Pitman Closest Estimators Based on Convex Linear Combinations of Two Contiguous Order Statistics

N. Balakrishnan, Katherine F. Davies, Jerome P. Keating and Robert L. Mason

Abstract Comparisons of best linear unbiased estimators with some other prominent estimators have been carried out over the last six decades since the ground breaking work of Lloyd [13]; see Arnold et al. [1] and David and Nagaraja [9] for elaborate details in this regard. Recently, Pitman closeness comparison of order statistics as estimators for population parameters, such as medians and quantiles, and their applications have been carried out by Balakrishnan et al. [3–5, 7]. In this paper, we discuss the Pitman closest estimators based on convex linear combinations of two contiguous order statistics, which sheds additional insight with regard to the estimation of the population median in the case of even sample sizes. We finally demonstrate the proposed method for the uniform, exponential, power function and Pareto distributions.

Keywords Order statistics · Pitman closeness · Probabilities of closeness · Convex linear estimator · Location-scale family

N. Balakrishnan (✉)
Department of Mathematics and Statistics, McMaster University,
Hamilton, ON L8S 4K1, Canada
e-mail: bala@mcmaster.ca

K.F. Davies
Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
e-mail: Katherine.Davies@UManitoba.CA

J.P. Keating
Department of Management Science and Statistics, University of Texas at San Antonio,
78249-0704 San Antonio, Texas, USA
e-mail: jerome.keating@utsa.edu

R.L. Mason
Southwest Research Institute, 78228-0510 San Antonio, Texas, USA
e-mail: robert.mason@swri.org

P.K. Choudhary et al. (eds.), *Ordered Data Analysis, Modeling and Health Research Methods*, Springer Proceedings in Mathematics & Statistics 149,
DOI 10.1007/978-3-319-25433-3_2

1 Introduction

The comparison of estimators under the Pitman closeness criterion has a long history since it was introduced by Pitman [17] and further discussed by Rao [18]. For estimation based on order statistics, Nagaraja [15] considered Pitman closeness of estimators and predictors for the two-parameter exponential distribution. In a similar light, Balakrishnan et al. [6] and Balakrishnan and Davies [2] considered Pitman comparison of estimators for the one-parameter exponential distribution based on Type-I and II censored samples, respectively. Recently, Balakrishnan et al. [3] carried out Pitman closeness comparisons between pairs of order statistics arising from a random sample of size n with regard to the estimation of population quantiles ξ_p. Specifically, with $X_1, \ldots, X_n$ denoting a random sample taken from a continuous population with probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$, and $X_{1:n}, \ldots, X_{n:n}$ denoting the corresponding order statistics, Balakrishnan et al. [3] derived formulas for the comparison of any two contiguous order statistics as estimators of population quantiles.

It is well known that (see David and Nagaraja [9] and Arnold et al. [1])

$$F\left(X_{i:n}\right) = U_{i:n} \sim \mathscr{B}\left(i, n-i+1\right), \tag{1}$$

where $\mathscr{B}(\alpha, \beta)$ denotes a beta random variable with shape parameters α and β; here, $U_{i:n}$ denotes the ith order statistic from a sample of size n from the Uniform(0, 1) distribution. Mean ranks in quantile-quantile plots are based on the relation

$$\boldsymbol{E}\left[F\left(X_{i:n}\right)\right] = \frac{i}{n+1} = e_{i:n}.$$

Similarly, if $m_{i:n}$ denotes the median of the beta random variable in (1), it is referred to as the median rank.

Definition 1 An estimator $\hat{\theta}$ will be said to overestimate a parameter θ if

$$\Pr\left(\hat{\theta} > \theta\right) > \frac{1}{2}.$$

This definition of *overestimation* is in the sense that the estimator $\hat{\theta}$ more frequently overestimates θ than it underestimates θ, or equivalently, the median of the distribution of $\hat{\theta}$ is less than θ.

In this paper, we discuss the Pitman closest estimation based on a convex linear combination of two contiguous order statistics. We then demonstrate the established results with uniform, exponential, power function and Pareto distributions.

2 Narrowing Down the Choices Among Order Statistics

In some cases, one may want to improve on the choice of order statistics given by Balakrishnan et al. [3], which provides the probability that a given order statistic is Pitman-closer to a specific population quantile ξ_p than any other order statistic from the same sample. The natural question that arises in this regard is whether one can improve on the estimation of ξ_p by using a linear combination of two contiguous order statistics. In some cases, no improvement can be made (e.g., the sample median in odd sample sizes as an estimator of the population median of a symmetric distribution is the Pitman-closest linear equivariant estimator of $\xi_{0.50}$), as shown in Balakrishnan et al. [4]. If we restrict our attention to convex linear combinations of two order statistics, then we can reduce the number of pairs to be considered to produce a Pitman-closer estimator and the following two lemmas facilitate this. For this specific purpose, we therefore want to bracket ξ_p so that we find the largest order statistic that underestimates ξ_p and the smallest order statistic that overestimates ξ_p, in the sense of Definition 1.

Lemma 1 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. For $p \geq 1 - 2^{-1/n}$, let $m_{j:n}$ be the largest median rank less than p. Then, the largest order statistic that does not overestimate ξ_p is $X_{j:n}$.*

Proof If $p < 1 - 2^{-1/n} = m_{1:n}$, then

$$\Pr\left(X_{1:n} < \xi_p\right) = \Pr\left[F\left(X_{1:n}\right) < F\left(\xi_p\right)\right] = \Pr\left(U_{1:n} < p\right) < \Pr\left(U_{1:n} < m_{1:n}\right) = \frac{1}{2}.$$

Thus all order statistics overestimate ξ_p whenever $p < 1 - 2^{-1/n}$. Next, let us consider the case when $p \in [m_{1:n}, 1)$. Since

$$m_{1:n} < m_{2:n} < \cdots < m_{n:n} < m_{n+1:n} = 1,$$

the spacings between the median ranks form a partition of the interval which immediately implies that there exists a j such that

$$m_{j:n} \leq p < m_{j+1:n}.$$

Thus, $U_{j:n}$ is the largest order statistic that underestimates p and consequently $X_{j:n}$ is the largest order statistic that underestimates ξ_p.

Note that j in the previous lemma does not depend on the underlying continuous distribution function $F(x)$, but only on the medians of order statistics from the Uniform(0,1) distribution, which can be determined numerically.

Lemma 2 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. For $p \leq 2^{-1/n}$, let $m_{\ell:n}$ be the largest median rank less than p. Then, the smallest order statistic that overestimates ξ_p is $X_{\ell:n}$.*

Notice that all order statistics underestimate ξ_p whenever $p > 2^{-1/n} = m_{n:n}$. The proof of this lemma proceeds in a similar way to that of Lemma 1. Determination of $X_{j:n}$ or $X_{j+1:n}$ does not require knowledge of the underlying distribution $F(x)$, but only needs the solution for j as a function of n and p through the medians of the beta distributions. We can combine Lemmas 1 and 2 to form the following theorem.

Theorem 1 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population with pdf $f(x)$ and cdf $F(x)$, and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. Then, there exists a largest order statistic $X_{j:n}$ that does not overestimate ξ_p and a smallest order statistic $X_{j+1:n}$ that overestimates ξ_p (in the sense of Definition 1) for $m_{1:n} \le p < m_{n:n}$.*

3 Pitman Closeness Criterion

We now introduce the comparison criterion known as Pitman closeness or Pitman nearness.

