The Annals of Statistics
1974, Vol. 2, No. 4, 815-818

ON SAMPLE QUANTILES FROM A REGULARLY VARYING DISTRIBUTION FUNCTION[1]

BY LAURENS DE HAAN

Stanford University[2]

A law of the iterated logarithm is proved for sample p-quantiles when the probability distribution function varies regularly at ξ with $F(\xi) = p$.

Introduction. Suppose $U_1, U_2, \cdots$ are independent random variables all with a uniform distribution on $[0, 1]$. Let $F_n(x)$ be the empirical distribution function based on $(U_1, U_2, \cdots, U_n)$, i.e. $nF_n(x) =$ number of U_i less or equal to $x (1 \leqq i \leqq n)$; let $V_{k,n}$ be a kth order statistic corresponding to $(U_1, U_2, \cdots, U_n)$ and take $0 < p < 1$. Bahadur (1966) proved that with probability one

$$V_{[np],n} + F_n(p) - 2p = O(n^{-\frac{3}{4}} \log n) \tag{1}$$

for $n \to \infty$ (here $[a]$ is the integral part of a). This result has been sharpened and extended by Kiefer (1967 and 1970). Ghosh (1971) gave a simple proof of a somewhat weaker result. Using the classical law of the iterated logarithm for Bernoulli variables, one gets from (1) that with probability one

$$\begin{aligned} \limsup_{n\to\infty} \frac{V_{[np],n} - p}{\alpha_n} &= (p(1-p))^{\frac{1}{2}} \\ \liminf_{n\to\infty} \frac{V_{[np],n} - p}{\alpha_n} &= -(p(1-p))^{\frac{1}{2}} \end{aligned} \tag{2}$$

where $\alpha_n = \{2n^{-1} \log \log n\}^{\frac{1}{2}}$ Bahadur also extended these results for a class of distribution functions F determined by: F is twice differentiable in a neighborhood of the point ξ for which $F(\xi) = p$, $F'(\xi)$ is positive and F'' is bounded in the neighborhood of ξ.

It will be shown that the iterated logarithm result (2) can be extended to a larger class of distribution functions including all df's with positive derivative $F'(\xi)$. For the proof we represent any order statistic from an arbitrary distribution as a function of the corresponding order statistic from the uniform distribution. The functions for which (2) carries over are the functions which vary regularly at $x = \xi$.

Transformation of order statistics. Let $X_1, X_2, \cdots$ be independent and identically distributed real-valued random variables with common distribution F. Suppose that the equation $F(\xi) = p$ has exactly one root ξ. Let $Y_{k,n}$ be the kth

Received January 1972; revised August 1973.

[1] Research supported by NSF Grant No. GP-24918.

[2] Now at Erasmus University, Rotterdam.

AMS 1970 *subject classifications.* 60F15, 62G30, 26A12.

Key words and phrases. Order statistics, law of the iterated logarithm, regular variation.

order statistic corresponding to $(X_1, X_2, \cdots, X_n)$; in case of equal order statistics the choice of $Y_{k,n}$ is arbitrary. For $0 < y < 1$ we define the function g by

$$g(y) = \inf\{t \mid F(t) \geqq y\}.$$

The set in the right-hand member is closed for all y, hence

$$g(y) \leqq x \Leftrightarrow y \leqq F(x),$$

i.e.

$$P\{g(U_1) \leqq x\} = P\{U_1 \leqq F(x)\} = F(x).$$

So $g(U_1)$ has the same distribution as X_1. As the validity of a law of the iterated logarithm only depends on the distribution function F, we may consider the sequence $g(U_1), g(U_2), \cdots$ instead of $X_1, X_2, \cdots$. Similarly, instead of $Y_{k,n}$ we will consider $g(V_{k,n})$.

