Canad. Math. Bull. Vol. **16** (2), 1973

SOME COMMENTS ON QUANTILES AND ORDER STATISTICS

BY
P. V. RAMACHANDRAMURTY AND M. SUDHAKARA RAO

0. **Summary.** A new concept—that of pseudoconsistency—which seems to be particularly appropriate for the estimation of a quantile is introduced. It is shown *without any conditions whatsoever on the underlying distribution* that the sample quantile is strongly pseudoconsistent for the corresponding population quantile. The asymptotic distribution of the sample quantiles and order statistics is derived when the underlying distribution is discrete.

1. **Introduction.** It is now a classical result that the $[np]$-th order statistic as well as the p-th quantile of the empirical distribution function $(0<p<1)$ is a consistent estimator of the p-th quantile of the population if it is unique (see, for example, [7, pp. 354–355]). Recently Feldman and Tucker [3] investigated the case of nonunique quantiles and proposed a sequence of estimators which are strongly consistent for the smallest and largest p-th quantiles of the population. These estimators seem to be artificial. We take a different point of view and show that the sample quantiles which are natural estimators of the corresponding population quantiles are strongly pseudoconsistent, no matter whether the population quantile is unique or not. The classical result mentioned above is a special case of our result.

Regarding the asymptotic distribution of the order statistics there is a large volume of literature (see, for example [2, pp. 199–224]). But all these assume many regularity conditions and they do not cover the simple and perhaps the most interesting case when the underlying distribution is discrete. So far as the authors are aware this case is touched upon only in [4] by Kabe. But Kabe confines himself to some particular discrete distributions and gives an example to show that the asymptotic distribution need not be the same as in the continuous case. In this paper we show that the asymptotic distribution of sample quantile is degenerate (for all norming constants) when it exists as a distribution function, with the sole assumption that the underlying distribution is discrete.

2. **Pseudo consistency.** Suppose (Ω, B, P) is a probability space and X_n is a sequence of i.i.d. r.v.'s on it with d.f. $F(x)=P(X_1\leq x)$. Let $X_{n,1}\leq X_{n,2}\leq\cdots\leq X_{n,n}$

Received by the editors June 11, 1971.

be the order statistics based on $X_1, X_2, \ldots, X_n$. Let $0<p<1$ and $S_p=\{x:F(x-0)\leq p, F(x)\geq p\}$. Let $a_p=\text{Inf}\{x:x\in S_p\}$, $b_p=\sup\{x:x\in S_p\}$. Any point in $[a_p, b_p]$ has as much claim as any other point in $[a_p, b_p]$ to be called p-th quantile. Hence we shall not try to fix a point in $[a_p, b_p]$ and call it the p-th quantile. Nevertheless we shall use the phrase "the p-th quantile" although strictly speaking we should say a p-th quantile. Let $F_n(x;\omega)$ be the empirical d.f. of $X_1, X_2, \ldots, X_n$ defined by

$$F_n(x;\omega) = \frac{1}{n}\{\# X_i(\omega)\leq x, i=1,2,\ldots,n\}.$$

Then the Glivenko-Cantelli theorem states:

$$P\{\omega:\sup_x |F_n(x;\omega)-F(x)|\to 0\} = 1.$$

Let $\delta_n(X_1, X_2, \ldots, X_n)$ be any sequence of measurable functions.

DEFINITIONS. δ_n is said to be *strongly pseudo consistent* for the p-th quantile of the population if $P(a_p\leq \liminf_n \delta_n \leq \limsup_n \delta_n\leq b_p)=1$. If $\forall\epsilon>0$, $P(a_p-\epsilon\leq\delta_n\leq b_p+\epsilon)\to 1$ as $n\to\infty$ then it is said to be *weakly pseudoconsistent*.

Note that if $a_p=b_p$ strong (weak) pseudoconsistency reduces to strong (weak) consistency for the unique quantile a_p.

LEMMA 1. (a) $F(x)>p\Rightarrow x\geq b_p$, $F(x)<p\Rightarrow x\leq a_p$, $F(x)=p\Rightarrow a_p\leq x\leq b_p$.

(b) *If* $d_n\to d$ *then* $F(d-0)\leq \liminf F(d_n)\leq \limsup F(d_n)\leq F(d)$.

(c) *For any sequence* c_n *for which* $F(c_n)\to p$ $a_p\leq \liminf c_n\leq \limsup c_n\leq b_p$.