Definition 2 Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be univariate estimators of a real-valued parameter θ based on a sample of size n. Then, Pitman Closeness (PC) is defined as

$$\boldsymbol{P}\left(\hat{\theta}_1, \hat{\theta}_2 | \theta, n\right) = \Pr\left(|\hat{\theta}_1 - \theta| < |\hat{\theta}_2 - \theta|\right).$$

Interested readers may refer to the monograph by Keating et al. [12] for pertinent details. The measure in Definition 2 quantifies the frequency with which one estimator is closer to the value of the parameter θ than a competing estimator; see, for example, [6, 10, 14, 16–18].

Definition 3 Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be univariate estimators of a real-valued parameter θ based on a sample of size n. Then, $\hat{\theta}_1$ is said to be Pitman-closer to θ, for a given value of θ, than $\hat{\theta}_2$ provided

$$\boldsymbol{P}\left(\hat{\theta}_1, \hat{\theta}_2 | \theta, n\right) \ge \boldsymbol{P}\left(\hat{\theta}_2, \hat{\theta}_1 | \theta, n\right).$$

Definition 4 The estimator $\hat{\theta}_1$ is said to be uniformly Pitman-closer than $\hat{\theta}_2$ if $\boldsymbol{P}(\hat{\theta}_1, \hat{\theta}_2 | \theta, n) \ge \boldsymbol{P}(\hat{\theta}_2, \hat{\theta}_1 | \theta, n)$ for all θ in the parameter space Θ, with strict inequality holding for at least one $\theta \in \Theta$. The estimator $\hat{\theta}_1$ is uniformly Pitman-closest among the estimators in a class $\mathscr{C}$ provided

$$\boldsymbol{P}(\hat{\theta}_1, \hat{\theta}_j | \theta, n) \ge \boldsymbol{P}(\hat{\theta}_j, \hat{\theta}_1 | \theta, n)$$

for all $\hat{\theta}_j$ in $\mathscr{C}$ and for all $\theta \in \Theta$, with strict inequality holding for at least one $\theta \in \Theta$.

Lemma 3 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population with pdf $f(x)$ and cdf $F(x)$, and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics.*

If $m_{1:n} < p$, then if j is such that $X_{j:n}$ is the largest order statistic that does not overestimate ξ_p, $X_{j:n}$ is Pitman-closer to ξ_p than any of $X_{1:n}, \ldots, X_{j-1:n}$.

Proof We have

$$\begin{aligned}
\boldsymbol{P}\left(X_{j:n}, X_{\ell:n} | \xi_p\right) &= \Pr\left[|X_{j:n} - \xi_p| < |X_{\ell:n} - \xi_p|\right] \\
&= \Pr\left[\left(X_{j:n} - \xi_p\right)^2 < \left(X_{\ell:n} - \xi_p\right)^2\right] \\
&= \Pr\left[X_{j:n}^2 - X_{\ell:n}^2 < 2\xi_p\left(X_{j:n} - X_{\ell:n}\right)\right] \\
&= \Pr\left[\left(X_{j:n} - X_{\ell:n}\right)\left(X_{j:n} + X_{\ell:n}\right) < 2\xi_p\left(X_{j:n} - X_{\ell:n}\right)\right] \\
&= \Pr\left[X_{j:n} + X_{\ell:n} < 2\xi_p\right] \\
&< \Pr\left[X_{\ell:n} < \xi_p\right] < 1/2.
\end{aligned}$$

Thus, it follows that $X_{j:n}$ is Pitman-closer to ξ_p than $X_{\ell:n}$ for all $\ell = 1, \ldots, j-1$.

In an analogous manner, we can establish the following lemma.

Lemma 4 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population with pdf $f(x)$ and cdf $F(x)$, and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. If $p < m_{n:n}$, then if j is such that $X_{j:n}$ is the largest order statistic that does not overestimate ξ_p, $X_{j+1:n}$ is Pitman-closer to ξ_p than any of $X_{j+2:n}, \ldots, X_{n:n}$.*

Now, let $p \in (m_{1:n}, m_{n:n})$. Then, due to Lemmas 3 and 4, it is evident that there exists a largest integer j such that $\Pr\left(X_{j:n} < \xi_p\right) \leq 1/2$ and $\Pr\left(X_{j+1:n} < \xi_p\right) > 1/2$, which is formally stated in the following theorem.

Theorem 2 *Let $X_1, \ldots, X_n$ be a random sample from a continuous population with pdf $f(x)$ and cdf $F(x)$, and $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. Then, there exists a largest order statistic $X_{j:n}$ such that $X_{j:n}$ is Pitman-closer to ξ_p than $X_{\ell:n}$ for $\ell = 1, \ldots, j-1$, and $X_{j+1:n}$ is Pitman-closer to ξ_p than $X_{\ell:n}$ for $\ell = j+2, \ldots, n$, when $m_{1:n} \leq p < m_{n:n}$.*

Consequently, in terms of comparisons of individual order statistics, the Pitman-closest one to ξ_p, for a given p, will depend on the comparison of $X_{j:n}$ and $X_{j+1:n}$. The better of these two in the sense of Pitman closeness will depend on the underlying distribution $F(x)$. For this reason, it will be reasonable to compare the largest order statistic that underestimates ξ_p with the smallest order statistic that overestimates ξ_p.

In fact, one can generalize the use of contiguous order statistics, $X_{j:n}$ and $X_{j+1:n}$, to any pair $X_{i:n}$ and $X_{k:n}$, where $1 \leq i \leq j$ and $j+1 \leq k \leq n$. These results imply that if we are to find a Pitman-closer estimator than any individual order statistic from a convex class based on two order statistics, then one order statistic must underestimate ξ_p and the other must overestimate ξ_p. Of course, a single order statistic may outperform any convex linear combination of all other order statistics.

4 Use of a Convex Class

Based on Theorem 2, we may consider some linear combination of these contiguous order statistics. The use of a convex linear combination, i.e.,

$$\hat{\xi}_p = wX_{j:n} + (1-w)X_{j+1:n}, w \in [0, 1],$$

produces a class of ordered estimators in the closed bounded interval $\left[X_{j:n}, X_{j+1:n}\right]$. Furthermore, in location-scale families, the individual order statistics are location invariant estimators of the location parameter and so convex linear combinations of order statistics are location invariant estimators as well. So, we wish to find a median unbiased estimator of ξ_p within the convex class given above. While the value of w, for which the convex linear combination has a median of ξ_p, may not be independent of the unknown parameters in the distributions of $X_{j:n}$ and $X_{j+1;n}$, certain special and important cases do exist in which the choice only depends on n and p. However, it should be kept in mind there is no certainty that this median unbiased convex linear combination of $X_{j:n}$ and $X_{j+1:n}$ will be the Pitman-closest median unbiased convex linear combination of any pair of order statistics $X_{i:n}$ and $X_{k:n}$, where $1 \le i \le j$ and $j+1 \le k \le n$.

In order to assess the median unbiased estimator within the class of convex linear combinations of two contiguous order statistics, we need the joint density of $X_{j:n}$ and $X_{j+1:n}$ given by (see Arnold et al. [1] and David and Nagaraja [9])

$$f(u, v) = \frac{n!}{(j-1)!(n-j-1)!}[F(u)]^{j-1}[1-F(v)]^{n-j-1} f(u) f(v), \text{ if } u < v, \tag{2}$$

for $j = 1, \ldots, n-1$. Since we have reduced our consideration to just contiguous order statistics mentioned in the sense of overestimating and underestimating, we can now consider a new class of estimators based on them.

