

ON RANDOM MATRICES

by

P. ERDŐS and A. RÉNYI

Introduction

In the present paper we deal with certain random 0 — 1 matrices. Let $\mathcal{M}(n, N)$ denote the set of all n by n square matrices among the elements of which there are exactly N elements ($n \leq N \leq n^2$) equal to 1, all the other elements are equal to 0. The set $\mathcal{M}(n, N)$ contains clearly $\binom{n^2}{N}$ such matrices; we consider a matrix M chosen at random from the set $\mathcal{M}(n, N)$, so that each element of $\mathcal{M}(n, N)$ has the same probability $\binom{n^2}{N}^{-1}$ to be chosen. We ask now how large N has to be, for a given large value of n , in order that the permanent of the random matrix M should be different from zero with probability $\geq \alpha$ where $0 < \alpha < 1$. By other words if $M = (\varepsilon_{jk})$ we want to evaluate asymptotically the probability $P(n, N)$ of the event that there exists at least one permutation j_1, j_2, \dots, j_n of the numbers $1, 2, \dots, n$ such that the product $\varepsilon_{1j_1} \cdot \varepsilon_{2j_2} \cdot \dots \cdot \varepsilon_{nj_n}$ should be equal to 1. A second way to formulate the problem is as follows: we shall say that two elements of a matrix are in independent position if they are not in the same row and not in the same column. Now our question is to determine the probability that the random matrix M should contain n elements which are all equal to 1 and are pairwise in independent position. A third way to state the problem is: what is the probability of the event that the permanent of the random 0 — 1 matrix M should be positive?

We prove in § 1 (Theorem 1) that if

$$(1) \quad N(n) = n \log n + cn + o(n)$$

where c is an arbitrary real constant, then

$$(2) \quad \lim_{n \rightarrow +\infty} P(n, N(n)) = e^{-2e^{-c}}.$$

This implies that if

$$(3) \quad \lim_{n \rightarrow +\infty} \frac{N_1(n) - n \log n}{n} = +\infty,$$

then

$$(4) \quad \lim_{n \rightarrow +\infty} P(n, N_1(n)) = 1,$$

while if

$$(5) \quad \lim_{n \rightarrow +\infty} \frac{N_2(n) - n \log n}{n} = -\infty,$$

then

$$(6) \quad \lim_{n \rightarrow +\infty} P(n, N_2(n)) = 0.$$

This result can be interpreted also in the following way, in terms of graph theory. Let $\Gamma_{n, N}$ be a bichromatic random graph containing n red and n blue vertices, and N edges which are chosen at random among the n^2 possible edges connecting two vertices having different colour (so that each of the $\binom{n^2}{N}$ possible choices has the same probability). Then $P(n, N)$ is equal to the probability that the random graph $\Gamma_{n, N}$ should contain a factor of degree 1, i.e. $\Gamma_{n, N}$ should have a subgraph which contains all vertices of $\Gamma_{n, N}$ and n disjoint edges, i.e. n edges which have no common endpoint.

Clearly if the permanent of a matrix M consisting of zeros and ones is positive, then the matrix M does not contain a row or column all elements of which are equal to 0 (called in what follows for the sake of brevity a 0-row resp. 0-column), but conversely, if M does not contain a 0-row, nor a 0-column, it is not sure that its permanent is different from 0. However, from our result it follows that this is "almost" sure. As a matter of fact, Theorem 1 can be interpreted as follows: if $P(n, N)$ denotes the probability that $\text{perm}(M) > 0$ and $Q(n, N)$ the probability that M does not contain a 0-row or a 0-column, then if $N = N(n)$ is chosen so that for $n \rightarrow \infty$ we should have $Q(n, N(n)) \rightarrow 1$, then we have also $P(n, N(n)) \rightarrow 1$.

One can state this result somewhat vaguely also in the following way: if the permanent of a random matrix with elements 0 and 1 is equal to 0, then under the conditions of Theorem 1 this in most cases is due to the presence of a 0-row or a 0-column.

In § 2 we deal with a somewhat simpler variant of the problem, when the elements ε_{ij} ($1 \leq i \leq n$, $1 \leq j \leq n$) of the matrix M are independent random variables each taking on the values 0 and 1 with probability $1 - p$ and p respectively. The results obtained are analogous to those of § 1. In § 3 we add some remarks and mention some unsolved problems.

