ON SOME PROBLEMS CONNECTED WITH THE GALTON-TEST

by

E. CSÁKI and I. VINCZE

Introduction

One of the oldest two-sample tests is that proposed by GALTON to DARWIN (see [6]), without however knowing the distribution of his statistic. As far as we know this distribution was determined for the first time in the work of K. L. CHUNG and W. FELLER [1]. Papers [2], [3], [4] and [5] are considering the same problem and its generalization respectively. As the Galton-test does not appear to be powerful, present paper aims at improving the test both by modification of the Galton-statistic and by forming a pair of statistics. Our considerations are closely connected with the method of N. V. SMIRNOV used for the determination of limiting distribution of his two sample test [9]. In this paper for equal sample-sizes some exact joint distributions are determined by elementary methods and the corresponding limiting distributions are determined as well. We wish to consider statistical problems and the case of different sample-sizes in a following paper.

Notations. Let us denote by $\xi_1, \xi_2, \ldots, \xi_n$ and $\eta_1, \eta_2, \ldots, \eta_n$ samples taken from populations with the common continuous distribution functions $F(x)$ and $G(x)$ resp. Let us arrange these samples in order of magnitude:

$$(1) \qquad \begin{aligned} &\xi_1^* < \xi_2^* < \ldots < \xi_n^*, \\ &\eta_1^* < \eta_2^* < \ldots < \eta_n^*. \end{aligned}$$

We introduce further the union of these ordered samples:

$$\zeta_1^* < \zeta_2^* < \ldots < \zeta_{2n}^*$$

and the random variables

$$\vartheta_i = \begin{cases} +1, & \text{if} \quad \zeta_i^* = \xi_j \\ -1, & \text{if} \quad \zeta_i^* = \eta_k. \end{cases}$$

The partial sum of the ϑ_i -s is denoted by s_i i. e.

$$s_i = \vartheta_1 + \vartheta_2 + \ldots + \vartheta_i, \; s_0 = 0, \quad i = 0, 1, 2, \ldots, 2n.$$

Under the assumption $F(x) \equiv G(x)$ each array $(\vartheta_1, \vartheta_2, \ldots, \vartheta_{2n})$ of the $n\,(+1)$ -s and $n\,(-1)$ -s has the same probability $\binom{2n}{n}^{-1}$. Each array corresponds to a random path starting at and returning after $2n$ steps to the origin.

Each path has the same probability. If the points (i, s_i) are represented in the plane and each of them is connected with the next one, then we obtain the usual illustrative figure of the paths. We shall denote by

E_{2n} a path from $(0, 0)$ to $(2n, 0)$,

T_i a point of the path, where either $(s_{i-1} = -1,\ s_i = 0,\ s_{i+1} = +1)$ or $(s_{i-1} = +1,\ s_i = 0,\ s_{i+1} = -1)$ occurs, and we join to these points $T_0 = (0,0)$, $T_{2n} = (2n, 0)$.

T_i^k a point of the path, where either

$$(s_{i-1} = k-1,\ s_i = k,\ s_{i+1} = k+1), \text{ or } (s_{i-1} = k+1,\ s_i = k,\ s_{i+1} = k-1)$$

holds.

We shall call in the following the points T_i and T_i^k intersection points.

E_{2n}^l an E_{2n}-path containing exactly $l+1$ T_i points; with other words an E_{2n}^l-path has l waves (halfwaves),

$E_{2n}^{g,l}$ an E_{2n}-path having $l+1$ T_i points and $2g$ steps above the axis.

$E_{2n,k}^l$ an E_{2n}-path having $l+1$ T_i^k points,

$E_{2n,k}^{g,l}$ an $E_{2n,k}^l$-path having $2g$ steps above the height k,

H_m^k a path starting at the origin and reaching for the first time the height k at the m-th step,

$N(A)$ the number of A paths or points $\left(\text{e.g. } N(E_{2n}) = \binom{2n}{n} \text{ or for an } E_{2n}^l\text{-path } l+1 = N(T_i).\right)$.

§ 1. The Galton statistic and the number of waves

1. We shall give two proofs for the following

Theorem 1.1

$$N(E_{2n}^l) = \frac{2l}{n}\binom{2n}{n-l} \qquad l = 1, 2, \ldots, n. \tag{2}$$

First proof. As it is known the number of E_{2m} paths, going throughout under or above the axis, is

$$\frac{1}{m+1}\binom{2m}{m}.$$

In consequence of this the number of E_{2n}-paths with intersection points $T_0, T_{2\alpha_1}, T_{2(\alpha_1+\alpha_2)}, \ldots, T_{2(\alpha_1+\alpha_2+\cdots+\alpha_{l-1})}, T_{2(\alpha_1+\alpha_2+\cdots+\alpha_l)}$ $(\alpha_1 + \alpha_2 + \ldots + \alpha_l = n,\ \alpha_i \geqq 1)$ is equal to

$$2\frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1}\frac{1}{\alpha_2+1}\binom{2\alpha_2}{\alpha_2}\cdots\frac{1}{\alpha_l+1}\binom{2\alpha_l}{\alpha_l}.$$

The factor 2 is due to the possibilities of starting in either positive or negative direction.

