ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIII, 3 (1973)

D. SZYNAL (Lublin)

ON THE LIMIT BEHAVIOUR OF A SEQUENCE OF QUANTILES OF A SAMPLE WITH A RANDOM NUMBER OF ITEMS

1. Introduction and preliminaries. One of the important problems of mathematical statistics concerns the limit behaviour of statistics. Investigations dealing with those problems in the case of an order statistics play a particulary important role in the statistical inference. It is partly because some properties of those statistics do not depend on the distribution from which the random sample is obtained. Limit behaviour of order statistics and, particularly, of quantiles of random samples of size n with $n \to \infty$ was investigated by many authors (see, e. g., [1], [5], [7], [8] and [4], where the bibilography of this subject may be found).

Now, we are going to consider the limit behaviour of quantiles of a random sample of size N, where N is a random variable taking positive integer values n, $n = 1, 2, \ldots$, with probabilities p_n depending on a parameter $\tau > 0$, i. e.,

$$p_n = p_n(\tau) = \mathrm{P}[N(\tau) = n], \quad n = 1, 2, \ldots \tag{1}$$

We need the following notation:

Let $(X_1, \ldots, X_n)$ be a random sample of size n from the distribution with distribution function $F(x)$. A quantile of that sample we denote by $Y_k^{(n)}$, where $k = [n\lambda]+1$ with $0 < \lambda < 1$, and $[x]$ stands for the largest part of x. Further on, we assume that for every λ of the interval $0 < \lambda < 1$ there exists at least one value a_λ such that

$$\lambda - \mathrm{P}[X = a_\lambda] \leqslant F(a_\lambda) \leqslant \lambda. \tag{2}$$

Of course, this value a_λ is a quantile of the distribution function $F(x)$.

2. Almost complete convergence of a sequence of quantiles of a sample. In [8] one can find the conditions under which a sequence of order statistics converges in probability and, even with probability 1, to a quantile of the distribution from which the random sample was obtained.

Some limit theorems dealing with random-sums can be obtained immediately.

THEOREM 1. *Let a random sample* $(X_1, X_2, \ldots, X_N)$ *be taken from the distribution with distribution function* $F(x)$ *such that there exists only one value* a_λ *satisfying inequality* (2). *If* $N \to \infty$ *in probability with* $\tau \to \infty$, *then the sequence* $\{Y_K^{(N)}\}$ $(K = [N\lambda]+1,\ 0 < \lambda < 1)$ *converges in probability to* a_λ, *i. e.*

$$(3)\qquad \lim_{\tau\to\infty} P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon] = 0 \quad \textit{for every } \varepsilon > 0.$$

Proof. It is known that the sequence $Y_k^{(n)} - a_\lambda$ converges to 0 with probability 1 [8]. Using Lemma 2 of [2], we obtain (3).

Now, we are going to prove a stronger result.

THEOREM 2. *Let a random sample* $(X_1, X_2, \ldots, X_N)$ *be taken from the distribution with distribution function* $F(x)$ *such that there exists only one value* a_λ *satisfying* (2). *If the parameter* τ *runs through the set of positive integer numbers* $n = 1, 2, \ldots$, *and if, for every* $\delta > 0$,

$$(4)\qquad \sum_{n=1}^{\infty} P\left[\left|\frac{N}{n} - L\right| \geqslant \delta\right] < \infty,$$

where L *is a positive random variable with values from the interval* (a, b), $0 < a < b < \infty$, *then*

$$(5)\qquad \sum_{n=1}^{\infty} P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon] < \infty \quad \textit{for every } \varepsilon > 0.$$

Proof. Let the event $E_n = [|N/n - L| < \delta]$. Then we have

$$\begin{aligned} P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon] &\leqslant P[\bar{E}_n] + P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon, E_n] \\ &\leqslant P[\bar{E}_n] + P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon, n(a-\delta) < N < n(b+\delta)] \\ &\leqslant P[\bar{E}_n] + P[\max_{[n(a-\delta)] \leqslant m \leqslant [n(b+\delta)]} |Y_k^{(m)} - a_\lambda| \geqslant \varepsilon] \\ &\leqslant P[\bar{E}_n] + \sum_{m=[n(a-\delta)]}^{[n(b+\delta)]} P[Y_k^{(m)} \geqslant a_\lambda + \varepsilon] + \sum_{m=[n(a-\delta)]}^{[n(b+\delta)]} P[Y_k^{(m)} \leqslant a_\lambda - \varepsilon]. \end{aligned}$$

By the definition of the empirical distribution function (see [4], p. 374), we obtain

$$(6)\qquad P[Y_k^{(m)} \geqslant a_\lambda + \varepsilon] = P\left[S_m(a_\lambda + \varepsilon) < \frac{k}{m}\right]$$

and

$$\text{(7)} \qquad P[Y_k^{(m)} \leqslant a_\lambda - \varepsilon] = P\left[S_m(a_\lambda - \varepsilon) \geqslant \frac{k-1}{m}\right],$$

where $S_m(x)$ denotes the empirical distribution function.