LEMMA 1. *Let g be a non-decreasing function on* $(0, 1)$ *and let* $\alpha > 0$. *If for some finite constant* $c > 0$

$$\lim_{t\downarrow 0} \frac{g(p+t) - g(p)}{g(p) - g(p-t)} = c \tag{3}$$

and for all $x > 0$

$$\lim_{t\downarrow 0} \frac{g(p+tx) - g(p)}{g(p+t) - g(p)} = x^\alpha, \tag{4}$$

then with probability one

$$\begin{aligned} &\limsup_{n\to\infty} \frac{g(V_{[np],n}) - g(p)}{g(p+\alpha_n) - g(p)} = \{p(1-p)\}^{\alpha/2} \\ &\liminf_{n\to\infty} \frac{g(V_{[np],n}) - g(p)}{g(p+\alpha_n) - g(p)} = -c^{-1} \cdot \{p(1-p)\}^{\alpha/2}. \end{aligned} \tag{5}$$

PROOF. As both sides of (4) are monotone functions of x and x^α is a continuous function of x, (4) holds uniformly on all finite intervals. Set $Z_n = \alpha_n^{-1}\{V_{[np],n} - p\}$, then with probability one

$$\begin{aligned} \limsup_{n\to\infty} \frac{g(V_{[np],n}) - g(p)}{g(p+\alpha_n) - g(p)} &= \limsup_{n\to\infty} \frac{g(p+\alpha_n Z_n) - g(p)}{g(p+\alpha_n) - g(p)} \\ &= (\limsup_{n\to\infty} Z_n)^\alpha = \{p(1-p)\}^{\alpha/2}. \end{aligned}$$

From (3) and (4), we get for all $x > 0$

$$\lim_{t\downarrow 0} \frac{g(p-tx) - g(p)}{g(p+t) - g(p)} = -c^{-1} \cdot x^\alpha.$$

This gives the lim inf statement. □

Condition (4) means that the function $U\colon \mathbb{R}^+ \to \mathbb{R}^+$ defined by $U(x) = g(p+x) - g(p)$ is regularly varying at $x = 0$ with exponent α (shorter: α-varying at $x = 0$). For the proof of our theorem we need two lemmas on regularly varying functions. They are very similar to Propositions 5 and 6, page 22 of de Haan (1970); we omit the proofs.

LEMMA 2. *Let* $U: \mathbb{R}^+ \to \mathbb{R}^+$ *be non-decreasing and* ρ*-varying at* $x = 0$ $(0 < \rho < \infty)$. *Define the function* $U^* : \mathbb{R}^+ \to \mathbb{R}^+$ *by*

$$U^*(y) = \inf\{t \mid U(t) \geqq x\}. \tag{6}$$

Then U^* *is* ρ^{-1}*-varying at* $x = 0$.

LEMMA 3. *Suppose* U_1 *and* U_2 (*both* $\mathbb{R}^+ \to \mathbb{R}^+$) *are non-decreasing and* ρ*-varying at* $x = 0$ $(0 < \rho < \infty)$. *Let* $A > 0$. *We have*

$$U_1(x) \sim A \cdot U_2(x) \qquad \text{for} \quad x \downarrow 0$$

if and only if

$$U_1^*(y) \sim A^{-1/\rho} \cdot U_2^*(y) \qquad \text{for} \quad y \downarrow 0,$$

where U_1^* *and* U_2^* *are defined as in* (6).

THEOREM. *Suppose* F *is a distribution function for which the equation* $F(\xi) = p$ *has exactly one root* ξ. *Let* $A, \rho > 0$. *If*

$$\lim_{t\downarrow 0} \frac{F(\xi + t) - F(\xi)}{F(\xi) - F(\xi - t)} = A \tag{7}$$

and for all $x > 0$

$$\lim_{t\downarrow 0} \frac{F(\xi + tx) - F(\xi)}{F(\xi + t) - F(\xi)} = x^\rho, \tag{8}$$

then with probability one

$$\limsup_{n\to\infty} \frac{Y_{[np],n} - \xi}{a_n} = \{p(1-p)\}^{1/2\rho}$$

$$\liminf_{n\to\infty} \frac{Y_{[np],n} - \xi}{a_n} = -\{A^2 \cdot p(1-p)\}^{1/2\rho}, \tag{9}$$

where for $n = 1, 2, \cdots$

$$a_n = \inf\{t \mid F(t) \geqq p + (2n^{-1} \log\log n)^{\frac{1}{2}}\} - \xi.$$

PROOF. Using the transformation g we see that (9) and (5) hold with the same probability, so it is sufficient to prove (3) and (4) with $c = A^{-1/\rho}$ and $\alpha = \rho^{-1}$. Relation (7) means that the function $U_1(x) = F(\xi + x) - F(\xi)$ (for $x > 0$) is regularly varying at $x = 0$ with exponent ρ. The inverse function of U_1 is $U_1^*(y) = g(p + y) - g(p)$. By Lemma 2 then (4) holds with $\alpha = \rho^{-1}$. On the other hand, by (7) and (8) the function $U_2(x) = F(\xi) - F(\xi - x)$ (for $x > 0$) is also ρ-varying at $x = 0$ and $U_2^*(y) = g(p) - g(p - y)$. Application of Lemma 3 then gives (3) with $c = A^{-1/\rho}$. □

REMARK. The proof shows that the lim inf and the lim sup may be treated separately.