Definition 5 Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from $F(x)$, which is strictly monotone on the support of X. Let ξ_p be the pth quantile of the distribution. Let j be the largest integer for which $m_{j:n} \le p$. Define the class $\mathscr{Q}$ as the collection of all convex linear combinations of $X_{j:n}$ and $X_{j+1:n}$, i.e.,

$$\mathscr{Q} = \left\{\hat{\xi}_p(w) | \hat{\xi}_p(w) = wX_{j:n} + (1-w)X_{j+1:n},\ w \in [0, 1]\right\}. \tag{3}$$

In general, determining the Pitman-closest estimator in the class $\mathscr{Q}$ can be difficult, and also can produce a random variable that depends on the unknown parameters of the distribution and consequently not an estimator. But, the determination of a best choice within the class is guaranteed for a location-scale family as shown below.

4.1 Location-Scale Family

Let us consider the location-scale family of distributions with the density function of X given by

$$f(x \mid \mu, \sigma) = (1/\sigma)g\left[(x-\mu)/\sigma\right], \tag{4}$$

where $g(z)$ is a continuous parameter-free density. The parameter space for these families is the upper half-plane $\Omega = \{-\infty < \mu < \infty, \sigma > 0\}$.

The cdf of X is

$$F(x \mid \mu, \sigma) = G\left[(x-\mu)/\sigma\right], \text{ where } G(t) = \int_{-\infty}^{t} g(u)du.$$

The $100p$ percentage point (or percentile) of the random variable X, denoted by ξ_p, is defined as $\xi_p = \inf\{x \in \mathbf{R} : F(x) \geq p\}$. The distribution function $G(t)$ is usually taken to be a parameter-free cdf. If the essential range, $\mathbf{R}$, of X is an open connected subset of $\mathbf{R}$, then ξ_p is unique for each $p \in (0, 1)$ with

$$\xi_p = \mu + G^{-1}(p)\sigma, \tag{5}$$

where $G^{-1}(\cdot)$ is the inverse function of $G(\cdot)$. One can see from (5) that within this family, percentiles are linear combinations of the parameters μ and σ. This family includes many well-known distributions such as normal, extreme-value, exponential, Laplace, Cauchy, uniform and logistic as members, but also includes several other distributions such as lognormal, log-uniform, inverse Gaussian, Pareto and Weibull through suitable transformations.

Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample of size n from a location-scale parameter density $f(x)$ in (4). The estimation of location and scale parameters as well as percentiles have been discussed quite extensively based on order statistics; see, for example, Balakrishnan and Cohen [8]. First define $Z_{1:n}, \ldots, Z_{n:n}$ as

$$Z_{i:n} = \frac{X_{i:n} - \mu}{\sigma}, \text{ for } i = 1, \ldots, n. \tag{6}$$

Theorem 3 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from a continuous location-scale parameter density $f(x)$ in (4). Let $Z_{1:n}, \ldots, Z_{n:n}$ be the corresponding order statistics as defined in (6). Then, we have*

$$\boldsymbol{P}\left(X_{j:n}, X_{\ell:n} | \xi_p\right) = \boldsymbol{P}\left(Z_{j:n}, Z_{\ell:n} | G^{-1}(p)\right).$$

Thus, within the class of location-scale families of distributions, the Pitman closeness of any two order statistics in the estimation of ξ_p is independent of the unknown parameters.

Theorem 4 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from $F(x)$, which is strictly monotone on the support of X. Let $p \in (m_{1:n}, m_{n:n})$. With*

$w \in [0, 1]$, *let us consider the class* $\mathcal{Q}$ *in (3) of estimators* ξ_p. *Then,* $\Pr\left(\hat{\xi}_p(w) < \xi_p\right)$ *is a continuous non-increasing function of* w.

Proof For $w \in [0, 1]$, let us define

$$Q_{n,p}(w) = \Pr\left(\hat{\xi}_p(w) < \xi_p\right).$$

Then, by Definition 5,

$$Q_{n,p}(1) = \Pr\left(X_{j:n} < \xi_p\right) > \frac{1}{2}, \quad Q_{n,p}(0) = \Pr\left(X_{j+1:n} < \xi_p\right) < \frac{1}{2}.$$

Since $F(x)$ is continuous, $Q_{n,p}(w)$ is continuous and so there exists a value $0 \leq w_0 \leq 1$ such that $Q_{n,p}(w_0) = 1/2$. For $0 \leq w_1 < w_2 \leq 1$, we have

$$\begin{gathered} w_1 X_{j:n} + (1 - w_1) X_{j+1:n} < w_2 X_{j:n} + (1 - w_2) X_{j+1:n}, \\ \Pr\left\{w_1 X_{j:n} + (1 - w_1) X_{j+1:n} < x\right\} > \Pr\left\{w_2 X_{j:n} + (1 - w_2) X_{j+1:n} < x\right\}, \\ Q_{n,p}(w_1) > Q_{n,p}(w_2). \end{gathered}$$

Therefore, $Q_{n,p}(w)$ is a continuous non-increasing function of w.

Corollary 1 *Under the conditions of Theorem 4, a median unbiased estimator of* ξ_p *exists within the considered convex class. If* $F(x)$ *is strictly increasing over its support, the median unbiased estimator is unique within this class.*

Proof Since $Q_{n,p}(1) > \frac{1}{2}$ and $Q_{n,p}(0) \leq \frac{1}{2}$, by the continuity of $Q_{n,p}(w)$, there exists a value $0 \leq w_0 \leq 1$ such that $Q_{n,p}(w_0) = 1/2$. Further, if $F(x)$ is strictly increasing over its support, $Q_{n,p}(w)$ will be strictly decreasing on [0,1) and so the solution w_0 will be unique.

Corollary 2 *Let* $X_{1:n}, \ldots, X_{n:n}$ *be the order statistics from a random sample from* $F(x)$*, which is strictly monotonically increasing on the support of* X. *Let* $F(x)$ *be a member of the location-scale family of distributions, and* $Z_{1:n}, \ldots, Z_{n:n}$ *be as defined in (6). Let* $p \in (m_{1:n}, m_{n:n})$. *With* $w \in [0, 1)$, *let us consider the class* $\mathcal{Q}$ *in (3) for the estimation of* ξ_p. *Then, there exists a unique Pitman-closest estimator of* ξ_p *within* $\mathcal{Q}$.

Proof The proof follows directly from Corollary 1. Within an ordered class of estimators of some parameter, say θ, the median unbiased estimator within the class will be the Pitman-closest estimator of θ. We are guaranteed that the class $\mathcal{Q}$, by its very construction, produces some estimators that overestimate ξ_p and some that underestimate ξ_p such that

$$\begin{aligned} Q_{n,p}(w) &= \Pr\left(\hat{\xi}_p(w) < \xi_p\right) \\ &= \Pr\left(w X_{j:n} + (1 - w) X_{j+1:n} < \xi_p\right) \\ &= \Pr\left(w Z_{j:n} + (1 - w) Z_{j+1:n} < G^{-1}(p)\right). \end{aligned}$$

The value of w such that $Q_{n,p}(w) = 1/2$ is unique does not involve the unknown parameters, and is only a function of n, p and the corresponding value of j. This choice of w produces an estimator that is median unbiased and is therefore Pitman-closer than all other estimators within $\mathscr{Q}$ since the class is completely ordered, and is therefore the Pitman-closest estimator of ξ_p in $\mathscr{Q}$. While this result guarantees the existence and uniqueness of a median unbiased estimator in $\mathscr{Q}$, it would be good to have a Rao-Blackwell type result that would provide a method for its construction.