Proof. (a) is trivial from the definitions of a_p and b_p and the properties of d.f.'s.

(b) follows from the observation that the limit of every convergent subsequence is either $F(d-0)$ or $F(d)$. Q.E.D.

(c) Suppose $c_{n_k}\to c$ and $F(c_{n_k})\to p$. Then from (b) $F(c-0)\leq p\leq F(c)$ and hence $a_p\leq c\leq b_p$. Thus limits of all convergent subsequences of c_n lie between a_p and b_p and the result follows.

We shall now prove the following main theorem.

THEOREM 1. $X_{n,[np]+k}$ *is strongly pseudoconsistent for the p-th quantile of the population for any fixed integer K (independent of n).*

Proof. Let

$$T = \left\{\omega:\sup_x |F_n(x,\omega)-F(x)|\to 0\right\}.$$

For all $\omega\in\Omega$, $\epsilon>0$, $\exists N_1, N_2$ (independent of ω) such that

$$n > N_1 \Rightarrow |F_n(X_{n,[np]+k})-p| < \frac{\epsilon}{2},$$

https://doi.org/10.4153/CMB-1973-048-1 Published online by Cambridge University Press

and

$$\omega \in T, n > N_2 \Rightarrow |F(X_{n,[np]+k}) - F_n(X_{n,[np]+k})| < \frac{\epsilon}{2}.$$

Thus

$$\omega \in T, n > \max(N_1, N_2) \Rightarrow |F(X_{n,[np]+k}) - p| < \epsilon \cdot ie.$$

For $\omega \in T$, $F(X_{n[n,p]+k}) \to p$. It follows from Lemma 1 (c) that $\omega \in T \Rightarrow a_p \leq \liminf X_{n,[np]+k} \leq \limsup X_{n,[np]+k} \leq b_p$. Since from the Glivenko-Cantelli theorem $P(T^c)=0$, the result follows. It may be mentioned that the above result is true if $X_{n,[np]+k}$ is replaced by X_{n,k_n} where $k_n/n \to p$. Q.E.D.

Considerations of Fisherian consistency as interpreted by Rao (see [7, pp. 281–282]) and statistical practice suggest the following estimators for the p-th quantile of the population. Let $S_{n,p} = \{x : F_n(x-0) \leq p, F_n(x) \geq p\}$, where for convenience we suppress ω. Let

$$a_{n,p} = \text{Inf}\{x : x \in S_{n,p}\}$$

$$b_{n,p} = \text{Sup}\{x : x \in S_{np}\}.$$

$[a_{n,p}, b_{n,p}]$ may be called the sample p-th quantile interval and any sequence of measurable functions $\delta_{n,p}(X_1, X_2, \ldots, X_n)$ such that $a_{n,p} \leq \delta_{n,p} \leq b_{n,p}$ a.s. is a "natural" estimator of the p-th quantile of the population.

COROLLARY 1. *$\delta_{n,p}$ is strongly pseudoconsistent for the p-th quantile of the population.*

Proof. Note that $F_n(X_{n,[np]-1}) = ([np]-1)/n < p$ and $F_n(X_{n,[np]+1}) = ([np]+1)/n > p$ and it follows from Lemma 1(a) that

$$X_{n,[np]-1} \leq a_{n,p} \leq \delta_{n,p} \leq b_{n,p} \leq X_{n,[np]+1} \; a.s. \tag{1}$$

Hence from Theorem 1

$$P[a_p \leq \liminf \delta_{n,p} \leq \limsup \delta_{n,p} \leq b_p] = 1 \qquad \text{Q.E.D.}$$

COROLLARY 2. $\liminf X_{n,[np]+k} = \liminf \delta_{n,p} = a_p$ a.s. *and* $\limsup X_{n,[np]+k} = \limsup \delta_{n,p} = b_p$ a.s.

Proof. It can be seen by applying the law of iterated logarithm as in Theorem 1 of [3] that $P[X_{n,[np]+k} \leq a_p \text{ i.o.}] = 1$.

$$\Rightarrow \liminf X_{n,[np]+k} \leq a_p \; a.s.$$

It then follows from our Theorem 1 that

$$\liminf X_{n,[np]+k} = a_p \; a.s.$$

and from (1) that $\liminf \delta_{n,p} = a_p$ a.s. Second part follows similarly. Q.E.D.

