JOINT ASYMPTOTIC DISTRIBUTION OF MARGINAL QUANTILES AND QUANTILE FUNCTIONS IN SAMPLE FROM A MULTIVARIATE POPULATION

BY

G. Jogesh Babu
Penn State University, University Park

and

C. Radhakrishna Rao
University of Pittsburgh, Pittsburgh

Center for Multivariate Analysis
University of Pittsburgh
Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population

C. Jogesh Babu and C. Radhakrishna Rao

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh, Pittsburgh, PA 15260

Air Force Office of Scientific Research
Department of the Air Force
Bolling Air Force Base, DC 20332

Approved for public release; distribution unlimited.

Asymptotic tests, Bootstrap, Quantiles, Quantile processes, Tests based on medians.

The joint asymptotic distributions of the marginal quantiles and quantile functions in samples from a p-variate population are derived. Of particular interest is the joint asymptotic distribution of the marginal sample medians, on the basis of which tests of significance for population medians are developed. Methods of estimating unknown nuisance parameters are discussed. The approach is completely nonparametric.
JOINT ASYMPTOTIC DISTRIBUTION OF MARGINAL QUANTILES AND QUANTILE FUNCTIONS IN SAMPLES FROM A MULTIVARIATE POPULATION*

BY

G. Jogesh Babu
Penn State University, University Park

and

C. Radhakrishna Rao
University of Pittsburgh, Pittsburgh

Technical Report No. 87-42

October 1987

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*Research sponsored by the Air Force Office of Scientific Research under contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
Summary: The joint asymptotic distributions of the marginal quantiles and quantile functions in samples from a p-variate population are derived. Of particular interest is the joint asymptotic distribution of the marginal sample medians, on the basis of which tests of significance for population medians are developed. Methods of estimating unknown nuisance parameters are discussed. The approach is completely nonparametric.

Key Words: Asymptotic tests, Bootstrap, Quantiles, Quantile processes, Tests based on medians

AMS classification: 62H15, 62G30
1. INTRODUCTION

Let $X = (x_1, \ldots, x_p)$ be a random vector with joint d.f. (distribution function) $F$, $i$-th marginal d.f. $F_i$, $(i,j)$-th marginal d.f. $F_{ij}$ and $i$-th marginal density function $f_i$. We denote the $i$-th marginal quantile function by

$$
\xi_i(q) = F_i^{-1}(q) = \inf\{x : F_i(x) \geq q\}, 0 < q < 1
$$

(1.1)

and, for convenience, a specific quantile say the $q_i$-th of $F_i$ by

$$
\theta_i = \xi_i(q_i).
$$

(1.2)

Further let

$$
n_{ij}(q, r) = F_{ij}(\xi_i(q), \xi_j(r))
$$

(1.3)

and denote for given $q_i$ and $q_j$

$$
\sigma_{ij} = n_{ij}(q_i, q_j) - q_i q_j = F_{ij}(\theta_i, \theta_j) - q_i q_j.
$$

(1.4)

The parameters (1.1) - (1.4) defined above refer to the d.f. of $X$.

Now let

$$
X_i = (x_{i1}, \ldots, x_{ip_i}), i = 1, \ldots, n
$$

(1.5)

be $n$ independent copies of $X$ and denote the empirical d.f. of $(X_i, i=1, \ldots, n)$ by $F(n)$ and the corresponding $i$-th and $(i,j)$-th marginal distributions by $F_i(n)$ and $F_{ij}(n)$ respectively. We denote the quantities (1.1) - (1.4) defined in terms of $F(n)$, $F_i(n)$ and $F_{ij}(n)$ by

$$
\xi_i^{(n)}(q), \theta_i^{(n)} \text{ and } \sigma_{ij}^{(n)}
$$

(1.6)
or simply as

\[ \hat{\xi}_1(q), \hat{\theta}_i \text{ and } \hat{\sigma}_{ij} \]  

as estimates of \( \xi_1(q), \theta_i \) and \( \sigma_{ij} \) respectively.

In this paper, we derive the asymptotic distribution of

\[ \hat{\theta}^i = (\hat{\theta}_1, \ldots, \hat{\theta}_p) = (\hat{\xi}_1(q_1), \ldots, \hat{\xi}_p(q_p)) \]  

for given \( q_1, \ldots, q_p \) and also the joint distribution of the marginal quantile processes

\[ \hat{\xi}_i(q), 0 < q < 1, i = 1, \ldots, p. \]

The asymptotic distributions of the empirical quantile process (Csörgő and Révész (1981)) and of a fixed set of specified quantiles (Mosteller (1946)) in one dimension are well known.