Hence we have

$$N(E_{2n}^l) = 2 \sum_{\substack{\alpha_1+\ldots+\alpha_l=n \\ \alpha_i \geqq 1}} \frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1} \cdots \frac{1}{\alpha_l+1}\binom{2\alpha_l}{\alpha_l}.$$

Let us denote the generating function of the $N(E_{2n}^l)$-s by $F_l(v)$, i. e.

$$F_l(v) = \sum_{n=1}^{\infty} N(E_{2n}^l)\, v^n. \tag{3}$$

Let us introduce further the notation for the known generating function

$$f(v) = \sum_{\alpha=1}^{\infty} \frac{1}{\alpha+1}\binom{2\alpha}{\alpha} v^\alpha = \frac{1-\sqrt{1-4v}}{2v} - 1. \tag{4}$$

As it is easy to see the relation

$$F_l(v) = 2\,[f(v)]^l = 2\, v^l \left(\frac{1-\sqrt{1-4v}}{2v}\right)^{2l}$$

holds. One of the authors [8] has determined the following generating function

$$\sum_{j=0}^{\infty} \binom{2l+2j}{j} \frac{l}{l+j} v^j = \left(\frac{1-\sqrt{1-4v}}{2v}\right)^{2l},$$

from which the relation

$$F_l(v) = 2 \sum_{j=0}^{\infty} \binom{2l+2j}{j} \frac{l}{l+j} v^{j+l} = 2 \sum_{n=1}^{\infty} \frac{l}{n} \binom{2n}{n-l} v^n$$

may be obtained, giving the proof of our theorem 1.1.

Second proof. There holds the following

Lemma.

$$N(E_{2n}^l) = 2\, N(H_{2n}^{2l}).$$

For the known relation (see eg. Feller [7] p. 71)

$$N(H_{2n}^{2l}) = \frac{l}{n}\binom{2n}{n-l}$$

the proof of the lemma gives us the proof of theorem 1.1 too.

As one half of the E_{2n}^l-paths is starting in the positive direction and the other half in the negative one, we may consider the paths with $s_1 = +1$ only. A one to one transformation of these paths into the H_{2n}^{2l}-paths will be given. This happens in the following way: Figure 1 shows a possible domain within which such $(E_{2n}^{l} \mid s_1 = +1)$ path must proceed. As it is seen each path is divided by the points T_i into l sections.

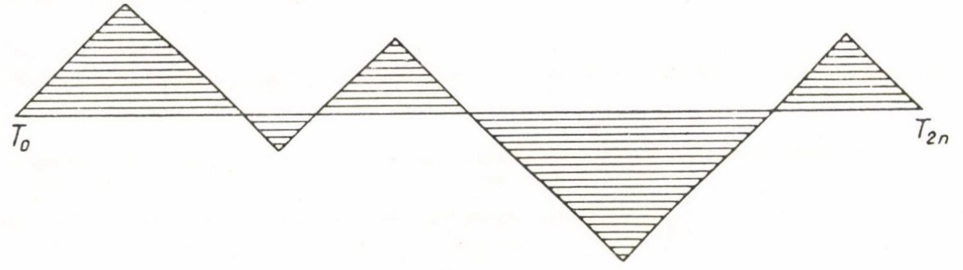

Figure 1.

Let us reflect the positive parts of the path (the positive waves) around the axis (see figure 2).

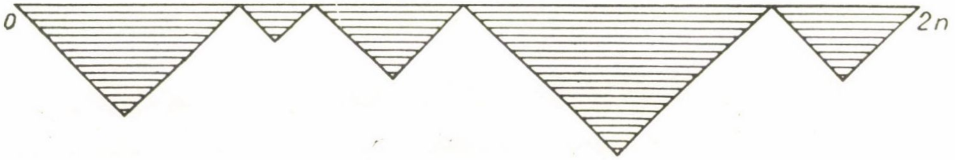

Figure 2.

Here we have for each $i = 1,\ 2\alpha_1 + 1,\ 2(\alpha_1 + \alpha_2) + 1, \dots s_i = -1$. Let us omit these first steps of each of the l sections and let us join to the end of each section a positive step. Figure 3 shows the domain containing the graph of a path after the mentioned modifications. Thus we obtained a H_{2n}^{2l} path.

