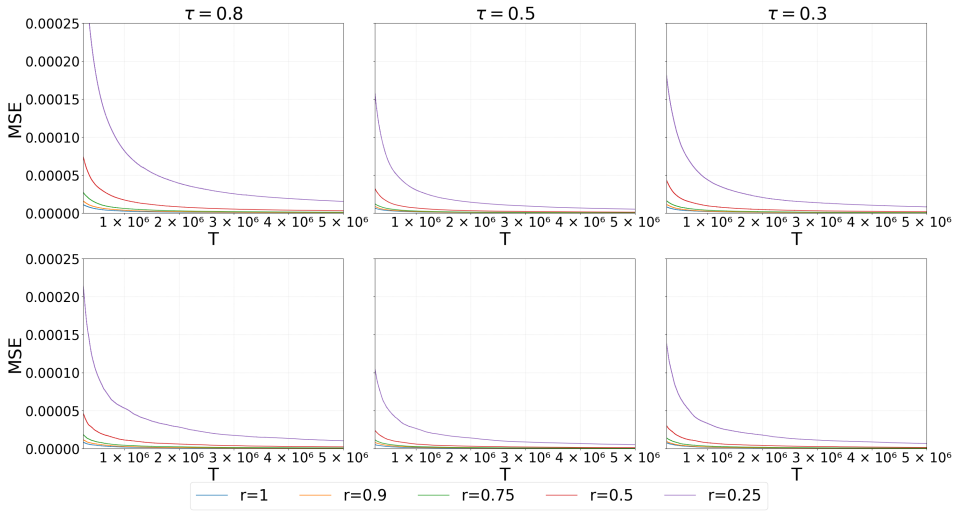

Figure A.10: MSE onstructed by the proposed method in Corollary 1 (on top panel) and [37]) (on bottom panel), when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $\tau = 0.3, 0.5, 0.8$ and $r = 0.25, 0.5, 0.75, 0.9, 1$.

A.3 The selection and sensitivity analysis of tuning parameters

The proposed method requires the selection of several tuning parameters. This subsection conducts a comprehensive sensitivity analyses to show that the results are robust to variations in these choices.

Note that our tuning parameters fall into three categories. The first includes SGD-based parameters (e.g., the learning rate–related parameter a). Selecting the learning rate is indeed a well-known challenge in practice: SGD can be sensitive to this choice, especially in high-dimensional sparse settings and in rare–frequent or heavy-tailed regimes; see [6, 12, 18]. Nevertheless, under appropriate conditions, for example, when the objective is convex and smooth, or when the initialization is sufficiently close to the true parameter, Polyak–Ruppert averaged SGD enjoys provable convergence with tolerable sensitivity to the learning rate [42, 48, 15]. As later reported in our sensitivity studies, our results are robust to reasonable variations in this hyperparameter. The second category includes tuning parameters related to time-uniform inference, such as the AsympCS starting index m and the hyperparameter ρ in the Gaussian mixture bound (equation (6)). The third category consists of tuning parameters specific to our proposed method, such as the number of chains $h(t)$. We find that the results are not sensitive to these parameters; thus, recommendations from the time-uniform inference literature [53] and the default setting provided in our paper (e.g., $h(t) = \lfloor 8 \log_{10}(t) \rfloor$) can serve as practical choices.

We next conduct comprehensive sensitivity analyses for the aforementioned tuning parameters (i.e., a, m, ρ, and $h(t)$). Specifically, we consider one of the simulation settings from Section 4, with a total sample size of $T = 5{,}000{,}000$, $1,000$ repetitions, truthful response rate $r = 0.75$, quantile level $\tau = 0.5$, and normally distributed data. We evaluate the time-uniform type I error for AsympCS across a range of hyperparameter choices. The results are summarized in the following Figures A.11 to A.14. The proposed methods maintain the nominal type I error rate (5%) for nearly all hyperparameter choices, demonstrating its insensitivity to these tuning parameters.