REMARK. The rate constants a_n are regularly varying in n as $n \to \infty$ with exponent $(2\rho)^{-1}$; that means e.g. $a_n \cdot n^{\alpha - 1/2\rho} \to 0$ or $\to \infty$ according to $\alpha < 0$ or $\alpha > 0$.

COROLLARY. *If $F'(\xi)$ exists and is positive, then with probability one*

$$\limsup_{n\to\infty} \frac{Y_{[np],n} - \xi}{\alpha_n} = \frac{(p(1-p))^{\frac{1}{2}}}{F'(\xi)}$$

$$\liminf_{n\to\infty} \frac{Y_{[np],n} - \xi}{\alpha_n} = -\frac{(p(1-p))^{\frac{1}{2}}}{F'(\xi)}$$

where $\alpha_n = \{2n^{-1} \log\log n\}^{\frac{1}{2}}$.

PROOF. Obviously, $F'(\xi) > 0$ implies (7) and (8) with $\rho = 1$ and $A = 1$. It also implies that the function $U(x) = F(\xi + x) - F(\xi)$ is asymptotic to $x \cdot F'(\xi)$ for $|x| \downarrow 0$; hence by Lemma 2 the function $g(p + y) - g(p)$ is asymptotic to $y\{F'(\xi)\}^{-1}$ for $|y| \downarrow 0$. This gives (9) with $a_n \sim \alpha_n \cdot \{F'(\xi)\}^{-1}$ for $n \to \infty$. □

EXAMPLE. Take $X_1, X_2, \cdots$ i.i.d. such that $1/X_1$ has a Student distribution with 2 degrees of freedom. Then $F(0) = \frac{1}{2}$, $F'(0) = 0$ but (7) and (8) hold with $A = 1$ and $\rho = 2$. In (9) we can take $a_n = (8n^{-1} \log\log n)^{\frac{1}{4}}$.

REMARK. Note that the conditions of our theorem are the same as Smirnov's necessary and sufficient conditions for the asymptotic normality of $\{g(p + n^{-\frac{1}{2}}) - g(p)\}^{-1}\{Y_{[np],n} - \xi\}$ as $n \to \infty$ (see Smirnov (1949), page 112). A corollary similar to ours can be stated to Smirnov's theorem. This means that the frequently used condition that F' has to be continuous in some neighborhood of ξ (see e.g. Rényi (1970), page 490) is superfluous. This could also be concluded from Ghosh's result ((1971), Theorem 1). On the other hand, it can be remarked that Ghosh's Theorem 1 holds also under the weaker conditions of our theorem (take $Y_{p_n,n} = M_{p_n} + b_n n^{\frac{1}{2}}\{G_n(M_p) - (1 - p)\} + R_n$ where $b_n = F^{-1}(p + n^{-1}) - F^{-1}(p)$ and $M_{p_n} = \inf\{x \mid F(x) \geqq p_n\}$, then $R_n/b_n \to 0$ in probability; in view of Smirnov's result the conditions (7) and (8) are also necessary for $R_n/b_n \to 0$).

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

DE HAAN, L. (1970). On regular variation and its application to the weak convergence of sample extremes. MC Tract 32, *Math. Centrum Amsterdam.*

KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342.

KIEFER, J. (1970). Deviations between the sample quantile process and the sample df. *Non-Parametric Techniques in Statistical Inference*, (J. Puri, ed.). Cambridge Univ. Press.

RÉNYI, A. (1970). *Probability Theory*. North-Holland Publishing, Amsterdam.

SMIRNOV, N. V. (1949). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl. Ser.* (1) **11** 82–143.

ECONOMETRIC INSTITUTE
ERASMUS UNIVERSITY
BURGEMEESTER OUDLAAN 50
ROTTERDAM-3016, NETHERLANDS