Besides elementary combinatorial and probabilistic arguments similar to that used by us in our previous work on random graphs (see [1], [2], [3], [4], [5]) our main tool in proving our results is the well-known theorem of D. KÖNIG (see [6]), which is nowadays well known in the theory of linear programming, according to which if M is an n by n matrix, every element of which is either 0 or 1, then the minimal number of lines (i.e. rows or columns) which contain all the 1-s, is equal to the maximal number of 1-s in independent position. As a matter of fact, for our purposes we need only the special case of this theorem, proved already by G. FROBENIUS [7], concerning the case when the maximal number of ones in independent position is equal to n .

§ 1. Random square matrices with a prescribed number of zeros and ones

Let $P(n, N)$ denote the probability of the event that the random matrix M ($M \in \mathcal{M}(n, N)$) has a positive permanent. According to the theorem of FROBENIUS—KÖNIG (see [6] and [7]) $1 - P(n, N)$ is equal to the probability that there exists a number k such that there can be found k rows and $n - k - 1$

columns of M which contain all the ones ($0 \leq k \leq n-1$). If we denote by $Q_k(n, N)$ the probability that there can be found k rows and $n-k-1$ columns or k columns and $n-k-1$ rows which contain all the ones, and k is the least number with this property, then clearly

$$(1.1) \quad 0 \leq 1 - P(n, N) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} Q_k(n, N).$$

Now we shall prove that if

$$(1.2) \quad N(n) = n \log n + cn + o(n)$$

where c is a real constant, then

$$(1.3) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} Q_k(n, N(n)) = 0,$$

further that

$$(1.4) \quad \lim_{n \rightarrow \infty} Q_0(n, N(n)) = 1 - e^{-2e^{-c}}.$$

Clearly (1.1), (1.3) and (1.4) imply that

$$(1.5) \quad \lim_{n \rightarrow \infty} P(n, N(n)) = e^{-2e^{-c}},$$

which is the result we want to prove. Thus it remains only to prove (1.3) and (1.4). Let us consider first (1.4). Clearly $1 - Q_0(n, N(n))$ is equal to the probability of the event that the random matrix M does not contain a 0-row or a 0-column. Thus we have

$$(1.6) \quad 1 - Q_0(n, N(n)) = \sum_{i=0}^{2n} (-1)^i S_i$$

where $S_0 = 1$ and

$$(1.7) \quad S_i = \sum_{h=0}^i \binom{n}{h} \binom{n}{i-h} \frac{\binom{(n-h)(n-i+h)}{N(n)}}{\binom{n^2}{N(n)}} \quad (i = 1, 2, \dots, 2n),$$

further for each $l \geq 0$

$$(1.8) \quad \sum_{i=0}^{2l+1} (-1)^i S_i \leq 1 - Q_0(n, N(n)) \leq \sum_{i=0}^{2l} (-1)^i S_i.$$

As clearly for each fixed value of i and for $n \rightarrow \infty$, if $N(n)$ is defined by (1.2) we have

$$(1.9) \quad S_i = \frac{2^i e^{-ci}}{i!} (1 + o(1)),$$

it follows that

$$(1.10) \quad \lim_{n \rightarrow \infty} (1 - Q_0(n, N(n))) = \sum_{i=0}^{\infty} (-1)^i \frac{2^i e^{-ci}}{i!} = e^{-2e^{-c}}.$$

Thus (1.4) is proved. Now let us prove (1.3).

Let us suppose that M is a matrix such that all the ones of M are contained in k columns and $n - k - 1$ rows ($k \geq 1$), and k is the least number with this property. Then the matrix M can be partitioned into four matrices A, B, C, D as shown by Fig. 1, so that D consists only of zeros. Then clearly each column of C contains at least two ones, because if a column of C would contain not more than a single 1, then by leaving out this column and adding the row in which this 1 is contained, we would get a system of $k - 1$ columns and $n - k$ rows which contain all the ones, in contradiction to our supposition of the minimum property of k .

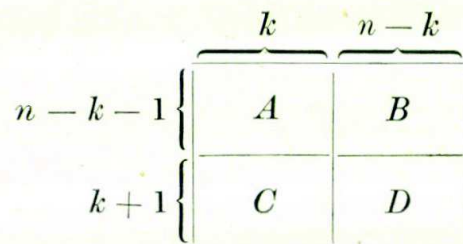


Fig. 1.

Thus it follows that

$$(1.11) \quad Q_k(n, N) \leq 2 \binom{n}{k} \binom{n}{k+1} \binom{k+1}{2}^k \frac{\binom{n(n-k-1) + k(k-1)}{N-2k}}{\binom{n^2}{N}}$$

and thus, that

$$(1.12) \quad Q_k(n, N(n)) \leq \left(\frac{A \log^2 n}{\sqrt{n}} \right)^k \quad \text{for } k = 1, 2, \dots, \left\lfloor \frac{n-1}{2} \right\rfloor$$

where A is a positive constant depending only on c . Thus we obtain

$$(1.13) \quad \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} Q_k(n, N(n)) \leq \frac{A \log^2 n}{\sqrt{n} - A \log^2 n}.$$

From (1.13) we obtain (1.3) and this completes the proof of (1.5).

