Inference for Quantile-Parametrized Families via CDF Confidence Bands

Srijan Chattopadhyay[1] **Siddhaarth Sarkar**[2] **Arun Kumar Kuchibhotla**[2]

[1]Indian Statistical Institute, Kolkata
[2]Department of Statistics & Data Science, Carnegie Mellon University

Abstract

Quantile-based distribution families are an important subclass of parametric families, capable of exhibiting a wide range of behaviors using very few parameters. These parametric models present significant challenges for classical methods, since the CDF and density do not have a closed-form expression. Furthermore, approximate maximum likelihood estimation and related procedures may yield non-$\sqrt{n}$ and non-normal asymptotics over regions of the parameter space, making bootstrap and resampling techniques unreliable. We develop a novel inference framework that constructs confidence sets by inverting distribution-free confidence bands for the empirical CDF through the known quantile function. Our proposed inference procedure provides a principled and assumption-lean alternative in this setting, requiring no distributional assumptions beyond the parametric model specification and avoiding the computational and theoretical difficulties associated with likelihood-based methods for these complex parametric families. We demonstrate our framework on Tukey Lambda and generalized Lambda distributions, evaluate performance through simulation studies, and illustrate practical utility with applications to a small-sample dataset (Twin Study) and a large-sample dataset (Spanish household incomes).

1 INTRODUCTION

Parametric models summarize data using a finite set of interpretable parameters, which makes them efficient and widely useful in statistics and machine learning. To achieve greater shape flexibility, some families are defined via quantile functions rather than densities. Such quantile-parameterized models arise naturally in a range of modern data settings, where inference for the underlying shape and tail parameters is important for interpretation, comparison, and downstream decisions. Here are a few examples:

Simulator-first / likelihood-free modeling: In many ML and scientific workflows, the model is accessed primarily through forward simulation, yet one still seeks low-dimensional distributional summaries with uncertainty quantification for their parameters [Cranmer et al., 2020].

Synthetic data (privacy and prototyping): Synthetic data is used to enable analysis, sharing, and prototyping while aiming to preserve the main distributional structure, making flexible parametric distributional families (with inferentially meaningful parameters) practically relevant [Jordon et al., 2022].

Production ML monitoring: Monitoring deployed ML applications emphasizes tracking data/model behavior during operation; interpretable distributional summaries provide a convenient interface for reporting and diagnostics, provided inference for the associated parameters is available [Schröder and Schulz, 2022].

Despite their practical appeal in these settings, quantile-based parametric families present major inference challenges. The flexibility of these models comes at a cost: their densities and CDFs are often unavailable in closed form, making standard estimation methods difficult to implement. In particular, practitioners have developed alternatives such as trimmed L-moments [Asquith, 2007, Dean, 2013], numerical MLE approximations [Su, 2007b,a], and quantile-matching techniques Su [2010].

Beyond implementation difficulties, the statistical behavior of estimators for quantile-based models can be irregular, unlike the standard asymptotic theory. Existing alternatives either depend on unavailable likelihood formulas or require large simulations. We propose a simpler framework designed specifically for quantile-defined models that gives reliable finite-sample guarantees.

Although our main development is based on univariate CDF

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:1034–1053.

bounds, the same principle can be extended to multivariate data by replacing intervals of the form $(-\infty, x]$ with richer classes of sets. The univariate CDF approach works by setting upper and lower bounds for $P(X \leq x)$ and finding the parameters that fit within them. This idea can easily be extended to multivariate data by placing bounds on more complex shapes. For instance, using Vapnik–Chervonenkis (VC) theory, we can establish upper and lower bounds for $P(X \in A)$ across all hyperrectangles A [Kpotufe, 2011], and then find the parameters that satisfy all these inequalities. While finding every valid parameter is computationally difficult in multivariate settings, hypothesis testing becomes much easier because we only need to check if a specific hypothesized parameter satisfies the inequalities. When VC inequalities can sometimes be too conservative in practice, we can use the theory of statistically equivalent blocks [Tukey, 1947] as an alternative to obtain tighter, distribution-free confidence bounds, much like in the univariate case.

2 INFERENCE FRAMEWORK

Assume that we have i.i.d. univariate data $X_1, \ldots, X_n$ from a parametric distribution F_{Λ_0}, characterized by a parameter Λ_0 (not necessarily one-dimensional) and a quantile function $Q_{\Lambda_0}(p)$. The key feature of our setting is that we have a closed-form expression for the quantile function Q_Λ, but not for the CDF F_Λ or the associated density.

We first construct a nonparametric confidence band for the unknown CDF F_{Λ_0} using the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$, so that with probability $(1-\alpha)$, $\ell_{n,\alpha,i} \leq F_{\Lambda_0}(X_{(i)}) \leq u_{n,\alpha,i}$, $i \in [n]$. This band can be inverted through the quantile function to obtain $Q_{\Lambda_0}(\ell_{n,\alpha,i}) \leq X_{(i)} \leq Q_{\Lambda_0}(u_{n,\alpha,i})$, $i \in [n]$, which links the data directly to the parameters. The confidence set is then

$$\widehat{\mathrm{CI}}_{n,\alpha} = \big\{\Lambda : Q_\Lambda(\ell_{n,\alpha,i}) \leq X_{(i)} \leq Q_\Lambda(u_{n,\alpha,i}),\ \forall i \in [n]\big\},$$

containing all parameter values consistent with the joint band. The resulting procedure (explained in Algorithm 1) satisfies the coverage guarantee:

$$\mathbb{P}_{\Lambda_0}\Big(\Lambda_0 \in \widehat{\mathrm{CI}}_{n,\alpha}\Big) \geq 1-\alpha,$$

ensuring that the true parameter lies within the confidence region with at least $1-\alpha$ probability (proof in Section B).

Reparametrization for Multi-parameter families. In multi-parameter families, the confidence set $\widehat{\mathrm{CI}}_{n,\alpha}$ can be difficult to interpret directly, especially if the initial parametrization in terms of Λ is difficult to interpret. As shall be seen in Section 3.2, this is the case for generalized lambda distribution (GLD) family with four parameters. In these cases, it may be beneficial to first consider inference for some easy-to-understand functionals of the underlying distribution and then transform the confidence set to get one for the original parametrization. We therefore introduce a reparametrized variant of Algorithm 1 that targets interpretable summaries of F_{Λ_0}. Let $\{\phi_j(F) : j \in \mathcal{J}\}$ be a collection of scalar functionals of a CDF F (in our applications, ϕ_j will be quantile-based, e.g., differences or ratios of quantiles). Using the same joint CDF band as above, we first form simultaneous confidence intervals for the vector $\big(\phi_j(F_{\Lambda_0})\big)_{j\in\mathcal{J}}$. We then transform these functional intervals through the model map $\Lambda \mapsto \big(\phi_j(F_\Lambda)\big)_{j\in\mathcal{J}}$ to obtain a confidence set for Λ. Algorithm 2 summarizes the procedure. In the context of GLD family, we choose $\phi_j(F), 1 \leq j \leq 4$ to be measures of location, scatter, skew, and kurtosis, which are significantly more interpretable than the initial parameterization of GLD. This procedure also provides confidence intervals with $(1-\alpha)$ coverage guarantee (proof in Section B).

This quantile-based inversion method thus provides a gen-

Algorithm 1 CDF-band inversion for quantile-parametric models

Require: Data $X_1, \ldots, X_n$, target $\alpha \in (0,1)$, confidence band type (e.g., DKW/DW) for F_{Λ_0}, quantile function $Q_\Lambda(\cdot)$, parameter constraints $\Lambda \in \mathcal{L}$.

1: Sort the sample to obtain order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$.

2: Build a CDF confidence band $[\ell_{n,\alpha}, u_{n,\alpha}]$ and set $\ell_{n,\alpha,i} = \ell_{n,\alpha}(X_{(i)})$, $u_{n,\alpha,i} = u_{n,\alpha}(X_{(i)})$ for $i \in [n]$.

3: Return the confidence set $\widehat{\mathrm{CI}}_{n,\alpha}$ defined as:

$$\big\{\Lambda \in \mathcal{L} : Q_\Lambda(\ell_{n,\alpha,i}) \leq X_{(i)} \leq Q_\Lambda(u_{n,\alpha,i}), \forall i \in [n]\big\}.$$

Algorithm 2 Reparametrization based parameter confidence intervals

Require: Data $X_1, \ldots, X_n$, target $\alpha \in (0,1)$, confidence band construction for F_{Λ_0}, model family $\{F_\Lambda : \Lambda \in \mathcal{L}\}$ (via Q_Λ), and functionals $\{\phi_j : j \in \mathcal{J}\}$.

1: Construct a joint CDF confidence band $[\ell_{n,\alpha}, u_{n,\alpha}]$ and define the corresponding set of plausible distributions

$$\mathcal{F}_{n,\alpha} := \big\{F :\ \ell_{n,\alpha}(x) \leq F(x) \leq u_{n,\alpha}(x) \text{ for all } x\big\}.$$

2: Compute a simultaneous confidence interval for $\phi_j(F_{\Lambda_0})$ by

$$L_{n,\alpha}(j) := \inf_{F \in \mathcal{F}_{n,\alpha}} \phi_j(F),\ U_{n,\alpha}(j) := \sup_{F \in \mathcal{F}_{n,\alpha}} \phi_j(F),$$

and set $\widehat{\mathrm{CI}}^{\phi}_{n,\alpha}(j) := \big[L_{n,\alpha}(j), U_{n,\alpha}(j)\big]$.

3: Return the parameter confidence set obtained by inversion:

$$\widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha} := \Big\{\Lambda \in \mathcal{L} :\ \phi_j(F_\Lambda) \in \widehat{\mathrm{CI}}^{\phi}_{n,\alpha}(j)\ \forall j \in \mathcal{J}\Big\}.$$

eral route to valid parametric inference in settings where the quantile function has a closed form, using distribution-free confidence bands for the univariate cumulative distribution function. Our inference framework is fully agnostic to the choice of confidence band. Although there are several choices of confidence bands available (Section 3.1, Sarkar and Kuchibhotla [2023] and the references therein), we focus on the DKW and DW bands [Dvoretzky et al., 1956, Massart, 1990, Dümbgen and Wellner, 2023]. These two confidence bands represent an analytically simple baseline and a more complex but optimal alternative. A detailed description of this is available in the Section A.1.

3 APPLICATION

We apply the proposed inference method to the Tukey lambda and generalized lambda models, which have closed-form quantile functions but intractable densities. For the Tukey lambda model, checking the constraints reduces to explicit inequalities and interval endpoints can be computed by root finding; see Section 3.1. For the generalized lambda model, we parameterize $(\mu, \tilde{\sigma})$ directly and solve for the shape parameters using normalized (scale-free) constraints; see Section 3.2. Empirical results show the framework adapts naturally to these models and outperforms classical approaches.

3.1 TUKEY LAMBDA DISTRIBUTION

The Tukey Lambda distribution, introduced in Tukey [1990, 1962], is a continuous symmetric distribution, defined by a quantile that is parameterized by the parameter $\lambda \in \mathbb{R}$ as follows:

$$Q\left(p, \lambda\right) \;=\; \begin{cases} \frac{1}{\lambda}\left(p^{\lambda} - (1-p)^{\lambda}\right), & \text{if } \lambda \neq 0\,, \\ \ln\left(\frac{p}{1-p}\right)\,, & \text{if } \lambda = 0\,. \end{cases} \tag{3.1}$$

The support of this distribution varies with λ, with it being $[-\frac{1}{\lambda}, \frac{1}{\lambda}]$ for $\lambda > 0$, and $\mathbb{R}$ for $\lambda < 0$. The distribution has several interesting properties, with the ability to represent different tail behaviors depending on the choice of λ (See Section A.2).

We apply the general framework from Section 2 to the Tukey Lambda distribution. Given i.i.d. samples $X_1, \ldots, X_n$ with quantile function $p \mapsto Q(p, \lambda_0)$ and a CDF confidence band $(\ell_{n,\alpha,i}, u_{n,\alpha,i})$, the confidence set is

$$\widetilde{\mathrm{CI}}_{n,\alpha} = \bigcap_{i=1}^{n} \widetilde{\mathrm{CI}}_{n,\alpha,i}, \tag{3.2}$$

where

$$\widetilde{\mathrm{CI}}_{n,\alpha,i} := \{\lambda : Q(\ell_{n,\alpha,i}, \lambda) \le X_{(i)} \le Q(u_{n,\alpha,i}, \lambda)\}.$$

Now, note that $p \mapsto Q(p, \lambda)$ is increasing and non-positive for $p < 1/2$, and is decreasing and non-negative for $p > 1/2$. Given these monotonicity properties, we can determine a closed-form expression for $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ (See Section A.2).