4.2 Transformation

In this case, consider the following transformation $R = X_{j:n}$ and $T = X_{j+1:n} - X_{j:n}$, for $1 \leq j \leq n-1$. It follows that

- R is a location invariant statistic and as noted before $Z_{j:n} = (X_{j:n} - \mu)/\sigma$ has a parameter-free distribution;
- $T = X_{j+1:n} - X_{j:n}$ is a scale invariant statistic and T/σ is a pivotal quantity for σ;
- $Z_{j:n}/T$ is a pivotal quantity for μ;
- $\left(R - \xi_p\right)/T$ is a pivotal quantity for ξ_p with a distribution that depends on $G^{-1}(p)$ and the sample size n.

Under this transformation, one can rewrite any convex class for which $X_{j:n}$ and $X_{j+1:n}$, respectively, underestimate and overestimate ξ_p, in the following way:

$$\begin{aligned}\mathscr{Q} &= \left\{\hat{\xi}_p(w) | \hat{\xi}_p(w) = X_{j+1:n} - w\left(X_{j+1:n} - X_{j:n}\right),\ w \in (0,1]\right\} \\ &= \left\{\hat{\xi}_p(c) | \hat{\xi}_p(c) = X_{j:n} + c\left(X_{j+1:n} - X_{j:n}\right),\ c = 1 - w \in (0,1]\right\}.\end{aligned}$$

One can now derive the median unbiased estimator within $\mathscr{Q}$ according to the following theorem.

Theorem 5 *In the context of Lemma 3, consider two order statistics $X_{j:n}$ and $X_{j+1:n}$ in a location-scale family where $X_{j:n}$ underestimates ξ_p and $X_{j+1:n}$ overestimates ξ_p (in the sense of Definition 1). Then, a median unbiased estimator within $\mathscr{Q}$ is given by*

$$\hat{x}_p = X_{j:n} + \boldsymbol{M}_{(0,1)}\left(\frac{G^{-1}(p) - R}{T}\right)T, \tag{7}$$

where $T = X_{j+1:n} - X_{j:n}$ and $\boldsymbol{M}_{(0,1)}(U)$ denotes the median of the random variable U when $\mu = 0$ and $\sigma = 1$.

Proof We have

$$
\begin{aligned}
\Pr\left\{\hat{\xi}_p(c) < \xi_p\right\} &= \Pr\left\{X_{j:n} + c\left(X_{j+1:n} - X_{j:n}\right) < \xi_p\right\} \\
&= \Pr\left\{c < \tfrac{\xi_p - X_{j:n}}{X_{j+1:n} - X_{j:n}}\right\} \\
&= \Pr\left\{c < \tfrac{G^{-1}(p) - Z_{j:n}}{Z_{j+1:n} - Z_{j:n}}\right\}.
\end{aligned}
$$

If the estimator is median unbiased, then the probability content of the interval is 1/2, and so

$$
c = \boldsymbol{M}_{(0,1)}\left(\frac{G^{-1}(p) - R}{T}\right).
$$

It follows that the median unbiased estimator within $\mathscr{Q}$ is the estimator in (7). Solving for c will require numerical methods and the fact that c must be in the interval (0, 1] would facilitate the use of the secant method. Incidentally, a naive and nonparametric estimate of c can be obtained as

$$
\hat{c} = \frac{p - m_{j:n}}{m_{j+1:n} - m_{j:n}}.
$$

4.3 Examples

Uniform distribution

In the case of the Uniform(0,1) distribution, consider $\Pr\left[R + cT < \xi_p\right]$, i.e., $\Pr\left[R + cT < p\right]$, which we need to set to 1/2. First, we note that we can express

$$
R + cT = X_{j:n} + c(X_{j+1:n} - X_{j:n}) = X_{j+1:n}\left(\frac{X_{j:n}}{X_{j+1:n}} + c\left(1 - \frac{X_{j:n}}{X_{j+1:n}}\right)\right) = VW,
$$

where $U = \frac{X_{j:n}}{X_{j+1:n}}$, $V = X_{j+1:n}$ and $W = U + c(1 - U)$. It is known that $U \sim Beta(j, 1)$ and $V \sim Beta(j+1, n-j)$ and that the two random variables are independent; see Arnold et al. [1]. Using the distribution of U, it can be shown that the pdf of W is given by

$$
f_W(w) = \frac{j}{(1-c)^j}(w - c)^{j-1} \text{ if } w \in (c, 1).
$$

We then have

$$
\Pr\left[R + cT < p\right] = \Pr(VW < p) = \Pr\left(V < \frac{p}{W}\right) = \int_c^1 \Pr\left(V < \frac{p}{w}\right) f_W(w)dw,
$$

where

$$
\Pr\left(V < \frac{p}{w}\right) = \begin{cases} I_{\frac{p}{w}}(j+1, n-j) & \text{if } \frac{p}{w} < 1 \\ 1 & \text{if } \frac{p}{w} \geq 1, \end{cases}
$$

and $I_q(a,b)$ is the incomplete beta ratio defined by $I_q(a,b) = \frac{1}{B(a,b)}\int_0^q t^{a-1}(1-t)^{b-1}dt$ and $B(a,b)$ is the complete beta function defined by $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Thus, we get

$$\Pr(VW < p) = \begin{cases} \int_c^1 I_{\frac{p}{w}}(j+1, n-j) f_W(w)dw & \text{if } \frac{p}{c} < 1 \\ \int_p^1 I_{\frac{p}{w}}(j+1, n-j) f_W(w)dw + \left(\frac{p-c}{1-c}\right)^j & \text{if } \frac{p}{c} \geq 1. \end{cases} \tag{8}$$

Equation (8) can be solved for various p to find c when p is away from the bounds 0 and 1. For p close to 0, c is found such that $\Pr(cX_{1:n} \leq \xi_p) = 1/2$, and similarly, for p close to 1, c is found such that $\Pr(cX_{n:n} \leq \xi_p) = 1/2$. However, in order to use $cX_{n:n}$, we need to need to check the validity of the determined c since it is possible that the estimator may exist outside the support. Since we choose c such that $\Pr(X_{n:n} \leq \frac{\xi_p}{c})$ $= 1/2$, i.e., $\Pr(X_{n:n} \leq \frac{p}{c}) = 1/2$, the desired c turns out to be $c = 2^{1/n}p$. Now, let $W = 2^{1/n}pX_{n:n}$. In this case, we find

$$\Pr(W \leq 1) = \Pr(2^{1/n}pX_{n:n} \leq 1) = \Pr\left(X_{n:n} \leq \frac{1}{2^{1/n}p}\right) = \left(\frac{1}{2}\right)\left(\frac{1}{p}\right)^n.$$

This is a valid probability if and only if $\left(\frac{1}{p}\right)^n \leq 2$, i.e., $-\log(p) \leq \frac{1}{n}\log(2)$. So, the corresponding entries in Table 1 have been checked accordingly.

