Median confidence regions in a nonparametric model

Edsel A. Peña[*] and Taeho Kim[†]

Department of Statistics
University of South Carolina
Columbia, SC 29208 USA
e-mail: pena@stat.sc.edu; taeho@email.sc.edu

Abstract: The nonparametric measurement error model (NMEM) postulates that $X_i = \Delta + \epsilon_i, i = 1, 2, \ldots, n; \Delta \in \Re$ with $\epsilon_i, i = 1, 2, \ldots, n$, IID from $F(\cdot) \in \mathfrak{F}_{c,0}$, where $\mathfrak{F}_{c,0}$ is the class of all continuous distributions with median 0, so Δ is the median parameter of X. This paper deals with the problem of constructing a confidence region (CR) for Δ under the NMEM. Aside from the NMEM, the problem setting also arises in a variety of situations, including inference about the median lifetime of a complex system arising in engineering, reliability, biomedical, and public health settings, as well as in the economic arena such as when dealing with household income. Current methods of constructing CRs for Δ are discussed, including the T-statistic based CR and the Wilcoxon signed-rank statistic based CR, arguably the two default methods in applied work when a confidence interval about the center of a distribution is desired. A 'bottom-to-top' approach for constructing CRs is implemented, which starts by imposing reasonable invariance or equivariance conditions on the desired CRs, and then optimizing with respect to their mean contents on subclasses of $\mathfrak{F}_{c,0}$. This contrasts with the usual approach of using a pivotal quantity constructed from test statistics and/or estimators and then 'pivoting' to obtain the CR. Applications to a real car mileage data set and to Proschan's famous air-conditioning data set are illustrated. Simulation studies to compare performances of the different CR methods were performed. Results of these studies indicate that the sign-statistic based CR and the optimal CR focused on symmetric distributions satisfy the confidence level requirement, though they tended to have higher contents; while three of the bootstrap-based CR procedures and one of the newly-developed adaptive CR tended to be a tad more liberal, but with smaller contents. A critical recommendation for practitioners is that, *under the NMEM*, the T-statistic based and Wilcoxon signed-rank statistic based CRs should **not** be used since they either have very degraded coverage probabilities or inflated contents under some of the allowable error distributions under the NMEM.

MSC 2010 subject classifications: Primary 62G15; secondary 62G09, 62G35.
Keywords and phrases: Bootstrap confidence region, BCa confidence region, confidence region, content of a confidence region, equivariant confidence region, expected values of differences of order statistics, invariant models, nonparametric confidence region, nonparametric measurement error model, randomized confidence region.

Received April 2019.

---

[*]E. Peña is Professor at the Department of Statistics, University of South Carolina.
[†]T. Kim is a PhD student at the Department of Statistics, University of South Carolina.

Contents

1. Introduction and motivation

Given a univariate distribution function H, the two most common measures of central tendency are the mean $\mu = \int x H(dx)$, provided it exists ($\int |x| H(dx) < \infty$), and the median $\Delta = \inf\{x \in \Re : \; H(x) \geq 1/2\}$. The mean need not always exist, whereas the median always exists. Under symmetric distributions, and when the mean exists, then the mean and the median coincide. This paper is concerned with making statistical inference about the median Δ of a distribution. A popular model leading to the problem of making inference about the median of a distribution is the so-called measurement error model. In this model Δ represents a quantity of interest which is unknown, and when one measures its value, the observed value x is a realization of the random variable

$$X = \Delta + \epsilon, \tag{1.1}$$

where ϵ represents a measurement error with a continuous distribution $F(\epsilon)$ whose median equals zero. As such, the distribution of X is $H(x) = F(x - \Delta)$. Typically, $F(\cdot)$ is assumed to be a zero-mean normal distribution, but this assumption may not tenable in many situations. For instance, in dealing with event times in biomedical, reliability, engineering, economic, and social settings, the error distribution need not even be symmetric. This is also the case when dealing with economic indicators such as per capita income, retirement savings, etc., or even when dealing with measures of research productivity such as the h-index, though the population in this case is discrete (see, for instance, [15]). As such, a general model is to simply assume that the error distribution F belongs

to the class of all continuous distributions with medians equal to zero. This class will be denoted by $\mathfrak{F}_{c,0}$.

Another situation where this problem arises is when dealing with a complex engineering system, such as the motherboard of a laptop computer or some technologically-advanced car (e.g., a Tesla Model S sedan). Such a system will be composed of many different components configured according to some structure function, with the components having different failure-time distributions and some of them possibly acting dependently on each other. Of main interest for such a system will be its time-to-failure (also called lifetime) denoted by X. Because of the complexity of the system, it may not be feasible to analyze the distribution of X by taking into account each of the failure time distributions of the components and the system's structure function which represents the configuration of the components to form the system. Thus a simplified and practically feasible viewpoint is to assume that the system's life distribution is some continuous distribution H. One may then be interested in the median Δ of this distribution H.

Thus, in these situations, the observable random variable X is assumed to have a distribution $H(x) = F(x - \Delta)$ with $F(\cdot) \in \mathfrak{F}_{c,0}$ and $\Delta \in \Re$ being the median of X. This will be referred to as the one-population nonparametric measurement error model, abbreviated NMEM. This is the simplest among the measurement error models. The goal is to infer about the parameter of interest Δ with $F(\cdot)$ acting as an infinite-dimensional nuisance parameter. We shall be interested in this paper in the construction of a confidence region (CR) for Δ based on a random sample of observations of X.

The problem of constructing confidence regions (usually intervals) for Δ in the NMEM is an old and well-trodden nonparametric statistical problem that has been addressed in many works. See, for instance, the textbooks [19, 9] which both discuss the confidence interval (CI) for Δ based on the sign statistic, a CI first presented in [20]. Many of the CIs for Δ are derived by starting with an appropriate test statistic for testing a hypothesis about Δ or an estimator of Δ and creating an appropriate pivotal quantity for Δ. Section 2 will briefly discuss several of these "off-the-shelf" CIs for Δ that have been proposed in the literature. More generally, quantiles instead of just the median may be of interest. The methods developed here could be adaptable to making inferences about quantiles.

It could be argued that, in many situations, a confidence region for a parameter is preferable than an associated point estimate, since it addresses *simultaneously* the issue of closeness to the truth (measured through the content of the region) and the sureness about such closeness to the truth (measured by the confidence region coefficient). Of course, as is typically done, one usually accompanies a point estimate (PE) by an estimate of its standard error (ESE), but then the user will still need to deduce closeness to the truth and assess the level of confidence about this closeness to the truth based on the PE and the ESE, a non-trivial activity if to be done *properly*. For more discussions on desirability of confidence regions see, for instance, the introduction in [5] and chapter 5 in [4].

We introduce some notations and definitions. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables (a random sample) from $F(x - \Delta)$, where $\Delta \in \Re$ and $F \in \mathfrak{F}_{c,0}$. The mathematical problem is to construct a confidence region (CR) for the parameter $\theta(F, \Delta) = \Delta$ with F an infinite-dimensional nuisance parameter. Denote by $\mathfrak{X}$ the range space of $\mathbf{X} = (X_1, \ldots, X_n)$ which will be endowed with a σ-field $\mathcal{X}$. We also denote by $\mathfrak{B}$ the Borel σ-field of $\Re$, and this will be endowed with the σ-field of subsets of $\mathfrak{B}$ consisting of its countable and co-countable subsets, with this σ-field denoted by $\mathcal{B}$.

Definition 1. *Fix an $\alpha \in (0, 1)$. Let $\mathbf{X} = (X_1, \ldots, X_n) \in \mathfrak{X}$ be IID from $F(x - \Delta)$. A measurable mapping $\Gamma : (\mathfrak{X}, \mathcal{X}) \to (\mathfrak{B}, \mathcal{B})$ is called a $100(1 - \alpha)\%$ region estimator or confidence region (CR) for Δ if $P_{(F(\cdot), \Delta)}\{\Delta \in \Gamma(\mathbf{X})\} \geq 1 - \alpha$ for every $(F(\cdot), \Delta) \in \mathfrak{F}_{c,0} \times \Re$.*

Remark: In Definition 1 we emphasize that we are dealing with the large class of distributions $\mathfrak{F}_{c,0}$ which places a stringent condition on the confidence level condition. One may restrict to a smaller subclass of $\mathfrak{F}_{c,0}$, such as the class of continuous *symmetric* at zero distributions, denoted by $\mathfrak{F}_{c,0}^{sym}$. Such a restriction may admit better CRs than those obtained under the larger class $\mathfrak{F}_{c,0}$ since any CR satisfying the confidence level condition under $\mathfrak{F}_{c,0}$ will also satisfy the confidence level condition under $\mathfrak{F}_{c,0}^{sym}$. $\parallel$

Remark: In later developments, we will allow the CR Γ to also depend on a randomizer U, a standard uniform random variable independent of $\mathbf{X}$. This is to be able to achieve exactly the desired confidence level $1 - \alpha$. In such a case, $\Gamma : (\mathfrak{X} \times [0, 1], \mathfrak{B} \otimes \sigma[0, 1]) \to (\mathfrak{B}, \mathcal{B})$ and $\Gamma(\mathbf{x}, u)$ will be the realized CR when $\mathbf{X} = \mathbf{x}$ and $U = u$. However, even if we allow for randomized CRs, we will usually suppress writing the U in $\Gamma(\mathbf{X}, U)$ and simply write $\Gamma(\mathbf{X})$. $\parallel$

Aside from satisfying the desired confidence coefficient in Definition 1, the quality of a CR depends on some measure of its content. Let $\nu(\cdot)$ be Lebesgue measure on $(\Re, \mathfrak{B})$. We will measure the content of a CR Γ for Δ via

$$\mathcal{C}[\Gamma; (F(\cdot), \Delta)] = E_{(F(\cdot), \Delta)} \nu[\Gamma(\mathbf{X})]. \tag{1.2}$$

In Definition 2 below we have the notion of uniformly best CRs. Our goal is to determine those CRs for Δ that possess such optimality properties.

Definition 2. *Let $\bar{\mathfrak{F}}_{c,0}$ be a subclass of $\mathfrak{F}_{c,0}$. A $100(1 - \alpha)\%$ CR Γ^* for Δ is a uniformly best CR for Δ under the subclass $\bar{\mathfrak{F}}_{c,0}$ if for any other $100(1 - \alpha)\%$ CR Γ,*

$$\mathcal{C}[\Gamma^*; (F(\cdot), \Delta)] \leq \mathcal{C}[\Gamma; (F(\cdot), \Delta)]$$

for all $(F(\cdot), \Delta) \in \bar{\mathfrak{F}}_{c,0} \times \Re$. If $\bar{\mathfrak{F}}_{c,0} = \mathfrak{F}_{c,0}$, then Γ^* will be said to be the uniformly best CR for Δ.

A major contribution of this work is a rigorous construction, starting from basic principles and considerations (what we call as a 'bottom-to-top' approach),

of $100(1 - \alpha)\%$ randomized region estimators. Thus, in contrast to the derivation of existing CIs which starts by creating pivotal quantities from known test statistics and/or estimators, our CRs arise from sufficiency, invariance, and optimality considerations. The region estimators we propose are of form

$$\Gamma(X, U) \equiv \Gamma(X, U; \alpha) = \left[\bigcup_{\{k \in \{0,1,\ldots,n\}:\ b(k;n,1/2) > c^* \hat{l}(k)\}} \left[X_{(k)}, X_{(k+1)} \right) \right] \bigcup$$

$$\left\{ \{U \leq \gamma\} \cap \left[\bigcup_{\{k \in \{0,1,\ldots,n\}:\ b(k;n,1/2) = c^* \hat{l}(k)\}} \left[X_{(k)}, X_{(k+1)} \right) \right] \right\},$$

where $b(k; n, 1/2) = \binom{n}{k} 2^{-n}$ and $\hat{l}(k)$ is an appropriate estimator of $l(k; F) = E\{X_{(k+1)} - X_{(k)}\}$, with $X_{(i)}$s the order statistics of the random sample, and U is a standard uniform variable independent of the X_is. The constant c^* is the infimum over all $c \in \Re$ satisfying $P\{b(B; n, 1/2) > c\hat{l}(B)\} \leq 1 - \alpha$, where B is a binomial random variable with parameters n and $1/2$; while $\gamma \in [0, 1]$ is the randomization probability determined by α. Both c^* and γ may also depend on the sample data $\mathbf{x}$, since $\hat{l}(k)$ may depend on $\mathbf{x}$. A specific form of $\hat{l}(k)$ that leads to a reasonable CR is given by

$$\hat{l}(k) = \binom{n}{k} \int_{-\infty}^{\infty} \hat{F}(w)^k [1 - \hat{F}(w)]^{n-k} dw$$

where $\hat{F}(w) = \sum_{i=1}^{n} I\{X_i - \hat{\Delta} \leq w\}/n$, the empirical distribution function of $X_i - \hat{\Delta}, i = 1, 2, \ldots, n$, with $\hat{\Delta}$ being the sample median. This specific CR will be developed in section 6. Prior to the development of these specific CRs, in section 3 we utilize invariance ideas to derive the general form of the almost-optimal equivariant CR for Δ under the NMEM but still under the assumption that F is known. Then, we address the question of how to deal with the fact that F is not actually known, leading to the region estimator above. Two other region estimators which are focused toward the class of symmetric distributions and the class of negative exponential distributions, but still valid under NMEM, will be developed in sections 4 and 5, respectively. We demonstrate these region estimators, together with existing 'off-the-shelf' confidence regions reviewed in Section 2, by applying them to two real data sets in section 7. Section 8 will present the results of simulation studies comparing the performance of these region estimators under different underlying distributions by examining their mean contents and their achieved coverage probabilities. In these studies, the procedure focused on symmetric distributions performed quite robustly under varied distributions, even for the non-symmetric distributions, in terms of coverage probability. Its mean content was also smaller than that for the CI developed from the sign statistic. Interestingly, this procedure, which is Γ_{10} in Table 2 on page 2370, has both its c^* and γ not dependent on $\mathbf{x}$. The simulation studies of the performance of the different CRs is the second major contribution of this work. The results demonstrate which CRs are viable under the NMEM. Section 9 will provide some concluding remarks.

2. Brief review of existing 'off-the-shelf' median CRs

In this section we briefly review existing methods for constructing frequentist-based $100(1 - \alpha)\%$ CRs for Δ under the NMEM. For a sample realization $\mathbf{x} = (x_1, \ldots, x_n)$, we define the usual sample statistics:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \quad s^2 = s^2(\mathbf{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2; \quad \hat{\Delta} = \mathrm{med}(\mathbf{x}).$$

$\bar{X}$, S^2, and $\hat{\Delta}$ will then represent the random versions of these sample statistics. We let $\phi(\cdot)$, $\Phi(\cdot)$, and $\Phi^{-1}(\cdot)$ denote the standard normal density, distribution, and quantile functions, respectively. We will use the conventional notation $z_\beta = \Phi^{-1}(1-\beta)$ for the $(1-\beta)$th quantile of $\Phi(\cdot)$. The functions $\mathcal{T}(\cdot; k)$ and $\mathcal{T}^{-1}(\cdot; k)$ will denote the distribution and quantile functions, respectively, of a Student's T random variable with degrees-of-freedom k, and its $(1-\beta)$th quantile will be denoted by $t_{k;\beta}$.

Arguably, the most commonly-used CR for the center of a distribution, which is Δ for symmetric distributions, is the T-based CR given by

$$\Gamma_1(\mathbf{X}) = \left[\bar{X} \pm t_{n-1;\alpha/2} \frac{S}{\sqrt{n}} \right]. \tag{2.1}$$

However, this CR is actually *not* valid under the NMEM since it does not satisfy the condition $P_{(F,\Delta)}\{\Delta \in \Gamma(X)\} \geq 1 - \alpha$ for *all* $(F(\cdot), \Delta) \in \mathfrak{F}_{c,0} \times \Re$. We still included this CR since it is typically the first choice to use by practitioners when constructing a confidence interval for μ or Δ and we would like to examine and compare its performance with other CRs under the NMEM.