Because the confidence bands are distribution-free, they do not use the symmetry of the Tukey Lambda distribution. We can improve efficiency by exploiting symmetry via the transformation $x \mapsto |x|$. The absolute values $\{|X_i|\}_{i=1}^{n}$ form a sufficient statistic for λ_0 and combine information from both sides of zero, leading to sharper inference. Importantly, the transformed variable still admits a closed-form quantile function $\widetilde{Q}$:

$$\widetilde{Q}\left(p, \lambda\right) = Q\left(\frac{1+p}{2}, \lambda\right) \tag{3.3}$$

Mimicking Eq (3.2), we get the following confidence interval for λ_0: $\widehat{\mathrm{CI}}_{n,\alpha} = \bigcap_{i=1}^{n} \widehat{\mathrm{CI}}_{n,\alpha,i}$, where

$$\widehat{\mathrm{CI}}_{n,\alpha,i} := \{\lambda : \widetilde{Q}(\ell_{n,\alpha,i}, \lambda) \le |X|_{(i)} \le \widetilde{Q}(u_{n,\alpha,i}, \lambda)\}.$$

As suggested by the sufficiency argument, the transformation is expected to yield a more powerful inference procedure, i.e., narrower confidence intervals. We compare the average width and coverage for the original and transformed data across values of λ_0 and sample sizes n (Fig. 1). The transformed procedure shows consistent improvement, empirically confirming the theoretical intuition. Theorem 3.1 and Remark 3.1 theoretically justify the observed shrinkage of the confidence intervals by showing that each constraint concentrates around λ_0 as $n \to \infty$.

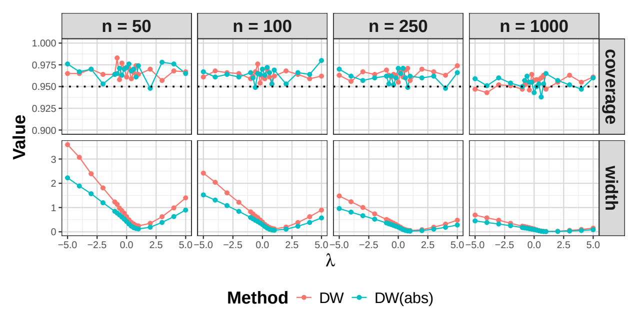

Figure 1: Comparing confidence interval generated by $\{X_i\}_{i\in[n]}$ vs. $\{|X|_i\}_{i\in[n]}$: across λ_0 and growing sample sizes. We choose the DW confidence band and coverage level of 95%. The number of replications is 500.

Theorem 3.1. *Let $Q(p, \lambda), \widetilde{Q}(p, \lambda)$ be as defined in Eq (3.1), & Eq (3.3) for $0 \le p \le 1$. Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. samples from Tukey Lambda distribution with parameter λ_0, i.e. have quantile function $Q(\cdot, \lambda_0)$. Let $|X|_{(i)}$ be the i-th order statistic of the absolute values of the observed data, so that $|X|_{(i)} \stackrel{d}{=} \widetilde{Q}(U_{(i)}, \lambda_0)$ $\forall i \in [n]$ for uniform order statistics $U_{(i)}, 1 \le$*

$i \leq n$. Then, for each $1 \leq i \leq n$,

$$\widehat{\mathrm{CI}}_{n,\alpha,i} \subseteq \left[\lambda_0 + \frac{\widetilde{Q}(U_{(i)}, \lambda_0) - \widetilde{Q}(\ell_{n,\alpha,i}, \lambda_0)}{\widetilde{Q}'(\ell_{n,\alpha,i}, \lambda_0)}, \right.$$
$$\left. \lambda_0 - \frac{\widetilde{Q}(u_{n,\alpha,i}, \lambda_0) - \widetilde{Q}(U_{(i)}, \lambda_0)}{\ln(p_{n,\alpha,i})\, \widetilde{Q}(U_{(i)}, \lambda_0)}\right].$$

where $p_{n,\alpha,i} = (1 + u_{n,\alpha,i})/2$ and $\widetilde{Q}'(p, \lambda_0)$ is the partial derivative of $\widetilde{Q}(p, \lambda)$ wrt λ at λ_0 for fixed p.

Remark 3.1. *From Section A.1, we know that $|U_{(i)} - \ell_{n,\alpha,i}| + |U_{(i)} - u_{n,\alpha,i}| = O_p(n^{-1/2})$ as $n \to \infty$ for any i, and hence, the length of $\widehat{\mathrm{CI}}_{n,\alpha,i}$ goes to 0 for each i.*

3.2 GENERALIZED LAMBDA DISTRIBUTION

The Tukey Lambda distribution is flexible but restricted to symmetry and zero median. The generalized Lambda distribution (GLD) removes these limitations by introducing location and asymmetry parameters, allowing a much broader class of shapes. In the common parameterization of Freimer et al. [1988], GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ has quantile function

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2}\left[\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4}\right],$$

where λ_1 is location, $\lambda_2 > 0$ scale, and λ_3, λ_4 control tail behavior. The Tukey Lambda arises when $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = \lambda_4$. However, these parameters must satisfy nontrivial constraints for Q to define a valid quantile function. Moreover, they are difficult to interpret. For these reasons, following Algorithm 2, we use the CSW parametrization from Chalabi et al. [2012] (See Section A.3). The CSW parameterization expresses the GLD in terms of interpretable location, scale, and shape components. The parameters $\tilde{\mu}$ and $\tilde{\sigma}$ represent the median and interquartile range, while bounded shape parameters $\chi \in (-1, 1)$ and $\xi \in (0, 1)$ control asymmetry and tail behavior. This separation simplifies inference: location and scale are obtained by inverting the CDF band, and shape is handled through normalized quantile ratios. We therefore construct confidence sets for $(\tilde{\mu}, \tilde{\sigma}, \chi, \xi)$ by exploiting this structure, avoiding an opaque four-dimensional region. Inference for the location parameter (median) follows readily from the confidence band; see Section A.3.

Scale parameter $(\tilde{\sigma})$ For $\tilde{\sigma}$, we first define a general quantile range functional $\mathrm{QR}(F, u_1, u_2)$ for $0 < u_2 < u_1 < 1$:

$$\mathrm{QR}(F, u_1, u_2) = F^{-1}(u_1) - F^{-1}(u_2).$$

Since L, U, and F are monotonic, we have $U^{-1}(x) \leq F^{-1}(x) \leq L^{-1}(x)$, which gives us the following inequality for QR:

$$U^{-1}(u_1) - L^{-1}(u_2) \leq \mathrm{QR}(F, u_1, u_2) \leq L^{-1}(u_1) - U^{-1}(u_2).$$

Using this relationship, we obtain the confidence interval for QR as:

$$\begin{aligned}
&\widehat{\mathrm{CI}}_{\mathrm{QR}}(u_1, u_2) \\
&:= \{\mathrm{QR}(G, u_1, u_2) : \ell_{n,\alpha}(x) \leq G(x) \leq u_{n,\alpha}(x)\ \forall x \in \mathbb{R}\} \\
&= [\max\{X_{(a_1)} - X_{(b_2)}, 0\}, X_{(b_1)} - X_{(a_2)}],
\end{aligned}$$

where

$$\begin{aligned}
a_1 &= \inf\{i : u_{n,\alpha,i} \geq u_1\}, \quad b_1 = \sup\{i : \ell_{n,\alpha,i} \leq u_1\}, \\
a_2 &= \inf\{i : u_{n,\alpha,i} \geq u_2\}, \quad b_2 = \sup\{i : \ell_{n,\alpha,i} \leq u_2\}.
\end{aligned}$$

The maximum with 0 ensures non-negativity, since $\mathrm{QR}(F, u_1, u_2) \geq 0$ for all $u_1 \geq u_2$. Any CDF in the confidence band will enforce this constraint when $X_{(a_1)} - X_{(b_2)}$ is negative. Substituting $u_1 = 3/4$ and $u_2 = 1/4$ yields the confidence interval for $\tilde{\sigma}$.

Shape parameters (χ, ξ) Having constructed confidence intervals for $\tilde{\mu}$ and $\tilde{\sigma}$, we now consider the shape parameters χ and ξ. We first define the rescaled shape statistic:

$$\tilde{s}(u_1, u_2 \mid \chi, \xi) = \frac{Q(u_2) - Q(u_1)}{Q(3/4) - Q(1/4)}$$

(See Equation (A.6).) This statistic represents the ratio of a quantile range to the IQR, effectively normalizing the shape characteristics.

Let $\widehat{\mathrm{CI}}_{\mathrm{QR}}(u_1, u_2) = [\widehat{\ell}_{\mathrm{QR}}(u_1, u_2), \widehat{u}_{\mathrm{QR}}(u_1, u_2)]$. The confidence interval for the shape statistic is:

$$\widehat{\mathrm{CI}}_s(u_1, u_2) := \left[\frac{\widehat{\ell}_{\mathrm{QR}}(u_1, u_2)}{\widehat{u}_{\mathrm{QR}}(3/4, 1/4)}, \frac{\widehat{u}_{\mathrm{QR}}(u_1, u_2)}{\widehat{\ell}_{\mathrm{QR}}(3/4, 1/4)}\right].$$

The joint confidence region for the shape parameters (χ, ξ) is, therefore,

$$\begin{aligned}
&\widehat{\mathrm{CI}}_{n,\alpha}(\chi, \xi; \mathcal{U}) \\
&:= \left\{(\chi, \xi) : \tilde{s}(u_1, u_2 \mid \chi, \xi) \in \widehat{\mathrm{CI}}_s(u_1, u_2)\ \forall (u_1, u_2) \in \mathcal{U}\right\},
\end{aligned} \tag{3.4}$$

where $\mathcal{U}$ is a collection of pairs with $1 \geq u_1 > u_2 \geq 0$. To compute $\widehat{\mathrm{CI}}_{n,\alpha}(\chi, \xi)$, we construct fine grids for $\chi \in (-1, 1)$ and $\xi \in (0, 1)$ and identify pairs (χ, ξ) for which all values of $\tilde{s}(u_1, u_2)$ fall within their corresponding confidence intervals.Clearly, enlarging the set $\mathcal{U}$ obtains shorter confidence intervals. However, this improvement comes at the cost of increased computational complexity. Thus, there exists an inherent trade-off between interval width and computational efficiency. An ideal procedure would achieve relatively short confidence intervals while maintaining a moderate computational burden. To illustrate this trade-off, we present a comparative analysis in Figure 2 (for true $\chi = 0, \xi = 0.3661$), where the impact of increasing $\mathcal{U}$ on both interval width and computational cost is examined. A brief discussion on choices of $\mathcal{U}$ can be found in Section A.4.

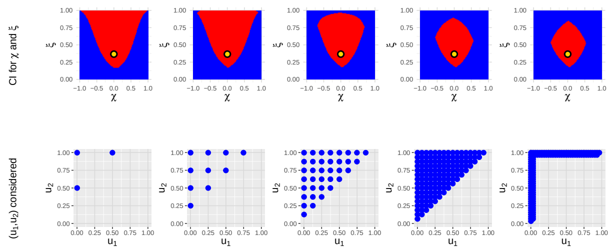

Figure 2: Choice of $\mathcal{U}$ and confidence region for $\chi = 0$ and $\xi = 0.3661$. Each column represents a choice of $\mathcal{U}$ and the corresponding confidence region. We use a coverage level of 95%.

4 SYNTHETIC DATA EXPERIMENTS

In the previous section, we have developed and described inference procedures for both the Tukey Lambda Distribution family and the Generalized Lambda Distribution (GLD) family. In this section, we do an empirical assessment of the proposed methodologies, by performing simulations under a wide range of parameter settings and sample sizes. The aim here is to demonstrate that our methods maintain coverage validity across different scenarios and to provide a comparison with existing techniques that have been proposed in the literature. In Section 4.1, we address the question of how to choose a confidence band. In particular, we compare two different confidence bands for our method in Tukey Lambda Distribution : the DKW and DW confidence bands. In Section 4.2, we compare our method for Tukey Lambda Distribution with various existing ones. In Section 4.3, we examine the proposed method for GLD and compare it with existing approaches. Finally, in Section 5, we perform a real data analysis, comparing it with bootstrap methods. A general conclusion from our simulations is that our proposed methods yield significantly robust validity performance compared to the existing methods.