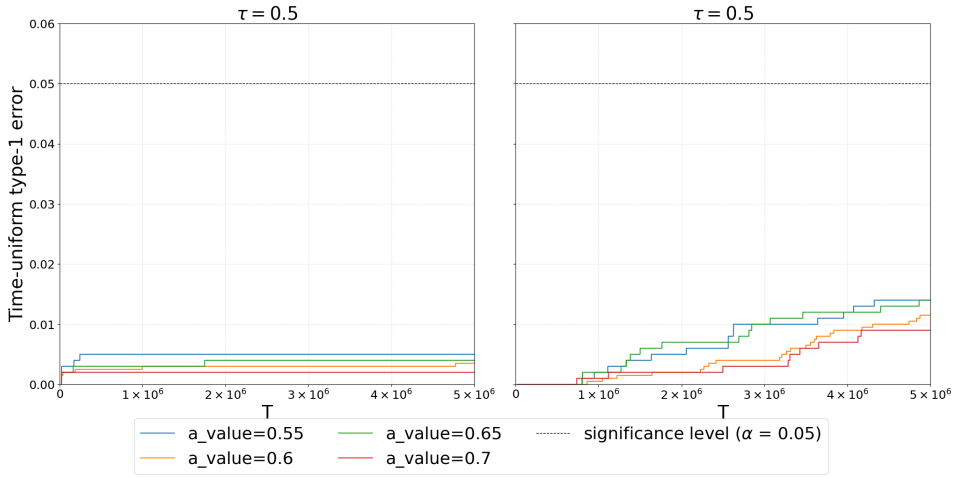

Figure A.11: Time-uniform type I error for AsypmCS constructed by (4) (on left panel) and (6) (on right panel),when confidential data come from standard Normal $\mathcal{N}(0, 1)$ with $a = 0.55, 0.6, 0.65, 0.7$.

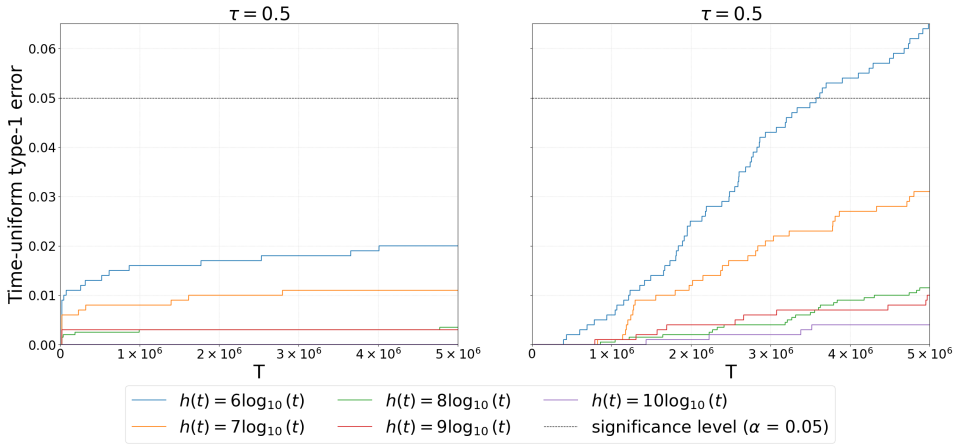

Figure A.12: Time-uniform type I error for AsypmCS constructed by (4) (on left panel) and (6) (on right panel),when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $h(t) = 6\log_{10}(t), 7\log_{10}(t), 8\log_{10}(t), 9\log_{10}(t), 10\log_{10}(t)$.

