ON THE TEST OF A HYPOTHESIS CONCERNING TWO INDEPENDENT FREQUENCY DISTRIBUTIONS

By D. N. Lal

Department of Statistics, Patna University, Patna

1. In this paper we shall develop a test of hypothesis that two independent samples given in sets of frequencies classified into the same k frequency classes may be said to belong to the same population, it being assumed that the samples are large and the law of the distribution of the population is known except for certain unspecified parameters.

It is inherent in the above problem that even when the samples belong to the different populations the nature of their distributions, *i.e.*, their mathematical forms remain the same, *e.g.*, if the law of distribution is known to be normal we assume that both the samples originate from normal populations—these populations may differ in their parameters, *i.e.*, in their means or in their variances or in both.

At first sight it may seem that the problem is identical with one due to Karl Pearson (*Biometrika*, 1911), but it will be realised that in Pearson's problem the " population is one however of which we have no *a priori* experience " while in our case we start with a knowledge of the nature of the population. K. Pearson referred to this problem of 1911 again in (*Biometrika*, 1932) and contradicted his previous results of 1911 in an attempt to solve the problem by minimising χ^2. This explains why the two results differ.

The problem under different conditions has been studied by Thompson (1938), Wald and Wolfowitz (1940), Dixon (1940). In all these it has been assumed that the nature of the distribution is not known or we are not interested in it but when we know the nature of the distribution these solutions will not meet the requirements of the problem.

Fundamentally, therefore, the problem is one of estimation of the unknown parameters (or their functions) of the population from the given samples.

2. This problem can be solved following the method due to Wilks (1938) but this seems to be difficult in practice. We proceed as follows in the case of s parameters.

7. Summary

By making use of certain results established by the author it has been shown that the discussion of a number of distributions considered by Tukey, Stevens, Mann, Dixon, Kermack and McKendrick, and Kendall and Babington Smith can be simplified to a considerable extent. The results obtained by these authors have been extended to the more general case where the samples consist of observations with values $\theta_1, \theta_2. \ldots \theta_k$ either with fixed probabilities $p_1, p_2, \ldots p_k$ or are such that there are $n_1, n_2, \ldots n_k$ observations belonging to these values and are subject to the condition $\Sigma_1^k n_r = n$, the total number of observations in the samples. A new distribution on the number of positive or negative differences between k samples has also been considered. This distribution would enable us to test the significance of the differences between k samples. It has been indicated that this method can be extended for examining a randomised block experiment with sufficient number of replications.

References

Dixon, W. J. .. "A criterion for testing the hypothesis that two samples are from the same population," *Ann. Math. Stat.*, 1940, **11**, 199.

Fréchet, M. .. "Les probabilities associeés á Un systéme d événents compatibles et dépendants," *Actualitiés Scientifiques et Industrilles*, Paris, 1940, No. 859.

Kendall, M. G. .. *The advanced Theory of Statistics*, I, Charless Griffin & Co., London, 1945, 388.

Krishna Iyer, P. V. .. "The theory of probability distribution of points on a line," *J. Ind. Soc. Agri. Stat.*, 1948, **1**, 173.

——— .. "The theory of probability distribution of points on a lattice," *Ann. Math. Stat.*, 1950, **21**, 198.

——— .. "Further contributions to the theory of probability distributions of points on a line," *J. Ind. Soc. Agri. Stat.*, 1950, **2** & **3**, 141 and 80.

Kermack, W. O. and McKendrick, A. G. "Tests for randomness in a series of numerical observations," *Proc. Roy. Soc.*, Edinburgh, 1937, **57**, 228 and 332.

Moore, G. H. and Wallis, W. A. "Time series significance tests based on signs of differences," *J. Amer. Stat. Assoc.*, 1943, **38**, 153.

Mann, H. B. .. "On a test of randomness based on signs of differences," *Ann. Math. Stat.*, 1945, **16**, 193.

Stevens, W. L. .. "Distribution of groups in a sequence of alternatives," *Ann. Eug.*, 1939, **9**, 10.