Thus we obtained the following

Theorem 1. Let $\mathcal{M}(n, N)$ denote the set of all n by n square matrices, among the n^2 elements of which N are equal to 1 and the other $n^2 - N$ to 0. Let M be selected at random from the set $\mathcal{M}(n, N)$ so that each of the $\binom{n^2}{N}$ elements of the set $\mathcal{M}(n, N)$ has the same probability $\left(\binom{n^2}{N} \right)^{-1}$ to be selected. Let $P(n, N)$ denote

the probability of the event that the permanent of the random matrix M is positive. Then if

$$N(n) = n \log n + cn + o(n)$$

where c is any real constant, we have

$$\lim_{n \rightarrow \infty} P(n, N(n)) = e^{-2e^{-c}}.$$

§ 2. Random matrices with independent elements

In this § we prove the following theorem which is a variant of Theorem 1.

Theorem 2. Let $M_n(p)$ be a random n by n matrix whose elements ε_{ij} ($1 \leq i \leq n$; $1 \leq j \leq n$) are independent random variables such that

$$(2.1) \quad \mathbf{P}(\varepsilon_{ij} = 1) = p \quad \text{and} \quad \mathbf{P}(\varepsilon_{ij} = 0) = 1 - p.$$

Let $P_n(p)$ denote the probability of the event that the permanent of the random matrix $M_n(p)$ is positive. Then we have for

$$(2.2) \quad p_n = \frac{\log n + c}{n} + o\left(\frac{1}{n}\right)$$

$$(2.3) \quad \lim_{n \rightarrow \infty} P_n(p_n) = e^{-2e^{-c}}.$$

Proof of Theorem 2. The proof follows step by step the proof of Theorem 1. We have

$$(2.4) \quad 0 \leq 1 - P_n(p) \leq \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} Q_{k,n}(p)$$

where $Q_{k,n}(p)$ denotes the probability that there can be found k rows and $n - k - 1$ columns, or k columns and $n - k - 1$ rows of $M_n(p)$ which contain all the 1-s, and k is the least number with this property. In this case we have

$$(2.5) \quad 1 - Q_{0,n}(p) = \sum_{i=0}^{2n} (-1)^i S_i^*$$

where $S_0^* = 1$ and

$$(2.6) \quad S_i^* = \sum_{h=0}^i \binom{n}{h} \binom{n}{i-h} (1-p)^{in-h(i-h)}.$$

Thus we have for each fixed value of i if (2.2) holds

$$(2.7) \quad \lim_{n \rightarrow \infty} S_i^* = \frac{2^i e^{-ic}}{i!}$$

and therefore

$$(2.8) \quad \lim_{n \rightarrow \infty} (1 - Q_{0,n}(p_n)) = e^{-2e^{-c}}.$$

On the other hand we have now for $k = 1, 2, \dots, \left\lfloor \frac{n-1}{2} \right\rfloor$

$$(2.9) \quad Q_{k,n}(p) \leq 2 \binom{n}{k} \binom{n}{k+1} \binom{k+1}{2}^k p^{2k} (1-p)^{(k+1)(n-k)}$$

and thus

$$(2.10) \quad Q_{k,n}(p_n) \leq \left(\frac{B \log^2 n}{\sqrt{n}} \right)^k \quad \text{for } k = 1, 2, \dots, \left\lfloor \frac{n-1}{2} \right\rfloor,$$

where the constant B depends on c only.

Thus

$$(2.11) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^n Q_{k,n}(p_n) = 0$$

and Theorem 2 follows.

§ 3. Some further remarks

The results of §§ 1 and 2 could be generalized for rectangular matrices of size m by n where $m < n$. In this case the question is: what is the probability that a random matrix of size m by n consisting of zeros and ones should contain m elements in independent position, which are all equal to 1?

Another possible generalization of our results would be to determine the probability distribution of the maximal number of ones in independent position in a random square matrix.

