The Canadian Journal of Statistics
Vol. 14, No. 3, 1986, Pages 251–255
La Revue Canadienne de Statistique

A note on the residual median process

J.K. GHOSH and C.K. MUSTAFI

Indian Statistical Institute, Calcutta
and *Indian Institute of Management, Calcutta*

Key words and phrases: Residual median, Brownian bridge, Kiefer process, law of iterated logarithm.
AMS 1980 subject classifications: Primary 60F10; secondary 62E20.

ABSTRACT

We study the residual median process, defined as the median of those observations which are greater than a number t. Using appropriate limit theorems, it is shown that the stochastic process converges in law to a Gaussian process defined in terms of a Brownian bridge.

RESUME

Nous étudions le processus stochastique de la médiane résiduelle, c'est-à-dire la médiane des observations dont la valeur est supérieure à un nombre t. Nous démontrons que ce processus stochastique converge en loi vers un processus gaussien défini au moyen d'un pont brownien.

1. INTRODUCTION

Consider a distribution function $F(\cdot)$ $(t > 0)$ which is twice differentiable with a density $f(\cdot)$. Let

$$G(t) = 1 - F(t) \tag{1}$$

be the probability of surviving after time t. The residual median m_t after time t is the solution of the equation

$$\frac{G(m_t)}{G(t)} = \frac{1}{2}. \tag{2}$$

The use of the residual median has some advantages over the residual mean in problems of reliability and biometry. It is not possible to compute the sample mean residual lifetime until the failure of all the units has been observed. With a "fat-tailed" distribution the estimated mean residual time is likely to be unstable due to its dependence on a few high observations. From a theoretical point of view there are situations where the mean residual life may not exist (Johnson and Kotz 1970, p. 234). The form of the residual median has also been used to characterise the form of $F(\cdot)$ (Arnold and Brockett 1983).

Consider n independent observations with distribution function $F(\cdot)$. Let

$$G_n(t) = \frac{\text{no. of observations} \geq t}{n}. \tag{3}$$

The sample residual median after time t is

$$y_t = \inf\left\{y_{t'}; \frac{G_n(y_{t'})}{G_n(t)} \le \frac{1}{2}, t' > t\right\}. \tag{4}$$

Note that y_t is well defined for $a \le t \le b$ if $G_n(b) > 0$, which is true for sufficiently large n a.s.

In this note we shall study the asymptotic behaviour of the process $\sqrt{n}(y_t - m_t)$ for t between any two fixed values $o < a < b < \infty$. It will be shown that the process $\sqrt{n}(y_t - m_t)$, $a < t < b$, converges in law to a Gaussian process which can be defined in terms of a Brownian bridge process. It will be assumed that condition (C) below is satisified for $F(\cdot)$:

(C) $F(\cdot)$ has a continuously differentiable density $f(\cdot)$ which is bounded away from zero and infinity on bounded sets.

It may be indicated here that a similar stochastic process for residual mean was studied by Yang (1978), who showed that the process converges in law to a Gaussian process.

2. THE LIMIT THEOREMS

As stated in the introduction, we assume throughout that F has a continuously differentiable density. This implies (1) that the inverse function $G^{-1}(t)$ also has a continuously differentiable derivative, bounded away from zero and infinity on bounded sets, and (2) m_t is unique, continuous, and strictly increasing.

We begin by observing an elementary but important fact, namely, that

$$G_n(y_t) - G(y_t) = \tfrac{1}{2}\{G_n(t) - G(t)\} - \{G(y_t) - G(m_t)\} + R_n(t), \tag{5}$$

where $R_n(t) = O(n^{-1})$ uniformly in t. The relation (5) follows from the definition of m_t and y_t. Observe that $2G(m_t) = G(t)$. Also, if $nG_n(t) = 2k$, then y_t is the kth order statistic among the $2k$ observations which are greater than or equal to t, and

$$G_n(y_t) = \tfrac{1}{2}G_n(t) + \frac{1}{n}.$$

If $nG_n(t) = 2k + 1$, then y_t is the $(k + 1)$th order statistic and

$$G_n(y_t) = \tfrac{1}{2}G_n(t) + \frac{1}{2n}.$$

Of course we have tacitly assumed that the observations are distinct, which is true with probability one, since F is continuous.