The nonparametric analog of the T-based CR is the CR constructed from the Wilcoxon signed-rank statistic W^+ ([19]). The Walsh averages associated with the random sample X_is are $W_{ij} = (X_i + X_j)/2$ for $i \leq j$. Then the W^+ statistic for testing the null hypothesis $H_0 : \Delta = \Delta_0$ can be represented in terms of these Walsh averages via $W^+(\Delta_0) = \sum_{i \leq j} I\{W_{ij} \leq \Delta_0\}$. Denoting by $W_{(1)} \leq W_{(2)} \leq \ldots \leq W_{((n)(n+1)/2)}$ the order statistics of the Walsh averages, the Wilcoxon signed-rank based nonparametric CR for Δ is given by

$$\Gamma_2(\mathbf{X}) = [W_{(k_1+1)}, W_{(k_2+1)}) \tag{2.2}$$

where, with $\mathcal{W}^+(\cdot)$ denoting the null distribution of W^+ and obtainable using the function object `psignrank` in R [18],

$$k_1 = \sup\{w : \mathcal{W}^+(w) \leq \alpha/2\} \quad \text{and} \quad k_2 = \inf\{w : \mathcal{W}^+(w) \geq 1 - \alpha/2\}.$$

The CR Γ_2 is valid under $\mathfrak{F}_{c,0}^{sym}$, but not under $\mathfrak{F}_{c,0}$. Just like the T-based CR Γ_1, we also include this in the comparisons since it is a CR that will also tend to be used in practical work.

In contrast, the nonparametric CR derived from the sign statistic is valid under $\mathfrak{F}_{c,0}$ (see [20, 19, 9]). As before, let $\mathfrak{B}(\cdot)$ be the binomial distribution with parameters n and $1/2$. Let

$$k_1 = \sup\{w : \mathfrak{B}(w) \leq \alpha/2\} \quad \text{and} \quad k_2 = \inf\{w : \mathfrak{B}(w) \geq 1 - \alpha/2\}.$$

Then, this sign statistic-based CR is

$$\Gamma_3(\mathbf{X}) = [X_{(k_1+1)}, X_{(k_2+1)}). \tag{2.3}$$

Another CR of Δ is developed from the asymptotic normality of the sample median $\hat{\Delta}$. For $X_1, X_2, \ldots, X_n$ IID from a distribution $F(\cdot)$ with density $f(\cdot)$, this asymptotic distribution (cf., [19]) is given by

$$\hat{\Delta} \sim AN\left[\Delta, \frac{1}{n}\frac{1}{4f(\Delta)^2}\right].$$

If $\hat{f}(\hat{\Delta}; \mathbf{X})$ is an estimator of $f(\Delta)$, then an asymptotic confidence interval for Δ is

$$\Gamma_4(\mathbf{X}) = \left[\hat{\Delta} \pm z_{\alpha/2}\left(\frac{1}{\sqrt{n}2\hat{f}(\hat{\Delta}; \mathbf{X})}\right)\right]. \tag{2.4}$$

Kernel-based estimators of the density f are available in R using the function object `density` ([18]), and coupled with the `approx` function object, $f(\Delta)$ could then be estimated. This is how we implemented this CR in the illustrations and in the simulation studies.

Instead of relying on asymptotic approximations, one may resort to bootstrapping approaches. Let $\mathbf{X}_k^* = (X_{k1}^*, \ldots, X_{kn}^*)$ be the kth bootstrap sample out of BREPS bootstrap samples. Denote its sample median by $\hat{\Delta}_k^*$. The basic bootstrap CR using the sample median obtains the $\alpha/2$th and $(1-\alpha)/2$th quantiles of $\{\hat{\Delta}_k^* - \hat{\Delta}, k = 1, 2, \ldots, BREPS\}$, denoted by $\kappa_{1-\alpha/2}^*$ and $\kappa_{\alpha/2}^*$, respectively, and constructs the CR (cf., [6]) via

$$\Gamma_5(\mathbf{X}) = [\hat{\Delta} - \kappa_{\alpha/2}^*, \hat{\Delta} - \kappa_{1-\alpha/2}^*]. \tag{2.5}$$

The next bootstrapped-based CR is derived using the studentized median as pivot and with its standard error estimated by S_{boot}^*, the standard deviation of $\{\hat{\Delta}_k^*, k = 1, 2, \ldots, BREPS\}$. The CR is constructed via

$$\Gamma_6(\mathbf{X}) = \left[\hat{\Delta} \pm t_{n-1;\alpha/2}(S_{boot}^*)\right]. \tag{2.6}$$

The bootstrap percentile CR also uses the bootstrap distribution of $\hat{\Delta}$. Denoting by

$$\mathcal{B}^*(t) = \frac{1}{BREPS}\sum_{k=1}^{BREPS} I\{\hat{\Delta}_k^* \le t\}$$

the (empirical) bootstrap distribution, the percentile bootstrap CR is

$$\Gamma_7(\mathbf{X}) = \left[\mathcal{B}^{*-1}(\alpha/2), \mathcal{B}^{*-1}(1 - \alpha/2)\right] \tag{2.7}$$

where $\mathcal{B}^{*-1}(\cdot)$ is the bootstrap quantile function.

This percentile bootstrap CR is usually bettered by so-called bias-corrected CRs. See, for instance, chapter 5 of [4] and chapter 11 of [6]. The first improvement is provided by the bias-corrected (BC) CR which is

$$\Gamma_8(\mathbf{X}) = \left[\mathcal{B}^{*-1} \left(\Phi(2z_0 + \Phi^{-1}(\alpha/2)) \right), \mathcal{B}^{*-1} \left(\Phi(2z_0 + \Phi^{-1}(1 - \alpha/2)) \right) \right] \quad (2.8)$$

where $p_0 = \mathcal{B}^*(\hat{\Delta})$ and $z_0 = \Phi^{-1}(p_0)$. On the other hand, the BCa method (bias-corrected and accelerated) CR takes the form

$$\begin{aligned}
\Gamma_9(\mathbf{X}) &= \left[\mathcal{B}^{*-1} \left(\Phi \left[z_0 + \frac{z_0 + \Phi^{-1}(\alpha/2)}{1 - a(z_0 + \Phi^{-1}(\alpha/2))} \right] \right), \right. \\
&\qquad \left. \mathcal{B}^{*-1} \left(\Phi \left[z_0 + \frac{z_0 + \Phi^{-1}(1 - \alpha/2)}{1 - a(z_0 + \Phi^{-1}(1 - \alpha/2))} \right] \right) \right]
\end{aligned} \quad (2.9)$$

where the acceleration coefficient a can be estimated using jackknifed samples estimates $\hat{\Delta}_{(i)}, i = 1, 2, \ldots, n$, via

$$\hat{a} = \frac{1}{6} \frac{\sum_{i=1}^{n} (\hat{\Delta}_{(i)} - \hat{\Delta}_{()})^2}{\left[\sum_{i=1}^{n} (\hat{\Delta}_{(i)} - \hat{\Delta}_{()})^2 \right]^{3/2}}$$

with $\hat{\Delta}_{()} = \frac{1}{n} \sum_{i=1}^{n} \hat{\Delta}_{(i)}$, where $\hat{\Delta}_{(i)}$ is the sample median of the ith jackknifed sample $X_{-i} \equiv (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

These CRs from Γ_1 to Γ_9, together with the CRs developed in the succeeding sections, will be used in the illustrations and in the simulation studies. For a peek, see Table 2 on page 2370.

3. Development of optimal CRs

In this section, we present the development of the general form of our CRs by invoking the Sufficiency Principle, Invariance Principle, and optimality considerations.

3.1. Invariant models and equivariant CRs

We first review the notions of invariant statistical models and equivariant CRs (see, for instance, [14]). We do this review in a more general framework than the concrete NMEM which is the focus of this paper. We note that sufficiency and invariance were major ideas utilized by Peter Hooper in several of his papers dealing with confidence sets and prediction sets, cf., [11, 10, 12].

Let X be an observable random element taking values in a sample space $\mathfrak{X}$. The class of probability models governing X is $\mathfrak{P}$, which consists of probability measures Ps on the measurable space $(\mathfrak{X}, \mathcal{F})$, with $\mathcal{F}$ a suitable σ-field of subsets of $\mathfrak{X}$. Let $\tau : \mathfrak{P} \to \mathfrak{T}$ be a functional, with $\tau(P)$ being the parameter of interest. Let $\mathcal{T}$ be a σ-field of subsets of $\mathfrak{T}$, and let $\sigma\mathcal{T}$ be the countable/co-countable

σ-field of subsets of $\mathcal{T}$. A $100(1-\alpha)\%$ region estimator or confidence region (CR) for $\tau(P)$ is a set-valued mapping $\Gamma : (\mathfrak{X}, \mathcal{F}) \to (\mathcal{T}, \sigma\mathcal{T})$, such that

$$P\{\tau(P) \in \Gamma(\mathbf{X})\} \geq 1 - \alpha, \ \forall P \in \mathfrak{P}. \tag{3.1}$$

Let $G = \{g : \mathfrak{X} \to \mathfrak{X}\}$ be a family of (measurable) transformations on $\mathfrak{X}$ that forms a group under an operation $\cdot$ and with identity element $1_G \equiv 1$. Let $\bar{G} = \{\bar{g} : \mathfrak{P} \to \mathfrak{P}\}$ be a group of transformations on $\mathfrak{P}$ such that there exists a homomorphism $\bar{h} : G \to \bar{G}$ and let $1 \equiv 1_{\bar{G}} = \bar{h}(1_G)$ be the identity element in $\bar{G}$. The statistical model is said to be $(G, \bar{G})$-invariant if

$$P\{gX \in A\} = \bar{g}P\{X \in A\}, \forall g \in G; A \in \mathcal{F}. \tag{3.2}$$

In addition, let $\tilde{G} = \{\tilde{g} : \mathfrak{T} \to \mathfrak{T}\}$ be a group of transformations on $\mathfrak{T}$ such that there exists a homomorphism $\tilde{h} : G \to \tilde{G}$. The parametric functional $\tau(P)$ is said to be $(\bar{G}, \tilde{G})$-equivariant if $\tau(\bar{g}P) = \tilde{g}\tau(P)$ for all $g \in G, P \in \mathfrak{P}$. Employing a decision-theoretic framework, define a loss function L on $\mathfrak{T} \times \mathcal{T}$ given by the $0/1$ loss function

$$L(\tau, C) = 1 - I\{\tau \in C\}.$$

We shall say that the loss function is $\tilde{G}$-invariant if $L(\tilde{g}\tau, \tilde{g}C) = L(\tau, C)$ for every $g \in G, \tau \in \mathfrak{T}$, and $C \in \mathcal{T}$. Given a confidence region $\Gamma(X)$, its risk function is

$$R(P, \Gamma) \equiv E_P\{L(\tau(P), \Gamma(X))\} = 1 - P\{\tau(P) \in \Gamma(X)\}.$$

As such, the condition for a $100(1-\alpha)\%$ confidence region $\Gamma(X)$ is equivalent to having $R(P, \Gamma) = E_P\{L(\tau(P), \Gamma(X))\} \leq \alpha$ for every $P \in \mathfrak{P}$. When a $(G, \bar{G})$-invariant statistical model is coupled with a $\tilde{G}$-invariant loss function, then we would say that the statistical problem of constructing a confidence region $\Gamma(\cdot)$ is $(G, \bar{G}, \tilde{G})$-invariant. A confidence region $\Gamma(X)$ is then said to be $(G, \tilde{G})$-equivariant if for every $g \in G$ and $x \in \mathfrak{X}$, we have that

$$\Gamma(gx) = \tilde{g}\Gamma(x) \equiv \{\tilde{g}t : t \in \Gamma(x)\}.$$

The Principle of Invariance then dictates that we should only utilize $(G, \tilde{G})$-equivariant confidence regions.

For an invariant confidence region problem, if $\Gamma(\cdot)$ is equivariant, then we have that, for every $g \in G$,

$$\begin{aligned} P\{\tau(P) \in \Gamma(X)\} &= E_P\{1 - L(\tau(P), \Gamma(X))\} = E_P\{1 - L(\tilde{g}\tau(P), \tilde{g}\Gamma(X))\} \\ &= E_P\{1 - L(\tau(\bar{g}P), \Gamma(gX)\} = E_{\bar{g}P}\{1 - L(\tau(\bar{g}P), \Gamma(X))\} \\ &= (\bar{g}P)\{\tau(\bar{g}P) \in \Gamma(X)\}. \end{aligned}$$

Furthermore, if the group $\bar{G}$ is transitive over $\mathfrak{P}$, meaning that for any given $P_0 \in \mathfrak{P}$ we have $\{\bar{g}P_0 : \bar{g} \in G\} = \mathfrak{P}$, then it suffices to consider an arbitrary element $P_0 \in \mathfrak{P}$ to determine $P\{\tau(P) \in \Gamma(X)\}$ for all $P \in \mathfrak{P}$ since this equals the value using the arbitrary P_0.

Recall that we also need to measure the quality of a confidence region by measuring its content using the quantity $\mathcal{C}(\Gamma, P) = E_P[\nu(\Gamma(X))]$, where $\nu(\cdot)$ is a measure on $(\mathfrak{T}, \mathcal{T})$, e.g., Lebesgue measure. We seek those confidence regions with small $\mathcal{C}(\Gamma, P)$. Observe that for an equivariant $\Gamma(\cdot)$ in an invariant statistical model, we have for every $g \in G$ that

$$\mathcal{C}(\Gamma, P) = E_P\left[\nu(\Gamma(X))\right] = E_P\left[\nu\left(\tilde{g}^{-1}\Gamma(gX)\right)\right] = E_{\bar{g}P}\left[\nu\left(\tilde{g}^{-1}\Gamma(X)\right)\right].$$

If it so happens that $\nu[\tilde{g}^{-1}\Gamma(x)] = \xi(g)\nu[\Gamma(x)]$ for all $x \in \mathfrak{X}$ and $g \in G$ and for some $\xi : G \to \Re$, then there is the possibility of finding a $\Gamma(\cdot)$ that satisfies the required confidence level and minimizes the content. We shall call this condition as quasi-invariance of ν with respect to $(G, \tilde{G})$. However, if $(G, \tilde{G})$-quasi-invariance of ν does not hold, then a uniformly best confidence region may not exist. But, a uniformly best confidence region on a *subfamily* $\mathfrak{P}_0 \subset \mathfrak{P}$ may still exist among the class of $100(1 - \alpha)\%$ confidence regions over $\mathfrak{P}$. In [11, 10] quasi-invariance of the measure ν was imposed, but in some settings this may be unnatural such as in the NMEM under consideration in the current paper.

3.2. Towards optimal CRs for the median

Consider now the problem of constructing a CR for the median Δ under the NMEM: $X_i = \Delta + \epsilon_i, i = 1, \ldots, n$; $\epsilon_i \overset{IID}{\sim} F(\cdot) \in \mathfrak{F}_{c,0}, \Delta \in \Re$. Prior to invoking invariance, we first reduce the problem via the Sufficiency Principle. Thus, we may assume that the observable random vector is $X_{()} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, the vector of order statistics which is a complete sufficient statistic under the class of distributions $\mathfrak{F}_{c,0}$ (cf., [14]). The appropriate sample space is therefore $\mathfrak{X} = \{(v_1, v_2, \ldots, v_n) : v_1 \leq v_2 \leq \ldots \leq v_n\}$. *A word on our notation:* even though we had reduced to $X_{()}$, in the sequel, when we write P_F or E_F, this means that the common distribution of the X_is is F. For measuring the content of a region for Δ we use Lebesgue measure ν on $\Re$.

