

# ON TESTING FOR NORMALITY

KÁROLY SARKADI

MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES  
and  
UNIVERSITY OF CALIFORNIA, BERKELEY

## 1. Introduction

The majority of the goodness-of-fit problems arising in practice involves nuisance parameters. On the other hand, the majority of the results which have appeared within this field deal with simple hypothesis testing only. Among the relatively few results concerning the first-mentioned, more general problem, the most known and applied one is the modification given by R. A. Fisher to Karl Pearson's  $\chi^2$ -test (see for example, [4], pp. 424–434). As it is well known, this modification consists of replacing the unknown parameters by their estimates; the distribution of the modified test statistic (at least its approximate distribution) was determined. The same way was followed by Kac, Kiefer, and Wolfowitz [7] and Darling [5] concerning the Cramér-von Mises test and the Kolmogorov test. Computational difficulties, however, prevented them from providing tables having sufficient range and accuracy for practical purposes.

But even if these difficulties could be overcome in the future, neither they nor Fisher's method work in some nonsimple sample cases.

Considering this fact as well as the disadvantages of the  $\chi^2$ -test (see [7], pp. 191–192), other solutions of problems of this type seem to be of particular interest. The straightforward way is to find an equivalent simple hypothesis.

The basic theory and the most important results of this approach are dealt with in another paper appearing in this volume [12].

Solutions of this type form the subject of the present paper, but it is confined in a rather special direction. This specialization may be characterized by the aim of avoiding theoretical and computational difficulties and of utilizing the known results of the theory of goodness-of-fit tests as much as possible.

Therefore, we are interested in such equivalent (substitute) hypotheses which are of the form of goodness-of-fit problems. In other words, we want to provide a set of random variables which are distribution-free and independently and identically distributed in the case of the null hypothesis.

As a further specialization, we require that each of these transformed values should represent, in the same way, one of the original sample elements. The purpose of this restriction is that the test statistic made with the transformed variables should approximate the correct test one could form with the knowledge of the unknown constants. Therefore, the properties of the combined test, consisting of the transformation and the testing of the simple hypothesis, will be

near that of the test concerning a simple original hypothesis, and this indirect information is the more important since exact investigations of properties of tests relating to the original hypotheses seem very difficult.

Since the test statistics which we can apply are of different types, the problem of what to consider a good representation is the matter of an arbitrary decision. We shall apply, in the cases to be considered (sections 2–5), the correlation coefficient of the original and transformed values (in the case of null hypothesis) as the measure of goodness of representation. Similarly, the minimum value of these measures will be considered the measure of the goodness of the representation if we replace the original set or its subset by a set of transformed variables. In these cases—and in any case when the nuisance parameters are location and scale parameters—this measure does not depend on the nuisance parameters. Furthermore, in these cases the maximization of the correlation coefficient is equivalent to the minimization of the square of difference between the transformed values and the standardized original values.

In section 8 the goodness of representation in a special case is numerically investigated.

The question whether the substitute simple hypothesis is strictly equivalent to the original null hypothesis in the sense that the latter is not only sufficient but also necessary to the former, is important from the point of view of the biasedness of the test. The conditions of this property, in the case of normal distribution with unknown variance, were stated by Prohorov [12].

Solutions of the problem using the method outlined above have been given until now by Durbin [6], Störmer [15], and the present author [13] independently. The purpose of this paper is the comparison, investigation, and improvement of these results.

The main difference between the method of Durbin and that of the other two authors is that Durbin uses random numbers. Consequently, the number of the transformed variables is the same as that of the original variables in Durbin's method, whereas the other method of transformation decreases the number of variables by the number of unknown constants.

Here we deal with the latter method mainly. But first, in section 2, we make a comparison between the two methods in the case where the underlying distribution is normal, with both parameters unknown, and where the transformation is linear for any sample realization. This is the case considered in [6] in detail.

In sections 3–5 some "optimal" transformations are derived for the case of normal distributions. In section 6 the possibilities of generalization are discussed, and section 7 deals with practical applications.

## 2. Comparison with Durbin's method

Durbin's method is the following [6]: let  $x_1, \dots, x_n$  be independent, normally distributed random variables with unknown expectation and variance,  $\mu$  and

$\sigma^2$ , respectively. Let  $\bar{x}$  and  $s^2$  be the sample mean and empirical variance, respectively.

Durbin considers the transformation where the transformed values  $y_1, \dots, y_n$  are independent  $N(0, 1)$  variables and fulfill the relation

$$(2.1) \quad \frac{y_i - \bar{y}}{s'} = \frac{x_i - \bar{x}}{s}, \quad (i = 1, \dots, n).$$

The variables  $\bar{y}$  and  $s'^2$  are independent and are distributed as the sample mean and empirical variance of a sample of size  $n$  from a standardized normal parent population, that is,  $\bar{y}$  has the distribution  $N(0, 1/n)$ , and  $(n - 1)s'^2$  has  $\chi^2$ -distribution with  $n - 1$  degrees of freedom.

Durbin generates the values of  $\bar{y}$  and  $s'$  by the help of random numbers. Similarly, they may be generated as functions of the sample elements (of those defined above and of additional ones) according to the alternative method.

