AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(3): 6763–6776.
DOI:10.3934/math.2023344
Received: 21 August 2022
Revised: 07 December 2022
Accepted: 14 December 2022
Published: 09 January 2023

Research article

On asymptotic correlation coefficient for some order statistics

Jinliang Wang[1], **Fang Wang**[1,*] **and Songbo Hu**[2]

[1] Science Department of Jiujiang University, Jiangxi Province, China
[2] Jiangxi Province Key Laboratory of Preventive Medicine, School of Public Health, Nanchang University, China

* **Correspondence:** Email: Wfang0219@163.com.

Abstract: By the relationship between a continuous population X and the uniform distribution $U[0, 1]$, we gain for a sample quantile an equivalent expression of its variance and for two different sample quantiles the asymptotic correlation coefficient. As the population of interest can have no expectation, the obtained conclusions are applicable to the location estimating problem of a Cauchy distribution. On that occasion, we finally obtained a quick and effective estimator established by a linear function of some sample quantiles. For similar problems, the presented approach is worthy of reference.

Keywords: asymptotic correlation coefficient; location estimating problem; lower bound of Cramér-Rao inequality; moment convergence; order statistic
Mathematics Subject Classification: 62F10, 62G30

1. Introduction

Let a random sample $X_1, ..., X_n$ be drawn from a population X according to a distribution $F(x, \theta)$ where θ is unknown. If a statistic $\widehat{\theta_n}$ is unbiased for θ and there are known positive constants b_n's such that the normalized random sequence $\{(\widehat{\theta_n} - \theta)/b_n, n \geq 1\}$ converges in both the first and the second moment to a known distribution of a random variable (RV), say ξ, with a known variance σ^2, then the effectiveness of the unbiased estimator $\widehat{\theta_n}$ can be assessed by the variance $b_n^2\sigma^2$, the less the variance, the more effective the estimator. However, the Cramér-Rao inequality indicates that to estimate the unknown parameter θ of the distribution $F(x, \theta)$, the unbiased estimator usually has a variance not less than $1/(nI(\theta))$ where $I(\theta)$ denotes the Fisher information. That indicates that an unbiased estimator with a variance reaching the lower bound $1/(nI(\theta))$ is sure of minimum variance.

Under a large sample size, the maximum likelihood estimate (m.l.e) method usually (but not always) yields a theoretically desirable estimator, say $\widehat{\theta_n}$, in a sense that $\widehat{\theta_n}$ has an asymptotic normal distribution $N(\theta, 1/(nI(\theta)))$ with a variance reaching the lower bound of the well-known Cramér-Rao inequality (see [1] as a reference).

For estimating a parameter of the distribution of a population that has no expectation, the classical moment estimate method is futile. Moreover, the classical method of m.l.e usually becomes invalid too in the sense that it doesn't have a closed solution. Under such a situation, especially in the case of estimating some parameters such as the location of a population, it is worth trying to investigate an unbiased estimator established by a linear function of some sample quantiles. That will be preferable if the efficiency is close to that of the theoretical m.l.e. To approximate the efficiency of the estimator, we need the following conclusions.

Theorem 1.1. *For a population X distributed according to a continuous pdf $f(x)$, let p and r be two numbers satisfying $0 < p \le r < 1$ and x_p and x_r be respectively the p-quantile and r-quantile of X satisfying $f(x_p)f(x_r) > 0$. Let $(X_1, ..., X_n)$ be a random sample derived from X. If there are constants $\omega > 0$ and $\nu \in (-\infty, \infty)$ such that the cdf $F(x)$ of $\omega X + \nu$ has an inverse function $Q(u)$ which possesses a continuous third-order derivative function $Q'''(u)$ in the interval $(0, 1)$ satisfying*

$$|Q'''(u)| \le Ku^{-A}(1-u)^{-A} \tag{1.1}$$

for some given constants $K > 0$, $A \ge 0$ and all $u \in (0, 1)$, then:

(1) *we have, as $n \to \infty$,*

$$E\left(\frac{f(x_p)(X_{i:n} - x_p)}{\sqrt{p(1-p)/n}}\right)^2 \sim E\left(\frac{f(x_r)(X_{j:n} - x_r)}{\sqrt{r(1-r)/n}}\right)^2 \to 1$$

provided $i/n = p + o(n^{-1/2})$ and $j/n = r + o(n^{-1/2})$ as $n \to +\infty$;

(2) *the correlation coefficient $corr(X_{i:n}, X_{j:n})$ between $X_{i:n}$ and $X_{j:n}$ satisfies*

$$\lim_{n\to\infty} corr(X_{i:n}, X_{j:n}) = \sqrt{\frac{p(1-r)}{r(1-p)}} \tag{1.2}$$

provided $i/n = p + o(1)$ and $j/n = r + o(1)$ as $n \to \infty$.

Under the conditions in Theorem 1.1 but without assumption (1.1), it is mentioned (without formal proof) in [2] that the same conclusions hold according to some equations given. We have found a gap there, that is, [2] uses a partial sum of a Taylor expansion of a function to approximate the function itself without rigorous proof. Here we have to apply the assumption (1.1) to fill that gap.

In exploring the measurement of dependence or independence between two order statistics (OSs), many research works based on the Copula function method are instructive. On that subject, Barakat led the research. For references, we can consult Barakat's [3–5] and Hürlimann's [6] as well.

Item (2) in Theorem 1.1 can be regarded as a corresponding exploration of the relation between two general OSs, where we measure the asymptotic dependence or independence by providing the limiting correlation coefficient for them. That has two main advantages. First, this preferred measurement has its advantage in that the respective normalization of both OSs has no effect. Second, as was discovered by Bahadur (see [7]) or as summarized by DasGupta in [8], the asymptotic joint distributions of some sample quantiles have a multivariate normal distribution while, for two RVs according to a bivariate normal distribution, being independent is equivalent to being uncorrelated.

Under the conditions of Theorem 1.1, we see that the OSs $X_{i:n}$ and $X_{j:n}$ are asymptotically dependent, which supports Barakat's corresponding conclusion in article [3]. Moreover, Theorem 1.1 also supports the exclamation in [9] stating that the dependence between $X_{i:n}$ and $X_{j:n}$ decreases as i and j draw apart.