The first invariance reduction is obtained through location-invariance. The problem is invariant with respect to translations with the groups of transformations being, for every $c \in \Re$,

$$x_{()} \mapsto x_{()} + c; \ (F, \Delta) \mapsto (F, \Delta + c); \text{ and } \theta \mapsto \theta + c.$$

A CR $\Gamma(X_{()})$ is location-equivariant if, for every $c \in \Re$, $\Gamma(x_{()} + c) = \Gamma(x_{()}) + c$, where $x_{()} + c = (x_{(1)} + c, \ldots, x_{(n)} + c)$. Observe that for a location-equivariant $\Gamma(\cdot)$, we have for every $c \in \Re$ that

$$
\begin{aligned}
P_{(F,\Delta)}\{\Delta \in \Gamma(X_{()})\} &= P_{(F,\Delta)}\{\Delta \in \Gamma(X_{()} + c) - c\} \\
&= P_{(F,\Delta)}\{\Delta + c \in \Gamma(X_{()} + c)\} = P_{(F,\Delta+c)}\{\Delta + c \in \Gamma(X_{()})\} \\
&= P_{(F,0)}\{0 \in \Gamma(X_{()})\}
\end{aligned}
$$

by taking $c = -\Delta$ to obtain the last equality. The problem has thus been reduced to considering $X_{()}$ to be the order statistics of size n from $F(\cdot)$ and we seek a

location-equivariant $\Gamma(x_{()})$ such that, for every $F \in \mathfrak{F}_{c,0}$, $P_F\{0 \in \Gamma(X_{()})\} \geq 1 - \alpha$. In addition, we seek to minimize $E_F\nu[\Gamma(X_{()})]$ over all $F \in \mathfrak{F}_{c,0}$. Note that Lebesgue measure ν in $\Re$ is location-invariant, that is, $\nu(B) = \nu(B + c)$ for every $B \in \mathfrak{B}$ and $c \in \Re$.

We remark at this stage that if we *know* the distribution $F(\cdot)$, then we could determine the optimal CR for Δ *under* this known distribution and no further invariance reduction will be needed. To demonstrate, suppose that F is the normal distribution with mean zero and variance σ^2 which could be taken to be $\sigma^2 = 1$, so $F(\cdot) = \Phi(\cdot)$. Then, we seek a location-equivariant $\Gamma^*(x_{()})$ satisfying $P_\Phi\{0 \in \Gamma^*(X_{()})\} \geq 1 - \alpha$ and with $E_\Phi \int_\Re I\{w \in \Gamma^*(X_{()})\}dw$ minimized. By order statistics theory (cf., [3]), under Φ, the joint density function of $X_{()}$ is given by $f(x_{()}) = n! \prod_{i=1}^n \phi(x_{(i)})I\{x_{()} \in \mathfrak{X}\}$. Thus, we want $P_\Phi\{0 \in \Gamma^*(X_{()})\} = \int I\{0 \in \Gamma^*(x_{()})\}f(x_{()})dx_{()} \geq 1 - \alpha$. On the other hand, we obtain

$$
\begin{aligned}
E_\Phi \int_\Re I\{w \in \Gamma^*(X_{()})\}dw &= \int_\mathfrak{X} \int_\Re I\{w \in \Gamma^*(x_{()})\}f(x_{()})dwdx_{()} \\
&= \int_\mathfrak{X} \int_\Re I\{0 \in \Gamma^*(x_{()} - w)\}f(x_{()})dwdx_{()} \\
&= \int_\mathfrak{X} I\{0 \in \Gamma^*(x_{()})\}h(x_{()})dx_{()},
\end{aligned}
$$

where

$$
h(x_{()}) = n! \frac{(2\pi)^{-(n-1)/2}}{\sqrt{n}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n (x_{(i)} - \bar{x})^2 \right\}
$$

obtained by letting $u = w; y_{(i)} = x_{(i)} - w, i = 1, 2, \ldots, n$, and noting that

$$
\int_\Re \prod_{i=1}^n \exp\left\{ -\frac{1}{2}(y_{(i)} + u)^2 \right\} du = \exp\left\{ -\frac{1}{2} \sum_{i=1}^n (y_{(i)} - \bar{y})^2 \right\} \frac{\sqrt{2\pi}}{\sqrt{n}}.
$$

The problem is then to find a location-equivariant $\Gamma^*(x_{()})$ that will minimize $\int_\mathfrak{X} I\{0 \in \Gamma^*(x_{()})\}h(x_{()})dx_{()}$ subject to the condition that

$$
\int I\{0 \in \Gamma^*(x_{()})\}f(x_{()})dx_{()} \geq 1 - \alpha.
$$

The solution to this constrained minimization problem (see the optimization result in Theorem 2 on page 2361) is the well-known z-confidence interval for the normal mean given by $\Gamma^*(x_{()}) = [\bar{x} \pm z_{\alpha/2}(1/\sqrt{n})]$ with $z_\alpha = \Phi^{-1}(1 - \alpha)$.

However, since F is known only to belong to $\mathfrak{F}_{c,0}$, a further invariance reduction is needed. This is achieved through strictly increasing continuous transformations with 0 as a fixed point. Let $\mathcal{M}$ denote the collection of functions $m(\cdot)$ that are strictly increasing continuous functions on $\Re$ with $m(0) = 0$. The groups of transformations are given by

$$
x_{()} \mapsto (m(x_{(1)}), m(x_{(2)}), \ldots, m(x_{(n)})); \ F \mapsto Fm^{-1}; \text{ and } \Delta \mapsto m(\Delta).
$$

$\Gamma(x_{()})$ is then equivariant with respect to these groups of transformations if

$$\Gamma(m(x_{(1)}), \ldots, m(x_{(n)})) = m\Gamma(x_{(1)}, \ldots, x_{(n)}) \equiv \{m(w) : w \in \Gamma(x_{()})\},$$

so that for every $m \in \mathcal{M}$ and $x_{()} \in \mathfrak{X}$, we have $\Gamma(x_{()}) = m^{-1}\Gamma(m(x_{()}))$. We then have that, for every $m \in \mathcal{M}$ and $F \in \mathfrak{F}_{c,0}$,

$$
\begin{aligned}
P_F\{0 \in \Gamma(X_{()})\} &= P_F\{0 \in m^{-1}\Gamma(m(X_{()}))\} \\
&= P_F\{0 \in \Gamma(m(X_{()}))\} \quad \text{since } m(0) = 0 \\
&= P_{Fm^{-1}}\{0 \in \Gamma(X_{()})\}.
\end{aligned}
$$

Observe, however, that

$$E_F\nu[\Gamma(X_{()})] = E_F\nu[m^{-1}\Gamma(m(X_{()}))] = E_{Fm^{-1}}\nu[m^{-1}\Gamma(X_{()})]$$

and we do *not* have in this situation quasi-invariance of the measure ν with respect to the groups of monotone transformations.

The group of transformations $\mathcal{M}$ with $F \mapsto Fm^{-1}$ is transitive over $\mathfrak{F}_{c,0}$. Thus, we may simply pick an arbitrary $F_0 \in \mathfrak{F}_{c,0}$, which could be taken to be $F_0 = U[-1, 1]$, the uniform distribution over $[-1, 1]$. Indeed, if $X \sim F \in \mathfrak{F}_{c,0}$, then with $m_F(v) = 2F(v) - 1$, we have $m(X) \sim U[-1, 1]$. Thus,

$$P_F\{0 \in \Gamma(X_{()})\} = P_{F_0}\{0 \in \Gamma(X_{()})\} \text{ and } E_F\nu[\Gamma(X_{()})] = E_{F_0}\nu[m^{-1}\Gamma(X_{()})].$$

We emphasize again that in the second equation we could *not* drop the term m^{-1} nor factor it out from inside the $\nu(\cdot)$ measure. This will prevent us from obtaining a uniformly (over $\mathfrak{F}_{c,0}$) best confidence region for Δ.

Next, we obtain a representation of $\Gamma(x_{()})$ by choosing a specific member of $\mathcal{M}$ that depends on $x_{()}$. For an $x_{()}$, define $m(x_{()})(w)$ for $w \in \{x_{(i)} - x_{(n)}, i = 1, 2, \ldots, n\}$ via

$$m(x_{()})(w) = \sum_{i=1}^{n} I\{x_{(i)} - x_{(n)} \leq w\} - n,$$

and for $w \in \Re$ define it such that it is strictly increasing and continuous over all $w \in \Re$. Observe that for $j = 1, 2, \ldots, n$,

$$m(x_{()})(x_{(j)} - x_{(n)}) = j - n \quad \text{and} \quad m(x_{()})(-x_{(n)}) = B(x_{()}) - n, \qquad (3.3)$$

where $B(x_{()}) = \sum_{j=1}^{n} I\{x_{(j)} \leq 0\} = \sum_{j=1}^{n} I\{x_j \leq 0\}$. Note that $m(x_{()})(0) = n - n = 0$ and observe that $m(x_{()})^{-1}(j - n) = x_{(j)} - x_{(n)}, j = 1, 2, \ldots, n$.

Lemma 1. *With $m(x_{()})(\cdot)$ defined as above, a location-equivariant (LE) and $\mathcal{M}$-equivariant (ME) $\Gamma(x_{()})$ has representation*

$$\Gamma(x_{()}) = m(x_{()})^{-1}[\Gamma_0 - n] + x_{(n)} \qquad (3.4)$$

where Γ_0 is some region in $\Re$. Thus, the LE and ME $\Gamma(\cdot)$s are determined by Γ_0s, which are subsets of $\Re$. In fact, given a $\Gamma_0 \subset \Re$, we have

$$\Gamma(x_{()}) = \bigcup_{k \in [\Gamma_0 \cap \{0, 1, \ldots, n\}]} [x_{(k)}, x_{(k+1)}) \qquad (3.5)$$

whose Lebesgue measure is $\nu[\Gamma(x_{()})] = \sum_{k \in [\Gamma_0 \cap \{0, 1, \ldots, n\}]} [x_{(k+1)} - x_{(k)}]$.

Proof. We utilize the location-equivariance (LE) and $\mathcal{M}$-equivariance (ME) of $\Gamma(\cdot)$. We have

$$
\begin{aligned}
\Gamma(x_{()}) &= \Gamma(x_{()} - x_{(n)}) + x_{(n)} \text{ (by LE)} \\
&= m(x_{()})^{-1}\Gamma(m(x_{()})(x_{()} - x_{(n)})) + x_{(n)} \text{ (by ME)} \\
&= m(x_{()})^{-1}\Gamma(1 - n, 2 - n, \ldots, (n-1) - n, n - n) + x_{(n)} \\
&= m(x_{()})^{-1}[\Gamma(1, 2, \ldots, n - 1, n) - n] + x_{(n)} \text{ (again, by LE)} \\
&= m(x_{()})^{-1}[\Gamma_0 - n] + x_{(n)},
\end{aligned}
$$

where $\Gamma_0 = \Gamma(1, 2, \ldots, n)$. To establish (3.5), given a $\Gamma_0 \subset \Re$, observe that

$$
\begin{aligned}
\{0 \in \Gamma(x_{()})\} &\iff \{0 \in m(x_{()})^{-1}[\Gamma_0 - n] + x_{(n)}\} \\
&\iff \{m(x_{()})(-x_{(n)}) \in \Gamma_0 - n\} \\
&\iff \{(B(x_{()}) - n) \in (\Gamma_0 - n)\} \text{ by (3.3)} \\
&\iff \{B(x_{()}) \in \Gamma_0\}.
\end{aligned}
$$

It now follows that

$$
\begin{aligned}
\{w \in \Gamma(x_{()})\} &\iff \{0 \in \Gamma(x_{()} - w)\} \text{ [by LE property]} \\
&\iff \{B(x_{()} - w) \in \Gamma_0\} \text{ [preceding result]} \\
&\iff \left\{ w \in \bigcup_{k \in \Gamma_0 \cap \{0,1,\ldots,n\}} \{v : B(x_{()} - v) = k\} \right\} \\
&\iff \left\{ w \in \bigcup_{k \in \Gamma_0 \cap \{0,1,\ldots,n\}} [x_{(k)}, x_{(k+1)}) \right\}.
\end{aligned}
$$

Thus, given a $\Gamma_0 \subset \Re$, $\Gamma(x_{()}) = \bigcup_{k \in \Gamma_0 \cap \{0,1,\ldots,n\}} [x_{(k)}, x_{(k+1)})$ establishing (3.5). The last result about the Lebesgue measure of $\Gamma(x_{()})$ is immediate since the intervals $\{[x_{(k)}, x_{(k+1)}), k \in \Gamma_0 \cap \{0, 1, \ldots, n\}\}$ are disjoint. □

3.3. Optimal CRs

Next, we tackle the problem of choosing an 'optimal' (properly defined) region Γ_0, which then determines $\Gamma(x_{()})$ via the representation in Lemma 1. Recall that the goal is to find $\Gamma(\cdot)$ such that with $F_0 = U[-1, 1]$, $P_{F_0}\{0 \in \Gamma(X_{()})\} \geq 1 - \alpha$ and, for every $F \in \mathfrak{F}_{c,0}$, $E_F\{\nu[\Gamma(X_{()})]\}$ is minimized, or if this is not possible, is made small. Note that under F_0, $B = B(X_{()}) = B(X) = \sum_{i=1}^{n} I\{X_i \leq 0\}$ has a binomial distribution with parameters $(n, 1/2)$, denoted by $\mathfrak{B}(\cdot; n, 1/2)$, and with associated probability mass function $b(k; n, 1/2) = \binom{n}{k}2^{-n}I\{k \in \{0, 1, \ldots, n\}\}$. From the proof of Lemma 1, we have that $\{0 \in \Gamma(X_{()})\} = \{B(X_{()}) \in \Gamma_0\}$, so that

$$
\begin{aligned}
1 - \alpha &\leq P_{F_0}\{0 \in \Gamma(X_{()})\} = P_{F_0}\{B \in \Gamma_0\} \\
&= \sum_{k=0}^{n} I\{k \in \Gamma_0\}b(k; n, 1/2) = \sum_{k=0}^{n} \delta_0(k)b(k; n, 1/2)
\end{aligned}
$$

where we let $\delta_0(k) = I\{k \in \Gamma_0\}, k = 0, 1, \ldots, n$. The expected Lebesgue measure of $\Gamma(X_{()})$ is

$$E_F\nu[\Gamma(X_{()})] = \sum_{k=0}^{n} I\{k \in \Gamma_0\}[E_F(X_{(k+1)}) - E_F(X_{(k)})] = \sum_{k=0}^{n} \delta_0(k)l(k; F)$$

with $X_{(0)} \equiv E_F[X_{(0)}] \equiv \inf\{v \in \Re : F(v) > 0\}$ and $X_{(n+1)} \equiv E_F[X_{(n+1)}] \equiv \sup\{v \in \Re : F(v) < 1\}$, and

$$l(k; F) = E_F(X_{(k+1)}) - E_F(X_{(k)}), k = 0, 1, \ldots, n. \tag{3.6}$$

Assume *first* that we know the values of $\{l(k) \equiv l(k; F), k = 0, 1, \ldots, n\}$. We now allow for randomized confidence regions in order to achieve optimality, that is, we allow for Γ_0 and Γ to depend on a randomizer U which is a standard uniform random variable independent of the X_is. We remark that in [11, 10, 12] randomized procedures were also allowed to enable achieving optimality, similarly to the Neyman-Pearson theory of most powerful tests (cf., [14]).

Define the right-continuous non-increasing $[0, 1]$-valued function, for $t \in \Re$,

$$G(t) = P\{b(B; n, 1/2) > tl(B; F)\} = \sum_{k=0}^{n} I\{b(k; n, 1/2) > tl(k; F)\}b(k; n, 1/2).$$

For a given $\alpha \in (0, 1)$, define

$$c = \inf\{t : G(t) \le 1 - \alpha\} \quad \text{and} \quad \gamma = \frac{(1 - \alpha) - G(c)}{G(c-) - G(c)}.$$

Define the function δ_0^* over $\{0, 1, \ldots, n\} \times [0, 1]$ via

$$\delta_0^*((k, u)) = I\{b(k; n, 1/2) > cl(k; F)\} + I\{b(k; n, 1/2) = cl(k; F)\}I\{u \le \gamma\}.$$

The optimal Γ_0 then satisfies $\delta_0^*(k, u) = I\{k \in \Gamma_0^*(u)\}$.