4.1 CHOICE OF CONFIDENCE BANDS FOR THE TUKEY LAMBDA DISTRIBUTION

The first consideration when applying our method is the choice of confidence band. We focus on two options: the DKW and DW bands, defined in Eq (A.1) and Eq (A.3), respectively. We compare their performance in constructing confidence intervals for λ_0 in the Tukey Lambda distribution, with the results shown in Fig. 3.

Both bands achieve comparable coverage across the parameter space and maintain nominal validity. However, the DW band consistently yields narrower confidence intervals than the DKW band, demonstrating a clear efficiency advantage. This is not particularly surprising given the fact that the DW band has optimal width across all order statistics which the DKW band does not. Based on this comparison, we used the DW band for the remaining simulation studies.

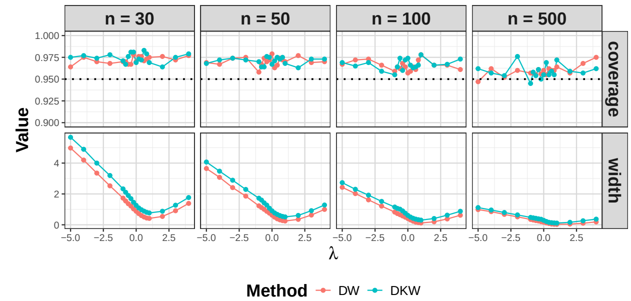

Figure 3: Comparison of DW and DKW confidence bands for the Tukey Lambda distribution across values of λ and sample sizes. Target coverage level is 95%, estimated with 500 replications.

4.2 COMPARISON OF METHODS FOR TUKEY LAMBDA DISTRIBUTION

An advantage of our method compared to the existing ones is that we do not need to construct an estimator of the true parameter to apply our method. All other existing methods start with an estimator and either rely on their limiting distribution or bootstrap to construct a confidence interval. We now compare our method for Tukey Lambda distribution (using the DW confidence band) to existing inference methods based on some method-of-moments estimator.

- *L-moments:* The L-moments [Hosking, 1990] method works only for $\lambda_0 > -1$. However, this restriction cannot be identified from the data, so we apply this method even when $\lambda_0 \leq -1$.
- *Quantile matching:* This method is valid for all ranges of λ. We take several quantiles and average the estimate of λ_0 obtained [Su, 2010].

Given these point estimators, bootstrap methods [Efron and Tibshirani, 1994] provide resampling-based procedures for statistical inference. The nonparametric bootstrap resamples with replacement from the observed data $X_1, \dots, X_n$ to generate bootstrap samples $X_1^*, \dots, X_n^*$, from which bootstrap estimates $\hat{\lambda}^*$ are computed. An alternative approach is the parametric bootstrap, which leverages the model's quantile function directly. Given an estimate $\hat{\lambda}$, bootstrap samples are generated as $X_i^* = Q_{\hat{\lambda}}(U_i)$, where $U_i \sim \text{Uniform}(0, 1)$. However, both bootstrap approaches lack theoretical justification when the underlying parameter estimators exhibit irregular asymptotic behavior.

We compare the performance of these methods across different values of the true parameter λ_0 and increasing sample sizes. The results are presented in Figure 4. As expected, the L-moment method performs poorly for $\lambda_0 \leq -1$, reflecting its known theoretical limitations in this regime. The quantile-based method achieves approximately nominal coverage for most combinations of λ_0 and n, although the parametric bootstrap exhibits undercoverage in small samples.

In finite samples, our proposed method produces narrower confidence intervals for all $\lambda_0 \geq 1$. When $\lambda_0 \leq 1$, the parametric bootstrap fails to provide valid coverage guarantees. Although the nonparametric bootstrap applied to quantile matching maintains coverage, it yields substantially wider intervals in small samples. Consequently, among the methods considered, only our approach simultaneously achieves finite-sample coverage and comparatively minimal interval width across a broad range of λ_0 values. As the sample size increases, the interval widths of all methods converge.

A particularly distinctive behavior emerges near $\lambda_0 = 0$. To investigate this region more closely, we perform an analysis using an increasing number of points in the neighborhood of zero, enabling a more detailed examination of potential irregularities in the competing methods. The results are shown in Figure 5 As we approach $\lambda_0 = 0$, the coverage of all other methods fluctuates heavily, while our method remains stable. Specifically, the L-moments method and the parametric bootstrap applied to quantile matching fail to give coverage for small samples. Although the nonparametric bootstrap for quantile matching gives valid coverage, its width is much higher compared to ours in small samples. This instability results from changes in the asymptotic behavior of the estimators near zero. This instability diminishes as the sample size increases.

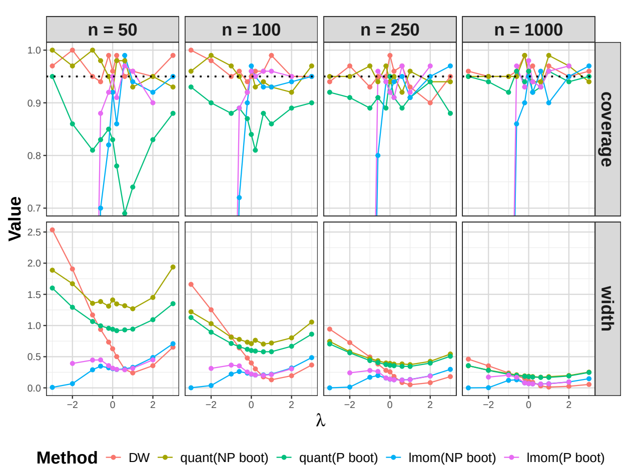

Figure 4: Method Comparison for Tukey Lambda: across λ and growing sample sizes. We choose a coverage level of 95%. The number of replications is 500.

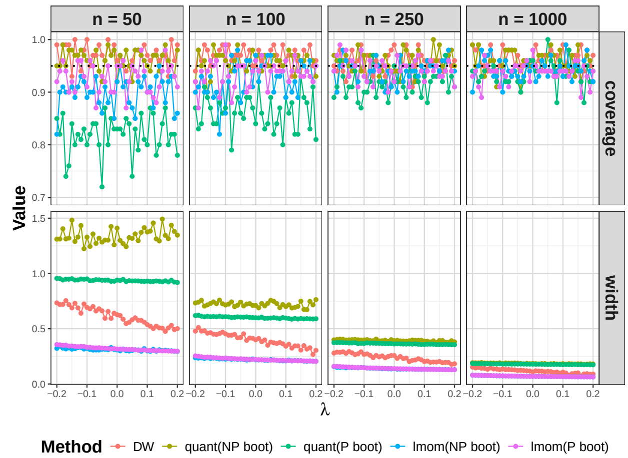

Figure 5: Method Comparison for Tukey Lambda: in a close neighbourhood around $\lambda = 0$ and growing sample sizes. We choose a coverage level of 95%. The number of replications is 500.

4.3 CONFIDENCE INTERVAL AND REGIONS FOR GLD

In this subsection, we study our proposed confidence intervals for Generalized Lambda Distribution via simulations. We first study the confidence intervals for the location parameter $\tilde{\mu}$ and scale parameter $\tilde{\sigma}$.

In Figure 6, we display the average width and coverage of the proposed confidence interval for $\mu \in \{-1, 0, 1\}$ as sample size increases. The confidence interval width decreases with sample size while maintaining nominal coverage across all values of μ. Since μ is a location parameter, variations in this parameter correspond to pure location changes, resulting in identical width behavior and consistent coverage probabilities across all μ values. The widths decrease rapidly with increasing sample size n.

In Figure 7, we display the average width and coverage of the proposed confidence interval for $\tilde{\sigma} \in \{1/2, 1, 2\}$ as sample size increases. The confidence interval width decreases with sample size while maintaining nominal coverage across all values of $\tilde{\sigma}$. Since $\tilde{\sigma}$ is a scale parameter, larger values correspond to increased variability in the distribution, resulting in correspondingly larger average interval widths across all sample sizes.

Finally, we examine the confidence regions for the asymmetry parameter χ and the steepness parameter ξ. In Figure 8, we display the average volume and coverage of the proposed confidence region for $\chi \in \{-0.8, -0.4, 0.0, 0.4, 0.8\}, \xi \in \{0.2, 0.35, 0.5, 0.65, 0.8\}$ as the sample size increases. We compute the volume and coverage for each of the pairs. The volume of the confidence region decreases with the sample size while maintaining the nominal coverage for each pair (χ, ξ). Since χ is the asymmetry parameter, we

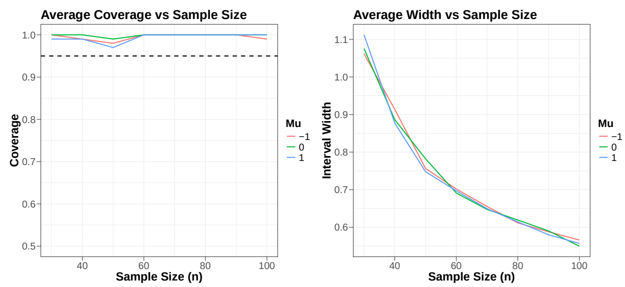

Figure 6: $\tilde{\mu}$ CI width and coverage: for different choices of $\tilde{\mu}$ and growing sample sizes. We choose a coverage level of 95%. The number of replications is 500.

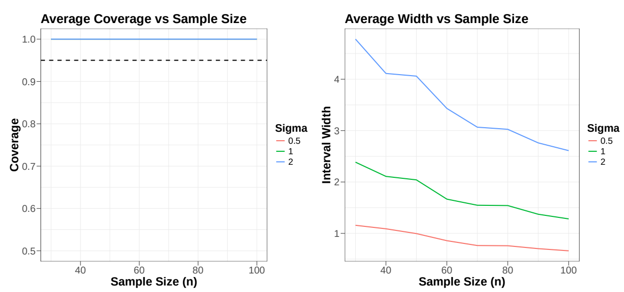

Figure 7: $\tilde{\sigma}$ CI width and coverage: for different choices of $\tilde{\sigma}$ and growing sample sizes. We choose a coverage level of 95%. The number of replications is 500.

observe a similar behavior for symmetric points around 0 for χ. Fig. 14 of Chalabi et al. [2012] suggests that as ξ increases, the distribution becomes increasingly steep. The steeper it is, the more concentrated it is. Hence, due to decreased uncertainty, the volume decreases as we increase ξ for each fixed χ. We now look a bit deeper into one sample instance of the confidence region for fixed (χ, ξ). We consider three representative scenarios with $(\chi, \xi) \in \{(0, 1/2 - 1/\sqrt{5}), (0, 0.3661), (0.2844, 0.3583)\}$ corresponding to approximating uniform, normal, and log-normal distributions. The results are presented in Fig. 9. As the sample size increases, the confidence regions contract around the true parameter values, demonstrating improved estimation precision.

Distribution	χ	ξ
Uniform	0	$\frac{1}{2} - \frac{1}{\sqrt{5}}$
Normal	0	0.3661
Log-normal	0.2844	0.3583

Table 1: Special cases of the GLD: (χ, ξ) values.

5 REAL-WORLD DATA EXPERIMENTS

In this section, we illustrate the practical utility of the proposed methods by applying them to real-world datasets. The aim is to demonstrate how the techniques developed in the

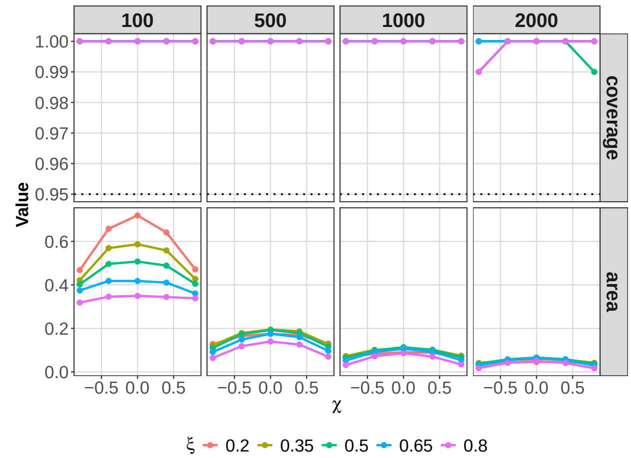

Figure 8: Comparison of Coverage and Area for different values of χ, ξ for different sample sizes. We choose a coverage level of 95%. The number of replications is 500.

