QUANTILES AND MEDIANS

By M. Rosenblatt-Roth

University of Bucharest

1. Introduction. The role played by the quantiles and in particular by the median of a random variable in probability theory and in mathematical statistics is well known. It is difficult to work with them, because they are not linear functionals of the random variables, even when these are independent. While the elementary properties of the expectation may be found in every text-book, the analogous properties are known neither for the quantiles nor for the median. This paper contains a contribution to this problem.

2. Definitions and notations. For every number $p(0 \leqq p \leqq 1; q = 1 - p)$ we define the *p-quantile* of the real valued random variable ξ as the real number $r(\xi; p) = r\xi$ for which the two inequalities

$$P\{\xi \leqq r\,\xi\} \geqq p, \qquad P\{\xi \geqq r\,\xi\} \geqq q \tag{1}$$

are simultaneously satisfied. The 0.5-quantile of ξ is his *median*, $m\xi$.

We denote by $M\xi$, $D\xi$ the expectation and the variance; by ξ^-, ξ^+ the infimum and the supremum of ξ.

Let us consider n random variables $\xi_i(1 \leqq i \leqq n)$; we denote

$$\begin{aligned}
&\sigma_n = \textstyle\sum_{i=1}^n \xi_i, \qquad \sigma_{1n} = \sum_{i=1}^n \xi_i^+, \qquad \sigma_{2n} = \sum_{i=1}^n \xi_i^-,\\
&S_n = r(\sigma_n; p) = r\sigma_n, \qquad S_n' = \textstyle\sum_{i=1}^n r(\xi_i; p) = \sum_{i=1}^n r\xi_i,\\
&\pi_n = \textstyle\prod_{i=1}^n \xi_i, \qquad P_n = r\pi_n, \qquad P_n' = \prod_{i=1}^n r\xi_i, \qquad D_n = \sum_{i=1}^n D\xi_i,\\
&U_i = S_i - (S_{i-1} + r\xi_i), \qquad \gamma = 9 + 6^{\frac{1}{2}}8, \qquad Q_1 = 2p^{-\frac{1}{2}} + 2^{\frac{1}{2}}q^{-\frac{1}{2}},\\
&Q_2 = 2^{\frac{1}{2}}p^{-\frac{1}{2}} + 2q^{-\frac{1}{2}}, \qquad Q_3 = 2p^{-\frac{1}{2}} + q^{-\frac{1}{2}}, \qquad Q_4 = p^{-\frac{1}{2}} + 2q^{-\frac{1}{2}}\\
&G_n^{(1)} = (n-1)^{\frac{1}{2}}Q_1, \qquad G_n^{(2)} = \gamma^{\frac{1}{2}}[\alpha^{(n-1)}]^{-\frac{1}{2}}Q_1, \qquad G_n^{(3)} = \gamma^{\frac{1}{2}}\rho^{-\frac{1}{2}}Q_1\\
&G_n^{(4)} = Q_3, \qquad H_n^{(1)} = -(n-1)^{\frac{1}{2}}Q_2, \qquad H_n^{(2)} = -\gamma^{\frac{1}{2}}[\alpha^{(n-1)}]^{-\frac{1}{2}}Q_2,\\
&H_n^{(3)} = -\gamma^{\frac{1}{2}}\rho^{-\frac{1}{2}}Q_2, \qquad H_n^{(4)} = Q_4, \qquad R_n^{(k)} = (n-1)G_n^{(k)},\\
&T_n^{(k)} = (n-1)H_n^{(k)}(1 \leqq k \leqq 4); \qquad I = (1, 2, 3, \cdots).
\end{aligned} \tag{2}$$

In the case where the sequence of random variables $\xi_i(1 \leqq i \leqq n)$ is a Markov chain, we denote by $\alpha_i(1 \leqq i \leqq n)$ the ergodic coefficient ([1], [2], [3], [4]) of its ith stochastic transition function. Let us denote

$$\alpha^{(n)} = \min_{1 \leqq i < n} \alpha_i. \tag{3}$$

The reader is referred to [3] for a survey of the definition and properties of α_i.

Received 7 April 1964.

