ON THE DISTRIBUTION OF QUANTILES OF RESIDUALS IN A LINEAR MODEL

MAR 78 R J CARROLL

AFOSR-75-2796
ON THE DISTRIBUTION OF QUANTILES OF RESIDUALS IN A LINEAR MODEL

Raymond J. Carroll

Institute of Statistics Nimeo Series No. 1161

March, 1978

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release AFML-TR-190-12 (7D).
Distribution is unlimited.
A. D. BLOSE
Technical Information Office
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 \( k_n \)th 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

\[
\begin{align*}
\Pr \left( \frac{1}{\sqrt{n}} (\bar{x}_n - (k_n/n - F_n(\xi))/(F(\xi))) - 0. 
\right)
\end{align*}
\]

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 \( z_i = x_i' \beta + y_i \), where \( y_1, \ldots, y_n \) are i.i.d. with distribution \( F \) continuous and strictly increasing in a neighborhood of \( \xi \), with \( F(\xi) = p \). If \( \hat{\beta}_n \) is an estimate of \( \beta \) (possibly the least squares estimate or one of the new robust estimates), the residuals are

\[ r_{in} = z_i - x_i' \hat{\beta}_n = y_i - x_i' (\hat{\beta}_n - \beta). \]

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 \( k_n \)th order statistic of the residuals. We make the following assumptions:

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

(A2) \( a_n^{-1} (\hat{\beta}_n - \beta_n) \to 0 \) (a.s.)
(A3) For some $\delta > 0$, $a_n^{1/2} - \delta \max\{|x_i|: i=1, \ldots, n\} > 0$.

(A4) There exists $x_o$ finite with $(\log n)n^{-1} \sum_{i=1}^{n} (x_i - x_o) > 0$.

**Theorem** Under conditions (A1) - (A4),

$$n^{1/2} |(V_n - c) - (k/n - F_n(\xi))/f(\xi) + x_o (\beta_n - \beta)| \rightarrow 0 \text{ (a.s.)}.$$  

**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 $n^{-1}X'X = \Sigma$ (Drygas (1976)), and this implies that if there is an intercept term there is an $x_o$ with $n^{-1}\sum_{i=1}^{n} (x_i - x_o) > 0$, so that (A4) is only a slight strengthening of a standard condition, which in fact is reduced to $n^{-1}\sum_{i=1}^{n} (x_i - x_o) > 0$ 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:

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

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

$$W_n(s_o) = n^{-1/2} \sum_{i=1}^{n} \left\{ I(y_i \leq \xi + a_n s_o + a_n t x_i) - I(y_i \leq \xi) \right\}$$

$$-F(\xi + a_n s_o + a_n t x_i) + F(\xi)$$
where \( I(A) \) is the indicator of the event \( A \). We note that from (A1) - (A4),

\[
\begin{align*}
\frac{1}{n} \sum_{i=1}^{n} \left\{ F(\xi \cdot s_{a_{n}}) - F(\xi \cdot s_{a_{n}} + t \cdot x_{i}) + t \cdot x_{i} \cdot f(\xi) \right\} = O(n^{-1/2})
\end{align*}
\]
uniformly for \( 0 \leq |s|, |t| \leq 1 \).

**Lemma 1** Under (A1) - (A4), \( H_{n} \to O(a.s.) \).

**Proof of Lemma 1** Since \( r_{n} = y_{1} - x_{1}(\beta_{n} \cdot \beta) \), by (A2) is suffices to show that \( \sup \{|W_{n}(s,t)| : 0 \leq s, t, \leq 1\} \to o \) (a.s.).

For this, we may consider only \( x_{i} \geq o \) by breaking the sum defining \( W_{n} \) into two parts, one with \( x_{i} \geq o \) and the other over the terms \( x_{i} < o \). By (A1) as in Bahadur's proof, it also suffices to consider points \( s, t = n_{r_{n}} = r/b_{n} \) \((r = 0, \ldots, b_{n})\). To see this, note that if

\[
|s - n_{r_{n}}| \leq b_{n}, \quad |t - n_{p_{n}}| \leq b_{n},
\]

then

\[
\begin{align*}
\frac{1}{n} \sum_{i=1}^{n} \left\{ F(\xi \cdot s_{a_{n}} + t \cdot x_{i}) - F(\xi \cdot n_{r_{n}, n_{p_{n}}}) + n_{p_{n}} \cdot n_{a_{n}} \cdot x_{i}) \right\}
\end{align*}
\]

\[
= O(a_{n} b_{n}^{-1}) = O(n^{-1/2}).
\]

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

\[
P_{x}(\{|W_{n}(s,t)| > \epsilon\} \leq c_{0} \exp(-c_{n}^{1/4})
\]

for some constants \( c_{0}, c \) depending on \( \epsilon \). Thus,
\[ P_x\{|W_n(s,t)| > \varepsilon \text{ for some } n \geq n_0, (s,t) = (\eta_{rn}, \eta_{pn})\} \]

\[ \leq \sum_{n=n_0}^{\infty} c_0 b_n 2^{\exp(-c_1 n^{1/4})} + o \text{ as } n_0 \to \infty. \]

This completes the proof of the Lemma.

Lemma 2  Almost surely as \( n \to \infty, V_n \in I_n \).

Proof of Lemma 2 We have

\[ \Pr\{V_n \geq \xi + a_n\} = \Pr\{ \sum_{i=1}^{n} I(y_i \leq \xi + a_n + a_n x_i) \leq k_n\} . \]

It suffices to show the existence of an \( \eta > 0 \) for which

\[ Q_{n_0}(n) \to 0, \text{ where} \]

\[ Q_{n_0}(n) = \Pr\{ \sum_{i=1}^{n} I(y_i \leq \xi + a_n + a_n x_i) \leq k_n \text{ for some } 0 \leq t \leq 1, n \geq n_0\} = \Pr(F_n(\xi + a_n) \leq k_n/n + n^{-1} \sum_{i=1}^{n} (F(\xi + a_n + a_n x_i) - F(\xi + a_n + a_n x_i)) ) . \]

By Lemma 1 and using equation (3),

\[ Q_n(n) = \Pr(F_n(\xi + a_n) - F(\xi + a_n) \leq -a_n f(\xi)(1 + x_0) + o(a_n)) + o(1) \]

for some \( 0 \leq t \leq 1, n \geq n_0 \} . \]

Choosing \( \eta > 0 \) small, \( Q_{n_0}(n) \to 0 \) 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, \( F(V_n) - F(\xi) = (V_n - \xi)f(\xi) + o(n^{-1/2}) \text{ (a.s.) and} \)

\[ F_n(\xi) - F(\xi) = F_n(\xi) - F(\xi) + W_n(0,a_n^{-1}(\beta_n - \beta)) + n^{-1} \sum_{i=1}^{n} (F(\xi + x_i(\beta_n - \beta)) - F(\xi)) . \]
Since $A_n^{-1}(\beta_n-\beta) \to o \text{ (a.s.)}$, the proof is complete by Lemma 1 and (A4).

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 $F(x_p) = p$,

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

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

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

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


quantiles, order statistics, tests for symmetry

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.