ON GALTON'S RANK ORDER TEST

KÁROLY SARKADI

§ 1. Introduction

In this note we shall give a simple proof for a theorem of CHUNG and FELLER [1], as well as for its generalized forms, and make some remarks on some connected problems.

Let us consider a random sequence of k $(+1)$'s and k (-1)'s and denote it[1] by $\mathbf{C}$. There are $\binom{2k}{k}$ possible sequences C and each of them is supposed to be equiprobable. Let $S_i = S_i\ (C)$ be the sum of the first i members of C. Let us denote by $N'_k = N'_k(C)$ the number of the positive members in the sequence $S_1, S_3, \ldots, S_{2k-1}$($N'_k$ is the so-called Galton-statistic).

The theorem of CHUNG and FELLER [1] states: $(\mathbf{N}'_k = N'_k(\mathbf{C}))$.

Theorem 1.

$$\mathsf{P}\{\mathbf{N}'_k = i\} = \frac{1}{k+1} \qquad (i = 0,1,\ldots,k)\ .$$

This theorem was extended first by GNEDENKO and MIHALEVICH. They considered a random sequence of $k = mp$ $(+1)'$ s and m $(-p)'$s where p is integer. S_i being defined as before, let us denote by N'_k the number of the positive members in the sequence

$$\begin{array}{lll} S_1, & S_2, & \ldots, S_p, \\ S_{p+2}, & S_{p+3} & \ldots, S_{2p+1} \\ S_{2p+3}, & S_{2p+4}, & \ldots, S_{3p+2} \\ \multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots} \\ S_{k+m-p}, & S_{k+m-p+1}, & \ldots, S_{k+m-1} \end{array}$$

The theorem of GNEDENKO and MIHALEVICH [2] states:

[1] Random variables will be distinguished from their actual values by printing their symbols in bold type.

Theorem 2.

$$\mathbf{P}\{\mathbf{N}'_k = i\} = \frac{1}{k+1} \qquad (i = 0,1,\ldots,k).$$

In 1953, SPARRE ANDERSEN proved the following theorem:

Let be $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ symmetrically dependent random variables[2)] with $\sum_{i=1}^{n} \mathbf{X}_i = 0$. Let us denote by $\mathbf{S}_i = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_i$ and let be $\mathbf{P}\{\mathbf{S}_1 \mathbf{S}_2 \ldots \mathbf{S}_{n-1} = 0\} = 0$. Let be $\mathbf{N}_n$ defined as the number of positive members in the sequence $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{n-1}$. Then([3], Theorem 2/1°)

Theorem 3.

$$\mathbf{P}\{\mathbf{N}_n = i\} = \frac{1}{n} \qquad (i = 0,1,\ldots,n-1).$$

We wish to mention that although Theorem 3 is closely connected to theorems 1 and 2 it is no generalization in the proper sense.

Several proofs are known for Theorem 1 [4], [5], the simplest one due to HODGES [6]. However, even this proof seems to be complicated compared to the simplicity of the theorem itself.

For the theorem of SPARRE ANDERSEN, SPITZER [7] gave an elegant, simple proof, but his method cannot be applied to prove the theorem of CHUNG and FELLER, nor that of GNEDENKO and MIHALEVICH.

In § 2 we give a method for proving each of the three theorems which is not more complicated than that of SPITZER.

In § 3 we consider some connected problems.

§ 2. The new proofs

Consider first the theorem of CHUNG and FELLER. Our proof — as well as that of HODGES — is based on a one by one transformation mapping the sequences C into sequences C' for which $N'_k(C') = N'_k(C) - 1$.

Let be $N_k(C) \geq 1$. We divide C into three subsequences of consecutive members: C_1, C_2, C_3. C_1 is the shortest beginning subsequence of C, which sums to 1 and C_2 the shortest subsequence immediately following C_1 which sums to —1. Our transformation consists of changing the order of the subsequences C_1 and C_2. (Fig. 1).

In the transformed sequence C_2 will be the shortest beginning subsequence which sums to —1 and C_1 the shortest subsequence immediately following C_2 which sums to 1, thus the transformation is a one by one.

Further let be $2x+1$ the length of C_1 and $2y+1$ the length of C_2. Then the first x members of the sequence $S_1(C)$, $S_3(C)$, $\ldots$, $S_{2k-1}(C)$ are negative, $S_{2x+1} = 1$ and the following y members are positive. On the other hand, the sequence $S_1(C')$, $S_3(C')$, $\ldots$, $S_{2k-1}(C')$ has y positive members at the beginning, $S_{2y+1} = -1$ and the following x members are negative. The

2) This means that their joint distribution function is a symmetrical function of its arguments [3]. Also the terms "equivalent random variables", "interchangeable random variables" are used for this notion in the literature.

remaining subsequences are identical from which $N_k'(C') = N_k'(C) - 1$ follows.

Now we turn to the theorem of SPARRE ANDERSEN. It is proved first the special case if the components of the vector $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ are the random permutations of the constants $\alpha_1, \alpha_2, \ldots, \alpha_n$. In this case let be C_1 the shortest beginning subsequence which sums to a positive value and C_2 the following subsequence having a negative partial sum. We obtain in the same manner as before that the change of order of C_1 and C_2 diminishes

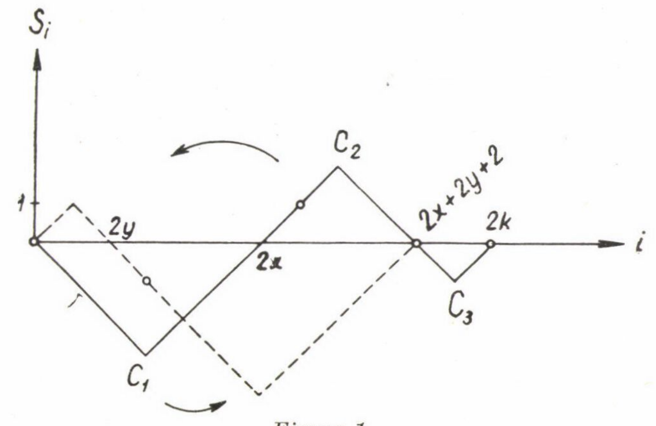

Figure 1.

the number of positive members in the sequence $S_1, S_2, \ldots, S_{n-1}$ by unity. Then the general form of the theorem can be proved by stating that the conditional distribution of $\mathbf{N}_n$ with respect to the order statistics of the $\mathbf{X}_i$'s does not depend on the condition.

