ON THE INDEPENDENCE IN THE LIMIT OF EXTREME AND CENTRAL ORDER STATISTICS

by

ALAJOS KREM

I. Introduction

Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with the same continuous distribution function $F(x) = \mathbf{P}(\xi_i < x)$. Let us arrange them according to their size and introduce the notation

$$\xi_k^* = R_k(\xi_1, \xi_2, \ldots, \xi_n) \qquad (1 \leqq k \leqq n)$$

where the function $R_k(x_1, x_2, \ldots, x_n)$ of n variables represents the k-th of the $x_1, x_2, \ldots, x_n$ arranged in order of magnitude $(k = 1, 2, \ldots, n)$. The random variables defined in this way are not independent, as the relation

$$\xi_1^* \leqq \xi_2^* \leqq \ldots \leqq \xi_n^*$$

holds.

We shall investigate the independence in limit of the variables ξ_k^* as $n \to \infty$. It is known that for constant h and k ξ_h^* and ξ_k^* are asymptotically independent [1]. Furthermore it is known that the asymptotic independence does not hold for the central members; more precisely if $\frac{h}{n} \to \lambda_1$, $\frac{k}{n} \to \lambda_2$, $0 < \lambda_1 < \lambda_2$, and if the limiting distribution of the vector (ξ_h^*, ξ_k^*) is normal, then ξ_h^* and ξ_k^* are not asymptotically independent [5].

A. RÉNYI has informed me that H. J. ROßBERG has proved the asymptotic independence of ξ_h^* and ξ_k^* under the condition $\frac{h^2}{k} \to 0$ and suggested to me to prove the same under more general conditions. According to his advices, I have proved in my B. Sc. paper in April 1961 among others the theorem of this paper, namely that ξ_h^* and ξ_k^* are asymptotically independent, if $\frac{h}{k} \to 0$. In the mean time H. J. ROßBERG has independently proved the same result see [4]. The proof given below uses the method of A. RÉNYI. The essence of the method is the following: Let us construct the variables

$$\eta_k = F(\xi_k) \quad \text{and} \quad \zeta_k = -\log \eta_k \qquad (k = 1, 2, \ldots, n),$$

it follows that

$$\eta_k^* = F(\xi_k^*) \quad \text{and} \quad \zeta_k^* = -\log \eta_{n-k+1}^* \qquad (k = 1, 2, \ldots, n).$$

The variables ζ_k have the exponential distribution function $1 - e^{-x}$ $(x \geqq 0)$, and

$$(1.1) \qquad \zeta_k^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_k}{n-k+1} \qquad (k = 1, 2, \ldots n)$$ [1]

where δ_i $(i = 1, 2, \ldots, n)$ are independent random variables of exponential distribution function with parameter 1. Thus the variables $\zeta_1^*, \zeta_2^*, \ldots, \zeta_n^*$ form an additive Markov chain. Making use of (1.1), it can be proved that the variables $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ form a Markov chain too.

II. Lemmas

Lemma 1. *Let be* $\xi_n' = \xi_n^{(1)} + \xi_n^{(2)}$ *where* $\lim_{n\to\infty} \mathbf{P}(|\xi_n^{(2)}| > \varepsilon) = 0$ *for arbitrary* $\varepsilon < 0$, *furthermore let* η_n' *be independent from* $\xi_n^{(1)}$. *If* $\xi_n^{1)}$ *has a distribution function* $F_n(x)$ *and* η_n' *has a distribution function* $G_n(y)$, *furthermore the limiting distributions* $\lim_{n\to\infty} F_n(x) = F(x)$ *and* $\lim_{n\to\infty} G_n(y) = G(y)$ *exist, then*

$$\lim_{n\to\infty} \mathbf{P}(\xi_n' < x, \eta_n' < y) = F(x) \cdot G(y)$$

that is ξ_n' *and* η_n' *are asymptotically independent.* ξ_n' *has also a limiting distribution function* $F(x)$.

The proof of this lemma is given e.g. in [3].

Lemma 2. *Put*

$$\bar{\zeta}_k = \frac{\zeta_k^* - \mathbf{M}(\zeta_k^*)}{\mathbf{D}(\zeta_k^*)}$$

where $\mathbf{M}(\xi)$ *denotes the expected value of* ξ *and* $\mathbf{D}^2(\xi)$ *the varience of* ξ. *If* $k \to \infty$, $n - k \to \infty$ *then*

$$\lim_{n\to\infty} \mathbf{P}(\bar{\zeta}_k < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Proof. Let us apply the central limit theorem under Ljapunov's conditions. Let $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n}$ be completely independent random variables for any value of n $(n = 1, 2, \ldots; k_n \geqq 1$ integer). $\mathbf{M}(\xi_{nk}) = M_{nk}$, $\mathbf{D}(\xi_{nk}) = D_{nk}$ and $H_{nk} = \sqrt[\infty]{\mathbf{M}(|\xi_{nk} - M_{nk}|^3)}$ $(k = 1, \ldots, k_n$ and put

$$S_n = \sqrt{\sum_{k=1}^{k_n} D_{nk}^2} \quad \text{and} \quad K_n = \sqrt[3]{\sum_{k=1}^{k_n} H_{nk}^3},$$

further

$$\zeta_n = \sum_{k=1}^{k_n} \xi_{nk}$$

[1] For the proof of (1.1) see A. Rényi [2].

