KONINKL. NEDERL. AKADEMIE VAN WETENSCHAPPEN – AMSTERDAM
Reprinted from Proceedings, Series A, 58, No. 2 and Indag. Math., 17, No. 2, 1955

MATHEMATICS

A TEST FOR THE EQUALITY OF PROBABILITIES AGAINST A CLASS OF SPECIFIED ALTERNATIVE HYPOTHESES, INCLUDING TREND [1]). I

BY

CONSTANCE VAN EEDEN AND J. HEMELRIJK
(Mathematical Centre, Amsterdam)

(Communicated by Prof. D. VAN DANTZIG at the meeting of January 29, 1955)

1. *Introduction*

The problem considered in this paper concerns k ($k \geqq 2$) independent series of independent trials, each trial resulting in a success or a failure. For each trial of the i-th series, $i=1, 2, \ldots, k$, the probability of a success is the same and this probability will be denoted by p_i, the probability of a failure being $q_i = 1-p_i$. Denoting the number of trials in the i-th series by n_i, the number of successes by $\boldsymbol{a}_i$ [2]) and the number of failures by $\boldsymbol{b}_i$ and defining

$$\mathbf{t}_1 \overset{def}{=} \sum_i \boldsymbol{a}_i, \quad \mathbf{t}_2 \overset{def}{=} \sum_i \boldsymbol{b}_i, \quad n \overset{def}{=} \sum_i n_i, \tag{1.1}$$

the situation may be summarized by means of the following table

TABLE 1

Series	Probability of succes	Numbers of successes	Numbers of failures	Numbers of trials
1	p_1	$\boldsymbol{a}_1$	$\boldsymbol{b}_1$	n_1
2	p_2	$\boldsymbol{a}_2$	$\boldsymbol{b}_2$	n_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
k	p_k	$\boldsymbol{a}_k$	$\boldsymbol{b}_k$	n_k
	total	$\mathbf{t}_1$	$\mathbf{t}_2$	n

The hypothesis to be tested is:

$$H_0: p_1 = p_2 = \ldots = p_k. \tag{1.2}$$

Usually the χ^2-test is applied to data of this kind. The class of alternative hypotheses for which this test is consistent is of a very

[1]) Report SP 39 of the Statistical Department of the Mathematical Centre.
[2]) Random variables will be distinguished from numbers (e.g. from the values they take in the experiment) by printing them in bold type.

general character, comprising e.g., when k is constant, all hypotheses for which H_0 is not satisfied.

Apart from general tests of this kind, specialized tests for smaller and more specific classes of alternative hypotheses are very useful.

The alternative hypothesis, which formed the original aim of this investigation, was the hypothesis of a trend of the p_i, defined by

$$\sum_{i<j} (p_i - p_j) = \sum_i (k+1-2i)\, p_i \neq 0. \tag{1.3}$$

This class of alternative hypotheses is, however, a special case of the class defined by

$$\theta \overset{def}{=} \sum_i g_i\, p_i \neq 0, \tag{1.4}$$

where $g_1, g_2, \ldots, g_k$ are given numbers ("weights"). This consideration led to the following *general formulation* of the problem treated in this paper.

Consider, for $\nu = 1, 2, \ldots$, a sequence $\{T_\nu\}$ of tables like table 1 with $T_\nu \subset T_{\nu+1}$, i.e. $T_{\nu+1}$ originates from T_ν by adding new independent trials. Let T_ν consist of k_ν series, containing, for $i = 1, 2, \ldots, k_\nu$, $n_{i,\nu}$ trials, with probabilities $p_{i,\nu}$ of success, with $\mathbf{a}_{i,\nu}$ successes and $\mathbf{b}_{i,\nu}$ failures and with

$$\mathbf{t}_{1,\nu} \overset{def}{=} \sum_i \mathbf{a}_{i,\nu}, \quad \mathbf{t}_{2,\nu} \overset{def}{=} \sum_i \mathbf{b}_{i,\nu}, \quad n_\nu \overset{def}{=} \sum_i n_{i,\nu}. \tag{1.6}$$

The problem under consideration is then to find a test for the series of hypotheses

$$H_{0,\nu} : p_{1,\nu} = p_{2,\nu} = \ldots = p_{k_\nu,\nu}, \tag{1.7}$$

a test which is consistent, for $\nu \to \infty$ and $n_\nu \to \infty$, for the following class of alternative hypotheses. Let $g_{1,\nu}, g_{2,\nu}, \ldots, g_{k_\nu,\nu}$ be k_ν freely chosen weights, satisfying, for convenience sake, the relations

$$\sum_i g_{i,\nu} = 0\,; \quad \sum_i |g_{i,\nu}| = 1 \text{ [3]} \qquad (\nu = 1, 2, \ldots), \tag{1.8}$$

then we want the (twosided) test to be consistent exclusively for hypotheses with

$$\liminf_{\nu\to\infty} |\theta_\nu| > 0 \qquad (\theta_\nu \overset{def}{=} \sum_i g_{i,\nu}\, p_{i,\nu}). \tag{1.9}$$

It will be proved that this problem can be solved by means of the series of statistics

$$\mathbf{W}_\nu \overset{def}{=} \sum_i n_{i,\nu}^{-1}\, g_{i,\nu}\, \mathbf{a}_{i,\nu}, \tag{1.10}$$

the maximum likelihood estimates of θ_ν if no restrictions are imposed on the $p_{i,\nu}$.

The results of this investigation are summarized in section 2, the special case of a trend in the p_i is treated in section 3, the sections

[3]) The second relation might be replaced by e.g. $\sum_i g_{i,\nu}^2 = 1$ without bringing about an essential change in the problem or its solution.

4, ..., 7 contain the proofs and the relations between this test and some other tests are discussed in section 8.