Now let us consider the theorem of GNEDENKO and MIHALEVICH. We prove the following generalization of this theorem:

Let be $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ symmetrically dependent random variables with $\sum_{i=1}^{n} \mathbf{X}_i = 0$; let be always exactly k of the $\mathbf{X}_i$'s equal to 1, and let the other $n-k$ take on negative integral values. Let us denote by $R_i = X_1 + X_2 + \ldots + X_{r_i}$ where r_i is the rank of the i-th $+1$ in the sequence $X_1, X_2, \ldots, X_n$.

Let be N_k' defined as the number of positive members in the sequence $R_1, R_2, \ldots, R_k$. Then we have

Theorem 4.

$$\mathbf{P}\{\mathbf{N}_k' = i\} = \frac{1}{k+1}$$

$$(i = 0, 1, \ldots, k).$$

Evidently if we apply in this case the same transformation and if the subsequences C_1 and C_2 contained x and y $(+1)$'s, respectively, i. e. the sequence $\{R_i\}$ began with $x-1$ negative terms and $y+1$ positive terms then after applying the above transformation it will begin with y positive and x negative terms which proves Theorem 4.

§ 3. Connected problems

Let us consider again the random sequence of k $(+1)$'s and k (-1)'s Let $T_k = T_k(C)$ be the first maximum place of $S_i(S_0 = 0)$. Another theorem of SPARRE ANDERSEN ([3], formula 5.10) states that

Theorem 5.

$$P\{\mathbf{T}_k = 2i\} = P\{\mathbf{T}_k = 2i+1\}$$

$$(i = 0,1,\ldots,n-1)$$

If we exclude the case $T_k = 0$, i. e. the first positive maximum place is considered, then the sign $+$ on the right side changes to $-$. This form of the theorem is due to VINCZE [8].

A modified, symmetric form of these theorems is the following:

Let be $\mathbf{U}_k$ a randomly chosen value from the maximum places (0 and $2k$ are regarded as the same place). Then we have

Theorem 6.

$$P\{\mathbf{U}_k = i\} = \frac{1}{2k} \qquad (i = 0,1,\ldots,2k-1).$$

This theorem can be simply proved by the aid of the transformation used by SPARRE ANDERSEN. This transformation consists of transposing the first element of the sequence to the end. Evidently this transformation shiftes every maximum place of S_k to left by unity *mod* $2k$, which proves theorem 6. Theorem 5 can be also simply verified in this way by stating that the transformation shiftes the first maximum place by unity if it was ≥ 1, on the other hand it will be odd if it was 0 before the transformation.

The generalization for the sequence of symmetrically dependent random variables ([3], Theorem 2/2°) can be treated in the same way.

(Received September 15, 1960.)

REFERENCES

[1] CHUNG, K. L.—FELLER, W.: „Fluctuations in coin tossing." *Proc. Nat. Acad. Sci. USA* **35** (1943) 605—608.

[2] Гнеденко, Б. В. — Михалевич, В. С.: «Две теоремы о поведении эмпирических функции распределения». *Доклады АН СССР* **85** (1952) 25—27.

[3] SPARRE ANDERSEN, E.: "On the fluctuations of sums of random variables." *Math. Scand.* **1** (1953) 263—285.

[4] FELLER, W.: *An Introduction to Probability Theory and its Applications*. 1. ed Wiley, New York, 1950; Chapter 12, problem 4., p. 263.
[5] Гнеденко, Б. В. — Михалевич, В. С.: «О распределении числа выходов одной эмпирической функции распределения над другой.» *Доклады АН СССР* **82** (1952) 841—843.
[6] HODGES, J. L.: ,,Galton's rank order test." *Biometrika* **42** (1952) 261—262.
[7] SPITZER, F.: ,,A combinatorial lemma and its application to probability theory." *Trans. Am. Math. Soc.* **82** (1956) 323—339.
[8] VINCZE, I.: ,,On some joint distributions and joint limiting distributions in the theory of order statistics, II." *Publ. Math. Inst. Hung. Acad. Sci.* **4** (1959) 29—47.

О РАНГОВОМ КРИТЕРИЙ GALTON-А

K. SARKADI

Резюме

В статье дается новое и простое доказательство одной теоремы CHUNG-а и FELLER-а [1], ее обобщения происходящего от ГНЕДЕНКО и МИХАЛЕВИЧ-а [2] а также одной теоремы SPARRE ANDERSEN-а [3], высказывающих равномерное распределение критерии GALTON-а и его обобщения. Сущность доказательства состоит в том, что фигурирующая там последовательность разлагается на три последовательности: на наиболее короткую начальную последовательность C_1 сумма которой положительна, на наиболее короткую подпоследовательность C_2, состоящую из элементов последующих после C_1 таких, что ее сумма отрицательна и на остальную подпоследовательность C_3. Перемена местами C_1 и C_2 уменьшает значение критерии GALTON-а на единицу и примененное при этом преобразование взаимно-однозначно.