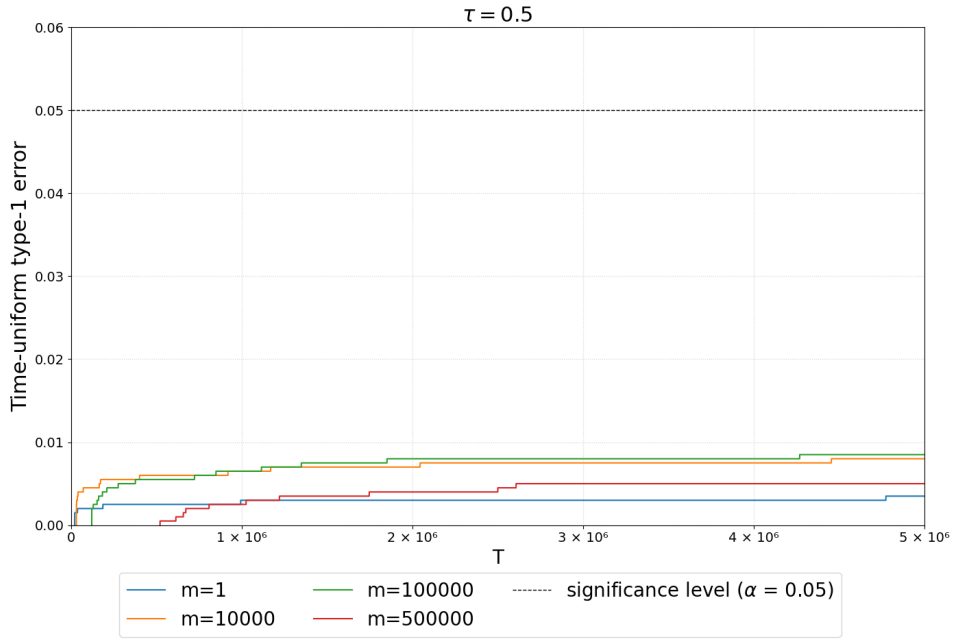

Figure A.13: Time-uniform type I error for AsypmCS constructed by (4), when confidential data come from standard Normal $\mathcal{N}(0,1)$ with $m = 1, 10000, 100000, 500000$.

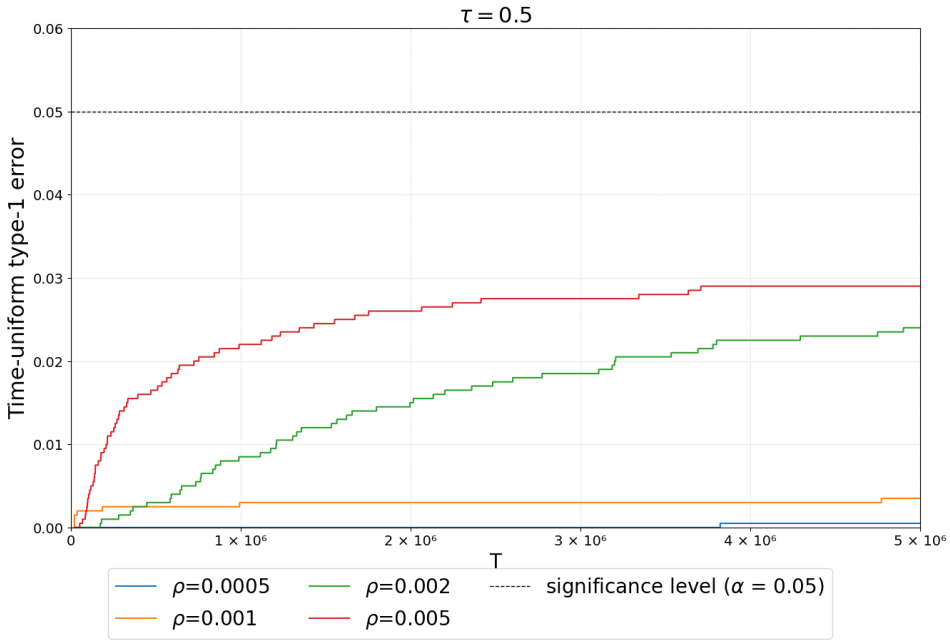

Figure A.14: Time-uniform type I error for AsypmCS constructed by (6), when confidential data come from standard Normal $\mathcal{N}(0, 1)$ with $0.0005, 0.001, 0.002, 0.005$.

B Proofs

This section includes detailed proofs of the theoretical results in the main article.

Elementary calculation shows that

$$g(x) := \mathbb{E}G(x,\zeta) = r\left\{F(x) - F(x^*)\right\},$$

which will be frequently used in our proofs.

Proofs of Theorem 1: Define the weight $\omega_k = T_k/T$. One rewrites

$$\frac{1}{T}\sum_{k=1}^{\kappa}\sum_{j=1}^{T_k}(x_{k,j}-x^*) = \sum_{k=1}^{\kappa}\frac{T_k}{T}\frac{1}{T_k}\sum_{j=1}^{T_k}(x_{k,j}-x^*) =: \sum_{k=1}^{\kappa}\omega_k\mathcal{T}_k.$$

