

Simultaneous Closeness among Order Statistics to Population Quantiles

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

Abstract We derive expressions for the probability that an individual order statistic is closest to the target parameter among all order statistics from a complete random sample. We then apply these general results to location-scale parameter families of distributions with specific applications to estimation of percentiles. In this case, simultaneous-closeness probabilities depend upon the parameters through the value of p in the percentile and the sample size, n . The results can be applied to distributions with both bounded and unbounded supports. We illustrate the results with the estimation of percentiles for Normal and exponential.

Key words: Order Statistics, Simultaneous Closeness, Percentile.

1 Introduction

In this paper, we determine among all order statistics in a random sample, which one is most-frequently closest among all other order statistics to a population quantile. In this regard, we consider all order statistics simultaneously as opposed to multiple paired comparisons as done in Balakrishnan et al. [4]. In computing the probability with which an order statistic is most-frequently closest among all those in a sample,

N. Balakrishnan
McMaster University, Hamilton ON, Canada, e-mail: bala@mcmaster.ca

Katherine F. Davies
University of Manitoba, Winnipeg MB, Canada e-mail: katherine.davies@umanitoba.ca

Jerome P. Keating
The University of Texas, San Antonio TX, US, e-mail: jerome.keating@utsa.edu

Robert L. Mason
Southwest Research Institute, San Antonio TX, US, e-mail: Robert.Mason@swri.org

we are computing the simultaneous version of Pitman's measure of closeness as formulated by Blyth [8]. We restrict our attention to location-scale parameter families.

Definition 1. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be univariate estimators of a real-valued parameter θ . Pitman Closeness (PC) is defined as

$$P(\hat{\theta}_1, \hat{\theta}_2 | \theta, n) = \Pr(|\hat{\theta}_1 - \theta| < |\hat{\theta}_2 - \theta|). \quad (1)$$

This measure quantifies the frequency with which one estimator is closer to the value of a parameter than a competing estimator. We refer to the monograph by Keating, Mason and Sen [11] for pertinent details about this definition, as well as other Pitman-related definitions.

Recently, Pitman closeness of order statistics to population median and quantiles have been discussed by Balakrishnan et al. [7] and Balakrishnan et al. [4], respectively. Pitman closeness comparison of estimators and predictors based on censored samples from exponential distribution have also been carried out by Balakrishnan et al. [5] and Balakrishnan et al. [6]. A similar discussion based on record values has been provided by Ahmadi and Balakrishnan [1] and Ahmadi and Balakrishnan [2].

1.1 Terminology

We introduce methods for partitioning the essential range of a random vector of observations into regions in which each order statistic is the best. In light of Blyth's criteria, these partitions have specific applications to the first simultaneous-closeness criterion. Much insight can be gained by an examination of the regions in which each order statistic has the smallest loss, in that we may develop better methods for improving upon competing estimators based on its value and its order within the sample. For this purpose, we state common conditions needed throughout the article.

Common Conditions: We assume $X_{1:n}, \dots, X_{n:n}$ are the order statistics obtained from a random sample of size n taken on a random variable, X . We assume X has an absolutely continuous pdf, $f(x)$, cdf, $F(x)$, and survival function, $\bar{F}(x) = 1 - F(x)$.

Then, the marginal density function of $X_{i:n}$ is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) [F(x)]^{i-1} [\bar{F}(x)]^{n-i} \quad (2)$$

and the joint density of $X_{i:n}$ and $X_{j:n}$ (for $i < j$) is given by

$$\begin{aligned} f_{i,j:n}(x, y) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(x) f(y) \\ &\quad \times [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [\bar{F}(y)]^{n-j} \end{aligned} \quad (3)$$

for $-\infty < x < \infty$ and $-\infty < x < y < \infty$, respectively; see, for example, Arnold, Balakrishnan and Nagaraja [3].

We define simultaneous-closeness probabilities as follows.

Definition 2. The simultaneous-closeness probability (SCP) of $X_{i:n}$, $i \in \{1, \dots, n\}$, among the order statistics, $X_{1:n}, \dots, X_{n:n}$, in the estimation of a population parameter θ is

$$\pi_{i:n}(\theta) = \Pr \left\{ |X_{i:n} - \theta| < \min_{j, j \neq i} |X_{j:n} - \theta| \right\}. \quad (4)$$

The results throughout the article are predicated on the support of the random variable X (i.e., whether the support is the entire real line or a bounded open interval). We assume the support is a connected open subset of the reals regardless. Moreover, we are interested in parameters with a parameter space that coincides with the support of X . We present general results for random variables with complete support (i.e., $\mathfrak{R}$) and similar ones for random variables with bounded support.