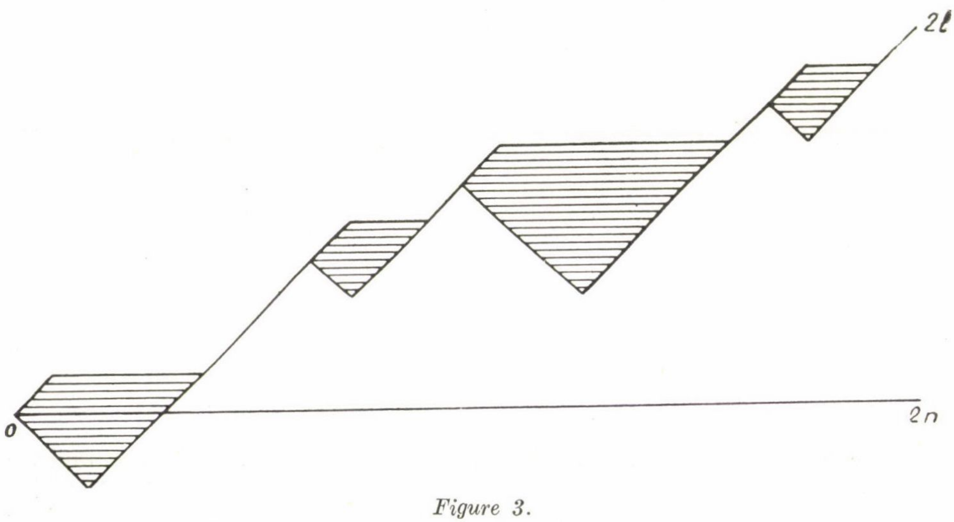

Figure 3.

The connection between a H_{2n}^{2l} and the corresponding $(E_{2n}^{l} \mid s_1 = +1)$ is given by the relation that the points where H_{2n}^{2l} reaches the height $2i$ for the first time correspond to the T_i points of E_{2n}^{l} $(i = 1, 2, \ldots, l)$. This construction may be carried out in the opposite direction as well and thus the proof of our lemma is given.

2. Let the random variable λ be the number of waves in the case of two samples defined in the above section. Then our theorem 1.1 gives imediately the following

Theorem 1.1'. *Under the null hypothesis*

$$\mathbf{P}(\lambda = l) = \frac{2l}{n} \frac{\binom{2n}{n-l}}{\binom{2n}{n}}, \qquad l = 1, 2, \ldots, n. \tag{5}$$

or

$$\mathbf{P}(\lambda < l) = 1 - 2\frac{\binom{2n-1}{n-l}}{\binom{2n}{n}}, \qquad l = 1, 2, \ldots, n.$$

For the limiting distribution we have

$$\lim_{n \to \infty} \mathbf{P}(\lambda < y\sqrt{2n}) = 1 - e^{-2y^2}, \qquad y \geqq 0. \tag{6}$$

As it can be seen the random variable λ is equal to the number of intersections the two empirical distribution functions. We wish to mention that a similar problem was considered by Mihalevich [4] but his definition of intersection is not equal to ours and he obtains different results.

3. Let us consider now the Galton-statistic. Let us denote it by γ, thus γ is equal to the number of ξ_i^*-s exceeding the corresponding $\eta_i^* (i = 1, 2, \ldots, n)$ and γ may be $0, 1, \ldots, n$ (see the array in (1) of section 1.) As it is known 2γ equals the "time" spent by the point walking randomly on the straight line above 0. The relation

$$\mathbf{P}(\gamma = g) = \frac{1}{n+1}, \qquad g = 0, 1, 2, \ldots, n.$$

is well known.

For the comparison of the two samples, i. e. for deciding whether the hypothesis $F(x) \equiv G(x)$ holds, or not we suggest the test based on the joint distribution of the pair of statistics $(\gamma; \lambda)$. In order to determine this distribution we prove the following

Theorem 1.2. *The number of $E_{2n}^{g,l}$-paths is equal to*

$$N(E_{2n}^{g,l}) =$$

$$= \begin{cases} \dfrac{l^2}{2g(n-g)}\dbinom{2g}{g-\frac{l}{2}}\dbinom{2n-2g}{n-g-\frac{l}{2}}, & \text{if } l \text{ is even}, \\ & 2 \leqq l \leqq 2g \leqq 2-l \\ \dfrac{l^2-1}{4g(n-g)}\left[\dbinom{2g}{g-\frac{l+1}{2}}\dbinom{2n-2g}{n-g-\frac{l-1}{2}}+\dbinom{2g}{g-\frac{l-1}{2}}\dbinom{2n-2g}{n-g-\frac{l+1}{2}}\right], & \\ & \text{if } l \text{ is odd}, \quad l-1 \leqq 2g \leqq 2n-l+1 \end{cases} \tag{7}$$