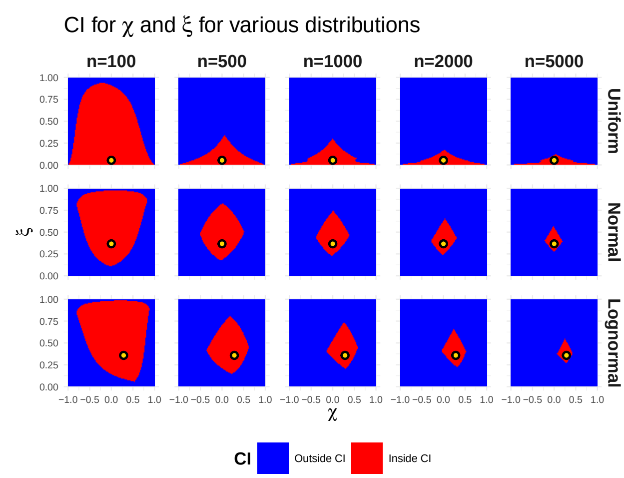

Figure 9: CI for χ and ξ for Uniform, Normal and Lognormal (see Table 1). We choose a coverage level of 95%. We choose edge set of equations for constructing the regions.

previous sections can be effectively used to draw meaningful statistical inferences and gain insights from empirical data. We compare our method with a bootstrap-based approach on a real dataset. For estimation, we use the estimates proposed in Chalabi et al. [2012].

$$\tilde{s} = \frac{S(3/4|\chi,\xi) + S(1/4|\chi,\xi) - 2S(2/4|\chi,\xi)}{S(3/4|\chi,\xi) - S(1/4|\chi,\xi)},$$
$$\tilde{\kappa} = \frac{S(7/8|\chi,\xi) - S(5/8|\chi,\xi) + S(3/8|\chi,\xi) - S(1/8|\chi,\xi)}{S(6/8|\chi,\xi) - S(2/8|\chi,\xi)}.$$

We compute the empirical version of $\tilde{s}, \tilde{\kappa}$, i.e.

$$\hat{\tilde{s}} = \frac{\hat{\pi}_{3/4} + \hat{\pi}_{1/4} - 2\hat{\pi}_{2/4}}{\hat{\pi}_{3/4} - \hat{\pi}_{1/4}}, \quad \hat{\tilde{\kappa}} = \frac{\hat{\pi}_{7/8} - \hat{\pi}_{5/8} + \hat{\pi}_{3/8} - \hat{\pi}_{1/8}}{\hat{\pi}_{6/8} - \hat{\pi}_{2/8}}$$

where $\hat{\pi}_q$ is the q th sample quantile of the data. We then estimate χ, ξ from the following set of nonlinear equations:

$$\tilde{s} = \hat{\tilde{s}}, \ \tilde{\kappa} = \hat{\tilde{\kappa}}$$

We use the `optimize` function in R to solve the nonlinear system of equations. We then generate bootstrap resamples and compute the corresponding estimates (χ^*, ξ^*) for each resample. From these estimates, we compute the coordinate-wise median and retain the 95% of bootstrap parameter estimates closest to this median in Euclidean distance. The bootstrap confidence region is taken to be the convex hull of these retained points. We illustrate the procedure using datasets discussed in Dedduwakumara et al. [2021].

5.1 SMALL SAMPLE: TWIN STUDY DATA

Consider the first example, a small-sample setting drawn from the Indiana Twin Study data, which consists of the birth weights of 123 twins. The data originally comes from the PhD thesis of Dr Cynthia Moore, Department of Medical and Molecular Genetics, Indiana University School of Medicine and has also been used in Karian and Dudewicz [2022] and Karian and Dudewicz [2000], for GLD parameter estimation. The CI of $\tilde{\mu}$ from the first set is [5.03, 5.81] and that from second set is [5.00, 5.88]. The CI for $\tilde{\sigma}$ from the first set is [0.78, 2.54], and that from the second set is [0.88, 2.28]. Here, we take $\mathcal{U} = \mathcal{U}_{RW}$ (Rivera and Walther [2013]) (See Equation (A.8)) while using our method. We provide the comparison of bootstrap and our method in Figure 10.

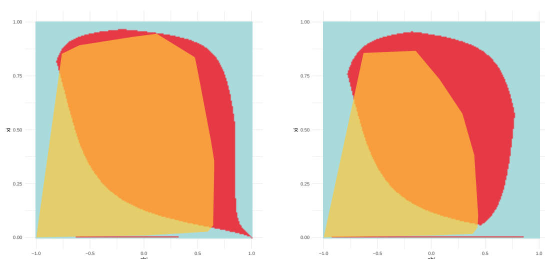

Figure 10: Comparison of Bootstrap vs our method in case of Twin data (Left - Set 1, Right - Set 2), Red one is our method and Yellow one is using Bootstrap.

The bootstrap procedure produces a confidence region that is quite symmetric with respect to χ, contrasting with the confidence region produced by our proposed method. This suggests that our procedure can capture more information about the distribution than the standard bootstrap procedure. Regarding volume, our method produces slightly larger confidence regions compared to bootstrap, but this is because it provides finite sample coverage guarantees, unlike bootstrap. Often, bootstrap can fail to provide the target coverage level for finite samples, as demonstrated in the simulations for the Tukey Lambda distribution (Section 4.2).

5.2 LARGE SAMPLE: SPANISH HOUSEHOLD INCOME DATA

Finally, we used our method on a large-sample dataset. In this example, we consider the total income data of Spanish households, collected in the 1980 Spanish Family Expenditure Survey (FES) described in Alonso-Colmenares et al. [1994]. This data set consists of 23,972 observations and total income is recorded with household characteristics and expenditure in several categories. This data set is available in the Ecdat package (Croissant and Graves [2016]) under the name "BudgetFood". We plot the density of the histogram of total income in Figure 11. The data exhibit a right-skewed and leptokurtic (peaked) distribution. The similarity of its shape to the various regimes illustrated in Fig. 14 of Chalabi et al. [2012] indicates that the GLD provides a suitable modeling framework for this data set. Using the proposed inference procedure, we obtained the confidence interval for the location parameter $\tilde{\mu}$ as $[719454, \ 742122]$, and for the scale parameter $\tilde{\sigma}$ as $[634159, \ 691116]$. In addition, the joint confidence region for the shape parameters by our method (χ, ξ) is presented in Figure 11. For practical time feasibility reasons, we only use the edge constraints (Eq. A.11) set while implementing our method. Unlike the twin-study data example, here we can see that the bootstrap and CDF based confidence regions are fairly small and similar, aligning with the large-sample trends we observed in our simulations (Fig. 8,9)

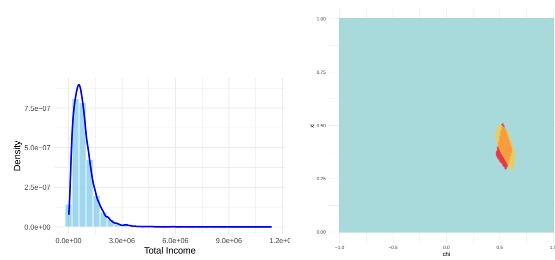

Figure 11: Histogram and Density of Total income of Spanish households (Left) and Comparison of Bootstrap vs our method for (χ, ξ) (Right), Red one is our method and Yellow one is using Bootstrap.

6 CONCLUSION

In conclusion, we have developed an inference framework for quantile-based parametric distributions that provides reliable finite-sample guarantees without requiring assumptions on where the true parameter lies within the parameter space. By inverting distribution-free confidence bands for the CDF through the known quantile function, the method avoids the irregular asymptotics that complicate likelihood-based approaches for these families.

Through simulations, we showed that the procedure yields valid finite-sample confidence intervals for the scaling parameter in the Tukey Lambda distribution across a wide range of parameter values and sample sizes. For the generalized Lambda distribution, the framework produces confidence intervals and joint confidence regions with nominal coverage and widths or volumes that decrease with the sample size, consistent with the theoretical behavior of the construction.

Several directions remain for further investigation. A detailed asymptotic analysis of the interval width would be useful for both the Tukey Lambda distribution and the GLD, particularly to understand whether the method adapts to parameter-specific irregularities in convergence rates. In addition, for the shape parameters of the GLD, a more principled strategy is needed to select $\mathcal{U}$ in Eq. (3.4). As illustrated in Fig. 2, the contributions of extreme quantiles appear to play a central role in determining an effective choice of $\mathcal{U}$.

References

M Dolores Alonso-Colmenares, Raquel Arévalo, Ana Lara, and Javier Ruiz-Castillo. *La encuesta de presupuestos familiares de 1980-81*. Universidad Carlos III Madrid (SP), 1994.

William H Asquith. L-moments and tl-moments of the generalized lambda distribution. *Computational Statistics & Data Analysis*, 51(9):4484–4496, 2007.

Robert H Berk and Douglas H Jones. Goodness-of-fit test statistics that dominate the kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 47(1):47–59, 1979.

Yohan Chalabi, David J. Scott, and Diethelm Wuertz. Flexible distribution modeling with the generalized lambda distribution. MPRA Paper 43333, University Library of Munich, Germany, 2012. URL https://mpra.ub.uni-muenchen.de/43333/.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Yves Croissant and Spencer Graves. *Ecdat: Data Sets for Econometrics*, 2016. URL https://CRAN.R-project.org/package=Ecdat. R package.

Benjamin Dean. Improved estimation and regression techniques with the generalised lambda distribution. *PhD. The University of Newcastle, Australia, Faculty of Science and Information Technology*, 2013.

Dilanka S Dedduwakumara, Luke A Prendergast, and Robert G Staudte. An efficient estimator of the parameters of the generalized lambda distribution. *Journal of Statistical Computation and Simulation*, 91(1):197–215, 2021.

Lutz Dümbgen and Jon A Wellner. A new approach to tests and confidence bands for distribution functions. *The Annals of Statistics*, 51(1):260–289, 2023.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. Chapman and Hall/CRC, 1994.

James J Filliben. The probability plot correlation coefficient test for normality. *Technometrics*, 17(1):111–117, 1975.

Marshall Freimer, Georgia Kollia, Govind S Mudholkar, and C Thomas Lin. A study of the generalized tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10):3547–3567, 1988.

Jonathan RM Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 52(1):105–124, 1990.

James Jordon, Lukasz Szpruch, Florimond Houssiau, Mirko Bottarelli, Giovanni Cherubin, Carsten Maple, Samuel N Cohen, and Adrian Weller. Synthetic data–what, why and how? *arXiv preprint arXiv:2205.03257*, 2022.

Zaven A Karian and Edward J Dudewicz. *Fitting statistical distributions: the generalized lambda distribution and generalized bootstrap methods*. Chapman and Hall/CRC, 2000.

Zaven A Karian and Edward J Dudewicz. Comparison of gld fitting methods: superiority of percentile fits to moments in lˆ 2 norm. *Journal of the Iranian Statistical Society*, 2 (2):171–187, 2022.

Samory Kpotufe. k-nn regression adapts to local intrinsic dimension. *Advances in neural information processing systems*, 24, 2011.

Housen Li, Axel Munk, Hannes Sieling, and Guenther Walther. The essential histogram. *Biometrika*, 107(2): 347–364, 2020.

Pascal Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

Camilo Rivera and Guenther Walther. Optimal detection of a jump in the intensity of a poisson process or in a density with likelihood ratio statistics. *Scandinavian Journal of Statistics*, 40(4):752–769, 2013.

Siddhaarth Sarkar and Arun Kumar Kuchibhotla. Post-selection inference for conformal prediction: Trading off coverage for precision. *arXiv preprint arXiv:2304.06158*, 2023.

Tim Schröder and Michael Schulz. Monitoring machine learning models: a categorization of challenges and methods. *Data Science and Management*, 5(3):105–116, 2022.

Steve Su. Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: Gldex in r. *Journal of Statistical Software*, 21:1–17, 2007a.