Exponential distribution

We again consider $\Pr(R + cT < \xi_p)$, which we can rewrite as

$$\Pr(R + cT < \xi_p) = \Pr[X_{j:n} + c(X_{j+1:n} - X_{j:n}) < \xi_p].$$

We then have this probability as

$$\begin{aligned} \Pr(R + cT < \xi_p) &= \Pr\left((n-j)(X_{j+1:n} - X_{j:n}) < (n-j)\left(\frac{\xi_p - X_{j:n}}{c}\right)\right) \\ &= \int_0^\infty \Pr\left(S_{j+1} < (n-j)\left(\frac{\xi_p - x_j}{c}\right)\right) f_{X_{j:n}}(x_j)dx_j \\ &= \int_0^{\xi_p}\left(1 - e^{-\frac{(n-j)(\xi_p - x_j)}{c}}\right)\frac{n!}{(j-1)!(n-j)!}(1-e^{-x_j})^{j-1}(e^{-x_j})^{n-j} \\ &\qquad \times e^{-x_j}dx_j \\ &= I_p(j, n-j+1) - e^{-\frac{(n-j)\xi_p}{c}}\int_0^{\xi_p} e^{\frac{(n-j)x_j}{c}}\frac{1}{B(j,n-j+1)}(1-e^{-x_j})^{j-1} \\ &\qquad \times (e^{-x_j})^{n-j}e^{-x_j}dx_j \\ &= I_p(j, n-j+1) - \frac{e^{-\frac{(n-j)\xi_p}{c}}}{B(j,n-j+1)}\sum_{k=0}^{j-1}(-1)^k\binom{j-1}{k}\int_{1-p}^1 u^{k+n-j-\frac{n-j}{c}}du, \end{aligned}$$

where S_{j+1} is the normalized spacing defined as $S_{j+1} = (n-j)(X_{j+1:n} - X_{j:n}) \sim Exp(1)$, and since it is known to be independent of $X_{j:n}$ (see Arnold et al. [1]), we have

$$\Pr(S_{j+1} < s) = \begin{cases} 0 & \text{if } s \leq 0, \\ 1 - e^{-s} & \text{if } s > 0. \end{cases}$$

Table 1 Values of j and c for the uniform distribution when $n = 10$ for various choices of p

p	j	c	p	j	c	p	j	c
0.01	1	0.1493	0.34	3	0.8336	0.67	7	0.2728
0.02	1	0.2987	0.35	3	0.9430	0.68	7	0.3763
0.03	1	0.4480	0.36	4	0.0528	0.69	7	0.4774
0.04	1	0.5973	0.37	4	0.1583	0.70	7	0.5770
0.05	1	0.7466	0.38	4	0.2612	0.71	7	0.6762
0.06	1	0.8960	0.39	4	0.3622	0.72	7	0.7765
0.07	1	0.0320	0.40	4	0.4621	0.73	7	0.8790
0.08	1	0.1312	0.41	4	0.5622	0.74	7	0.9847
0.09	1	0.2238	0.42	4	0.6635	0.75	8	0.0982
0.10	1	0.3169	0.43	4	0.7668	0.76	8	0.2094
0.11	1	0.4142	0.44	4	0.8726	0.77	8	0.3170
0.12	1	0.5163	0.45	4	0.9813	0.78	8	0.4211
0.13	1	0.6234	0.46	5	0.0899	0.79	8	0.5221
0.14	1	0.7354	0.47	5	0.1954	0.80	8	0.6206
0.15	1	0.8517	0.48	5	0.2986	0.81	8	0.7180
0.16	1	0.9722	0.49	5	0.3998	0.82	8	0.8166
0.17	2	0.0817	0.50	5	0.0000	0.83	8	0.9183
0.18	2	0.1834	0.51	5	0.6002	0.84	9	0.0278
0.19	2	0.2820	0.52	5	0.7014	0.85	9	0.1483
0.20	2	0.3794	0.53	5	0.8046	0.86	9	0.2646
0.21	2	0.4779	0.54	5	0.9101	0.87	9	0.3766
0.22	2	0.5789	0.55	6	0.0187	0.88	9	0.4837
0.23	2	0.6830	0.56	6	0.1274	0.89	9	0.5858
0.24	2	0.7906	0.57	6	0.2332	0.90	9	0.6831
0.25	2	0.9018	0.58	6	0.3365	0.91	9	0.7762
0.26	3	0.0153	0.59	6	0.4378	0.92	9	0.8688
0.27	3	0.1210	0.60	6	0.5379	0.93	9	0.9680
0.28	3	0.2235	0.61	6	0.6378	0.94	10	1.0075
0.29	3	0.3238	0.62	6	0.7388	0.95	10	1.0182
0.30	3	0.4230	0.63	6	0.8417	0.96	10	1.0289
0.31	3	0.5226	0.64	6	0.9472	0.97	10	1.0396
0.32	3	0.6237	0.65	7	0.0570	0.98	10	1.0503
0.33	3	0.7272	0.66	7	0.1664	0.99	10	1.0611

Furthermore, we have

$$\int_{1-p}^{1} u^{k+n-j-\frac{n-j}{c}}\,du$$

$$= \begin{cases} \frac{1}{k+n-j+1-\frac{n-j}{c}} \left(1-(1-p)^{k+n+j+1-\frac{n-j}{c}}\right) & \text{if } \frac{n-j}{c}-(k+n-j) \neq 1 \\ -log(1-p) & \text{if } \frac{n-j}{c}-(k+n-j) = 1. \end{cases}$$

Values of c and j for various choices of p were numerically determined in this case and are presented in Table 2.

Pareto and power function distributions
Let $X \sim$ Power Function(θ), i.e.,

$$f_X(x) = \theta x^{\theta-1} \text{if } 0 < x < 1 \tag{9}$$

for $\theta > 0$. In this case, by proceeding as in the uniform case, it can be shown that

$$\Pr(VW < \xi_p) = \Pr\left(V < \frac{\xi_p}{W}\right) = \int_w \Pr\left(V < \frac{\xi_p}{w}\right) f_W(w)dw, \tag{10}$$

where $F_V(v) = \Pr(V \leq v) = I_{v^\theta}(j+1, n-j)$ and

$$f_W(w) = \theta j \frac{(w-c)^{\theta j-1}}{(1-c)^{\theta j}} \text{if } c < w < 1. \tag{11}$$

Next, let us consider $X \sim$ Pareto(θ), i.e.,

$$f_X(x) = \nu x^{-\nu-1} \text{if } x \geq 1$$

for $\nu > 0$. Then, the joint density of $X_{j:n}$ and $X_{j+1:n}$ is obtained from (2) as

$$f(x_j, x_{j+1}) = \frac{n!}{(j-1)!(n-j-1)!}(1-x_j^{-\nu})^{j-1}(x_{j+1}^{-\nu})^{n-j-1}\nu x_j^{-\nu-1}\nu x_{j+1}^{-\nu-1}, \\ \text{if } 1 < x_j < x_{j+1} < \infty.$$

Let $U = \frac{X_{j+1:n}}{X_{j:n}}$ and $V = X_{j:n}$. In this case, it is known that U and V are independent with $U \sim$ Pareto($(n-j)\nu$) and the pdf of V is

$$f_V(v) = \frac{n!}{(j-1)!(n-j)!}(1-v^{-\nu})^{j-1}(v^{-\nu})^{n-j}\nu v^{-\nu-1} \text{if } v \geq 1; \tag{12}$$