Remark on notation

The notation of the rest of this paper will be simplified by omitting, in general, the index ν, because every symbol in (1.6), ..., (1.10) bears this index. Omitting it will, therefore, not lead to confusion, if it is understood, that all asymptotic relations, if not explicitly mentioned otherwise, are for $\nu \to \infty$. Since it is evidently necessary to let $n \to \infty$ when $\nu \to \infty$, these asymptotic relations may also be said to be valid for $n \to \infty$. The indices i and j will always be understood to take the values 1, 2, ..., k (i.e. 1, 2, ..., k_ν) if not explicitly mentioned otherwise. If no confusion is to be feared the $\sum$- and max-signs will be written without the index i.

2. *Description of the test*

To test the hypothesis H_0, given by (1.2), the statistic

$$\mathbf{W} \overset{def}{=} \sum n_i^{-1} g_i \mathbf{a}_i \tag{2.1}$$

with

$$\sum g_i = 0, \quad \sum |g_i| = 1 \tag{2.2}$$

is introduced. Under H_0 and under the condition $\mathbf{t}_1 \overset{def}{=} \sum \mathbf{a}_i = t_1$, $\mathbf{W}$ will be proved to have the following mean and variance

$$E(\mathbf{W}|t_1, H_0) = 0 \tag{2.3}$$

and

$$\sigma^2(\mathbf{W}|t_1, H_0) = \{n(n-1)\}^{-1} t_1 t_2 \sum n_i^{-1} g_i^2. \tag{2.4}$$

Dropping the condition $\mathbf{t}_1 = t_1$, $\sigma^2(\mathbf{W}|\mathbf{t}_1, H_0)$ becomes a random variable, which will be denoted by $\mathbf{s}^2$:

$$\mathbf{s}^2 \overset{def}{=} \{n(n-1)\}^{-1} \mathbf{t}_1 \mathbf{t}_2 \sum n_i^{-1} g_i^2. \tag{2.5}$$

The *test statistic* $\mathbf{V}$ is now defined as follows

$$\mathbf{V} \overset{def}{=} \mathbf{s}^{-1} \mathbf{W}, \tag{2.6}$$

where $\mathbf{s}$ is the positive root of $\mathbf{s}^2$.

The following onesided and twosided *critical regions* are used

$$\begin{cases} Z_1: \ V \geqq \xi_\alpha, \\ Z_2: \ V \leqq -\xi_\alpha, \\ Z_3: \ |V| \geqq \xi_{\frac{1}{2}\alpha}, \end{cases} \tag{2.7}$$

where ξ_ε is the $(1-\varepsilon)$-quantile of the standard normal distribution, i.e.

$$(2\pi)^{-\frac{1}{2}} \int_{\xi_\varepsilon}^{\infty} e^{-\frac{1}{2}u^2} du = \varepsilon. \tag{2.8}$$

The following properties will be proved.

If for $n \to \infty$ either the *conditions*

$$(2.9) \qquad \begin{cases} 1. & (\sum n_i^{-1} g_i^2)^{-\frac{1}{2}r} n^{\frac{1}{2}r-1} \sum n_i^{1-r} g_i^r = O(1) \\ 2. & \sum n_i p_i \to \infty, \ \sum n_i q_i \to \infty \end{cases} \quad \text{for each integer } r > 2,$$

or the *conditions*

$$(2.10) \qquad \begin{cases} 1. & (\sum n_i^{-1} g_i^2)^{-1} \max (n_i^{-2} g_i^2) = o(1) \\ 2. & (\sum n_i p_i)^{-1} (\sum n_i q_i) = O(1), \ (\sum n_i q_i)^{-1} \sum n_i p_i = O(1), \end{cases}$$

are satisfied, then

$$(2.11) \qquad \lim P[\mathbf{V} \geqq \xi_\alpha | H_0] = \lim P[\mathbf{V} \leqq -\xi_\alpha | H_0] = \lim P[|\mathbf{V}| \geqq \xi_{\frac{1}{2}\alpha} | H_0] = \alpha$$

and the tests are consistent for the following classes of alternatives.

Let

$$(2.12) \qquad \theta = \sum g_i p_i$$

then the tests based on Z_1 and Z_2 respectively are consistent if

$$(2.13) \qquad \lim\inf \theta > 0$$

or

$$(2.14) \qquad \lim\sup \theta < 0$$

respectively. They are, on the other hand, not consistent, if

$$(2.15) \qquad \lim\inf \theta < 0$$

or

$$(2.16) \qquad \lim\sup \theta > 0$$

respectively and, if

$$(2.17) \qquad \lim\inf \theta = 0$$

or

$$(2.18) \qquad \lim\sup \theta = 0$$

respectively, they are not consistent either, if α is chosen sufficiently small and if (for both cases) the following condition holds

$$(2.19) \qquad \lim\sup (\sum n_i p_i \sum n_i q_i \sum n_i^{-1} g_i^2)^{-1} n^2 \sum n_i^{-1} g_i^2 p_i q_i < \infty.$$

The test based on Z_3 is consistent if

$$(2.20) \qquad \lim\inf |\theta| > 0$$

and not consistent if

$$(2.21) \qquad \lim\inf |\theta| = 0$$

provided that (2.19) holds and α is chosen sufficiently small.

Remarks

1. The test has been formulated above as an unconditional test, based on $\mathbf{V}$. It may, however, also be interpreted as a conditional test, based on $\mathbf{W}$ and on the condition $\mathbf{t}_1 = t_1$, where t_1 is the value found in the experiment. As will be proved in section 6 (theorem II), the conditional asymptotic distribution of $\mathbf{W}$, under the condition $\mathbf{t}_1 = t_1$ is asymptotically normal; the conditional mean and variance are given by (2.3) and (2.4).