3. Results.

THEOREM 1. *The inequalities*

$$T_n^{(k)} D_n^{\frac{1}{2}} \leqq S_n - S_n' \leqq R_n^{(k)} D_n^{\frac{1}{2}} \tag{4}$$

are true: for arbitrarily dependent random variables $\xi_i (1 \leqq i \leqq n)$ *if* $k = 1$; *for random variables connected in a Markov chain with* $n\alpha^{(n)} \to \infty (n \to \infty)$ *if* $k = 2$; *for random variables connected in a Markov chain with* $\alpha_i > \rho > 0 (1 \leqq i < n)$ *if* $k = 3$; *for independent random variables if* $k = 4$.

THEOREM 2. *For arbitrarily dependent random variables* $\xi_i (1 \leqq i \leqq n)$,

$$(1 - n^{-1})\sigma_{2n} \leqq S_n - n^{-1} S_n' \leqq (1 - n^{-1})\sigma_{1n}. \tag{5}$$

If $\xi_i \geqq 0 (1 \leqq i \leqq n)$, *the inequalities*

$$(\pi_n^{-})^{1-n^{-1}} \leqq (P_n')^{-n^{-1}} \cdot P_n \leqq (\pi_n^{+})^{1-n^{-1}} \tag{6}$$

are true.

THEOREM 3. *If the sequence of arbitrarily dependent random variables* $\xi_n (n \in I)$ *converge uniformly everywhere to the random variable* ξ, *then* $r\xi_n \to r\xi (n \to \infty)$.

LEMMA 1. *Let* ξ, η *be two arbitrarily dependent random variables and c a constant. The following relations are true. If* $\xi \leqq \eta$, *then*

$$r\xi \leqq r\eta. \tag{7}$$

Let $f(x)$ *be a real valued function of a real variable; if* $f(x)$ *is non-decreasing, then*

$$r[f(\xi); p] = f[r(\xi; p)]; \tag{8}$$

if it is non-increasing, then

$$r[f(\xi); q] = f[r(\xi; p)]. \tag{9}$$

For every ξ, η, *we have also*

$$r\xi + \eta^{-} \leqq r(\xi + \eta) \leqq r\xi + \eta^{+}. \tag{10}$$

If $\xi \geqq 0$, $\eta \geqq 0$, *then*

$$(r\xi)\eta^{-} \leqq r(\xi\eta) \leqq (r\xi)\eta^{+}. \tag{11}$$

CONSEQUENCES OF LEMMA 1.

$$r(\xi + c) = r\xi + c \tag{12}$$

$$r(c\xi) = c \cdot r\xi, \qquad (c > 0) \tag{13}$$

$$r(\xi; p) + r(-\xi; q) = 0 \tag{14}$$

$$r(\xi; p) \cdot r(\xi^{-1}; q) = 1. \tag{15}$$

LEMMA 2. *If* ξ, η *are arbitrarily dependent random variables, there are true the inequalities*

$$Q_2(D\xi + D\eta)^{\frac{1}{2}} \leqq r(\xi + \eta) - (r\xi + r\eta) \leqq Q_1(D\xi + D\eta)^{\frac{1}{2}}. \tag{16}$$

If ξ, η *are independent, we may take* Q_3, Q_4 *instead of* Q_1, Q_2 *respectively.*

LEMMA 3. *The inequalities*

$$H_n^{(k)} D_n^{\frac{1}{2}} \leqq U_i \leqq G_n^{(k)} D_n^{\frac{1}{2}} \qquad (2 \leqq i \leqq n) \tag{17}$$

are true: for arbitrarily dependent random variables $\xi_i (1 \leqq i \leqq n)$ *if* $k = 1$; *for random variables connected in a Markov chain with* $n\alpha^{(n)} \to \infty \, (n \to \infty)$ *if* $k = 2$; *for random variables connected in a Markov chain with* $\alpha_i > \rho > 0 (1 \leqq i < n)$ *if* $k = 3$; *for independent random variables if* $k = 4$.