There are two possible cases for the value of T_k: either $T_k \in (T/\kappa - 1, T/\kappa + 1)$ for all $1 \le k \le \kappa$ (case 1), or $T_1 = T_2 = \cdots = T_{\kappa_0} \ge T_{\kappa_0+1} = \cdots = T_{\kappa-1} \ge T_\kappa$ and $|T_{\kappa_0} - T_{\kappa_0+1}| \le 1$, $T_k \asymp T/\kappa$ for any $1 \le k \le \kappa - 1$ (case 2). Define

$$\varepsilon_{k,t} = g(x_{k,t-1}) - G(x_{k,t-1},\zeta_{k,t}),\ \ \widetilde{\varepsilon}_{k,t} = g(x^*) - G(x^*,\zeta_{k,t}).$$

Elementary calculation shows that

$$\begin{aligned}\mathbb{E}(\varepsilon_{k,t}^2|\mathcal{F}_{k,t-1}) &= \frac{1+r-2rF(x_{k,t-1})}{2} - \left(\frac{1+r-2rF(x_{k,t-1})}{2}\right)^2\\ &= \frac{1-r^2(2F(x_{k,t-1})-1)^2}{4}\\ &\xrightarrow{\mathbb{P}} \frac{1-r^2(2\tau-1)^2}{4},\end{aligned}$$

where the convergence in probability holds by the consistency of the quantile estimation and the continuous mapping theorem. Denote $\gamma_{k,t} = x_{k,t} - x^*$, $H = rf(x^*)$, $B_t = 1 - \eta_t H$, $A_j^t = \sum_{s=j}^t \left(\prod_{i=j+1}^s B_i\right)\eta_i$ for any $j \le t$. We decompose that

$$\begin{aligned}\mathcal{T}_k &= \frac{1}{T_k}\sum_{j=1}^{T_k}(x_{k,j}-x^*) = \frac{1}{T_k}\sum_{j=1}^{T_k}\gamma_{k,j}\\ &= \frac{1}{T_k}A_0^{T_k-1}B_0\gamma_{k,0} + \frac{1}{T_k}\sum_{j=0}^{T_k-1}A_j^{T_k-1}r_{k,j} + \frac{1}{T_k}\sum_{j=0}^{T_k-1}\left(A_j^{T_k-1} - H^{-1}\right)\varepsilon_{k,j+1}\\ &+ \frac{1}{T_k}\sum_{j=0}^{T_k-1}H^{-1}\left(\varepsilon_{k,j+1} - \widetilde{\varepsilon}_{k,j+1}\right) + \frac{1}{T_k}\sum_{j=0}^{T_k-1}H^{-1}\widetilde{\varepsilon}_{k,j+1}\\ &=: \mathcal{T}_{k,1} + \mathcal{T}_{k,2} + \mathcal{T}_{k,3} + \mathcal{T}_{k,4} + \mathcal{T}_{k,5},\end{aligned}$$

in which

$$r_{k,j} = H(x_{k,j} - x^*) - g(x_{k,j}).$$

For $\mathcal{T}_{k,1}$: According to Lemma C.4 of [53], one has $\left|A_0^{t-1}\right| \le C_0$ uniformly for all $t \ge 1$. Further observe that $\gamma_{k,0} \equiv x_0 - x^*$ for all $1 \le k \le \kappa$, thus one obtains

$$\left|\sum_{k=1}^{\kappa}\omega_k\mathcal{T}_{k,1}\right| = \mathcal{O}\left(\kappa T^{-1}\right).$$

For $\mathcal{T}_{k,2}$: Theorem 5 of [24] shows that

$$\max_{1\le k\le\kappa}\mathbb{E}\left|x_{k,t} - x^*\right|^2 \lesssim \eta_t.$$

Since $|r_{k,t}| \lesssim |x_{k,t} - x^*|^2$ by Assumption (A2), we show that

$$\begin{aligned}\sum_{t=2}^{\infty} \frac{\mathbb{E}\sum_{k=1}^{\kappa} \omega_k |r_{k,t}|}{t^{1-a}\log^{1+\epsilon}(t)} &\lesssim \sum_{t=2}^{\infty} \frac{\sum_{k=1}^{\kappa} \omega_k \mathbb{E}\left|x_{k,t} - x^*\right|^2}{t^{1-a}\log^{1+\epsilon}(t)} \\ &\lesssim \sum_{t=2}^{\infty} \frac{\max_{1\le k\le \kappa} \mathbb{E}\left|x_{k,t} - x^*\right|^2}{t^{1-a}\log^{1+\epsilon}(t)} \\ &\lesssim \sum_{t=2}^{\infty} \frac{1}{t\log^{1+\epsilon}(t)} < \infty.\end{aligned} \tag{7}$$