2. It seems to the authors that tests, involving samples of different sizes should satisfy the *general principle* that the set of alternatives for which the test is consistent does not depend on the ratios of the sample sizes, except, if necessary, for asymptotic relations on these ratios as e.g. their boundedness for $n \to \infty$. For the test under consideration this means that the g_i in (2.12) should not depend on the ratios $n_j^{-1} n_{j'}$ $(j \neq j')$. In section 8 an example of a related test (T. J. TERPSTRA [10]), not satisfying this rule, is given and its drawback, owing to this fact, is illustrated by an example. A small adjustment is, however, sufficient to avoid this difficulty.

On the other hand, as may be seen from (2.19), the authors have not succeeded completely in keeping the relations concerning consistency free from the n_i; (2.19) is, however, only a sufficient condition and it might be possible to eliminate the n_i completely from the consistency relations. Also the case, when (2.19) is not satisfied, has not yet been investigated. It may be pointed out, that (2.19) follows from (2.10.2) and is therefore not very restrictive. Nevertheless further work on these points seems desirable.

3. The power of the test for finite values of n has not been investigated as yet. It may be surmised, that for the classes of alternative hypotheses, for which the test is consistent, this power might be larger than that of the χ^2-test, the latter being consistent for a much wider class of alternatives. This has, however, not been proved as yet and on this point too further investigations seem worth while.

4. In practice it may often be convenient to multiply all weights g_i by the same function of k and $n_1, n_2, \ldots, n_k$. This has, of course, no influence on $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{s}$ being multiplied by this same function.

3. *A trend as alternative hypothesis*

Taking, in accordance with (2.2), the weights

$$g_{i,T} = 2K^{-1}(k+1-2i) \qquad (i = 1, 2, \ldots, k) \tag{3.1}$$

for the g_i, where $K = k^2$ if k is even and $K = k^2 - 1$ if k is odd, θ assumes the form

$$\theta_T \overset{def}{=} \sum g_{i,T}\, p_i = 2K^{-1} \sum_{i<j} (p_i - p_j). \tag{3.2}$$

For $\mathbf{W}$, $\mathbf{s}^2$ and $\mathbf{V}$ this gives, omitting the factor $2K^{-1}$ (cf. remark 4):

$$(3.3) \qquad \mathbf{W}_T = \sum n_i^{-1}(k+1-2i)\, \mathbf{a}_i = \sum_{i<j} (n_i\, n_j)^{-1}(n_j\, \mathbf{a}_i - n_i\, \mathbf{a}_j),$$

$$(3.4) \qquad \mathbf{s}_T^2 = \{n(n-1)\}^{-1}\, \mathbf{t}_1\, \mathbf{t}_2 \sum n_i^{-1}(k+1-2i)^2,$$

$$(3.5) \qquad \mathbf{V}_T = \mathbf{s}_T^{-1}\, \mathbf{W}_T.$$

The two-sided test is now consistent if lim inf $|\theta_T| > 0$, which may be taken as a definition of a trend of the p_i.

In two special cases considerable simplification occurs. First if $n_i = m$ for all i, the conditions (2.9.1) and (2.10.1) are automatically satisfied and $\mathbf{W}_T$ reduces, omitting a factor m^{-1}, to

$$(3.6) \qquad \mathbf{W}_T = \sum_{i<j} (\mathbf{a}_i - \mathbf{a}_j),$$

with

$$(3.7) \qquad \mathbf{s}_T^2 = \{3(n-1)\}^{-1}\, \mathbf{t}_1\, \mathbf{t}_2 (k^2-1),$$

which, if $m=1$, reduces to

$$(3.8) \qquad \mathbf{s}_T^2 = 3^{-1}\, \mathbf{t}_1\, \mathbf{t}_2 (k+1),$$

k being equal to n in that case.

Secondly, if k is bounded for $n \to \infty$, the conditions (2.9.1) and (2.10.1) respectively reduce to

$$(3.9) \qquad n_i^{-1}\, n = O(1) \quad \text{for} \quad i \neq 2^{-1}(k+1)$$

and

$$(3.10) \qquad n_i \to \infty \quad \text{for} \quad i \neq 2^{-1}(k+1)$$

respectively, no restrictions being imposed, if k is odd, on the n_i with $i = 2^{-1}(k+1)$.

Remark 5. In section 8 it will be shown, that if $n_i = m$ for $i = 1, 2, \ldots, k$ and if the g_i are given by (3.1), the test may be interpreted as an application of WILCOXON's [12] two sample test. This means, that for $m=1$ and small values of t_1 and t_2 the *exact* tables of the distribution of WILCOXON's test statistic may be used to obtain an exact test for H_0 against the alternative of a trend of the p_i as defined above.

4. *The conditional mean and variance of* $\mathbf{W}$ *under the hypothesis* H_0

This section contains the proof of (2.3) and (2.4).

Theorem I

If H_0 *is true the mean and variance of* $\mathbf{W}$ *under the condition* $\mathbf{t}_1 = t_1$ *are*

$$(4.1) \qquad E(\mathbf{W}|t_1, H_0) = 0,$$

$$(4.2) \qquad \sigma^2(\mathbf{W}|t_1, H_0) = \{n(n-1)\}^{-1}\, t_1\, t_2 \sum n_i^{-1}\, g_i^2.$$

Proof

The simultaneous distribution of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ under the hypothesis H_0

and under the condition $\mathbf{t}_1 = t_1$ is a $(k-1)$-dimensional hypergeometric distribution:

$$(4.3)\qquad \mathrm{P}\,[\boldsymbol{a}_1 = a_1,\ \boldsymbol{a}_2 = a_2,\ \ldots,\ \boldsymbol{a}_k = a_k \,|\, t_1,\, H_0] = \binom{n}{t_1}^{-1} \prod \binom{n_i}{a_i}.$$