2 Simultaneous Closeness of Order Statistics

Assume that the common conditions hold for a random variable X . Let θ be a population parameter associated with the random variable X such that the range space of θ is the support of X . Our objective is to determine the probability with which each order statistic, $X_{i:n}$ for each $i \in \{1, 2, \dots, n\}$, is simultaneously-closest to θ when compared with the remaining order statistics. Using Voroni regions, as done by Fountain et al., [9] this probability can be found by carrying out just two comparisons. We arrive at

$$\pi_{i:n}(\theta) = P(X_{i:n}, X_{i-1:n} | \theta, n) - P(X_{i+1:n}, X_{i:n} | \theta, n). \quad (5)$$

In the case where θ is a population quantile, expressions for this probability can be found in Balakrishnan et al. [4]. They can therefore be used to express the simultaneous-closeness probabilities in the context of the estimation of population quantiles.

Theorem 1. Assume that the common conditions hold for X , with support $\mathfrak{R}$. Then for $i = 2, \dots, n-1$, the simultaneous-closeness probability, $\pi_{i:n}(\theta)$, of $X_{i:n}$ to θ is given by

$$\begin{aligned} \pi_{i:n}(\theta) &= \binom{n}{i-1} [F(\theta)]^{i-1} [1 - F(\theta)]^{n-i+1} \\ &+ \int_0^{F(\theta)} \frac{n!}{(i-1)!(n-i)!} [\bar{F}(2\theta - F^{-1}(u))]^{n-i} u^{i-1} du \\ &- \int_0^{F(\theta)} \frac{n!}{(i-2)!(n-i+1)!} [\bar{F}(2\theta - F^{-1}(u))]^{n-i+1} u^{i-2} du. \end{aligned} \quad (6)$$

In the cases of $i=1$ and $i=n$, simplified versions of the simultaneous-closeness probabilities can be derived.

2.1 Bounded Support

Theorem 2. Assume that the common conditions hold for X , where the support of X is bounded on the interval (a, b) . Then, the simultaneous-closeness probability $\pi_{i:n}(\theta)$ is given by

$$\begin{aligned} \pi_{i:n}(\theta) = n \binom{n-1}{i-2} \int_a^{b^*} f(x) [F(x)]^{i-2} \\ \times \left\{ [F(h_2(x))]^{n-i+1} - [F(h_1(x))]^{n-i+1} \right\} dx, \end{aligned} \quad (7)$$

where $b^* = \min(b, 2\theta - a)$, $h_1(x) = \max(a, x)$ and $h_2(x) = \min(b, 2\theta - x)$.

3 Applications

Let us proceed as in Balakrishnan et al. [4], where $X_1, \dots, X_n$ is a random sample of size n from a continuous population with a cdf $F(x)$, pdf $f(x)$, and the corresponding order statistics denoted by $X_{1:n} \leq \dots \leq X_{n:n}$. For $p \in (0, 1)$, let x_p be the p^{th} quantile of $F(x)$. If $F(\cdot)$ belongs to the location-scale family, viz.,

$$F(x) = G\left(\frac{x-\mu}{\sigma}\right) \quad \text{and} \quad f(x) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right) \quad \forall x \in \mathfrak{R},$$

where $\mu \in \mathfrak{R}$ is the location parameter and $\sigma > 0$ is the scale parameter. Then, $z_p = (x_p - \mu)/\sigma$ is the p^{th} quantile of $G(z)$. Finally, let $Z_1, \dots, Z_n$ be a random sample of size n from the standard population with cdf $G(z)$, pdf $g(z)$, and $Z_{1:n} \leq \dots \leq Z_{n:n}$ denote the corresponding order statistics.

3.1 Complete Support

If we set $\theta = x_p$, then $F(x_p) = p$. When we substitute these into Theorem 1, we obtain expressions for the simultaneous-closeness probabilities, which depend on p and n , and not on μ and σ .