Theorem 2. *Let $F \in \mathfrak{F}_{c,0}$ and let $\{l(k; F), k = 0, 1, \ldots, n\}$ be as defined in (3.6). Then $E_F[\delta_0^*(B, U)] = 1 - \alpha$. Furthermore, if δ_0 is any other $\{0, 1\}$-valued function in $\{0, 1, \ldots, n\} \times [0, 1]$ with $E_F[\delta_0(B, U)] \ge 1 - \alpha$, then*

$$E\left\{\sum_{k=0}^{n} [\delta_0^*(k, U)l(k; F)]\right\} \le E\left\{\sum_{k=0}^{n} [\delta_0(k, U)l(k; F)]\right\},$$

where the expectation is with respect to the randomizer U.

Proof. From the form of δ_0^*, we have

$$
\begin{aligned}
E_F[\delta_0^*(B, U)] &= P\{b(B; n, 1/2) > cl(B; F)\} + \gamma P\{b(B; n, 1/2) = cl(B; F)\} \\
&= G(c) + \left[\frac{(1 - \alpha) - G(c)}{G(c-) - G(c)}\right][G(c-) - G(c)] = 1 - \alpha.
\end{aligned}
$$

Let δ_0 be any other function on $\{0, 1, \ldots, n\} \times [0, 1]$ with $E_F[\delta_0(B, U)] \geq 1 - \alpha$. From the definition of δ_0^*, we observe that for each $(k, u) \in \{0, 1, \ldots, n\} \times [0, 1]$,

$$[b(k; n, 1/2) - cl(k; F)][\delta_0^*(k, u) - \delta_0(k, u)] \geq 0.$$

Summing over $k = 0, 1, \ldots, n$, and integrating over $u \in [0, 1]$, we find that

$$E_F\{\delta_0^*(B, U) - \delta_0(B, U)\}$$
$$\geq c \left[\sum_{k=0}^{n} \int_0^1 l(k; F)\delta_0^*(k, u)du - \sum_{k=0}^{n} \int_0^1 l(k; F)\delta_0(k, u)du \right].$$

Since $c \geq 0$ and by condition we have $E_F\{\delta_0^*(B, U) - \delta_0(B, U)\} \leq 0$, then

$$\sum_{k=0}^{n} \int_0^1 l(k; F)\delta_0^*(k, u)du \leq \sum_{k=0}^{n} \int_0^1 l(k; F)\delta_0(k, u)du$$

which completes the proof of the theorem. □

Remark: We note that this proof is similar to that of the Neyman-Pearson Lemma except for the fact that the $\{l(k; F) : k = 0, 1, \ldots, n\}$ is not a distribution function. ‖

Therefore, the optimal Γ_0, possibly using a randomizer $U \sim U[0, 1]$, is

$$\Gamma_0^*(u) = [\{k \in \{0, 1, \ldots, n\} : b(k; n, 1/2) > cl(k; F)\}] \bigcup$$
$$[\{u \leq \gamma\} \cap \{k \in \{0, 1, \ldots, n\} : b(k; n, 1/2) = cl(k; F)\}].$$

The associated optimal confidence region for Δ, possibly randomized, is

$$\Gamma^*(X_{()}, U) = \left[\bigcup_{\{k \in \{0,1,\ldots,n\}: \ b(k;n,1/2)>cl(k;F)\}} [X_{(k)}, X_{(k+1)}) \right] \bigcup$$
$$\left[\{U \leq \gamma\} \cap \left\{ \bigcup_{\{k \in \{0,1,\ldots,n\}: \ b(k;n,1/2)=cl(k;F)\}} [X_{(k)}, X_{(k+1)}) \right\} \right]. \quad (3.7)$$

Note that if we drop the term in the pair of brackets containing $\{U \leq \gamma\}$ in (3.7), we obtain

$$\Gamma^*(X_{()}) = \bigcup_{\{k \in \{0,1,\ldots,n\}: \ b(k;n,1/2)\geq cl(k;F)\}} [X_{(k)}, X_{(k+1)}), \quad (3.8)$$

which is a conservative confidence region for Δ in the sense that $P_{(F,\Delta)}\{\Delta \in \Gamma^*(X_{()})\} \geq 1 - \alpha$. If $100(1 - \alpha)\%$ is a *natural* confidence coefficient (cf., [19]) associated with the binomial distribution, then we may still obtain an exact confidence region.

3.4. Implementation aspects

The optimal confidence region required knowledge of the $l(k; F)$s, or at the very least the ordering of the $k \in \{0, 1, \ldots, n\}$ for inclusion in the Γ_0^*, which is determined by the magnitude of the ratios $r(k; F) \equiv b(k; n, 1/2)/l(k; F), k = 0, 1, \ldots, n$. In general, $l(k; F)$ will depend on the unknown true distribution F, hence the values of $l(k; F)$s or the $r(k; F)$s will not be known. In this case, we will not be able to determine $\Gamma^*(\cdot)$. We describe two approaches to circumvent this problem.

(i) Restrict F to belong to a subclass of the family of continuous distributions, say $\widetilde{\mathfrak{F}}_{c,0} \subset \mathfrak{F}_{c,0}$ and determine the $l(k; F)$s or the $r(k; F)$s for this class. This is then tantamount to satisfying the confidence level condition over the *whole* of $\mathfrak{F}_{c,0}$, but *focusing* only on the subclass $\widetilde{\mathfrak{F}}_{c,0}$ for minimizing the expected Lebesgue measure of the confidence region.

(ii) Utilize the observed sample data to estimate $l(k; F)$ by $\hat{l}(k)$, then use these $\hat{l}(k)$s in the expression of $\Gamma^*(X_{()}, U)$. There are several ways to accomplish this which are discussed below. However, it should be pointed out that when estimates are plugged-in, data double-dipping ensues and the achieved confidence level may not anymore satisfy the condition of being at least equal to $1 - \alpha$.

The next two sections will deal with these two approaches towards developing the region estimators.

4. Optimal CRs focused on symmetric distributions

4.1. For uniform distributions

We illustrate the different approaches for dealing with the situation of unknown $l(k; F), k = 0, 1, \ldots, n$. Let us consider approach (i) first. Suppose that we consider the subfamily of uniform distributions. It suffices to consider $F = U[-\alpha, \alpha]$ for $\alpha > 0$. Note that $X \sim U[-\alpha, \alpha]$ iff $X \overset{d}{=} (2V - 1)\alpha$ with $V \sim U[0, 1]$. Thus, $X_{(k)} \overset{d}{=} (2V_{(k)} - 1)\alpha$ and it is well-known that $E(V_{(k)}) = k/(n + 1)$. As such $E(X_{(k)}) = (2k/(n + 1) - 1)\alpha$, hence

$$l(k; F) = \frac{2\alpha}{n+1}[(k+1) - k] = \frac{2\alpha}{n+1}, k = 0, 1, \ldots, n.$$

Consequently, the ratios of interest are

$$r(k; F) = \frac{b(k; n, 1/2)}{l(k; F)} = \frac{\binom{n}{k}2^{-n}}{(2\alpha/(n+1))} \propto \binom{n}{k}, k = 0, 1, \ldots, n.$$

Hence, the optimal Γ_0^* is of form $\Gamma_0^* = \{k \in \{0, 1, \ldots, n\} : \binom{n}{k} > c\}$ where c is the smallest value such that $P\{B \in \Gamma_0^*\} \leq 1 - \alpha$. Since the mapping from

$k \in \{0, 1, \ldots, n\}$ into $\binom{n}{k}$ is symmetric about $n/2$ and decreases as $|k - n/2|$ increases, then $\Gamma_0^* = \{k \in \{0, 1, \ldots, n\} : |k - n/2| < d\}$ for d satisfying

$$
\begin{aligned}
d &= \sup\left\{e : P\{|B - \frac{n}{2}| < e\} \le 1 - \alpha\right\} \\
&= \sup\left\{e : P\{B < \frac{n}{2} + e\} \le 1 - \alpha/2\right\}.
\end{aligned}
$$

This implies that $(n/2) + d$ is the $(1 - \alpha/2)$th quantile of $\mathfrak{B}(\cdot; n, 1/2)$, denoted by $b_{n,1/2;\alpha/2}$, obtainable via the `qbinom` function in `R` ([18]). Letting $k_2 = b_{n,1/2;\alpha/2}$ and $k_1 = n - k_2$, then

$$
P\{k_1 < B < k_2\} \le 1 - \alpha \le P\{k_1 \le B \le k_2\}.
$$

With

$$
\gamma = \frac{(1 - \alpha) - P\{k_1 < B < k_2\}}{2 \Pr\{B = k_2\}},
$$

the resulting randomized confidence region for Δ is

$$
\Gamma_{10}^*(X_{()}, U) = \begin{cases} \left[X_{(k_1+1)}, X_{(k_2)}\right) & \text{if} \quad U > \gamma \\ \left[X_{(k_1)}, X_{(k_2+1)}\right) & \text{if} \quad U \le \gamma \end{cases}. \tag{4.1}
$$

This Γ_{10}^* CR procedure is the *randomized* version of the sign-statistic based CR, given in (2.3) and denoted by Γ_3, which was developed by Thompson [20]. Observe that c and γ for this Γ_{10}^* CR do not depend on the data $\mathbf{x}$, but they do depend on the sample size n.

4.2. For general symmetric distributions

We assumed the uniform family of distributions in the preceding subsection. A question arises whether we obtain the same CR if F belongs to the subfamily $\mathfrak{F}_{c,0}^{sym}$ of $\mathfrak{F}_{c,0}$ of continuous symmetric at zero distributions. For instance, the classes of normal, Cauchy, logistic, double exponential, symmetric mixtures of distributions all belong to this subclass. It turns out that Γ_{10}^* is also optimal for this larger (relative to the class of uniform distributions) subclass as a consequence of Theorem 3 below. The CR based on the Wilcoxon signed-rank statistic is valid for $\mathfrak{F}_{c,0}^{sym}$ [see (2.2)]; however, this CR is *not* a legitimate competitor of $\Gamma_{10}^*(X_{()}, U)$ under the class $\mathfrak{F}_{c,0}$ since it does *not* satisfy the confidence level requirement under non-symmetric distributions, which are allowed under the NMEM.

Theorem 3. *Let $X_1, \ldots, X_n$ be IID from $F \in \mathfrak{F}_{c,0}^{sym}$ and define $l(k; F) = E_F(X_{(k+1)}) - E_F(X_{(k)}), k = 0, 1, \ldots, n$, with $X_{(0)} = \inf\{x \in \Re : F(x) > 0\}$ and $X_{(n+1)} = \sup\{x \in \Re : F(x) < 1\}$. Let*

$$
r(k; F) = \frac{\binom{n}{k} 2^{-n}}{l(k; F)}, k = 0, 1, \ldots, n.
$$

(i) *If n is even, then $r(k; F)$ is maximized at $k = n/2$ and it decreases from the maximum as $|k - n/2|$ increases.*

(ii) *If n is odd, then $r(k; F)$ is maximized at $k \in \{(n-1)/2, (n+1)/2\}$ and decreases when $(n-1)/2 - k$ increases for $k \leq (n-1)/2$ or when $k - (n+1)/2$ increases for $k \geq (n+1)/2$.*

The results stated in Theorem 3 appear intuitive when the distribution $F \in \mathfrak{F}_{c,0}^{sym}$ is unimodal. What is surprising is that the results also hold when F is bi-modal or multi-modal. For instance, if we take a symmetric mixture of a $N(-\mu, \sigma)$ and a $N(\mu, \sigma)$, even when μ is quite large relative to σ, the results still hold true. We present a mathematical proof of Theorem 3. To do so we first establish an identity analogous to that given by Pearson in a paper of [7].

Lemma 4. *Let $X_{(1)} < \ldots < X_{(n)}$ be the associated order statistics for a random sample of size n from a continuous distribution F. For $k = 0, 1, 2, \ldots, n-1, n$, and provided that the expectations are well-defined with possibly a value of ∞, then*

$$l(k; F) \equiv E_F(X_{(k+1)}) - E_F(X_{(k)}) = \binom{n}{k} \int_{-\infty}^{\infty} F(x)^k (1 - F(x))^{n-k} dx.$$

Proof. Recall that for a positive-valued continuous random variable W, we have $E(W) = \int_0^{\infty} [1 - F_W(w)] dw$, hence for a general continuous W,

$$E(W) = \int_0^{\infty} [1 - F_W(w)] dw - \int_{-\infty}^0 F_W(w) dw.$$

For $k \in \{1, 2, \ldots, n\}$, we also recall that $P\{X_{(k)} \leq y\} = \sum_{j=k}^n \binom{n}{j} F(y)^j [1 - F(y)]^{n-j}$. Using these expressions, we immediately find that

$$
\begin{aligned}
l(k; F) &= E[X_{(k+1)}] - E[X_{(k)}] = E[X_{(k+1)}^+ - X_{(k)}^+] - E[X_{(k+1)}^- - X_{(k)}^-] \\
&= \int_0^{\infty} [(1 - P\{X_{(k+1)} \leq y\}) - (1 - P\{X_{(k)} \leq y\})] dy - \\
&\quad \int_{-\infty}^0 [P\{X_{(k+1)} \leq y\} - P\{X_{(k)} \leq y\}] dy \\
&= \binom{n}{k} \left[\int_0^{\infty} F(y)^k [1 - F(y)]^{n-k} dy + \int_{-\infty}^0 F(y)^k [1 - F(y)]^{n-k} dy \right] \\
&= \binom{n}{k} \int_{-\infty}^{\infty} F(y)^k [1 - F(y)]^{n-k} dy.
\end{aligned}
$$

$\square$

We now prove Theorem 3.

Proof. Using the representation in Lemma 4, we have for $k = 0, 1, \ldots, n$, that

$$r(k; F) = 2^{-n} \left\{ \int_{-\infty}^{\infty} (1 - F(x))^{n-k} F(x)^k dx \right\}^{-1}.$$

Let $Q(k; F) = \int_{-\infty}^{\infty} (1 - F(x))^{n-k} F(x)^k dx$. We prove the results by showing that $Q(k; F)$ is symmetric about $n/2$ and that it first decreases then increases with the maximum occurring at values of k that depends on whether n is odd or even. Making the transformation $u = F(x)$ in the expression of $Q(k; F)$ and denoting by f the density function of F, we have

$$
\begin{aligned}
Q(k; F) &= \int_0^1 \frac{(1-u)^{n-k} u^k}{f[F^{-1}(u)]} du \\
&= \int_0^{1/2} \frac{(1-u)^{n-k} u^k}{f[F^{-1}(u)]} du + \int_{1/2}^1 \frac{(1-u)^{n-k} u^k}{f[F^{-1}(u)]} du.
\end{aligned}
$$

In the last integral, let $w = 1 - u$ and note that since $F \in \mathfrak{F}_{c,0}^{sym}$, $F^{-1}(u) = -F^{-1}(1 - u)$ so that $f[F^{-1}(u)] = f[-F^{-1}(u)] = f[F^{-1}(1 - u)]$. Consequently,

$$
Q(k; F) = \int_0^{1/2} \frac{(1-u)^{n-k} u^k + (1-u)^k u^{n-k}}{f[F^{-1}(u)]} du.
$$

Letting $c(u) = 1/f[F^{-1}(u)]$ and $D(k; F) = Q(k; F) - Q(k + 1; F)$, it follows that

$$
\begin{aligned}
D(k; F) &= \int_0^{1/2} c(u)(1 - 2u) \left[(1-u)^{n-k-1} u^k - (1-u)^k u^{n-k-1} \right] du \\
&= \int_0^{1/2} c(u)(1 - 2u)(1-u)^{n-k-1} u^k \left[1 - \left(\frac{u}{1-u} \right)^{n-(2k+1)} \right] du.
\end{aligned}
$$

In the last integral, all terms in the integrand outside the brackets are nonnegative. Let $b(u, \alpha) = 1 - [u/(1 - u)]^\alpha$ for $\alpha \in \Re, u \in [0, 1/2]$. For this function, we have

$$
\lim_{u \downarrow 0} b(u; \alpha) = \begin{cases} 1 & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \\ -\infty & \text{if } \alpha < 0 \end{cases} \quad \text{and} \quad b(1/2; \alpha) = 0 \text{ all } \alpha.
$$