LEMMA 4.

$$-p^{-\frac{1}{2}} \cdot (D\xi)^{\frac{1}{2}} \leqq r(\xi; p) - M\xi \leqq q^{-\frac{1}{2}} \cdot (D\xi)^{\frac{1}{2}}. \tag{18}$$

LEMMA 5. *For any real valued random variable* ξ *and any real number* a,

$$P\{\xi \leqq a\} \geqq p \tag{19}$$

implies $r(\xi; p) \leqq a$ *and*

$$P\{\xi \geqq a\} \geqq q \tag{20}$$

implies $r(\xi; p) \geqq a$.

4. Proofs.

PROOF OF LEMMA 5. Considering the random events $A = \{\xi > a\}, B = \{\xi \geqq r\xi\}$ from (1) and (19) we obtain $P(A) < q$, $P(B) \geqq q$ from which follows $A \subset B$, which is equivalent with $r(\xi; p) \leqq a$. Considering the random events $A = \{\xi < a\}, B = \{\xi \leqq r\xi\}$, from (1) and (20) we obtain $P(A) < p, P(B) \geqq p$ from which follows the relation $A \subset B$ which is equivalent with $r(\xi; p) \geqq a$.

PROOF OF LEMMA 1 AND ITS CONSEQUENCES. Let us consider the random events $A = \{\xi \geqq r\xi\}, B = \{\eta \geqq r\xi\}$; if $\xi \leqq \eta$, then $A \subset B$ i.e. $P(B) \geqq q$. With the help of Lemma 5, (7) follows.

If $f(x)$ is a non-decreasing function and $\eta = f(\xi)$, let us consider the random events $A = \{\xi \leqq r\xi\}, A_1 = \{\eta \leqq f(r\xi)\}, B = \{\xi \geqq r\xi\}, B_1 = \{\eta \geqq f(r\xi)\}$. From (1) we obtain $P(A_1) = P(A) \geqq p, P(B_1) = P(B) \geqq q$ and from Lemma 5, (8) follows. (9) may be derived similarly. Equations (10) through (15) follow immediately.

PROOF OF LEMMA 4. If $\epsilon_1 = q^{-\frac{1}{2}} \cdot (D\xi)^{\frac{1}{2}}$, $\epsilon_2 = p^{-\frac{1}{2}} \cdot (D\xi)^{\frac{1}{2}}$, by means of Chebychev's inequality we obtain the relations $P\{\xi \leqq M\xi + \epsilon_1\} \geqq P\{|\xi - M\xi| \leqq \epsilon_1\} \geqq 1 - \epsilon_1^{-2} \cdot D\xi = p$ and $P\{\xi \geqq M\xi - \epsilon_2\} \geqq P\{|\xi - M\xi| \leqq \epsilon_2\} \geqq 1 - \epsilon_2^{-2} \cdot D\xi = q$. Using Lemma 5, the result follows.

PROOF OF LEMMA 2. From Lemma 4, we obtain

$$\begin{aligned} U = r(\xi + \eta) - (r\xi + r\eta) &= [r(\xi + \eta) - M(\xi + \eta)] - [r\xi - M\xi] \\ &\quad - [r\eta - M\eta] \leqq q^{-\frac{1}{2}}(D(\xi + \eta))^{\frac{1}{2}} + p^{-\frac{1}{2}}((D\xi)^{\frac{1}{2}} + (D\eta)^{\frac{1}{2}}) \\ &\leqq q^{-\frac{1}{2}}(2(D\xi + D\eta))^{\frac{1}{2}} + 2p^{-\frac{1}{2}}(D\xi + D\eta)^{\frac{1}{2}} = Q_1(D\xi + D\eta)^{\frac{1}{2}} \end{aligned}$$

and also

$$\begin{aligned} U \geqq -p^{-\frac{1}{2}}(D(\xi + \eta))^{\frac{1}{2}} - q^{-\frac{1}{2}}((D\xi)^{\frac{1}{2}} + (D\eta)^{\frac{1}{2}}) &\geqq -p^{-\frac{1}{2}}(2(D\xi + D\eta))^{\frac{1}{2}} \\ &\quad - 2q^{-\frac{1}{2}}(D\xi + D\eta)^{\frac{1}{2}} = -Q_2(D\xi + D\eta)^{\frac{1}{2}}. \end{aligned}$$