To our studies, conditions in Theorem 1.1 are met for almost all continuous populations including the situation discussed in [10] from which we see that the correlation coefficient between a sample maximum and a sample minimum has a limiting value 0 as the sample size n tends to infinity. Here Theorem 1.1 deals with correlation coefficients for common OSs relevant to some general sample quantiles.

Now we use symbol $\lceil z \rceil$ for the integer part of a positive number z and $m_{n,p}$ for the p-quantile of the random sample $(X_1, \cdots, X_n)$. Namely $m_{n,p} = (X_{pn:n} + X_{pn+1:n})/2$ if pn is an integer and $m_{n,p} = X_{\lceil pn+1 \rceil:n}$ otherwise.

Remark 1.1. *Assume the condition (1.1) of Theorem 1.1, then:*

(1) *Corresponding to the central limit theorem for sample quantiles, the following second moment convergence conclusion holds as $n \to \infty$,*

$$E\left(\frac{f(x_p)(m_{n,p} - x_p)}{\sqrt{p(1-p)/n}}\right)^2 \to 1;$$

(2) *The asymptotic correlation coefficient $corr(m_{n,p}, m_{n,r})$ for $m_{n,p}$ and $m_{n,r}$ satisfies*

$$\lim_{n\to\infty} corr(m_{n,p}, m_{n,r}) = \sqrt{\frac{p(1-r)}{r(1-p)}}.$$

Remark 1.2. *For a given sample $(X_1, ..., X_n)$, it is obvious that the correlation coefficient for two different sample quantiles is relevant to the distribution of the population X from which the sample is drawn. Here Theorem 1.1 indicates that, as the sample size n tends to infinity, the mentioned correlation coefficient is eventually free from the distribution of X.*

Corollary 1.1. *Under the condition (1.1) of Theorem 1.1, if we use the sample quantile $m_{n,p}$ as an estimator for the corresponding population quantile x_p, then we have the following variance equivalence*

$$Var(m_{n,p}) \sim \frac{p(1-p)}{n(f(x_p))^2}.$$

Moreover, if we further assume that $F(x_r) = r$ and $F'(x_r) = f(x_r) > 0$ where $0 < p \leq r < 1$, then the covariance between $m_{n,p}$ and $m_{n,r}$ is

$$cov(m_{n,p}, m_{n,r}) = corr(m_{n,p}, m_{n,r}) \sqrt{Var(m_{n,p}) \cdot Var(m_{n,r})} \sim \frac{p(1-r)}{nf(x_p)f(x_r)}. \tag{1.3}$$

Generally, for real numbers $u_1, u_2, ..., u_k$ and $p_1, p_2, ..., p_k$ satisfying $0 < p_1 < ... < p_k < 1$ and $f(x_{p_1})f(x_{p_2})...f(x_{p_k}) > 0$, the following analogous expression holds

$$Var\left(\sum_{i=1}^{k} u_i m_{n,p_i}\right) \sim \sum_{i=1}^{k} \frac{u_i^2 p_i(1-p_i)}{n(f(x_{p_i}))^2} + \sum_{1\leq i<j\leq k} \frac{2u_i u_j p_i(1-p_j)}{nf(x_{p_i})f(x_{p_j})}. \tag{1.4}$$

For a positive integer r, we mean that a random sequence $\{\xi_n, n \geq 1\}$ converges in an r-th order of moment if the number sequence $\{E\xi_n^r, n \geq 1\}$ converges. In reference [11], Wang et al. investigated moment convergence conclusions for some OSs connected to a general continuous population. It is found that not only the sequence of sample quantiles but also the corresponding standardized sequence converges in some positive order of moments even when the population of interest has no expectation. Here Theorem 1.1 is a subsequent exploration.

2. Preparation for the main proof

Lemma 2.1. *(see [2]). Let Y be a population uniformly distributed over the interval* $[0,1]$*; Let* $Y_{i:n}$ *be the corresponding i-th OS of a random sample* $(Y_1, \cdots, Y_n)$ *from Y. For nonnegative integers u, v, i and j satisfying* $1 \leq i < j \leq n$*,*

$$E(Y_{i:n}^u Y_{j:n}^v) = \frac{n!}{(i-1)!} \frac{(u+i-1)!}{(u+j-1)!} \frac{(v+u+j-1)!}{(v+u+n)!}.$$

Remark 2.1. *Setting u, v, i and j to be some specified nonnegative integers, we can gain*

$$EY_{i:n}^u = \frac{n!}{(i-1)!} \frac{(u+i-1)!}{(u+n)!} \tag{2.1}$$

and

$$cov(Y_{i:n}, Y_{j:n}) = \frac{(j+1)i}{(2+n)(1+n)} - \frac{ij}{(n+1)^2} = \frac{i(n-j+1)}{(n+2)(n+1)^2}. \tag{2.2}$$

Lemma 2.2. *Now we denote* $\mu_{k:n} = EY_{k:n} = k/(n+1)$*. Under the conditions of Lemma 2.1 and by Eq (2.1), we can conclude that for integer k satisfying* $k/n \to \rho \in (0,1)$*, as* $n \to \infty$*,*

$$E(Y_{k:n} - \mu_{k:n})^2 \sim \frac{\rho(1-\rho)}{n}; \quad E(Y_{k:n} - \mu_{k:n})^4 \sim \frac{3\rho^2(1-\rho)^2}{n^2} \tag{2.3}$$

and

$$E(Y_{k:n} - \mu_{k:n})^6 \sim \frac{15\rho^3(1-\rho)^3}{n^3}. \tag{2.4}$$

3. Main proof

3.1. The proof of Theorem 1.1

Proof. Denoting $Y = F(X)$ and $Y_i = F(X_i)$, we see that Lemmas 2.1 and 2.2 are applicable.