Corollary 1. We have, for $i = 2, \dots, n-1$,

$$\begin{aligned} \pi_{i:n}(p) = & \binom{n}{i-1} p^{i-1} (1-p)^{n-i+1} \\ & + \frac{n!}{(i-1)!(n-i)!} \int_0^p \{1 - G[2G^{-1}(p) - G^{-1}(u)]\}^{n-i} u^{i-1} du \\ & - \frac{n!}{(i-2)!(n-i+1)!} \int_0^p \{1 - G[2G^{-1}(p) - G^{-1}(u)]\}^{n-i+1} u^{i-2} du. \end{aligned} \quad (8)$$

The parameter-free result suggests that the probabilities can be computed and cataloged for various choices of p and n for different families such as normal, extreme-value, Cauchy, Laplace, logistic, and so on. These may then be used in developing efficient goodness-of-fit methods and also as “good” plotting points in graphical model validity methods. Work is currently under progress in this direction and we hope to report these findings in a future paper.

The following corollary simplifies the computation in the case when the standard distribution is symmetric about the origin.

Corollary 2. *If the standard pdf is symmetric about the origin (i.e., $g(z) = g(-z)$, $G(-z) = 1 - G(z)$ and $G^{-1}(p) = -G^{-1}(q)$, where $q = 1 - p$), then we have*

$$\pi_{i:n}(p) = \pi_{n-i+1:n}(q) \quad (9)$$

for $i = 1, \dots, n$.

3.1.1 Normal

In the normal distribution, the standard forms are

$$g(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad G(z) = \Phi(z) \text{ and } G^{-1}(p) = \Phi^{-1}(p).$$

When we substitute these expressions into Eq. (8) and use the fact that $\Phi(-z) = 1 - \Phi(z)$, we obtain

$$\begin{aligned} \pi_{i:n}(p) &= \binom{n}{i-1} p^{i-1} (1-p)^{n-i+1} \\ &+ \frac{n!}{(i-1)!(n-i)!} \int_0^p \{\Phi[\Phi^{-1}(u) - 2\Phi^{-1}(p)]\}^{n-i} u^{i-1} du \\ &- \frac{n!}{(i-2)!(n-i+1)!} \int_0^p \{\Phi[\Phi^{-1}(u) - 2\Phi^{-1}(p)]\}^{n-i+1} u^{i-2} du. \end{aligned}$$

We have tabulated these probabilities for values of $i = 1, \dots, 10$, $p = 0.05(0.05)0.50$ and $n = 10$ in Table 1. Then, the values for $p = 0.55(0.05)0.95$ can be found readily by using the symmetry property in Corollary 2. Table 1 presents a clear picture of the most-preferred order statistic for the p^{th} quantile and we do observe from this table that the most-preferred order statistic, with index i^* , increases from 1 to 10 as p increases from 0.05 to 0.95. In this even sample-size case, we also note that the fifth and sixth order statistics have the highest simultaneous-closeness probability when $p = 0.50$ as one would expect.

Table 1 SCP of order statistics for Normal distribution when $n = 10$ for various values of p and i

$p \backslash i$	1	2	3	4	5	6	7	8	9	10
0.05	0.78921	0.18365	0.02478	0.00222	0.00014	0.00001	0.00000	0.00000	0.00000	0.00000
0.10	0.52231	0.34097	0.11082	0.02252	0.00308	0.00029	0.00002	0.00000	0.00000	0.00000
0.15	0.31927	0.37855	0.21151	0.07179	0.01613	0.00247	0.00026	0.00002	0.00000	0.00000
0.20	0.18403	0.33583	0.28126	0.14078	0.04616	0.01024	0.00153	0.00015	0.00001	0.00000
0.25	0.10058	0.25937	0.30255	0.20883	0.09375	0.02833	0.00576	0.00076	0.00006	0.00000
0.30	0.05210	0.18049	0.28047	0.25626	0.15155	0.06011	0.01598	0.00274	0.00028	0.00001
0.35	0.02548	0.11483	0.23100	0.27212	0.20682	0.10513	0.03573	0.00783	0.00100	0.00006
0.40	0.01169	0.06712	0.17152	0.25607	0.24613	0.15794	0.06765	0.01865	0.00300	0.00022
0.45	0.00499	0.03599	0.11543	0.21606	0.26008	0.20878	0.11177	0.03848	0.00773	0.00069
0.50	0.00195	0.01758	0.07031	0.16406	0.24609	0.24609	0.16406	0.07031	0.01758	0.00195