The estimations given in [4], p. 379, allow us to write

$$\text{(8)} \qquad P\left[S_m(a_\lambda + \varepsilon) < \frac{k}{m}\right] \leqslant P\left[|S_m(a_\lambda + \varepsilon) - F(a_\lambda + \varepsilon)| > \frac{1}{2}\eta_1\right]$$

and

$$\text{(9)} \qquad P\left[S_m(a_\lambda - \varepsilon) \geqslant \frac{k-1}{m}\right] \leqslant P\left[|S_m(a_\lambda - \varepsilon) - F(a_\lambda - \varepsilon)| > -\frac{1}{2}\eta_2\right],$$

where $F(a_\lambda + \varepsilon) - \lambda = \eta_1 > 0$, and $F(a_\lambda - \varepsilon) - \lambda = \eta_2 < 0$.

Taking into account that $S_m(x)$ for fixed x is the frequency in a Bernoulli scheme with probability of success $p = F(x)$ and using the estimate given in [3], we obtain

$$\text{(10)} \qquad P[|S_m(a_\lambda + \varepsilon) - F(a_\lambda + \varepsilon)| \geqslant \tfrac{1}{2}\eta_1] \leqslant 2\exp\{-m\varepsilon_1^2\},$$

where $\varepsilon_1 = \min\left(\frac{1}{2}\eta_1, F(a_\lambda + \varepsilon), 1 - F(a_\lambda + \varepsilon)\right)$, and, analogously,

$$\text{(11)} \qquad P[|S_m(a_\lambda - \varepsilon) - F(a_\lambda - \varepsilon)| \geqslant -\tfrac{1}{2}\eta_2] \leqslant 2\exp\{-m\varepsilon_2^2\},$$

where $\varepsilon_2 = \min\left(-\frac{1}{2}\eta_2, F(a_\lambda - \varepsilon), 1 - F(a_\lambda - \varepsilon)\right)$.

According to (6)-(11), we obtain

$$\text{(12)} \qquad P[|Y_K^{(N)} - a_\lambda| \geqslant \varepsilon] \leqslant P[\bar{E}_n] + 2\left([n(b+\delta)] - [n(a-\delta)]\right)\left(\exp\{-[n(a-\delta)]\varepsilon_1^2\} + \exp\{-[n(a-\delta)]\varepsilon_2^2\}\right).$$

Now, observing that the right-hand side of inequality (12) is the term of a convergent series, we obtain (5).

Example. Let

$$N = \sum_{k=1}^{n} Z_k,$$

where Z_k, $k = 1, 2, \ldots, n$, are random variables having the same Poisson distribution with parameter μ. From Hsu-Robbins' theorem [6] results

$$\sum_{n=1}^{\infty} P\left[\left|\frac{\sum_{k=1}^{n} Z_k}{n} - \mu\right| \geqslant \delta\right] < \infty \qquad \text{for every } \delta > 0,$$

so (4) holds. Thus, for such a random variable, one can use Theorem 2.

3. Limit distributions of quantiles of a random sample. Now, we are going to investigate the limit behaviour of quantiles of a random sample $(X_1, X_2, \ldots, X_N)$ of size N from a continuous distribution with distribution function $F(x)$ and probability density $f(x)$.

The following theorem holds:

THEOREM 3. *Let $F(x), f(x)$ and $a_\lambda (0 < \lambda < 1)$ be the distribution function, density function and quantile of the random variable X, respectively. Next, let $Y_K^{(N)}$ $(K = [N\lambda]+1, 0 < \lambda < 1)$ be the quantile of a random sample $(X_1, X_2, \ldots, X_N)$, where N is a positive integer-valued random variable independent of X. If the density function $f(x)$ is continuous and positive at the point $x = a_\lambda$, and if, for every $\varepsilon > 0$,*

$$\lim_{\tau\to\infty} \mathrm{P}\left[\left|\frac{N-a}{a}\right| \geqslant \varepsilon\right] = 0, \tag{13}$$

where $a = EN \to \infty$ with $\tau \to \infty$, then the quantile of the random sample $(X_1, X_2, \ldots, X_N)$ is asymptotically normally distributed with parameters a_λ and $\big(\lambda(1-\lambda)/[a]\big)^{1/2}/f(a_\lambda)$.