Proof. Using our above notations let be the ordinate of the T_i points $0, 2\alpha_1, 2(\alpha_1+\alpha_2), \ldots, 2(\alpha_1+\alpha_2+\ldots+\alpha_{l-1}), 2(\alpha_1+\alpha_2+\ldots+\alpha_l) = 2n$. If $s_1 = +1$, then $g = \alpha_1+\alpha_3+\alpha_5+\ldots$ must hold. Hence for the number of $E_{2n}^{g,l}$ paths starting in the positive direction we have

$$\sum{}^{*} \frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1}\cdots\frac{1}{\alpha_l+1}\binom{2\alpha_l}{\alpha_l}, \tag{8}$$

where Σ^* means summation for $\alpha_1+\alpha_3+\ldots = g$, $\alpha_2+\alpha_4+\ldots = n-g$ and $\alpha_i \geqq 1$, $i = 1, 2, \ldots, l$. In the same way we have for the number of $E_{2n}^{g,l}$ paths starting in negative direction

$$\sum{}^{**} \frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1}\cdots\frac{1}{\alpha_e+1}\binom{2\alpha_l}{\alpha_l}, \tag{9}$$

where now the summation holds for $\alpha_1+\alpha_3+\ldots = n-g$, $\alpha_2+\alpha_4+\ldots = g$ and $\alpha_i \geqq 1$, $i = 1, 2, \ldots, l$.

$N(E_{2n}^{g,l})$ equals the sum of the above two expressions (8) and (9).

Formula (8) may be written in the form

$$\left(\sum_{\alpha_1+\alpha_3+\ldots=g}\frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1}\frac{1}{\alpha_3+1}\binom{2\alpha_3}{\alpha_3}\cdots\right)\times \tag{8'}$$
$$\times\left(\sum_{\alpha_2+\alpha_4+\ldots=n-g}\frac{1}{\alpha_1+1}\binom{2\alpha_2}{\alpha_2}\frac{1}{\alpha_4+1}\binom{2\alpha_4}{\alpha_4}\cdots\right),$$

and expression (9) equals

$$\left(\sum_{\alpha_1+\alpha_3+\ldots=n-g}\frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha}\frac{1}{\alpha_3+1}\binom{2\alpha_3}{\alpha_3}\cdots\right)\times \tag{9'}$$
$$\times\left(\sum_{\alpha_2+\alpha_4+\ldots=g}\frac{1}{\alpha_2+1}\binom{2\alpha_2}{\alpha_2}\frac{1}{\alpha_4+1}\binom{2\alpha_4}{\alpha_4}\cdots\right).$$

If now l is even, then we have for the first factor in (8') the value $\frac{1}{2} N\left(E_{2g}^{\frac{l}{2}}\right)$ and for the second $\frac{1}{2} N\left(E_{2(n-g)}^{\frac{l}{2}}\right)$. But the same holds — in inverted order — for the factors in (9'). These give the statement of our theorem 1.2 for even l.

If l is odd, then we have for the two factors in (8'),

$$\sum_{\alpha_1+\alpha_3+\ldots=g} \frac{1}{\alpha_1+1}\binom{2\alpha_1}{\alpha_1}\frac{1}{\alpha_3+1}\binom{2\alpha_3}{\alpha_3}\cdots\frac{1}{\alpha_l+1}\binom{2\alpha_l}{\alpha_l} = \frac{1}{2} N\left(E_{2g}^{\frac{l+1}{2}}\right)$$

and

$$\sum_{\alpha_2+\alpha_4+\ldots=n-g} \frac{1}{\alpha_2+1}\binom{2\alpha_2}{\alpha_2}\frac{1}{\alpha_4+1}\binom{2\alpha_4}{\alpha_4}\cdots\frac{1}{\alpha_{l-1}+1}\binom{2\alpha_{l-1}}{\alpha_{l-1}} = \frac{1}{2} N\left(E_{2(n-g)}^{\frac{l-1}{2}}\right).$$

We have analogous expressions for the factors in (9') and thus we obtained the complete proof of our theorem 1.2.