From (4.3) it follows

$$(4.4)\qquad E\,(\boldsymbol{a}_i \,|\, t_1,\, H_0) = n^{-1}\, n_i\, t_1,$$

$$(4.5)\qquad \sigma^2\,(\boldsymbol{a}_i \,|\, t_1,\, H_0) = \{n^2\,(n-1)\}^{-1}\, n_i\,(n-n_i)\, t_1\, t_2,$$

$$(4.6)\qquad cov\,(\boldsymbol{a}_i,\, \boldsymbol{a}_j \,|\, t_1,\, H_0) = -\,\{n^2\,(n-1)\}^{-1}\, n_i\, n_j\, t_1\, t_2 \qquad i \neq j.$$

Consequently

$$(4.7)\quad E\,(\mathbf{W}\,|\,t_1,\, H_0) = \sum n_i^{-1}\, g_i\, E\,(\boldsymbol{a}_i\,|\,t_1,\, H_0) = n^{-1}\, t_1 \sum g_i = 0 \quad (\text{cf } (2.2)),$$

$$(4.8)\quad \begin{cases} \sigma^2(\mathbf{W}|t_1,H_0) = \sum n_i^{-2} g_i^2\, \sigma^2(\boldsymbol{a}_i|t_1,H_0) + \sum\limits_{i\neq j} (n_i n_j)^{-1} g_i g_j\, cov\,(\boldsymbol{a}_i, \boldsymbol{a}_j|t_1,H_0) = \\ \quad = \{n^2\,(n-1)\}^{-1}\, t_1\, t_2\, \{\sum n_i^{-1}\,(n-n_i)\, g_i^2 - \sum\limits_{i\neq j} g_i\, g_j\} = \\ \quad = \{n^2\,(n-1)\}^{-1}\, t_1\, t_2\, \{n \sum n_i^{-1}\, g_i^2 - (\sum g_i)^2\} = \\ \quad = \{n\,(n-1)\}^{-1}\, t_1\, t_2 \sum n_i^{-1}\, g_i^2. \qquad (\text{cf } (2.2.)) \end{cases}$$

5. *Lemma's*

In this section some lemma's needed for the proofs of the theorems, are given.

Lemma I

If either condition (2.9.1) *or condition* (2.10.1) *is satisfied then*

$$(5.1)\qquad \lim \sum n_i^{-1} g_i^2 = 0.$$

Proof

From (2.2) it follows

$$\sum n_i^{-1}\, g_i^2 \leqq \max\,(n_i^{-1}|g_i|) \cdot \sum |g_i| = \max\,(n_i^{-1}|g_i|).$$

Consequently

$$\{\sum n_i^{-1}\, g_i^2\}^{-1} \max\,(n_i^{-2}\, g_i^2) \geqq \sum n_i^{-1}\, g_i^2.$$

Therefore if (2.10.1) is satisfied

$$0 \leqq \liminf \sum n_i^{-1}\, g_i^2 \leqq \limsup \sum n_i^{-1}\, g_i^2 \leqq$$
$$\leqq \limsup\, \{\sum n_i^{-1}\, g_i^2\}^{-1} \max\,(n_i^{-2}\, g_i^2) = 0,$$

whence $\lim \sum n_i^{-1}\, g_i^2 = 0$.

On the other hand if r is even

$$\{\sum n_i^{-1}\, g_i^2\}^{-\frac12 r}\, n^{\frac12 r-1} \sum n_i^{1-r}\, g_i^r \geqq \{\sum n_i^{-1}\, g_i^2\}^{-\frac12 r}\, n^{\frac12 r-1} \max\,(n_i^{-r}\, g_i^r) =$$
$$= n^{\frac12 r-1}\, \{(\sum n_i^{-1}\, g_i^2)^{-1} \max\,(n_i^{-2}\, g_i^2)\}^{\frac12 r}.$$

Consequently (2.10.1) follows from (2.9.1) which completes the proof.[4])

Lemma II

If

$$(5.2) \qquad 0 < \varepsilon < 1$$

and

$$(5.3) \quad \begin{cases} \tau_1^2 \overset{def}{=} \{n(n-1\}^{-1}(\sum n_i p_i \sum n_i q_i - \varepsilon^{-\frac{1}{2}} n \sqrt{\sum n_i p_i q_i} + \varepsilon^{-1} \sum n_i p_i q_i) \sum n_i^{-1} g_i^2, \\ \tau_2^2 \overset{def}{=} \{n(n-1\}^{-1}(\sum n_i p_i \sum n_i q_i + \varepsilon^{-\frac{1}{2}} n \sqrt{\sum n_i p_i q_i} + \varepsilon^{-1} \sum n_i p_i q_i) \sum n_i^{-1} g_i^2, \end{cases}$$

then

$$(5.4) \qquad \tau_1^2 < \mathbf{s}^2 < \tau_2^2 \quad \textit{except for a probability } \varepsilon.$$

Proof

From the inequality of Bienaymé–Tschebycheff it follows

$$|\mathbf{t}_1 - \sum n_i p_i| = |\mathbf{t}_2 - \sum n_i q_i| < \varepsilon^{-\frac{1}{2}} \sqrt{\sum n_i p_i q_i} \quad \text{except for a probability } \varepsilon.$$

Consequently

$$(5.5) \quad n(n-1)\tau_1^2 < \mathbf{t}_1 \mathbf{t}_2 \sum n_i^{-1} g_i^2 < n(n-1)\tau_2^2 \text{ except for a probability } \varepsilon.$$

Lemma II follows from (5.5) and (2.5).

From lemma I and the definition of τ_1^2 and τ_2^2 follows

Lemma III

$$(5.6) \qquad \lim \tau_1^2 = \lim \tau_2^2 = 0.$$

[4]) (2.9.1) entails even

$$(\sum n_i^{-1} g_i^2)^{-1} n \max (n_i^{-2} g_i^2) = O(1),$$

as is seen by raising both members to the power $2r^{-1}$ and passing to the limit $r \to \infty$.

