An RP-based Resampling Method for the Logistic Distribution and Its Application

Cuiran Shi[1], Xindan Liang[2] and Jiajuan Liang[3,4,*]

[1]*Department of Biostatistics, State University of New York at Buffalo, 401 Kimball Tower, Buffalo, New York 14214, USA*

[2]*Weill Cornell Medicine, Division of Biostatistics, Department of Population Health Sciences, 402 East 67th Street, LA-007, New York, New York 10065, USA*

[3]*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China*

[4]*Guangdong Provincial/Zhuhai Key Laboratory of Interdisciplinary Research and Application for Data Science, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China*

Abstract: This paper proposes a representative point-based bootstrap (RP-bootstrap) to improve confidence interval estimation for the logistic distribution. The method replaces the traditional empirical distribution with a smoothed approximation constructed from statistically optimal representative points (RPs), leading to a more stable resampling distribution. We integrate the RP-bootstrap with the bootstrap-t, percentile, and BC_a methods to construct intervals for the location and scale parameters. Its performance is compared to the classical nonparametric bootstrap via comprehensive Monte Carlo simulations and two real-data applications. The results show that the RP-bootstrap delivers noticeable improved finite-sample performance, particularly for small samples where standard bootstrapping often under-covers. It achieves recognizably higher coverage probabilities while maintaining shorter or comparable expected interval lengths. These improvements are strongest for the bootstrap-t interval and are consistent for both parameters, though more marked for the location. In conclusion, the RP-bootstrap is a computationally efficient and reliable alternative for logistic inference, offering enhanced accuracy, especially in small-sample scenarios.

Purpose: Construction of confidence intervals under small sample size is frequently encountered in statistical inference, such as estimating some treatment effect in medical research with limited number of patients. Traditional nonparametric bootstrap methods often suffer from undercoverage in such settings. To address this limitation, we propose the RP-bootstrap—a resampling procedure that draws samples from an approximated distribution formed by representative points (RPs) of the logistic distribution.

Methods: The RP-bootstrap is developed for constructing confidence intervals for the mean and variance of the logistic distribution. The algorithm generates weighted samples from the estimated RPs. The RP-bootstrap method is applied to construction of different types of confidence intervals (CIs) like the bootstrap-t, percentile, and {\rm BC_a} CIs. Its performance and comparison with the traditional nonparametric bootstrap are evaluated through Monte Carlo simulation and real-data application.

Results: Based on the Monte Carlo study under a set of small sample sizes, the RP-bootstrap achieves noticeable higher empirical coverage probability and competitive or shorter expected interval lengths compared with the nonparametric bootstrap. The improvements are much noticeable for small sample sizes like n<30 and for the bootstrap-t confidence intervals, where the nonparametric bootstrap frequently shows undercoverage of the true population parameter.

Contribution: This study demonstrates that representative points provide a stable and efficient alternative to resampling methods from logistic models. The RP-bootstrap offers a practical method for reliable small-sample inference and yields confidence intervals with improved accuracy and reduced variability relative to the traditional nonparametric bootstrap method.

Keywords: RP-bootstrap, nonparametric bootstrap, logistic distribution, representative points, bootstrap confidence intervals.

1. INTRODUCTION

The logistic distribution family is a symmetric location–scale distribution family with similar properties to those of the normal distribution $N(\mu,\sigma^2)$. It has been widely applied to statistical models for demographic studies [22, 24-25], biological growth modeling [28], bio-assay data analysis [4], survival studies [26, 30], agricultural and industrial processes [24, 28], income distribution modeling [17], and public-health research [18], etc. Its increasing importance in modern statistical and machine-learning applications, particularly through logistic regression and sigmoid-based neural network components, has further motivated the need for finite-sample inference procedures. There is a lot of literature focused on parameter estimation for the logistic distribution. For example, Ogawa [23] proposed best linear unbiased estimators derived from sample

*Address correspondence to this author at the Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China; E-mail: jiajuanliang@bnbu.edu.cn

quantiles; Blom [7] and Jung [20] introduced estimators based on order statistics; Harter and Moore [19] showed that maximum likelihood (ML) estimators retain strong properties under various censoring scenarios. Although ML estimators are asymptotically unbiased, they often exhibit non-negligible finite-sample bias, and closed-form expressions for their sampling variances are typically unavailable. Consequently, simulation-based tools have played a crucial role in examining their finite-sample behavior [19]. Interval estimation, which is critical in many practical settings, has been studied by Antle, Klimko, and Harkness [2] and later refined by Schafer and Sheffield [27] that developed exact small-sample interval procedures.

The bootstrap provides a flexible and broadly applicable framework for estimating the sampling distribution of estimators when analytic derivations are difficult or impossible. Introduced by Efron in 1979 [10], the bootstrap quickly became a cornerstone of modern statistical inference. The nonparametric bootstrap generates subsamples with replacement from the empirical distribution, enabling the construction of confidence intervals and the estimation of standard errors. Efron [12] later proposed various bootstrap confidence interval methods—including the bootstrap-t, percentile, and bias-corrected and accelerated (BC_a) intervals—addressing skewness and bias in resampling distributions. Subsequently, numerous studies refined bootstrap methodologies, such as Babu and Singh [3], Efron and Tibshirani [13], and DiCiccio and Romano [9]. Despite their versatility, traditional bootstrap methods may perform poorly in small-sample scenarios because the empirical distribution may inadequately represent the underlying population. To address this limitation, a new resampling scheme was developed by Xu, Li, and Fang (2024, [31]), which is based on the empirical distribution from a set of representative points (RPs) from the underlying population distribution. The idea is similar to that for the traditional bootstrap sampling, which is a simple random sampling (or equal probability sampling) from the original empirical distribution function constructed from the ordered sample, while RP-sampling is an unequal probability resampling obtaining samples from the empirical distribution function constructed from the RPs of the underlying distribution.

The RP idea was first introduced by Cox [8], who used a finite set of points to approximate the normal distribution for minimizing mean squared error. Max [21] and Anderberg [1] expanded this concept through quantization techniques for univariate distributions, an approach further developed in multivariate settings by Bofinger [7], Fang [14], and Zador [32]. A systematic algorithm for generating representative points, known as the Fang–He algorithm, was later proposed by Fang and He [15]. Their algorithm can be applied to any continuous probability distributions with finite second moments. A major theoretical advancement was recently made by Xu, Li, and Fang [31]. They established a rigorous framework for resampling via representative points. Their results show that RP-based approximating distributions can achieve higher-order moment convergence and provide a consistent approximation to sampling distributions under both Kolmogorov and Mallows–Wasserstein metrics. This theoretical development opens a new direction for improving bootstrap accuracy, especially in small-sample settings where traditional Monte Carlo resampling may be less accurate.

Motivated by the theoretical foundation for RP-based resampling in [31], we propose an RP-bootstrap method tailored for inference on the logistic distribution in this paper. Section 2 gives a simple description on the location-scale logistic distribution family and its representative points. Section 3 describes the method of RP-based resampling and the construction of RP-based confidence interval (CI) and some ordinary bootstrap-based CIs. Section 4 presents the Monte Carlo study on the comparisons between our proposed RP-bootstrap resampling and some ordinary bootstrap CIs using the criteria of coverage probability and the width of CI. Some concluding remarks are given in the last section.