One may ask what can be said about the distribution of the *value* of the permanent of a random square matrix, under conditions of Theorems 1 and 2? It is easy to compute in both cases the mean value of the permanent $\text{perm}(M)$; we have evidently under conditions of Theorem 1

$$\mathbf{E}(\text{perm}(M)) = n! \frac{\binom{n^2 - n}{N(n) - n}}{\binom{n^2}{N(n)}}$$

and under conditions of Theorem 2

$$\mathbf{E}(\text{perm}(M_n(p_n))) = n! p_n^n.$$

It is easy to see, that these expressions are of the form $e^{n \log \log n + O(n)}$ and thus tend rather rapidly to $+\infty$. However one can not draw any conclusion from this fact, because as is easily seen, the variance of the permanent is still much larger than the square of the mean value. An interesting related problem is of course to evaluate under the conditions of Theorem 1 and 2 the probability of the determinant of the random matrix being different from 0.

Another problem arises in connection with the graph-theoretical interpretation of the questions discussed in the present paper: To compute the probability that a random graph having n vertices and N edges should contain a factor of the first degree? We hope to return to these problems in another paper.

(Received November 11, 1963)

REFERENCES

- [1] ERDŐS, P.—RÉNYI, A.: "On random graphs, I." *Publicationes Mathematicae (Debrecen)* **6** (1959) 290—297.
- [2] ERDŐS, P.—RÉNYI, A.: "On the evolution of random graphs." *Publ. of the Math. Inst. of the Hung. Acad. Sci.* **5** (1960) 17—61.
- [3] ERDŐS, P.—RÉNYI, A.: "On the evolution of random graphs." *International Stat. Inst.*, 32. Session, Tokyo, 1960, 119. 1—5.
- [4] ERDŐS, P.—RÉNYI, A.: "On the strength of connectedness of a random graph." *Acta Math. Acad. Sci. Hung.* **12** (1961) 261—267.
- [5] ERDŐS, P.—RÉNYI, A.: "Asymmetric graphs." *Acta Math. Acad. Sci. Hung.* (in print).
- [6] KÖNIG, D.: "Graphok és matrixok." *Matematikai és Fizikai Lapok* **38** (1931) 116—119.
- [7] FROBENIUS, G.: "Über zerlegbare Determinanten." *Sitzungsberichte der Berliner Akademie* 1917, S. 274—277.

О СЛУЧАЙНЫХ МАТРИЦАХ

P. ERDŐS и А. RÉNYI

Резюме

Пусть $\mathcal{M}(n, N)$ обозначает множество всех матриц порядка $(n \times n)$, все элементы которых равны 0, за исключением N элементов, равных 1. Выберем случайным образом из множества $\mathcal{M}(n, N)$ матрицу M так, чтобы каждый из $\binom{n^2}{N}$ элементов множества мог быть выбран с одинаковой вероятностью. Главным результатом статьи является следующая

Теорема. Если $P(n, N)$ обозначает вероятность того, что перманент случайным образом выбранной матрицы M является положительным, и если

$$N(n) = n \log n + cn + o(n),$$

где c — любое действительное число, тогда

$$\lim_{n \rightarrow +\infty} P(n, N(n)) = e^{-2e^{-c}}.$$

ÜBER DAS ASYMPTOTISCHE VERHALTEN DER RAND- UND ZENTRALGLIEDER EINER VARIATIONSREIHE I

von
HANS-JOACHIM ROßBERG¹

Wir betrachten n unabhängige Zufallsgrößen x_1, \dots, x_n mit der Verteilungsfunktion $\mathbf{P}\{x_i < x\} = F(x)$. Bei einem Experiment an diesen x_i mögen die Werte X_i ($i = 1, \dots, n$) eingetreten sein. Dann ist die Ranggröße $\xi_k = R_k(X_1, \dots, X_n)$ als der k -te Wert unter den X_i definiert, wenn diese der Größe nach angeordnet sind. Die so definierten Ranggrößen, die infolge der Beziehung $\xi_1 \leq \xi_2 \leq \dots \leq \xi_n$ voneinander abhängig sind, bilden eine sog. Variationsreihe. Unter den genannten Voraussetzungen hat die Theorie des Grenzverhaltens für $n \rightarrow \infty$ der Ranggrößen in vieler Hinsicht abschließende Resultate erzielt. Kaum behandelt wurden jedoch bisher die Beziehungen zwischen Rand- und Zentralgliedern einer Variationsreihe, die wir hier untersuchen wollen.