According to the Taylor expansion formula,

$$Q(t) = Q(t_0) + Q'(t_0)(t-t_0) + \frac{1}{2!}Q''(t_0)(t-t_0)^2 + \frac{1}{3!}Q'''(\Delta)(t-t_0)^3$$

where $\Delta \in [\min(t, t_0), \max(t, t_0)]$, there exists some RVs $\tau_{i:n}$ satisfying

$$\tau_{i:n} \in [min(\mu_{i:n}, Y_{i:n}), max(\mu_{i:n}, Y_{i:n})]$$

such that

$$\begin{aligned} X_{i:n} = Q(Y_{i:n}) &= Q(\mu_{i:n}) + Q'(\mu_{i:n})(Y_{i:n} - \mu_{i:n}) + \frac{Q''(\mu_{i:n})(Y_{i:n} - \mu_{i:n})^2}{2} + \frac{Q'''(\tau_{i:n})(Y_{i:n} - \mu_{i:n})^3}{6} \\ &=: Q(\mu_{i:n}) + part_1 + part_2 + part_3. \end{aligned} \tag{3.1}$$

By the equivalent expressions in Lemma (2.2), we have

$$E(part_1^2) = (Q'(\mu_{i:n}))^2 E(Y_{i:n} - \mu_{i:n})^2 \sim (Q'(p))^2 \frac{p(1-p)}{n} = O(n^{-1}) \tag{3.2}$$

and

$$E(part_2^2) = \frac{1}{4}(Q''(\mu_{i:n}))^2 E(Y_{i:n} - \mu_{i:n})^4 \sim (Q''(p))^2 \frac{3p^2(1-p)^2}{4n^2} = O(n^{-2}). \tag{3.3}$$

Moreover, by the assumption $|Q'''(u)| \leq Ku^{-A}(1-u)^{-A}$, no matter if $0 < Y_{i:n} \leq \tau_{i:n} \leq \mu_{i:n} \leq 1$ or $0 < \mu_{i:n} \leq \tau_{i:n} \leq Y_{i:n} \leq 1$, we have

$$|Q'''(\tau_{i:n})| \leq K\left[{Y_{i:n}}^{-A}(1-\mu_{i:n})^{-A} \cdot {\mu_{i:n}}^{-A}(1-Y_{i:n})^{-A}\right].$$

Noting that the pdf of $Y_{i:n}$ is $n!/((i-1)!(n-i)!)x^{i-1}(1-x)^{n-i}I_{[0,1]}(x)$, we see that

$$\begin{aligned} E[Q'''(\tau_{i:n})^4] &\leq K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A} \cdot E\left[{Y_{i:n}}^{-4A}(1-Y_{i:n})^{-4A}\right]. \\ &= \frac{K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A}n!}{(i-1)!(n-i)!}\int_0^1 x^{i-4A-1}(1-x)^{n-i-4A}dx \\ &= \frac{K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A}n!}{(i-1)!(n-i)!}B(i-4A, n-i-4A+1) \\ &= \frac{K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A}n!}{(i-1)!(n-i)!}\frac{\Gamma(i-4A)\cdot\Gamma(n-i-4A+1)}{\Gamma(n-8A+1)}. \end{aligned} \tag{3.4}$$

Now let M be the nonnegative integer satisfying $u = M - 4A \in [0,1)$. By the formula (see [12]) $\Gamma(n+\alpha) \sim n^\alpha(n-1)!$ where $\alpha > 0$, we have for $i/n \to p \in (0,1)$ as $n \to \infty$,

$$\begin{aligned} E[Q'''(\tau_{i:n})^4] &\leq \frac{K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A}n!}{(i-1)!(n-i)!}\frac{\Gamma(i-M+u)\cdot\Gamma(n+1-i-M+u)}{\Gamma(n+1-2M+2u)} \\ &\sim \frac{K^4(1-\mu_{i:n})^{-4A}{\mu_{i:n}}^{-4A}n!}{(i-1)!(n-i)!}\frac{(i-M)^u(i-M-1)!(n+1-i-M)^u(n-i-M)!}{(n+1-2M)^{2u}(n-2M)!} \\ &\sim \frac{K^4(1-p)^{-4A}p^{-4A}n!}{(i-1)!(n-i)!}\frac{(i-M-1)!\cdot(n-i-M)!}{(n-2M)!}\frac{(i-M)^u(n+1-i-M)^u}{(n+1-2M)^{2u}} \\ &\sim \frac{K^4(1-p)^{-4A}p^{-4A}n!}{(i-1)!(n-i)!}\frac{(i-M-1)!\cdot(n-i-M)!}{(n-2M)!}p^u(1-p)^u \\ &\sim \frac{K^4(1-p)^{u-4A}p^{u-4A}n!}{(i-1)!(n-i)!}\frac{(i-M-1)!\cdot(n-i-M)!}{(n-2M)!}. \end{aligned} \tag{3.5}$$

Furthermore, we can utilize the Stirling formula $m! \sim (m/e)^m\sqrt{2\pi m}$ to obtain

$$\frac{K^4(1-p)^{u-4A}p^{u-4A}n!}{(i-1)!(n-i)!}\frac{(i-M-1)!\cdot(n-i-M)!}{(n-2M)!}$$

$$= K^4(1-p)^{u-4A}p^{u-4A}\frac{n(n-1)...(n-(2M-1))}{(i-1)...(i-M)(n-i)...(n-i-(M-1))} \tag{3.6}$$

$$\sim K^4(1-p)^{u-4A}p^{u-4A}\frac{n^{2M}}{i^M(n-i)^M} \tag{3.7}$$

$$\to K^4(1-p)^{u-4A}p^{u-4A}(p(1-p))^{-M} = K^4(1-p)^{-8A}p^{-8A}. \tag{3.8}$$

According to (3.5) and (3.8) and by using Liapunov's inequality in the form $E(\xi^2) \le [E(\xi^4)]^{1/2}$, we see that there exists a positive constant $R > 0$ such that the inequality

$$E[Q'''(\tau_{i:n})^4] \le R \tag{3.9}$$

holds uniformly with respect to $n \ge 1$. By the Cauchy-Schwarz inequality $[E(\xi\eta)]^2 \le E\xi^2 \cdot E\eta^2$ and Lemma 2.2 as well as the fact $|Y_{i:n} - \mu_{i:n}| \le 1$, we see that

$$\begin{aligned} E(part_3^2) &= \frac{1}{36}E\left\{[Q'''(\tau_{i:n})]^2(Y_{i:n}-\mu_{i:n})^6\right\} \le \frac{1}{36}E\left\{[Q'''(\tau_{i:n})]^2|Y_{i:n}-\mu_{i:n}|^3\right\} \\ &\le \frac{1}{36}\sqrt{E[Q'''(\tau_{i:n})]^4E(Y_{i:n}-\mu_{i:n})^6} = O(n^{-3/2}). \end{aligned} \tag{3.10}$$