The distribution of the pair of random variables is given by

Theorem 1.2'. *In the case $F(x) \equiv G(x)$*

$$(10)\quad \mathbf{P}(\gamma=g, \lambda=l) = \begin{cases} \dfrac{1}{\binom{2n}{n}}\left[\dfrac{l^2}{2g(n-g)}\binom{2g}{g-\frac{l}{2}}\binom{2n-2g}{n-g-\frac{l}{2}}\right], & \text{if } l \text{ is even} \\[2ex] \dfrac{1}{\binom{2n}{n}}\left[\dfrac{l^2-1}{4g(n-g)}\left\{\binom{2g}{g-\frac{l+1}{2}}\binom{2n-2g}{n-g-\frac{l-1}{2}}+\right.\right. \\ \qquad\left.\left.+\binom{2g}{g-\frac{l-1}{2}}\binom{2n-2g}{n-g-\frac{l+1}{2}}\right\}\right], & \text{if } l \text{ is odd} \end{cases}$$

here if $g = 0$, or n, then $l = 1$,

if $1 \leq g \leq n-1$, then $l = 2, 3, \ldots, \min\,(2g+1, 2n-2g+1)$

and for the limiting distribution

$$(11)\quad \lim_{n\to\infty} \mathbf{P}(\gamma \leq zn,\ \lambda \leq y\sqrt{2n}) = \sqrt{\frac{2}{\pi}} \int_0^y \int_0^z \frac{u^2}{[v(1-v)]^{3/2}} e^{-\frac{u^2}{2v(1-v)}}\, du\, dv\,.$$

As mentioned in the introduction the statistical questions of the test based on the above statistics will be treated in a second paper. We shall prove there that *the test based on the statistic λ is asymptotically consistent against all continuous alternatives.*

§ 2. Extension of the Galton statistic and the number of waves

1. In this paragraph the distinguished role of the height $k = 0$ i. e. the horisontal axis is abolished and the situation of the random path relative to the horisontal line of height $k>0$ is regarded. The number of intersections and the length of time spent above this height will be considered.

Theorem 2.1. *The number of $E^l_{2n,k}$-paths, is equal to*

$$N(E^l_{2n,k}) = \binom{2n+2}{n-k-l}\frac{k+l+1}{n+1}, \quad k>0, \quad l>0 \ \textit{odd}. \tag{12}$$

Proof. We shall give a one to one transformation of the $E^l_{2n,k}$-paths into the $H^{2k+2l+2}_{2n+2}$-paths. Let be r_1 the first and r_2 the last T^k_i point. (See figure 4.)

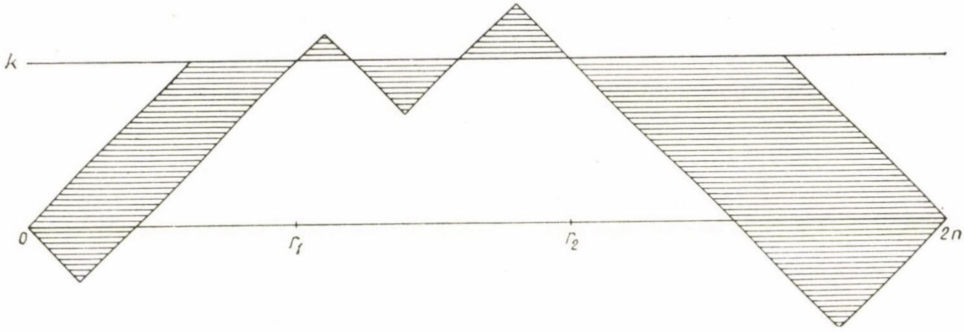

Figure 4.

Let us leave the section $(0, r_1)$ of the path unchanged. According to the second proof of theorem 1.1, the section between r_1 and r_2 corresponds to a path starting at the point (r_1, k) and reaching after $r_2 - r_1$ steps for the first time the height $k + 2l$. Concerning the section between r_2 and $2n$ let us first alter the signs and then the direction, i.e. we replace $\vartheta_{r_2}, \ldots, \vartheta_{2n}$ by $-\vartheta_{2n}, \ldots, \vartheta_{r_2}$, and let us attach this transformed section to the end of the previous one (see figure 5).

Finally let us now insert both between ϑ_{r_1} and ϑ_{r_1+1} and after ϑ_{2n} a $(+1)$. Thus we obtain a $H^{2k+2l+2}_{2n+2}$-path. By reversing this procedure it may be seen that this transformation is a one to one.

We wish to mention that by writing $|k|$ instead of k the formula (12) is valid in the case of negative k as well. From formula (12) there follows the following

Theorem 2.1′. *In the case $F(x) \equiv G(x)$ denoting by $\lambda_k + 1$ the number of $T^{(k)}_i$ points and by $\varkappa$ the maximal distance between the two empirical distribution function*

$$\varkappa = n \cdot \max_{(x)} \big(F_n(x) - G_n(x)\big) = \max_{0 \le i \le 2n} s_i$$

the relation

$$\mathbf{P}(\varkappa > k, \lambda_k = l) = \frac{k+l+1}{n+1}\,\frac{\binom{2n+2}{n-k-l}}{\binom{2n}{n}}, \quad k>0, \quad l \geq 1 \ \textit{odd}.$$

and for the limiting distribution

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a, \frac{\lambda_k}{\sqrt{2n}} < y\right) = e^{-2a^2} - e^{-2(a+y)^2}, \ a \geq 0, \ y \geq 0 \ \textit{holds}.$$