Hence, with probability one,

$$\sum_{t=2}^{\infty} \frac{\sum_{k=1}^{\kappa} \omega_k |r_{k,t}|}{t^{1-a}\log^{1+\epsilon}(t)} < \infty.$$

According to the uniform boundedness of $\left|A_j^{t-1}\right|$, for case 1, one shows that

$$\begin{aligned}&\left|\sum_{k=1}^{\kappa} \omega_k \mathcal{T}_{k,2}\right| \lesssim \sum_{k=1}^{\kappa} \omega_k \frac{1}{T_k} \sum_{j=0}^{T_k-1} |r_{k,j}| \\ &\le \frac{\lfloor T/\kappa + 1\rfloor}{\lceil T/\kappa - 1\rceil} \frac{1}{\lfloor T/\kappa + 1\rfloor} \sum_{j=0}^{\lfloor T/\kappa+1\rfloor - 1} \sum_{k=1}^{\kappa} \omega_k \left|r_{k,j}\right| \\ &= o_{a.s.}\left((\kappa/T)^a \log^{1+\epsilon}(T)\right).\end{aligned}$$

For case 2, one shows that

$$\begin{aligned}&\left|\sum_{k=1}^{\kappa} \omega_k \mathcal{T}_{k,2}\right| \lesssim \sum_{k=1}^{\kappa} \omega_k \frac{1}{T_k} \sum_{j=0}^{T_k-1} |r_{k,j}| \\ &\le \sum_{k=1}^{\kappa_0} \omega_k \frac{1}{T_k} \sum_{j=0}^{T_k-1} |r_{k,j}| + \sum_{k=\kappa_0+1}^{\kappa-1} \omega_k \frac{1}{T_k} \sum_{j=0}^{T_k-1} |r_{k,j}| + \frac{1}{T} \sum_{j=0}^{T_\kappa-1} |r_{\kappa,j}| \\ &= o_{a.s.}\left((\kappa/T)^a \log^{1+\epsilon}(T)\right).\end{aligned}$$

For $\mathcal{T}_{k,3}$: For any fixed $p > 0$ (large enough), note that $\max_{1\le k\le\kappa} \mathbb{E}|\varepsilon_{k,j}|^{2p} = \mathbb{E}|\varepsilon_{1,j}|^{2p}$ is bounded. Following the arguments in [53], one has $\|\mathcal{T}_{3,k}\|_{2p} = \mathcal{O}\left(T_k^{-1+a/2}\right)$. Then, using the Lemma A in Chapter 9.2.6 of [45] and the independence over k, one has that

$$\left\|\sum_{k=1}^{\kappa} \omega_k \mathcal{T}_{k,3}\right\|_{2p} = \mathcal{O}\left(\kappa^{-1/2}(T/\kappa)^{-1+a/2}\right),$$

which implies that

$$\left|\sum_{k=1}^{\kappa} \omega_k \mathcal{T}_{k,3}\right| = \mathcal{O}_{a.s.}\left(\kappa^{-1/2}(T/\kappa)^{-1+a/2} T^{1/(2p)} \log^{1/(2p)+\epsilon}(T)\right).$$

For $\mathcal{T}_{k,4}$: Observe that $\varepsilon_{k,j} - \widetilde{\varepsilon}_{k,j} = g(x_{k,j-1}) - G(x_{k,j-1}, \zeta_{k,j}) + G(x^*, \zeta_{k,j})$, and

$$\begin{aligned}\mathbb{E}\left(\varepsilon_{k,j} - \widetilde{\varepsilon}_{k,j}\right)^2 &\lesssim \mathbb{E}g^2(x_{k,j-1}) + \mathbb{E}\left\{G(x_{k,j-1}, \zeta_{k,j}) - G(x^*, \zeta_{k,j})\right\}^2 \\ &\lesssim \mathbb{E}|x_{k,j-1} - x^*|^2 + \mathbb{E}|x_{k,j-1} - x^*| \\ &\lesssim \mathbb{E}|x_{k,j-1} - x^*|^2 + \left\{\mathbb{E}|x_{k,j-1} - x^*|^2\right\}^{1/2} \lesssim \eta_t^{1/2},\end{aligned}$$

where the last inequality holds by Theorem 5 of [24], and the constant does not depend on k.