(To be continued)

KONINKL. NEDERL. AKADEMIE VAN WETENSCHAPPEN – AMSTERDAM
Reprinted from Proceedings, Series A, 58, No. 3 and Indag. Math., 17, No. 3, 1955

MATHEMATICS

A TEST FOR THE EQUALITY OF PROBABILITIES AGAINST A CLASS OF SPECIFIED ALTERNATIVE HYPOTHESES, INCLUDING TREND. II

BY

CONSTANCE VAN EEDEN AND J. HEMELRIJK
(Mathematical Centre, Amsterdam)

(Communicated by Prof. D. VAN DANTZIG at the meeting of January 29, 1955)

6. *The conditional asymptotic distribution of* $\mathbf{W}$ *under the hypothesis* H_0

In this section we shall prove (2.11). For this proof we use the following theorem.

Theorem II

If for $n \to \infty$ *either the conditions*

$$(6.1) \quad \begin{cases} 1. & \{\sum n_i^{-1} g_i^2\}^{-\frac{1}{2}r} n^{\frac{1}{2}r-1} \sum n_i^{1-r} g_i^r = O(1) \quad \textit{for each integer } r > 2 \\ 2. & t_1 \to \infty, \ t_2 \to \infty \end{cases}$$

or the conditions

$$(6.2) \quad \begin{cases} 1. & \{\sum n_i^{-1} g_i^2\}^{-1} \max \ (n_i^{-2} g_i^2) = o(1), \\ 2. & t_1^{-1} t_2 = O(1), \quad t_2^{-1} t_1 = O(1) \end{cases}$$

are satisfied the random variable

$$\{\sigma(\mathbf{W} \mid t_1, H_0)\}^{-1} \mathbf{W}$$

is, under the condition $\mathbf{t}_1 = t_1$ *and under the hypothesis* H_0, *for* $n \to \infty$ *asymptotically normally distributed with mean* 0 *and variance* 1.

Proof

The proof is based on a theorem by A. WALD and J. WOLFOWITZ [11], improved by G. E. NOETHER [9] and simplified by W. HOEFFDING [5].

If $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are n independent random variables, where the values taken by these variables form a permutation of n numbers $c_1, c_2, \ldots, c_n$ and

$$(6.3) \qquad \mathbf{L} \overset{\text{def}}{=} \sum_{\lambda} d_\lambda \mathbf{y}_\lambda, \text{ [5]}$$

where $d_1, d_2, \ldots, d_n$ is a given row of numbers, the following theorem follows from the abovementioned theorems. If

[5]) Unless explicitly stated otherwise λ and μ take the values 1, 2, ..., n.

(6.4)
1. all permutations of the values $c_1, c_1, \ldots, c_n$ assumed by $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ have the same probability,
2. the row $\{d_\lambda\}$ satisfies the condition [6]) $\{M_2(d)\}^{-\frac{1}{2}r} M_r(d) = O(1)$ for $n \to \infty$ and each integer $r > 2$,
3. the row $\{c_\lambda\}$ satisfies the condition $\{n M_2(c)\}^{-1} \max_\lambda (c_\lambda - \bar{c})^2 = o(1)$ for $n \to \infty$

then the random variable

$$\{\sigma(\mathbf{L})\}^{-1} \{\mathbf{L} - E(\mathbf{L})\}$$

is, for $n \to \infty$, asymptotically normally distributed with mean 0 and variance 1.

In order to apply this theorem to our problem we consider the n trials as n independent observations, one of each of n random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ where

$$\text{(6.5)} \quad \mathrm{P}[\mathbf{x}_\lambda = 1] = p_i, \mathrm{P}[\mathbf{x}_\lambda = 0] = q_i \begin{cases} n_1 + n_2 + \ldots + n_{i-1} < \lambda \leqq n_1 + n_2 + \ldots + n_i \\ i = 1, 2, \ldots, k; \quad \lambda = 1, 2, \ldots, n. \end{cases}$$

Then

$$\text{(6.6)} \qquad \mathbf{a}_i = \sum_{\lambda = n_1 + \ldots + n_{i-1} + 1}^{n_1 + \ldots + n_i} \mathbf{x}_\lambda \qquad (i = 1, 2, \ldots, k).$$

The theorem will be applied twice. First substitute $\mathbf{x}_\lambda$ for $\mathbf{y}_\lambda (\lambda = 1, 2, \ldots, n)$ and $n_i^{-1} g_i$ for d_λ $(n_1 + n_2 + \ldots + n_{i-1} < \lambda \leqq n_1 + n_2 + \ldots + n_i;\ i = 1, 2, \ldots, k;$ $\lambda = 1, 2, \ldots, n)$ and let the row $\{c_\lambda\}$ consist of t_1 times the number 1 and t_2 times the number 0.

Then

$$\text{(6.7)} \qquad \mathbf{L} = \sum n_i^{-1} g_i \mathbf{a}_i = \mathbf{W},$$

and, given the independence of the trials, condition (6.4.1) is equivalent to H_0 if $\mathbf{t}_1 = t_1$. Condition (6.4.2) reduces to (6.1.1) and (6.4.3) to

$$\text{(6.8)} \qquad (n\, t_1 t_2)^{-1} \max\ (t_1^2, t_2^2) = o(1),$$

which is equivalent to (6.1.2). This proves the first part of theorem II.