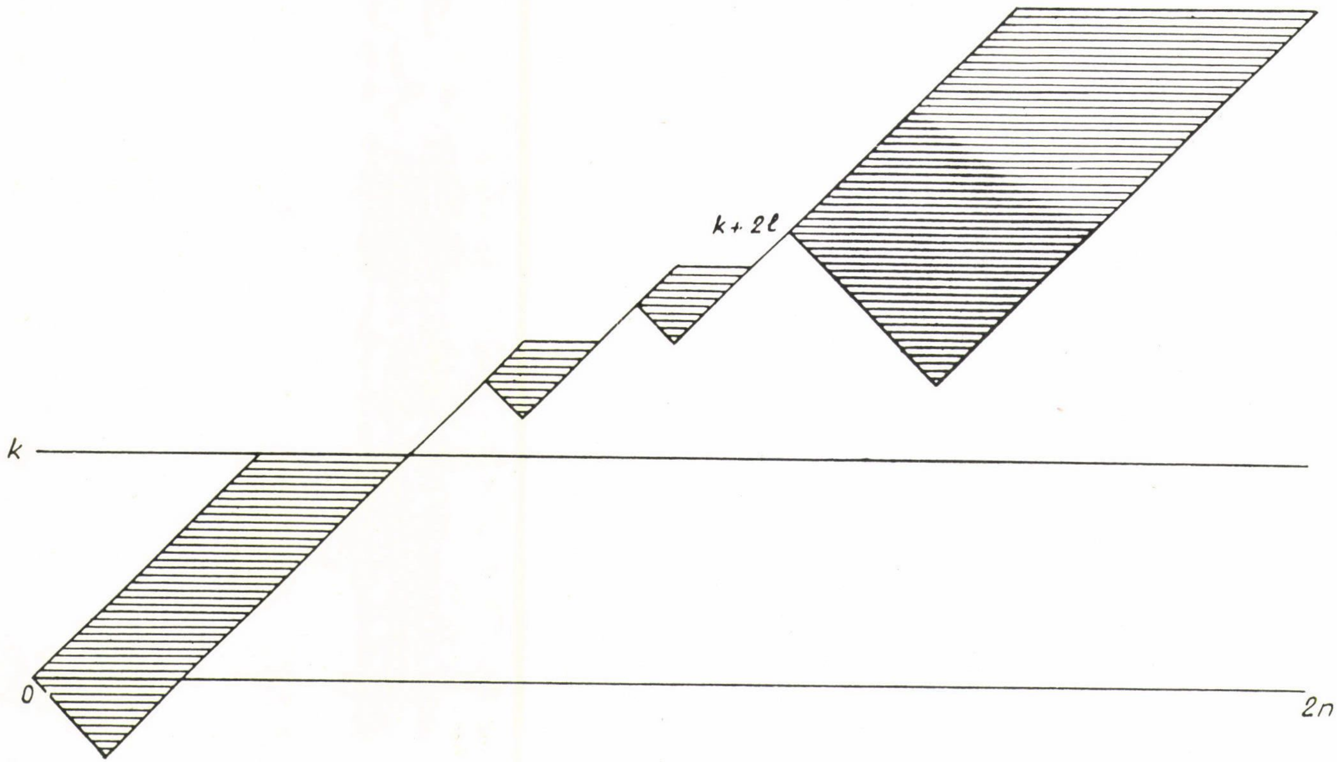

Figure 5.

2. Let us regard now the modification of the Galton-statistic.

As mentioned above the Galton-statistic γ is equal to the number of the positive members of the array

$$\xi_1^* - \eta_1^*, \xi_2^* - \eta_2^*, \ldots, \xi_i^* - \eta_i^*, \ldots, \xi_n^* - \eta_n^* .$$

If one of the samples is removed in one direction the number γ_k of the positive members of the array

$$\xi_{k+1}^* - \eta_1^*, \xi_{k+2}^* - \eta_2^*, \ldots, \xi_n^* - \xi_{n-k}^*$$

may be considered. It is easy to see that $2\gamma_k$ is equal to the "time" spent by the point walking randomly on the straight line above the height k. MIHALEVICH has derived an equivalent problem to this [4] and obtained a result equivalent to the following:

$$\mathbf{P}(\varkappa > k, \gamma_k = g) = \frac{1}{\binom{2n}{n}} \sum_{r=g}^{n-k} \binom{2r}{r} \frac{1}{r+1} \binom{2n-2r}{n+k-r} \frac{k}{n-r} . \tag{13}$$

3. We now prove the

Theorem 2.2.

$$N(E_{2n,k}^{g,l}) = \frac{(k+1)(l^2-1)}{4g} \binom{2g}{g - \frac{l+1}{2}} \sum_{(r)} \frac{1}{(r-g)(n+1-r)} \times \tag{14}$$
$$\times \binom{2(r-g)}{r-g-\frac{l-1}{2}} \binom{2n+2-2r}{n-r-k},$$

where the summation is for $2n - 2k \geqq 2r \geqq \max(2g + l - 1, 2l)$.