If ξ, η are independent, we obtain $U \leqq q^{-\frac{1}{2}}(D(\xi + \eta))^{\frac{1}{2}} + p^{-\frac{1}{2}}((D\xi)^{\frac{1}{2}} + (D\eta)^{\frac{1}{2}}) \leqq Q_3(D\xi + D\eta)^{\frac{1}{2}}$ and $U \geqq -p^{-\frac{1}{2}}(D(\xi + \eta))^{\frac{1}{2}} - q^{-\frac{1}{2}}((D\xi)^{\frac{1}{2}} + (D\eta)^{\frac{1}{2}}) \geqq -Q_4(D\xi + D\eta)^{\frac{1}{2}}$.

Proof of Lemma 3. For arbitrarily dependent random variables, using Lemma 2, we obtain

$$U_i \leqq Q_1(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} \leqq Q_1((i-1)D_{i-1} + D\xi_i)^{\frac{1}{2}} \leqq (i-1)^{\frac{1}{2}} Q_1 D_i^{\frac{1}{2}} \leqq (n-1)^{\frac{1}{2}} Q_1 D_n^{\frac{1}{2}}$$

$$U_i \geqq -Q_2(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} \geqq -(n-1)^{\frac{1}{2}} Q_2 D_n^{\frac{1}{2}}.$$

For random variables connected in a Markov chain, using Lemma 2 and ([3], 9; [4], 1) it follows that

$$U_i \leqq Q_1(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} \leqq Q_1(\gamma[\alpha^{(i-1)}]^{-1} D_{i-1} + D\xi_i)^{\frac{1}{2}} \leqq Q_1(\gamma[\alpha^{(i-1)}]^{-1} D_i)^{\frac{1}{2}} \leqq \gamma^{\frac{1}{2}} [\alpha^{(n-1)}]^{-\frac{1}{2}} Q_1 D_n^{\frac{1}{2}}$$

$$U_i \geqq -Q_2(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} \geqq -\gamma^{\frac{1}{2}}[\alpha^{(n-1)}]^{-\frac{1}{2}} Q_2 D_n^{\frac{1}{2}}.$$

If we observe that from $\alpha_i > \rho > 0 (1 \leqq i < n)$ it follows that $\alpha^{(n-1)} > \rho > 0$, then from these inequalities we obtain the wanted results. For independent random variables, obviously

$$-Q_4 D_n^{\frac{1}{2}} \leqq -Q_4 D_i^{\frac{1}{2}} = -Q_4(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} \leqq U_i \leqq Q_3(D\sigma_{i-1} + D\xi_i)^{\frac{1}{2}} = Q_3 D_i^{\frac{1}{2}} \leqq Q_3 D_n^{\frac{1}{2}}.$$

Proof of Theorem 1. The proof follows from Lemma 3 upon observing that

$$S_n - S_n' = \sum_{i=2}^{n} U_i .$$

Proof of Theorem 2. By repeated use of (10), for any fixed index i, we obtain the inequalities

$$\sigma_{2n} - \xi_i^- + r\xi_i \leqq r\sigma_n \leqq r\xi_i + \sigma_{1n} - \xi_i^+ \qquad (1 \leqq i \leqq n)$$

from which by summation for all these values of i, (5) follows. By repeated use of (11), for any fixed index i we obtain the inequalities $\pi_n^-(\xi_i^-)^{-1} \cdot r\xi_i \leqq r\pi_n \leqq (r\xi_i) \cdot \pi_n^+(\xi_i^+)^{-1}$, $(1 \leqq i \leqq n)$, from which by multiplication for all these values of i, (6) follows.

Proof of Theorem 3. From the uniform convergence of ξ_n to ξ, it follows that for any given $\epsilon > 0$, there exists a number N so that $|\xi_n - \xi| < \epsilon$, for $n > N$. If we denote $A = \{\xi_n \leqq r\xi_n\}$, $A_1 = \{\xi - \epsilon \leqq r\xi_n\}$, $B = \{\xi_n \geqq r\xi_n\}$, $B_1 = \{\xi + \epsilon \geqq r\xi_n\}$, we obtain $A \subset A_1$, $B \subset B_1$; from (1) we obtain $P(A_1) \geqq p$, $P(B_1) \geqq q$ and using Lemma 5 it follows $|r\xi_n - r\xi| < \epsilon$ for $n > N$, which proves our theorem.