Secondly, to arrive at the conditions (6.2), take $d_1 = d_2 = \ldots = d_{t_1} = 1$ and $d_{t_1+1} = d_{t_1+2} = \ldots = d_n = 0$ and substitute $n_i^{-1} g_i$ for $c_\lambda (n_1 + n_2 + \ldots + n_{i-1} < \lambda \leqq n_1 + n_2 + \ldots + n_i;\ i = 1, 2, \ldots, k;\ \lambda = 1, 2, \ldots, n)$. It is easy to see that this leads to a random variable $\mathbf{L}$, which has under condition (6.4.1) the same probability distribution as $\mathbf{W}$ under H_0, given $\mathbf{t}_1 = t_1$. Thus the

---

[6]) For a row $w_1, w_2, \ldots, w_n$, $M_r(w)$ is defined by

$$M_r(w) \overset{def}{=} n^{-1} \sum_\lambda (w_\lambda - \bar{w})^r,$$

where

$$\bar{w} \overset{def}{=} n^{-1} \sum_\lambda w_\lambda.$$

asymptotic normality of $\boldsymbol{L}$ implies the conditional asymptotic normality of $\boldsymbol{W}$. Condition (6.4.2) reduces in this case to

$$(6.9) \quad \{n(t_1 t_2)^{\frac{1}{2}r}\}^{-1}\{t_1 t_2^r + t_2(-t_1)^r\} = \mathrm{O}(1) \quad \text{for each integer } r > 2,$$

which is equivalent to (6.2.2) and (6.4.3) reduces to (6.2.1). This proves the second part of the theorem.

The validity of (2.11) may now be proved as follows. According to the BOREL-CANTELLI lemma (cf. e.g. W. FELLER [2], p. 155) condition (2.9.2) implies (6.1.2) except for a probability 0 and the condition (6.2.2). holds with probability 1 if (2.10.2) is true as may be seen from the strong law of large numbers (cf. e.g. [2] p. 156).

Consequently, if (2.9) or (2.10) are satisfied

$$(6.10) \qquad \lim \mathrm{P}[\boldsymbol{W} \geqq \xi_\alpha \sigma(\boldsymbol{W}|t_1, H_0)|t_1, H_0] = \alpha.$$

From (6.10) and

$$(6.11) \begin{cases} \mathrm{P}[\boldsymbol{V} \geqq \xi_\alpha | H_0] = \mathrm{P}[\boldsymbol{W} \geqq \xi_\alpha \mathbf{s} | H_0] = \\ \qquad = \sum_{t_1} \mathrm{P}[\mathbf{t}_1 = t_1 | H_0] \cdot \mathrm{P}[\boldsymbol{W} \geqq \sigma(\boldsymbol{W}|t_1, H_0)|t_1, H_0] \end{cases}$$

follows

$$(6.12) \qquad \lim \mathrm{P}[\boldsymbol{V} \geqq \xi_\alpha | H_0] = \alpha.$$

The other relations of (2.11) may be proved analogously.

7. *The consistency of the test*

Theorem III

If either the conditions (2.9) *or the conditions* (2.10) *are satisfied the test consisting of rejecting* H_0 *if and only if* $\boldsymbol{V} \geqq \xi_\alpha$ *is*

1. *consistent for the class of those alternative hypotheses for which*

$$(7.1) \qquad \lim\inf \theta > 0,$$

2. *consistent for no alternative hypothesis of the class*

$$(7.2) \qquad \lim\inf \theta < 0,$$

and

3. *consistent for no alternative hypothesis satisfying*

$$(7.3) \begin{cases} \lim\inf \theta = 0 \\ \lim\sup \{\sum n_i p_i \sum n_i q_i \sum n_i^{-1} g_i^2\}^{-1} n^2 \sum n_i^{-1} g_i^2 p_i q_i < \infty \end{cases}$$

if α *is sufficiently small.*

Proof

Define

$$(7.4) \qquad \sigma^2 \overset{def}{=} \sigma^2(\boldsymbol{W}|H) = \sum n_i^{-1} g_i^2 p_i q_i,$$

where H denotes the hypothesis $\{p_1, p_2, \ldots, p_k\}$, then from (2.2) it follows

$$|\theta| = |\sum g_i p_i| \leqq \sum |g_i| p_i \leqq \sum |g_i| = 1 \tag{7.5}$$

and from lemma I

$$\lim \sigma^2 = 0. \tag{7.6}$$

For those alternative hypotheses H for which

$$\lim\inf \theta > 0 \tag{7.7}$$

we have except for a probability ε (cf (5.4))

$$\left\{\begin{aligned} &\mathrm{P}[\mathbf{V} < \xi_\alpha | H] = \mathrm{P}[\mathbf{W} < \xi_\alpha \mathbf{s} | H] \leqq \\ &\qquad \leqq \mathrm{P}[\mathbf{W} < \xi_\alpha \tau_2 | H] = \mathrm{P}[\mathbf{W} - \theta < \xi_\alpha \tau_2 - \theta | H]. \end{aligned}\right. \tag{7.8}$$

From lemma III and (7.7) it follows that $\xi_\alpha \tau_2 - \theta$ is negative and bounded away from 0 for sufficiently large n; this leads, by means of the inequality of BIENAYMÉ–TSCHEBYCHEFF, to the relation

$$\mathrm{P}[\mathbf{W} - \theta < \xi_\alpha \tau_2 - \theta | H] \leqq (\xi_\alpha \tau_2 - \theta)^{-2} \sigma^2 \to 0 \qquad \text{(cf. (7.6))}. \tag{7.9}$$

Consequently, ε being an arbitrary small positive number, the probability of not rejecting H_0 if (7.7) is true can be made arbitrarily small by choosing n sufficiently large.

This proves the first part of theorem III.