We rewrite (5) as

$$G(y_t) - G(m_t) = \tfrac{1}{2}\{G_n(t) - G(t)\} - \{G_n(y_t) - G(y_t)\} + O(n^{-1}) \tag{6}$$

and estimate $y_t - m_t$ in our first lemma.

LEMMA 1.

$$\sup_{a \le t \le b} |y_t - m_t| = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \qquad \text{a.s.} \tag{7}$$

Proof. By the law of the iterated logarithm for empirical distribution functions (Sen 1981, p. 39),

$$\sup_{0 \le t < \infty} |G_n(t) - G(t)| = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}, \qquad \text{a.s.}$$

So the RHS of (6) is almost surely $O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ uniformly in t. The same is therefore true of the LHS of (6), namely $G(y_t) - G(m_t)$. But

$$y_t - m_t = G^{-1}(G(y_t)) - G^{-1}(G(m_t)),$$

and G^{-1} has a bounded derivative in an open interval containing all m_t $(a \le t \le b)$ by the observation made at the beginning of this section. The lemma follows immediately. Q.E.D.

LEMMA 2. *Assume condition* (C). *Then there exists a probability space with a Kiefer process* $\{K(x, s); o \le x \le 1, s \ge o\}$ *on it so that*

$$\sup_{a \le t \le b} \left| n^{\frac{1}{2}}\{G(y_t) - G(m_t)\} - \frac{\frac{1}{2}K(F(t), n) - K(F(m_t), n)}{n^{\frac{1}{2}}} \right|$$
$$= O(n^{-\frac{1}{4}}(\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}) \quad \text{a.s.} \quad (8)$$

Proof. Let $F_n(t) = 1 - G_n(t)$ be the empirical distribution function of a random sample of size n on F. Then by (6) we have

$$n^{\frac{1}{2}}\{G(y_t) - G(m_t)\} = \tfrac{1}{2}n^{\frac{1}{2}}\{F(t) - F_n(t)\} - n^{\frac{1}{2}}\{F(y_t) - F_n(y_t)\} + O(n^{-1}). \quad (9)$$

Hence by Komlós, Major, and Tusnády (1975) (cf. Theorem 4.4.3 in Csörgő and Révész 1981), the left-hand side of the equality in (8) is almost surely bounded above by

$$O(n^{-\frac{1}{2}}\log^2 n) + \sup_{a \le t \le b} \frac{|K(F(y_t), n) - K(F(m_t), n)|}{n^{\frac{1}{2}}}. \quad (10)$$

Next, with $\epsilon_n = \sup_{a \le t \le m_b}|F(y_t) - F(m_t)|$, we have

$$\sup_{a \le t \le b} |K(F(y_t), n) - K(F(m_t), n)|$$
$$\le \sup_{a \le t \le m_b} \sup_{0 \le s \le \epsilon_n} |K(F(m_t) + s, n) - K(F(m_t), n)|$$
$$\le \sup_{a \le t \le m_b} \sup_{0 \le s \le O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})} |K(F(m_t) + s, n) - K(F(m_t), n)|$$

almost surely, for all but a finite number of n, by (7). Consequently, by taking $h_n = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ in Theorem 1.15.2 in Csörgő and Révész (1981), we obtain that

$$\sup_{a \le t \le b} \frac{|K(F(y_t), n) - K(F(m_t), n)|}{n^{\frac{1}{2}}} = O(n^{-\frac{1}{4}}(\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}) \quad \text{a.s.} \quad (11)$$

The latter combined with (10) yields (8). Q.E.D.