Similarly

$$|E(part_3)| = \frac{1}{6}|EQ'''(\tau_{i:n})(Y_{i:n}-\mu_{i:n})^3| \le \frac{1}{6}\sqrt{E[Q'''(\tau_{i:n})]^2 \cdot E(Y_{i:n}-\mu_{i:n})^6} \tag{3.11}$$

$$\le \frac{1}{6}\sqrt{\sqrt{R} \cdot E(Y_{i:n}-\mu_{i:n})^6} = O(n^{-3/2}). \tag{3.12}$$

Combining the conclusions (3.1) and (3.11), we get

$$EX_{i:n} = EQ(Y_{i:n}) = Q(\mu_{i:n}) + \frac{Q''(\mu_{i:n})}{2}Var(Y_{i:n}) + o(n^{-1}). \tag{3.13}$$

Similarly to (3.1), there exists some RV $\alpha_{j:n} \in [min(\mu_{j:n}, Y_{j:n}), max(\mu_{j:n}, Y_{j:n})]$ such that

$$\begin{aligned} X_{j:n} = Q(Y_{j:n}) &= Q\left(\mu_{j:n}\right) + Q'(\mu_{j:n})(Y_{j:n}-\mu_{j:n}) + \frac{Q''(\mu_{j:n})(Y_{j:n}-\mu_{j:n})^2}{2} + \frac{Q'''(\alpha_{j:n})(Y_{j:n}-\mu_{j:n})^3}{6} \\ &= : Q\left(\mu_{j:n}\right) + PART_1 + PART_2 + PART_3. \end{aligned} \tag{3.14}$$

Replacing i with j in (3.2), (3.3), (3.11) and (3.13) yields

$$E(PART_1^2) \sim (Q'(p))^2\frac{r(1-r)}{n} = O(n^{-1}), \tag{3.15}$$

$$E(PART_2^2) \sim (Q''(p))^2\frac{3r^2(1-r)^2}{4n^2} = O(n^{-2}), \tag{3.16}$$

$$E(PART_3) = O(n^{-3/2}), \quad E(PART_3^2) = O(n^{-3/2}) \tag{3.17}$$

and

$$EX_{j:n} = EQ(Y_{j:n}) = Q\left(\mu_{j:n}\right) + \frac{Q''\left(\mu_{j:n}\right)}{2}Var(Y_{j:n}) + o(n^{-1}). \tag{3.18}$$

Moreover, we see that

$$|cov\,(part_3, PART_3)\,| \le \sqrt{E(part_3^2)} \cdot \sqrt{E(PART_3^2)} = O(n^{-3/2}),$$

hence

$$cov(part_3, PART_3) = o(n^{-1}). \tag{3.19}$$

That results in the following conclusion according to Eqs (3.1) and (3.14):

$$\begin{aligned} cov(X_{i:n}, X_{j:n}) &= \sum_{1\le s\le 3; 1\le t\le 3} cov(part_s, PART_t) \\ &= \sum_{2\le t\le 3} cov(part_1, PART_t) + \sum_{1\le t\le 3} cov(part_2, PART_t) \\ &\quad + \sum_{1\le t\le 2} cov(part_3, PART_t) + cov(part_1, PART_1) + o(n^{-1}). \end{aligned} \tag{3.20}$$

Now noting the equations numbered (3.2), (3.3), (3.11) and those from (3.15) to (3.17), we derive

$$\begin{aligned} &\left|\sum_{2\le t\le 3} cov(part_1, PART_t) + \sum_{1\le t\le 3} cov(part_2, PART_t) + \sum_{1\le t\le 2} cov(part_3, PART_t)\right| \\ &\le \sum_{2\le t\le 3} \sqrt{E(part_1^2)\cdot E(PART_t^2)} + \sum_{1\le t\le 3} \sqrt{E(part_2^2)\cdot E(PART_t^2)} \\ &+ \sum_{1\le t\le 2} \sqrt{E(part_3^2)\cdot E(PART_t^2)} = o(n^{-1}), \end{aligned} \tag{3.21}$$

from which we conclude that

$$\sum_{2\le t\le 3} cov(part_1, PART_t) + \sum_{1\le t\le 3} cov(part_2, PART_t) + \sum_{1\le t\le 2} cov(part_3, PART_t) = o(n^{-1}).$$

Substituting the corresponding part in (3.20) by the just obtained above result, we have

$$cov(X_{i:n}, X_{j:n}) = cov(part_1, PART_1) + o(n^{-1}) = Q'(\mu_{i:n})Q'(\mu_{j:n})\frac{i(n-j+1)}{(n+2)(n+1)^2} + o(n^{-1}). \tag{3.22}$$

Referring to the procedure in obtaining conclusion (3.22), we can also reach the following conclusions

$$Var(X_{i:n}) = Var(part_1) + o(n^{-1}) = [Q^{'}(\mu_{i:n})]^2\frac{i(n+1-i)}{(n+2)(n+1)^2} + o(n^{-1}) \tag{3.23}$$

and

$$Var(X_{j:n}) = Var(PART_1) + o(n^{-1}) = [Q'(\mu_{j:n})]^2\frac{j(n-j+1)}{(n+2)(n+1)^2} + o(n^{-1}). \tag{3.24}$$