2. THE LOGISTIC DISTRIBUTION AND ITS REPRESENTATIVE POINTS

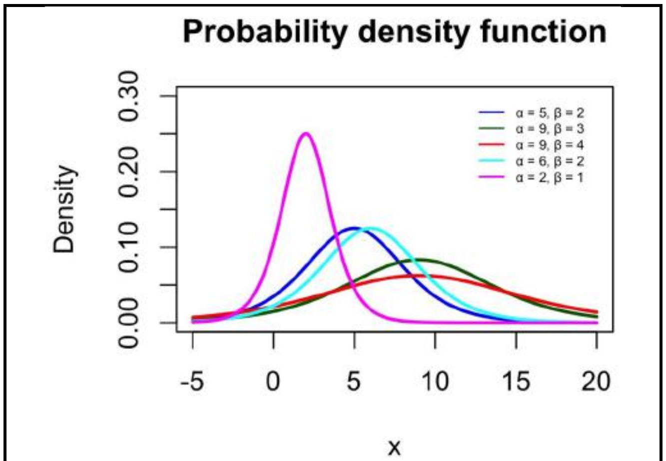
2.1. The Logistic Distribution and ML Estimation

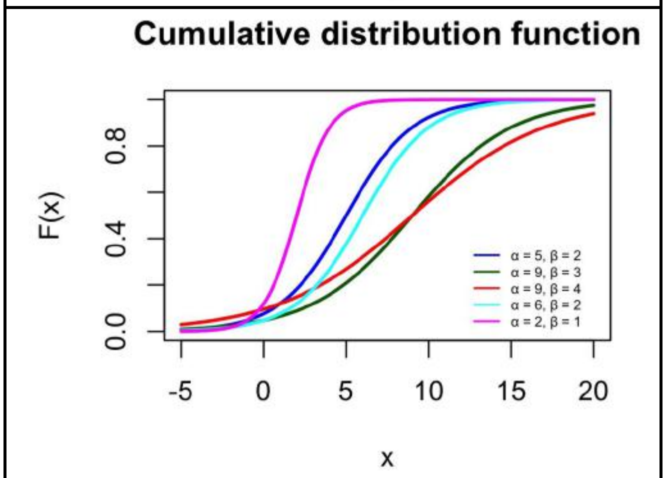
A random variable X with a location-scale logistic distribution, denote by $X\sim\text{Logistic}(\alpha,\beta)$, has the basic properties summarized in the following table.

It is noted that the maximum likelihood estimators (MLE) $\hat{\alpha}$ and $\hat{\beta}$ for the location and scale parameters α and β, respectively, have to be obtained by solving nonlinear equations and therefore no simple algebraic expressions are available.

2.2. Representative Points

Statistical representative points (RPs) provide a discrete approximation to a continuous probability distribution. They have been widely studied in





Parameters	α, location (real) $\beta > 0$, scale (real)
Support	$x \in (-\infty, \infty)$
PDF	$\frac{\exp\left(-\frac{x-\alpha}{\beta}\right)}{\beta\left(1+\exp\left(-\frac{x-\alpha}{\beta}\right)\right)}$
CDF	$\frac{1}{1+\exp\left(-\frac{x-\alpha}{\beta}\right)}$
Mean	α
Median	α
Mode	α
Variance	$\frac{\pi^2\beta^2}{3}$
$X = \alpha + \beta Z,$	$Z \sim \text{Logistic}(0,1)$

quantization and optimal discretization literature originating from Cox [8], Max [21], and later extended by Bofinger [7], Fang [14], and Zador [32]. For a continuous random variable X with a density function $f(\cdot)$, a set of RPs $\{R_1, \ldots, R_m : m \geq 1\}$ for f is a set of points that minimize the mean-squared-error (MSE) defined by the function

$$L(x_1, \ldots, x_m) = \frac{1}{\text{Var}(X)} \int_{-\infty}^{\infty} \min_{1 \leq i \leq m} (x - x_i)^2 f(x)\, dx \tag{1}$$

The corresponding weight p_i for each RP R_i is the probability given by

$$p_i = \int_{(R_{i-1}+R_i)/2}^{(R_i+R_{i+1})/2} f(x)\, dx, i = 1, \ldots, m, R_0 = -\infty, R_{m+1} = +\infty. \tag{2}$$

Fang–He algorithm [15] provides an algorithm for computing the MSE representative points for a broad class of distributions. The following website provides the RPs $\{R_1, \ldots, R_m\}$ and their associated probabilities $\{p_1, \ldots, p_m\}$ as defined by (2) for a number of standard location-scale probability distributions:

https://fst.uic.edu.cn/isci_en/Representative_Points/MSE_Representative_Points_for_Different_Statistica.htm.

A key property of RPs is that the discretized distribution Y_m defined by

Y_m	R_1	$\cdots$	R_m	
Probability	p_1	$\cdots$	p_m	(3)

preserves the population mean and has smaller variance than the original distribution [31]:

$$E(Y_m) = E(X), \text{Var}(Y_m) \leq \text{Var}(X).$$

This property provides the possibility of higher accuracy when sampling from the distribution of Y_m compared with sampling from the original distribution of *X*. In practice, the RPs for the general location-scale logistic distribution Logistic(α, β) are estimated from those of the standardized logistic distribution [16]:

$$\hat{R}_i = \hat{\alpha} + \hat{\beta} R_i, i = 1, \ldots, m, \tag{4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the ML estimates, $\{R_i : i = 1, \ldots, m\}$ is a set of RPs from Logistic(0,1) and $\{\hat{R}_i : i = 1, \ldots, m\}$ is taken as a set of approximate RPs from Logistic(α, β). Th

3. METHODOLOGY OF THE RP-BOOTSTRAP

The RP-bootstrap from a location-sale distribution is a probability-resampling procedure that draws samples from the approximated discrete distribution:

$\hat{Y}_m$	$\hat{R}_1$	$\cdots$	$\hat{R}_m$	
Probability	p_1	$\cdots$	p_m	(5)

where $\{\hat{R}_i : i = 1, \ldots, m\}$ is a set of estimated RPs from (4). The discrete distribution defined by (5) can be considered as an estimated version of the traditional empirical distribution (3) from the original underlying distribution. Classical bootstrap theory focuses on resampling from the empirical distribution (3) of an i.i.d. (independent identically distributed) sample. Recent work by Xu, Li, and Fang [31] established strong convergence properties for RP-based resampling from (5). It provides the theoretical foundation for the RP-bootstrap procedure.

3.1. Algorithm

Let $\{x_1, \dots, x_n\}$ be a set of i.i.d. sample from Logistic(α, β). The RP-bootstrap is a probability resampling process that takes samples from the estimated RP-empirical distribution (5) with probability p_i to select the RP $\hat{R}_i$ $(i = 1, \dots, m)$ for any given number of RPs *m*. The algorithm for the resampling is summarized as follows.