Steve Su. Numerical maximum log likelihood estimation for generalized lambda distributions. *Computational statistics & data analysis*, 51(8):3983–3998, 2007b.

Steve Su. Fitting glds and mixture of glds to data using quantile matching method. In *Handbook of Fitting Statistical Distributions with R*, pages 557–583. Chapman and Hall/CRC, 2010.

John W Tukey. Non-parametric estimation ii. statistically equivalent blocks and tolerance regions–the continuous case. *The Annals of Mathematical Statistics*, pages 529–539, 1947.

John W Tukey. The future of data analysis. In *Breakthroughs in Statistics: Methodology and Distribution*, pages 408–452. Springer, 1962.

John W. Tukey. The practical relationship between the common transformations of percentages or fractions and of amounts. In Colin L. Mallows, editor, *The Collected Works of John W. Tukey, Volume VI: More Mathematical*, pages 211–219. Wadsworth and Brooks/Cole, Pacific Grove, CA, 1990. Original work published 1960.

Inference for Quantile-Parametrized Families via CDF Confidence Bands (Supplementary Material)

Srijan Chattopadhyay[1] **Siddhaarth Sarkar**[2] **Arun Kumar Kuchibhotla**[2]

[1]Indian Statistical Institute, Kolkata
[2]Department of Statistics & Data Science, Carnegie Mellon University

A ADDITIONAL THEORETICAL DETAILS

A.1 CDF CONFIDENCE BANDS

Given an IID sample $X_1, \ldots, X_n$ from this distribution F and $\alpha \in (0,1)$, a *CDF confidence band* construction refers to constructing $\ell_{n,\alpha,i}, u_{n,\alpha,i}$ such that

$$\ell_{n,\alpha,i} \le F(X_{(i)}) \le u_{n,\alpha,i}, \quad \forall i \in [n] \text{ with probability } 1-\alpha.$$

Note that these quantities are fully deterministic. A standard approach uses the empirical CDF $\widehat{F}_n(x) = n^{-1}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}$ for all $x \in \mathbb{R}$. Plugging $\{X_{(i)}\}_{i\in[n]}$ into the classic Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [Dvoretzky et al., 1956, Massart, 1990] gives us $[\ell^{\mathrm{DKW}}_{n,\alpha,i}, u^{\mathrm{DKW}}_{n,\alpha,i}]$ with

$$\ell^{\mathrm{DKW}}_{n,\alpha,i} = i/n - \sqrt{\frac{\ln(2/\alpha)}{2n}}, \quad u^{\mathrm{DKW}}_{n,\alpha,i} = i/n + \sqrt{\frac{\ln(2/\alpha)}{2n}}, \ \forall i \in [n]. \tag{A.1}$$

This is a fixed-width confidence band, as $u_{n,\alpha,i} - \ell_{n,\alpha,i} = \sqrt{2\ln(2/\alpha)/n}$ does not depend on i. However, variable-width confidence bands can achieve better performance. Asymptotically, we know that $\sqrt{n}(F(X_{(i)}) - i/n) \xrightarrow{d} N\big(0, i(n-i)/n\big)$ as $n \to \infty$, suggesting that the width should scale as $\sqrt{i(n-i)/n}$. The DKW inequality only achieves a $n^{-1/2}$ rate without the standard deviation factor. Moreover, when $i = \mathcal{O}(1/n)$, $nF(X_{(i)})$ has a non-degenerate Poisson distribution limit, suggesting that the width for small (or large) values of i can be asymptotically smaller than $\mathcal{O}(n^{-1/2})$. To address these limitations, Dümbgen and Wellner [2023] introduced a modified Berk-Jones statistic that achieves rate-optimal confidence bands based on adjusted KL divergence, where $K(a,b) = a\log(a/b) + (1-a)\log((1-a)/(1-b))$ is the KL divergence between Bernoulli random variables. They propose the following statistic, defined with a tuning parameter $\nu > 3/4$:

$$T^{\mathrm{DW}}_{n,\nu} := \sup_{z\in\mathbb{R}} \left\{ n \cdot K(\widehat{F}_n(z), F(z)) - C_\nu(\widehat{F}_n(z), F(z)) \right\}, \tag{A.2}$$

where $C_\nu(u,v) := \min\{C(t) + \nu D(t) : \min(u,v) \le t \le \max(u,v)\}$ with $C(t) := \log\left(\log\left(\frac{e}{4t(1-t)}\right)\right)$ and $D(t) := \log(1 + C(t)^2)$. Theorem 2.1 of Dümbgen and Wellner [2023] shows that $T^{\mathrm{DW}}_{n,\nu}$ has a non-degenerate limiting distribution for any $\nu > 3/4$ when F is continuous. Furthermore, $T^{\mathrm{DW}}_{n,\nu}$ is distribution-free for any continuous F, and for any discontinuous F it is stochastically dominated by its continuous counterpart. Thus, we can use Monte Carlo simulations with standard uniform random variables to obtain accurate estimates of the quantiles. Let $\kappa^{\mathrm{DW}}_n(\alpha)$ be the $(1-\alpha)$-th quantile of $T^{\mathrm{DW}}_{n,\nu}$. This yields the confidence band $[\ell^{\mathrm{DW}}_{n,\alpha,i},\ u^{\mathrm{DW}}_{n,\alpha,i}]$ as follows:

$$\begin{aligned} \ell^{\mathrm{DW}}_{n,\alpha,i} &:= \min\left\{u \in [0, i/n) : n \cdot K(i/n, u) - C_\nu(i/n, u) \le \kappa^{\mathrm{DW}}_n(\alpha)\right\} \\ u^{\mathrm{DW}}_{n,\alpha,i} &:= \max\left\{u \in (i/n, 1] : n \cdot K(i/n, u) - C_\nu(i/n, u) \le \kappa^{\mathrm{DW}}_n(\alpha)\right\}. \end{aligned} \tag{A.3}$$

Compared to the DKW confidence band, the DW confidence band can be much smaller at both endpoints of the distribution. Dümbgen and Wellner [2023, Thms. 3.5–3.7] show that $u_{n,\alpha,i} - \ell_{n,\alpha,i} \lesssim n^{-1/2}$ for all i, similar to the DKW confidence band. However, similar to the Berk and Jones [1979] confidence band, $u_{n,\alpha,i} - \ell_{n,\alpha,i} \lesssim (\log\log n)/n$ if $\min\{i, n-i\} \lesssim \log\log n$.

A.2 TUKEY'S LAMBDA DISTRIBUTION

The quantile function of Tukey's Lambda distribution is described in Section 3.1. A major application of this distribution is to construct probability plot correlation coefficient (PPCC) plots, a diagnostic tool to identify appropriate distributional models, particularly for symmetric data [Filliben, 1975]. This method takes advantage of the shape parameter λ to characterize the behavior of the tail and suggest candidate distributions based on the maximum correlation achieved. (The correlation is computed between the observed order statistics and the theoretical quantiles of the Tukey Lambda distribution.)

Some important λ parameter values and their corresponding distributional behaviour are given in Table 2.

λ Value	Approximate Distribution	Tail Behavior
-1	Cauchy	Extremely heavy tails
0	Logistic	Heavy tails
0.14	Normal (Gaussian)	Moderate tails
0.5	U-shaped	Very light tails
1	Uniform$(-1, 1)$	Bounded support

Table 2: Tukey Lambda Distribution: Parameter Values and Corresponding Distributional Behaviors

Given these monotonicity properties of $Q(p, .)$, as described earlier, we can determine a closed-form expression for $\widetilde{\mathrm{CI}}_{n,\alpha,i}$. The complete characterization is given in Table 3. Thus, for each inequality corresponding to a particular order statistic, we

Table 3: Characterization for upper and lower boundaries of $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ based on other conditions.

Inequality	**Condition on** $\ell_{n,\alpha,i}, u_{n,\alpha,i}$	**Condition on** $X_{(i)}$	**Solution in** λ
$Q(\ell_{n,\alpha,i}, \lambda) \leq X_{(i)}$	$\ell_{n,\alpha,i} \leq 1/2$	$X_{(i)} < 0$	$\lambda \leq u_i^L$
$Q(\ell_{n,\alpha,i}, \lambda) \leq X_{(i)}$	$\ell_{n,\alpha,i} \leq 1/2$	$X_{(i)} \geq 0$	$\lambda \in \mathbb{R}$
$Q(\ell_{n,\alpha,i}, \lambda) \leq X_{(i)}$	$\ell_{n,\alpha,i} \geq 1/2$	$X_{(i)} \leq 0$	$\emptyset$
$Q(\ell_{n,\alpha,i}, \lambda) \leq X_{(i)}$	$\ell_{n,\alpha,i} \geq 1/2$	$X_{(i)} > 0$	$\lambda \geq \ell_i^L$
$Q(u_{n,\alpha,i}, \lambda) \geq X_{(i)}$	$u_{n,\alpha,i} \leq 1/2$	$X_{(i)} \geq 0$	$\emptyset$
$Q(u_{n,\alpha,i}, \lambda) \geq X_{(i)}$	$u_{n,\alpha,i} \leq 1/2$	$X_{(i)} < 0$	$\lambda \geq \ell_i^U$
$Q(u_{n,\alpha,i}, \lambda) \geq X_{(i)}$	$u_{n,\alpha,i} \geq 1/2$	$X_{(i)} \leq 0$	$\lambda \in \mathbb{R}$
$Q(u_{n,\alpha,i}, \lambda) \geq X_{(i)}$	$u_{n,\alpha,i} \geq 1/2$	$X_{(i)} > 0$	$\lambda \leq u_i^U$

get a confidence interval $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ of the following form.

$$\widetilde{\mathrm{CI}}_{n,\alpha,i} = \begin{cases} [L_i^u, U_i^\ell] & \text{if } \ell_{n,\alpha,i}, u_{n,\alpha,i} \leq 1/2, X_{(i)} < 0 \\ (-\infty, U_i^\ell] & \text{if } \ell_{n,\alpha,i} \leq 1/2 \leq u_{n,\alpha,i}, X_{(i)} < 0 \\ \mathbb{R} & \text{if } \ell_{n,\alpha,i} \leq 1/2 \leq u_{n,\alpha,i}, X_{(i)} = 0 \\ (-\infty, U_i^u] & \text{if } \ell_{n,\alpha,i} \leq 1/2 \leq u_{n,\alpha,i}, X_{(i)} > 0 \\ [L_i^\ell, U_i^u] & \text{if } \ell_{n,\alpha,i}, u_{n,\alpha,i} \geq 1/2, X_{(i)} > 0 \\ \emptyset & \text{otherwise.} \end{cases} \tag{A.4}$$

Here L_i^u, U_i^u, L_i^ℓ and U_i^ℓ are all functions of $X_{(i)}, \ell_{n,\alpha,i}$ and $u_{n,\alpha,i}$. Therefore, each inequality corresponding to every order statistic contributes to the final confidence interval. An important question to consider: Which inequalities actually determine the final confidence set? For example, close to the median, we observe an interesting phenomenon. If $\ell_{n,\alpha,i} \leq 1/2 \leq u_{n,\alpha,i}$, $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ is unbounded from below. Hence, order statistic inequalities near the median ($\ell_{n,\alpha,i} \leq 1/2 \leq u_{n,\alpha,i}$) fail to provide nontrivial lower bounds on λ_0, yielding unbounded or semi-bounded intervals. This suggests that such inequalities may be less informative for inference compared to those concentrated on one side of the median. The informativeness of individual inequalities can be effectively assessed by visualizing $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ separately for each inequality constraint. Fig. 13 illustrates this for $\lambda_0 \in \{-2, 0, 2\}$ with sample size $n = 30$. The figure shows clearly that inequalities corresponding to extreme order statistics (near the tails) are the ones that dictate the final confidence interval. As we argued earlier via the sufficiency argument, we also would expect the transformation to absolute values to lead to a more powerful inference procedure (in other words, to produce narrow confidence intervals). Similarly to Fig. 13, we plot $\widehat{\mathrm{CI}}_{n,\alpha,i}$ with $\lambda \in \{-2, 0, 2\}$ and sample size $n = 30$ in Fig. 12. In particular, none of the intervals is unbounded in this case.