We rewrite

$$\sum_{k=1}^{\kappa}\omega_k\mathcal{T}_{k,4}=\frac{t_T}{T}\frac{1}{t_T}\sum_{t=1}^{t_T}\sum_{k=1}^{\kappa_t}H^{-1}\left(\varepsilon_{k,t}-\widetilde{\varepsilon}_{k,t}\right),$$

where $t_T=\max_{1\leq k\leq\kappa}T_k\asymp T/\kappa$ and $\kappa_t=|\{k:T_k\geq t\}|\leq\kappa$. Notice that

$$\text{Var}\left\{\sum_{k=1}^{\kappa_t}H^{-1}\left(\varepsilon_{k,t}-\widetilde{\varepsilon}_{k,t}\right)\right\}\lesssim\kappa_t\text{Var}\left(\varepsilon_{k,t}-\widetilde{\varepsilon}_{k,t}\right)\lesssim\kappa\eta_t^{1/2}.$$

Hence,

$$\sum_{t=2}^{\infty}\text{Var}\left(\frac{\sum_{k=1}^{\kappa_t}H^{-1}\left(\varepsilon_{k,t}-\widetilde{\varepsilon}_{k,t}\right)}{\kappa^{1/2}t^{1/2-a/4}\log^{1/2+\epsilon}(t)}\right)\lesssim\sum_{t=2}^{\infty}\frac{1}{t\log^{1+2\epsilon}(t)}<\infty,$$

which implies that (by Kronecker's lemma),

$$\left|\sum_{k=1}^{\kappa}\omega_k\mathcal{T}_{k,4}\right|=o_{a.s.}\left(\kappa^{-1/2}(T/\kappa)^{-1/2-a/4}\log^{1/2+\epsilon}(T)\right).$$

For $\mathcal{T}_{k,5}$: Elementary calculation shows that

$$\mathbb{E}\widetilde{\varepsilon}_{k,j}^2=\frac{1-r^2(2\tau-1)^2}{4}=:S.$$

Applying Theorem 2.6.7 of [13] with $H(x)=x^{2p}$ and $x_n=n^{\beta_0}$, there exist i.i.d. standard normal $\widetilde{Z}_{k,j}$'s and some $a,C>0$ (depending on the distribution of $H^{-1}\widetilde{\varepsilon}_{k,j}$) such that

$$\mathbb{P}\left(\left|\sum_{k=1}^{\kappa}\sum_{t=1}^{T_k}\frac{H^{-1}\widetilde{\varepsilon}_{k,t}}{\sqrt{H^{-1}SH^{-1}}}-\sum_{i=1}^{T}\widetilde{Z}_i\right|>T^{\beta_0}\right)\leq Ca^{-2p}T^{1-2p\beta_0}.$$

Thus,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{k=1}^{\kappa}\sum_{t=1}^{T_k}\frac{H^{-1}\widetilde{\varepsilon}_{k,t}}{\sqrt{H^{-1}SH^{-1}}}-\frac{1}{T}\sum_{i=1}^{T}\widetilde{Z}_i\right|>T^{-1+\beta_0}\right)\lesssim T^{1-2p\beta_0}.$$

For $p>2$, one selects $\beta_0\in(1/p,1/2)$, the Borel-Cantelli lemma leads to

$$\left|\sum_{k=1}^{\kappa}\omega_k\mathcal{T}_{k,5}-\frac{1}{T}\sum_{i=1}^{T}Z_i\right|=o_{a.s.}\left(T^{-1+\beta_0}\right),$$

where Z_i's are i.i.d. normal r.v.'s with mean zero and covariance $H^{-1}SH^{-1}$.

Therefore, we obtain that

$$\left|\frac{1}{T}\sum_{k=1}^{\kappa}\sum_{t=1}^{T_k}(x_{k,t}-x^*)-\frac{1}{T}\sum_{i=1}^{T}Z_i\right|=o_{a.s.}\left(\sqrt{\frac{\log\log T}{T}}\right),$$

which completes the proof.

Proofs of Theorem 2: Recall that the weight $\omega_k=T_k/T$. We rewrite

$$\widehat{\sigma}_T^2=\sum_{k=1}^{\kappa}\omega_k\left\{\frac{1}{\sqrt{T_k}}\sum_{j=1}^{T_k}(x_{k,j}-x^*)\right\}^2-\left\{\sum_{k=1}^{\kappa}\omega_k\frac{1}{\sqrt{T_k}}\sum_{j=1}^{T_k}(x_{k,j}-x^*)\right\}^2.$$

Recall the definitions of $\mathcal{T}_{k,j}$, $1\leq j\leq 5$.