Step 1: Estimating MSE-optimal RPs from the sample

The ML estimates of the location and scale parameters are first obtained using the sample data from the general logistic density function by solving two simultaneous nonlinear equations. The MSE-optimal representative points $R_1, \dots, R_m$ for the standard logistic distribution $\text{Logistic}(0,1)$ are obtained from the website as given in subsection 2.2. The estimated RPs from Logistic(α, β) are obtained by (4) with their associated probabilities $p_1, \dots, p_m$ that are given when computing the RPs for $\text{Logistic}(0,1)$, and

$$\hat{p}_i = \int_{\frac{\hat{R}_{i-1}+\hat{R}_i}{2}}^{\frac{\hat{R}_i+\hat{R}_{i+1}}{2}} f\left(\frac{x-\hat{\alpha}}{\hat{\beta}}\right) dx = \int_{\hat{\alpha}+\hat{\beta}\frac{R_{i-1}+R_i}{2}}^{\hat{\alpha}+\hat{\beta}\frac{R_i+R_{i+1}}{2}} f\left(\frac{x-\hat{\alpha}}{\hat{\beta}}\right) dx$$
$$= \int_{\frac{R_{i-1}+R_i}{2}}^{\frac{R_i+R_{i+1}}{2}} f(y)dy = p_i \qquad (6)$$

for $i = 1, \dots, m, R_0 = -\infty, R_{m+1} = +\infty$, where $f(y) = \frac{e^{-y}}{1+e^{-y}}$ $(-\infty < y < +\infty)$ is the density function of the standard logistic distribution Logistic(0,1). Equation (6) implies that the probability weight for the estimated RP $\hat{R}_i$ is the same as that for R_i $(i = 1, \dots, m)$. Based on equations (4)-(6), it is easy to obtain the estimated RPs $\{\hat{R}_i : i = 1, \dots, m\}$ and their associated probabilities $\{p_1, \dots, p_m\}$ using the website in subsection 2.2.

Step 2: Generate weighted RP-bootstrap samples

An RP-bootstrap sample is obtained by drawing *n* observations from the RP-estimated empirical distribution (5) with probability p_i to select the RP $\hat{R}_i$ $(i = 1, \dots, m)$. This can be realized by generating $U \sim \text{Uniform}(0,1)$ and selecting $\hat{R}_i$ if

$$\sum_{j=1}^{i-1} p_j \leq U < \sum_{j=1}^{i} p_j \quad (p_0 = 0). \qquad (7)$$

Repeating this sampling step *n* times yields one RP-bootstrap sample with size *n*. For each RP-bootstrap sample with size *n*, repeating the same procedure *B* times gives *B* sets of RP-bootstrap samples with size *n* for each set, denote by $S^{*1}, \dots, S^{*B}$.

Step 3: Construct confidence intervals

For each RP-bootstrap sample, the statistic of interest (mean or variance) is recomputed to form the bootstrap distribution. Confidence intervals are constructed using any of the standard bootstrap methods, including the bootstrap-t, percentile, or BC_a intervals [12-13]. To evaluate the empirical performance of each bootstrap sampling method, the bootstrap procedure is repeated N Monte Carlo replications. Coverage probability (CP) and expected interval length (EL) are computed by

$$CP = \frac{\#\{\text{intervals containing the true value}\}}{N} \text{ and } EL = \frac{1}{N}\sum_{i=1}^{N}(UCL_i - LCL_i), \qquad (8)$$

respectively, where UCL_i stands for the upper confidence limit, and LCL_i for the lower confidence limit from the *i*-th bootstrap sample.

3.2. RP-bootstrap Confidence Intervals

3.2.1. RP Bootstrap-t Interval

For each RP-bootstrap sample S^{*b}, $b = 1, \dots, B$, we compute the RP-bootstrap estimate of the mean or variance, denoted $\hat{\theta}^*(b)$, and its estimated standard error $s_e^*(b)$. The bootstrap-t statistic is then

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}_n}{s_e^*(b)}, \qquad (9)$$

where $\hat{\theta}_n$is the corresponding estimate (sample mean or sample variance) from the original data. Let $z^{*(1-\alpha)}$and $z^{*(\alpha)}$be the empirical $(1-\alpha)$ and α quantiles of the $Z^*(b)$ values. The RP bootstrap-t confidence interval for $\theta = E(X)$ or $\theta = \text{Var}(X)$ is [13]:

$$[\hat{\theta}_n - z^{*(1-\alpha)}\widehat{\text{se}}, \hat{\theta}_n - z^{*(\alpha)}\, \widehat{\text{se}}], \qquad (10)$$

where $\widehat{\text{se}}$ is the standard error of the estimator computed from the nonparametric bootstrap sample (see page 46, equation (6.3) in [13]). For logistic distribution $X \sim$Logistic(α, β), both $E(X)$ and Var(X) have simple closed-form as given in the table in section 2.1, the bootstrap-t intervals are directly applicable to both parameters α and β.

3.2.2. RP Percentile Interval

The RP percentile interval is constructed directly from the empirical distribution of the RP-bootstrap estimates. Let $\hat{\theta}^*(1), \ldots, \hat{\theta}^*(B)$ be the RP-bootstrap estimates of the parameter of interest θ (mean or variance). Denote by $\hat{G}$ the empirical distribution function of these bootstrap estimates. The RP percentile interval is defined by [13]:

$$[\hat{G}^{-1}(\delta), \hat{G}^{-1}(1-\delta)], \tag{11}$$

where $\hat{G}^{-1}(\cdot)$ denotes the bootstrap quantile function, and $0 < \delta < 1$ is any given significance level. This method is simple to implement and requires no studentization or standard-error estimation.

3.2.3. RP BC$_a$ Interval

The BC$_a$ interval improves the percentile interval by adjusting both bias and skewness in the bootstrap distribution [13]. Using the RP-bootstrap estimates $\hat{\theta}^*(b)$, the BC$_a$ interval is defined as

$$[\hat{G}^{-1}(\alpha_1), \hat{G}^{-1}(\alpha_2)], \tag{12}$$

where $\hat{G}^{-1}(\cdot)$ is the same as in (11). The adjusted percentiles α_1 and α_2 are given by the standard BC$_a$ formulas:

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\delta)}}{1 - \hat{a}(\hat{z}_0 + z^{(\delta)})}\right),$$
$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\delta)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\delta)})}\right) \tag{13}$$