Proof. Let

$$Z_k^{(n)} = \sqrt{\frac{n}{\lambda(1-\lambda)}}\, f(a_\lambda)(Y_k^{(n)} - a_\lambda),$$

and let $h_{kn}(z)$ be the density function of $Z_k^{(n)}$.

It is sufficient to prove that, for any arbitrary pair of real numbers z_1, z_2 such that $z_1 < z_2$,

$$\lim_{\tau\to\infty} \mathrm{P}[z_1 < Z_K^{(N)} < z_2] = \lim_{\tau\to\infty} \int_{z_1}^{z_2} h_{k[a]}(z)\,dz = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} \exp\{-z^2/2\}\,dz.$$

Simple calculations give

$$\mathrm{P}[z_1 < Z_K^{(N)} < z_2] = \sum_{n=1}^{\infty} \mathrm{P}[z_1 < Z_k^{(n)} < z_2]p_n = \sum_{n=1}^{\infty} p_n \int_{z_1}^{z_2} h_{kn}(z)\,dz$$

$$= \sum_{|n-a|\geqslant a\varepsilon} p_n \int_{z_1}^{z_2} h_{kn}(z)\,dz + \sum_{|n-a|<a\varepsilon} p_n \int_{z_1}^{z_2} h_{kn}(z)\,dz = \sum_{|n-a|\geqslant a\varepsilon} p_n \int_{z_1}^{z_2} h_{kn}(z)\,dz +$$

$$+\mathrm{P}\left[\left|\frac{N-a}{a}\right| < \varepsilon\right] \int_{z_1}^{z_2} h_{k[a]}(z)\,dz + \sum_{|n-a|<\varepsilon a} p_n \int_{z_1}^{z_2} \big(h_{kn}(z) - h_{k[a]}(z)\big)\,dz.$$

Now, we observe that according to (13)

$$\sum_{|n-a|\geqslant a\varepsilon} p_n \int_{z_1}^{z_2} h_{kn}(z)\,dz \leqslant \sum_{|n-a|\geqslant a\varepsilon} p_n = \mathrm{P}\left[\left|\frac{N-a}{a}\right| \geqslant \varepsilon\right] \to 0 \quad \text{as } \tau \to \infty,$$

and

$$\lim_{\tau\to\infty} \mathrm{P}\left[\left|\frac{N-a}{a}\right| < \varepsilon\right] \int_{z_1}^{z_2} h_{k[a]}(z)\,dz = \lim_{\tau\to\infty} \int_{z_1}^{z_2} h_{k[a]}(z)\,dz = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} \exp\{-z^2/2\}\,dz.$$

The last equality follows from the known theorem concerning the asymptotical distribution of a quantile of a sample [8].

To complete the proof, we must still estimate

$$\sum_{|n-a|<\varepsilon a} p_n \int_{z_1}^{z_2} \big(h_{kn}(z) - h_{k[a]}(z)\big)\,dz$$
$$\leqslant \mathrm{P}\left[\left|\frac{N-a}{a}\right| < \varepsilon\right] \int_{z_1}^{z_2} \max_{a(1-\varepsilon)\leqslant n \leqslant a(1+\varepsilon)} |h_{kn}(z) - h_{k[a]}(z)|\,dz$$
$$\leqslant \mathrm{P}\left[\left|\frac{N-a}{a}\right| < \varepsilon\right] \int_{z_1}^{z_2} \max_{[a(1-\varepsilon)]\leqslant n \leqslant [a(1+\varepsilon)]} |h_{kn}(z) - h_{k[a]}(z)|\,dz$$
$$\leqslant \mathrm{P}\left[\left|\frac{N-a}{a}\right| < \varepsilon\right] \max \left\{ \int_{z_1}^{z_2} \big(h_{n[a]}(z) - \min_{[a(1-\varepsilon)]\leqslant n \leqslant [a(1+\varepsilon)]} h_{kn}(z)\big)\,dz, \right.$$
$$\left. \int_{z_1}^{z_2} \big(\max_{[a(1-\varepsilon)]\leqslant n \leqslant [a(1+\varepsilon)]} h_{kn}(z) - h_{k[a]}(z)\big)\,dz \right\}.$$