Wir beweisen einen Satz über die Unabhängigkeit im Limes von Ranggrößen. Was die Randglieder einer Variationsreihe betrifft, d. h. Ranggrößen ξ_h und ξ_k mit $h = \text{const}$ und $n - k = \text{const}$, so wurde ihre Unabhängigkeit im Limes (unter gewissen Voraussetzungen) von E. J. GUMBEL [2] festgestellt. Einen exakten Beweis dieser Tatsache und eine Verallgemeinerung lieferte T. HOMMA [3]. In einem sehr starken Sinne bewies J. GEFFROY [1] die Unabhängigkeit im Limes von Randgliedern. Den Fall zweier Zentralglieder $\left(\frac{h}{n} \rightarrow \lambda_1, \frac{k}{n} \rightarrow \lambda_2, 0 < \lambda_1 < \lambda_2 < 1, \text{ d.h. } \lim \frac{h}{k} > 0\right)$ behandelte N. W. SMIRNOFF [6]. Er zeigte, daß Unabhängigkeit im Limes nicht auftritt, wenn der Vektor (ξ_h, ξ_k) im Limes einer zweidimensionalen Normalverteilung unterworfen ist. Wir zeigen nun, daß Rand- und Zentralglieder stets im Limes unabhängig voneinander sind. In Bezug auf Formulierung und Beweis von Satz 1 verdanke ich Herrn Professor A. RÉNYI wertvolle Hinweise.

Satz 1. *Die Rangnummern h und k mögen derart von n abhängen, daß $\lim_{n \rightarrow \infty} \frac{h}{k} = 0$. Dann sind die Ranggrößen ξ_h und ξ_k im Limes unabhängig, d. h. für*

$$\sigma_n = \sup_{\substack{-\infty < u < \infty \\ -\infty < v < \infty}} |\mathbf{P}(\xi_h < u, \xi_k < v) - \mathbf{P}(\xi_h < u) \mathbf{P}(\xi_k < v)|$$

gilt $\lim_{n \rightarrow \infty} \sigma_n = 0$. In derselben Weise sind auch ξ_{n+1-h} und ξ_{n+1-k} im Limes voneinander unabhängig.

¹ Berlin.

Beweis. Die letzte Behauptung läßt sich auf die erste zurückführen, indem man die Zufallsgrößen $\bar{x}_i = -x_i$ heranzieht, die auf die Variationsreihe $\bar{\xi}_i = -\xi_{n+1-i}$ ($i = 1, \dots, n$) führen.

Wir definieren neue unabhängige Zufallsgrößen durch die Transformation

$$y_i = -\log(1 - F(x_i)) \quad (i = 1, \dots, n).$$

Sie geben Anlaß zu der Variationsreihe

$$\eta_i = -\log(1 - F(\xi_i)) \quad (i = 1, \dots, n).$$

Mit \mathfrak{M} bezeichnen wir die Menge der Zahlen y , für die mindestens eine Zahl \hat{y} existiert, so daß $y = -\log(1 - F(\hat{y}))$. Für $y \in \mathfrak{M}$ gilt

$$\begin{aligned} \mathbf{P}\{y_i < y\} &= G(y) = \mathbf{P}\{-\log(1 - F(x_i)) < -\log(1 - F(\hat{y}))\} = \\ &= \mathbf{P}\{x_i < \hat{y}\} = F(\hat{y}) = 1 - e^{-y}. \end{aligned}$$

Alle Zahlen $y \notin \mathfrak{M}$ liegen in Konstanzintervallen von $G(y)$. Wenn $F(x)$ stetig ist, haben somit die Zufallsgrößen y_i eine exponentielle Verteilungsfunktion. Im anderen Falle bildet der Wertevorrat von $G(y)$ nur eine echte Teilmenge des Intervalls $[0, 1]$. Nun ist

$$\begin{aligned} \sigma_n &= \sup_{u, v} |\mathbf{P}\{\eta_h < -\log(1 - F(u)), \eta_k < -\log(1 - F(v))\} - \\ &\quad - \mathbf{P}\{\eta_h < -\log(1 - F(u))\} \mathbf{P}\{\eta_k < -\log(1 - F(v))\}| = \\ &= \sup_{x, y \in \mathfrak{M}} |\mathbf{P}\{\eta_h < x, \eta_k < y\} - \mathbf{P}\{\eta_h < x\} \mathbf{P}\{\eta_k < y\}|. \end{aligned}$$