Tukey, J. W. .. "Moments of random group size distribution," *Ann. Math. Stat.*, 1949, **20**, 523.

Let our two independent samples classified into the same k different categories be as follows:

Samples	Limits of class frequencies						Total
	$l_1 - l_2$	$l_2 - e_3$	. . .	$l_1 - l_{i+1}$	. . .	$l_h - l_{h+1}$	
First ..	n_{11}	n_{12}	. . .	n_{1i}	. . .	n_{1h}	N_1
Second ..	n_{21}	n_{22}	. . .	n_{2i}	. . .	n_{2h}	N_2
Total ..	n_1	n_1	. . .	n_i	. . .	n_p	N

To simplify writing let us put

n_{ji} = number of observations of j-th sample belonging to i-th class frequency.

N_j = total number of observations in the j-th sample.

p_{ji} = probability of an observation of j-th sample falling in the i-th class frequency.

$\phi\,(x, \theta_{j1}, \theta_{j2}, \ldots \theta_{js})$ be the assumed law of distribution of the j-th sample; θ_{jr} $(r = 1, \ldots s)$ being its s unknown independent parameters necessary to specify the law.

As we are dealing with two observed samples classified into the same k different categories suffixes j and i will run over $j = 1, 2$ and $i = 1, 2, \ldots k$.

Thus we have

$$n_i = \sum_{j=1}^{2} n_{ji} \qquad (i = 1, 2, \ldots k)$$

$$N_j = \sum_{i=1}^{k} n_{ji} \qquad (j = 1, 2)$$

$$N = \sum_j N_j = \sum_{j,i} n_{ji} = \sum_i n_i$$

and

$$p_{ji} = \int_{l_i}^{l_{i+1}} \phi\,(x, \theta_{j1}, \theta_{j2}, \ldots \theta_{js})\,dx \qquad (j = 1, 2,;\; i = 1, 2, \ldots$$

$$= p_{ji}\,(\theta_{j1}, \ldots \theta_{js}) \text{ say.}$$

3. The number of independent parameters, s, necessary to specify he law is subject to the restriction $s \leqslant k - 1$, k being the number of different categories. This restriction on the value of s will be made clear in Section 5.

(*a*) *First method.*—Assuming n_{ji} to be large, the probability P of drawing the two given samples is denoted by

$$2 \log P = 2L \doteqdot \text{const.} - \sum_{j} \sum_{i} \frac{x_{ji}^{2}}{N_j p_{ji}}, \qquad (1)$$

where $x_{ji} = n_{ji} - N_i p_{ji}$.

Hence

$$\left.\begin{aligned} \frac{\partial L}{\partial \theta_{jr}} &\doteqdot \sum_{i=0}^{k} \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_{ir}}\right) x_{ji} \\ \text{and} \qquad \frac{\partial^2 L}{\partial \theta_{jm} \partial \theta_{jr}} &\doteqdot - N_j \sum_{i=1}^{k} \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_{jm}}\right)\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right) \end{aligned}\right\} \qquad (2)$$

$$j = 1, 2$$

$$m, r = 1, 2, \ldots s$$

neglecting terms proportional to $0\left(\frac{1}{\sqrt{N_j}}\right)$.

Also to estimate θ_{jr} we solve the following set of s equations simultaneously, namely,

$$\frac{\partial L}{\delta \theta_{jr}} = 0 \qquad (r = 1, 2, \ldots s). \qquad (3)$$

Thus if $\check{\theta}_{1r}$ and $\check{\theta}_{2r}$ $(r = 1, 2, \ldots s)$ are the maximum likelihood estimates of θ's it follows that these are respectively the solutions the sets of equations

$$\left.\begin{aligned} \frac{\partial L}{\partial \theta_{1r}} &= 0 \qquad (r = 1, 2, \ldots s) \\ \text{and} \qquad \frac{\partial L}{\partial \theta_{2r}} &= 0 \qquad (r = 1, 2, \ldots s) \end{aligned}\right\} \qquad (4)$$