Given  $\bar{y}$  and  $s'$ ,  $y_1, \dots, y_n$  can be determined by (2.1).

Durbin proved that the  $y_1, \dots, y_n$  so defined are, in fact, independent  $N(0, 1)$  variables. Below we give an alternative, simple proof for this fact. We suppose that  $\bar{y}$  and  $s'$  are independent, have the above mentioned distribution, and  $[y, s']$  is independent of  $[(x_1 - \bar{x})/s, \dots, (x_n - \bar{x})/s]$ .

Let

$$(2.2) \quad \begin{aligned} \mathbf{x} &= [x_1, \dots, x_n], \\ \mathbf{y} &= [y_1, \dots, y_n], \\ \mathbf{v} &= \left[ \frac{x_1 - \bar{x}}{s}, \dots, \frac{x_n - \bar{x}}{s} \right], \end{aligned}$$

and let us denote by  $\mathbf{1}_n$  an  $n$ -way vector with unit components. As it is known,  $\bar{x}$ ,  $s$ , and  $\mathbf{v}$  are completely independent. Let us consider the two following identities:

$$(2.3) \quad \mathbf{y} = \mathbf{1}_n \bar{y} + \mathbf{v} s',$$

$$(2.4) \quad \frac{(\mathbf{x} - \mu \mathbf{1}_n)}{\sigma} = \frac{\mathbf{1}_n (\bar{x} - \mu)}{\sigma} + \frac{\mathbf{v} s}{\sigma}.$$

If  $\mathbf{v}$  and  $[\bar{y}, s']$  are independent, the right-hand sides are identical in distribution; therefore,  $\mathbf{y}$  has the same distribution as  $(\mathbf{x} - \mu \mathbf{1}_n)/\sigma$ , that is,  $y_1, \dots, y_n$  are independent  $N(0, 1)$  variables.

Clearly, condition (2.1) can be expressed equivalently as follows: for any given sample realization, there is a linear relation between the sequences  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$ . This seems a very reasonable property of the transformation which preserves the shape of the empirical distribution function.

Denoting  $v_i = (x_i - \bar{x})/s$ , we obtain for the correlation coefficient of  $x_i$  and  $y_i$ ,

$$(2.5) \quad r(x_i, y_i) = \frac{\text{cov}(\bar{x}, \bar{y})}{\sigma} + D^2(v_i) \left[ E^2(s') + \frac{\text{cov}(s, s')}{\sigma} \right].$$

In the case of Durbin's transformation,  $\text{cov}(\bar{x}, \bar{y}) = \text{cov}(s, s') = 0$ . Both quantities are positive by appropriate choice of the functions defining  $\bar{y}$  and  $s'$

according to the alternative method. This means that the use of the additional sample elements is not merely an alternative way to generate random numbers, the information involved in them is utilized to improve the goodness of the transformation.

### 3. The case of unknown expectation

In this section we formulate our problem somewhat more generally than in the other cases. Namely, we include cases where the difference of the number of the original and transformed values is larger than 1, the number of unknown parameters.

Let us suppose we have  $n + k$  random variables  $x_1, \dots, x_{n+k}$  which are independently normally distributed with common unknown expectation  $\mu$  and known variance  $\sigma^2$ , and we want to provide the transformed variables  $y_1^{(1)}, \dots, y_n^{(1)}$  which have the expectation and variance  $\sigma^2$ .

Similarly to (2.1), a possible way of defining the variables  $y_i^{(1)}$  is to define them by the equation

$$(3.1) \quad y_i^{(1)} - \bar{y}^{(1)} = x_i - \bar{x}$$

where  $\bar{y}^{(1)}$  has the distribution  $N(0, (\sigma^2/n))$  and  $\bar{x} = (x_1 + \dots + x_n)/n$  as before. The variable  $\bar{y}^{(1)}$  may be suitably defined by the formula

$$(3.2) \quad \bar{y}^{(1)} = (\bar{x} - \bar{x}_k) \sqrt{\frac{k}{n+k}},$$

whereas

$$(3.3) \quad \bar{x}_k = \frac{x_{n+1} + \dots + x_{n+k}}{k}.$$

It is shown below that the transformation (3.1) is optimal in the adopted sense.

**THEOREM 3.1.** *Let  $x_1, \dots, x_{n+k}$  be independent random variables with common distribution, having the expectation  $\mu$  and variance  $\sigma^2$ , and let  $y_1^{(1)}, \dots, y_n^{(1)}$  be defined by (3.1) and (3.2). Let  $z_i = z_i(x_1, \dots, x_{n+k})$ , ( $i = 1, \dots, n$ ) functions of  $x_1, \dots, x_{n+k}$  such that if  $x_1, \dots, x_{n+k}$  are normally distributed, then  $z_1, \dots, z_n$  are independent with the common distribution  $N(0, \sigma^2)$ .*

*Then if  $x_1, \dots, x_{n+k}$  are normally distributed or  $z_1, \dots, z_n$  are linear functions of  $x_1, \dots, x_{n+k}$ , then*

$$(3.4) \quad \min_{1 \leq i \leq n} r(z_i, x_i) \leq \min_{1 \leq i \leq n} r(y_i^{(1)}, x_i).$$

*If in (3.4) equality holds, then it is valid not only for the minima, but for any  $i$ .*

**PROOF.** Suppose  $x_1, \dots, x_{n+k}$  are normally distributed. It follows from a theorem of Lukács [8] that  $z_i$  has constant regression on the mean

$$(3.5) \quad \bar{X} = \frac{x_1 + \dots + x_{n+k}}{n+k},$$

and therefore,  $E(z_i \bar{X}) = E(\bar{X} E(z_i)) = 0$ .