For $\mathcal{T}_{k,1}$: According to Lemma C.4 of [53], one has $\left|A_0^{t-1}\right|\leq C_0$ uniformly for all $t\geq 1$. Further observe that $\gamma_{k,0}\equiv x_0-x^*$ for all $1\leq k\leq\kappa$, thus one obtains $\|\mathcal{T}_{k,1}\|_2=\mathcal{O}\left(T_k^{-1}\right)$, where the constant does not depend on k.

For $\mathcal{T}_{k,2}$: Consider that

$$\left|\frac{1}{T_k}\sum_{j=0}^{T_k-1} A_j^{T_k-1} r_{k,j}\right| \lesssim \frac{1}{T_k}\sum_{j=0}^{T_k-1}|r_{k,j}| \lesssim \frac{1}{T_k}\sum_{j=0}^{T_k-1}|x_{k,j}-x^*|^2.$$

Then,

$$\left\|\frac{1}{T_k}\sum_{j=0}^{T_k-1} A_j^{T_k-1} r_{k,j}\right\|_2 \lesssim \frac{1}{T_k}\sum_{j=0}^{T_k-1}\|x_{k,j}-x^*\|_4^2.$$

Applying Theorem 5 of [24], we have

$$\max_{1\le k\le\kappa}\mathbb{E}\,|x_{k,j}-x^*|^4 \lesssim \eta_j^2.$$

Since $\eta_j \asymp j^{-a}$ with $a > 1/2$, it follows that $\|\mathcal{T}_{k,2}\|_2 = o\left(T_k^{-1/2}\right)$.

For $\mathcal{T}_{k,3}$: As shown in the proof of Theorem 1, one has $\|\mathcal{T}_{k,3}\|_{2p} = \mathcal{O}\left(T_k^{-1+a/2}\right) = o\left(T_k^{-1/2}\right)$, since $a < 1$.

For $\mathcal{T}_{k,4}$: Observe that $\sum_{j=1}^{T_k} H^{-1}\left(\varepsilon_{k,j}-\widetilde{\varepsilon}_{k,j}\right)$ is a martingale for each k (independent over $1 \le k \le \kappa$), Burkholder's inequality entails that

$$\begin{aligned}\left\|\sum_{j=1}^{T_k} H^{-1}\left(\varepsilon_{k,j}-\widetilde{\varepsilon}_{k,j}\right)\right\|_2 &\lesssim \left\{\sum_{j=1}^{T_k}\left\|H^{-1}\left(\varepsilon_{k,j}-\widetilde{\varepsilon}_{k,j}\right)\right\|_2^2\right\}^{1/2} \\ &\lesssim \left(\sum_{j=1}^{T_k}\eta_j^{1/2}\right)^{1/2} \lesssim T_k^{\{1-a/2\}/2}.\end{aligned}$$

Hence, for any $1 \le k \le \kappa$,

$$\|\mathcal{T}_{k,4}\|_2 = \left\|\frac{1}{T_k}\sum_{j=1}^{T_k} H^{-1}\left(\varepsilon_{k,j}-\widetilde{\varepsilon}_{k,j}\right)\right\|_2 = \mathcal{O}\left(T_k^{-1/2-a/4}\right).$$

For $\mathcal{T}_{k,5}$: Applying Theorem 2.6.7 of [13] with $H(x) = x^{2p}$ and $x_n = vn^{\beta_0}$, there exist i.i.d. standard normal $\widetilde{Z}_{k,j}$'s and some $a_k, C_k > 0$ (depending on the distribution of $H^{-1}\widetilde{\varepsilon}_{k,j}$) such that

$$\mathbb{P}\left(\left|\sum_{j=1}^{T_k}\frac{H^{-1}\widetilde{\varepsilon}_{k,j}}{\sqrt{H^{-1}SH^{-1}}} - \sum_{j=1}^{T_k}\widetilde{Z}_{k,j}\right| > vT_k^{\beta_0}\right) \le C_k a_k^{-2p} v^{-2p} T_k^{1-2p\beta_0}.$$