On the basis of the considerations given in [4], p. 380, we get

$$(14) \qquad h_{kn}(z) \cong \left(\frac{1}{\sqrt{2\pi}} \frac{n-k+1}{n} \frac{1}{1-\lambda}\right) \left(\frac{f\left(a_\lambda + \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{z}{f(a_\lambda)}\right)}{f(a_\lambda)}\right).$$
$$\cdot \exp\left\{\left(\left(\frac{k}{n} - \lambda\right)\sqrt{n} + \frac{\lambda - 1}{\sqrt{n}}\right) \frac{1+Q(n,z)}{\sqrt{\lambda(1-\lambda)}} z - \right.$$
$$\left. - \frac{(k-1)c^2 + (n-k)d^2}{n} \frac{z^2}{2} + no\left(\frac{1}{n}\right)\right\},$$

where $k = [n\lambda]+1$, $0 < \lambda < 1$, and

$$c = \sqrt{\frac{1-\lambda}{\lambda}}\big(1+Q(n,z)\big), \qquad d = \sqrt{\frac{\lambda}{1-\lambda}}\big(1+Q(n,z)\big),$$
$$\lim_{n\to\infty} Q(n,z) = 0;$$

and a similar formula for $h_{k[a]}(z)$ obtained from (14) by putting there $[a]$ instead of n.

According to (14) and to the similar formula for $h_{k[a]}(z)$, we obtain

$$\lim_{\tau\to\infty}\Big(h_{k[a]}(z)-\min_{[a(1-\varepsilon)]\leqslant n\leqslant[a(1+\varepsilon)]} h_{kn}(z)\Big)=0$$

and

$$\lim_{\tau\to\infty}\Big(\max_{[a(1-\varepsilon)]\leqslant n\leqslant[a(1+\varepsilon)]} h_{kn}(z)-h_{k[a]}(z)\Big)=0.$$

Hence

$$\lim_{\tau\to\infty}\sum_{|n-a|<\varepsilon a} p_n\int_{z_1}^{z_2}\big(h_{kn}(z)-h_{k[a]}(z)\big)dz=0,$$

which completes the proof of statement of Theorem 3.

Remark. One can easily see that, for the random variable N of the example of Section 2, $a = n\mu$ and (13) holds. Thus, Theorem 3 can be applied in this case.

References

[1] R. Cogburn, *Lois limites des termes variationelles des sommes normées*, C. R. 246 (1958), p. 3408-3410.

[2] M. Csörgö and R. Fischler, *Departure from dependence: the strong law, standard and random-sum central limit theorems*, Acta Math. Acad. Sci. Hung. 21 (1970), p. 105-114.

[3] R. P. Dilworth, *Note on the strong law of large numbers*, Amer. Math. Monthly 56 (1949), p. 249-251.

[4] M. Fisz, *Probability theory and mathematical statistics*, third edition, New York-London 1963.

[5] B. W. Gnedenko, *Sur la distribution limite du terme maximum d'une série aléatoire*, Ann. Math. 44 (1943), p. 423-453.

[6] P. L. Hsu and H. Robbins, *Complete convergence and the law of large numbers*, Proc. Nat. Acad. Sci. 33 (1946), p. 25-31.

[7] M. Loève, *Ranking limit problems*, Proc. Third Berkeley Symp. 2 (1956), p. 174-194.

[8] Н. В. Смирнов, *Предельные законы распределения для членов вариационного ряда*, Москва 1949.

MATHEMATICAL INSTITUTE
M. CURIE-SKŁODOWSKA UNIVERSITY
LUBLIN

Received on 27. 5. 1972

D. SZYNAL (Lublin)

O GRANICZNYCH WŁASNOŚCIACH CIĄGU KWANTYLI Z PRÓBKI O WIELKOŚCI LOSOWEJ

STRESZCZENIE

W pracy podano pewne graniczne własności kwantyli z próby zawierającej losową liczbę elementów. Własności te ujęte są w trzech twierdzeniach, będących uogólnieniami twierdzeń Gniedenki i Smirnowa o granicznych własnościach kwantyli z próby o liczebności n, gdzie $n \to \infty$.