Since  $z_1, \dots, z_n$  are independent random variables, we have for any random variable  $u$ ,

$$(3.6) \quad \sum_{i=1}^n [r(z_i, u)]^2 \leq 1.$$

It follows that for some  $i$ , ( $1 \leq i \leq n$ ),

$$(3.7) \quad r(z_i, u) \leq n^{-1/2}$$

where strict inequality holds unless equality holds for any  $i$ . We choose

$$(3.8) \quad u = \sum_{j=1}^n x_j - n\bar{X}.$$

We have

$$(3.9) \quad r(x_i - \bar{X}, u) = \sqrt{\frac{k}{n(n+k-1)}},$$

$$(3.10) \quad r(z_i, x_i - \bar{X}) = [E(x_i, z_i)/\sigma] \sqrt{\frac{n+k}{n+k-1}}.$$

Using the triangle inequality which is valid to any triplet of random variables (see [13], p. 271)

$$(3.11) \quad \arccos r(x_i - \bar{X}, z_i) \geq |\arccos r(u, z_i) - \arccos r(u, x_i - \bar{X})|,$$

we obtain

$$(3.12) \quad r(x_i, z_i) = E(x_i z_i) \sigma \leq \frac{\sqrt{k} + (n-1)\sqrt{n+k}}{n\sqrt{n+k}} = 1 - \frac{1}{n+k + \sqrt{k(n+k)}},$$

which proves (3.4), provided  $x_1, \dots, x_{n+k}$  are normally distributed. If  $z_1, \dots, z_n$  are linear functions of  $x_1, \dots, x_n$ , the left-hand side of (3.4) does not depend on the distribution of  $x_1, \dots, x_n$ . Thus theorem 3.1 is proved.

In the case of transformation (3.1) the substitute simple hypothesis is equivalent to the original one; in other words,  $y_i^{(1)}, \dots, y_n^{(1)}$  are normally distributed if and *only if*  $x_1, \dots, x_{n+k}$  are. This follows easily from the theorem of Cramér.

We remark that the normality of  $x_1, \dots, x_{n+k}$  is necessary for the independence of  $y_i^{(1)}, \dots, y_n^{(1)}$  as well. This follows from a theorem of Skitovich [14].

#### 4. The case of unknown variance

Let us consider now the case when (according to the null hypothesis)  $x_1, \dots, x_{n+1}$  are independently normally distributed with common expectation 0 and unknown variance  $\sigma^2 (> 0)$ . We want to obtain independent  $N(0, 1)$  normally distributed variables  $y_i^{(2)}, \dots, y_n^{(2)}$ . The way corresponding to (2.1) is to define them by the equation

$$(4.1) \quad \frac{y_i^{(2)}}{s'_0} = \frac{x_i}{s_0} \quad (i = 1, \dots, n)$$

where

$$(4.2) \quad s_0 = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}},$$

and  $s'_0$  is a function of  $s_0$  and  $x_{n+1}$ , and  $ns_0'^2$  has  $\chi^2$ -distribution with  $n$  degrees of freedom.

Heuristic arguments lead to the choice

$$(4.3) \quad s'_0 = \psi_n \left( \frac{|x_{n+1}|}{s_0} \right)$$

where  $\psi_n(t)$  is a monotone decreasing function. This condition, and the distribution of  $|x_{n+1}|/s_0$  and  $s'_0$ , determine the function  $\psi_n(t)$  completely (see formula (7.4)).

The transformation gives an indefinite result in the case when  $x_1 = \cdots = x_n = 0$ ; the probability of this event is 0 under the null hypothesis. For this case let us define  $y_1^{(2)} = \cdots = y_n^{(2)} = 0$ .

The result we give below concerning the optimality of (4.1) is of weaker character than that of the preceding section. Now the class of admitted alternatives is more restricted; we suppose both linearity for any sample realization and scale invariance. The latter supposition means, equivalently, that the transformed values are of structure  $d'_{n+1}$  (see [2]).

**THEOREM 4.1.** *Let  $x_1, \cdots, x_{n+1}$  be independent  $N(0, \sigma^2)$  random variables, and let  $y_1^{(2)}, \cdots, y_n^{(2)}$  be defined by (4.1), (4.2) and (4.3), and let  $z_1, \cdots, z_n$  be defined by*

$$(4.4) \quad \frac{z_i}{s_0''} = \frac{x_i}{s_0}, \quad (i = 1, \cdots, n)$$

where  $s_0'' = s_0''(x_1, \cdots, x_{n+1})$  is a scale invariant function of  $x_1, \cdots, x_{n+1}$ ; that is,

$$(4.5) \quad s_0''(x_1, \cdots, x_{n+1}) \equiv s_0''(cx_1, \cdots, cx_{n+1}) \quad \text{for any } c > 0,$$