Of particular interest is the joint asymptotic distribution of the marginal sample medians

\[ (\hat{\xi}_1(\frac{1}{2}), \ldots, \hat{\xi}_p(\frac{1}{2})) \]

using which we develop tests of significance for the population medians analogous to tests for the means in the multivariate case (see Rao (1973), pp. 543-573). An early work on the joint asymptotic distribution of the sample medians is due to Mood (1941), where he assumes the existence of the density function for the vector variable \( X \). We obtain the distribution in the general case in a form convenient for practical applications.
2. DISTRIBUTION OF THE MARGINAL SAMPLE QUANTILES

We prove the following theorem concerning the joint asymptotic distribution of

\[(\hat{\theta}_1, \ldots, \hat{\theta}_p) = (\hat{\xi}_1(q_1), \ldots, \hat{\xi}_p(q_p)) \quad (2.1)\]

the sample \(q_1\)-th, \ldots, \(q_p\)-th quantiles of the marginal empirical distributions of \(x_1, \ldots, x_p\) respectively.

**Theorem 2.1.** Let \(F\) be continuously twice differentiable in a neighborhood of \(\theta_i\) and \(\delta_i = f_i(\xi_i(q_i)) = f_i'\theta_i > 0, i = 1, \ldots, p\), where \(f_i\) denotes the derivative of \(F\). Then the asymptotic distribution of

\[Y_n = \sqrt{n}(\hat{\theta}_1 - \theta_1, \ldots, \hat{\theta}_p - \theta_p) \quad (2.2)\]

is \(p\)-variate normal with mean vector zero and variance-covariance matrix

\[
\Sigma = \begin{pmatrix}
\frac{\sigma_1(1-q_1)}{\delta_1} & \frac{\sigma_{12}}{\delta_1 \delta_2} & \cdots & \frac{\sigma_{1p}}{\delta_1 \delta_p} \\
\frac{\sigma_{12}}{\delta_1 \delta_2} & \frac{\sigma_{22}}{\delta_2} & \cdots & \frac{\sigma_{2p}}{\delta_2 \delta_p} \\
\frac{\sigma_{1p}}{\delta_1 \delta_p} & \frac{\sigma_{2p}}{\delta_2 \delta_p} & \cdots & \frac{\sigma_p(1-q_p)}{\delta_p^2}
\end{pmatrix} \quad (2.3)
\]

where \(\sigma_{ij}\) are as defined in (1.4).

**Proof.** By Bahadur's representation of the sample quantiles (see Bahadur (1966)).

\[(\log n)^{-1} n^{3/4} |(\hat{\theta}_i - \theta_i)| - \delta_i^{-1} |(r_i - q_i)| \to^d 0, \quad i = 1, \ldots, p \quad (2.4)\]

where \(r_i = F_i^{(n)}(\theta_i)\). Then, it follows that
\[ y_n = \sqrt{n} (\hat{\theta}_1 - \theta_1, \ldots, \hat{\theta}_p - \theta_p) \]  
(2.5)

and

\[ z_n = \sqrt{n} (\delta^{-1}_1(r_1 - q_1), \ldots, \delta^{-1}_p(r_p - q_p)) \]  
(2.6)

have the same asymptotic distribution. By the multivariate central limit theorem, \( z_n \) weakly converges to a \( p \)-variate normal distribution with mean vector zero and covariance matrix as given in (2.3). This proves Theorem 2.1.

For practical applications we need a consistent estimate of \( \Sigma \) as defined in (2.3). There are two sets of unknown \( \{\sigma_{ij}\} \) and \( \{\delta^{-1}_{ij}\} \) in \( \Sigma \). A consistent estimate of \( \sigma_{ij} \) is provided by \( \hat{\sigma}_{ij} \) as shown in Theorem 2.2.

**Theorem 2.2.** Let \( F_{ij} \) be continuous at \( (\theta_i, \theta_j) = (\xi_i(q_i), \xi_j(q_j)) \). Then

\[ \hat{\sigma}_{ij} = F_{ij}(\hat{\theta}_i, \hat{\theta}_j) \]

\[ = \sigma_{ij} = F_{ij}(\theta_i, \theta_j) \text{ a.e. as } n \to \infty. \]  
(2.7)

**Proof**

\[ |F_{ij}(\theta_i, \theta_j) - F_{ij}(\hat{\theta}_i, \hat{\theta}_j)| \]

\[ \leq |F_{ij}(\theta_i, \theta_j) - F_{ij}(\hat{\theta}_i, \hat{\theta}_j)| \]

\[ + \sup_{x,y} |F_{ij}(x, y) - F_{ij}(n)(x, y)|. \]  
(2.8)

Since \( F_{ij} \) is continuous at \( (\theta_i, \theta_j) \) and

\[ \sup_{x,y} |F_{ij}(x, y) - F_{ij}(n)(x, y)| \to a.e. \]  
(2.9)

it follows that the expression on the left hand side of (2.8) \( \to 0 \) a.e. which
establishes the result (2.7) of Theorem 2.2.

The result (2.7) implies that \( \sigma_{ij} \) in (2.3) can be consistently estimated by its sample equivalent \( \hat{\sigma}_{ij} \).

There exist several methods for the estimation of \( \delta_i \) (see Krieger and Pickands III (1981) and the references there in). Recently, a consistent and efficient estimator of \( \delta_i^{-1} \) based on a sample of size \( n \) has been proposed by Babu (1986a) under the assumption that \( f_i \) is continuously differentiable at \( \xi_i(q_i) \). There is a possibility of this estimate taking negative values, and when this happens some modification of the estimate may have to be made. Using consistent estimates of \( \hat{\sigma}_{ij} \) and \( \hat{\delta}_i^{-1} \), a consistent estimate of \( \sigma_{ij}/\delta_i \delta_j \), the \((i,j)\)-th element of \( \Sigma \), can be obtained as \( \hat{\sigma}_{ij}/\hat{\delta}_i \hat{\delta}_j \).

Another possibility is to obtain a direct estimate of \( \sigma_{ij}/\delta_i \delta_j \) by the bootstrap method

\[
\hat{\sigma}_{ij}/\hat{\delta}_i \hat{\delta}_j = E^* \{ n(\hat{\theta}_i - \bar{\theta}_i)(\hat{\theta}_j - \bar{\theta}_j) \}
\] (2.10)

where \( E^* \) is the expectation under the bootstrap distribution function. The consistency of the estimator (2.10) can be proved on the same lines as those given by Babu (1986b) for the bootstrap estimate of the variance of the sample median.

3. TESTS OF SIGNIFICANCE BASED ON MEDIANS

Let

\[
\hat{\theta}^*_i = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{ip_i}), \hat{\tau}_i
\] (3.1)

be the marginal sample medians and an estimate of \( \Sigma \) (as defined in (2.3))
obtained from a sample of size \( n_i \) from a \( p \)-variate population \( \Pi_i \), \( i=1,...,k \).

Further let \( \theta_i = (\theta_{i1},...,\theta_{ip}) \) be the true value of the marginal medians for \( \Pi_i \). To test the hypothesis

\[
\theta_1 = \ldots = \theta_p
\]

(3.2)

we can use the statistic

\[
X^2 = \text{trace} \left[ \sum_{i=1}^{k} \frac{1}{n_i} \frac{\hat{\Sigma}_i^{-1}}{\hat{\Sigma}_i^{-1} - (\hat{\Sigma}_i^{-1}) \hat{\theta_i} \hat{\theta_i}' } \right]
\]

(3.3)

where

\[
\hat{\theta}_i = (\sum_{i=1}^{k} n_i \Sigma_i^{-1})^{-1} \sum_{i=1}^{k} n_i \Sigma_i^{-1} \bar{\theta}_i
\]

(3.4)

as chi-square on \( p(k-1) \) degrees of freedom, provided the individual sample sizes \( n_1,...,n_k \) are large.

In cases where a common \( \Sigma \) for the \( k \) populations can be assumed, we have the problem of estimating \( \Sigma \) from the combined sample. For this purpose we consider the residual vectors by replacing each observed vector by its difference from the sample median vector computed from the sample to which the observed vector belongs. There are altogether \( n = (n_1+\ldots+n_k) \) residual vectors, arising out of the \( k \) different samples, from which we construct a \( p \)-dimensional empirical distribution function \( E \) with the marginal medians as zeros. Then \( \sigma_{ij} \) can be estimated from \( E_{ij} \), the \( (i,j) \)-th marginal d.f. of \( E \) as indicated in (2.7) and \( \delta_i \) from \( E_i \), the \( i \)-th marginal d.f. of \( E \) using any of the methods described at the end of Section 2. If we denote a common estimate of \( \Sigma \) by \( \hat{\Sigma} \), then we can develop tests of significance concerning the structure of the median vectors \( \theta_i, i=1,...,k \), as in the case of mean values (see Rao (1973), p. 558). For this purpose we compute the "between populations"
matrix

\[ S = \sum_{i=1}^{k} n_i \hat{\Theta}_1 - n \hat{\Theta}_i \]  

(3.5)

where \( n \hat{\Theta} = n_1 \hat{\Theta}_1 + \ldots + n_k \hat{\Theta}_k \), and set up the determinental equation

\[ |S - \lambda \hat{E}| = 0. \]  

(3.6)

The roots of the equation (3.7) can be used as in the table on p. 558 of Rao (1973) to test the dimensionality of the configuration of median values.

4. JOINT DISTRIBUTION OF THE MARGINAL QUANTILE PROCESSES

In Section 2 of the paper, we derived the joint asymptotic distribution of specified marginal quantiles. We now derive the weak limits of the entire marginal quantile processes after suitable scaling. More specifically we consider the processes \( \{Z_n\} \) indexed by \((q_1, \ldots, q_p) \in (0,1)^p\) where

\[ Z_n(q_1, \ldots, q_p) = \sqrt{n}[f_1(\xi_1(q_1))(\xi_1^{(n)}(q_1) - \xi_1(q_1)), \ldots, \]

\[ f_p(\xi_p(q_p))(\xi_p^{(n)}(q_p) - \xi_p(q_p))]. \]  

(4.1)

We first simplify the problem using the following result which is essentially a restatement of Theorem 5.2.2 of Csörgő and Révész (1981).

**Theorem 3.1.** Suppose that for \( i=1, \ldots, p \), the marginal d.f. \( F_i \) is twice differentiable on \((a_i, b_i)\), where
\[- \infty < a_i = \sup \{x: F_i(x) = 0\}\]
\[\infty > b_i = \inf \{x: F_i(x) = 1\}\]
and \(F_i = f_i \neq 0\) on \((a_i, b_i)\). Further assume that
\[
\max \sup_{a_i < x < b_i} \frac{|f_i'(x)|}{F_i(x)[1 - F_i(x)]} < \infty
\]
and \(f_i\) is non-decreasing (non-increasing) on an interval to the right of \(a_i\)(to the left of \(b_i\)). Let
\[
Y^*_n(q_1, \ldots, q_p) = \sqrt{n}(V_{11}(n)(q_1) - q_1, \ldots, V_{pp}(n)(q_p) - q_p)
\]
where \(V_i(n)\) is the empirical d.f. of the uniform variables
\[u_{ij} = F_i(x_{ij}), \ j = 1, \ldots, n.\]
Then
\[
\sup_{q \in (0, 1)^p} |Y^*_n(q) - Z_n(q)| \to 0 \ a.e. \quad (4.2)
\]
Hence \(\{Y^*_n\}\) and \(\{Z_n\}\) have the same limit.

Note that the marginals of \(\{Y^*_n\}\) converge weakly to a Brownian bridge on \(C[0, 1]\) (see Brillingsley (1968)). Since the paths of the limiting process are continuous, we define a new process \(Y_n\) close to \(Y^*_n\) as follows. Let \(D_1^{(n)(t)}\) be, as a function of \(t \in [0, 1]\), the d.f. corresponding to a uniform distribution of mass \((n+1)^{-1}\) over each of the \((n+1)\) intervals \([d_{j-1}, d_j]\), \(j = 1, \ldots, n + 1\), where \(d_0 = 0, d_{n+1} = 1\) and \(d_1, \ldots, d_n\) are the values of \(u_{i1}, \ldots, u_{in}\) arranged in...
increasing order. Clearly

\[ |V_1^{(n)}(t) - D_1^{(n)}| \leq \frac{1}{n}, 0 \leq t \leq 1 \text{ a.e.} \]

So if

\[ Y_n(q) = \sqrt{n}(D_1^{(n)}(q_1) - q_1, \ldots, D_p^{(n)}(q_p) - q_p) \]

then

\[ ||Y_n(q) - Y_n^*(q)|| \leq n^{-\frac{1}{2}} \forall q \in [0,1]^p \text{ a.e.} \]

As a consequence, \( \{Y_n\} \) and \( \{Z_n\} \) have the same weak limits and the marginals of \( Y_n \) are continuous functions. Note that

\[ Y_n \in B = \{ h : h(q) = (h_1(q_1), \ldots, h_p(q_p)), 0 \leq h \leq 1 \} \]

is a continuous function on \([0,1], \mathbb{R}^p\).

Clearly \( B \) is a separable closed linear subspace of the Banach space \( C \) of continuous functions on \([0,1]^p \) into \( \mathbb{R}^p \).

We shall show that \( \{Y_n\} \) converges weakly to a Gaussian measure on \( B \). A probability measure \( \mu \) on \( B \) is called Gaussian if for every \( H \in B^* \), the space of real continuous linear functionals on \( B \), \( \mu H^{-1} \) is Gaussian on the line (see Aranjo and Giné (1980), pp. 140-142, 28 and problem 2 on p. 33).

To characterize \( B^* \), let \( H \) be a real continuous linear functional on \( B \). Then

\[ H(h_1, \ldots, h_p) = H(h_1, 0, \ldots, 0) + \ldots + H(0, 0, \ldots, h_p) \]

\[ = H_1(h_1) + \ldots + H_p(h_p), \text{ say.} \quad (4.3) \]
The zeroes in the first line of (4.3) refer to the zero function. Clearly, each $H_i$ is a real continuous linear functional on $C[0,1]$. It then follows that $B^*$ is the k-fold direct sum of the dual space $C^*$ of $C[0,1]$. By Riesz's representation theorem, for any $L \in C^*$, there exists a signed measure $\nu$ on $[0,1]$ such that

$$L(f) = \int_0^1 f(x) d\nu(x)$$

for any $f \in C[0,1]$, (see Dunford and Schwartz (1958)). Thus for every $H \in B^*$, there exist signed measures $\nu_1, \ldots, \nu_p$ on $[0,1]$ such that for $f = (f_1, \ldots, f_p) \in B$,

$$H(f) = \sum_{i=1}^p \int_0^1 f_i(x) d\nu_i(x).$$

Now let

$$A = \left\{ \sum_{j=1}^r \alpha_j \varepsilon_{x_j} : 0 < x_j \leq 1, x_j, \alpha_j \text{ rational, } j=1, \ldots, r, \right\}$$

where $\varepsilon_x$ is the probability measure putting all its mass at $x$. It is easily seen that $A$ is dense in $C^*$ and is countable. We now state the main result.

**Theorem 3.2.** \{\(Y_n\)\} converges weakly to a Gaussian random element $W = (W_1, \ldots, W_k)$ in $B$, where $W_i$ is a Brownian Bridge for each $i$ and

$$E(W_i(t)W_j(s)) = P(F_i(x_i) < t, F_j(x_j) < s) - ts$$

(4.4)

for all $i, j$ and $0 < t, s < 1$.

**Proof.** Since \(\sqrt{n} (D_i^{(n)}(t)-t) : 0 \leq t \leq 1\) is tight for each $i$ in $C[0,1]$, it follows that \(\{Y_n\}\) is tight in $B$. Since $A$ is dense in $C^*$, in order to show that
\{Y_n\} has a weak limit it is enough to show that for any \(q_{11}, \ldots, q_{1r}, \ldots, q_{p1}, \ldots, q_{pr}\) in \([0,1]\) and \(a_{ij}\) real

\[
\sum_{i=1}^{p} \sum_{j=1}^{r} a_{ij} \frac{1}{\sqrt{n}} (D_i^{(n)}(V_{ij}) - q_{ij})
\]

converges weakly. This holds because of the central limit theorem and the fact that

\[
\sup_{0 \leq t \leq 1} |D_i^{(n)}(t) - V_i^{(n)}(t)| \leq \frac{1}{n} \quad \text{a.e.}
\]

To complete the proof it is enough to show the existence of \(W\) satisfying (4.4).

Since \(\{Y_n\}\) is tight, there exists a random element \(Y\) on \(B\) and a subsequence \(\{Y_{n'}\}\) such that \(Y_{n'}\), converges weakly to \(Y = (Y^{(1)}, \ldots, Y^{(p)})\). Further, from the above arguments

\[
\sum_{i=1}^{p} \sum_{j=1}^{r} a_{ij} Y^{(i)}(q_{ij}) \quad \text{and} \quad \sum_{i=1}^{p} \sum_{j=1}^{r} a_{ij} W_i(q_{ij})
\]

have the same distribution as that of normal random variables. So it follows that \(Y\) satisfies the properties of \(W\) mentioned in (4.4) and \(Y\) is Gaussian.

Thus \(Y_n\) converges weakly to \(W\), and in view of Theorem 3.1, \(\{Z_n\}\) converges to \(W\).


END
DATE
FILMED
7-88
DTIC