5. Remarks.

Remarks on Lemma 1.

1. From (10) and (11) one obtains

$$r\xi \leqq r(\xi + \eta), \quad (\eta \geqq 0); \qquad r(\xi + \eta) \leqq r\xi, \quad (\eta \leqq 0) \tag{21}$$

$$r\xi \leqq r(\xi\eta), \quad (\eta \geqq 1); \qquad r(\xi\eta) \leqq r\xi, \quad (\eta \leqq 1). \tag{22}$$

2. From (13) and (14) it follows that

$$r(c\xi;\, p) = cr(\xi;\, q) \qquad (c < 0). \tag{23}$$

Indeed if $c' = -c$, we have

$$r(c\xi;\, p) = r(-c'\xi;\, p) = c'r(-\xi;\, p) = -c'\cdot r(\xi;\, q) = c\cdot r(\xi;\, q).$$

3. If $p = \frac{1}{2}$, from (13) and (23) it follows that

$$m(c\xi) = c\cdot m\xi. \tag{24}$$

Remarks on Lemma 4.

1. From (18) it is easy to obtain the inequality

$$|r(\xi;\, p) - M\xi| \leqq (\beta^{-1}\cdot D\xi)^{\frac{1}{2}}, \qquad \beta^{-1} = \min\,(p,\, q). \tag{25}$$

This result can be obtained also directly in the same manner as (18), taking $\epsilon = (\beta D\xi)^{\frac{1}{2}}$ in Chebychev's inequality, because we have the relation

$$\min\,[P\{\xi \leqq M\xi + \epsilon\},\, P\{\xi \geqq M\xi - \epsilon\}] \geqq 1 - \epsilon^{-2}\cdot D\xi = \max\,(p,\, q).$$

2. If $p = \frac{1}{2}$, from (18) we obtain as a particular case the known relation $|m\xi - M\xi| \leqq (2D\xi)^{\frac{1}{2}}$.

Remarks on Theorem 1.

1. If $n\alpha^{(n)}$ does not converge to infinity for $n \to \infty$, we may find an infinite sequence of natural numbers m and a number A so that $m\alpha^{(m)} = A + o(1)$. From ([3], 9; [4], 1) with $\gamma' = \gamma[A^{-1} + o(1)]$ we deduce the relation $D\sigma_m \leqq \gamma' m D_m$, which is asymptotically equivalent with the well known inequality

$$D\sigma_m \leqq mD_m\,. \tag{26}$$

2. We may observe that if $n\alpha^{(n)} \to \infty$, it follows that $R_n^{(2)} = o(R_n^{(1)})$, $T_n^{(2)} = o(T_n^{(1)})$.

3. To use (25) instead (18) in the proof of our theorem (and also in Lemmas 2 and 3) is equivalent to taking $Q_1^* = (2 + 2^{\frac{1}{2}})\beta^{-\frac{1}{2}}$ instead of Q_1, Q_2 and $Q_2^* = 3\beta^{-\frac{1}{2}}$ instead of Q_3, Q_4. Obviously $\max\,(Q_1, Q_2) \leqq Q_1^*$, $\max\,(Q_3, Q_4) \leqq Q_2^*$.

REFERENCES

[1] Dynkin, E. B. (1954). On some limit theorems for Markov chains. *Ukrain. Mat. Ž.* **6** 21–27.

[2] Dobrushin, R. L. (1956). The central limit theorem for non-homogeneous Markov Chains. *Teor. Veroyatnost. i Primenen.* **1** 72–89, 365–425.

[3] Rosenblatt-Roth, M. (1963). Some theorems concerning the law of large numbers for non-homogeneous Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 433–445.

[4] Rosenblatt-Roth, M. (1963). Sur la dispersion des sommes de variables enchaînées. *C. R. Acad. Sci. Paris* **256** 5499–5501.