and the function is such as to assure that  $s_0''$  is independent of the vector  $[x_1/s_0, \cdots, x_n/s_0]$  and its distribution is a  $\chi^2$ -distribution with  $n$  degrees of freedom. Then

$$(4.6) \quad r(z_i, x_i) \leq r(y_i^{(2)}, x_i), \quad (i = 1, \cdots, n).$$

**PROOF.** Let us denote

$$(4.7) \quad \frac{x_i}{s_0} = \xi_i, \quad (i = 1, \cdots, n+1)$$

and

$$(4.8) \quad \zeta = ns_0^2 + x_{n+1}^2.$$

It follows from (4.5) that

$$(4.9) \quad s_0''(x_1, \cdots, x_{n+1}) = s_0''(\xi_1, \cdots, \xi_{n+1}).$$

It follows further that the following pairs of random variables (vectors) are pairwise independent:

$$(4.10) \quad \begin{aligned} &\zeta \text{ and } [\xi_1, \dots, \xi_{n+1}], \\ &\xi_i \text{ and } [\xi_{n+1}, \zeta], \quad (i = 1, \dots, n), \\ &\xi_i \text{ and } s_0''(\xi_1, \dots, \xi_{n+1}), \quad (i = 1, \dots, n). \end{aligned}$$

(However,  $\xi_i$  and the vector  $[\xi_{n+1}, s_0'']$  do not have to be independent.)

One has

$$(4.11) \quad \begin{aligned} x_i &= \xi_i \sqrt{\frac{\zeta}{n + \xi_{n+1}^2}}, \\ y_i^{(2)} &= \xi_i \psi_n(|\xi_{n+1}|), \\ z_i &= \xi_i s_0''(\xi_1, \dots, \xi_{n+1}). \end{aligned}$$

Therefore,

$$(4.12) \quad E(x_i y_i^{(2)}) = E(\xi_i^2) E(\zeta^{1/2}) E\left(\frac{\psi_n(|\xi_{n+1}|)}{\sqrt{n + \xi_{n+1}^2}}\right),$$

$$(4.13) \quad E(x_i z_i) = E(\xi_i^2) E(\zeta^{1/2}) E\left(E\left(\frac{s_0''(\xi_1, \dots, \xi_{n+1})}{\sqrt{n + \xi_{n+1}^2}} \mid \xi_i\right)\right).$$

The distribution of  $\psi_n(|\xi_{n+1}|) = s_0'$  agrees with the distribution of

$$(4.14) \quad s_0''(\xi_1, \dots, \xi_{n+1}).$$

Since  $\xi_i$  is independent of  $s_0''(\xi_1, \dots, \xi_{n+1})$ , the conditional distribution of (4.14) given  $\xi_i$  is the same. Therefore, since  $E(\xi_i^2) > 0$ ,  $E(\zeta^{1/2}) > 0$ , it suffices to prove for the conditional correlation coefficient of (4.14) and  $(n + \xi_{n+1}^2)^{-1/2}$  given  $\xi_i$  that

$$(4.15) \quad r(s_0''(\xi_1, \dots, \xi_{n+1}), (n + \xi_{n+1}^2)^{-1/2} \mid \xi_i) \leq r(\psi_n(|\xi_{n+1}|), (n + \xi_{n+1}^2)^{-1/2}).$$

This follows, however, from a theorem of Fréchet and Bass ([1], p. 640) according to which in case of a two-way variable having given marginal distributions, the correlation coefficient is maximal if one of the marginal variables is a monotone increasing function of the other. This is, however, the case between  $\psi_n(|\xi_{n+1}|)$  and  $(n + \xi_{n+1}^2)^{-1/2}$ , and thus, (4.15) is true, and theorem 4.1 is proven.

**5. The case of two unknown parameters**

Let  $n$  be given and let  $k = 2$ . Let us join to the transformation (3.1), where  $\bar{y}^{(1)}$  is defined by (3.2), the equality

$$(5.1) \quad y_{n+1}^{(1)} = 2^{-1/2}(x_{n+2} - x_{n+1}),$$

and let us denote this transformation by  $T_1$ ,

$$(5.2) \quad [y_1^{(1)}, \dots, y_{n+1}^{(1)}] = T_1[x_1, \dots, x_{n+2}].$$

Let us denote by  $T_2$  the transformation (4.1) where  $s_0'$  is defined by (4.3), that is

$$(5.3) \quad [y_1^{(2)}, \dots, y_n^{(2)}] = T_2[x_1, \dots, x_{n+1}].$$

Let  $y_1^{(3)}, \dots, y_n^{(3)}$  be defined by the formula

$$(5.4) \quad [y_1^{(3)}, \dots, y_n^{(3)}] = T_2 T_1[x_1, \dots, x_{n+2}].$$

Alternatively, (5.4) can be written in the form of (2.1):

$$(5.5) \quad \frac{y_i^{(3)} - \bar{y}^{(3)}}{s^{(3)}} = \frac{x_i - \bar{x}}{s}, \quad (i = 1, \dots, n)$$

where

$$(5.6) \quad \begin{aligned} \bar{y}^{(3)} &= (2\bar{x} - x_{n+1} - x_{n+2})q(2[n + 2])^{-1/2}, \\ s^{(3)} &= sq, \end{aligned}$$

while

$$(5.7) \quad n\bar{x} = \sum_{i=1}^n x_i, \quad (n - 1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$(5.8) \quad q = \psi_n \left( |x_{n+2} - x_{n+1}| \sqrt{\frac{n(n+2)}{2(n-1)(n+2)s^2 + n(2\bar{x} - x_{n+1} - x_{n+2})^2}} \right) \\ \times \sqrt{\frac{2n(n+2)}{2(n-1)(n+2)s^2 + n(2\bar{x} - x_{n+1} - x_{n+2})^2}}.$$

Now we have the following result characterizing the property of the above transformation.