In addition, we have

$$
b'(u; \alpha) \equiv \frac{d}{du} b(u; \alpha) = -\alpha \left(\frac{u}{1-u} \right)^{\alpha-1} \frac{1}{(1-u)^2} \begin{cases} < 0 & \text{if } \alpha > 0 \\ = 0 & \text{if } \alpha = 0 \\ > 0 & \text{if } \alpha < 0 \end{cases}.
$$

Therefore, on $u \in (0, 1/2]$, $u \mapsto b(u; \alpha)$ is decreasing when $\alpha > 0$; constant at 0 when $\alpha = 0$; and increasing when $\alpha < 0$. Consequently,

$$
D(k; \alpha) = Q(k; F) - Q(k + 1; F) = \begin{cases} < 0 & \text{if } k < (n - 1)/2 \\ = 0 & \text{if } k = (n - 1)/2 \\ > 0 & \text{if } k > (n - 1)/2 \end{cases}.
$$

This completes the proof of the theorem. $\square$

5. Optimal CRs focused on exponential distributions

Next, we consider the situation where the focused class of distributions is the negative exponential family, a right-skewed class of distributions, in contrast to the symmetric distributions considered above. Let $X_1, \ldots, X_n$ be IID from an exponential distribution $(\mathrm{Exp}(\lambda))$ $F(x; \lambda) = [1 - \exp(-\lambda x)]I\{x \geq 0\}$ so the common median is $\Delta = \lambda^{-1} \log(2)$. From the normalized spacings theory (see, for instance, [1]) we have that

$$X_{(k)} = \sum_{j=1}^{k} D_j/(n - j + 1) \stackrel{d}{=} \sum_{j=1}^{k} X_j/(n - j + 1), k = 1, 2, \ldots, n,$$

where $D_j = (n-j+1)[X_{(j)}-X_{(j-1)}], j = 1, 2, \ldots, n$, are the normalized spacings statistics, which are also IID from $\mathrm{Exp}(\lambda)$. As such, with $X_{(0)} \equiv 0$, we obtain

$$l(k; F) = E[X_{(k+1)}] - E[X_{(k)}] = \frac{1}{\lambda(n - k)}, k = 0, 1, 2, \ldots, n.$$

It follows that under this negative exponential distribution model,

$$r(k; F) = \frac{\binom{n}{k} 2^{-n}}{[\lambda(n - k)]^{-1}} \propto \binom{n - 1}{k}, k = 0, 1, 2, \ldots, n.$$

The optimal Γ_0 will therefore be of form

$$\Gamma_0^* = \left\{ k \in \{0, 1, \ldots, n\} : \binom{n - 1}{k} > c \right\}$$

with c chosen to be the smallest value such that $P\{B \in \Gamma_0^*\} \leq 1 - \alpha$. Note that this subset Γ_0^* will never include n since $l(n; F) = \infty$, but it could include 0. We shall denote by $\Gamma_{11}^*(X_{()}, U)$ the resulting randomized CR for Δ under this exponentially-distributed focused case. We note that c and γ for this Γ_{11}^* CR also do not depend on $\mathbf{x}$, similarly to Γ_{10}^*.

We illustrate the resulting CR for concrete values of n. We start with $n = 10$, an even sample size. Using R [18] we obtain for each $k \in \{0, 1, 2, \ldots, n\}$ the values of $r(k; F)$ (up to proportionality) and $P\{B = k\}$ given in Table 1. Observe that the way we start including values of k into Γ_0^* is according to the value of $r(k)$. Thus we start by first including k-values of 4 and 5; then 3 and 6; etc. Observe the asymmetry in the process of including the k-values into Γ_0^* with a bias in favor of the lower k-values. The intuition behind this is that since the exponential distribution is highly right-skewed, then the expected lengths between successive order statistics increases as k increases, which is formally indicated by the $l(k; F) = 1/[\lambda(n-k)]$ expression. Thus, to shorten the interval, there is preference for the lower order statistics. For a 95% confidence region, from the table we find that $\Gamma_0^* = \{2, 3, 4, 5, 6, 7\}$ which yields $P\{B \in \Gamma_0^*\} = .9346$. The randomization probability then becomes

$$\gamma = \frac{.95 - .9346}{.0537} = .2867.$$

Values of $(k, r(k), P(B = k))$ for $k = 0, 1, 2, \ldots, n$ for the case $n = 10$ when F is assumed to be a negative exponential distribution.

k	$r(k)$	$P(B = k)$
0	1	0.0009765625
1	9	0.0097656250
2	36	0.0439453125
3	84	0.1171875000
4	126	0.2050781250
5	126	0.2460937500
6	84	0.2050781250
7	36	0.1171875000
8	9	0.0439453125
9	1	0.0097656250
10	0	0.0009765625

The 95% (randomized) confidence region for the median will then be

$$\Gamma_{11}^*(X, U) = [X_{(2)}, X_{(8)})I\{U > .2867\} + [X_{(1)}, X_{(9)})I\{U \le .2867\}.$$

When $n = 11$, an odd sample size, by following the same calculations as for $n = 10$ and with the first k-value to enter being $k = 5$, we find the 95% (randomized) confidence region for the median to be

$$\Gamma_{11}^*(X, U) = [X_{(3)}, X_{(8)})I\{U > .6758\} + [X_{(2)}, X_{(9)})I\{U \le .6758\}.$$

6. Data-adaptive methods

Next we demonstrate approach (ii). In this situation we do not know the $l(k; F)$s so we instead estimate these quantities using the observed data. Recall that $l(k; F) = E_F[X_{(k+1)} - X_{(k)}]$, so without any knowledge of the underlying F we will not have closed-form expressions for these $l(k; F)$s. However, given the sample data, we could estimate $l(k; F)$, *unbiasedly*, by

$$\hat{l}(k) = X_{(k+1)} - X_{(k)}, k = 0, 1, 2, \ldots, n, \tag{6.1}$$

which is the method-of-moments estimator. However, we surmise, just based on intuitive considerations, that this estimator may be generally unstable or inefficient relative to an alternative estimator obtained by replacing F in $l(k; F)$ by its empirical distribution $\hat{F}$ – see the estimators in (6.2). Our intuition is borne out of the fact that $\hat{F}$ is equivalent to the complete and sufficient statistic under the model $\mathfrak{F}_{c,0}$, hence the plug-in approach where F is replaced by $\hat{F}$ will lead to good estimators. Using the MM-based estimators in (6.1), we may then order the $k \in \{0, 1, \ldots, n\}$ in terms of priority of entry into Γ_0^* according to the quantities

$$\hat{r}(k) \propto \frac{\binom{n}{k}}{X_{(k+1)} - X_{(k)}}, k = 0, 1, 2, \ldots, n.$$

As such the form of the 'optimal' Γ_0 will be

$$\Gamma_0^* = \left\{ k \in \{0, 1, 2, \ldots, n\} : \frac{\binom{n}{k}}{X_{(k+1)} - X_{(k)}} > c \right\}.$$

We shall denote by $\Gamma_{12}^*(X_{()}, U)$ the resulting randomized CR. The constant c will be chosen to be the smallest value such that $P\{B \in \Gamma_0^* | X\} \leq 1 - \alpha$. This value of c will depend on the sample data X since the ordering of entry of the k-values into Γ_0^* will depend on X, so that the randomization probability γ will also depend on X. Because of this dependence of c and γ on the sample data X, it is possible that the achieved confidence coefficient of the resulting confidence region will *not* anymore be at least $100(1 - \alpha)\%$. In the simulation studies in section 8 we will indeed see that there is a non-negligible degradation in terms of the achieved confidence level for this CR.

As indicated in the preceding paragraph, a potentially better adaptive approach may be obtained by utilizing the representation of $l(k; F)$ in Lemma 4 and replacing the unknown $F(\cdot)$ in the expression of $l(k; F)$ by its empirical distribution function (EDF), $\hat{F}$, based on the $X_i - \hat{\Delta}, i = 1, 2, \ldots, n$. This may lead to a better procedure since the resulting estimators of $l(k; F)$s may be more stable compared to the MM estimators in (6.1). The EDF is $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i - \hat{\Delta} \leq t\}, t \in \Re$, so the estimators of $l(k; F)$s are, for $k = 1, 2, \ldots, n$,

$$
\begin{aligned}
\hat{l}(k) &= \binom{n}{k} \int_{-\infty}^{\infty} (1 - \hat{F}(x))^{n-k} \hat{F}(x)^k dx \\
&= \binom{n}{k} \sum_{i=2}^{n} \left[1 - \frac{i-1}{n} \right]^{n-k} \left[\frac{i-1}{n} \right]^k (X_{(i)} - X_{(i-1)}). \quad (6.2)
\end{aligned}
$$

Observe that we could also have used the EDF of the X_is instead of the $(X_i - \hat{\Delta})$s since the $\hat{\Delta}$ cancels out. The ordering of $k \in \{0, 1, \ldots, n\}$ in terms of entry into Γ_0^* is then based on

$$\hat{r}(k) \propto \frac{\binom{n}{k}}{\hat{l}(k)} = \left\{ \sum_{i=2}^{n} \left[1 - \frac{i-1}{n} \right]^{n-k} \left[\frac{i-1}{n} \right]^k [X_{(i)} - X_{(i-1)}] \right\}^{-1}.$$

The resulting randomized CR for Δ will be denoted by $\Gamma_{13}^*(X_{()}, U)$.

These adaptive procedures are totally nonparametric in the sense that no knowledge of the underlying distribution F is required. If we do have knowledge of the families of distributions in which F belongs, then we may be able to provide better estimators of the $l(k; F)$s as in the subsections dealing with symmetric Fs and exponential F. It should be noted, however, that these adaptive procedures may not anymore satisfy the confidence level requirement due to the data-dependent plug-in step. Section 8, which presents results of simulation studies, provides some insights regarding the empirical properties of these procedures. Looking ahead, based on these simulation studies, the impact of this data double-dipping on Γ_{13}^* in terms of its coverage probability could be characterized as slight.

List of the confidence regions (CRs) considered in the illustrations and simulations.

Method Label	Description or Type or Basis of CR
Γ_1	T Statistic Based
Γ_2	Wilcoxon Signed-Rank Statistic Based
Γ_3	Sign Statistic Based
Γ_4	Asymptotic Distribution of Median Based
Γ_5	Basic Median Bootstrap
Γ_6	Median with Bootstrapped SE
Γ_7	Percentile Bootstrapped
Γ_8	Bias-Corrected (BC) Bootstrapped
Γ_9	Bias-Corrected Accelerated (BCa) Bootstrapped
Γ_{10}	Optimally Focused for Symmetric
Γ_{11}	Optimally Focused for Exponential
Γ_{12}	Adaptive with $l(k)$ Unbiasedly Estimated
Γ_{13}	Adaptive with $l(k)$ Estimated using EDF

7. Illustration using real data sets

Together with the 'off-the-shelf' CRs for Δ discussed in section 2, we now add the CRs that were just developed (we drop the superscript '*' in the notation): Γ_{10}, the CR optimized for symmetric distributions [see equation (4.1)]; Γ_{11}, the CR optimized for exponential distributions (see section 5); Γ_{12}, the adaptive CR using the crude estimator for $l(k; F)$; and Γ_{13}, the adaptive CR using the empirical distribution in the estimator of $l(k; F)$ (see section 6). These different CRs, which are implemented in the illustrations and the simulation studies, are summarized in Table 2. We note that among these thirteen CRs, those that are location-equivariant and equivariant to monotone transformations are Γ_3, Γ_{10}, Γ_{11}, Γ_{12}, and Γ_{13}.

For our first illustration we use a real data set gathered by the first author about the mileage efficiency of his car during a period [from October 20, 1996 to January 27, 1999] when he was commuting between Ann Arbor, Michigan and Bowling Green, Ohio. Efficiency is measured in terms of the miles traveled per gallon of gasoline. In the data set there were $n = 205$ observations, with each observation recorded at each gas fill-up. The histogram and time plot of these observations are shown in Figure 1. The full data set, which contained other relevant variables, was also used in demonstrating global validation procedures of linear model assumptions in [16]. For this sample data set the mean is $\bar{x} = 29.291$, median is $\hat{\Delta} = 29.462$, the first and third sample quartiles are 28.275 and 30.575, respectively, and the extreme values are 18.374 and 33.472. From the histogram we notice that there is a noticeable left-skewness in the distribution. The sample standard deviation is 1.887 and the inter-quartile range is 2.299. Letting $X_i, (i = 1, 2, \ldots, n)$, be the random observable denoting the fuel efficiency preceding the ith gas fill-up, we assume that these X_i's are IID from some unknown distribution H whose median is Δ. The distribution H could be viewed as a mixture of subpopulation distributions corresponding to summer highway driving, winter highway driving, summer city driving, winter city driving, and a combination of highway and city driving. The assumption that the
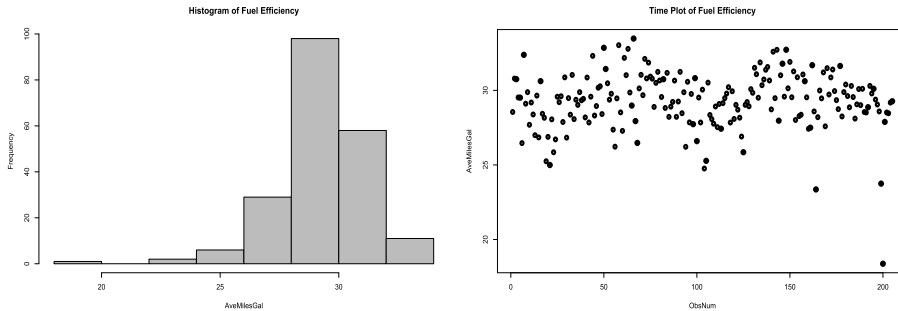
Fɪɢ 1. *Histogram and Time Plot of 205 observations for the variable Average Miles Per Gallon (*`AveMilesGal`*) for the data set gathered by the first author pertaining to his car.*

X_i's are identically distributed is an approximate one since the fuel efficiency could have decreased as the car got older (though not evident from the data), when lubricant oil quality degraded between in-between maintenances, or most plausibly it was affected by the driving season (indeed, one could discern a seasonal trend in the time plot in Figure 1). The independence assumption of the X_is, however, may still be somewhat tenable since there is no reason for X_{i-1} to have an effect on X_i. The Pearson correlation coefficient of the bivariate data $\{(i, x_i), i = 1, 2, \ldots, n\}$ was 0.0113, leading to a z-score of 0.1607, with associated p-value of 0.8723 for testing $H_0 : \rho = 0$, where a Fisher's z-transformation was applied. The Spearman's correlation coefficient was 0.0634 with p-value of 0.3663. We recognize that these tests are merely for association between ranked time and fuel efficiency, not for the independence among the X_is. We also tested for the presence of first-order autocorrelation using Bartel's [2] nonparametric rank version of von Neumann's ratio test for randomness. For the test of the null hypothesis that there was no serial correlation, the p-value was 0.14 using a normal approximation mentioned in [2].

For our illustration, we seek a confidence region for the population median Δ, where the population could be thought of as the hypothetical set of all values of `AveMilesGal` if the car has been observed forever. Figure 2 depicts the 95% confidence regions produced by the thirteen different methods for the `AveMilesGal` data set. Except for CR Γ_{12}, all of them produced intervals. It is possible that method Γ_{13} may also produce a region that is not an interval, while CRs Γ_1 to Γ_{11} will all produce confidence *intervals* by construction. In addition, observe that only the CRs produced by Γ_4 and Γ_6 turned out to be symmetric about the sample median.