By Taylor's theorem we have

$$\left(\frac{\partial L}{\partial \theta_{jr}}\right)_{\check{\theta}_{jr}} = \left(\frac{\partial L}{\partial \theta_{jr}}\right) + \sum_{m=1}^{s} (\check{\theta}_{jm} - \theta_{jm}) \frac{\partial^2 L}{\partial \theta_{jm}\, \partial \theta_{jr}} \quad \text{approximately}$$

$$(j = 1, 2;\ r = 1, 2, \ldots s). \qquad (5)$$

Hence from (5) with the help of (4) we get

$$\sum_{m=1}^{s} (\check{\theta}_{jm} - \theta_{jm}) \frac{\partial^2 L}{\partial\theta_{jm}\partial\theta_{jr}} = -\frac{\partial L}{\partial\theta_{jr}} \qquad (6)$$

$$(j = 1, 2;\ r = 1, 2, \ldots s).$$

We have from (6)

$$[\check{\theta}_{jr} - \theta_{jr}] = A_j^{-1}\left[\frac{\partial L}{\partial\theta_{jr}}\right] \quad (j = 1, 2;\ r = 1, 2, \ldots s), \qquad (7)$$

where

$$A_j = \left[-\frac{\partial^2 L}{\partial\theta_{jm}\,\partial\theta_{jr}}\right] \qquad (m, r = 1, 2, \ldots s).$$

In the result (7) we have assumed that A_j is non-singular and we also assume that

$$\frac{\partial^2 L}{\partial\theta_{jm}\,\partial\theta_{jr}} = \frac{\partial^2 L}{\partial\theta_{jr}\,\partial\theta_{jm}}\,.$$

Thus A_j is symmetrical and non-singular matrix. With the help of (2)

$$A_j = N_j P_j \qquad (j = 1, 2), \qquad (8)$$

where

$$P_j = \left[\Sigma\,\frac{1}{p_{ji}}\left(\frac{\partial p_{ji}}{\partial\theta_{jm}}\right)\left(\frac{\partial p_{ji}}{\partial\theta_{jr}}\right)\right] \qquad (m, r = 1, 2, \ldots s).$$

Thus (7) can be written as

$$\left[\check{\theta}_{jr} - \theta_{jr}\right] = \frac{1}{N_j} P_j^{-1}\left[\frac{\partial L}{\partial\theta_{jr}}\right] \quad (j = 1, 2;\ r = 1, 2, \ldots s) \ (9)$$

which has been given by Cramer (1946). Hence from the two sets of equations in (9) we get by subtraction and by using the result (2)

$$\left[(\check{\theta}_{1r} - \check{\theta}_{2r}) - (\theta_{1r} - \theta_{2r})\right] = \frac{1}{N_1} P_1^{-1}\left[\frac{\partial L}{\partial\theta_{1r}}\right] - \frac{1}{N_2} P_2^{-1}\left[\frac{\partial L}{\partial\theta_{2r}}\right]$$

$$= P_1^{-1}\left[\Sigma_i\,\frac{1}{p_{1i}}\left(\frac{\partial p_{1i}}{\partial\theta_{1r}}\right)\frac{x_{1i}}{N_1}\right] - P_2^{-1}\left[\Sigma_i\,\frac{1}{p_{2i}}\left(\frac{\partial p_{2i}}{\partial\theta_{2r}}\right)\frac{x_{2i}}{N_2}\right]$$

$$(r = 1, 2, \ldots s) \qquad (10)$$

On the hypothesis H_0 we have

$$\theta_{1r} = \theta_{2r} = \theta_r \text{ (say)} \qquad (r = 1, 2, \ldots s)$$

therefore

$$p_{1i} = p_{2i} = p_i \text{ (say)} \qquad (i = 1, 2, \ldots k)$$

and thence

$$P_1 = P_2 = P \text{ (say)}.$$

Therefore on the hypothesis H_0 and using the relation

$$x_{ji} = n_{ji} - N_j p_{ji}$$

we have from (10)

$$\left[\breve{\theta}_{1r} - \breve{\theta}_{2r}\right] = P^{-1}\left[\Sigma_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right], \quad (r = 1, 2, \ldots s), \tag{11}$$

where

$$P = \left[\Sigma_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_m}\right)\left(\frac{\partial p_i}{\partial \theta_r}\right)\right] \qquad (m, r = 1, 2, \ldots s).$$