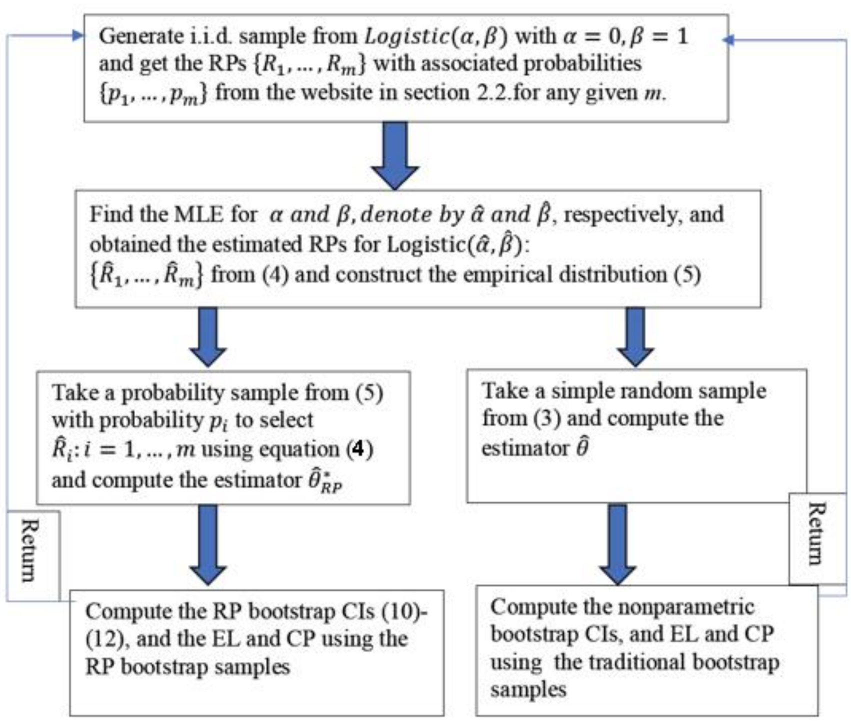
where $\Phi(\cdot)$ is the standard normal CDF (cumulative distribution function) and $z^{(\delta)}$ is the normal δ-quantile satisfying $\Phi(z^{(\delta)}) = \delta, (0 < \delta < 1)$, $\hat{a}$ and $\hat{z}_0$ are estimated through equations (14.9)-(14.15) in Efron [13] (Chapter 14, pp. 185-186).

The above algorithm can be summarized by the workflow as follows.

3.3. Comparison with the Nonparametric Bootstrap

The simulation procedure for the nonparametric bootstrap is similar to that of the RP-bootstrap, except for the source of resampling. The RP-bootstrap draws weighted samples from the estimated representative points defined by (4)-(5), whereas the nonparametric bootstrap draws simple random samples directly from the empirical distribution of the original data. Figure **1** illustrates the approximate sampling distributions generated by the two methods. For the RP-bootstrap, the distribution of a statistic $T(X_1, \ldots, X_n; F)$ under F is approximated by the distribution of $T(Y_1, \ldots Y_m; F_{mse,m})$ under $F_{mse,m}$, where $F_{mse,m}$ is an approximating

Algorithm flow chart



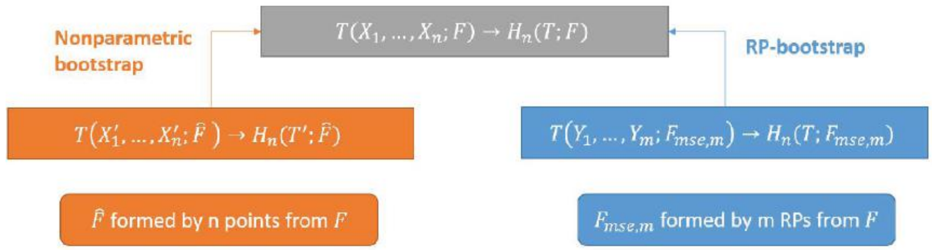


Figure 1: Approximate sampling distribution comparison.

distribution formed by m representative points that minimize the mean squared error. For the nonparametric bootstrap, the distribution of $T(X_1, \ldots, X_n; F)$ under F is approximated by the distribution of $T(X'_1, \ldots X'_n; \hat{F})$ under the empirical distribution $\hat{F}$. Figure **2** describes their respective simulation workflows. For the RP-bootstrap, we first estimate the distribution parameters from the sample S using maximum likelihood estimation to obtain the transformed RPs. We then draw weighted bootstrap resamples from these transformed RPs to approximate the sampling distribution of the test statistic and construct confidence intervals. For the nonparametric bootstrap, we draw equal-weight bootstrap resamples directly from the original sample to approximate the sampling distribution of the test statistic and construct confidence intervals.

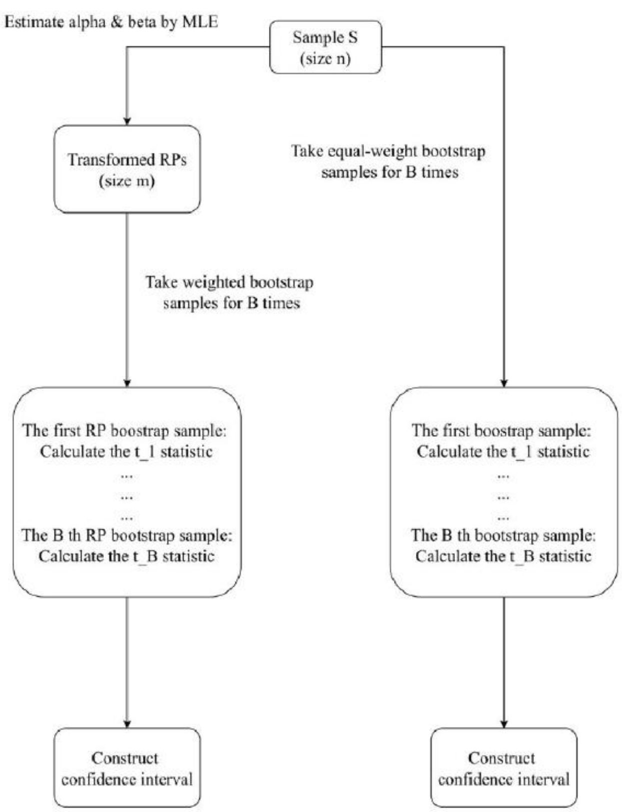


Figure 2: Comparison of the simulation procedure.

4. MONTE CARLO COMPARISONS

Without loss of generality, we conducted Monte Carlo simulations to assess the finite-sample performance of the proposed RP-bootstrap under the standard logistic distribution. For each sampling design, $N = 500$ samples were generated with sample sizes n from $\{5,10,15,20,30,50,100,150,200\}$. Although there is no definite rule for determining the number of RPs in practical application, we chose the number of RPs that is equal to the sample size for a fair comparison between the RP-bootstrap and the traditional nonparametric bootstrap. For every sample, we constructed 95% confidence intervals for the mean $E[X]$ and the variance $\mathrm{Var}(X)$ using both the RP-bootstrap and the nonparametric bootstrap, with a bootstrap sample size of $B = 1000$. Both bootstrap-t and percentile intervals were computed. The estimated RPs were computed by (4) with MLEs for the location and scale parameters from solving two simultaneous nonlinear equations following the traditional maximum likelihood approach. Coverage probability (CP) and expected interval length (EL) were obtained by (8). Table **1** summarizes the simulation outcomes for the location parameter with the comparisons (CP and EL) between RP-bootstrap and Efron's original nonparametric bootstrap based on three types of confidence intervals: Bootstrap-t, Percentile-bootstrap, and BC_a-bootstrap. Table **2** summarizes the simulation outcomes for the scale parameter with the same types of comparison as in Table **1**. The following observations can be summarized.