Die Funktion, deren Supremum hier gebildet wird, läßt sich aber in der Form $Q_{hk}(G(x); G(y))$ darstellen. Die Verteilungsfunktion von η_k unter der Bedingung $\eta_h = u$ ist nämlich gleich der Verteilungsfunktion der Ranggröße mit der Nummer $k - h$ in einer Stichprobe vom Umfang $n - h$ mit der Grundverteilung

$$\frac{G(y) - G(u)}{1 - G(u)} \quad (y \geq u)$$

(vgl. [5], § 2). Daher gilt für $x \leq y$

$$\begin{aligned} \mathbf{P}\{\eta_h < x, \eta_k < y\} &= \int_0^x \mathbf{P}\{\eta_k < y | \eta_h = u\} d\mathbf{P}\{\eta_h < u\} = \\ &= \binom{n}{h} \binom{n-h}{k-h} \int_0^{G(x)} \left\{ \int_0^{\frac{G(y)-w}{1-w}} (1-s)^{n-k} ds^{k-h} \right\} d \int_0^w (1-t)^{n-h} dt^h \end{aligned}$$

und für $x > y$

$$\mathbf{P}(\eta_h < x, \eta_k < y) = \mathbf{P}\{\eta_k < y\} = \binom{n}{k} \int_0^{G(y)} (1-t)^{n-k} dt^k.$$

Führen wir unabhängige Zufallsgrößen z_i mit $\mathbf{P}(z_i < z) = 1 - e^{-z}$ ($z \geq 0$, $i = 1, \dots, n$) ein, die die Variationsreihe ζ_1, \dots, ζ_n liefern, setzen wir ferner

$$\mathbf{P}\{\zeta_h < x, \zeta_k < y\} - \mathbf{P}\{\zeta_h < x\} \mathbf{P}\{\zeta_k < y\} = R_{hk}(x, y),$$

so gilt die Abschätzung

$$\sigma_n \leq \sup_{x,y} |R(1 - e^{-x}, 1 - e^{-y})| = \sup_{x,y} |R_{hk}(x, y)|.$$

Damit ist unser allgemeines Problem auf ein viel einfacheres zurückgeführt, das wir mit Hilfe des folgenden Lemmas behandeln werden.

Lemma 1. Wenn $h \rightarrow \infty, \frac{h}{k} \rightarrow 0, n - k \rightarrow \infty$, so ist gilt gleichmäßig bezüglich x und y

$$R_{hk}(x, y) = \mathbf{P}\{\zeta_h < x, \zeta_k < y\} - \mathbf{P}\{\zeta_h < x\} \mathbf{P}\{\zeta_k < y\} \rightarrow 0.$$

Beweis. Da die Differenz $\zeta_k - \zeta_h$ unabhängig von ζ_h ist, (s. z. B. [4]), gilt

$$\begin{aligned} R_{hk}(x, y) &= \int_0^x [\mathbf{P}\{\zeta_k < y \mid \zeta_h = u\} - \mathbf{P}\{\zeta_k < y\}] d\mathbf{P}\{\zeta_h < u\} = \\ (1) \quad &= \int_0^x [\mathbf{P}\{\zeta_k - \zeta_h < y - u\} - \mathbf{P}\{\zeta_k < y\}] d\mathbf{P}\{\zeta_h < u\}. \end{aligned}$$

Man erkennt leicht, daß $R_{hk}(x, y) \geq 0$ für alle x und y .

A. RÉNYI [4] hat eine Methode eingeführt, die es gestattet, Grenzwertsätze für Ranggrößen in einer sehr eleganten Weise zu gewinnen. Wir sind jetzt in der Lage, sie anzuwenden. Es gilt die Darstellung

$$(2) \quad \zeta_k = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \dots + \frac{\delta_k}{n+1-k},$$

wo die Zufallsgrößen $\delta_1, \dots, \delta_k$ unabhängig sind und die Verteilungsfunktion $\mathbf{P}\{\delta_j < y\} = 1 - e^{-y}$ ($y \geq 0$) haben, so daß für ihre mathematischen Erwartungen und Dispersionen

$$\mathbf{M} \delta_j = \mathbf{D}^2 \delta_j = 1 \quad (j = 1, \dots, k)$$

gilt. Daraus folgt

$$\begin{aligned} (3) \quad M_k &= \mathbf{M} \zeta_k = \sum_{i=n+1-k}^n \frac{1}{i} = M_{kh} + M_h \\ S_k^2 &= \mathbf{D}^2 \zeta_k = \sum_{i=n+1-k}^n \frac{1}{i^2} = S_{kh}^2 + S_h^2, \end{aligned}$$

und wegen $\mathbf{M} |\delta_j - 1|^3 = \int_0^\infty |x - 1|^3 e^{-x} dx < \int_0^\infty e^{-x} dx + \int_0^\infty x^3 e^{-x} dx \leq 5$ gilt ferner

$$T_k^3 = \sum_{i=1}^k \mathbf{M} \left| \frac{\delta_i - 1}{n + 1 - i} \right|^3 \leq 5 \sum_{i=n+1-k}^n \frac{1}{i^3} < 5 \int_{n-k}^n \frac{dt}{t^3} < \frac{5k}{n(n-k)^2}.$$