**THEOREM 5.1.** *Let  $x_1, \dots, x_{n+2}$  be independent  $N(\mu, \sigma^2)$  random variables, and let  $y_1^{(3)}, \dots, y_n^{(3)}$  be defined by (5.4) and let  $z_1, \dots, z_n$  be defined by the relation*

$$(5.9) \quad z_i = \frac{u_i}{(u_1^2 + \dots + u_n^2)^{1/2}} \varphi(u_1, \dots, u_{n+1}), \quad (i = 1, \dots, n)$$

where  $u_1, \dots, u_{n+1}$  are linear functions of  $x_1, \dots, x_{n+2}$  such that they are independently distributed with the common distribution  $N(0, \sigma^2)$ , and  $\varphi(u_1, \dots, u_{n+1})$  is a scale invariant function (for any  $c > 0, \varphi(u_1, \dots, u_{n+1}) = \varphi(cu_1, \dots, cu_{n+1})$ ), and such that  $z_1, \dots, z_n$  is a sequence of independent  $N(0, 1)$  random variables. Then

$$(5.10) \quad \min_{1 \leq i \leq n} r(z_i, x_i) \leq \min_{1 \leq i \leq n} r(y_i^{(3)}, x_i).$$

If in this formula equality holds, then it is valid not only for the minima but for any  $i$ .

**PROOF.** Let us introduce the following notation:

$$(5.11) \quad \xi_i = \frac{y_i^{(1)}}{s_0}, \quad (i = 1, \dots, n + 1)$$

where  $y_1^{(1)}, \dots, y_n^{(1)}$  are defined by formula (4.1),  $y_{n+1}^{(1)}$  by formula (5.1),

$$(5.12) \quad \begin{aligned} \xi'_i &= \frac{u_i n^{1/2}}{(u_1^2 + \dots + u_n^2)^{1/2}}, \quad (i = 1, \dots, n + 1), \\ \bar{X} &= \frac{x_1 + \dots + x_{n+2}}{n + 2}, \\ \zeta &= \sum_{i=1}^{n+2} x_i^2 - n\bar{X}^2. \end{aligned}$$

It follows (as in the proof of theorem 3.1) that  $\bar{X}$  is independent of  $u_1, \dots, u_{n+1}$  which implies that  $x_i$  may be written in the form

$$(5.13) \quad x_i = \sum_{j=1}^{n+1} a_{ij}u_j + a_{in}\bar{X}, \quad (i = 1, \dots, n + 2)$$

and that  $u_1^2 + \dots + u_{n+1}^2 = \zeta$ .

Moreover, we have (see the proof of the preceding theorem) that

$$(5.14) \quad \varphi(u_1, \dots, u_{n+1}) = \varphi(\xi'_1, \dots, \xi'_{n+1}).$$

And further, the following pairs of random variables (vectors) are independent for  $(i = 1, \dots, n)$ :

$$(5.15) \quad \begin{aligned} &[\bar{X}, \zeta] \quad \text{and} \quad [\xi_1, \dots, \xi_{n+1}], \\ &[\bar{X}, \zeta] \quad \text{and} \quad [\xi'_1, \dots, \xi'_{n+1}], \\ &\xi_i \quad \text{and} \quad [\xi_{n+1}, \zeta], \\ &\xi'_i \quad \text{and} \quad \xi'_{n+1}, \\ &\xi'_i \quad \text{and} \quad \varphi(\xi'_1, \dots, \xi'_{n+1}). \end{aligned}$$

Evidently  $E(\xi_i \xi_j) = E(\xi'_i \xi'_j) = 0$  for  $i \neq j; i, j = 1, \dots, n + 1$ .

We have

$$(5.16) \quad \begin{aligned} x_i &= \left( \xi_i - \frac{\sum_{j=1}^n \xi_j}{n + 2 + [2(n + 2)]^{1/2}} \right) \left( \frac{\zeta}{n + \xi_{n+1}^2} \right)^{1/2} + \bar{X} \\ &= \sum_{j=1}^{n+1} a_{ij} \xi'_j \left( \frac{\zeta}{n + \xi_{n+1}^2} \right)^{1/2} + \bar{X}, \end{aligned}$$

$$(5.17) \quad y_i^{(3)} = \xi_i \psi_n(|\xi_{n+1}|)$$

$$(5.18) \quad z_i = \xi'_i \varphi(\xi'_1, \dots, \xi'_{n+1}) / n^{1/2}, \quad (i = 1, \dots, n).$$