Table 2 Values of j and c for the exponential distribution when $n = 10$ for various choices of p

p	j	c	p	j	c	p	j	c
0.01	1	0.1450	0.34	3	0.8183	0.67	7	0.2167
0.02	1	0.2915	0.35	3	0.9373	0.68	7	0.3079
0.03	1	0.4394	0.36	4	0.0455	0.69	7	0.4035
0.04	1	0.5889	0.37	4	0.1382	0.70	7	0.5047
0.05	1	0.7400	0.38	4	0.2314	0.71	7	0.6123
0.06	1	0.8927	0.39	4	0.3264	0.72	7	0.7270
0.07	1	0.0288	0.40	4	0.4243	0.73	7	0.8495
0.08	1	0.1185	0.41	4	0.5258	0.74	7	0.9806
0.09	1	0.2057	0.42	4	0.6316	0.75	8	0.0669
0.10	1	0.2967	0.43	4	0.7422	0.76	8	0.1471
0.11	1	0.3935	0.44	4	0.8579	0.77	8	0.2308
0.12	1	0.4967	0.45	4	0.9789	0.78	8	0.3196
0.13	1	0.6062	0.46	5	0.0757	0.79	8	0.4149
0.14	1	0.7219	0.47	5	0.1673	0.80	8	0.5181
0.15	1	0.8434	0.48	5	0.2603	0.81	8	0.6304
0.16	1	0.9705	0.49	5	0.3557	0.82	8	0.7530
0.17	2	0.0730	0.50	5	0.4545	0.83	8	0.8873
0.18	2	0.1655	0.51	5	0.5575	0.84	9	0.0140
0.19	2	0.2580	0.52	5	0.6653	0.85	9	0.0775
0.20	2	0.3528	0.53	5	0.7784	0.86	9	0.1451
0.21	2	0.4513	0.54	5	0.8970	0.87	9	0.2188
0.22	2	0.5543	0.55	6	0.0150	0.88	9	0.3011
0.23	2	0.6624	0.56	6	0.1039	0.89	9	0.3944
0.24	2	0.7757	0.57	6	0.1940	0.90	9	0.5013
0.25	2	0.8942	0.58	6	0.2864	0.91	9	0.6252
0.26	3	0.0134	0.59	6	0.3820	0.92	9	0.7701
0.27	3	0.1070	0.60	6	0.4819	0.93	9	0.9416
0.28	3	0.2002	0.61	6	0.5867	0.94	10	1.0406
0.29	3	0.2944	0.62	6	0.6972	0.95	10	1.1081
0.30	3	0.3911	0.63	6	0.8138	0.96	10	1.1906
0.31	3	0.4912	0.64	6	0.9371	0.97	10	1.2970
0.32	3	0.5955	0.65	7	0.0432	0.98	10	1.4470
0.33	3	0.7044	0.66	7	0.1289	0.99	10	1.7034

see Arnold et al. [1]. So, the probability of interest is

$$\begin{aligned}\Pr[(1-c)X_{j:n} + cX_{j+1:n} \le \xi_p] &= \Pr\left[X_{j:n}\left\{(1-c) + c\left(\frac{X_{j+1:n}}{X_{j:n}}\right)\right\} \le \xi_p\right] \\ &= \Pr\left[V\left\{(1-c) + cU\right\} \le \xi_p\right].\end{aligned}$$

The pdf of $W = (1 - c) + cU$ is

$$f_W(w) = (n-j)\nu \frac{(w-(1-c))^{-\nu(n-j)-1}}{c^{-\nu(n-j)}} \text{if } 1 < w < \infty. \tag{13}$$

Consequently, the probability becomes

$$\Pr(VW < \xi_p) = \Pr\left(V < \frac{\xi_p}{W}\right) = \int_w \Pr\left(V < \frac{\xi_p}{w}\right) f_W(w)dw,$$

where $F_V(v) = \Pr(V \leq v) = I_{1-\nu^{-\nu}}(j, n-j+1)$ from (12) and $f_W(w)$ is as given in (13).

5 Some Heuristic Attempts

One may be tempted to estimate the value of j in the preceding discussions without inspecting of the underlying median ranks. It certainly seems plausible to attempt to estimate j by the largest integer less than or equal to $(n+1)p$. Such approximations can lead to order statistics that are upper and lower bounds, just as $X_{j:n}$ and $X_{j+1:n}$ were in the preceding discussion. However, the order statistics are no longer contiguous. All the methodology developed in the preceding sections can be reapplied here except that the order statistics, used to form the convex class, are no longer contiguous.

Lemma 5 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from $F(x)$, which is strictly monotone on the support of X. Let p be a real-number in the interval (0, 1) such that $p \in \left(\frac{1}{n+1}, \frac{n}{n+1}\right)$, and ξ_p be the pth quantile of $F(x)$. Then, there exists a $j \in \{1, \ldots, n\}$ such that $\Pr\left(X_{j:n} < \xi_p\right) > \frac{1}{2}$ and $\Pr\left(X_{n-j+1:n} < \xi_p\right) < \frac{1}{2}$.*

Proof Define j as $j = [\![(n+1)p]\!]$, where $[\![x]\!]$ denotes the largest integer less than or equal to x. Observe that

$$\Pr\left(X_{j:n} < \xi_p\right) = \Pr\left[F\left(X_{j:n}\right) < F\left(\xi_p\right)\right] = \Pr\left(U_{j:n} < p\right),$$

where $U_{j:n}$ is the jth order statistic from a random sample of size n from the Uniform(0,1) distribution. As mentioned in Sect. 1, $U_{j:n} \sim \mathscr{B}(j, n-j+1)$, where $\mathscr{B}(\alpha, \beta)$ denotes a beta distribution with shape parameters α and β. Without loss of generality, we assume that $p \leq \frac{1}{2}$ and so $j \leq \frac{n+1}{2}$. If $j < \frac{n+1}{2}$, then $U_{j:n}$ is unimodal and positively skewed and therefore satisfies the mode-median-mean inequality. Hence, it follows that

$$\Pr\left(U_{j:n} < p\right) \geq \Pr\left(U_{j:n} < \frac{j}{n+1}\right) \geq \frac{1}{2}.$$

This means that the order statistic $X_{j:n}$ underestimates ξ_p. Notice that this result is nonparametric in the sense that it does not depend on the form of the distribution function $F(x)$, only that it be strictly monotone on its support.

Using symmetry arguments, we can then prove that

$$\Pr\left(X_{n-j+1:n} > \xi_p\right) \geq \frac{1}{2},$$

which means that $X_{n-j+1:n}$ overestimates ξ_p. Hence, the required result.

Lemma 6 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from $F(x)$, which is strictly monotone on the support of X. Let $p \in \left(\frac{1}{n+1}, \frac{n}{n+1}\right)$. With $w \in (0,1)$, let us consider a class of estimators $\hat{\xi}_p(w) = wX_{j:n} + (1-w)X_{n-j+1:n}$ for the pth quantile ξ_p, where $j = [(n+1)p]$ and $j < n - j + 1$. Then, $\Pr\left(\hat{\xi}_p(w) < \xi_p\right)$ is a continuous increasing function of w.*

Proof With $j = [(n+1)p]$, let us define

$$Q_{n,p}(w) = \Pr\left(\hat{\xi}_p(w) < \xi_p\right). \tag{14}$$

By Lemma 5, we have $Q_{n,p}(1) = \Pr\left(X_{j:n} < \xi_p\right) > \frac{1}{2}$ and $Q_{n,p}(0) = \Pr(X_{n-j+1:n} < \xi_p) < \frac{1}{2}$. Since $F(x)$ is continuous, $Q_{n,p}(w)$ is continuous. Hence, for $0 < w_1 < w_2 < 1$, we have

$$\begin{gathered} w_1 X_{j:n} + (1-w_1) X_{n-j+1:n} < w_2 X_{j:n} + (1-w_2) X_{n-j+1:n}, \\ \Pr\left\{w_1 X_{j:n} + (1-w_1) X_{n-j+1:n} < x\right\} < \Pr\left\{w_2 X_{j:n} + (1-w_2) X_{n-j+1:n} < x\right\}, \\ Q_{n,p}(w_1) < Q_{n,p}(w_2). \end{gathered}$$

Thus $Q_{n,p}(w)$ is a continuous increasing function, as desired.

