ON THE DISTRIBUTION OF QUANTILES OF RESIDUALS IN A LINEAR MODEL

by

Raymond J. Carroll*

University of North Carolina at Chapel Hill

ABSTRACT

The representation for quantiles discussed by Bahadur (1966) and Ghosh (1971) is extended to quantiles from the residuals in a linear model. As an application, a robust test for symmetry is proposed.

AMS 1970 Subject Classifications: Primary 62G30; Secondary 62J05.

Key Words and Phrases: Quantiles, Order Statistics, Tests for Symmetry.

---

*This work was supported by the Air Force Office of Scientific Research under contract AFOSR-75-2796.

I. Introduction

Let $y_1,\ldots,y_n$ be independent and identically distributed with common distribution F, and let $F(\xi) = p$. Let k_n be a sequence of integers $k_n = np + o\,(n^{1/2} \log n)$, and let V_n be the k_nth sample quantile. Bahadur (1966) and Ghosh (1971) have shown that if F_n is the empirical distribution function of $y_1,\ldots,y_n$, then

$$(1) \quad n^{1/2}\,(V_n-\xi-(k_n/n - F_n(\xi))/F'(\xi)) \overset{p}{\to} 0.$$

Actually, Bahadur's result is an almost sure result, but (1) suffices for many statistical applications. Our goal is to extend (1) to the quantiles of the residuals in a simple linear model, using this to conduct a quick test for symmetry.

We consider the simple linear model $z_i = \underline{x}_i'\,\underline{\beta} + y_i$, where $y_1,\ldots,y_n$ are i.i.d. with distribution F continuous and strictly increasing in a neighborhood of ξ, with $F(\xi) = p$. If $\underline{\beta}_n$ is an estimate of β (possibly the least squares estimate or one of the new robust estimates), the residuals are

$r_{in} = z_i - \underline{x}_i'\,\underline{\beta}_n = y_i - \underline{x}_i'\,(\underline{\beta}_n - \underline{\beta})$. Let $E_n(F_n)$ be the empirical distribution function of the residuals (errors), and choose $a_n = n^{-1/2}\log n$, $b_n = (\log n)^2$. Let k_n be as above, let $I_n = (\xi-a_n,\xi+a_n)$, and define V_n to be the k_nth order statistic of the residuals. We make the following assumptions:

(A1) F has two bounded derivatives in a neighborhood of ξ, and $F'(\xi) = f(\xi) > 0$.

(A2) $a_n^{-1}\,(\underline{\beta}_n-\underline{\beta}_n) \to 0$ (a.s.)

(A3) For some $\delta > 0$, $a_n^{1/2 - \delta} \max\{|x_i| : i=1,\ldots,n\} \to 0$.

(A4) There exists $\underline{x}_o$ finite with $(\log n) n^{-1} \sum_1^n (\underline{x}_i - \underline{x}_o) \to 0$.

Theorem Under conditions (A1) - (A4),

$$n^{1/2} | (V_n - \xi) - (k_n/n - F_n(\xi))/f(\xi) + \underline{x}_o' (\underline{\beta}_n - \underline{\beta}) | \to 0 \ (a.s.) \quad . \tag{2}$$

Remarks Condition (A1) is the same as Bahadur's, while (A2) is quite reasonable. Condition (A3) guarantees that the design elements will not vary too wildly. The usual condition on the design matrix X is that $n^{-1} X'X \to \Sigma$ (Drygas (1976)), and this implies that if there is an intercept term there is an $\underline{x}_o$ with $n^{-1} \sum_1^n (\underline{x}_i - \underline{x}_o) \to 0$, so that (A4) is only a slight strengthening of a standard condition, which in fact is reduced to $n^{-1} \sum_1^n (\underline{x}_i - \underline{x}_o) \to 0$ at the end of Section 2. Note that (A3) and (A4) are guaranteed in the problem considered by Bahadur and Ghosh.

2. Proof of the Theorem

We first consider simple linear regression without an intercept $z_i = x_i \beta + y_i$. The problems of simple and multiple linear regression with an intercept are easy consequences of the proof presented here. Define the following processes:

$$G_n(x) = n^{-1} \sum_{i=1}^{n} \{ I(r_i \le x) - I(y_i \le \xi) - (F(x + x_i(\beta_n - \beta)) - F(x)) \}$$

$$H_n = n^{1/2} \sup\{ |G_n(x)| : x \in I_n \}$$

$$W_n(s,) = n^{-1/2} \sum_{i=1}^{n} \left\{ \begin{array}{l} I(y_1 \le \xi + a_n s + a_n t x_i) - I(y_1 \le \xi) \\ -F(\xi + a_n s + a_n t x_i) + F(\xi) \end{array} \right\}$$

where I(A) is the indicator of the event A. We note that from (A1) - (A4),

$$(3) \quad n^{-1}\sum_{i=1}^{n}\{F(\xi+sa_n)-F(\xi+sa_n+ta_nx_i)+ta_nx_of(\xi)\} = O(n^{-1/2})$$

uniformly for $0\le|s|,|t|\le 1$.

Lemma 1 Under (A1) - (A4), $H_n \to 0$(a.s.) .