Proof. If r_1 and r_2 denote the first and the last $T_i^{(k)}$ points, resp., the section between r_1 and r_2 is an $E_{2r}^{g,l}$-path starting in positive direction where $2r = r_2 - r_1$.

The first section corresponds to a $H_{r_1+1}^{k+1}$-path, the last one to a $H_{2n-r_2+1}^{k+1}$-path.

Thus

$$N(E_{2n,k}^{g,l}) = \sum_{r_1, r_2} N(H_{r_1+1}^{k+1}) N(E_{r_2-r_1}^{g,l}, s_1 = +1) N(H_{2n-r_2+1}^{k+1}) =$$
$$= \sum_{(r)} \sum_{r_1=k}^{2n-2r-k} N(H_{r_1+1}^{k+1}) N(E_{2r}^{g,l}, s_1 = +1) N(H_{2n-2r-r_1+1}^{k+1}) =$$
$$= \sum_{(r)} \left[N(E_{2r}^{g,l}, s_1 = +1) \sum_{r_1=k}^{2n-2r-k} N(H_{r_1+1}^{k+1}) N(H_{2n-2r-r_1+1}^{k+1}) \right] =$$
$$= \sum_{(r)} \frac{l^2-1}{4g(r-g)} \binom{2g}{g-\frac{l+1}{2}} \binom{2(r-g)}{r-g-\frac{l-1}{2}} \sum_{r_1=k}^{2n-2r-k} \frac{k+1}{r_1+1} \times$$

$$\times \binom{r_1+1}{\frac{r_1+k}{2}+1} \frac{k+1}{2n-2r-r_1+1} \binom{2n-2r-r_1+1}{\frac{2n-2r-r_1+k}{2}+1},$$

but using the method of generating functions it can be seen that

$$\sum_{r_1=k}^{2n-2r-k} \frac{k+1}{r_1+1} \binom{r_1+1}{\frac{r_1+k}{2}+1} \frac{k+1}{2n-2r-r_1+1} \binom{2n-2r-r_1+1}{\frac{2n-2r-r_1+k}{2}+1} =$$

$$= \frac{k+1}{n-r+1} \binom{2n-2r+2}{n-r-k},$$

proving our Theorem 2.2.

Theorem 2.2′. *In the case* $F(x) \equiv G(x)$ *for the random variables* $\varkappa$, λ_k *and* γ_k *the joint distribution law*

$$\mathbf{P}(\varkappa > k, \gamma_k = g, \lambda_k = l) = \frac{1}{\binom{2n}{n}} N(E_{2n,k}^{g,l})$$

holds and we have the joint limiting distribution

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a, \frac{\lambda_k}{\sqrt{2n}} < y, \frac{\gamma_k}{n} < z\right) =$$

$$= \sqrt{\frac{2}{\pi}} e^{-2a^2} \int_0^y \int_0^z \frac{u^2 + 2au}{[v(1-v)]^{3/2}} e^{-\frac{(u+2av)^2}{2v(1-v)}} du\, dv$$

$$a \geqq 0, y \geqq 0, 0 \leqq z \leqq 1.$$

Proof. For the finite distribution we refer to the expression (14).

In order to obtain the limiting distribution the following notations are introduced

$$\frac{k}{\sqrt{2n}} \sim a, \frac{l}{\sqrt{2n}} \sim y, \frac{g}{n} \sim z, \frac{r}{n} \sim w$$

$$a \geqq 0, y \geqq 0, 1 \geqq w \geqq z \geqq 0.$$

from which follow

$$dy \sim \frac{2}{\sqrt{n}} (l \text{ is odd!}), dz \sim \frac{1}{n}, dw \sim \frac{1}{n}.$$

Further we have

$$\frac{l+1}{2} \sim \frac{1}{2} \frac{y}{\sqrt{z}} \sqrt{2g} \sim \frac{1}{z} \frac{y}{\sqrt{w-z}} \sqrt{2(r-g)},$$

$$k \sim \frac{a}{\sqrt{1-w}} \sqrt{2(n-r)}.$$

From these relations and by using the well known asymptotic formulae we obtain the relation