From Table **1**:

1) The RP-bootstrap-t ELs for the location parameter are generally smaller than those of the nonparametric bootstrap-t EL, the negative percentages at the row "% difference (RP vs NP)" indicate the shortened EL for each case, implying the RP-bootstrap-t dominates over the nonparametric bootstrap-t of the CI (confidence interval) in the sense of EL;

Table 1: Coverage Probability and Expected Length (*θ* = *E(X)*)

Metric	n=5	n=10	n=15	n=20	n=30	n=50	n=100	n=150	n=200
Bootstrap-t									
EL (RP-bootstrap)	2.202	1.745	1.447	1.270	1.057	0.840	0.592	0.486	0.417
EL (Nonparametric)	2.321	1.800	1.454	1.271	1.041	0.846	0.594	0.487	0.413
% difference (RP vs NP)	-5.127%	-3.056%	-0.481%	-0.079%	1.537%	-0.709%	-0.337%	-0.205%	0.969%
CP (RP-bootstrap)	0.812	0.838	0.888	0.868	0.904	0.904	0.894	0.934	0.886
CP (Nonparametric)	0.894	0.980	0.986	0.990	0.998	1.000	1.000	1.000	1.000
% difference (RP vs NP)	-9.172%	-14.490%	-9.939%	-12.323%	-9.419%	-9.600%	-10.600%	-6.600%	-11.400%
Percentile									
EL (RP-bootstrap)	2.202	1.745	1.447	1.270	1.057	0.840	0.592	0.486	0.417
EL (Nonparametric)	2.181	1.723	1.433	1.260	1.041	0.833	0.592	0.485	0.416
% difference (RP vs NP)	0.963%	1.277%	0.977%	0.794%	1.537%	0.840%	0.000%	0.206%	0.240%
CP (RP-bootstrap)	0.808	0.838	0.888	0.870	0.888	0.916	0.902	0.942	0.884
CP (Nonparametric)	0.788	0.820	0.862	0.828	0.866	0.896	0.884	0.932	0.864
% difference (RP vs NP)	2.538%	2.195%	3.016%	5.072%	2.540%	2.232%	2.036%	1.073%	2.315%
BCa									
EL (RP-bootstrap)	2.198	1.745	1.446	1.270	1.058	0.840	0.593	0.486	0.417
EL (Nonparametric)	2.239	1.761	1.450	1.273	1.047	0.835	0.592	0.485	0.416
% difference (RP vs NP)	-1.831%	-0.909%	-0.276%	-0.236%	1.051%	0.599%	0.169%	0.206%	0.240%
CP (RP-bootstrap)	0.808	0.844	0.886	0.870	0.880	0.910	0.902	0.946	0.892
CP (Nonparametric)	0.774	0.804	0.860	0.808	0.856	0.892	0.878	0.926	0.862
% difference (RP vs NP)	4.393%	4.975%	3.023%	7.673%	2.804%	2.018%	2.733%	2.160%	3.480%

Note:

3) negative percentages in EL imply shorter EL from RP-bootstrap CI. Positive percentages in CP imply higher CP from RP-bootstrap CI.
4) The CPs in the nonparametric bootstrap t-CIs (they reach 100% for large sample size $n \geq 50$) are generally higher than those of RP bootstrap t-CI. This is caused by the fact that standard error from simple random sampling is always larger than that of the RP-bootstrap sampling [31], resulting in wider CIs that can cover the true parameter with 100%. This means higher CP sacrifices the accuracy in the sense of EL.

2) The RP-bootstrap-t CPs for the location parameter are generally smaller than those of the nonparametric bootstrap-t, the negative percentages at the row "% difference (RP vs NP)" indicate the decreased CP for each case, implying the RP-bootstrap-t CIs are weaker than those of the nonparametric bootstrap-t CIs in the sense of CP. This validates the general principle that a shorter CI usually results in a smaller CP;

5) The RP-bootstrap percentile ELs for the location parameter are generally comparable with those of the nonparametric percentile, the positive percentages at the row "% difference (RP vs NP)" indicate the increased EL for each case, implying the RP-bootstrap percentile cannot improve the nonparametric percentile in constructing CIs for the location parameter in the sense of EL;

6) The RP-bootstrap percentile CPs are noticeably higher than those of the nonparametric percentile, the positive percentages at the row "% difference (RP vs NP)" indicate the increased CP for each case, implying the RP-bootstrap percentile noticeably dominates over the nonparametric percentile in the sense of CP. This validates the general principle that a wider CI usually results in a higher CP;

7) The RP-bootstrap BC_a ELs are slightly smaller than those of the nonparametric BC_a in the small sample cases ($n < 30$), the negative percentages at the row "% difference (RP vs NP)" indicate the shortened EL

for the small sample cases of $n < 30$, while the positive percentages at the row "% difference (RP vs NP)" indicate the increased EL for the sample cases $(n \geq 30)$, implying the RP-bootstrap BC_a dominates over the nonparametric BC_a in the sense of EL under small sample sizes $(n < 30)$. For the sample sizes of $n \geq 30$, the two methods are almost comparable with the negligible percentage change;

10) The RP-bootstrap BC_a's CPs for the location parameter are generally higher than those of the nonparametric bootstrap BC_a, the positive percentages at the row "% difference (RP vs NP)" indicate the increased CP for each case implying the RP-bootstrap BC_a generally improves the CIs compared to the nonparametric bootstrap BC_a in the sense of CP.

From Table **2**, it can be observed that the RP-bootstrap shows remarkable improvements over the nonparametric bootstrap (NP) for all sample designs for the two types of CIs: percentile bootstrap and BC_a in the sense of CP but RP-bootstrap displays some weaknesses in the ELs of these two types of CIs with observed positive percentage changes. The RP-bootstrap t-CIs are generally worse than NP CIs with most positive percentage changes in the ELs and negative percentage changes in the CPs, indicating the RP-bootstrap t-CI is not recommended for constructing CIs for the logistic population variance.

In addition to the EL and CP comparison between the RP bootstrap and the traditional bootstrap, computational complexity and run time comparisons can be also carried out. For the simple location-scale logistic distribution, we found no noteworthy run time difference between the two methods because RP bootstrap replaces the simple random sampling in traditional bootstrap with a probability sampling. The computational complexity from the RP bootstrap comes from computing the RPs, but this complexity can be solved by the aid of the website in section 2.2.