Durch derartige Integralabschätzungen (nach unten und oben) der Summen M_h und S_k^2 findet man weiter ($0 < \vartheta < 1$)

$$(4) \quad M_h = (1 + o(1)) \log \frac{n}{n-h} = \frac{h}{n} (1 + o(1))$$

$$S_k^2 = \frac{k}{(n-k)n} \left(1 + \vartheta \frac{n}{(n-k)k} \right) = \frac{k}{(n-k)n} (1 + o(1))$$

und somit

$$\frac{T_k^3}{S_k^3} \leq 5 \sqrt{\frac{n}{(n-k)k}} \rightarrow 0.$$

Daher läßt sich auf die Summe (2) der Satz von Ljapunoff anwenden, d. h. es gilt

$$(5) \quad \lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\zeta_k - M_k}{S_k} < x \right\} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Wegen $M_{kh} = \mathbf{M}(\zeta_k - \zeta_h) = \sum_{i=n+1-k}^{n-h} \frac{1}{i}$, $S_{kh}^2 = \mathbf{D}^2(\zeta_k - \zeta_h) = \sum_{i=n+1-k}^{n-h} \frac{1}{i^2}$ braucht man in den obigen Rechnungen nur n und k durch $n-h$, bzw. $k-h$ zu ersetzen, um mit Hilfe von $\frac{h}{k} \rightarrow 0$ die Beziehungen

$$(6) \quad S_{kh}^2 = S_k^2 (1 + o(1))$$

und

$$(7) \quad \lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\zeta_k - \zeta_h - M_{kh}}{S_{kh}} < x \right\} = \Phi(x)$$

zu gewinnen. Da die Konvergenz in (5) und (7) gleichmäßig bezüglich x ist, ergibt sich aus (1)

$$(8) \quad R_{hk}(x, y) = \int_0^x \left[\Phi\left(\frac{y-u-M_{kh}}{S_{kh}}\right) - \Phi\left(\frac{y-u-M_k}{S_k}\right) \right] d\mathbf{P}(\zeta_h < u) + r_n(x, y) =$$

$$= J_n(x, y) + r_n(x, y)$$

und das Restglied r_n konvergiert gleichmäßig in x und y gegen 0.

Bei beliebigem $y \geq 0$ bestimmen wir die Zahl $\hat{x}(y)$ durch die Gleichung

$$\frac{y - \hat{x} - M_{kh}}{S_{kh}} = \frac{y - M_k}{S_k}, \quad \text{oder} \quad \hat{x} = \frac{y - M_k}{S_k} (S_k - S_{kh}) + M_h.$$

Nach (3) gilt $S_k - S_{kh} > 0$. Falls $\frac{y - M_k}{S_k} \rightarrow -\infty$ und dabei $\hat{x} < 0$, folgt

wegen $\frac{y - M_{kh}}{S_{kh}} < \frac{y - M_k}{S_k}$ aus (8) die Behauptung des Lemmas. Wenn

$\frac{y - M_k}{S_k} > -C^2$, ist für genügend große n $\hat{x} > 0$, denn mit (4) und (6) folgt $\frac{S_k - S_{kh}}{M_h} = \frac{S_h^2}{M_h(S_k + S_{kh})} < k^{-\frac{1}{2}}$. Für $\hat{x} > 0$ aber ist bei festem y $\sup_x J_n(x, y) = J_n(\hat{x}, y)$. Wenn $\frac{y - M_k}{S_k} \rightarrow \infty$, folgt nun aus (8) unmittelbar die Behauptung.

Es bleibt noch übrig, den Fall $\frac{y - M_k}{S_k} < C^2$, $\hat{x} > 0$ zu betrachten. Bei beliebigem $\eta > 0$ gilt die Abschätzung

$$(9) \quad J_n(\hat{x}, y) = \int_0^{\hat{x} - \eta S_{kh}} \dots + \int_{\hat{x} - \eta S_{kh}}^{\hat{x}} \dots \leq \mathbf{P}\{\zeta_h < \hat{x} - \eta S_{kh}\} + \left[\Phi\left(\frac{y - M_k}{S_k} + \eta\right) - \Phi\left(\frac{y - M_k}{S_k}\right) \right].$$

Das zweite Glied kann mit η gleichmäßig beliebig klein gemacht werden; das erste aber ist kleiner als $\mathbf{P}\left\{\frac{\zeta_h - M_h}{S_h} < C \frac{S_k - S_{kh}}{S_h} - \eta \frac{S_{kh}}{S_h}\right\}$ und verschwindet nach dem Satz von Ljapunoff für $n \rightarrow \infty$. Die Gleichung (5) gilt nämlich auch mit h an Stelle von k , und ebenso können wir in (4) k durch h ersetzen. Daher folgt mit Hilfe von (6) für genügend große n

$$\frac{S_k - S_{kh}}{S_h} = \frac{S_k^2 - S_{kh}^2}{S_h(S_k + S_{kh})} = \frac{S_h}{2S_k} (1 + o(1)) < \sqrt{\frac{h}{k}} \rightarrow 0$$

und

$$\frac{S_{kh}}{S_h} = \frac{S_k}{S_h} (1 + o(1)) \rightarrow \infty.$$

Damit ist das Lemma bewiesen.