Theorem 6 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from a random sample from $F(x)$, which is strictly monotone on the support of X. Let $p \in \left(\frac{1}{n+1}, \frac{n}{n+1}\right)$. With $w \in (0,1)$, let us consider a class of estimators $\hat{\xi}_p(w) = wX_{j:n} + (1-w)X_{n-j+1:n}$ for the pth quantile ξ_p, where $j = [(n+1)p]$. Then, there exists a unique value w_0 $(0 < w_0 < 1)$ such that $\Pr\left(\hat{\xi}_p(w_0) < \xi_p\right) = \frac{1}{2}$.*

6 Applications

In this section, we illustrate the results of the last section for the special cases of uniform and exponential distributions.

6.1 Uniform Distribution

Use of $\lfloor (n+1)p \rfloor$ **for** j

In order to evaluate $Q_{n,p}(w)$, we must develop an expression for the cdf of the convex linear combination of $X_{j:n}$ and $X_{n-j+1:n}$. Of special interest is the Uniform(0,1) distribution since in this case the subsequent estimator will be an L-estimator of ξ_p.

Suppose $U_{1:n}, \ldots, U_{n:n}$ are the order statistics from the uniform Uniform(0,1) distribution. Then, the joint density function of $U_{j:n}$ and $U_{n-j+1:n}$ is given by

$$f(u, v) = \frac{n!}{[(j-1)!]^2 (n-2j)!} u^{j-1} (v-u)^{n-2j} (1-v)^{j-1} \text{ if } 0 < u < v < 1.$$

Performing the transformation $\bar{w} = 1 - w$ and $q = wu + (1-w)v$, we arrive at

$$f(u, q; w) = \frac{n!}{[(j-1)!]^2 (n-2j)! \bar{w}^{n-j}} u^{j-1} (q-u)^{n-2j} (\bar{w} + wu - q)^{j-1},$$

if $0 < u < q < wu + \bar{w} < 1$. By noting the ranges of integration as $0 < u < q$ for $0 \le q \le 1-w$ and $\frac{q-\bar{w}}{w} < u < q$ for $1 - w \le q \le 1$ and once again making use of binomial expansions, we find the density of q to be:

$$f(q; w) = \begin{cases} \sum_{r=0}^{j-1} (-1)^r \binom{n-2j+r}{r} \frac{w^r}{\bar{w}^{n-2j+r+1}} \frac{q^{n-j+r}(1-q)^{j-1-r}}{B(n-j+r+1, j-r)} & \text{if } 0 \le q \le 1-w, \\ \sum_{r=0}^{j-1} \binom{2j-r-2}{j-r-1} \frac{\bar{w}^{j-r-1}}{w^{n-j}} \frac{(q-\bar{w})^r (1-q)^{n-r-1}}{B(r+1, n-r)} & \text{if } 1 - w < q \le 1. \end{cases}$$

Therefore, the distribution function of q is

$$F(q; w) = \begin{cases} \sum_{r=0}^{j-1} \Big\{ (-1)^r \binom{n-2j+r}{r} \frac{w^r}{\bar{w}^{n-2j+r+1}} & \\ \qquad \times I_q(n-j+r+1, j-r) \Big\} & \text{if } 0 \le q \le 1-w, \\ \sum_{r=0}^{j-1} (-1)^r \binom{n-2j+r}{r} \frac{w^r}{\bar{w}^{n-2j+r+1}} I_{\bar{w}}(n-j+r+1, j-r) & \\ \quad + \sum_{r=0}^{j-1} \binom{2j-r-2}{j-r-1} w^j \bar{w}^{j-1-r} I_{\frac{q-\bar{w}}{w}}(r+1, n-r) & \text{if } 1-w < q \le 1, \end{cases}$$

respectively, where, as defined earlier, $I_q(a, b)$ is the incomplete beta ratio and $B(a, b)$ is the complete beta function. For fixed n and p, there exists a unique value of w for which

$$F(p; w) = Q_{n,p}(w) = \frac{1}{2}, \tag{15}$$

where $Q_{n,p}(w)$ is as given in (14). Thus, we can regard the corresponding $Q_{n,p}(w)$ as a nonparametric competitor to the Harrell-Davis [11] estimator, which is a robust L_1-estimator. Of course, for this purpose, we need to solve (15) for w, for given values of n and p.

Table 3 Values of w satisfying (15), for different choices of n and p

		p				
n	j	$\frac{j}{2j+1}$	$\frac{j+\frac{1}{4}}{2j+1}$	$\frac{j+\frac{1}{2}}{2j+1}$	$\frac{j+\frac{3}{4}}{2j+1}$	$\frac{j+1}{2j+1}$
4	2	0.9313	0.7056	0.5000	0.2944	0.0687
6	3	0.9509	0.7178	0.5000	0.2822	0.0491
8	4	0.9619	0.7248	0.5000	0.2752	0.0381
10	5	0.9689	0.7293	0.5000	0.2707	0.0311
12	6	0.9738	0.7325	0.5000	0.2675	0.0262

Special Case
If $n = 2m$ and $p \in \left(\frac{m}{2m+1}, \frac{m+1}{2m+1}\right)$, then $j = [(n+1)p] = [(2m+1)/2] = \left[m + \frac{1}{2}\right] = m$. In this case, by solving (15) for varying n, we found the values of w for different choices of p, and these are presented in Table 3.

Use of the largest order statistic that underestimates p
Suppose $U_{1:n}, \ldots, U_{n:n}$ are the order statistics from the uniform $\mathcal{U}(0, 1)$ distribution. Earlier, we considered $i = |(n+1)p|$, but it is possible that $U_{i:n}$ may underestimate p with a probability of at least 1/2. Yet, that does not guarantee that $U_{i+1:n}$ overestimates p with probability of at least 1/2. However, such an i does exist, but it just may not correspond to $|(n+1)p|$.

Consider the joint density of $U_{i:n}$ and $U_{i+1:n}$ given by

$$f(u, v) = \frac{n!}{(i-1)!)(n-i-1)!} u^{i-1}(1-v)^{n-i-1}, \; 0 < u < v < 1.$$

Letting $U = u$ and $Q = wU + (1-w)V$, then the joint density becomes

$$f(u, q; w) = \frac{n!}{(i-1)!(n-i-1)!\bar{w}} u^{i-1} (\bar{w} - q + wu)^{n-i-1} \text{ if } 0 < u < q < wu + \bar{w} < 1$$

where again $\bar{w} = 1 - w$. As in the previous case, noting the ranges of integration as $0 < u < q$ for $0 \le q \le 1-w$ and $\frac{q-\bar{w}}{w} < u < q$ for $1 - w < q \le 1$, we find the corresponding density and distribution functions of q to be:

$$f(q; w) = \begin{cases} \sum_{r=0}^{n-i-1} (-1)^r \frac{w^r}{\bar{w}^{r+1}} \frac{q^{r+i}(1-q)^{n-i-1-r}}{B(r+i+1, n-i-r}) & \text{if } 0 \le q \le 1-w, \\ \frac{n!}{(i-1)!(n-i-1)!} \sum_{r=0}^{i-1} \binom{i-1}{r} \frac{\bar{w}^{i-1-r}}{w^i} \frac{(q-\bar{w})^r (1-q)^{n-r-1}}{n-r-1} & \text{if } 1 - w < q \le 1, \end{cases}$$

and so

$$F(q;w)=\begin{cases}\sum_{r=0}^{n-i-1}(-1)^r\frac{w^r}{\bar{w}^{r+1}}I_q(r+i+1,n-i-r) & \text{if } 0\le q\le 1-w,\\ \sum_{r=0}^{n-i-1}\frac{w^r}{\bar{w}^{r+1}}I_{1-w}(r+i+1,n-i-r) & \\ +\sum_{r=0}^{i-1}\sum_{s=0}^{n-r-1}\Big\{(-1)^s\binom{n}{s}\binom{n-r-2}{i-1-r}\frac{\bar{w}^{i-1-r+s}}{w^i} & \\ \quad\times I_{q-\bar{w}}(r+1,n-r-s)\Big\} & \text{if } 1-w<q\le 1.\end{cases}$$