Therefore, for  $(i = 1, \dots, n)$ ,

$$(5.19) \quad \begin{aligned} E(x_i y_i^{(3)}) &= [1 - (n + 2 + [2(n + 2)]^{1/2})^{-1}] E(\xi_i^2) E(\zeta^{1/2}) E \left( \frac{\psi_n(|\xi_{n+1}|)}{(n + \xi_{n+1}^2)^{1/2}} \right), \end{aligned}$$

$$(5.20) \quad E(x_i z_i) = a_{ii} E(\xi_i'^2) E(\zeta^{1/2}) E \left( E \left( \frac{\varphi(\xi'_1, \dots, \xi'_{n+1})}{[n(n + \xi_{n+1}^2)]^{1/2}} \mid \xi'_i \right) \right).$$

Now  $a_{ii} = r(u_i, x_i)$  and  $E(\xi_i^2) = E(\xi_i'^2)$ . Using (3.4) and (4.15) we obtain (5.10). The last assertion of the theorem follows from the similar assertion of theorem 3.1.

### 6. The general case

This section deals briefly with the possibilities of extending the method to more general cases.

In general, let  $x_1, \dots, x_{n+k}$  be independent random variables, having under the null hypothesis the distribution  $F(x, \mu_i)$ ,  $(i = 1, \dots, n + k)$ . Suppose the shape of the function  $F(x, \cdot)$  is known, and the parameter vectors  $\mu_i$  are to be known functions of the unknown constant vector  $\theta$ ,

$$(6.1) \quad \mu_i = \mu_i(\theta)$$

where  $\theta$  is a point of the parameter space  $\Theta$ , ( $\theta \in \Theta$ ). Let us suppose there exists a  $\theta \in \Theta$  for which  $\mu_i(\theta_0)$  does not depend on  $i$ , that is

$$(6.2) \quad \mu_i(\theta_0) = \mu_0, \quad (i = 1, \dots, n+k).$$

In other words, if  $\theta = \theta_0$ , the variables  $x_1, \dots, x_{n+k}$  are independent, identically distributed random variables.

Let us suppose, further, the existence of a distribution-free statistic on the first  $n$  sample elements. Let us denote it by  $\mathbf{T} = \mathbf{T}(\mathbf{X})$ , where  $\mathbf{X} = [x_1, \dots, x_n]$ .

Let  $\Omega$  be a random vector whose distribution is known. Let us define the function  $\mathbf{Y}(\mathbf{T}, \Omega)$  in such a way that (cf. [6])

$$(6.3) \quad \begin{aligned} & \text{“cond. distr. of } \mathbf{Y}(\mathbf{T}, \Omega) \text{ given } T\text{”} \\ & = \text{“cond. distr. of } \mathbf{X} \text{ given } \mathbf{T}, \theta_0\text{.”} \end{aligned}$$

This is in general possible, in many ways. Computational difficulties, however, may arise in determining the conditional distribution of  $\mathbf{X}$ .

Since, according to (6.3), the conditional distributions of  $\mathbf{X}$  and  $\mathbf{Y}$  agree for all given values of  $\mathbf{T}$ , the unconditional distributions have to agree too; that is, its components  $y_1, \dots, y_n$  are independent, identically distributed random variables having the distribution  $F(x, \mu_0)$ .

The random vector  $\Omega$  may be generated by random numbers, or it may be a function of the variables  $x_1, \dots, x_{n+k}$ , such that its distribution is independent of  $\mathbf{T}$  and  $\theta$ .

Thus the original null hypothesis may be replaced by the simple hypothesis that the variables  $y_1, \dots, y_n$  are independently distributed with the distribution  $F(x, \theta_0)$ .

The adequateness of representation of the original hypothesis by this substitute may be investigated in the particular cases. Clearly, it depends on the forms of the function  $F(x, \mu)$  and the distribution-free statistic  $\mathbf{T}$ .

EXAMPLES. (a) *Gamma-distribution*. Let  $x_1, \dots, x_{n+1}$  be independent random variables with the common density function

$$(6.4) \quad F(x, \mu) = \begin{cases} 0 & \text{for } x \leq 0, \\ \frac{\mu^\lambda}{\Gamma(\lambda)} \int_0^x t^{\lambda-1} e^{-\mu t} dt & \text{for } x > 0. \end{cases}$$

Here  $\lambda$  is known and  $\mu$  is the nuisance parameter.

The appropriate transformation is

$$(6.5) \quad y_i^{(4)} = \frac{x_i}{x_1 + \dots + x_n} \psi_{\lambda, n} \left( \frac{x_{n+1}}{x_1 + \dots + x_{n+1}} \right), \quad (i = 1, 2, \dots, n)$$

where the function  $\psi_{\lambda, n}(t)$  is a monotone decreasing function assuring that

$$(6.6) \quad \psi_{\lambda, n} \left( \frac{x_{n+1}}{x_1 + \dots + x_{n+1}} \right)$$

has the same distribution as  $x_1 + \dots + x_n$  if  $\mu = 1$ . This condition determines the function  $\psi_{\lambda,n}(t)$  completely (see formula (7.13)).