Table 2: Coverage Probability and Expected Length (θ = var(X))

Metric	n=5	n=10	n=15	n=20	n=30	n=50	n=100	n=150	n=200
Bootstrap-t									
EL (RP-bootstrap)	25.791	12.111	8.385	6.578	4.886	3.524	2.238	1.752	1.470
EL (Nonparametric)	157.544	13.277	7.983	6.473	4.346	3.187	2.115	1.669	1.440
% difference (RP vs NP)	-83.620%	-8.780%	5.030%	1.620%	12.470%	10.550%	5.820%	4.970%	2.080%
CP (RP-bootstrap)	0.958	0.968	0.980	0.970	0.966	0.952	0.930	0.928	0.938
CP (Nonparametric)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
% difference (RP vs NP)	-4.200%	-3.200%	-2.000%	-3.000%	-3.400%	-4.800%	-7.000%	-7.200%	-6.200%
Percentile									
EL (RP-bootstrap)	5.746	5.215	4.419	3.928	3.329	2.714	1.915	1.567	1.347
EL (Nonparametric)	4.470	4.195	3.711	3.443	2.865	2.486	1.809	1.508	1.289
% difference (RP vs NP)	28.550%	24.310%	19.080%	14.090%	16.210%	9.170%	5.860%	3.910%	4.500%
CP (RP-bootstrap)	0.654	0.790	0.832	0.860	0.888	0.910	0.890	0.910	0.914
CP (Nonparametric)	0.524	0.694	0.712	0.750	0.782	0.836	0.852	0.866	0.874
% difference (RP vs NP)	24.810%	13.830%	16.850%	14.670%	13.560%	8.850%	4.460%	5.080%	4.580%
BCa									
EL (RP-bootstrap)	6.008	5.998	4.930	4.321	3.567	2.849	1.963	1.594	1.363
EL (Nonparametric)	4.491	4.454	3.951	3.646	3.009	2.585	1.855	1.537	1.309
% difference (RP vs NP)	33.780%	34.680%	24.790%	18.550%	18.550%	10.200%	5.820%	3.710%	4.130%
CP (RP-bootstrap)	0.692	0.838	0.864	0.886	0.912	0.914	0.896	0.906	0.922
CP (Nonparametric)	0.560	0.714	0.740	0.772	0.824	0.836	0.862	0.864	0.896
% difference (RP vs NP)	23.570%	17.370%	16.760%	14.770%	10.680%	9.330%	3.940%	4.860%	2.900%

Note:

8) Negative percentages in EL imply shorter EL from RP-bootstrap CI. Positive percentages in CP imply higher CP from RP-bootstrap CI.

9) The CPs in the nonparametric bootstrap t-CIs (they reach 100% for all sample sizes) are generally higher than those of RP bootstrap t-CI. This is caused by the fact that standard error from simple random sampling is always larger than that of the RP-bootstrap sampling [31], resulting in wider CIs that can cover the true parameter with 100%. This means higher CP sacrifices the accuracy in the sense of EL.

4. ILLUSTRATIVE EXAMPLES

Example 1. The data set consists of the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli. The data set is obtained from the work of Usman, Haq, and Talib [30]. The dataset was originally analyzed under a log-logistic distribution; therefore, we apply a logarithmic transformation so that the transformed data more closely follow a logistic distribution. The original data set is as follows: 0.1, 0.33, 0.44, 0.56, 0.59, 0.72, 0.74, 0.77, 0.92, 0.93, 0.96, 1.0, 1.0, 1.0, 1.02, 1.05, 1.07, 1.08, 1.08, 1.08, 1.09, 1.12, 1.13, 1.15, 1.16, 1.2, 1.21, 1.22, 1.22, 1.24, 1.3, 1.34, 1.36, 1.39, 1.44, 1.46, 1.53, 1.59, 1.6, 1.63, 1.63, 1.68, 1.71, 1.72, 1.76, 1.83, 1.95, 1.96, 1.97, 2.02, 2.13, 2.15, 2.16, 2.22, 2.3, 2.31, 2.4, 2.45, 2.51, 2.53, 2.54, 2.54, 2.78, 2.93, 3.27, 3.42, 3.47, 3.61, 4.02, 4.32, 4.58, 5.55.

The sample mean and variance of the transformed data are 0.425 and 0.433 respectively. We construct the RP-bootstrap and nonparametric 95% confidence

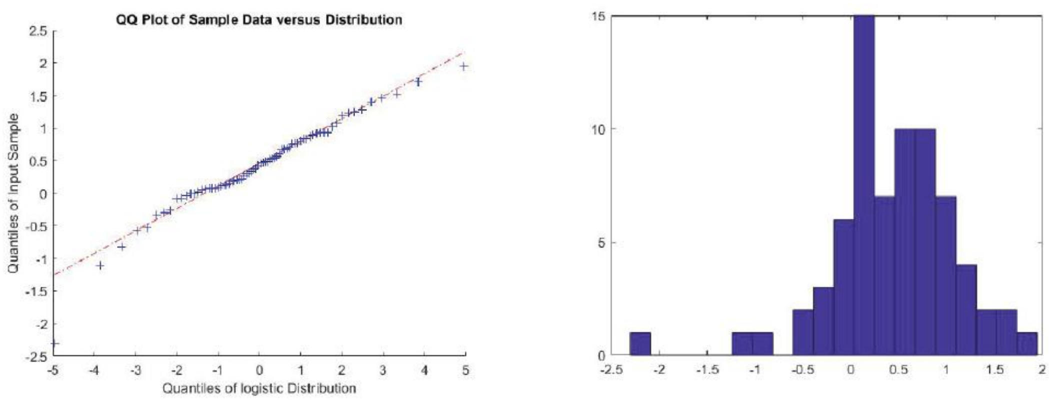


Figure 3: The Q-Q plot and histogram of the transformed pig data set.

Table 3: Results of the Pig Data Set

Metric	LCL	UCL	EL = UCL − LCL
Bootstrap-t			
E(X): RP-bootstrap	0.303	0.548	0.246
E(X): Nonparametric	0.278	0.523	0.245
E(X): % diff (RP vs NP)	-	-	+0.408%
Var(X): RP-bootstrap	0.256	0.568	0.313
Var(X): Nonparametric	0.268	0.800	0.532
Var(X): % diff (RP vs NP)	-	-	-41.165%
Percentile			
E(X): RP-bootstrap	0.292	0.538	0.246
E(X): Nonparametric	0.274	0.518	0.245
E(X): % diff (RP vs NP)	-	-	+0.408%
Var(X): RP-bootstrap	0.249	0.503	0.254
Var(X): Nonparametric	0.248	0.595	0.347
Var(X): % diff (RP vs NP)	-	-	-26.796%
BC_a			
E(X): RP-bootstrap	0.283	0.531	0.248
E(X): Nonparametric	0.259	0.511	0.253
E(X): % diff (RP vs NP)	-	-	-1.976%
Var(X): RP-bootstrap	0.261	0.520	0.259
Var(X): Nonparametric	0.260	0.623	0.363
Var(X): % diff (RP vs NP)	-	-	-28.646%

intervals for the underlying population mean and variance. The results in Table **3** show that the RP-bootstrap consistently produces narrower confidence intervals than those of the nonparametric bootstrap. Figure **3** shows the histogram of the data and the Q-Q (quantile-quantile) plot using the logistic quantiles from $L(\hat{\alpha}, \hat{\beta})$, where $\hat{\alpha}$ and $\hat{\beta}$ are the MLEs of the location and scale parameter from the sample data, respectively. It shows that the logistic distribution $L(\hat{\alpha}, \hat{\beta})$ is a good fit for the data, indicating the logistic RP-bootstrap resampling approach makes sense in comparing with the nonparametric bootstrap approach.