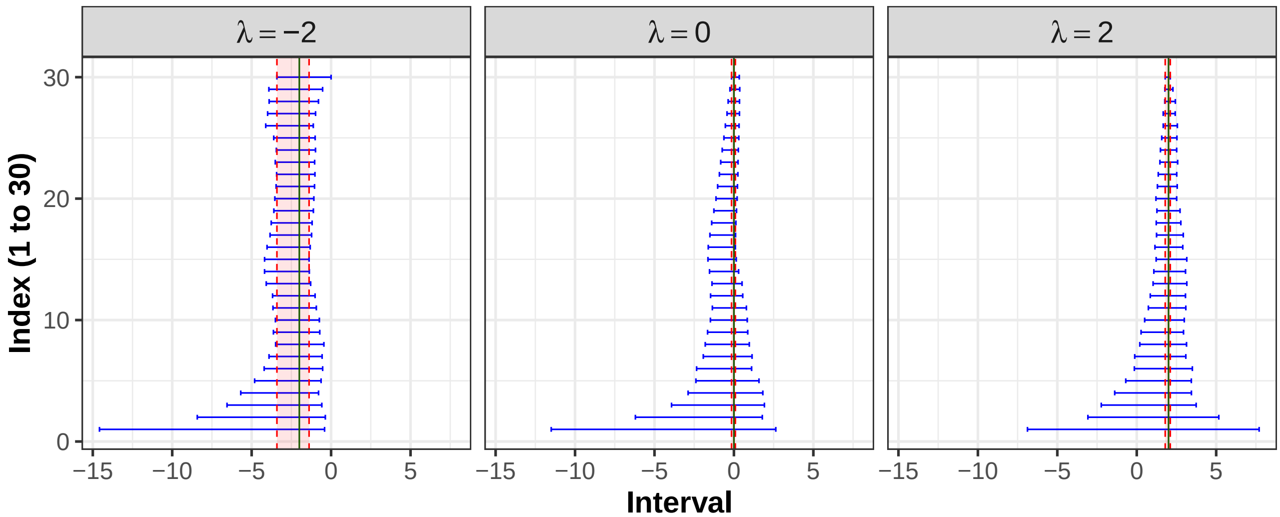

Figure 12: Comparing $\widehat{\mathrm{CI}}_{n,\alpha,i}$ vs. $i \in [n]$ for $\lambda \in \{-2, 0, 2\}$ and $n = 30$. The green line is the true λ and the red band represents the final confidence interval. We use the DW confidence band in Eq (A.3).

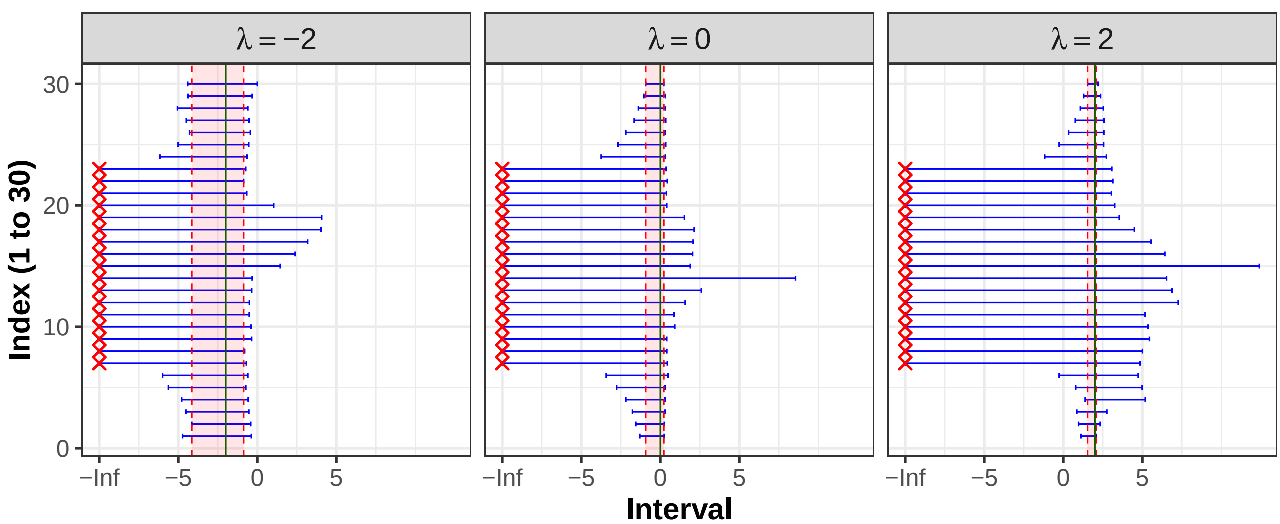

Figure 13: Comparing $\widetilde{\mathrm{CI}}_{n,\alpha,i}$ vs. $i \in [n]$ for $\lambda_0 \in \{-2, 0, 2\}$ and $n = 30$. A red cross represents $-\infty$. The green line is the true λ_0 and the red band represents the final confidence interval. We use the DW confidence band in Eq (A.3).

A.3 CSW PARAMETRIZATION OF GLD

Chalabi et al. [2012] proposed a parameterization by transforming the parameters of the earlier parameterization. First, note that the usual quantile function can be written in the following form:

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2} S(u|\lambda_3, \lambda_4),$$

where

$$S(u|\lambda_3,\lambda_4) = \begin{cases} \ln(u) - \ln(1-u) & \text{if } \lambda_3 = 0,\ \lambda_4 = 0, \\ \ln(u) - \frac{1}{\lambda_4}\left[(1-u)^{\lambda_4} - 1\right] & \text{if } \lambda_3 = 0,\ \lambda_4 \neq 0, \\ \frac{1}{\lambda_3}(u^{\lambda_3} - 1) - \ln(1-u) & \text{if } \lambda_3 \neq 0,\ \lambda_4 = 0, \\ \frac{1}{\lambda_3}(u^{\lambda_3} - 1) - \frac{1}{\lambda_4}\left[(1-u)^{\lambda_4} - 1\right] & \text{otherwise.} \end{cases}$$

In the limiting case $u = 0$:

$$S(0|\lambda_3,\lambda_4) = \begin{cases} -\dfrac{1}{\lambda_3} & \text{if } \lambda_3 > 0, \\ -\infty & \text{otherwise.} \end{cases}$$

In the limiting case $u = 1$:

$$S(1|\lambda_3,\lambda_4) = \begin{cases} \dfrac{1}{\lambda_4} & \text{if } \lambda_4 > 0, \\ \infty & \text{otherwise.} \end{cases}$$

The median, $\tilde{\mu}$, and the interquartile range, $\tilde{\sigma}$, can now be used to represent the location and scale parameters. These are defined as

$$\tilde{\mu} = Q(1/2),\ \tilde{\sigma} = Q(3/4) - Q(1/4).$$

The parameters λ_1 and λ_2 can therefore be expressed in terms of the median and interquartile range as follows.

$$\lambda_1 = \tilde{\mu} - \frac{1}{\lambda_2} S\left(\frac{1}{2}\middle|\lambda_3, \lambda_4\right),$$

$$\lambda_2 = \frac{1}{\tilde{\sigma}}\left[S\left(\frac{3}{4}\middle|\lambda_3,\lambda_4\right) - S\left(\frac{1}{4}\middle|\lambda_3,\lambda_4\right)\right].$$

The unbounded parameters λ_3, λ_4 are transformed to a bounded scale using the asymmetry parameter, χ, and the steepness parameter, ξ, defined as

$$\begin{aligned} \chi &= \frac{\lambda_3 - \lambda_4}{\sqrt{1 + (\lambda_3 - \lambda_4)^2}} \in (-1, 1) \\ \xi &= \frac{1}{2} - \frac{\lambda_3 + \lambda_4}{2\sqrt{1 + (\lambda_3 + \lambda_4)^2}} \in (0, 1). \end{aligned} \tag{A.5}$$

The function S can now be formulated in terms of the shape parameters χ and ξ. Given the definitions of $\tilde{\mu}, \tilde{\sigma}, \chi, \xi$, and S, the quantile function of the GLD becomes

$$Q_{\text{CSW}}(u|\tilde{\mu}, \tilde{\sigma}, \chi, \xi) = \tilde{\mu} + \tilde{\sigma}\frac{S(u|\lambda_3(\chi,\xi), \lambda_4(\chi,\xi)) - S\left(\frac{1}{2}|\lambda_3(\chi,\xi), \lambda_4(\chi,\xi)\right)}{S\left(\frac{3}{4}|\lambda_3(\chi,\xi), \lambda_4(\chi,\xi)\right) - S\left(\frac{1}{4}|\lambda_3(\chi,\xi), \lambda_4(\chi,\xi)\right)}, \tag{A.6}$$

where $\lambda_3(\chi, \xi), \lambda_4(\chi, \xi)$ are solutions to (A.5); see Equation (7,8,9) of Chalabi et al. [2012].

Inference for Location parameter $(\tilde{\mu})$ From Chalabi et al. [2012], we have the following relationships:

$$\tilde{\mu} = Q(1/2), \quad \tilde{\sigma} = Q(3/4) - Q(1/4).$$

Since $\tilde{\mu} = Q(\frac{1}{2}) = F^{-1}(\frac{1}{2})$, we can invert the CDF band to obtain confidence intervals for any quantile $Q(u)$:

$$\widehat{\text{CI}}_Q(u) = [X_{(a)}, X_{(b)}], \tag{A.7}$$

where $a = \inf\{i : u_{n,\alpha,i} \geq u\}$ and $b = \sup\{i : \ell_{n,\alpha,i} \leq u\}$. Substituting $u = 1/2$ gives the confidence interval for $\tilde{\mu}$.

A.4 CHOICE OF $\mathcal{U}$

In Eq (3.4), one must determine which pairs (u_1, u_2) (i.e. $\mathcal{U}$) to consider. Although this initially appears to involve infinitely many choices, the problem reduces to a finite set since $\widehat{\mathrm{CI}}_s(u_1, u_2)$ only changes when $u_1, u_2 \in \{\ell_{n,\alpha,i}\}_{i\in[n]} \cup \{u_{n,\alpha,i}\}_{i\in[n]}$, i.e., at the confidence band values corresponding to order statistics. This yields at most $\mathcal{O}(n^2)$ pairs to consider.

One tractable approach draws from the scan statistic literature, specifically the Rivera-Walther confidence set (Rivera and Walther [2013]), which we use in the current context to provide an optimal choice of pairs.

$$\mathcal{U}_{\mathrm{RW}} = \bigcup_{l=2}^{l_{\max}} \mathcal{U}(l), \text{ where } \mathcal{U}(l) = \big\{(k/n, j/n] : j, k \in \{1 + i \cdot d_l, i \in \mathbb{N}_0\}, \tag{A.8}$$

$$m_l < k - j < 2m_l\big\}. \tag{A.9}$$

where $l_{\max} = \lfloor \log_2(n/\log n) \rfloor$, and for $2 \le l \le l_{\max}$, $m_l = n2^{-l}$, and $d_l = \lceil m_l/(6l^{1/2}) \rceil$. This collection has a size of $\mathcal{O}(n)$. When used to choose which pairs of order statistics to consider, it has been shown to achieve optimal rates while remaining computationally efficient in feature detection [Rivera and Walther, 2013] and histogram construction [Li et al., 2020]. Therefore, $\mathcal{U}_{\mathrm{RW}}$ provides a natural choice for $\mathcal{U}$.

However, a more refined approach is possible. Consider the analogy with the Tukey lambda inference for the scale parameter λ. Selecting $\mathcal{U}$ parallels choosing which order statistic inequalities are most effective in constructing the final confidence interval for λ. In that context, extreme statistics proved crucial because tail behavior fundamentally determines distributional shape parameters. We observe a similar phenomenon here.

To better understand this behavior, we examine how the choice of $\mathcal{U}$ affects the confidence region using a single sample of size $n = 500$ from GLD$(0, 1, 0, 0.3661)$ (which approximates $N(0, 1)$) (see Fig. 2). For this sample, we construct $\mathcal{U}_k$ as follows:

- Let k denote the number of grid points in the interval $(0, 1)$:

  $$\mathrm{seq}_k = \left\{ \left\lfloor \frac{(n-1)i}{k} \right\rfloor / n \;\middle|\; i = 1, 2, \ldots, k \right\},$$

  where n is the sample size.
- From this sequence, we form all possible combinations of 2-elements (u_1, u_2) with $u_1 > u_2$:

$$\mathcal{U}_k = \{(u_1, u_2) \mid u_1, u_2 \in \mathrm{seq}_k,\ u_1 > u_2\}. \tag{A.10}$$

The higher values of k produce finer grids seq_k and thus more combinations of (u_1, u_2). We also consider $\mathcal{U}_k^{\mathrm{edge}}$, defined as:

$$\mathcal{U}_k^{\mathrm{edge}} = \{(u_1, u_2) \in \mathcal{U}_k : u_1 \le 2/k \text{ or } u_2 \ge 1 - (2/k)\}. \tag{A.11}$$

This subset of $\mathcal{U}_k$ focuses on points near the edges of $[0, 1]^2$, examining the rescaled shape statistic $\tilde{s}$ at the data extremes.