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a, y \leq \frac{\lambda_k}{\sqrt{2n}} < y + dy, z \leq \frac{\gamma_k}{n} < z + dz\right) =$$

$$= \frac{2}{\pi} \frac{y^2 a}{z^{3/2}} e^{-\frac{y^2}{2z}} \int_z^1 \frac{1}{[(w-z)(1-w)]^{3/2}} e^{-\left(\frac{y^2}{2(w-z)} + \frac{2a^2}{1-w}\right)} dw\,dy\,dz\,.$$

Applying for the variable w under the integral the transformation

$$\frac{1-w}{1-r} = \frac{t}{1+t}$$

it follows that

$$\frac{2}{\pi} \frac{y^2 a}{z^{3/2}(1-z)^2} e^{-\frac{y^2}{2z(1-z)} - \frac{2a^2}{1-z}} \int_0^\infty \frac{1+t}{t^{3/2}} e^{-\frac{4a^2+y^2t^2}{t}} dt\,dy\,dz\,.$$

The integral can be evaluated with the aid of the known formula

$$\int_0^\infty \frac{1}{t^{1/2}} e^{-\frac{1+t^2}{2bt}} dt = \int_0^\infty \frac{1}{t^{3/2}} e^{-\frac{1+t^2}{2bt}} dt = \frac{\sqrt{2b\pi}}{e^{1/b}}$$

and we obtain the limiting density function corresponding to our statement in theorem 2.2′.

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a, y \leq \frac{\lambda_k}{\sqrt{2n}} < y + dy, z \leq \frac{\gamma_k}{n} < z + dz\right) =$$

$$= \sqrt{\frac{2}{\pi}} \frac{y^2 + 2ay}{[z(1-z)]^{3/2}} e^{-2a^2} e^{-\frac{(y+2az)^2}{2z(1-z)}} dy\,dz\,.$$

Integration in respect of z from 0 to y leads to the relation

$$\mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a, y \leq \frac{\lambda_k}{\sqrt{2n}} < y + dy\right) = 4(a+y)\, e^{-2(a+y)^2} dy$$

corresponding to our theorem 2,1′, and in respect of y from 0 to ∞ to

$$\mathbf{P}\left(\frac{\varkappa}{\sqrt{2n}} > a,\ z \leq \frac{\gamma_k}{n} < z + dz\right) = \sqrt{\frac{2}{\pi}} \int_z^1 \frac{a}{[u(1-u)]^{3/2}} e^{-\frac{2a^2}{1-u}} du\,,$$

which is the limiting form of the distribution (13) of MIHALEVICH.

(Received July 27, 1960.)

REFERENCES

[1] Chung, K. L.—Feller, W.: „Fluctuations in coin tossing" *Proceedings National Academy of Sciences USA* **35** (1949) 605—608.

[2] Гнеденко, Б. В. — Михалевич, В. С.: «О распределении числа выходов одной эмпирической функции распределения над другой.» *ДАН СССР* **82** (1952) 6, 841—843.

[3] Гнеденко, Б. В. — Михалевич, В. С.: «Две теоремы о поведении эмпирической функции распределения». *ДАН СССР* **85** (1952) 1, 25—27.

[4] Михалевич, В. С.: «О взаимном расположении двух эмпирических функций распределения.» *ДАН СССР* **85** (1952) 3, 485—488.

[5] Гнеденко, Б. В.: «Проверка неизменности распределения вероятностей в двух независимых выборках.» *Mathematische Nachrichten* **12** 1 (2) (1954) 29—63.

[6] Hodges, J. L.: „Galton's rank-order test." *Biometrika* **42** (1955) 261—262.

[7] Feller, W.: *An introduction to probability theory and its applications*, Vol. **I.** (2nd edition) John Wiley and Sons, New York, 1957.

[8] Vincze, I.: „On some joint distributions and joint limiting distributions in the theory of order statistics, II." *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **4** (1959) 29—47.

[9] Смирнов, Н. В.: «Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках.» *Бюллетень Московского государственного Университета Математика* **2** (1939) 2, 3—14.

О НЕКОТОРЫХ ПРОБЛЕМАХ СВЯЗАННЫХ С КРИТЕРИЕМ ГАЛЬТОНА

E. CSÁKI и I. VINCZE

Резюме

В работе рассматривается случайное блуждание по линии, где блуждающаяся частица «шагает» единицу вперёд или назад с вероятностью $\frac{1}{2}$ и вернётся после $2n$ шагов в исходное положение. Пусть обозначает $2\gamma_k$ число шагов, после которых координата положения частицы больше чем k (γ_0 — критерий Гальтона), и $\lambda_k + 1$ число переходов в точке k. Определяются распределения, совместное распределение и соответствующие предельные распределения случайных величин γ_k и λ_k. Эти распределения могут быть использованы для конструирования критериев для сравнения двух выборок. — Исследования связаны с работами Смирнова [9], Гнеденко и Михалевича [2], [3], [4], [5].