Equality (6.5) yields independent gamma variables with the original shape parameter and unit scale parameter, that is, having the distribution function

$$(6.7) \quad F(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{1}{\Gamma(\lambda)} \int_0^x t^{\lambda-1} e^{-t} dt & \text{if } x > 0. \end{cases}$$

A theorem similar to theorem 4.1 can be proved concerning the optimality of this transformation.

(b) *Small subsamples.* In the case of normal null hypothesis, if we have several small subsamples with different nuisance parameters, the transformations given in sections 3-5 can be separately applied for each subsample.

(c) *Analysis of variance model.* In [13] a transformation was given for the case of a two-way classification, one observation per cell, when the distribution of the error term is normal, according to the null hypothesis [13], (2). If the variance is unknown as well, we may apply transformation (4.1) as a second step.

Also, this transformation enables, as mentioned there, to test the homogeneity of variances as well in the mentioned cases. In the cases of alternative hypothesis, the variances of the transformed variables will differ from that of the corresponding original ones, but the transformation preserves the magnitude order of the variances so that the hypothesis concerning the homogeneity of variances in the original and transformed data are equivalent.

An alternative solution for this problem has been given by N. L. Johnson [3].

### 7. Practical applications

It is convenient to make some slight modifications on the formulae given in sections 3-7 for the purpose of practical application. In the modified formulae we denote by  $n$  the original sample size.

(a) If we have a random simple sample  $x_1, \dots, x_n$ , which comes, according to the null hypothesis, from a normal distribution (expectation and variance being unspecified), apply the following transformation:

$$(7.1) \quad y_i = \frac{x_i - \bar{x}''}{S} \psi_{n-2} \left( \frac{|x_{n-1} - x_n| \sqrt{n-2}}{S\sqrt{2}} \right), \quad (i = 1, \dots, n-2)$$

where

$$(7.2) \quad \bar{x}'' = \frac{\sum_{i=1}^n x_i + \sqrt{\frac{n}{2}} (x_{n-1} + x_n)}{n + \sqrt{2n}},$$

$$(7.3) \quad S = \sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 - \frac{1}{2} (x_{n-1} - x_n)^2},$$

and the function  $\psi_\nu(t)$  is defined by the relation

$$(7.4) \quad Q([\psi_\nu(t)]^2|\nu) = 2P(t|\nu) - 1,$$

while  $\psi_\nu(t) \geq 0$ ,

$$(7.5) \quad P(t|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}} du,$$

$$(7.6) \quad Q(t|\nu) = \frac{1}{2^{1/2}\Gamma\left(\frac{\nu}{2}\right)} \int_t^\infty e^{-\frac{u}{2}} u^{\frac{\nu-1}{2}} du.$$

The functions  $P(t|\nu)$  and  $1 - Q(t|\nu)$  are the distributions of the Student distribution and the  $\chi^2$ -distribution, respectively, each with  $\nu$  degrees of freedom. The notations ( $P(t|\nu)$  and  $Q(t|\nu)$ ) agree with that of the *Biometrika Tables* [9], tables 9 and 7, respectively, where they are tabulated. These tables may be used in applications of (7.1). (The formula (7.1) is an equivalent form of formula (5.4)).

(b) If, according to the null hypothesis, our random simple sample  $x_1, \dots, x_n$  comes from a normal distribution with specified variance (expectation being unspecified), then we have to use the transformation

$$(7.7) \quad y_i = x_i - \bar{x}'$$

where

$$(7.8) \quad \bar{x}' = \frac{\sum_{i=1}^n x_i + x_n \sqrt{n}}{n + \sqrt{n}}.$$

This transformation was given in [13] as formula (1) and is the special case of formula (3.1) of the present paper.

(c) If, according to the null hypothesis, our random simple sample  $x_1, \dots, x_n$  comes from a normal distribution with expectation 0 (variance being unspecified), then the transformation to be used is

$$(7.9) \quad y_i = \frac{x_i}{x_i^2 + \dots + x_{n-1}^2} \psi_{n-1} \left( \frac{|x_n| \sqrt{n-1}}{\sqrt{x_1^2 + \dots + x_{n-1}^2}} \right),$$

the definition of  $\psi_\nu(t)$  being given in (7.4).

(d) Let  $x_{i1}, \dots, x_{in_i}$ , ( $i = 1, 2, \dots, m$ ) be independent random variables. For each given  $i$  the  $n_i$  random variables have the common distribution function  $F_i(x)$ . Some of the variances of the distributions are supposed to be equal but unknown.

The hypothesis to be tested is that the distributions  $F_i(x)$  are normal. The unknown variance is not specified in the null hypothesis.