If, on the other hand, H' is a hypothesis with

$$\lim\ \inf\ \theta < 0 \tag{7.10}$$

there is a subsequence $\{\nu'\}$ of the sequence $\nu = 1, 2, \ldots,$ such that

$$\lim_{\nu' \to \infty} \sup\ \theta < 0. \tag{7.11}$$

For this subsequence we have, again using the lemmas I, II and III, except for a probability ε (cf. (5.4)):

$$\left\{\begin{aligned} &\mathrm{P}[\mathbf{V} \geqq \xi_\alpha | H'] = \mathrm{P}[\mathbf{W} \geqq \xi_\alpha \mathbf{s} | H'] \leqq \mathrm{P}[\mathbf{W} \geqq \xi_\alpha \tau_1 | H'] = \\ &\qquad = \mathrm{P}[\mathbf{W} - \theta \geqq \xi_\alpha \tau_1 - \theta | H'] \leqq (\xi_\alpha \tau_1 - \theta)^{-2} \sigma^2 \to 0 \end{aligned}\right. \tag{7.12}$$

for $\nu' \to \infty$, because of (7.6) and (7.11). This proves the second part of the theorem.

Finally if [7])

$$\lim\ \inf\ \theta = 0 \tag{7.13}$$

there is a subsequence $\{\nu''\}$ of the sequence $\nu = 1, 2, \ldots,$ such that

$$\lim_{\nu'' \to \infty} \theta = 0. \tag{7.14}$$

---

[7]) This proof of the last part of theorem III is based on a method which may be found in D. VAN DANTZIG [1].

Then if

$$\text{(7.15)} \qquad \limsup \tau_1^{-2} \sigma^2 < \infty$$

and

$$\text{(7.16)} \qquad \xi_\alpha^2 > \limsup \tau_1^{-2} \sigma^2$$

we have, except for a probability ε

$$\text{(7.17)} \quad \limsup_{\nu''\to\infty} \mathrm{P}[\boldsymbol{V} \geqq \xi_\alpha] \leqq \limsup_{\nu''\to\infty} \mathrm{P}[\boldsymbol{W} \geqq \xi_\alpha \tau_1] \leqq \limsup_{\nu''\to\infty} (\xi_\alpha \tau_1)^{-2} \sigma^2 < 1.$$

The condition (7.15) is equivalent to

$$\text{(7.18)} \quad \limsup \{\sum n_i p_i \sum n_i q_i \sum n_i^{-1} g_i^2\}^{-1} n^2 \sum n_i^{-1} g_i^2 p_i q_i < \infty,$$

which proves the last part of theorem III.

Analogous theorems about the other onesided test and the twosided test follow easily. Thus the proof of the statements of section 2 is complete.

8. *Relations to other tests*

1. T. J. TERPSTRA [10] has developed a test against trend for groups of observations. This test could be applied to the data of table 1 as follows. Consider the n_i trials of the i-th series as independent observations of a random variable $\mathbf{z}_i$, which takes the values 0 and 1 respectively with probabilities $q_i = 1 - p_i$ and p_i respectively $(i = 1, 2, \ldots, k)$. Then H_0 is identical with the hypothesis that $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k$ possess the same probability distribution and this is the null hypothesis for TERPSTRA's test. In general the $\mathbf{z}_i$ will not take the values 0 and 1 only and, as a matter of fact, TERPSTRA supposes the $\mathbf{z}_i$ to have continuous distributions; i.e. he proves the asymptotic normality of his test statistic under this condition. His test statistic $\boldsymbol{T}$ may now be defined as follows. Let $\boldsymbol{U}_{i,j}$ denote WILCOXON's test statistic for the i-th and j-th series of observations if $i<j$ and let $\boldsymbol{U}_{i,j} = n_i n_j - \boldsymbol{U}_{j,i}$ [8]). Then, defining

$$\text{(8.1)} \qquad \boldsymbol{W}_{i,j} \overset{def}{=} 2\{\boldsymbol{U}_{i,j} - E(\boldsymbol{U}_{i,j} | H_0)\},$$

$\boldsymbol{T}$ satisfies the relation

$$\text{(8.2)} \qquad \boldsymbol{W}_{T'} \overset{def}{=} 2\{\boldsymbol{T} - E(\boldsymbol{T} | H_0)\} = \sum_{i<j} \boldsymbol{W}_{i,j}.$$

For the case considered in this paper

$$\text{(8.3)} \qquad \boldsymbol{W}_{i,j} = n_j \boldsymbol{a}_i - n_i \boldsymbol{a}$$

and thus

$$\text{(8.4)} \qquad \boldsymbol{W}_{T'} = \sum_{i<j} (n_j \boldsymbol{a}_i - n_i \boldsymbol{a}_j).$$

---

[8]) For two samples $u_1, u_2, \ldots, u_N$ and $v_1, v_2, \ldots, v_M$ the value of WILCOXON's U is defined as the sum of the number of pairs (u_h, v_k) with $u_h > v_k$ and half the number of pairs (u_h, v_k) with $u_h = v_k$ $(h = 1, 2, \ldots, N;\ k = 1, 2, \ldots, M)$. In our case this gives:

$$U_{i,j} = a_i b_j + \tfrac{1}{2}(a_i b_i + a_j b_j).$$

Comparing $\mathbf{W}_{T'}$ with $\mathbf{W}$ (cf. (2.1)) it is found that $\mathbf{W}_{T'} = \mathbf{W}$ if the following weights $g_{i,T'}$ are substituted for the g_i:

$$\left.\begin{aligned} &(8.5) \qquad g_{i,T'} = \{\sum n_i |D_i|\}^{-1} n_i D_i \\ &\text{with} \\ &(8.6) \qquad D_i \overset{def}{=} \sum_{j>i} n_j - \sum_{j<i} n_j \end{aligned}\right\} \quad (i=1, 2, \ldots, k).$$

The asymptotic normality, under the conditions (2.7) or (2.8), makes it possible to use these weights and θ becomes

$$\theta_{T'} = \{\sum n_i |D_i|\}^{-1} \sum_{i<j} n_i n_j (p_i - p_j). \tag{8.7}$$