(1) Now we notice that as $n \to \infty$,

$$Q(\mu_{i:n}) = Q\left(\frac{i}{n+1}\right) = Q(p) + Q^{'}(p)\left(\frac{i}{n+1} - p\right) + o\left((\frac{i}{n+1} - p)\right),$$

therefore according to Eq (3.13), we have

$$\lim_{n\to\infty}\left(\frac{f(x_p)(EX_{i:n}-x_p)}{\sqrt{p(1-p)/n}}\right)^2=\lim_{n\to\infty}\left(\frac{f(x_p)[Q(\mu_{i:n})+\frac{Q''(\mu_{i:n})}{2}Var(Y_{i:n})+o(n^{-1})-x_p]}{\sqrt{p(1-p)/n}}\right)^2$$
$$=\lim_{n\to\infty}\left(\frac{f(x_p)[\frac{Q''(\mu_{i:n})}{2}Var(Y_{i:n})+o(n^{-1})+Q(\mu_{i:n})-Q(p)]}{\sqrt{p(1-p)/n}}\right)^2$$
$$=\lim_{n\to\infty}\left(\frac{f(x_p)[\frac{Q''(\mu_{i:n})}{2}Var(Y_{i:n})+o(n^{-1})+Q^{'}(p)(\frac{i}{n+1}-p)+o((\frac{i}{n+1}-p))]}{\sqrt{p(1-p)/n}}\right)^2=0$$

provided $i/n=p+o(n^{-1/2})$ which is equivalent to $i/(n+1)=p+o(n^{-1/2})$. Consequently we see that

$$\begin{aligned}\lim_{n\to\infty}E\left(\frac{f(x_p)(X_{i:n}-x_p)}{\sqrt{p(1-p)/n}}\right)^2 &= \lim_{n\to\infty}E\left(\frac{f(x_p)[(X_{i:n}-EX_{i:n})+(EX_{i:n}-x_p)]}{\sqrt{p(1-p)/n}}\right)^2\\
&= \lim_{n\to\infty}\left[E\left(\frac{f(x_p)(X_{i:n}-EX_{i:n})}{\sqrt{p(1-p)/n}}\right)^2+\left(\frac{f(x_p)(EX_{i:n}-x_p)}{\sqrt{p(1-p)/n}}\right)^2\right]\\
&= \lim_{n\to\infty}E\left(\frac{f(x_p)(X_{i:n}-EX_{i:n})}{\sqrt{p(1-p)/n}}\right)^2=\lim_{n\to\infty}\frac{(f(x_p))^2\cdot Var(X_{i:n})}{p(1-p)/n}\\
&= \lim_{n\to\infty}(f(x_p))^2\frac{[Q^{'}(\mu_{i:n})]^2\frac{i(n+1-i)}{(n+2)(n+1)^2}+o(n^{-1})}{p(1-p)/n}=1\end{aligned}$$

according to Eq (3.23). Here the reason for the last equation is that the continuous function $Q'(u)$ is positive according to the deduction $Q'(u)=1/F'(x)=1/f(x)>0$ at $x=x_p$.

(2) Combining the three conclusions (3.22)–(3.24), we get the asymptotic correlation coefficient $corr(X_{i:n},X_{j:n})$ by the following procedures:

$$\begin{aligned}corr(X_{i:n},X_{j:n}) &= \frac{cov(X_{i:n},X_{j:n})}{\sqrt{Var(X_{i:n})}\sqrt{Var(X_{j:n})}}\\
&= \frac{Q'(\mu_{i:n})Q'(\mu_{j:n})\frac{i(n-j+1)}{(n+2)(n+1)^2}+o(n^{-1})}{\sqrt{[Q'(\mu_{i:n})]^2\frac{i(n-i+1)}{(n+2)(n+1)^2}+o(n^{-1})}\sqrt{[Q'(\mu_{j:n})]^2\frac{j(n-j+1)}{(n+2)(n+1)^2}+o(n^{-1})}}\\
&= \frac{Q'(\mu_{i:n})Q'(\mu_{j:n})\frac{i(n-j+1)}{(n+2)(n+1)^2}+o(n^{-1})}{\left|Q'(\mu_{i:n})Q'(\mu_{j:n})\right|\sqrt{\frac{i(n-i+1)}{(n+2)(n+1)^2}+o(n^{-1})}\sqrt{\frac{j(n-j+1)}{(n+2)(n+1)^2}+o(n^{-1})}}\\
&\overset{n\to\infty}{\to} \frac{Q'(p)Q'(r)p(1-r)}{|Q'(p)Q'(r)|\sqrt{p(1-p)}\sqrt{r(1-r)}}=\sqrt{\frac{p(1-r)}{r(1-p)}}\end{aligned}\tag{3.25}$$

provided $i/n=p+o(1)$ and $j/n=r+o(1)$. □

4. Examples

To continue our discussions, we give the following two propositions beforehand:

Proposition 1. *If the inverse function $Q(u)$ of a cdf $F(x)$ has a third-order derivative $Q'''(u)$, then*

$$Q'''(u) = \frac{-f''(x)f(x) + 3(f'(x))^2}{(f(x))^5}, \tag{4.1}$$

where $x = Q(u)$ and $f(x) = F'(x)$.

Proposition 2. *For a function $-\ln(u^A(1-u)^A)$ (where $A \geq 0$ is a constant) and any specified constant $\varepsilon > 0$, there exists a corresponding number $C(\varepsilon) > 0$ such that the inequality $-\ln(u^A(1-u)^A) \leq C(\varepsilon)(u(1-u))^{-\varepsilon}$ holds for all $u \in (0, 1)$.*

4.1. One example of Theorem 1.1

Although almost all commonly applied continuous types of populations satisfy the conditions in Theorem 1.1, due to length concerns in this section, we will present only one example of Theorem 1.1.