The confidence intervals in Table **3** show that the RP-bootstrap t-CI, percentile CI, and the BC_a CI for the population mean have about the same EIs as those from the traditional nonparametric bootstrap CIs but RP-bootstrap significantly reduced the EIs for all three types of CIs. This means that traditional nonparametric bootstrap method is very likely to overestimate the population variance. An over-estimated population variance can lead to overly conservative confidence intervals—specifically, intervals that are wider than necessary. This reduces the precision of the estimate and inflates the interval's coverage probability beyond the nominal level (e.g., exceeding 95%), potentially masking true effects or differences. This issue is particularly pronounced in percentile-based and basic bootstrap intervals, as they directly rely on the spread of the bootstrap distribution, which may itself be inflated if the original sample overestimates the true population variance. Over-estimation of variance in a biological or medical context is not just a statistical error; it directly impacts scientific conclusions and clinical decision-making with potentially serious consequences. The core implication is a loss of statistical power and an increased risk of Type II errors (failing to detect a true effect). Examples may include:

1) clinical drug trials: In a Phase III trial for a new oncology drug, an over-estimated variance in tumor shrinkage (the outcome measure) would produce an overly wide confidence interval for the treatment effect. This could cause the interval to cross the "no effect" line (e.g., zero difference), leading to the false conclusion that the drug is not statistically significantly better than the standard of care. A potentially life-extending therapy might be abandoned or delayed due to this miscalculation.

2) biomarker discovery: In a genomic study seeking to identify genes differentially expressed between healthy and diseased tissue, over-estimated variance within groups inflates the standard error. This makes it harder for statistical tests to distinguish true signal from noise. As a result, truly important biomarker genes may fail to reach significance and be filtered out of downstream analysis, halting a promising diagnostic or therapeutic research pathway.

In essence, over-estimation of variance acts as a conservative bias that obscures real biological signals, wasting resources, missing opportunities for medical advancement, and ultimately failing patients who could benefit from more accurate research.

Example 2. The dataset consists of the heights (to the nearest inch) of 12 pollen-fertile maize plants (Bliss [5]). The original dataset is shown as follows: 92, 107, 98, 97, 95, 94, 92, 96, 98, 104, 97, 89. This is a small-sample dataset. The sample mean and variance of are 96.588 and 25.174 respectively. The results in Table 4 show the RP-bootstrap produces slightly narrower confidence intervals than the nonparametric bootstrap, indicating improved interval efficiency. Figure 4 shows the histogram of the data and the Q-Q plot using the logistic quantiles from $L(\hat{\alpha}, \hat{\beta})$. Similar to Figure 3, the

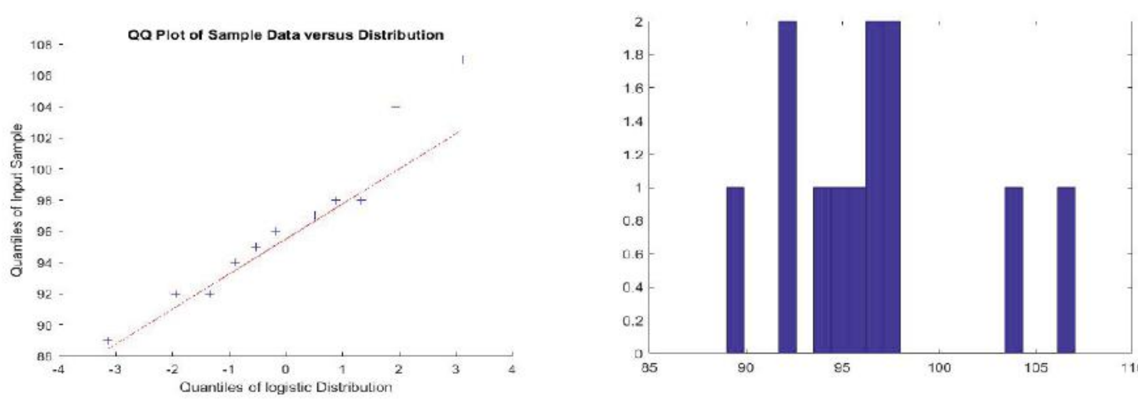


Figure 4: The Q-Q plot and histogram of the plant data set.

Table 4: Results of the Plant Data Set

Metric	LCL	UCL	EL = UCL − LCL
Bootstrap-t			
E(X): RP-bootstrap	4.544	4.588	0.044
E(X): Nonparametric	4.547	4.595	0.048
E(X): % diff (RP vs NP)	-	-	-8.333%
Var(X): RP-bootstrap	0.001	0.008	0.007
Var(X): Nonparametric	0.001	0.008	0.007
Var(X): % diff (RP vs NP)	-	-	+0.000%
Percentile			
E(X): RP-bootstrap	4.543	4.587	0.044
E(X): Nonparametric	4.544	4.591	0.048
E(X): % diff (RP vs NP)	-	-	-8.333%
Var(X): RP-bootstrap	0.001	0.005	0.004
Var(X): Nonparametric	0.001	0.004	0.003
Var(X): % diff (RP vs NP)	-	-	+33.333%
BC_a			
E(X): RP-bootstrap	4.545	4.589	0.044
E(X): Nonparametric	4.548	4.593	0.046
E(X): % diff (RP vs NP)	-	-	-4.348%
Var(X): RP-bootstrap	0.001	0.005	0.004
Var(X): Nonparametric	0.001	0.005	0.004
Var(X): % diff (RP vs NP)	-	-	+0.000%

Q-Q plot shows that the logistic distribution $L(\hat{\alpha}, \hat{\beta})$ is a feasible fit for the data, indicating the logistic RP-bootstrap resampling approach makes sense in comparing with the nonparametric bootstrap approach.

From Table **4**, it is observed that the ELs from RP-bootstrap mean CIs are improved remarkably with minimum 4.348% EL reduction, while the RP-bootstrap CIs for the population variance are about the same as those from the nonparametric bootstrap CIs except that for the percentile bootstrap with EL increase of about one thirds from the RP-bootstrap percentile CI. This observation is consistent with that from the Monte Carlo outcomes shown in Table 2: the RP-bootstrap percentile CIs generally exceed those of the nonparametric bootstrap CIs. This double validates the assertion that RP-bootstrap is not recommended for constructing the percentile bootstrap CIs.

5. CONCLUDING REMARKS

This paper introduces a representative point (RP) based bootstrap method for inference on the logistic distribution. Unlike the traditional nonparametric bootstrap, which resamples directly from the empirical distribution, the RP-bootstrap draws samples from an approximate distribution constructed using mean squared error (MSE)-optimal representative points. This yields a more structured and stable resampling foundation, particularly for small samples. The method's performance was evaluated using bootstrap-t, percentile, and BC_a methods to construct confidence intervals for the location and scale parameters. The contribution of our paper can be summarized from three aspects:

5.1. Methodological Contribution

Our primary contribution is the novel RP-bootstrap framework, which mitigates the instability inherent in direct empirical resampling. By leveraging an optimal representative-point approximation of the underlying distribution, the method provides a computationally efficient and theoretically grounded alternative. It is theoretically established that the sampling distribution from the RP empirical distribution has a smaller variance than that from the traditional empirical distribution.