Fig. 2 illustrates this behavior: the second row shows $\mathcal{U}_k$ for $k = 3, 5, 9, 17$, the final column displays $\mathcal{U}_{17}^{\mathrm{edge}}$, and the first row presents the corresponding confidence regions for (χ, ξ). As k increases, the constraint set expands, and the confidence region (in red) shrinks. Moreover, comparing the last two columns reveals that the confidence regions appear identical when restricting (u_1, u_2) to the edge values alone. This suggests that extreme statistics provide the most important constraints in determining the confidence region.

B COVERAGE GUARANTEES

We establish finite-sample coverage guarantees for the reparametrized confidence set of Algorithm 2, and then recover the guarantee for Algorithm 1 as a special case.

Theorem B.1 (Coverage of Algorithm 2)**.** *Let $X_1, \ldots, X_n$ be i.i.d. from F_{Λ_0} for some $\Lambda_0 \in \mathcal{L}$. Let $[\ell_{n,\alpha}, u_{n,\alpha}]$ be a $(1-\alpha)$-confidence band for F_{Λ_0} satisfying*

$$P_{\Lambda_0}(\ell_{n,\alpha}(x) \;\le\; F_{\Lambda_0}(x) \;\le\; u_{n,\alpha}(x), \quad \forall\, x \in \mathbb{R}) \;\ge\; 1 - \alpha.$$

Let $\{\phi_j : j \in \mathcal{J}\}$ be a collection of scalar functionals of a CDF, and define simultaneous confidence intervals $\widehat{\mathrm{CI}}^{\phi}_{n,\alpha}(j) = [L_{n,\alpha}(j),\, U_{n,\alpha}(j)]$ and the confidence set $\widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}$ as in Steps 2–3 of
Algorithm 2. Then

$$P_{\Lambda_0}\Big(\Lambda_0 \in \widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}\Big) \;\geq\; 1-\alpha.$$

Proof. Define the event

$$\mathcal{E} \;=\; \{\ell_{n,\alpha}(x) \leq F_{\Lambda_0}(x) \leq u_{n,\alpha}(x), \quad \forall\, x \in \mathbb{R}\}\,,$$

which satisfies $P_{\Lambda_0}(\mathcal{E}) \geq 1-\alpha$ by assumption. It suffices to show that on the event $\mathcal{E}$ we have $\Lambda_0 \in \widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}$, i.e.,

$$\phi_j(F_{\Lambda_0}) \;\in\; \big[L_{n,\alpha}(j),\, U_{n,\alpha}(j)\big] \quad \text{for all } j \in \mathcal{J}.$$

Fix $j \in \mathcal{J}$. Recall from Algorithm 2 that

$$L_{n,\alpha}(j) = \inf\big\{\phi_j(F) : \ell_{n,\alpha}(x) \leq F(x) \leq u_{n,\alpha}(x),\ \forall\, x \in \mathbb{R}\big\},$$
$$U_{n,\alpha}(j) = \sup\big\{\phi_j(F) : \ell_{n,\alpha}(x) \leq F(x) \leq u_{n,\alpha}(x),\ \forall\, x \in \mathbb{R}\big\}.$$

In event $\mathcal{E}$, the true CDF F_{Λ_0} belongs to the feasible set

$$\mathcal{F}_{n,\alpha} \;=\; \big\{F : \ell_{n,\alpha}(x) \leq F(x) \leq u_{n,\alpha}(x),\ \forall\, x \in \mathbb{R}\big\}.$$

Hence $\phi_j(F_{\Lambda_0})$ is one of the values over which the infimum and supremum are taken, so

$$L_{n,\alpha}(j) \;\leq\; \phi_j(F_{\Lambda_0}) \;\leq\; U_{n,\alpha}(j).$$

Since this holds for every $j \in \mathcal{J}$ simultaneously in $\mathcal{E}$, we have $\phi_j(F_{\Lambda_0}) \in \widehat{\mathrm{CI}}^{\phi}_{n,\alpha}(j)$ for all $j \in \mathcal{J}$. Inverting, Λ_0 satisfies all constraints in the definition of $\widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}$, so $\Lambda_0 \in \widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}$. Therefore,

$$P_{\Lambda_0}\Big(\Lambda_0 \in \widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}\Big) \;\geq\; P_{\Lambda_0}(\mathcal{E}) \;\geq\; 1-\alpha.$$

■

Corollary B.1 (Coverage of Algorithm 1)**.** *Let $X_1, \ldots, X_n$ be i.i.d. from F_{Λ_0}, and let $[\ell_{n,\alpha}, u_{n,\alpha}]$ be a $(1-\alpha)$-confidence band for F_{Λ_0}. Then the confidence set*

$$\widehat{\mathrm{CI}}^{c}_{n,\alpha} \;=\; \big\{\Lambda \in \mathcal{L} : Q_{\Lambda}(\ell_{n,\alpha,i}) \leq X_{(i)} \leq Q_{\Lambda}(u_{n,\alpha,i}),\ \forall\, i \in [n]\big\}$$

satisfies $P_{\Lambda_0}\Big(\Lambda_0 \in \widehat{\mathrm{CI}}^{c}_{n,\alpha}\Big) \geq 1-\alpha$.

Proof. We show that Algorithm 1 is the special case of Algorithm 2 with index set $\mathcal{J} = [n]$ and functionals

$$\phi_i(F) \;=\; F^{-1}(p_i), \qquad i \in [n],$$

where $p_i = F(X_{(i)})$ denotes the probability level associated with the i-th order statistic.

Under this choice, and using the fact that quantile inversion is order-reversing in the CDF argument, the infimum and supremum over $\mathcal{F}_{n,\alpha}$ reduce to

$$L_{n,\alpha}(i) \;=\; Q_{\Lambda_0}(\ell_{n,\alpha,i}), \qquad U_{n,\alpha}(i) \;=\; Q_{\Lambda_0}(u_{n,\alpha,i}),$$

where $\ell_{n,\alpha,i} = \ell_{n,\alpha}(X_{(i)})$ and $u_{n,\alpha,i} = u_{n,\alpha}(X_{(i)})$ are the band values evaluated at the i-th order statistic. Substituting into the definition of $\widehat{\mathrm{CI}}^{\mathcal{J}}_{n,\alpha}$ from Theorem B.1 gives exactly $\widehat{\mathrm{CI}}^{c}_{n,\alpha}$, and the coverage guarantee $P_{\Lambda_0}(\Lambda_0 \in \widehat{\mathrm{CI}}^{c}_{n,\alpha}) \geq 1-\alpha$ follows immediately. ■

C PROOFS OF THEOREMS

Lemma C.1. *Let $\widetilde{Q}(x,\lambda)$ be as defined in Equation (3.3) for $0 \le x \le 1, \lambda \in \mathbb{R}$. Then, $\widetilde{Q}(x,\lambda)$ is decreasing and convex as a function of λ for each $x \in [0,1]$.*

Proof. First let's do the analysis for $\lambda \neq 0$. Suppose $p = \frac{1+x}{2}, q = \frac{1-x}{2} = 1-p$. Then, $p \ge \frac{1}{2}, q \le \frac{1}{2}$. Then, we can write $\widetilde{Q}(x,\lambda)$ as

$$\widetilde{Q}(x,\lambda) = \frac{A(\lambda)}{\lambda}$$

where

$$A(\lambda) = p^\lambda - q^\lambda$$

We denote $\widetilde{Q}'(x,\lambda), \widetilde{Q}''(x,\lambda)$ as 1st and 2nd partial derivatives of $\widetilde{Q}(x,\lambda)$ with respect to λ for fixed x. Then,

$$\begin{aligned}
\widetilde{Q}'(x,\lambda) &= \frac{\lambda A'(\lambda) - A(\lambda)}{\lambda^2} \\
\widetilde{Q}''(x,\lambda) &= \frac{\lambda^2 A''(\lambda) - 2\lambda A'(\lambda) + 2A(\lambda)}{\lambda^3}
\end{aligned}$$

For showing, $\widetilde{Q}(x,\lambda)$ is decreasing as a function of λ, it is enough to show that $\lambda A'(\lambda) \le A(\lambda)$. So, for that, it is enough to show that

$$\begin{aligned}
&\lambda A'(\lambda) \le A(\lambda) \\
\iff\ &\lambda(p^\lambda \ln p - q^\lambda \ln q) \le p^\lambda - q^\lambda \\
\iff\ &p^\lambda(\lambda \ln p - 1) \le q^\lambda(\lambda \ln q - 1)
\end{aligned}$$

Now, note that $p \ge q$. Let, $a = \ln p, b = \ln q$, so clearly $0 > a > b$. So, it is enough to show that

$$e^{a\lambda}(a\lambda - 1) \le e^{b\lambda}(b\lambda - 1)$$

So, we will show $g(x) = e^{x\lambda}(x\lambda - 1)$ is decreasing for each $\lambda \in \mathbb{R}$ for $x < 0$. For that, it is enough to check the sign of $g'(x)$. So, for $x < 0$,

$$g'(x) = \lambda e^{x\lambda} + \lambda e^{x\lambda}(x\lambda - 1) = x\lambda^2 e^{x\lambda} < 0$$

Hence, g is decreasing, implying $\widetilde{Q}(x,\lambda)$ is a decreasing function of λ for $\lambda \neq 0$. Now, let's look into the case when $\lambda = 0$. Then,

$$\begin{aligned}
\widetilde{Q}'(x,0) &= \lim_{\lambda\to 0} \frac{\frac{1}{\lambda}\left[\left(\frac{1+x}{2}\right)^\lambda - \left(\frac{1-x}{2}\right)^\lambda\right] - \ln\left(\frac{1+x}{1-x}\right)}{\lambda} \\
&= \lim_{\lambda\to 0} \frac{\left(\frac{1+x}{2}\right)^\lambda - \left(\frac{1-x}{2}\right)^\lambda - \lambda\ln(1+x) + \lambda\ln(1-x)}{\lambda^2} \\
&= \lim_{\lambda\to 0} \frac{\left(\frac{1+x}{2}\right)^\lambda \ln\left(\frac{1+x}{2}\right) - \ln(1+x) - \left(\frac{1-x}{2}\right)^\lambda \ln\left(\frac{1-x}{2}\right) + \ln(1-x)}{2\lambda} \quad \text{[L Ho'pital]} \\
&= \lim_{\lambda\to 0} \frac{\left(\frac{1+x}{2}\right)^\lambda \left(\ln\left(\frac{1+x}{2}\right)\right)^2 - \left(\frac{1-x}{2}\right)^\lambda \left(\ln\left(\frac{1-x}{2}\right)\right)^2}{2} \quad \text{[L Ho'pital]} \\
&= \frac{\left(\ln\left(\frac{1+x}{2}\right)\right)^2 - \left(\ln\left(\frac{1-x}{2}\right)\right)^2}{2}
\end{aligned}$$

Now, since

$$0 > \ln\left(\frac{1+x}{2}\right) > \ln\left(\frac{1-x}{2}\right),$$

we clearly have $\widetilde{Q}'(x,\lambda) \le 0$. Also, note that $\lim_{\lambda\to 0} \widetilde{Q}'(x,\lambda) = \widetilde{Q}'(x,0)$. So, this is a smooth extension to $\lambda = 0$. Hence, $\widetilde{Q}(x,\lambda)$ is decreasing as a function of λ for each $x \in [0,1]$. Now, next we will show that it is convex as well. For that, it is enough to show that $\widetilde{Q}''(x,\lambda) \ge 0$. First assume $\lambda \neq 0$. So, we will show

$$\lambda^2 A''(\lambda) \ge 2\lambda A'(\lambda) - 2A(\lambda), \quad \text{for } \lambda \ge 0$$
$$\lambda^2 A''(\lambda) \le 2\lambda A'(\lambda) - 2A(\lambda), \quad \text{for } \lambda \le 0$$