3.2 Bounded Support

From Theorem 2, we have developed a general expression for $\pi_{i:n}(\theta)$ when the support of X is bounded on the interval (a, b) . If X is bounded on the interval (a, b) , then $Z = (X - \mu)/\sigma$ is bounded on the interval (a', b') , where $b' = (b - \mu)/\sigma$ and $a' = (a - \mu)/\sigma$. Let $z_p = (x_p - \mu)/\sigma$. The simultaneous-closeness probability of $X_{i:n}$ to x_p in this case is given by

$$\pi_{i:n}(p) = n \binom{n-1}{i-2} \int_{a'}^{b^{**}} g(z) [G(z)]^{i-2} \times \left\{ [G(h_2(z))]^{n-i+1} - [G(h_1(z))]^{n-i+1} \right\} dz, \quad (10)$$

where $b^{**} = \min(b', 2z_p - a')$, $h_1(z) = \max(a', z)$ and $h_2(z) = \min(b', 2z_p - z)$.

3.2.1 Exponential

In the case of the standard exponential distribution, we have $a' = 0$ and $b' = \infty$, and the standard forms, for $z > 0$, in this case are

$$g(z) = e^{-z}, \quad G(z) = 1 - e^{-z} \quad \text{and} \quad G^{-1}(z) = -\ln(1 - z).$$

We can use expressions from Balakrishnan et al. [4] to express the simultaneous probabilities for x_p as

$$\begin{aligned} \pi_{i:n}(p) &= P(X_{i:n}, X_{i-1:n} | p, n) + P(X_{i:n}, X_{i+1:n} | p, n) - 1 \\ &= \pi_{(i)i-1}(p) + \pi_{(i)i+1}(p) - 1. \end{aligned} \quad (11)$$

where $\pi_{(\ell),i}(p)$ is given in Result 7 (in Eq. (23) for $i = \ell + 1, \dots, n$ and Eq. (25) for $i = 1, \dots, \ell - 1$) of Balakrishnan et al. [4]. The combined expression for the simultaneous-closeness probabilities in the exponential distribution can be simpli-

fied using the following integral:

$$I_{\ell,m}(p) = \int_0^p \frac{u^{\ell-1}}{(1-u)^m} du \quad \text{for } \ell, m = 1, 2, \dots \quad (12)$$

With the integral in (12), $\pi_{i:n}(p)$ can be expressed in this case as

$$\pi_{i:n}(p) = \binom{n}{i-1} p^{i-1} (1-p)^{n-i+1} + a_{i:n} q^{2(n-i)} I_{i,n-i}(p) - b_{i:n} q^{2(n-i+1)} I_{i-1,n-i+1}(p), \quad (13)$$

where $a_{i:n} = \frac{n!}{(i-1)!(n-i)!}$ and $b_{i:n} = \frac{n!}{(i-2)!(n-i+1)!}$. We have tabulated these values for $n = 10$ and $i = 1, \dots, 10$ for incremental values of $p = 0.05(0.05)0.95$ in Table 2. Quite interestingly, we find the simultaneous-closest order statistic to the p^{th} -quantile in this case to be quite close to that of the normal case and the corresponding simultaneous-closeness probabilities to be quite close as well. This reveals that the concept discussed here is naturally quite robust.

In Table 2, we observe the lack of symmetry in the simultaneous-closeness probabilities since the standard exponential is skewed to the right. In addition, while the probabilities $\pi_{i:n}(p)$ and $\pi_{n-i+1:n}(q)$ are not equal, the range of percentiles over which an order statistic is preferred. We also see preference among the order statistics moves from 1 to 10 as p increases from 0.05 to 0.95, as one would expect.