We can now solve for w, using successive order statistics, instead of the earlier approach when the two order statistics are determined by the mean rank approach.

6.2 Exponential Distribution

Use of $|(n+1)p|$ for j
In the case of exponential distribution, by proceeding in a manner analogous to the uniform case, we can show that the cdf of q is

$$F(q)=\frac{n!}{[(j-1)!]^2(n-2j)!\bar{w}}\sum_{r=0}^{j-1}\sum_{s=0}^{n-2j}\frac{(-1)^{r-s}\binom{j-1}{r}\binom{n-2j}{s}}{r+n-2j-s-\frac{w}{\bar{w}(s+j-1}}$$
$$\times\left[\frac{\bar{w}}{s+j-1}\left(1-e^{-\left(\frac{s+j-1}{\bar{w}}\right)q}\right)-\frac{1}{r+n-j-1}\left(1-e^{-(r+n-j-1)q}\right)\right],$$
$$0<q<\infty.$$

Use of the largest order statistic that underestimates p
Here again, for the case of exponential distribution, by proceeding in a manner analogous to the uniform case, we can show that the cdf of q is

$$F(q)=\frac{n!}{(n-j)!\bar{w}}\frac{\Gamma\left(\frac{(n-j)w}{\bar{w}}+1\right)}{\Gamma(j+\frac{(n-j)w}{\bar{w}}+1)}\left\{1-I_{e^{-q}}\left(j,\frac{(n-j)w}{\bar{w}}+1\right)\right\}e^{-\frac{(n-j)}{\bar{w}}q},$$
$$\text{if } 0<q<\infty, \tag{16}$$

where $I_q(a,b)$ is the incomplete beta ratio defined earlier. We may use (16) to determine w such that

$$F(p;w)=\frac{1}{2}\text{ and } j \text{ is the greatest integer such that } m_{j:n}\le p.$$

7 Concluding Remarks

Pitman closeness of order statistics to population parameters such as quantiles have been discussed in the literature. Here, we have discussed Pitman closest estimation based on convex linear combinations of two contiguous order statistics. We have then illustrated the developed results for the uniform, exponential, power function and Pareto distributions. As done in the case of quantile estimation, one may also propose convex linear combinations of two contiguous order statistics as Pitman closest predictors of a future failure time. This work is currently under progress and we hope to report these findings in a future paper.

Acknowledgments The authors thank the editors, Drs. Pankaj Choudhary, Chaitra H. Nagaraja and Tony Ng, for their kind invitation to present this paper for the volume. Our sincere thanks go to an anonymous reviewer whose valuable comments and suggestions on an earlier version of this manuscript led to this significantly improved version. We also take this opportunity to congratulate Dr. H.N. Nagaraja for his accomplishments so far and hopefully for many more productive years in the future!

References

1. Arnold, B.C., N. Balakrishnan, and H.N. Nagaraja. 1992. *A first course in order statistics*. New York, NY: Wiley.
2. Balakrishnan N., K. Davies. 2013. Pitman closeness results for type-I censored data from exponential distribution. *Statistics and Probability Letters*, 2693–2698.
3. Balakrishnan, N., K. Davies, and J.P. Keating. 2009. Pitman closeness of order statistics to population quantiles. *Communications in Statistics - Simulation and Computation* 38: 802–820.
4. Balakrishnan, N., G. Iliopoulos, J.P. Keating, and R.L. Mason. 2009. Pitman closeness of sample median to population median. *Statistics and Probability Letters* 79: 1759–1766.
5. Balakrishnan, N., K.F. Davies, J.P. Keating, and R.L. Mason. 2010. Simultaneous closeness among order statistics to population quantiles. *Journal of Statistical Planning and Inference* 140: 2408–2415.
6. Balakrishnan, N., K.F. Davies, J.P. Keating, and R.L. Mason. 2011. Pitman closeness, monotonicity and consistency of best linear unbiased and invariant estimators for exponential distribution under Type-II censoring. *Journal of Statistical Computation and Simulation* 81: 985–999.
7. Balakrishnan, N., K. Davies, J.P. Keating, and R.L. Mason. 2012. Computation of optimal plotting points based on Pitman closeness with an application to goodness-of-fit for location-scale families. *Computational Statistics and Data Analysis* 56: 2637–2649.
8. Balakrishnan, N., and A.C. Cohen. 1991. *Order statistics and inference: Estimation methods*. San Diego, CA: Academic Press.
9. David, H.A., and H.N. Nagaraja. 2003. *Order statistics*, 3rd ed. Hoboken, NJ: Wiley.
10. Fountain, R.L. 1991. Pitman closeness comparison of linear estimators: A canonical form. *Communications in Statistics - Theory and Methods* 20: 3535–3550.
11. Harrell, F.E., and C.E. Davis. 1982. A new distribution-free quantile estimator. *Biometrika* 69: 635–640.
12. Keating, J.P., R.L. Mason, and P.K. Sen. 1983. *Pitman's measure of closeness*. Philadelphia: Society for Industrial and Applied Mathematics.

13. Lloyd, E.H. 1952. Least squares estimation of location and scale parameters using order statistics. *Biometrika* 39: 88–95.
14. Mason, R.L., J.P. Keating, P.K. Sen, and N.W. Blaylock. 1990. Comparison of linear estimators using Pitman's measure of closeness. *Journal of the American Statistical Association* 85: 579–581.
15. Nagaraja, H.N. 1986. Comparison of estimators and predictors from two-parameter exponential distribution. *Sankyha Series B*, 10–18.
16. Peddada, S.D., and R. Khattree. 1991. Comparison of estimators of the location parameter using Pitman's closeness criterion. *Communications in Statistics - Theory and Methods* 20: 3525–3534.
17. Pitman, E.J.G. 1937. The closest estimates of statistical parameters. *Proceedings of the Cambridge Philosophical Society* 33: 212–222.
18. Rao, C.R. 1981. Some comments on the minimum mean square error as a criterion in estimation. *Statistics and Related Topics*, 123–143. Amsterdam: North-Holland.

Springer

http://www.springer.com/978-3-319-25431-9

Ordered Data Analysis, Modeling and Health Research Methods
In Honor of H. N. Nagaraja's 60th Birthday
Choudhary, P.K.; Nagaraja, C.H.; Ng, H.K.T. (Eds.)
2015, XVII, 265 p. 39 illus., 30 illus. in color., Hardcover
ISBN: 978-3-319-25431-9