Our other illustration uses Proschan's [17] famous Boeing air-conditioning data set. Figure 3 provides a histogram of the 212 observations in this data set, with each observation being a time between successive failures. To un-tie tied observations, we perturbed the original values in tied sets of observations by uniform variates over $[-.001, .001]$. Observe that this data set is highly right-skewed. Proschan also demonstrated that this data set did not come from an
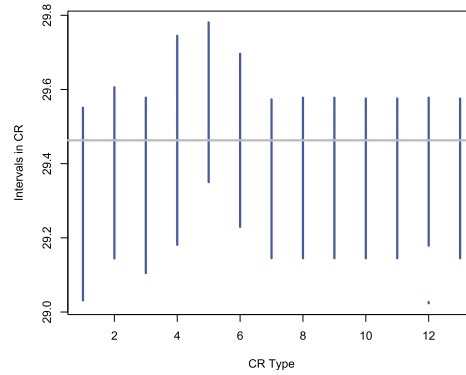
FIG 2. *The 95% confidence regions for the median of the* `AveMilesGal` *population produced by the thirteen different methods in Table 2. The horizontal line depicts the sample median.*
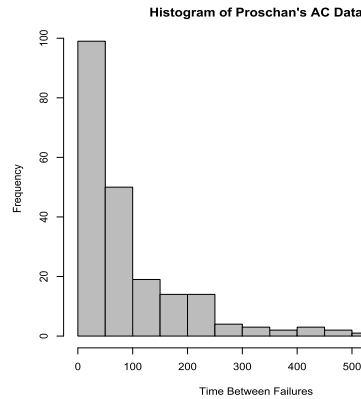
FIG 3. *Histogram of Proschan's Boeing air-conditioning data set. The 212 observations are times between failures, and there were several airplanes that were monitored.*

exponential distribution, but came from a mixture of exponential distributions, inducing a decreasing failure rate property. Our goal in this illustration is to provide a confidence region for the median inter-failure time given this data set. Figure 4 presents the comparative plots of the CRs for the thirteen methods. Notice that the T-based CR, as well as the Wilcoxon signed-rank based CR, are very different from the other eleven CRs. In fact, both excluded the sample median. At the same time, we are somewhat unfair in the manner in which we applied these two procedures since it would have been more appropriate to first perform a transformation to approximately symmetrize the distribution prior to applying these procedures, and then re-transforming back. Also, notice that the Γ_{12} CR is composed of one big interval together with several smaller intervals. This could be attributed to our surmised instability of the method-of-moments estimator of the $l(k; F)$'s. Interestingly, this CR has the smallest content, but
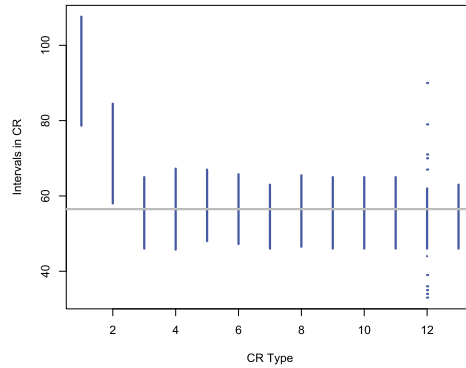
FIG 4. *The 95% confidence regions for the median based on the thirteen methods for the Proschan's Boeing air-conditioning data set. The horizontal line represents the sample median.*

we will see in the simulation studies in section 8 that its achieved coverage probability of the median tends to be non-negligibly lower than the nominal confidence level.

8. Comparison of methods via simulations

In this section we present the results of simulation studies to compare the thirteen CR methods listed in Table 2 in terms of their mean *standardized* contents and coverage probabilities under the NMEM. Different error distributions [normal, Cauchy, uniform, Laplace, logistic, normal mixture, sinh-arcsinh (SAS), gamma, and Weibull] and varied sample sizes,

$$n \in \{10, 15, 20, 25, 30, 40, 50, 75, 100\},$$

were utilized in the simulations. The authors are indebted to one of the referees for his/her suggestion to include the SAS distribution – the *sinh-arcsinh* distribution proposed by Jones and Pewsey [13] – to model a skewed distribution whose support is the whole real line $\Re$. The SAS distribution has four parameters: $(\mu, \sigma, \delta, \epsilon)$, and its distribution function is given by

$$F(x; \mu, \sigma, \delta, \epsilon) = \Phi\left[\sinh(\delta \operatorname{arcsinh}((x - \mu)/\sigma) - \epsilon)\right], \tag{8.1}$$

which has median of $\Delta = \mu + \sigma \sinh(\epsilon/\delta)$. The computer programs for the simulation studies were coded in R [18]. For each combination of error distribution and sample size, 20000 simulation replications were run, and with 5000 bootstrap replications for the bootstrap-based methods.

The results of these simulation studies are presented in a series of tables and figures: Tables 3–11 and Figures 5–7. We use the results in the tables to determine the CRs that are not competitive under the NMEM, and then present in graphical form the performances of the competitive CRs in the figures. Table

TABLE 3. *Simulation results under an F that is a standard normal distribution.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	95.1	95.0	94.9	94.8	95.0	95.2	94.8	95.2	95.2	4.4	4.2	4.1	4.1	4.1	4.0	4.0	4.0	4.0
2	95.2	95.4	95.3	94.9	95.1	95.2	95.0	95.1	95.3	4.5	4.3	4.2	4.2	4.1	4.1	4.1	4.1	4.1
3	97.9	96.5	96.0	95.5	95.8	96.1	96.7	96.6	96.7	6.3	5.5	5.3	5.2	5.2	5.3	5.4	5.3	5.3
4	92.6	93.4	94.8	94.6	95.2	95.3	95.3	95.4	95.6	5.0	5.1	5.1	5.1	5.2	5.2	5.2	5.1	5.1
5	82.1	79.8	83.0	84.8	84.3	85.5	86.3	86.1	88.3	4.7	4.9	4.7	5.2	4.7	4.8	4.9	4.8	4.9
6	95.4	95.2	94.7	94.9	94.5	94.2	94.1	94.3	94.2	5.5	5.6	5.3	5.4	5.2	5.1	5.1	5.1	5.0
7	94.0	92.8	93.9	95.4	94.2	94.6	94.8	94.2	95.1	4.7	4.9	4.7	5.2	4.7	4.8	4.9	4.8	4.9
8	91.9	92.2	94.2	93.3	94.6	94.4	94.3	93.8	94.9	4.6	4.8	4.8	5.0	4.8	4.8	4.8	4.8	4.9
9	92.0	92.2	94.2	93.2	94.6	94.4	94.4	93.8	94.9	4.6	4.8	4.8	5.0	4.8	4.8	4.8	4.8	4.9
10	95.1	95.0	95.2	94.9	95.0	94.9	95.0	95.1	95.2	5.6	5.2	5.1	5.1	5.0	5.0	5.0	5.0	5.0
11	95.1	94.9	95.2	94.8	95.1	94.9	94.8	95.2	95.1	7.1	5.3	5.1	5.4	5.0	5.0	5.0	5.0	5.0
12	91.7	91.3	91.4	90.5	90.7	90.2	89.7	89.4	89.6	5.2	4.9	4.8	4.7	4.6	4.6	4.6	4.5	4.5
13	93.7	94.1	94.6	94.3	94.5	94.2	94.1	94.5	94.6	5.4	5.2	5.1	5.0	5.0	5.0	5.0	4.9	4.9

TABLE 4. *Simulation results under an F that is a standard Cauchy distribution.*

Mtd	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	98.1	98.0	98.1	97.9	98.1	98.2	97.9	98.0	98.0	130.1	449.8	149.6	162.1	172.9	201.2	1178.8	228.2	391.0
2	95.2	95.2	95.0	95.0	95.1	95.1	95.3	95.0	95.0	143.5	16.0	11.4	10.0	9.2	8.5	8.2	7.8	7.6
3	98.0	96.4	95.8	95.6	95.5	96.3	96.9	96.4	96.5	18.9	10.1	8.5	7.9	7.6	7.4	7.5	7.0	7.0
4	96.7	97.4	97.9	98.0	98.1	98.2	98.4	98.1	98.1	9.5	8.9	8.6	8.4	8.3	8.1	8.0	7.8	7.7
5	93.5	88.2	90.2	90.9	89.5	90.1	90.5	88.7	90.8	11.9	8.5	7.3	7.9	6.8	6.7	6.6	6.3	6.4
6	98.9	98.3	97.3	97.2	96.4	95.9	95.8	95.4	95.3	36.6	10.1	8.3	8.0	7.4	7.1	6.9	6.7	6.6
7	94.3	92.8	93.7	95.5	94.0	94.8	94.9	94.2	94.9	11.9	8.6	7.3	7.9	6.8	6.7	6.7	6.3	6.4
8	92.5	92.0	94.0	93.3	94.2	94.5	94.4	93.7	94.8	10.6	8.3	7.6	7.6	7.0	6.7	6.6	6.4	6.4
9	92.5	92.0	94.0	93.3	94.1	94.5	94.5	93.6	94.8	10.6	8.3	7.6	7.6	7.0	6.7	6.6	6.4	6.4
10	95.2	94.9	94.9	95.0	94.8	95.2	95.0	95.0	95.0	15.5	9.5	8.2	7.7	7.3	7.0	6.9	6.6	6.5
11	94.9	94.8	94.9	95.0	94.9	95.2	95.0	95.1	95.0	113.4	9.5	8.2	8.5	7.3	7.0	6.9	6.6	6.5
12	95.5	94.2	93.3	92.8	92.2	91.8	91.3	90.8	90.5	12.1	8.6	7.5	7.1	6.7	6.4	6.2	6.0	5.9
13	96.9	96.0	95.8	95.7	95.4	95.7	95.5	95.2	95.1	13.9	9.4	8.1	7.6	7.3	6.9	6.7	6.5	6.4

TABLE 5. *Simulation results under an F that is a standard uniform distribution.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	94.4	94.5	94.8	94.8	94.9	95.0	95.2	95.0	94.9	2.6	2.5	2.4	2.4	2.4	2.3	2.3	2.3	2.3
2	95.0	95.0	95.1	95.1	95.0	95.0	95.4	95.2	95.0	2.8	2.7	2.6	2.6	2.5	2.5	2.5	2.4	2.4
3	97.8	96.4	96.0	95.5	95.8	96.1	96.7	96.3	96.4	4.0	3.9	3.8	3.8	3.9	4.0	4.2	4.1	4.2
4	88.0	88.3	90.5	90.6	91.3	92.0	92.7	93.1	93.5	3.3	3.5	3.6	3.6	3.7	3.7	3.8	3.8	3.8
5	74.2	73.8	78.6	80.8	81.1	83.1	84.7	84.8	87.7	3.1	3.5	3.4	3.8	3.6	3.7	3.8	3.7	3.9
6	93.0	93.5	93.2	93.6	93.1	93.4	93.6	93.8	93.7	3.8	4.1	3.9	4.0	3.9	3.9	4.0	4.0	4.0
7	94.1	92.7	93.7	95.4	94.3	94.6	94.8	94.1	94.8	3.1	3.5	3.4	3.8	3.6	3.7	3.8	3.7	3.9
8	91.7	92.0	94.0	93.3	94.4	94.3	94.4	93.8	94.5	3.1	3.4	3.5	3.7	3.6	3.7	3.7	3.7	3.8
9	91.8	92.0	93.9	93.3	94.4	94.2	94.3	93.7	94.6	3.1	3.4	3.5	3.7	3.6	3.7	3.7	3.7	3.8
10	94.8	94.9	95.1	94.9	95.0	94.9	95.0	94.9	94.9	3.6	3.7	3.7	3.8	3.8	3.8	3.9	3.9	3.9
11	95.1	94.8	95.1	94.6	95.0	95.0	94.9	94.9	94.9	4.1	3.7	3.7	4.0	3.8	3.8	3.9	3.9	3.9
12	88.9	89.1	89.2	89.2	89.5	89.3	89.5	89.3	89.4	3.3	3.4	3.5	3.5	3.5	3.5	3.5	3.5	3.5
13	91.6	92.9	93.8	93.5	93.8	93.6	93.8	94.1	94.1	3.5	3.6	3.7	3.7	3.8	3.8	3.8	3.8	3.9

TABLE 6. *Simulation results under an F that is a standard Laplace or double exponential distribution.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	95.9	95.5	95.3	95.2	95.1	95.1	95.1	95.1	95.3	6.0	5.8	5.7	5.7	5.7	5.6	5.6	5.6	5.6
2	95.1	95.1	95.2	95.0	94.9	95.0	95.2	94.9	95.1	6.1	5.5	5.2	5.1	5.0	4.9	4.9	4.7	4.7
3	97.8	96.5	95.9	95.7	95.7	96.1	96.8	96.2	96.5	7.8	6.1	5.6	5.3	5.2	5.1	5.2	4.9	4.8
4	95.3	96.4	97.1	97.1	97.4	97.6	97.7	97.7	98.0	5.6	5.6	5.5	5.5	5.4	5.4	5.3	5.2	5.1
5	89.8	86.3	88.6	90.2	89.3	90.2	90.6	90.1	91.7	5.5	5.3	4.8	5.3	4.7	4.7	4.6	4.4	4.4
6	97.3	97.3	96.7	96.6	96.1	96.0	95.8	95.8	95.6	6.3	6.0	5.4	5.4	5.1	4.9	4.8	4.6	4.5
7	94.2	93.0	93.7	95.6	94.1	94.6	94.9	94.0	95.1	5.5	5.3	4.8	5.3	4.7	4.7	4.6	4.4	4.4
8	92.1	92.1	94.1	93.2	94.2	94.5	94.3	93.5	94.8	5.3	5.2	5.0	5.1	4.8	4.7	4.6	4.4	4.4
9	92.1	92.0	94.1	93.3	94.2	94.5	94.5	93.6	94.8	5.3	5.2	5.0	5.1	4.8	4.7	4.6	4.4	4.4
10	94.9	95.0	95.1	95.0	94.8	94.9	95.1	94.9	95.1	6.8	5.8	5.4	5.2	5.1	4.9	4.8	4.6	4.5
11	94.8	95.1	95.0	95.0	95.0	95.0	95.0	94.9	95.1	9.6	5.8	5.4	5.6	5.1	4.9	4.8	4.6	4.5
12	93.7	93.1	92.8	92.4	92.2	92.1	91.9	91.3	91.1	6.1	5.4	5.0	4.8	4.6	4.4	4.3	4.1	4.0
13	95.1	95.3	95.2	95.2	94.9	95.0	95.2	95.1	95.1	6.4	5.7	5.3	5.1	5.0	4.8	4.7	4.5	4.4

TABLE 7. *Simulation results under an F that is a standard logistic distribution.*

	Coverage Percentage									Standardized Mean Content								
Method	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	95.4	95.1	95.1	95.0	95.2	94.9	95.1	95.0	95.0	7.9	7.6	7.5	7.4	7.3	7.3	7.2	7.2	7.2
2	95.1	95.2	95.3	95.1	94.8	94.8	95.3	95.1	95.2	8.0	7.5	7.3	7.2	7.1	7.0	7.0	6.9	6.9
3	97.9	96.4	96.0	95.5	95.6	96.2	96.8	96.2	96.4	10.9	9.2	8.7	8.5	8.4	8.5	8.7	8.5	8.5
4	93.5	94.2	95.3	95.1	95.6	95.9	96.0	95.5	95.9	8.4	8.4	8.5	8.5	8.5	8.5	8.4	8.4	8.4
5	84.2	81.5	83.8	86.1	84.8	86.3	87.0	86.0	88.5	7.9	8.1	7.6	8.4	7.6	7.8	7.9	7.7	7.9
6	96.1	95.8	94.9	94.9	94.6	94.4	94.3	94.4	94.0	9.3	9.2	8.6	8.7	8.4	8.3	8.2	8.2	8.1
7	94.4	92.9	93.7	95.3	94.1	94.6	94.9	93.8	94.9	7.9	8.1	7.6	8.4	7.6	7.8	7.8	7.7	7.9
8	92.5	92.2	93.9	93.2	94.4	94.4	94.5	93.4	94.6	7.7	7.9	7.8	8.1	7.8	7.8	7.7	7.8	7.8
9	92.5	92.2	93.8	93.2	94.4	94.4	94.4	93.5	94.7	7.7	7.9	7.8	8.1	7.8	7.8	7.7	7.7	7.8
10	95.2	94.9	95.1	94.8	94.9	95.0	95.1	94.8	94.9	9.6	8.7	8.4	8.3	8.2	8.1	8.1	8.0	8.0
11	94.9	94.8	95.2	94.7	94.9	94.8	95.2	94.9	95.0	12.6	8.7	8.4	8.9	8.2	8.1	8.1	8.0	8.0
12	92.4	91.5	91.3	90.7	90.4	90.3	90.3	89.8	89.7	8.8	8.1	7.8	7.7	7.5	7.4	7.3	7.3	7.2
13	94.4	94.3	94.5	94.3	94.5	94.4	94.4	94.2	94.5	9.2	8.6	8.4	8.2	8.1	8.1	7.9	7.9	7.9

TABLE 8. *Simulation results under a mixture of normal distributions. The mixture is* $.6 * Normal(-5, 3) + .4 * Normal(5, 2)$.