Table 2 SCP of order statistics for exponential distribution when $n=10$ for various values of p and i

$p \backslash i$	1	2	3	4	5	6	7	8	9	10
0.05	0.85064	0.12850	0.01894	0.00180	0.00012	0.00001	0.00000	0.00000	0.00000	0.00000
0.10	0.59691	0.28721	0.09362	0.01930	0.00268	0.00026	0.00002	0.00000	0.00000	0.00000
0.15	0.37591	0.35232	0.19064	0.06422	0.01444	0.00222	0.00023	0.00002	0.00000	0.00000
0.20	0.21907	0.33210	0.26540	0.13020	0.04232	0.00936	0.00140	0.00014	0.00001	0.00000
0.25	0.11966	0.26652	0.29531	0.19842	0.08774	0.02629	0.00532	0.00070	0.00005	0.00000
0.30	0.06152	0.19003	0.28078	0.24896	0.14444	0.05656	0.01490	0.00254	0.00025	0.00001
0.35	0.02975	0.12273	0.23570	0.26931	0.20031	0.10023	0.03367	0.00731	0.00093	0.00005
0.40	0.01348	0.07236	0.17753	0.25737	0.24188	0.15247	0.06438	0.01754	0.00280	0.00020
0.45	0.00567	0.03895	0.12074	0.21997	0.25894	0.20397	0.10742	0.03646	0.00723	0.00064
0.50	0.00219	0.01905	0.07410	0.16882	0.24792	0.24318	0.15927	0.06714	0.01653	0.00181
0.55	0.00077	0.00836	0.04075	0.11602	0.21257	0.25985	0.21190	0.11114	0.03402	0.00463
0.60	0.00024	0.00324	0.01981	0.07071	0.16233	0.24854	0.25378	0.16663	0.06384	0.01087
0.65	0.00006	0.00108	0.00832	0.03754	0.10896	0.21100	0.27263	0.22665	0.11001	0.02375
0.70	0.00001	0.00029	0.00292	0.01685	0.06275	0.15616	0.25967	0.27827	0.17438	0.04869
0.75	0.00000	0.00006	0.00080	0.00609	0.02976	0.09752	0.21416	0.30402	0.25325	0.09434
0.80	0.00000	0.00001	0.00016	0.00162	0.01081	0.04845	0.14623	0.28690	0.33239	0.17344
0.85	0.00000	0.00000	0.00002	0.00027	0.00261	0.01707	0.07561	0.21977	0.38173	0.30291
0.90	0.00000	0.00000	0.00000	0.00002	0.00031	0.00328	0.02407	0.11796	0.35384	0.50052
0.95	0.00000	0.00000	0.00000	0.00000	0.00001	0.00015	0.00241	0.02732	0.20140	0.76871

References

1. Ahmadi, J. and Balakrishnan, N. Pitman closeness of record values to population quantiles, *Statistics & Probability Letters*, 79, pp. 2037-2044 (2009)
2. Ahmadi, J. and Balakrishnan, N. On Pitman's measure of closeness of k -records, *Journal of Statistical Computation and Simulation* (to appear).
3. Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. *A First Course in Order Statistics*, Classics in Applied Mathematics, Number 54, Society for Industrial and Applied Mathematics, Philadelphia (2008)
4. Balakrishnan, N., Davies, K. and Keating, J.P. Pitman closeness of order statistics to population quantiles, *Communications in Statistics - Theory & Methods*, 38, pp. 802-820 (2009)
5. Balakrishnan, N. Davies, K., Keating, J.P. and Mason, R.L. Pitman closeness of best linear unbiased and invariant predictors for exponential distribution in one- and two-sample situations, *Communications in Statistics - Theory & Methods* (to appear).
6. Balakrishnan, N. Davies, K., Keating, J.P. and Mason, R.L. 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* (to appear).
7. Balakrishnan, N., Iliopoulos, G., Keating, J.P. and Mason, R.L. Pitman closeness of sample median to population median, *Statistics & Probability Letters*, 79, pp. 1759-1766 (2009)
8. Blyth, C.R. Some probability paradoxes in choice from among random alternatives, *Journal of the American Statistical Association*, 67, pp. 366-381 (1972)
9. Fountain, R.L., Keating, J.P. and Maynard, H.B. The simultaneous comparison of estimators, *Mathematical Methods of Statistics*, 5, pp. 187-198 (1996)
10. Keating, J.P. and Mason, R.L. Pitman's measure of closeness, *Sankhyā, Series B*, 47, pp. 22-32 (1985)
11. Keating, J.P., Mason, R.L., and Sen, P.K. *Pitman's Measure of Closeness: A Comparison of Statistical Estimators*, Society for Industrial and Applied Mathematics, Philadelphia (1993)