REMARKS. We note that although $X_{n,[np]+k}$ oscillates as pointed out in [3] the oscillation is confined to $(a-\epsilon, b+\epsilon)$ a.s. in the sense $P[X_{n,[np]+k}\leq a-\epsilon$ i.o.$]=0$ and $P[X_{n,[np]+k}\geq b+\epsilon$ i.o.$]=0\forall\epsilon>0$, as can be seen from Theorem 1.

3. **The asymptotic distribution in discrete case.** In the notation of §2 it is of interest to examine the asymptotic distribution of $X_{n,j}$ for fixed j and X_{n,k_n} when $k_n/n=p+0(1/n)(0<p<1)$. Both of these are thoroughly explored with regularity conditions which imply the continuity of F[2]. We shall now prove the following.

THEOREM 2. *Suppose* (i) $c_n\to\infty$, $k_n/n=p+0(1/n)$ *as* $n\to\infty$; (ii) F *is discrete* (*in the sense of Cramer* [1]) *and* (iii) $\exists c \ni F(c-0)<p$ *and* $F(c)>p$. *Then*

$$H_n(x) = P[c_n(X_{n,k_n}-c) \leq x] \to \begin{cases} 0 & \text{if} \quad x<0 \\ 1 & \text{if} \quad x \geq 0. \end{cases}$$

Proof. Since $X_{n,k_n}\leq x \Leftrightarrow nF_n(x)\geq k_n \forall x$

$$P[X_{n,k_n} \leq c+c_n^{-1}x] = P[nF_n(c+c_n^{-1}x) \geq k_n].$$

For n fixed $nF_n(c+c_n^{-1}x)$ is binomial with parameters n and $F(c+c_n^{-1}x)$. Furthermore $F(c+c_n^{-1}x)=F(c)$ or $F(c-0)$ according as $x\geq 0$ or $x<0$ for sufficiently large n due to the discreteness of F. We get by using uniform normal approximation to the binomial (which for example follows from the normal approximation theorem of [5, p. 288])

$$\lim_{n\to\infty}\left|H_n(x)-\Phi\left\{\frac{nF(c\pm 0)-\dfrac{k_n}{n}}{(F(c\pm 0)(1-F(c\pm 0)))^{1/2}}\right\}\right| = 0$$

according as $x\geq 0$ or $x<0$ where Φ is the c.d.f. of standard normal. Since $k_n/n=p+0(1/n)$ it follows

$$H_n(x) \to \begin{cases} 0 & \text{if} \quad x<0 \\ 1 & \text{if} \quad x \geq 0 \end{cases}, \quad \text{Q.E.D.}$$

REMARKS. Note that condition (iii) is equivalent to the uniqueness of p-th quantile. If it is not unique it can be proved similarly that for all $c_n\to\infty$ and c the limit of $H_n(x)$ is not a distribution function. This solves the problem of finding the asymptotic distribution of $c_n(X_{n,k_n}-c)$ for all $c_n\to\infty$ and c when F is discrete. It follows from (1) that $G_n(x)=P[c_n(\delta_{n,p}-c)\leq x]$ has the same limit as $H_n(x)$.

By using $X_{n,j}\leq x\Leftrightarrow nF_n(x)\geq j$ for any fixed j and following methods similar to the above one can show that

$$P[c_n(X_{n,j}-c) \leq x] \to 1 \;\forall\, x \quad \text{if} \quad F(c) > 0.$$

References

1. H. Cramer, *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N.J., 1951.
2. H. A. David, *Order statistics*, Wiley, New York, 1970.
3. D. Feldman and H. G. Tucker, *Estimation of non-unique quantiles*, Ann. Math. Stat. **37** (1966), 451–457.
4. D. G. Kabe, *Some distribution problems of order statistics from discrete populations*, Ann. Inst. Math. Statist. (1970), 551–556.
5. M. Loève, *Probability theory*, Van Nostrand, Princeton, N.J., 1960.
6. F. Mosteller, *On some useful inefficient statistics*, Ann. Math. Stat., **17** (1946), 377–408.
7. C. R. Rao, *Linear statistical inference and its applications*, Wiley, New York, 1965.

University of Alberta,
Edmonton, Alberta

https://doi.org/10.4153/CMB-1973-048-1 Published online by Cambridge University Press