So, for $\lambda \ge 0$,

$$\begin{aligned}
&\lambda^2 A''(\lambda) \ge 2\lambda A'(\lambda) - 2A(\lambda) \\
&\iff \lambda^2 p^\lambda (\ln p)^2 - \lambda^2 q^\lambda (\ln q)^2 \ge 2\lambda p^\lambda \ln p - 2\lambda q^\lambda \ln q - 2p^\lambda + 2q^\lambda \\
&\iff \lambda^2 p^\lambda (\ln p)^2 - 2\lambda p^\lambda \ln p + 2p^\lambda \ge \lambda^2 q^\lambda (\ln p)^2 - 2\lambda q^\lambda \ln p + 2q^\lambda
\end{aligned}$$

So, assuming $a = \ln p, b = \ln q$, $a > b$. Then, it is enough to show that

$$e^{a\lambda}(\lambda^2 a^2 - 2\lambda a + 2) \ge e^{b\lambda}(\lambda^2 b^2 - 2\lambda b + 2)$$

For that, we will show that $h(x) = e^{x\lambda}(\lambda^2 x^2 - 2\lambda x + 2)$ is increasing for $\lambda \ge 0$. For that,

$$\begin{aligned}
h'(x) &= e^{x\lambda}\lambda(\lambda^2 x^2 - 2\lambda x + 2) + e^{x\lambda}(2x\lambda^2 - 2\lambda) \\
&= \lambda^3 x^2 e^{x\lambda} \ge 0
\end{aligned}$$

Now, for $\lambda \le 0$, it is enough to show that

$$e^{a\lambda}(\lambda^2 a^2 - 2\lambda a + 2) \le e^{b\lambda}(\lambda^2 b^2 - 2\lambda b + 2)$$

For that, we will show that $h(x) = e^{x\lambda}(\lambda^2 x^2 - 2\lambda x + 2)$ is decreasing for $\lambda \le 0$. For that,

$$\begin{aligned}
h'(x) &= e^{x\lambda}\lambda(\lambda^2 x^2 - 2\lambda x + 2) + e^{x\lambda}(2x\lambda^2 - 2\lambda) \\
&= \lambda^3 x^2 e^{x\lambda} \le 0
\end{aligned}$$

Finally, let's do for $\lambda = 0$. For $\lambda \neq 0$,

$$\widetilde{Q}'(x,\lambda) = \frac{\lambda\left(p^\lambda \ln p - q^\lambda \ln q\right) - \left(p^\lambda - q^\lambda\right)}{\lambda^2}$$

Since, $\widetilde{Q}$ is smoothly extended at $\lambda = 0$,

$$\begin{aligned}
\widetilde{Q}''(x,0) &= \lim_{\lambda\to 0} \frac{\widetilde{Q}'(x,\lambda) - \widetilde{Q}'(x,0)}{\lambda} \\
&= \lim_{\lambda\to 0} \frac{\lambda(p^\lambda \ln p - q^\lambda \ln q) - (p^\lambda - q^\lambda) - \frac{\lambda^2(\ln p)^2}{2} + \frac{\lambda^2(\ln q)^2}{2}}{\lambda^3} \\
&= \lim_{\lambda\to 0} \frac{(p^\lambda \ln p - q^\lambda \ln q) + \lambda(p^\lambda(\ln p)^2 - q^\lambda(\ln q)^2) - (p^\lambda \ln p - q^\lambda \ln q) - \lambda(\ln p)^2 + \lambda(\ln q)^2}{3\lambda^2} \\
&= \lim_{\lambda\to 0} \frac{p^\lambda(\ln p)^2 - q^\lambda(\ln q)^{-}(\ln p)^2 + (\ln q)^2}{3\lambda} \\
&= \lim_{\lambda\to 0} \frac{p^\lambda(\ln p)^3 - q^\lambda(\ln q)^3}{3} \\
&= \frac{(\ln p)^3 - (\ln q)^3}{3}
\end{aligned}$$

Since, $p > q$, $\widetilde{Q}''(x,0) \ge 0$. Hence, $\widetilde{Q}''(x,\lambda) \ge 0$ for each $x \in [0,1]$, implying $\widetilde{Q}(x,\lambda)$ is a convex function of λ. ∎

Now, we are done proving the lemma. So, we finally we prove the Theorem 3.1.

Proof of Theorem 3.1. For simplicity of notation, we will use l_i in place of $\ell_{n,\alpha,i}$ and u_i in place of $u_{n,\alpha,i}$ for this proof. Let's first look at the case for $\lambda_0 \neq 0$. Then, for $0 \leq p \leq 1$,

$$\widetilde{Q}(p, \lambda) = \frac{1}{\lambda}\left[\left(\frac{1+p}{2}\right)^{\lambda} - \left(\frac{1-p}{2}\right)^{\lambda}\right].$$

Consider the constraint

$$A \ = \ \left\{\lambda : \widetilde{Q}(l_i, \lambda) \ \leq \ \widetilde{Q}(U_{(i)}, \lambda_0)\right\}$$

Here $U_{(i)}$ is the i^{th} uniform order statistic, and λ_0 is the truth. Now, by Lemma C.1, $\widetilde{Q}(x, \lambda)$ is increasing as a function of x, and decreasing as a function of λ. Hence, we have, for $\lambda \geq \lambda_0$,

$$\widetilde{Q}(l_i, \lambda) \leq \widetilde{Q}(U_{(i)}, \lambda) \leq \widetilde{Q}(U_{(i)}, \lambda_0)$$

Now, consider $\lambda < \lambda_0$. Let λ_1 denote the smallest such value satisfying

$$\widetilde{Q}(l_i, \lambda_1) \ = \ \widetilde{Q}(U_{(i)}, \lambda_0), \qquad \lambda_1 < \lambda_0.$$

By convexity,

$$\widetilde{Q}(l_i, \lambda_1) \ \geq \ \widetilde{Q}(l_i, \lambda_0) \ + \ \widetilde{Q}'(l_i, \lambda_0)\,(\lambda_1 - \lambda_0),$$

where $\widetilde{Q}'$ denotes the derivative with respect to λ. Since $\widetilde{Q}$ is decreasing as a function of λ (i.e., $\widetilde{Q}'(l_i, \lambda) \leq 0$),

$$\lambda_1 \geq \lambda_0 \ + \ \frac{\widetilde{Q}(U_{(i)}, \lambda_0) - \widetilde{Q}(l_i, \lambda_0)}{\widetilde{Q}'(l_i, \lambda_0)}.$$

It is clear from the previous analysis, that $g'(p)$ (as in the proof of Lemma C.1) is nonzero unless $\lambda = 0$ or $p = 0$. For $\lambda = 0$, as well, $\widetilde{Q}'$ is not 0 unless $p = 0$. Hence the superset is of the form $[\lambda_1, \infty)$. Now, consider the other side. Define

$$B = \{\lambda : \widetilde{Q}(U_{(i)}, \lambda_0) \leq \widetilde{Q}(u_i, \lambda)\}$$

Similarly, $\lambda \leq \lambda_0$ will automatically be part of the superset, since for $\lambda \leq \lambda_0$,

$$\widetilde{Q}(U_{(i)}, \lambda_0) \leq \widetilde{Q}(U_{(i)}, \lambda) \leq \widetilde{Q}(u_i, \lambda)$$

Then consider $\lambda > \lambda_0$. Let λ_2 be the greatest such value satisfying

$$\widetilde{Q}(u_i, \lambda_2) = \widetilde{Q}(U_{(i)}, \lambda_0), \qquad \lambda_2 > \lambda_0$$

We have $\widetilde{Q}(u_i, \lambda_0) \geq \widetilde{Q}(U_{(i)}, \lambda_2)$ because $\lambda_2 \geq \lambda_0$. By convexity, we get

$$\widetilde{Q}(u_i, \lambda_0) \geq \widetilde{Q}(u_i, \lambda_2) + \widetilde{Q}'(u_i, \lambda_2)(\lambda_0 - \lambda_2),$$

which implies

$$\lambda_2 \leq \lambda_0 - \frac{\widetilde{Q}(u_i, \lambda_0) - \widetilde{Q}(u_i, \lambda_2)}{\widetilde{Q}'(u_i, \lambda_2)}.$$

This is hard to solve in explicit form because of $\widetilde{Q}'(u_i, \lambda_2)$ in the denominator. We rather try to give bounds. Let's first look at the following result.

Claim: For $\lambda_2 \neq 0, \quad p_i = \frac{1+u_i}{2}, \ q_i = \frac{1-u_i}{2}$

$$\frac{p_i^{\lambda_2} \ln p_i - q_i^{\lambda_2} \ln q_i}{\lambda_2} \leq \left(\frac{1}{\lambda_2} + \ln(p_i)\right) \widetilde{Q}(u_i, \lambda_2).$$

Proof. Note that, if we can show

$$g_i(x) = \frac{p^{\lambda_2} \ln p}{\lambda_2} - \left(\frac{1}{\lambda_2} + \ln p_i\right) \frac{p^{\lambda_2}}{\lambda_2}$$

is decreasing in $[q_i, p_i]$, that will prove the claim. But, this is easy to see since

$$g_i'(x) = p^{\lambda_2 - 1} \ln x + \frac{x^{\lambda_2 - 1}}{\lambda_2} - \left(\frac{1}{\lambda_2} + \ln p_i\right) x^{\lambda_2 - 1} = x^{\lambda_2 - 1} \ln\left(\frac{x}{p_i}\right) \leq 0$$

since $x \leq p_i$. That proves the claim. ∎

Hence, till now we proved that for $\lambda_2 \neq 0$,

$$\frac{p_i^{\lambda_2} \ln p_i - q_i^{\lambda_2} \ln q_i}{\lambda_2} \leq \left(\frac{1}{\lambda_2} + \ln(p_i) \right) \widetilde{Q}(u_i, \lambda_2).$$

Note that

$$\widetilde{Q}'(u_i, \lambda_2) = \frac{p_i^{\lambda_2} \ln p_i - q_i^{\lambda_2} \ln q_i}{\lambda_2} - \frac{1}{\lambda_2} \widetilde{Q}(u_i, \lambda_2), \quad \text{with} \quad p_i = \frac{1+u_i}{2},\ q_i = \frac{1-u_i}{2}.$$

Hence, for $\lambda_2 \neq 0$,

$$\widetilde{Q}'(u_i, \lambda_2) \leq \left(\frac{1}{\lambda_2} + \ln(p_i) \right) \widetilde{Q}(u_i, \lambda_2) - \frac{1}{\lambda_2} \widetilde{Q}(u_i, \lambda_2) = \ln(p_i) \widetilde{Q}(u_i, \lambda_2)$$

But, for $\lambda_2 = 0$,

$$\begin{aligned} &\frac{(\ln p_i)^2 - (\ln q_i)^2}{2} \leq (\ln p_i)(\ln p_i - \ln q_i) \\ \iff &(\ln p_i)^2 - (\ln q_i)^2 \leq 2(\ln p_i)^2 - 2 \ln p_i \ln q_i \\ \iff &(\ln p_i - \ln q_i)^2 \geq 0, \end{aligned}$$

which is true. Hence, we get for $\lambda_2 \in \mathbb{R}$,

$$\widetilde{Q}'(u_i, \lambda_2) \leq \ln(p_i) \widetilde{Q}(u_i, \lambda_2)$$

So, we finally get

$$\lambda_2 \leq \lambda_0 - \frac{\widetilde{Q}(u_i, \lambda_0) - \widetilde{Q}(U_{(i)}, \lambda_0)}{\ln(p_i) \widetilde{Q}(U_{(i)}, \lambda_0)}$$

So, finally we get the following superset:

$$\left[\lambda_0 + \frac{\widetilde{Q}(U_{(i)}, \lambda_0) - \widetilde{Q}(\ell_{n,\alpha,i}, \lambda_0)}{\widetilde{Q}'(\ell_{n,\alpha,i}, \lambda_0)}, \lambda_0 - \frac{\widetilde{Q}(u_{n,\alpha,i}, \lambda_0) - \widetilde{Q}(U_{(i)}, \lambda_0)}{\ln(p_i) \widetilde{Q}(U_{(i)}, \lambda_0)} \right]$$

∎