Example. For a population X with a gamma distribution (including special cases such as the Exponential as well as the Chi-square distributions), the pdf is

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{[0,\infty)}(x), \alpha > 0, \beta > 0.$$

Corresponding to the case $\omega = \beta > 0$ and $\nu = 0$, we now assume $x = Q(u)$ to be the inverse function of the cdf $u = F(x)$ of βX, the pdf of which can be worked out as

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} I_{[0,\infty)}(x).$$

On that occasion, we can easily see that for $x > 0$,

$$f'(x) = \frac{\alpha - 1 - x}{x} f(x) \text{ and } f''(x) = \frac{1 - \alpha + (1 - \alpha + x)^2}{x^2} f(x). \tag{4.2}$$

Noting that it is easy to verify that Theorem 1.1 is applicable in the case $\alpha = 1$, we now assume that $\alpha \in (0, 1) \cup (1, +\infty)$.

As a condition like (1.1) is equivalent to verifying the existence of a positive number q such that

$$\lim_{u \to 0+} u^q Q'''(u) = \lim_{u \to 1-} (1-u)^q Q'''(u) = 0$$

and by Proposition 1, it is sufficient to verifying the conditions

$$\lim_{x \to 0} (F(x))^q f''(x) f(x)/f^5(x) = \lim_{x \to 0} (F(x))^q (f'(x))^2/f^5(x) = 0 \tag{4.3}$$

and

$$\lim_{x \to \infty} (1 - F(x))^q f''(x) f(x)/f^5(x) = \lim_{x \to \infty} (1 - F(x))^q (f'(x))^2/f^5(x) = 0. \tag{4.4}$$

Now we use the notation $g(x) \asymp h(x)$ to mean that there are positive constants $a < b$ such that $a|g(x)| \leq h(x) \leq b|g(x)|$ as $x \to 0+$ or $x \to \infty$, according to context. Then

- **Case 1.** $x \to 0+$. We only need consider x confined in a sufficiently small interval $(0, \delta]$. Obviously $F(x) \asymp x^{\alpha}$, $f(x) \asymp x^{\alpha-1}$, $f'(x) \asymp x^{\alpha-2}$ and $f''(x) \asymp x^{\alpha-3}$. It follows that conditions (4.3) are satisfied if a positive constant q satisfies $q > 3 - 1/\alpha$.
- **Case 2.** $x \to \infty$. On that occasion, we see according to (4.2) that the following relations hold simultaneously for a positive number $q > 3$:

$$(1 - F(x))^q f''(x) f(x)/f^5(x) \sim (1 - F(x))^q/f^3(x) \sim (1 - F(x))^{q-3} \to 0$$

and

$$(1 - F(x))^q (f'(x))^2/f^5(x) \sim (1 - F(x))^q/f^3(x) \sim (1 - F(x))^{q-3} \to 0.$$

Consequently we see the realization of (4.4).

As above analyzed, gamma distributions satisfy condition (1.1).

4.2. One application example

The Cauchy distribution has a wide range of applications in physics, economics as well as in the medical domain. We may perceive its important application in physics by a simple model depicted as what follows: in a coordinate plane, if we place at a point (θ_1, θ_2) (where $\theta_2 > 0$) a radioactive material emitting a particle at a random angle U uniformly distributed over an interval $[0, 2\pi]$, then we can show that the particle will reach the abscissa axis at a point X distributed according to a pdf

$$f(x, \theta_1, \theta_2) = \frac{\theta_2}{\pi[\theta_2^2 + (x - \theta_1)^2]}, -\infty < x < +\infty \tag{4.5}$$

which is the pdf of a Cauchy distribution. The relevant kinds of literature are huge. For general introduction we recommend [13] whereas, for some elegant studies on a similar topic to this article, we consult references [14] and [15].

There also is a considerable literature on L-estimation, including determining optimal weights. Some of this is in the robustness literature. See [2] and [16] for more references.

On estimating the location θ_1 in (4.5), Sen verified in [17] that the so-called mid-range $(m_{n,0.56} + m_{n,0.44})/2$ is more effective than the sample median $m_{n,0.5}$. By rejecting a fixed number of the largest and the smallest OSs to avoid a large mean squared error of the parameter estimator, Pekasiewicz utilized in [18] a method named the truncated quantile least squares method to estimate the location parameter θ_1. Recently, Krykun [19] investigated estimating both θ_1 and θ_2, by resorting to an arctangent regression function and rejecting some fraction of the largest and the smallest OSs. Some ideal simulated results are obtained in [19]. Comparatively, what we present in the following exploration is a third way using optimal linear combinations of some sample quantiles.

To estimate θ_1 in (4.5), we will, without loss of generality, set $\theta_2 = 1$.

Let $(X_1, ..., X_n)$ be a random sample from a population X according to the pdf

$$f(x, \theta) = \frac{1}{\pi((x - \theta)^2 + 1)} \tag{4.6}$$

with an unknown θ. As finding the uniformly minimum variance unbiased estimator (UMVUE) for θ is hopeless, we now think of the estimator

$$R_n(p) := \frac{m_{n,p} + m_{n,1-p}}{2} \tag{4.7}$$

which is named as sample quasi-midrange (see [20]). It is trivial to see that $R_n(r)$ is unbiased in estimating θ. According to Theorem 1.1, we see that

$$\begin{aligned} Var(R_n(p)) &= \frac{Var(m_{n,p}) + Var(m_{n,1-p}) + 2corr(m_{n,p}, m_{n,1-p})\sqrt{Var(m_{n,p}) \cdot Var(m_{n,1-p})}}{4} \\ &= \frac{1}{2(1-p)}Var(m_{n,p}) \sim \frac{1}{2(1-p)}\frac{p(1-p)}{n(f(x_p))^2} = \frac{p}{2n(f(x_p))^2} \\ &= \frac{p\pi^2(1+x_p^2)^2}{2n} = \frac{p\pi^2\left(1+\frac{cos^2(\pi p)}{sin^2(\pi p)}\right)^2}{2n} = \frac{p\pi^2}{2n \cdot sin^4(\pi p)}. \end{aligned} \tag{4.8}$$