5.2. Key Empirical Findings

Extensive Monte Carlo simulations and real-data applications reveal several key outcomes:

a. For small samples ($n < 30$), the RP-bootstrap consistently achieves higher coverage probabilities while maintaining shorter or comparable interval lengths, effectively countering the under-coverage tendency of the nonparametric bootstrap.

b. The improvement is most pronounced for the bootstrap‑t interval but is observable across all three interval construction methods.

c. Gains are consistent for both parameters but more marked for the location parameter.

d. While the RP-bootstrap offers remarkable improvements in expected length (EL), this can come at a slight cost to coverage probability (CP), a typical trade-off for shorter intervals.

e. For moderate to large samples ($n \geq 50$), the RP-bootstrap remains competitive but the necessity of its use diminishes, as performance converges with conventional methods.

5.3. Practical Recommendations

For applied researchers working with logistic models, we recommend the RP-bootstrap for constructing confidence intervals, especially when sample sizes are limited ($n < 30$), as it provides superior accuracy and reliability without sacrificing computational efficiency. In larger-sample scenarios, it remains a robust and viable option.

Future work should extend this approach to other probability distributions and more complex inferential settings. A critical open question is the method's robustness to model misspecification. We conjecture that the RP-bootstrap will remain competitive even under slight departures from the logistic model due to its lower-variance foundation; however, further research is necessary to formally investigate and validate this conjecture.

REFERENCES

[1] Anderberg MR. The broad view of cluster analysis. Cluster Analysis for Applications 1973; 1(1): 1-9. https://doi.org/10.1016/B978-0-12-057650-0.50007-7

[2] Antle C, Klimko L, Harkness W. Confidence intervals for the parameters of the logistic distribution. Biometrika 1970; 57(2): 397-402. https://doi.org/10.1093/biomet/57.2.397

[3] Babu GJ, Singh K. Inference on means using the bootstrap. The Annals of Statistics 1983; 11(3): 999-1003. https://doi.org/10.1214/aos/1176346267

[4] Berkson J. Application of the logistic function to bio-assay. Journal of the American Statistical Association 1944; 39(227): 357-365. https://doi.org/10.1080/01621459.1944.10500699

[5] Bliss, CI. Statistics in Biology: Statistical Methods for Research in the Natural Sciences (Vol. 1). McGraw-Hill, New York, 1967.

[6] Blom G. Statistical estimates and transformed beta-variables (Doctoral dissertation). Almqvist & Wiksell, 1958.

[7] Bofinger E. Maximizing the correlation of grouped observations. Journal of the American Statistical Association 1970; 65(332): 1632-1638. https://doi.org/10.1080/01621459.1970.10481193

[8] Cox DR. Note on grouping. Journal of the American Statistical Association 1957; 52(280): 543-547. https://doi.org/10.1080/01621459.1957.10501411

[9] DiCiccio TJ, Romano JP. On bootstrap procedures for second-order accurate confidence limits in parametric models. Statistica Sinica 1995: 141-160.

[10] Efron B. Bootstrap methods: another look at the jackknife. The Annals of Statistics 1979; 7: 1-26. https://doi.org/10.1214/aos/1176344552

[11] Efron B. The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, 1982. https://doi.org/10.1137/1.9781611970319

[12] Efron B. Better bootstrap confidence intervals. Journal of the American Statistical Association 1987; 82(397): 171-185. https://doi.org/10.1080/01621459.1987.10478410

[13] Efron B, Tibshirani RJ. An introduction to the Bootstrap. Chapman and Hall/CRC, 1993. https://doi.org/10.1007/978-1-4899-4541-9

[14] Fang KT. Application of the theory of the conditional distribution for the standardization of clothes. Acta Mathematicae Applicatae Sinica 1976; 2: 62-67.

[15] Fang KT, He SD. The problem of selecting a specified number of representative points from a normal population. Acta Mathematicae Applicatae Sinica 1984; 17: 293-306.

[16] Fang KT, Ye H, Zhou Y. Representative Points of Statistical Distributions -Applications in Statistical Inference. CRC Press, 2025. https://doi.org/10.1201/9781003589389

[17] Fisk PR. The graduation of income distributions. Econometrica: Journal of the Econometric Society 1961: 171-185. https://doi.org/10.2307/1909287

[18] Grizzle JE. A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics 1961; 17(3): 372-385. https://doi.org/10.2307/2527832

[19] Harter HL, Moore AH. Maximum likelihood estimation from censored samples of the parameters of a logistic distribution. Journal of the American Statistical Association 1967; 62(318): 675-684. https://doi.org/10.1080/01621459.1967.10482940

[20] Jung J. On linear estimates defined by a continuous weight function. Arkiv Mat 1954; 3: 199-209. https://doi.org/10.1007/BF02589406

[21] Max J. Quantizing for minimum distortion. IRE Transactions on Information Theory 1960; 6(1): 7-12. https://doi.org/10.1109/TIT.1960.1057548

[22] Meade N. A modified logistic model applied to human populations. Journal of the Royal Statistical Society （Series A，Statistics in Society) 1988; 151(3): 491-498. https://doi.org/10.2307/2982996

[23] Ogawa J. Contributions to the theory of systematic statistics. Osaka University, Institute of Statistical Mathematics, Japan, 1951.

[24] Oliver F. Methods of estimating the logistic growth function. Journal of the Royal Statistical Society: Series C (Applied Statistics) 1964; 13(2): 57-66. https://doi.org/10.2307/2985696

[25] Pearl R, Reed LJ. On the rate of growth of the population of the United States since 1790 and its mathematical representation. Proceedings of the National Academy of Sciences 1920; 6(6): 275-288. https://doi.org/10.1073/pnas.6.6.275

[26] Plackett RL. The analysis of life test data. Technometrics 1959; 1(1): 9-19. https://doi.org/10.1080/00401706.1959.10489845

[27] Schafer R, Sheffield T. Inferences on the parameters of the logistic distribution. Biometrics 1973; 449-455. https://doi.org/10.2307/2529168

[28] Schultz H. The standard error of a forecast from a curve. Journal of the American Statistical Association 1930; 25(170): 139-185. https://doi.org/10.1080/01621459.1930.10503117

[29] Stepanova M, Thomas L. Survival analysis methods for personal loan data. Operations Research 2002; 50(2): 277-289. https://doi.org/10.1287/opre.50.2.277.426

[30] Usman RM, Haq M, Talib J. Kumaraswamy half-logistic distribution: properties and applications. Journal of Statistical Applications and Probability 2017; 6: 597-609. https://doi.org/10.18576/jsap/060315

[31] Xu LH, Li Y, Fang KT. The resampling method via representative points. Statistical Papers. 2024; 65(6): 3621-3649. https://doi.org/10.1007/s00362-024-01536-2

[32] Zador PL. Development and Evaluation of Procedures for Quantizing Multivariate Distributions. Stanford University, 1964.

Received on 21-10-2025 Accepted on 22-11-2025 Published on 25-12-2025

https://doi.org/10.6000/1929-6029.2025.14.74