	Coverage Percentage									Standardized Mean Content								
Method	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	92.3	90.2	87.4	85.3	82.6	78.1	73.2	61.2	50.4	24.8	23.7	23.2	22.9	22.7	22.5	22.3	22.2	22.1
2	93.2	90.8	87.8	85.7	82.8	78.3	73.8	61.6	51.2	26.9	26.3	25.6	25.1	24.6	24.1	24.0	23.3	23.0
3	97.9	96.6	95.8	95.7	95.9	96.4	96.7	96.1	96.6	37.8	37.8	38.7	40.0	41.3	43.9	46.5	47.6	48.9
4	82.3	81.4	83.8	83.5	85.4	86.9	88.3	90.0	91.5	32.0	32.9	34.4	34.8	35.8	36.7	37.4	38.4	39.1
5	65.4	64.0	69.1	71.8	73.0	76.3	79.2	82.0	85.3	30.2	34.4	35.1	39.8	38.2	40.8	42.4	43.4	45.1
6	90.7	91.5	90.1	91.1	90.5	91.6	91.9	93.1	93.6	38.3	42.6	41.6	44.4	43.7	45.1	46.0	47.5	47.0
7	94.2	93.2	93.4	95.6	94.1	94.9	94.9	94.0	95.0	30.2	34.3	35.1	39.9	38.3	40.8	42.4	43.3	45.1
8	91.4	92.5	93.2	93.3	94.0	94.5	94.2	93.7	94.7	30.0	34.3	36.0	39.4	39.3	41.2	42.6	45.3	45.5
9	91.5	92.5	93.1	93.3	94.0	94.4	94.2	93.8	94.7	30.0	34.3	36.0	39.4	39.4	41.1	42.6	45.3	45.5
10	94.9	95.3	94.8	95.0	95.0	95.3	95.0	94.9	95.2	34.7	36.4	37.8	39.2	40.5	42.2	43.5	45.1	45.8
11	94.9	95.1	94.9	95.0	95.2	95.2	95.0	94.7	95.2	39.3	36.2	37.7	39.7	40.3	41.8	42.7	44.2	45.2
12	87.5	87.6	87.1	87.5	87.5	88.4	88.7	89.2	89.3	31.9	33.5	34.6	35.5	36.4	37.6	38.4	39.6	39.9
13	91.8	93.0	92.4	92.7	92.3	92.7	93.1	93.4	93.5	33.4	35.6	37.2	38.4	39.5	41.0	42.2	43.8	44.3

TABLE 9. *Simulation results under a SAS distribution with $\mu = 0, \sigma = 1, \delta = 1, \epsilon = 1$. See (8.1) for the SAS distribution formula.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	93.7	90.2	85.8	81.5	77.1	68.3	58.4	40.2	24.6	7.0	6.7	6.6	6.6	6.5	6.5	6.4	6.4	6.4
2	94.0	92.6	90.8	89.8	88.2	85.8	83.2	76.6	69.5	7.2	7.0	6.8	6.8	6.7	6.7	6.7	6.6	6.6
3	97.9	96.6	95.9	95.7	95.5	96.1	96.8	96.4	96.5	9.8	8.6	8.1	8.0	8.0	8.2	8.4	8.1	8.2
4	91.3	92.4	93.1	93.2	93.7	94.0	94.2	94.0	94.2	7.4	7.4	7.5	7.5	7.5	7.5	7.5	7.5	7.5
5	79.3	76.5	80.0	82.3	81.9	83.8	85.2	84.6	87.7	7.2	7.5	7.2	8.0	7.3	7.5	7.5	7.4	7.6
6	95.0	95.1	94.2	94.4	93.9	94.1	94.0	94.0	93.9	8.7	8.8	8.2	8.4	8.0	8.0	7.9	8.0	7.8
7	94.1	93.0	93.6	95.6	93.8	94.7	95.0	94.2	95.0	7.2	7.5	7.2	8.0	7.3	7.5	7.5	7.4	7.6
8	92.0	92.2	93.7	93.0	94.0	94.5	94.4	93.8	94.7	7.5	7.8	7.6	8.2	7.6	7.6	7.5	7.6	7.6
9	91.9	92.2	93.7	93.1	94.1	94.5	94.5	93.8	94.7	7.5	7.8	7.6	8.2	7.6	7.6	7.5	7.6	7.6
10	95.1	95.2	95.1	95.1	94.7	95.0	95.0	95.1	95.0	8.7	8.1	7.9	7.9	7.8	7.8	7.8	7.7	7.7
11	95.1	95.1	95.0	95.0	94.9	94.9	95.1	95.1	95.0	8.7	8.0	7.9	7.9	7.8	7.7	7.6	7.6	7.6
12	91.3	90.7	90.3	90.1	90.1	90.2	89.5	89.5	89.4	7.7	7.4	7.2	7.2	7.1	7.0	7.0	6.9	6.9
13	93.6	93.9	93.7	93.9	93.9	93.9	94.4	94.4	94.5	8.1	7.8	7.7	7.7	7.6	7.6	7.6	7.6	7.6

TABLE 10. *Simulation results under a gamma distribution with shape parameter 2 and scale parameter 1.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	94.3	92.0	89.8	86.0	82.5	75.8	68.5	51.3	37.3	6.1	5.9	5.8	5.7	5.7	5.6	5.6	5.6	5.6
2	94.3	93.5	93.1	91.9	91.0	89.1	88.0	83.1	78.8	6.3	5.9	5.8	5.7	5.6	5.5	5.5	5.5	5.5
3	97.7	96.7	95.8	95.7	95.8	96.1	96.8	96.5	96.4	8.3	7.2	6.8	6.7	6.7	6.7	6.9	6.7	6.8
4	91.6	92.7	93.8	93.7	94.2	94.3	94.9	94.7	95.0	6.2	6.3	6.3	6.4	6.4	6.4	6.4	6.4	6.4
5	80.0	78.7	81.4	83.6	83.5	84.8	86.2	85.8	87.9	6.1	6.3	6.0	6.7	6.1	6.2	6.2	6.1	6.3
6	95.2	95.3	94.5	94.8	94.0	93.9	94.3	94.3	94.3	7.3	7.3	6.8	6.9	6.6	6.6	6.5	6.5	6.4
7	94.1	92.9	93.7	95.6	94.2	94.4	95.0	94.3	95.0	6.1	6.3	6.0	6.7	6.1	6.2	6.2	6.1	6.3
8	91.8	92.3	94.0	93.1	94.3	94.2	94.6	93.8	94.8	6.3	6.5	6.3	6.7	6.3	6.2	6.2	6.2	6.2
9	91.8	92.2	94.0	93.2	94.4	94.3	94.5	93.7	94.8	6.3	6.5	6.3	6.7	6.3	6.2	6.2	6.2	6.2
10	94.8	95.3	94.8	95.0	95.1	94.9	95.2	95.2	94.9	7.4	6.8	6.6	6.5	6.5	6.4	6.4	6.4	6.3
11	95.0	95.2	95.0	95.1	95.1	95.0	95.2	95.2	94.8	7.4	6.7	6.6	6.6	6.4	6.4	6.3	6.3	6.3
12	91.5	91.3	90.8	90.6	90.3	89.7	90.1	89.6	89.6	6.6	6.3	6.1	6.0	5.9	5.8	5.8	5.7	5.7
13	93.8	94.1	94.1	94.1	94.2	93.9	94.5	94.6	94.4	6.8	6.6	6.5	6.4	6.4	6.3	6.3	6.3	6.3

TABLE 11. *Simulation results under a Weibull distribution with shape parameter .5 and scale parameter 1.*

Method	Coverage Percentage									Standardized Mean Content								
	10	15	20	25	30	40	50	75	100	10	15	20	25	30	40	50	75	100
1	93.2	82.2	67.1	52.1	38.4	19.8	8.9	1.4	0.3	15.0	15.1	15.5	15.5	15.7	16.0	16.1	16.4	16.6
2	90.8	85.2	80.1	75.8	69.9	61.6	52.7	35.9	23.3	16.0	11.3	10.2	9.6	9.3	8.9	8.7	8.5	8.3
3	98.0	96.3	96.0	96.0	95.7	96.1	96.8	96.3	96.6	13.3	8.9	7.7	7.1	6.9	6.7	6.8	6.3	6.3
4	90.2	91.0	92.0	91.9	92.4	91.7	92.2	90.9	90.9	6.8	6.3	6.0	5.7	5.6	5.3	5.2	4.9	4.8
5	73.8	71.4	76.7	77.8	78.8	80.6	82.5	83.4	86.0	9.0	7.6	6.6	7.1	6.2	6.1	6.0	5.7	5.7
6	97.0	96.1	95.5	95.5	95.1	94.3	94.5	94.4	94.3	11.3	9.2	7.9	7.6	7.1	6.6	6.4	6.2	6.0
7	93.8	92.6	93.8	95.8	94.2	94.5	94.9	94.1	94.9	9.0	7.6	6.6	7.1	6.2	6.1	6.0	5.7	5.7
8	92.0	91.9	93.7	93.6	94.3	94.2	94.1	93.6	94.7	10.0	8.7	7.4	8.0	6.7	6.3	6.2	6.1	5.8
9	92.0	91.9	93.7	93.6	94.3	94.3	94.2	93.7	94.7	10.0	8.7	7.4	8.1	6.8	6.3	6.2	6.1	5.8
10	95.1	94.7	95.1	95.3	94.9	94.9	95.0	94.9	95.0	11.2	8.4	7.5	6.9	6.7	6.4	6.2	5.9	5.8
11	95.3	94.8	95.0	95.2	95.0	94.9	95.0	95.0	95.1	8.2	7.9	7.3	6.3	6.6	6.1	5.9	5.7	5.7
12	93.0	91.3	91.3	91.2	90.7	90.2	90.1	89.7	89.5	7.9	6.6	6.1	5.8	5.7	5.4	5.4	5.2	5.1
13	94.9	94.2	94.3	94.5	94.1	94.2	94.3	94.2	94.4	8.2	7.0	6.5	6.2	6.1	5.8	5.8	5.7	5.6

[3](#) summarizes the standardized mean lengths and coverage probabilities of the different CRs for each of the sample sizes under a normal error distribution. Observe that the T-statistic based CR, Γ_1, performs best in terms of mean length with a coverage probability close to the desired nominal confidence level of 95%. This domination is expected from theory. Methods Γ_5 and Γ_{12} have coverage probabilities that are way below the nominal confidence level. Table 4 presents the results under the Cauchy error distribution. Again, Γ_5 and Γ_{12} did not perform well in terms of their coverage probabilities, while Γ_1 has an unacceptable mean length. Γ_2 and Γ_{11} also have inflated mean lengths when the sample size is small. Table 5 is for a uniform error distribution and this reveals that Γ_4, Γ_6 and Γ_{13} have slightly lower coverage probabilities relative to the nominal confidence level. Table 6, which is for the Laplace distribution, and Table 7, which is for the logistic distribution, both indicate that Γ_5 and Γ_{12} are not competitive. Table 8 for the normal mixture distribution leads to the elimination of CRs Γ_1, Γ_2, and Γ_4 due to their poor coverage probabilities. Γ_6 and Γ_{13} also have slightly degraded coverage probabilities. Table 9, which is for the SAS distribution, a skewed distribution on the whole real line, eliminates Γ_1 and Γ_2, and Γ_6's coverage probability is again slightly degraded. Table 10, which is for the gamma distribution, again demonstrates that Γ_1 and Γ_2 should not be used at all when the distribution is skewed, and Table 11, which is for the Weibull distribution, also leads to the elimination of Γ_4 due to its poor coverage probability. Therefore, examining these tables, the CRs that are competitive *under the NMEM* are Γ_3, Γ_7, Γ_8, Γ_9, Γ_{10}, and Γ_{13}.

To get a better picture of the performances of these six competitive CRs, we plotted their standardized mean contents and coverage probabilities with respect to the sample size, and these plots are presented in Figure 5, Figure 6, and Figure 7. Examining these plots we see that the sign-based test, Γ_3, and the procedure focused on symmetric distributions, Γ_{10}, are consistently satisfying the confidence level requirement among all the distributions considered, including the skewed SAS, gamma, and Weibull distributions, with Γ_3 tending to be more conservative. Γ_{10}, which as pointed out earlier is a randomized version of Γ_3, has coverage probability that is very close to the nominal confidence level, hence it also has smaller content compared to Γ_3. The PCT-Bootstrap Γ_7, the BC-Bootstrap Γ_8, and the BCa-Bootstrap Γ_9, tended to be liberal, and so is the adaptive Γ_{13}, though the latter could sometimes be conservative such as when the distribution is Cauchy, or it could also be more liberal than Γ_7, Γ_8 and Γ_9 as in the case for the mixture of normal distribution. Content-wise, Γ_7, Γ_8, Γ_9, and Γ_{13} almost have the same performances, with Γ_9 appearing to be having a tad smaller content. Due to the lower achieved coverage probabilities, these four CRs tended to have smaller contents compared to Γ_3 and Γ_{10}, with Γ_3 possessing the largest content since it is the most conservative. We note the impact of the plug-in procedure for Γ_{13}. Since we did not know $l(k; F)$ we plugged-in an estimator for this function which utilized the empirical distribution function and which also used the same sample data, hence the data double-dipping. The impact of this plug-in and double-dipping is to make the procedure more liberal, except under the Cauchy distribution. Note that in contrast, the randomized
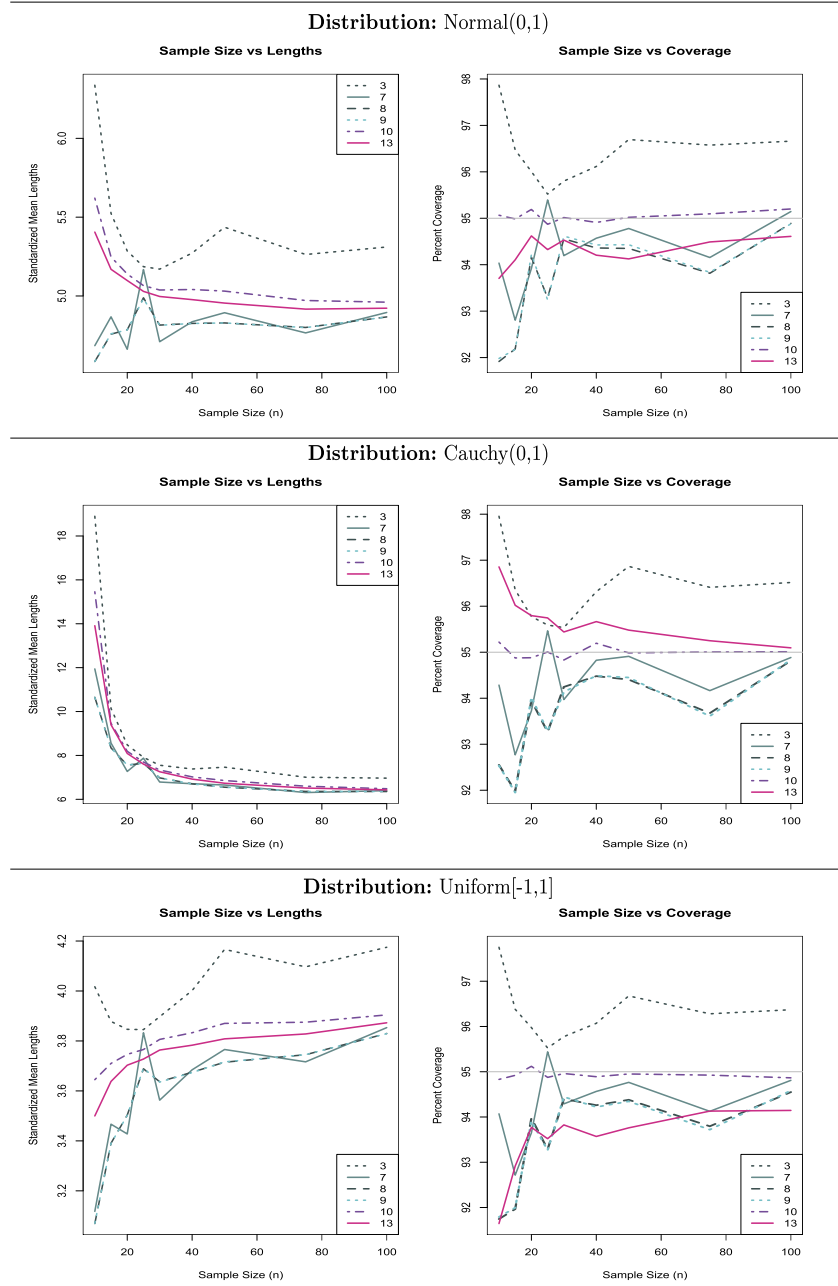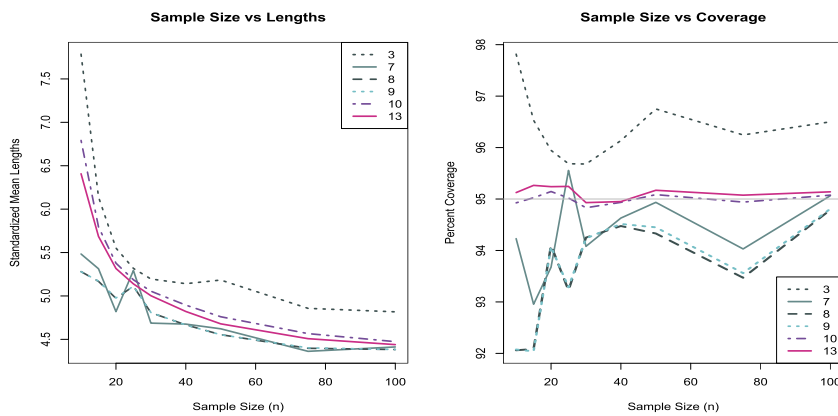
Distribution: Normal(0,1)