As we can easily see that the equivalence for the variance of the sample median

$$Var(m_{n,0.5}) \sim \frac{\pi^2}{4n} \approx \frac{2.467401016}{n},$$

the result of (4.8) seems to indicate that the unbiased estimator $R_n(r)$ will be more effective than the sample median $m_{n,0.5}$ if we can diminish the value $r/sin^4(\pi r)$. As the minimum value of $r/sin^4(\pi r)$ exists but can not be obtained as an explicit expression, here we make an approximation of the minimum value of $r/sin^4(\pi r)$ as 0.4724417292 when $r = 0.4435$. By the equivalence (1.3) in Corollary 1.1, the estimator $R_n(0.4435)$ is preferable for θ because the equivalent corresponding variance

$$Var\left(\frac{m_{n,0.4435} + m_{n,1-0.4435}}{2}\right) \sim \frac{0.4435\pi^2}{2n \cdot sin^4(0.4435\pi)} = \frac{2.332}{n}$$

is a bit smaller than that of the sample median. That is exactly the conclusion drawn in [17]. Moreover, for $0 < p < r \leq 0.5$ and $t \in (-\infty, +\infty)$, we see that the estimator $tR_n(p) + (1-t)R_n(r)$ is also unbiased for θ and

$$\begin{aligned} &Var(tR_n(p) + (1-t)R_n(r)) = t^2Var(R_n(p)) + (1-t)^2Var(R_n(r)) + 2t(1-t)cov(R_n(p), R_n(r)) \\ &\sim \frac{t^2p\pi^2}{2n \cdot sin^4(\pi p)} + \frac{(1-t)^2r\pi^2}{2n \cdot sin^4(\pi r)} + \frac{t(1-t)}{2}cov(m_{n,p} + m_{n,1-p}, m_{n,r} + m_{n,1-r}); \end{aligned}$$

According to equivalence (1.4) and by noting that $f(x_p) = f(x_{1-p})$ and $f(x_r) = f(x_{1-r})$, we obtain

$$cov(m_{n,p} + m_{n,1-p}, m_{n,r} + m_{n,1-r}) \sim \frac{2p\pi^2}{nsin^2(\pi p)sin^2(\pi r)} \tag{4.9}$$

and thus for large n,

$$\begin{aligned} Var(tR_n(p) + (1-t)R_n(r)) &\sim \frac{t^2p\pi^2}{2n \cdot sin^4(\pi p)} + \frac{(1-t)^2r\pi^2}{2n \cdot sin^4(\pi r)} + \frac{t(1-t)p\pi^2}{nsin^2(\pi p)sin^2(\pi r)} \\ &= \frac{\pi^2}{2n}\left(\frac{t^2p}{sin^4(\pi p)} + \frac{(1-t)^2r}{sin^4(\pi r)} + \frac{2t(1-t)p}{sin^2(\pi p)sin^2(\pi r)}\right). \end{aligned}$$

Generally, for two sequences of real numbers $t_1, ..., t_m$ and $p_1, ..., p_m$ respectively satisfying $t_m = 1 - \sum_{i=1}^{m-1} t_i$ and $0 \leq p_1 < p_2 < ... < p_m \leq 0.5$, the linear combination $\sum_{i=1}^{m} t_iR_n(p_i)$ is an unbiased estimator for θ and the corresponding asymptotic variance is

$$Var\left(\sum_{i=1}^{m} t_iR_n(p_i)\right) = \sum_{i=1}^{m} Var(t_iR_n(p_i)) + 2\sum_{1\leq i<j\leq m} t_it_jcov(R_n(p_i), R_n(p_j))$$

$$\sim \quad \frac{\pi^2}{2n}\left[\sum_{i=1}^{m}\frac{t_i^2 p_i}{sin^4(\pi p_i)}+\sum_{1\le i<j\le m}\frac{2t_it_jp_i}{sin^2(\pi p_i)sin^2(\pi p_j)}\right]. \tag{4.10}$$

For the unknown θ in the pdf of (4.6), to find an unbiased estimator of the form $\widehat{\theta}_{m,n} = \sum_{i=1}^{m} t_i R_n(p_i)$ with minimum variance, what is left is just a matter of some calculations of finding the t_i's and p_i's such that the expression (4.10) attains its minimum value. For instance, by putting $m = 5$ in (4.10) and by some numerical calculations, we obtain such an estimator defined by

$$\begin{aligned} E_{5,n} = &-0.0192R_n(0.0632) - 0.0747R_n(0.1347) + 0.2953R_n(0.3577) \\ &+ 0.3799R_n(0.4199) + 0.4187R_n(0.4739). \end{aligned}$$

With the aid of Matlab software, the asymptotic variance can be shown to be $Var(E_{5,n}) \sim 2.0314/n$. The estimator $\widehat{\theta}_{5,n}$ is unbiased and is better than the estimator $R_n(0.4435)$, which was named the optimum mid-range estimator and was admitted in [17] as a superior estimator to the sample median in estimating θ. As $p_1 = 0.4435$ can be determined numerically for the case $m = 1$, among unbiased estimators $R_n(p)$ in (4.7), $R_n(0.4435)$ is the most efficient one such that (4.10) has a minimum variance when $m = 1$ is specified.

The Fisher information $I(\theta) = 1/2$ for the Cauchy pdf (4.6), so we see that even if the UMVUE, say $\widehat{\theta}_n^*$ for θ exists, the theoretical variance $Var(\widehat{\theta}_n^*)$ can not be smaller than $\frac{2}{n}$ according to the well-known Cramér-Rao inequality.

Noting that the quotient $\frac{2}{2.0314} \approx 0.9845$ is close to 1, we see that the quick unbiased estimator $E_{5,n}$ is close to the theoretical ideal unbiased estimator.

To compare the effectiveness of estimating θ_1 by the three mentioned estimators, namely, the median $m_{n,0.5}$, the quasi-midrange $R_n(0.4435)$ in (4.7) and the just discussed estimator $E_{5,n}$, by the aid of Matlab software, we simulate 30 times a random sample of size $n = 200$ drawn from a specified Cauchy distribution $f(x, \theta_1, \theta_2) = \frac{\theta_2}{\pi[\theta_2^2+(x-\theta_1)^2]}$ with respective true values $\theta_1 = 0.75$ and $\theta_2 = 2$. According to the simulated results, Figure 1 shows the effectiveness of the three estimators in estimating θ_1. The averaged squared errors for the three estimators are respectively, 0.0030, 0.0025 and 0.0019.