and

$$\zeta_n' = \frac{\zeta_n - \mathbf{M}(\zeta_n)}{\mathbf{D}(\zeta_n)} = \frac{\zeta_n - \sum_{k=1}^{k_n} M_{nk}}{S_n}$$

and let us denote by $F_n(x)$ the distribution function of the variable ζ_n'. Assuming that the Ljapunov's condition

$$\lim_{n\to\infty} \frac{K_n}{S_n} = 0$$

holds, we have

$$\lim_{n\to\infty} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt .$$

Now $\dfrac{\delta_j}{n-j+1}$ takes the role of ξ_{nj}, and $\mathbf{M}(\delta_j) = \mathbf{D}^2(\delta_j) = 1$ $(1 \leq j \leq k(n))$, therefore

$$\mathbf{M}(\zeta_k^*) = \frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-k+1},$$

$$S_n^2 = \mathbf{D}^2(\zeta_k^*) = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-k+1)^2}$$

and

$$K_n^3 = \sum_{j=1}^{k} \mathbf{M}\left(\left|\frac{\delta_j - 1}{n-j+1}\right|^3\right) \leq 3 \cdot \sum_{j=1}^{k} \frac{1}{(n-j+1)^3} .$$

It is easy to see that

$$S_n^2 > \frac{k}{(n+1)(n-k+1)}, \tag{2.1}$$

on the other hand

$$K_n^3 < \frac{3}{n-k} \sum_{j=1}^{k} \frac{1}{(n-j+1)^2} < \frac{3}{n-k} \cdot \frac{k}{n(n-k)} = \frac{3k}{n(n-k)^2} . \tag{2.2}$$

From (2.1) and (2.2) we obtain

$$\frac{K_n^6}{S_n^6} < \frac{9k^2(n+1)^3(n-k+1)^3}{n^2(n-k)^4 k^3} \sim 9\left(\frac{1}{k} + \frac{1}{n-k}\right) \to 0$$

as $n \to \infty$, because $k \to \infty$ and $n - k \to \infty$.

Since the condition of Ljapunov holds, by the central limit theorem the lemma follows.

Lemma 3. *If k is constant, the limiting distributions of $\bar{\zeta}_k$ and $\bar{\zeta}_{n-k+1}$ exist, namely*

$$\lim_{n\to\infty} \mathbf{P}(n\,\zeta_k^* < x) = \int_0^x \frac{t^{k-1}\,e^{-t}}{(k-1)!}\,dt \tag{2.3}$$

and

$$\lim_{n\to\infty} \mathbf{P}(\zeta_{n-k+1}^* - \log n < x) = \int_{-\infty}^x \frac{e^{-uk-e^{-u}}}{(k-1)!}\,du\,. \tag{2.4}$$

Proof. Applying the above method, we have (see e.g. A. RÉNYI [2])

$$\lim_{n\to\infty} \mathbf{P}(n\,\zeta_k^* < x) = \lim \mathbf{P}(n(1-\eta_{n-k+1}^*) < n) = \int_0^x \frac{t^{k-1}\,e^{-t}}{(k-1)!}\,dt$$

and because of the symmetry of the uniform distribution

$$\lim_{n\to\infty} \mathbf{P}(n\,\eta_k^* < x) = \int_0^x \frac{t^{k-1}\,e^{-t}}{(k-1)!}\,dt\,.$$

On the other hand

$$\mathbf{P}(n\eta_k^* < x) = \mathbf{P}\left(\zeta_{n-k+1}^* > \log\frac{n}{x}\right) = \mathbf{P}(\zeta_{n-k+1}^* - \log n > -\log x)\,,$$

therefore

$$\lim_{n\to\infty} \mathbf{P}(\zeta_{n-k+1}^* - \log n > -\log x) = \int_0^x \frac{t^{k-1}\,e^{-t}}{(k-1)!}\,dt$$

and from this we have

$$\lim_{n\to\infty} \mathbf{P}(\zeta_{n-k+1}^* - \log n < x) = \int_{-\infty}^x \frac{e^{-uk-e^{-u}}}{(k-1)!}\,du\,.$$

III. Proof of the asymptotic independence

Theorem. *If $\frac{h}{k} \to 0$ the following relation is valid*

$$\lim_{n\to\infty} \mathbf{P}(\bar{\zeta}_h < x, \bar{\zeta}_k < y) = \lim_{n\to\infty} \mathbf{P}(\bar{\zeta}_h < x)\cdot \lim_{n\to\infty} \mathbf{P}(\bar{\zeta}_k < y)\,.$$