Proof of Lemma 1 Since $r_{in} = y_i - x_i(\beta_n-\beta)$, by (A2) is suffices to show that $\sup\{|W_n(s,t)| : o \le s,t, \le 1\} \to o$ (a.s.) . For this, we may consider only $x_i \ge o$ by breaking the sum defining W_n into two parts, one with $x_i \ge o$ and the other over the terms $x_i < o$. By (A1) as in Bahadur's proof, it also suffices to consider points $s,t = \eta_{r,n} = r/b_n$ $(r = o,\dots,b_n)$. To see this, note that if $|s-\eta_{r,n}| \le b_n$, $|t-\eta_{p,n}| \le b_n$, then

$$|n^{-1}\sum_{i=1}^{n}\{F(\xi+sa_n+ta_nx_i) - F(\xi+\eta_{r,n}a_n+\eta_{p,n}a_n x_i)\}|$$

$$= O(a_nb_n^{-1}) = o(n^{-1/2}).$$

Now, by Bernstein's inequality (Hoeffding, eq. 2.13)),

$$P_r\{|W_n(s,t)|>\varepsilon\} \le c_o\exp\{-c,n^{1/4}\}$$

for some constants c_o,ç depending on ε. Thus,

$$P_r\{|W_n(s,t)|>\varepsilon \text{ for some } n\ge n_o,\ (s,t) = (\eta_{rn},\eta_{pn})\}$$

$$\le \sum_{n=n_o}^{\infty} c_o b_n{}^2 \exp\{-c_1 n^{1/4}\} \to o \text{ as } n_o \to \infty \ .$$

This completes the proof of the Lemma. □

Lemma 2 Almost surely as $n \to \infty$, $V_n \varepsilon I_n$.

Proof of Lemma 2 We have

$$\Pr\{V_n \ge \xi + a_n\} = \Pr\{ \sum_1^n I(y_i \le \xi + a_n + x_i(\beta_n - \beta)) \le k_n\} \ .$$

It suffices to show the existence of an $\eta > 0$ for which $Q_{n_o}(\eta) \to 0$, where

$$Q_{n_o}(\eta) = \Pr\{ \textstyle\sum_1^n I(y_i \le \xi + a_n + t a_n x_i) \le k_n \text{ for some } 0 \le t \le \eta\ ,\ n \ge n_o\}$$

$$= \Pr\{F_n(\xi+a_n) \le k_n/n + n^{-1} \textstyle\sum_1^n (F(\xi+a_n) - F(\xi+a_n+ta_nx_i))$$

$$-n^{-1/2} W_n(1,t) \text{ for some } 0\le t\le\eta\ ,\ n\ge n_o\} \ .$$

By Lemma 1 and using equation (3),

$$Q_n(\eta) = \Pr\left\{\begin{array}{l} F_n(\xi+a_n) - F(\xi+a_n) \le -a_n f(\xi)(1+tx_o) + o(a_n) \\ \qquad\qquad\qquad\qquad \text{for some } 0\le t\le\eta,\ n\ge n_o \end{array}\right\} + o(1) \ .$$

Choosing $\eta > 0$ small, $Q_{n_o}(\eta) \to 0$ as in the proof of Bahadur's Lemma 2. □

Proof of the Theorem Since $E_n(k_n) = k_n/n$, Lemmas 1 and 2 may be applied to show that

$$n^{1/2}|F(V_n) - F(\xi) - (k_n/n - E_n(\xi))| \to 0 \text{ (a.s.) } .$$

Further, $F(V_n) - F(\xi) = (V_n - \xi)f(\xi) + o(n^{-1/2})$ (a.s.) and

$$E_n(\xi) - F(\xi) = F_n(\xi) - F(\xi) + W_n(0, a_n{}^{-1}(\beta_n - \beta)) + n^{-1} \sum_{i=1}^n \{F(\xi + x_i(\beta_n - \beta)) - F(\xi)\} \ .$$

Since $A_n^{-1}(\beta_n - \beta) \to o$ (a.s.), the proof is complete by Lemma 1 and (A4). $\square$

3. An Application

Hinkley (1975) suggested a method based on order statistics for estimating the parameter of a power transformation; his method is most applicable in the location parameter case. The Theorem can be used to construct a quick test for symmetry in more complicated models. Let $p \neq 1/2$ and define $V_n^{(p)}$ to be the $[np]^{th}$ order residual. For $F(\xi_p) = p$,

$$(4) \quad n^{1/2}(V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)} - (\xi_p + \xi_{1-p} - 2\xi_{1/2}))$$

has a limiting normal distribution (with mean zero and variance σ^2) independent of the limit distribution of β_n. Under the hypothesis of symmetry, $\xi_p + \xi_{1-p} = 2\xi_{1/2}$, so one rejects the hypothesis if

$$(5) \quad n^{1/2}|V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)}|/\sigma$$

is too large. To estimate σ one needs estimates of the density at the quantiles ξ_p, ξ_{1-p}, $\xi_{1/2}$, which can be accomplished as in Bloch and Gastwirth (1968) by using the Theorem.

References

1. Bahadur, R. R. (1966). A note on quantiles in large samples. Ann. Math. Statist. 37, 577-580.

Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function. Ann. Math. Statist. 39, 1083-1085.

Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. Z. Wahrscheninlichkeitsteorie Verw. Gebiete 34, 119-127.

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

Hinkley, D. V. (1975). On power transformations to symmetry. Biometrika 62, 101-111.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Distribution of Quantiles of Residuals in a Linear Model		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL 6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1161
7. AUTHOR(s) Raymond J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, Washington, D.C.		12. REPORT DATE March, 1978 13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quantiles, Order Statistics, Tests for Symmetry

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The representation for quantiles discussed by Bahadur (1966) and Ghosh (1971) is extended to quantiles from the residuals in a linear model. As an application, a robust test for symmetry is proposed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE