Journal of Mathematical Sciences, Vol. 99, No. 3, 2000

ON THE STABILITY OF CHARACTERIZATION BY THE DISTRIBUTION OF ANY ORDER STATISTIC*

A. P. Ushakova and V. G. Ushakov (Chernogolovka, Russia) UDC 519.2

Some upper and lower estimates of stability of characterization of distribution by the distribution of any order statistic are obtained.

1. Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed random variables. Let $X'_{r:n}$ denote the rth order statistic. Johnson and Kotz [1] showed that the distribution of $X'_{r:n}$ characterizes (uniquely determines) the distribution of X_i. In this paper, we prove that this characterization is stable in the uniform metric and obtain some lower and upper estimates of stability.

Let $X_1, \dots, X_n$ and $Z_1, \dots, Z_n$ be two groups of independent random variables identically distributed in each group, $1 \le r \le n$, and $X'_{r:n}$ and $Z'_{r:n}$ be the corresponding rth order statistics. Denote the cumulative distribution functions of X_i, Z_i, $X'_{r:n}$, and $Z'_{r:n}$, respectively, by $F(x)$, $G(x)$, $F^{(r)}(x)$, and $G^{(r)}(x)$.

2. Auxiliary Results

First we prove some elementary estimates connected with the integral of the form

$$\int t^{r-1}(1-t)^{n-r}\,dt, \quad 1 \le r \le n.$$

Below we use the following notations:

$$I_{n,r}^{(1)} = \int_0^y t^{r-1}(1-t)^{n-r}\,dt \tag{1}$$

and

$$I_{n,r}^{(2)} = \int_{1-y}^1 t^{r-1}(1-t)^{n-r}\,dt, \tag{2}$$

where $0 < y < 1$.

LEMMA 1. *For any $0 < y < 1$ and all $r = 1, 2, \dots, n$, the following inequalities hold:*

$$y \ge [rI_{n,r}^{(1)}]^{1/r}, \tag{3}$$

$$y \ge [(n-r+1)I_{n,r}^{(2)}]^{1/(n-r+1)}. \tag{4}$$

If, in addition, $2 \le r \le n-1$, then

$$y \ge 1 - (1 - nI_{n,r}^{(i)})^{1/n}, \quad i = 1, 2. \tag{5}$$

*Supported by the Russian Foundation for Fundamental Research (grant Nos. 96-01-01919 and 97-01-00273) and INTAS-RFBR (grant No. IR-98-0537).

Proceedings of the Seminar on Stability Problems for Stochastic Models, Vologda, Russia, 1998, Part I.

1072-3374/00/9903-1341$25.00 © 2000 Kluwer Academic/Plenum Publishers

Proof. We have

$$I^{(1)}_{n,r} \le \int_0^y t^{r-1}\,dt = \frac{1}{r}y^r,$$

which implies (3), and, analogously,

$$I^{(2)}_{n,r} \le \int_{1-y}^{1} (1-t)^{n-r}\,dt = \frac{1}{n-r+1}y^{n-r+1},$$

which implies (4).

Now suppose that $2 \le r \le n-1$. Let us prove (5) for $i=1$ (for $i=2$ the proof is similar). Set

$$p = \frac{n-1}{r-1}, \qquad q = \frac{n-1}{n-r};$$

then $1/p + 1/q$, and, due to the Hölder inequality,

$$I^{(1)}_{n,r} = \int_0^y t^{r-1}(1-t)^{n-r}\,dt \le \left[\int_0^y t^{n-1}\,dt\right]^{1/p}\left[\int_0^y (1-t)^{n-1}\,dt\right]^{1/q} = n^{-1/p}(y^n)^{1/p}n^{-1/q}(1-(1-y)^n)^{1/q}.$$

This implies that either

$$nI^{(1)}_{n,r} \le y^n$$

or

$$nI^{(1)}_{n,r} \le 1-(1-y)^n. \tag{6}$$

Since $y^n \le 1-(1-y)^n$, we come to the conclusion that (6) holds anyway; (5) is a simple consequence of (6).

Lemma 2. *If* $0 < y \le 1/2$, *then*

$$y \le r^{1/r}2^{(n-r)/r}[I^{(i)}_{n,r}]^{1/r}, \quad i=1,2. \tag{7}$$

If $1/2 < y < 1$, *then*

$$y \le 2^{(r-1)/(n-r+1)}(n-r+1)^{1/(n-r+1)}[I^{(i)}_{n,r}]^{1/(n-r+1)}, \quad i=1,2. \tag{8}$$

Proof. Let $0 < y \le 1/2$; then

$$I^{(1)}_{n,r} \ge \frac{1}{2^{n-r}}\int_0^y t^{r-1}\,dt = \frac{1}{r}\frac{1}{2^{n-r}}y^r,$$

which implies (7) for $i=1$, and, analogously,

$$I^{(2)}_{n,r} \ge \frac{1}{2^{n-r}}\int_{1-y}^{1} t^{r-1}\,dt = \frac{1}{r}\frac{1}{2^{n-r}}(1-(1-y)^r),$$

which implies (7) for $i=2$.

Now let $1/2 < y < 1$; then

$$I^{(1)}_{n,r} \ge \frac{1}{2^{r-1}}\int_0^y (1-t)^{n-r}\,dt = \frac{1}{2^{r-1}}\int_{1-y}^{1} t^{n-r}\,dt = \frac{1}{2^{r-1}(n-r+1)}(1-(1-y)^{n-r+1}),$$

which implies (8) for $i=1$, and, analogously,

$$I^{(2)}_{n,r} \ge \frac{1}{2^{r-1}}\int_{1-y}^{1} (1-t)^{n-r}\,dt = \frac{1}{2^{r-1}}\int_0^y t^{n-r}\,dt = \frac{1}{2^{r-1}(n-r+1)}y^{n-r+1},$$

which implies (8) for $i = 2$.

Lemma 3. *Let $f(x)$, $0 \le x \le 1$, be a nonnegative integrable function. If $f(x)$ is unimodal, i.e., there exists $x_0 \in [0,1]$ such that $f(x)$ increases for $0 \le x \le x_0$ and decreases for $x_0 \le x \le 1$, then*

$$\int_a^b f(x)\,dx \ge \min\left\{ \int_0^{b-a} f(x)\,dx,\ \int_{1-(b-a)}^{1} f(x)\,dx \right\} \tag{9}$$

for any $0 \le a < b \le 1$.

Proof. Without loss of generality, suppose that $f(x)$ is continuous. Set

$$F(x) = \int_x^{x+b-a} f(t)\,dt, \quad 0 \le x \le 1-(b-a).$$

Then $F(x)$ is a differentiable function and

$$F'(x) = f(x+b-a) - f(x).$$

Let us consider two cases: (1) $b-a \le x_0$ and (2) $b-a > x_0$.

(1) $b-a \le x_0$. It is easy to see that, in this case, $F'(x) \ge 0$ for $0 \le x \le x_0-(b-a)$ and $F'(x) \le 0$ for $x_0 \le x \le 1-(b-a)$. Consider the interval $(x_0-(b-a), x_0)$. In this interval, $f(x+b-a)$ decreases and $f(x)$ increases; therefore, there exists $x_1 \in (x_0-(b-a), x_0)$ such that $F'(x) \ge 0$ for $x_0-(b-a) \le x \le x_1$ and $F'(x) \le 0$ for $x_1 \le x \le x_0$. Thus, the function $F(x)$ increases for $0 \le x \le x_1$ and decreases for $x_1 \le x \le 1-(b-a)$, i.e., $F(x)$ is unimodal in the interval $(0, 1-(b-a))$. This obviously implies that

$$F(x) \ge \min\{F(0), F(1-(b-a))\}$$

for all $x \in (0, 1-(b-a))$. Putting $x = a$, we obtain (9).

(2) $b-a > x_0$. In this case, $f(x+b-a) \le f(x)$ and, therefore, $F(x)$ decreases for $x \in (0, 1-(b-a))$, i.e.,

$$F(x) \ge F(1-(b-a)).$$

Putting $x = a$, we obtain

$$\int_a^b f(x)\,dx \ge \int_{1-(b-a)}^{1} f(x)\,dx,$$

which implies (9).

3. Upper Estimates

In this section, we obtain some upper estimates of stability of characterization by the distribution of any order statistics.

Theorem 1. *If*

$$\sup_x |F^{(r)}(x) - G^{(r)}(x)| \le \varepsilon \le 1, \tag{10}$$

then

$$\sup_x |F(x) - G(x)| \le c(n,r)\varepsilon^{1/\max(r,n-r+1)}, \tag{11}$$

where $c(n,r)$ is a constant, depending only on n and r.

Proof. We have (see [2])

$$F^{(r)}(x) = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,dt$$

and an analogous representation for $G^{(r)}(x)$; therefore,

$$|F^{(r)}(x) - G^{(r)}(x)| \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \left| \int\limits_{F(x)}^{G(x)} t^{r-1}(1-t)^{n-r}\,dt \right|. \tag{12}$$

Let us fix an arbitrary x. It is easy to see that the function $f(t) = t^{r-1}(1-t)^{n-r}$ is unimodal in the interval $[0,1]$; hence, from Lemma 3 and (12) we obtain

$$\min\left\{ \int\limits_{0}^{|F(x)-G(x)|} t^{r-1}(1-t)^{n-r}\,dt, \int\limits_{1-|F(x)-G(x)|}^{1} t^{r-1}(1-t)^{n-r}\,dt \right\} \le \frac{\Gamma(r)\Gamma(n-r+1)}{\Gamma(n+1)}\varepsilon. \tag{13}$$

If $|F(x) - G(x)| \le 1/2$, then, due to (7) of Lemma 2, (13) implies that

$$|F(x) - G(x)| \le \left[\frac{r!\,(n-r)!}{n!}\right]^{1/r} 2^{(n-r)/r}\varepsilon^{1/r}. \tag{14}$$

If $|F(x) - G(x)| > 1/2$, then, due to (8) of Lemma 2, (13) implies that

$$|F(x) - G(x)| \le \left[\frac{(r-1)!\,(n-r+1)!}{n!}\right]^{1/(n-r+1)} 2^{(r-1)/(n-r+1)}\varepsilon^{1/(n-r+1)}. \tag{15}$$

Since x is arbitrary, (14) and (15) imply (11).

Remark. In Theorem 1, one can take

$$c(n,r) = \max(c_1(n,r), c_2(n,r)),$$

where

$$c_1(n,r) = \left[\frac{r!\,(n-r)!}{n!}\right]^{1/r} 2^{(n-r)/r}$$

and

$$c_2(n,r) = \left[\frac{(r-1)!\,(n-r+1)!}{n!}\right]^{1/(n-r+1)} 2^{(r-1)/(n-r+1)}.$$

From Theorem 1 the following uniform (not depending on r) estimate can be derived.

Corollary. *If*

$$\sup_x |F^{(r)}(x) - G^{(r)}(x)| \le \varepsilon \le 1,$$

then

$$\sup_x |F(x) - G(x)| \le 2^{n-1}\varepsilon^{1/n}.$$

4. Lower Estimates

In this section, we derive some lower estimates of stability of characterization of a continuous distribution by the distribution of any order statistic. Let us introduce the following notations. Denote the set of all continuous cumulative distribution functions by $\mathcal{A}$. Let $F(x) \in \mathcal{A}$. Denote the set of all continuous cumulative distribution functions $G(x)$ satisfying the condition

$$\sup_x |F^{(r)}(x) - G^{(r)}(x)| \le \varepsilon$$

by $\mathcal{A}_r(F,\varepsilon)$.

Theorem 2. *For any $F(x) \in \mathcal{A}$ the following estimate holds:*

$$\sup_{G\in\mathcal{A}_r(F,\varepsilon)} \sup_x |F(x) - G(x)| \ge c_0(n,r)\varepsilon^{1/\max(r,n-r+1)}, \tag{16}$$

where $c_0(n,r)$ is a constant depending only on r and n.

Proof. Let $F(x) \in \mathcal{A}$. Without loss of generality, suppose that $\varepsilon < 1/2$. Denote by $x_\varepsilon^{(1)}$ and $x_\varepsilon^{(2)}$ points of the real line such that

$$F^{(r)}(x_\varepsilon^{(1)}) = \varepsilon, \qquad F^{(r)}(x_\varepsilon^{(2)}) = 1-\varepsilon$$

(if these points are nonunique, then any of them can be taken). Set

$$H_\varepsilon(x) = \begin{cases} 0 & \text{for } x < x_\varepsilon^{(1)}, \\ \dfrac{F^{(r)}(x)-\varepsilon}{1-2\varepsilon} & \text{for } x_\varepsilon^{(1)} \le x \le x_\varepsilon^{(2)}, \\ 1 & \text{for } x > x_\varepsilon^{(2)}. \end{cases}$$

It is easy to see that

$$\sup_x |H_\varepsilon(x) - F^{(r)}(x)| = \varepsilon. \tag{17}$$

Indeed,

$$\sup_{x<x_\varepsilon^{(1)}} |H_\varepsilon(x) - F^{(r)}(x)| = F^{(r)}(x_\varepsilon^{(1)}) = \varepsilon,$$

$$\sup_{x>x_\varepsilon^{(2)}} |H_\varepsilon(x) - F^{(r)}(x)| = 1 - F^{(r)}(x_\varepsilon^{(2)}) = \varepsilon,$$

and

$$\sup_{x_\varepsilon^{(1)}\le x\le x_\varepsilon^{(2)}} |H_\varepsilon(x) - F^{(r)}(x)| = \frac{\varepsilon}{1-2\varepsilon}|2F^{(r)}(x)-1| \le \varepsilon,$$

because $\varepsilon \le F^{(r)}(x) \le 1-\varepsilon$ for $x_\varepsilon^{(1)} \le x \le \varepsilon^{(2)}$ and, hence, $-(1-2\varepsilon) \le 2F^{(r)}(x) - 1 \le 1-2\varepsilon$ for these x.

Let us consider the following equation with respect to the variable x:

$$\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int\limits_0^x t^{r-1}(1-t)^{n-r}\,dt = a.$$

The solution of this equation exists for any $a \in [0,1]$. In addition, if x_1 and x_2 are two solutions corresponding to the right-hand sides a_1 and a_2, respectively, and $a_1 < a_2$, then $x_1 < x_2$. This implies that there exists a cumulative distribution function $G(x)$ such that

$$H_\varepsilon(x) = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int\limits_0^{G(x)} t^{r-1}(1-t)^{n-r}\,dt$$

for all x. In other words, $H_\varepsilon(x) = G^{(r)}(x)$.

Since

$$H_\varepsilon(x_\varepsilon^{(1)}) = 0, \qquad H_\varepsilon(x_\varepsilon^{(2)}) = 1,$$

we obtain

$$G(x_\varepsilon^{(1)}) = 0, \qquad G(x_\varepsilon^{(2)}) = 1,$$

which implies

$$\sup_x |F(x) - G(x)| \ge \max\{|F(x_\varepsilon^{(1)}) - G(x_\varepsilon^{(1)})|, |F(x_\varepsilon^{(2)}) - G(x_\varepsilon^{(2)})|\} = \max\{F(x_\varepsilon^{(1)}), 1 - F(x_\varepsilon^{(2)})\}. \tag{18}$$

We have

$$\varepsilon = F^{(r)}(x_\varepsilon^{(1)}) - G^{(r)}(x_\varepsilon^{(1)}) = F^{(r)}(x_\varepsilon^{(1)}) = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int\limits_0^{F(x_\varepsilon^{(1)})} t^{r-1}(1-t)^{n-r}\,dt;$$

therefore, due to (3) of Lemma 1,

$$F(x_\varepsilon^{(1)}) \ge \left[\frac{r!\,(n-r)!}{n!}\right]^{1/r} \varepsilon^{1/r} \tag{19}$$

and, analogously,

$$\varepsilon = G^{(r)}(x_\varepsilon^{(2)}) - F^{(r)}(x_\varepsilon^{(2)}) = 1 - F^{(r)}(x_\varepsilon^{(2)})$$

$$= \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int\limits_{F(x_\varepsilon^{(2)})}^{1} t^{r-1}(1-t)^{n-r}\,dt = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} \int\limits_{1-(1-F(x_\varepsilon^{(2)}))}^{1} t^{r-1}(1-t)^{n-r}\,dt,$$

which, due to (4) of Lemma 1, implies that

$$1 - F(x_\varepsilon^{(2)}) \geq \left[\frac{(r-1)!\,(n-r+1)!}{n!}\right]^{1/(n-r+1)} \varepsilon^{1/(n-r+1)}. \tag{20}$$

Then (18)–(20) imply that

$$\sup_x |F(x) - G(x)| \geq \max\left\{ \left[\frac{r!\,(n-r)!}{n!}\right]^{1/r} \varepsilon^{1/r}, \left[\frac{(r-1)!\,(n-r+1)!}{n!}\right]^{1/(n-r+1)} \varepsilon^{1/(n-r+1)} \right\}$$

$$\geq c_0(n,r)\varepsilon^{1/\max(r,n-r+1)}, \tag{21}$$

where one can take

$$c_0(n,r) = \min\left\{ \left[\frac{r!\,(n-r)!}{n!}\right]^{1/r}, \left[\frac{(r-1)!\,(n-r+1)!}{n!}\right]^{1/(n-r+1)} \right\}.$$

From (17) and (21) we finally obtain (16).

Using (5) of Lemma 1 instead of (3) and (4), we obtain the following estimates, which have worse order with respect to ε than estimates given by Theorem 2 but better constants and, therefore, are sometimes more accurate.

Theorem 3. *For any $F(x) \in \mathcal{A}$ the following estimate holds:*

$$\sup_{G \in \mathcal{A}_r(F,\varepsilon)} \sup_x |F(x) - G(x)| \geq 1 - \left(1 - \frac{(r-1)!\,(n-r)!}{(n-1)!}\varepsilon\right)^{1/n}.$$

5. Conclusions

We have proved (and estimated) the stability of characterization for a certain metric (uniform metric for distribution functions). If we take other metrics, the stability can fail. For instance, "local" stability (stability for density functions) does not hold. More exactly, for any $\varepsilon > 0$ and any $M > 0$ there exist two probability density functions $f(x)$ and $g(x)$ such that

$$\sup_x |f^{(r)}(x) - g^{(r)}(x)| \leq \varepsilon$$

but

$$\sup_x |f(x) - g(x)| > M.$$

Let us construct the corresponding example:

$$f(x) = g(x) = 0, \quad x \leq -1; \qquad f(x) = g(x), \quad x \geq n\varepsilon^{n-1};$$

$$f(x) = \frac{\varepsilon^{2-n}}{n}, \quad 0 \leq x \leq n\varepsilon^{n-1}; \qquad g(x) = \frac{\varepsilon}{1 + n\varepsilon^{n-1}}, \quad -1 \leq x \leq n\varepsilon^{n-1}.$$

Then

$$\sup_x |f^{(n)}(x) - g^{(n)}(x)| \leq \varepsilon$$

but

$$\sup_x |f(x) - g(x)| > M(n)\varepsilon^{2-n}.$$

The right-hand side of the estimate given by Theorem 1 rapidly becomes close to 1 when n increases. On the other hand, Theorem 2 shows that this estimate is sharp (up to the constant in front of $\varepsilon^{1/\max(r,n-r+1)}$) for *any* distribution. This means that, although the distribution of any order statistic uniquely determines the distribution of a sample, it *always* "forgets" the latter very fast when the size of the sample increases. Thus, in speaking of applications, one should keep in mind that extremal values of samples carry "essential" information about the distribution of samples only if the samples are "small."

REFERENCES

1. N. L. Johnson and S. Kotz, "Characterization of an absolutely continuous distribution of any order statistic," *Indian Math. J.*, **32**, No. 2, 193–194 (1984).
2. H. A. David, *Order Statistics*, Wiley, New York (1981).

Institute of Microelectronics Technology,
Russian Academy of Sciences,
142432 Chernogolovka,
Russia