In this case we apply the transformation (7.7) for all but one subsample. It is suitable to choose the notation so that  $n_m \geq n_i$ , ( $i = 1, \dots, m-1$ ) and to

apply transformation (7.7) for each of the series  $x_{i1}, \dots, x_{in_i}$  ( $i = 1, 2, \dots, m - 1$ ) and the modified form of (3.1) with  $k = 2$ , that is the transformation

$$(7.10) \quad \begin{aligned} y_i &= x_i - \bar{x}'' \\ y_{n-1} &= \frac{x_{n-1} - x_n}{\sqrt{2}} \end{aligned}$$

for the series  $x_{m1}, \dots, x_{mn_m}$ . Here  $\bar{x}''$  is the same as in (7.2). As a second step we apply (7.9) for the sequence of the resultant  $\sum_{i=1}^m n_i - m$  variables

$$(7.11) \quad y_{11}, \dots, y_{1n_1-1}, y_{21}, \dots, y_{mn_m-2}, y_{mn_m-1}.$$

(e) *Gamma distribution.* The suitable form of the transformation (6.5) is the following:

$$(7.12) \quad y_i = \frac{x_i}{x_1 + \dots + x_{n-1}} \psi_{\lambda, (n-1)} \left( \frac{x_n}{x_1 + \dots + x_n} \right)$$

where the function  $\psi_{\lambda, \nu}(t)$  is defined by the relation

$$(7.13) \quad 1 - I \left( \frac{\psi_{\lambda, \nu}(t)}{\sqrt{\nu\lambda} - 1}, \nu\lambda - 2 \right) = I_t(\lambda, \nu\lambda),$$

whereas

$$(7.14) \quad I(u, p) = \frac{\int_0^{u\sqrt{p+1}} e^{-t^{p+1}} dt}{\int_0^\infty e^{-t^{p+1}} dt}$$

is a gamma distribution function, tabulated in [11], and

$$(7.15) \quad I_t(p, q) = \frac{\int_0^t x^{p-1}(1-x)^{q-1} dx}{\int_0^1 x^{p-1}(1-x)^{q-1} dx}$$

is a beta distribution function, tabulated in [10].

### 8. A comparative example

The aim of the following example is to illustrate in a special case that the transformation fulfilling our optimality criterion gives, in fact, better representation than the previously known ones.

Let  $x_{i1}, x_{i2}$ , ( $i = 1, \dots, m$ ) be independent random variables with distributions

$$(8.1) \quad \begin{aligned} P(x_{i1} = a_i) &= \frac{2}{3}, \\ P(x_{i2} = -2a_i) &= \frac{1}{3}, \quad (a_i > 0), \quad (i = 1, \dots, m; j = 1, 2). \end{aligned}$$

The first three columns of table I show the distributions of the transformed values yielded, by (4.1) and (3) of [15], respectively, ( $n = 2$ ):

TABLE I

Probability	Possible values of the transformed values using transformation	
	(4.1)	(3) of [15]
2/9	-1.047	-1.447
1/9	-.675	-.319
2/9	.378	.457
4/9	.675	1.150

(The second factor in the right-hand side of formula (4) of [15] should be read  $\varphi_m(y_m/(y_1^2 + \dots + y_m^2)^{1/2})$ .)

If, according to [6], we use random numbers, the distribution of the transformed values will be a continuous distribution, the mixture of four  $\chi$ -distributions. The density of this distribution is

$$(8.2) \quad \frac{5x}{18} e^{-5x^2/8} + \frac{2x}{9} e^{-x^2/4} \quad \text{for } x \geq 0,$$

$$(8.3) \quad -\frac{5x}{36} e^{-5x^2/16} - \frac{x}{18} e^{-x^2/4} \quad \text{for } x < 0.$$

Apparently the first of these three distributions gives the best representation for the original ones.

## REFERENCES

- [1] JEAN BASS, "Sur la compatibilité des fonctions de répartition," *C. R. Acad. Sci. Paris*, Vol. 240 (1955), pp. 839-841.
- [2] C. B. BELL, "Some basic theorems of distribution-free statistics," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 150-156.
- [3] T. CALINSKY, "On a certain statistical method of investigating interaction in serial experiment with plant varieties," *Bull. Acad. Polon. Sci. Sér. Sci. Biol.*, Vol. 8 (1960), pp. 565-568.
- [4] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.
- [5] D. A. DARLING, "The Cramér-Smirnov test in the parametric case," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 1-20.
- [6] S. DURBIN, "Some methods of constructing exact tests," *Biometrika*, Vol. 48 (1961), pp. 41-55.
- [7] M. KAC, J. KIEFER, and J. WOLFOWITZ, "On tests of normality and other tests on goodness of fit based on distance methods," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 189-211.
- [8] E. LUKÁCS, "On distribution-free partition statistics for the normal family," *Bull. Inst. Internat. Statist.*, Vol. 36 (1957), pp. 37-42.
- [9] E. S. PEARSON and H. O. HARTLEY, *Biometrika Tables for Statisticians*, Vol. I, Cambridge, Cambridge University Press, 1956.
- [10] KARL PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge, Cambridge University Press, 1934.
- [11] ———, *Tables of the Incomplete  $\Gamma$ -Function*, Cambridge, Cambridge University Press, 1934.

- [12] YU. PROHOROV, "Some characterization theorems in statistics," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1966, Vol. I, pp. 341-349.
- [13] K. SARKADI, "On testing for normality," *Publ. Math. Inst. Hungar. Acad. Sci.*, Vol. 5 (1960), pp. 269-275.
- [14] V. P. SKITOVICH, "Linear forms of independent random variables and the normal distribution," *Izv. Acad. Nauk USSR*, Vol. 18 (1954), pp. 185-200. (In Russian.)
- [15] HORAND STÖRMER, "Ein Test zum Erkennen von Normalverteilungen," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 2 (1964), pp. 420-428.