Jetzt müssen wir uns noch von den Voraussetzungen $h \rightarrow \infty, n - k \rightarrow \infty$ befreien. Wir beginnen mit der letzteren, betrachten also die Funktion $R_{hl}(x, y)$ für $h \rightarrow \infty, \frac{h}{l} \rightarrow 0, n - l \rightarrow \infty$. Dann existiert eine Zahlenfolge $k(n)$, so daß $\frac{h}{k} \rightarrow 0, n - k \rightarrow \infty, k \leq l$. Da die Ranggrößen ζ_1, \dots, ζ_n eine Markoffsche Kette bilden, sind ζ_h und ζ_l unter der Bedingung $\zeta_k = u$ unabhängig, und es gilt

$$R_{hl}(x, y) = \int_0^y \mathbf{P}\{\zeta_l < x \mid \zeta_k = u\} [\mathbf{P}\{\zeta_h < x \mid \zeta_k = u\} - \mathbf{P}\{\zeta_h < x\}] d\mathbf{P}\{\zeta_k < u\}.$$

Natürlich brauchen wir nur Werte von y zu betrachten, für die der Integrand im Intervall $(0, y)$ positiv ist. Für solche Werte aber gilt

$$R_{hl}(x, y) \leq \int_0^y [\mathbf{P}\{\zeta_h < x \mid \zeta_k = u\} - \mathbf{P}\{\zeta_h < x\}] d\mathbf{P}\{\zeta_k < u\} = R_{hk}(x, y).$$

Auf Grund des Lemmas 1 ist damit der betrachtete Fall erledigt. Mit Hilfe dieses Resultats läßt sich aber in der gleichen Weise auch der Fall $h \leftrightarrow \infty$ behandeln. Damit ist Satz 1 bewiesen.

(Eingegangen: 28. August 1960; in umgearbeiteter Form: 1. April 1962.)

LITERATURVERZEICHNIS

- [1] GEFFROY, J.: Publ. Inst. Stat. Paris 1958 und 1959.
- [2] GUMBEL, E. J.: Ann. Math. Stat. 1946.
- [3] HOMMA, T.: Reports of statistical application research 1, Tokio, 1951.
- [4] RÉNYI, A.: Acta Math. Acad. Sci. Hung. 1953.
- [5] ROBBERG, H. J.: Math. Nachrichten 1960.
- [6] SMIRNOFF, N. W.: Bull. de l'Université d'État à Moscou, Ser. Inter., Section B, Math. et Méc. 1, 1937.

ОБ АСИМПТОТИЧЕСКОМ ПОВЕДЕНИИ ЦЕНТРАЛЬНОГО И КРАЙНЕГО ЧЛЕНОВ ПОРЯДКОВЫХ СТАТИСТИК

Н. J. ROBBERG

Резюме

Используя метод А. РЭНИИ автор доказывает, что упорядоченные статистики ξ_h и ξ_k являются асимптотически независимыми, если $n \rightarrow \infty$, $\frac{h}{k} \rightarrow 0$.

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

by
ALAJOS KREM

I. Introduction

Let $\xi_1, \xi_2, \dots, \xi_n$ 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

$$\xi_k^* = R_k(\xi_1, \xi_2, \dots, \xi_n) \quad (1 \leq k \leq n)$$

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

$$\xi_1^* \leq \xi_2^* \leq \dots \leq \xi_n^*$$

holds.

We shall investigate the independence in limit of the variables ξ_k^* as $n \rightarrow \infty$. 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 $\frac{h}{n} \rightarrow \lambda_1, \frac{k}{n} \rightarrow \lambda_2, 0 < \lambda_1 < \lambda_2$, 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. ROBBERG has proved the asymptotic independence of ξ_h^* and ξ_k^* under the condition $\frac{h^2}{k} \rightarrow 0$ 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 $\frac{h}{k} \rightarrow 0$. In the mean time H. J. ROBBERG 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

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

it follows that

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