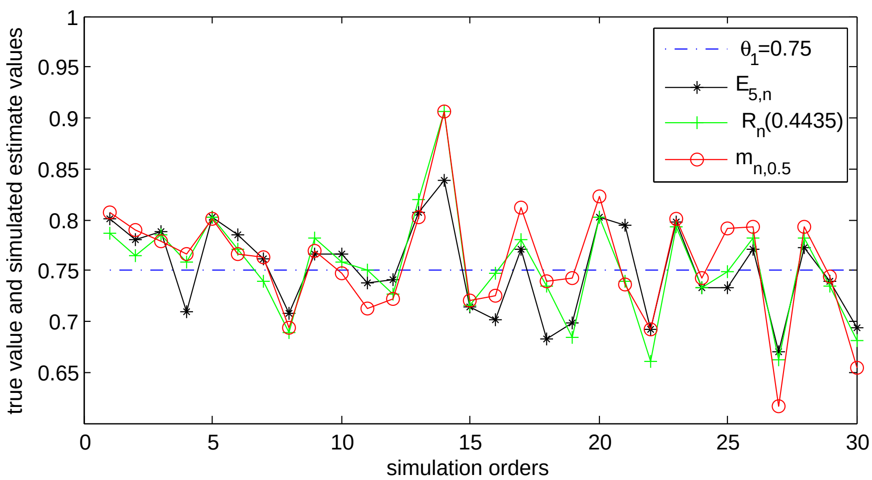

Figure 1. Comparing estimators $m_{n,0.5}$, $E_{5,n}$ and $R_n(0.4435)(n = 200)$ in estimating $\theta_1 = 0.75$.

As is indicated by the simulated results, among the three estimators $m_{n,0.5}$, $E_{5,n}$ and $R_n(0.4435)$, the estimator $E_{5,n}$ is the most effective under the assumption of a large sample size.

Acknowledgments

This work was supported by the Science and Technology Plan Project of Jiangxi Province Health Commission (Grant NO. 202311165) and the National Natural Science Foundation of China (Grant NO. 81960618).

Conflict of interest

There exists no conflict of interest between authors.

References

1. J. Shao, *Mathematical statistics*, New York: Springer, 2010.
2. H. David, H. Nagaraja, *Order statistics*, New Jersey: Wiley, 2003. http://dx.doi.org/10.1002/0471722162
3. H. Barakat, Comments on the rate of convergence to asymptotic independence between order statistics, *Stat. Probabil. Lett.*, **76** (2006), 35–38. http://dx.doi.org/10.1016/j.spl.2005.07.009
4. H. Barakat, Measuring the asymptotic dependence between generalized order statistics, *J. Stat. Theory Appl.*, **6** (2007), 106–117.
5. H. Barakat, A nonparametric general criterion of asymptotic dependence between order statistics, *Commun. Stat.-Theor. M.*, **38** (2009), 1960–1968. http://dx.doi.org/10.1080/03610920802689385
6. W. Hürlimann, On the rate of convergence to asymptotic independence between order statistics, *Stat. Probabil. Lett.*, **66** (2004), 355–362. http://dx.doi.org/10.1016/j.spl.2003.10.020
7. R. Bahadur, A note on quantiles in large samples, *Ann. Math. Statist.*, **37** (1966), 577–580. http://dx.doi.org/10.1214/aoms/1177699450
8. A. Dasgupta, *Asymptotic theory of statistics and probability*, New York: Springer, 2008. http://dx.doi.org/10.1007/978-0-387-75971-5
9. J. Averousa, C. Genest, S. Kochar, On the dependence structure of order statistics, *J. Multivariate Anal.*, **94** (2005), 159–171. http://dx.doi.org/10.1016/j.jmva.2004.03.004
10. J. Wang, C. Deng, J. Li, M. Zhou, On variances and covariances of a kind of extreme order statistics, *Commun. Stat.-Theor. M.*, **45** (2016), 3274–3282. http://dx.doi.org/10.1080/03610926.2014.901373
11. J. Wang, C. Deng, J. Li, On moment convergence for some order statistics, *AIMS Mathematics*, **7** (2022), 17061–17079. http://dx.doi.org/10.3934/math.2022938
12. V. Zorich, *Mathematical analysis II*, Berlin: Springer, 2016. http://dx.doi.org/10.1007/978-3-662-48993-2
13. S. Kotz, N. Balakrishnan, N. Johnson, *Continuous multivariate distributions: models and applications*, New York: John Wiley & Sons, 2000. http://dx.doi.org/10.1002/0471722065

14. Z. Chen, A simple method for estimating parameters of the location-scale distribution family, *J. Stat. Comput. Sim.*, **81** (2011), 49–58. http://dx.doi.org/10.1080/00949650903177497

15. O. Kravchuk, P. Pollett, Hodges-Lehmann scale estimator for Cauchy distribution, *Commun. Stat.-Theor. M.*, **41** (2012), 3621–3632, http://dx.doi.org/10.1080/03610926.2011.563016

16. B. Arnold, N. Balakrishnan, H. Nagaraja, *A first course in order statistics*, New York: SIAM, 1992. http://dx.doi.org/10.1137/1.9780898719062

17. P. Sen, On some properties of the asymptotic variance of the sample quantiles and mid-ranges, *J. R. Stat. Soc. B*, **23** (1961), 453–459. http://dx.doi.org/10.1111/j.2517-6161.1961.tb00428.x

18. D. Pekasiewicz, Application of quantile methods to estimation of Cauchy distribution parameters, *Statistics in Transition*, **15** (2014), 133–144.

19. I. Krykun, The arctangent regression and the estimation of parameters of the Cauchy distribution, *J Math. Sci.*, **249** (2020), 739–753. http://dx.doi.org/10.1007/s10958-020-04970-3

20. K. Raghunandanan, R. Srinivasan, Simplified estimation of parameters in a logistic distribution.*Biometrika*, **57** (1970), 677–679. http://dx.doi.org/10.1093/biomet/57.3.677