Thus,

$$\mathbb{P}\left(\left|\frac{1}{T_k}\sum_{j=1}^{T_k}\frac{H^{-1}\widetilde{\varepsilon}_{k,j}}{\sqrt{H^{-1}SH^{-1}}} - \frac{1}{T_k}\sum_{j=1}^{T_k}\widetilde{Z}_{k,j}\right| > vT_k^{-1+\beta_0}\right) \lesssim v^{-2p}T_k^{1-2p\beta_0}.$$

Since $\mathbb{E}X^p = \int_0^\infty pv^{p-1}\mathbb{P}(|X| > v)dv$, we also have

$$\left\|\mathcal{T}_{k,5} - \frac{1}{T_k}\sum_{j=1}^{T_k} Z_{k,j}\right\|_2 = \mathcal{O}\left(T_k^{-1+\beta_0}\right).$$

According to the above results, we show that

$$\left\|\frac{1}{T_k}\sum_{j=1}^{T_k}\left(x_{k,j}-x^*\right) - \frac{1}{T_k}\sum_{j=1}^{T_k} Z_{k,j}\right\|_2 = o\left(T_k^{-1/2}\right),$$

which implies

$$\left\| \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\|_2 = \left\| \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} Z_{k,j} \right\|_2 + \mathcal{O}(1) < \infty,$$

$$\mathbb{E} \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) = \mathbb{E} \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} Z_{k,j} + \mathcal{O}(1) = \mathcal{O}(1).$$

The SLLN (independent but not identically distributed) further yields that

$$\sum_{k=1}^{\kappa} \omega_k \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \xrightarrow{a.s.} 0.$$

It sufficed to show

$$\sum_{k=1}^{\kappa} \omega_k \left\{ \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\}^2 \xrightarrow{a.s.} \mathbb{E} Z_{k,j}^2 = \frac{1 - r^2(2\tau - 1)^2}{4r^2 f^2(x^*)}.$$

For case 1, $T_1 = T_2 = \cdots = T_{\kappa_0} = T_{\kappa_0} + 1 = \cdots = T_\kappa + 1$. The SLLN (i.i.d.) implies that

$$\sum_{k=1}^{\kappa_0} \omega_k \left\{ \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\}^2 \xrightarrow{a.s.} \sum_{k=1}^{\kappa_0} \frac{T_k}{T} \mathbb{E} Z_{k,j}^2 = \sum_{k=1}^{\kappa_0} \frac{T_k}{T} \frac{1 - r^2(2\tau - 1)^2}{4r^2 f^2(x^*)},$$

$$\sum_{k=\kappa_0+1}^{\kappa} \omega_k \left\{ \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\}^2 \xrightarrow{a.s.} \sum_{k=\kappa_0+1}^{\kappa} \frac{T_k}{T} \mathbb{E} Z_{k,j}^2 = \sum_{k=\kappa_0+1}^{\kappa} \frac{T_k}{T} \frac{1 - r^2(2\tau - 1)^2}{4r^2 f^2(x^*)}.$$

The result is obtained by adding the above two expressions.

For case 2, the SLLN (i.i.d.) entails that

$$\begin{aligned} &\sum_{k=1}^{\kappa-1} \omega_k \left\{ \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\}^2 - \sum_{k=1}^{\kappa-1} \omega_k \mathbb{E} Z_{k,j}^2 \\ &= \sum_{k=1}^{\kappa-1} \omega_k \left\{ \frac{1}{\sqrt{T_k}} \sum_{j=1}^{T_k} (x_{k,j} - x^*) \right\}^2 - \left(1 - \frac{T_\kappa}{T}\right) \mathbb{E} Z_{k,j}^2 \xrightarrow{a.s.} 0, \end{aligned}$$

As $T_\kappa / T = \mathcal{O}(1)$, it completes the proof of the consistency of $\widehat{\sigma}^2$.

Proofs of Theorem 3: According to the law of iterated logarithm, the rate of the bound of any confidence sequence for the unknown mean of Gaussian random variables with unit variance is at least $\sqrt{T^{-1} \log \log T}$. On the one hand, Theorem 1 shows that Conditions G-1 and G-3 in [51] are satisfied. On the other hand, Theorem 2 ensures Condition G-4 in [51]. Hence, we apply Theorem 2.4 in [51] to complete the proof.