Distribution: Cauchy(0,1)

Distribution: Uniform[-1,1]

FIG 5. *Plots of the results of simulation studies with 20000 replications over a set of sample sizes ($n \in \{10, 15, 20, 25, 30, 40, 50, 75, 100\}$) for normal, Cauchy, uniform, Laplace, logistic, normal mixture, SAS, gamma(2,1), and Weibull(.5,1) distributions. The plots pertaining to the lengths (contents) utilized the* standardized mean length, *which is the mean length multiplied by $\sqrt{n}$. These are for the six chosen CRs (Γ_3, Γ_7, Γ_8, Γ_9, Γ_{10}, and Γ_{13}) out of the thirteen CRs. Refer to Table 2 for their specific description.*

Distribution: Laplace(0,1)

Sample Size vs Lengths

Sample Size vs Coverage

Distribution: Logistic(0,1)

Sample Size vs Lengths

Sample Size vs Coverage
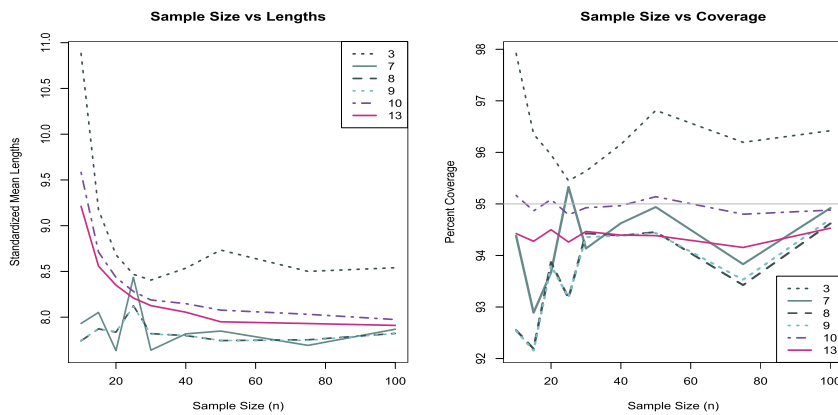
Distribution: .6*Normal(-5,3) + .4*Normal(5,2)

Sample Size vs Lengths

Sample Size vs Coverage

FIG 6. *Simulation results ... continued1*

Distribution: SAS($\mu = 0, \sigma = 1, \delta = 1, \epsilon = 1$)

Distribution: Gamma(2,1)
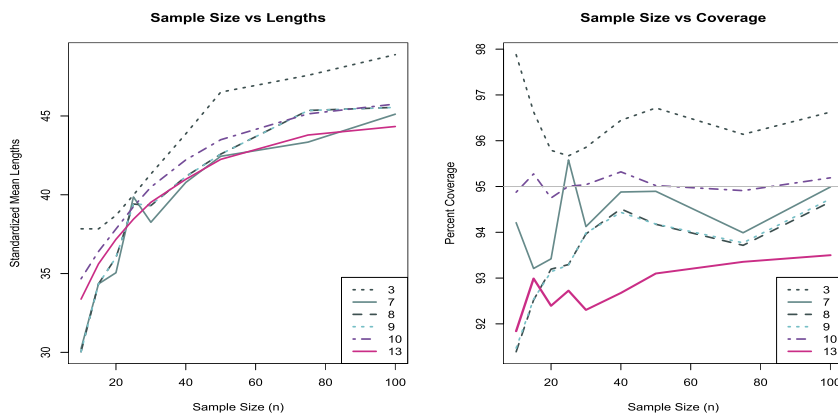
Distribution: Weibull(.5,1)
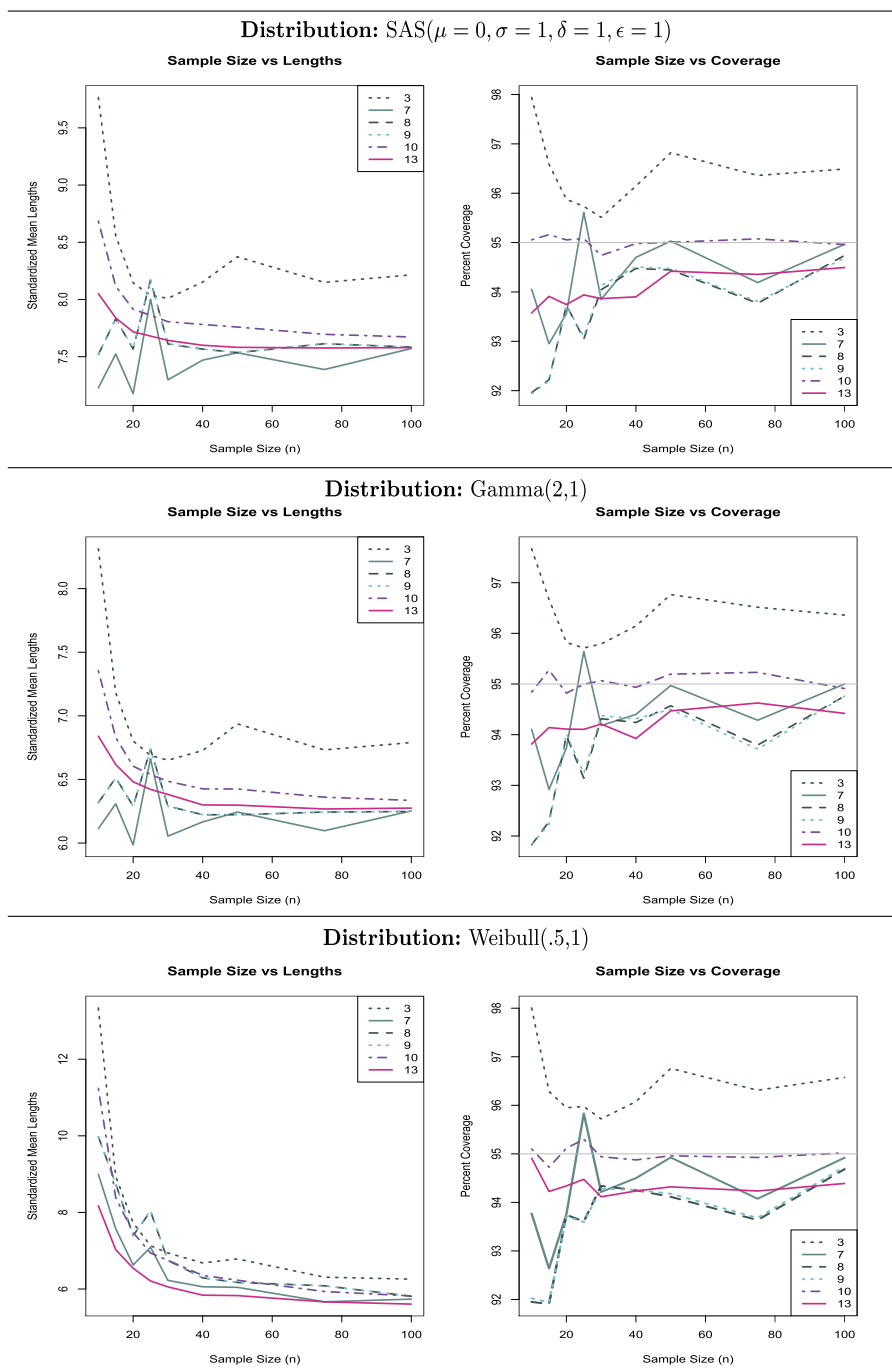
FIG 7. Simulation results ... continued2

CR Γ_{10} did not need this plug-in procedure and we observe that its achieved coverage probability is very close to the nominal confidence level for all the distributions considered.

9. Some concluding remarks

This paper revisits the classical problem of constructing CRs for the median under the NMEM. This problem is also relevant in the study of complex engineering systems where it is difficult to determine the exact functional form of the system's lifetime distribution, hence one is forced to utilize a nonparametric model for this distribution. In addition, this problem could also arise in economic settings where interest could be on the median of an economic indicator, such as household income which has a highly right-skewed distribution. Several existing nonparametric CRs for the median were reviewed. Included in these 'off-the-shelf' methods are the T-statistic based CR and the Wilcoxon signed-rank statistic based CR. These two methods are arguably the default methods by those trying to infer about the 'center' of a population distribution. Also included are the computationally-intensive bootstrap-based CRs such as the percentile, BC, and BCa methods (see [6]).

The adequacy of a CR is usually based on its expected content, which in one-dimensional settings is its Lebesgue measure, aside from the requirement that it satisfies a specified nominal confidence level. In our development of the CR for the median, we therefore aspired to have small mean content aside from fulfilling the confidence level requirement. Under the NMEM, invariance under location-shift and monotone transformation were invoked to reduce the problem to simply finding good equivariant CRs. Within the class of equivariant CRs, we obtained the best CRs by minimizing the expected content for subclasses of the class of all continuous distribution functions. Thus, there is a best CR when focused on the subclass of symmetric distributions (Γ_{10}) and a best CR for the subclass of exponential distributions (Γ_{11}). We also developed fully nonparametric data-adaptive CRs under the NMEM. Our development of these new CRs relied on a 'bottom-to-top' approach since we just start by imposing reasonable equivariance properties on these CRs and then optimizing their mean contents. This approach is in contrast with the conventional one of first constructing pivotal quantities from test statistics and/or estimators of the median and then 'pivoting' to construct the CRs.

Based on the simulation studies, we found that both the T-statistic based CR, Γ_1, and the Wilcoxon signed-rank statistic based CR, Γ_2, should **not** be employed *under the NMEM*. The sign-statistic based CR, Γ_3, and the optimal procedure focused towards symmetric distributions, Γ_{10}, both fulfill the confidence level requirement, with Γ_3 being somewhat conservative, hence tending to have higher content. Among the other CR methods, the bootstrap procedures Γ_7 (PCT), Γ_8 (BC), and Γ_9 (BCa), and the adaptive procedure Γ_{13}, were the most competitive among all scenarios, but these four also tended to be a tad more liberal than either Γ_3 and Γ_{10}, hence possessed smaller contents. If one

is to insist that the desired nominal confidence level should be achieved, then Γ_{10} appears to be the best, or if one is opposed to using randomized CRs, then Γ_3 should be preferred. However, it is our opinion that abhorrence to the use of randomized procedures is not mathematically justifiable nor defensible. The use of randomized procedures allows for better methods and such randomized strategies are in fact the bulwark of statistical decision theory and also game theory. See also the discussion in section 3 of [8] on the use and implementation of randomized p-values and test functions. On the other hand, if one could tolerate a small degradation in the achieved coverage probability, then Γ_7–Γ_9 and Γ_{13} appear to be reasonable choices among these different CR procedures.

Finally, we close by noting the power of invariance in the development of statistical procedures, specifically the construction of CRs for the median in a nonparametric setting which was done in this paper. Invariance arguments are also applicable in developing CRs for the median when the family of distributions is restricted to those that are symmetric. It also holds promise when dealing with more complex data structures, such as the presence of censored data which usually arise in biomedical situations. But, of particular interest to us and a major motivation for this paper, is the use of invariance to enable the construction of optimal simultaneous confidence regions when there are multiple parameters of interest. This is an ongoing research project that we are currently exploring.

Acknowledgements

Funding

References

[1] Richard E. Barlow and Frank Proschan. *Statistical theory of reliability and life testing*. Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London, 1975. Probability models, International Series in Decision Processes, Series in Quantitative Methods for Decision Making. MR0438625

[2] Robert Bartels. The rank version of von Neumann's Ratio Test for Randomness. *Journal of the American Statistical Association*, 77(377):40–46, 1982.

[3] George Casella and Roger L. Berger. *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990. MR1051420

[4] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997. With 1 IBM-PC floppy disk (3.5 inch; HD). MR1478673

[5] Thomas J. DiCiccio and Bradley Efron. Bootstrap confidence intervals. *Statist. Sci.*, 11(3):189–228, 1996. With comments and a rejoinder by the authors. MR1436647

[6] Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, New York, 2016. Algorithms, evidence, and data science. MR3523956

[7] F. Galton. The most suitable proportion between the value of first and second prizes. *Biometrika*, 1:385–395, 1902.

[8] Joshua D. Habiger and Edsel A. Peña. Randomised P-values and nonparametric procedures in multiple testing. *J. Nonparametr. Stat.*, 23(3):583–604, 2011. MR2836278

[9] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric statistical methods*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2014. MR3221959

[10] Peter M. Hooper. Invariant confidence sets with smallest expected measure. *Ann. Statist.*, 10(4):1283–1294, 1982. MR0673664

[11] Peter M. Hooper. Sufficiency and invariance in confidence set estimation. *Ann. Statist.*, 10(2):549–555, 1982. MR0653529

[12] Peter M. Hooper. Invariant prediction regions with smallest expected measure. *J. Multivariate Anal.*, 18(1):117–126, 1986. MR0827172

[13] M. C. Jones and Arthur Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780, 2009. MR2564489

[14] E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005. MR2135927

[15] C. C. Malesios and S. Psarakis. Comparison of the h-index for different fields of research using bootstrap methodology. *Qual. Quant.*, 48:521–545, 2014.

[16] Edsel A. Peña and Elizabeth H. Slate. Global validation of linear model assumptions. *J. Amer. Statist. Assoc.*, 101(473):341–354, 2006. MR2268051

[17] F. Proschan. Theoretical explanation of observing decreasing failure rate. *Technometrics*, 5:375–383, 1963.

[18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[19] Ronald H. Randles and Douglas A. Wolfe. *Introduction to the theory of non-*

parametric statistics. John Wiley & Sons, New York-Chichester-Brisbane, 1979. Wiley Series in Probability and Mathematical Statistics. MR0547836

[20] William R. Thompson. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. Math. Statist.*, 7(3):122–128, 09 1936.