Proof. Let us start from the identity

$$\bar{\zeta}_k = \frac{\mathbf{D}(\zeta_h^*)}{\mathbf{D}(\zeta_k^*)}\cdot\bar{\zeta}_h + \frac{\zeta_k^* - \zeta_h^* - \mathbf{M}(\zeta_k^* - \zeta_h^*)}{\mathbf{D}(\zeta_k^*)}\,. \tag{3.1}$$

We apply Lemma 1 for the sum (3.1). In this case:

$$(3.2)\qquad \begin{cases} \xi_n^{(1)} = \dfrac{\zeta_k^* - \zeta_h^* - \mathbf{M}(\zeta_k^* - \zeta_h^*)}{\mathbf{D}(\zeta_k^*)} & \xi_n^{(2)} = \dfrac{\mathbf{D}(\zeta_h^*)}{\mathbf{D}(\zeta_h^*)} \cdot \bar{\zeta}_h \\ \eta_n' = \bar{\zeta}_h & \xi_n' = \bar{\zeta}_k \,. \end{cases}$$

We must show that the following three assertions are true:

1. $\bar{\zeta}_k$ and $\bar{\zeta}_h$ have limiting distributions,
2. the second member on the right side in (3.1) is independent from $\bar{\zeta}_h$,
3. the first member on the right side in (3.1) converges stochastically to 0.

The first assertion follows from Lemmas 2 and 3. The second assertion follows from formula (1.1). In order to prove the third assertion, it is sufficient to show that

$$\frac{\mathbf{D}(\zeta_h^*)}{\mathbf{D}(\zeta_k^*)} \to 0 \,.$$

Let us now consider (1.1):

$$\zeta_h^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_n}{n-h+1} \qquad (h = 1, 2, \ldots, n)$$

where $\delta_1, \delta_2, \ldots, \delta_n$ are independent random variables of exponential distribution with expected value 1 and covariance 1; therefore we have

$$\mathbf{D}^2(\zeta_h^*) = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{(n-h+1)^2} < \frac{h}{n(n-h)}$$

and

$$\mathbf{D}^2(\zeta_k^*) = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{(n-k+1)^2} > \frac{k}{(n+1)(n-k+1)} \,.$$

From these follows that

$$\frac{\mathbf{D}^2(\zeta_h^*)}{\mathbf{D}^2(\zeta_k^*)} < \frac{h(n+1)(n-k+1)}{n(n-h)\,k} < \frac{h(n+1)}{n\,k} \to 0,$$

therefore

$$\frac{\mathbf{D}(\zeta_h^*)}{\mathbf{D}(\zeta_k^*)} \to 0 \,.$$

So the theorem follows from Lemma 1. The assertion of the theorem is valid also if the variables ξ_k have an arbitrary distribution. The proof follows easily from our theorem by applying a transformation on the variables ξ_k.

I wish to take this opportunity to thank Prof. A. Rényi for his valuable advices and for reading the manuscript.

(Received November 1, 1961)

REFERENCES

[1] GEFFROY, J.: "Contribution à la théorie des valeurs extrêmes." *Publ. Inst. Paris* **8** (1959) 3.
[2] RÉNYI, A.: "On the theory of order statistics". *Acta Math. Acad. Sci. Hung.* **4** (1953) 191—231.
[3] RÉNYI, A.: *Valószínűségszámítás*. Tankönyvkiadó, Budapest, 1954. p. 589.
[4] ROßBERG, H. J.: "Über das asymptotische Verhalten der Rand- und Zentralglieder einer Variationsreihe". *A MTA Mat. Kut. Int. Közleményei* **8** (1963) 461—466.
[5] SMIRNOV, N. V.: "Sur la dépendence des membres d'une séries de variations". *Bull. Univ. État. Moscou*, Sér. In., Sect. A. Math. et Mécan. **1** Fasc. 4 (1937) 319.

О ПРЕДЕЛЬНОЙ НЕЗАВИСИМОСТИ ЗНАЧЕНИЙ КРАЙНИХ И СРЕДНИХ ЭЛЕМЕНТОВ ПОРЯДКОВЫХ СТАТИСТИК

A. KREM

Резюме

Пусть случайные величины ξ_k $(k = 1, 2 \ldots, n)$ независимы и одинаково распределены и пусть $F(x) = \mathbf{P}(\xi < x)$ их общая функция распределения. Упорядочим ξ_i согласно их величине, ξ_k^* $(k = 1, 2 \ldots, n)$ означает k-тую по порядке.

Автор в своей работе показывает, что если для индексов $h = h(n)$ и $k = k(n)$ выполняется условие $\frac{h}{k} \to 0$, тогда ξ_k^* и ξ_h^* будут асимптотически независимы. В теореме используется следующая формула, предложенная A. RÉNYI [2]:

$$\zeta_k^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_k}{n-k+1}, \tag{1.1}$$

где $\zeta_k^* = -\log F(\xi_{n-k+1}^*)$ $(k = 1, 2, \ldots, n)$ и $\delta_1, \delta_2, \ldots, \delta_n$ независимые, экспоненциально распределённые случайные величины с математическим ожиданием, равным 1.