According to (2.13) and (2.14) the onesided critical region Z_1 (cf. (2.7)) gives a consistent test if

$$\liminf \theta_{T'} > 0 \tag{8.8}$$

and Z_2, if

$$\limsup \theta_{T'} < 0. \tag{8.9}$$

This means, however, as has been stated in remark 2 in section 2, that the n_i have an undue influence on the consistency. This may be illustrated by a simple example. Let $k=3$ and $p_1=p_3$, but $p_2<p_1$. Then, if $n_1>n_3$, (8.7) gives

$$\theta_{T'} = \{2\, n_1 (n_2 + n_3)\}^{-1} n_2 (n_1 - n_3)(p_1 - p_2)$$

and if $n_1<n_3$

$$\theta_{T'} = \{2\, n_3 (n_1 + n_2)\}^{-1} n_2 (n_1 - n_3)(p_1 - p_2).$$

Thus the sign of $\theta_{T'}$ depends on whether $n_1>n_3$ or $n_1<n_3$ and, keeping the proportions $n_i^{-1} n_j$ constant for $n \to \infty$ for all $i \neq j$, (8.8) is satisfied in the first case and (8.9) in the second case. This means, that for such values of p_i, where no trend is present at all, a positive or a negative trend respectively might be statistically established at will by choosing $n_1=cn_3$ with $c>1$ or $c<1$ respectively.

This drawback of Terpstra's test can, however, be avoided by means of a small modification. As a matter of fact this has been done in section 3, by choosing the weigths (3.1). Expressing $\mathbf{W}_T$ (cf (3.3)) in the $\mathbf{W}_{i,j}$ of (8.3) gives

$$\mathbf{W}_T = \sum_{i<j} (n_i n_j)^{-1} \mathbf{W}_{i,j}. \tag{8.10}$$

The $g_{i,T}$ being independent of the n_i, the classes of hypotheses for which this test is consistent, do not depend on the n_i. This adjustment of Terpstra's test is also useful for the general case considered in his paper. If $n_i=m$ for all i no adjustment is necessary. Another distribution-free test of a similar character, where such an adjustment might be desirable is the k sample test proposed by W. H. Kruskal [6].

2. The special case treated in section 3 can, if $n_i = m$ for all i, also be interpreted as an application of WILCOXON's [12] distributionfree test for the problem of two samples (cf. remark 5). Consider two samples A and B, which, taken together, contain m times the value i ($i = 1, 2, \ldots, k$) and let $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ respectively be the number of observations i in sample A and B respectively. Then, if $\boldsymbol{U}$ represents WILCOXON's statistic for these two samples, the relation

$$2\,\boldsymbol{U} = m\,\boldsymbol{W}_T + t_1 t_2 \tag{8.11}$$

holds, where $\boldsymbol{W}_T$ is defined by (3.6). It is easy to prove (cf. J. HEMELRIJK [4]) that the hypothesis H_0 is, under the condition $\mathbf{t}_1 = t_1$, equivalent with the hypothesis that sample A is a random sample taken without replacement from the two samples together. This means that WILCOXON's test may be applied. [9]) The well known formulas for the mean and the variance of $\boldsymbol{U}$ (cf. e.g. W. H. KRUSKAL and W. A. WALLIS [7]) under the null hypothesis lead again to (2.3) and (3.7).

For the case $n_i = m$ this test coincides with TERPSTRA's test treated above. If the n_i are unequal the same reasoning may be applied, again leading to TERPSTRA's test applied to the problem under investigation. The proposed modification of this test can in general not be interpreted in this way.

If $m = 1$ *exact* tables for WILCOXON's test are available (cf. e.g. H. B. MANN and D. R. WHITNEY [8]) and by means of these tables the test may thus be performed in an exact way as a test for two samples, one of which consists of the values of i which correspond with successes, while the other sample contains the rest of the numbers $1, 2, \ldots, n$.

3. If $k = 2$ we arrive at a 2×2-table and then all tests treated in this paper are identical. This includes the χ^2-test. The proper *exact* treatment has, for that case, been indicated by R. A. FISHER [3].

The authors want to thank Prof. Dr D. VAN DANTZIG for his constructive and stimulating criticism.

---

[9]) This result, derived in another way by A. BENARD and CONSTANCE VAN EEDEN, gave rise to the present investigation.

REFERENCES

1. DANTZIG, D. VAN, On the consistency and power of Wilcoxon's two sample test, Proc. Kon. Ned. Akad. van Wet. **A 54**, 1–8 (1951).
2. FELLER, W., An introduction to probability theory and its applications (London, 1950).
3. FISHER, R. A., Statistical methods for Research workers (London, 1948).
4. HEMELRIJK, J., Exemple d'application des méthodes non-paramétriques et un nouveau test pour l'égalité de plusieurs probabilités, Report SP 38 of the Mathematical Centre (Amsterdam, 1954).
5. HOEFFDING, W., A combinatorial central limit theorem, Ann. Math. Stat. 22, 558–566 (1951).

6. KRUSKAL, W. H., A non-parametric test for the several sample problem, Ann. Math. Stat. **23**, 525–539 (1952).
7. ——— and W. A. WALLIS, Use of ranks in one-criterion analysis of variance, J. Am. Stat. Ass. **47**, 583–621 (1952).
8. MANN, H. B. and D. R. WHITNEY, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. **18**, 50–60 (1947).
9. NOETHER, G. E., On a theorem by Wald and Wolfowitz, Ann. Math. Stat. 455–461 (1949).
10. TERPSTRA, T. J., The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking, Proc. Kon. Ned. Akad. v. Wet. **A 55**, 327–333 (1952).
11. WALD, A. and J. WOLFOWITZ, Statistical tests based on permutations of the observations, Ann. Math. Stat. **15**, 358–372 (1944).
12. WILCOXON, F., Individual comparisons by ranking methods, Biometrics **1**, 80–82 (1945).