THEOREM. *Assume condition* (C). *Then on the probability of space of Lemma 2 and with its Kiefer process* $K(\cdot\cdot\cdot)$ *we have*

$$\sup_{a \le t \le b} \left| n^{\frac{1}{2}}(y_t - m_t) - \frac{K(F(m_t), n) - \frac{1}{2}K(F(t), n)}{f(m_t)n^{\frac{1}{2}}} \right|$$
$$= O(n^{-\frac{1}{4}}(\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}) \quad \text{a.s.} \quad (12)$$

Proof. By a two-term Taylor expansion of the LHS of (6) and applying (7) and condition (C), we obtain

$$\sup_{a \le t \le b} |n^{\frac{1}{2}}\{G(y_t) - G(m_t)\} - n^{\frac{1}{2}}(y_t - m_t)\{-f(m_t)\}| = O(n^{-\frac{1}{4}} \log \log n) \quad \text{a.s.} \quad (13)$$

Hence by (8) and (13) we get (12).

COROLLARY 1. *Assume condition* (C). *Then on the probability space of Lemma 2 there exists a sequence of Brownian bridges*

$$\{B_n(x); o \le x \le 1\}$$

such that

$$\sup_{a \le t \le b} \left| n^{\frac{1}{2}}(y_t - m_t) - \left(\frac{1}{f(m_t)}\{B_n(F(m_t)) - \tfrac{1}{2} B_n(F(t))\}\right)\right| = o_p(1). \quad (14)$$

Proof. Since for each $n \ge 1$

$$\{n^{-\frac{1}{2}} K(x, n); o \le x \le 1\} \overset{D}{=} \{B_n(x); o \le x \le 1\},$$

by (12) we get (14). Q.E.D.

By (14) we of course have also shown that the $D[a, b]$-valued random process $\{n^{\frac{1}{2}}(y_t - m_t); a \le t \le b\}$ converges in distribution in Skorohod topology to the Gaussian process

$$\left\{\frac{1}{f(m_t)}\{B(F(m_t)) - \tfrac{1}{2}B(F(t))\}; a \le t \le b\right\}, \quad (15)$$

where $\{B(x); o \le x \le 1\}$ is a Brownian bridge.

Although in this article we have studied the stochastic process related to the median residual time, the theorems proved in Section 2 can be extended to a general percentile residual lifetime (Arnold and Brockett 1983) by introducing an appropriate definition corresponding to Equation (2).

Finally, as an illustration we consider the following distribution function (Johnson and Kotz 1970, p. 234):

$$F(t) = 1 - \left(\frac{k}{t}\right)^a, \qquad a > 0, \quad t \ge k > 0. \quad (16)$$

It is easy to see that the mean residual lifetime does not exist if $a < 1$. The median residual life m_t, however, exists for all values of a and can be seen to be equal to

$$m_t = t\,2^{1/a}. \quad (17)$$

ACKNOWLEDGEMENTS

The authors are grateful to the referee for providing elegant proofs of several results included in the manuscript, leading to improved presentation. The research of the first author was done partly at Iowa State University and supported by a grant from the U.S. Army Research Office. The work of the second author was sponsored by the Centre of Management Development Studies, Indian Institute of Management Calcutta.

REFERENCES

Arnold, B., and Brockett, P. (1983). When does βth percentile residual life function determine the distribution. *Oper. Res.*, 31, 391–396.

Csörgő, M., and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.

Johnson, N.L., and Kotz, S. (1970). *Continuous Univariate Distributions—I*. Wiley, New York.
Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent random variables and the sample distribution function. *Z. Wahrsch. Verw. Gebiete*, 32, 111–131.
Sen, P.K. (1981). *Sequential Nonparametrics*. Wiley, New York.
Yang, G.L. (1978). Estimation of a biometric function. *Ann. Statist.*, 6, 112–116.

Received 13 December 1985
Revised 29 March 1986
Accepted 29 March 1986

Operations Management
Indian Institute of Management
Joka, Diamond Harbour Road
Post Box 16757, Alipore Post Office
Calcutta
India