To study the joint distribution of $(\breve{\theta}_{1r} - \breve{\theta}_{2r})$, $(r = 1, 2, \ldots s)$ we obtain the quadratic form in these s normal variates, namely

$$[\breve{\theta}_{1r} - \breve{\theta}_{2r}]' V^{-1} [\breve{\theta}_{1r} - \breve{\theta}_{2r}] \qquad (r = 1, 2, \ldots s), \tag{12}$$

where V is the variance matrix of the variates involved. It is clear that

$$V = V_1 + V_2, \tag{13}$$

where V_j is the variance matrix of $\breve{\theta}_{jr}$ $(j = 1, 2;\ r = 1, 2, \ldots s)$ we also note that

$$\begin{aligned} V_j &= [E(A_j)]^{-1} \\ &= \frac{1}{N_j} P_j^{-1} \\ &= \frac{1}{N_j} P^{-1} \text{ (on .he hypothesis } H_0\text{)}, \end{aligned} \tag{14}$$

where $[E(A_j)]$ means the matrix obtained by replacing the elements of A_j by their expected values and in the case of a single sample will be the elements of A_j themselves.

From (13) and (14) we get

$$V = \sum_{j=1}^{2} \frac{1}{N_j} P_j^{-1} \tag{15}$$

$$= \left(\frac{1}{N_1} + \frac{1}{N_2}\right) P^{-1}, \text{ on the hypothesis } H_0.$$

Thus

$$V^{-1} = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} P, \text{ on the hypothesis } H_0. \tag{16}$$

Hence the quadratic form (12) with the help of (11) and (16) can be written on the hypothesis H_0 as

$$\left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1}\left[\mathop{\Sigma}_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right]' P^{-1}\left[\mathop{\Sigma}_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right] \quad (r=1,2,\ldots s) \tag{17}$$

and this behaves as χ^2 with s d.f., *i.e.*, $\chi^2_{(s)}$—our test criterion. The values of p_i's and their differential coefficients involved in (17) can be obtained by estimating θ_r's from the following equations obtained on the hypothesis H_0, namely,

$$\mathop{\Sigma}_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(x_{1i}+x_{2i}\right)=0 \qquad (r=1,2,\ldots s)$$

i.e.,

$$\mathop{\Sigma}_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(n_{1i}+n_{2i}\right)=0 \qquad \text{for all } r. \tag{18}$$

These s equations in (18) must be solved simultaneously to get the values of θ_r's and these must be substituted in the appropriate expressions of p_i's and their differential coefficients.

Thus the method is applicable in all cases where we can get the maximum likelihood estimates.

(*b*) The method of likelihood ratio also gives the same test criterion. It can be shown as follows:

As before

$$2L \doteq \text{const.} - \mathop{\Sigma\Sigma}_{j\ i} \frac{x_{ji}^2}{N_j p_{ji}} \tag{i}$$

To get the optimum estimates of θ_{jr} we solve independently the two sets of s equations each, namely,

$$\frac{\partial L}{\partial \theta_{1r}}=0, \qquad (r=1,2,\ldots s)$$

and

$$\frac{\partial L}{\partial \theta_{2r}}=0, \qquad (r=1,2,\ldots s)$$

i.e.,

$$\left.\begin{array}{ll} \mathop{\Sigma}_i \dfrac{1}{p_{ji}}\left(\dfrac{\partial p_{ji}}{\partial \theta_{jr}}\right)x_{ji}=0 & (j=1,2;\ r=1,2,\ldots s) \\ \text{also from definition} & \\ \mathop{\Sigma}_i x_{ji}=0 & (j=1,2) \end{array}\right\} \tag{ii}$$

Putting

$$z_{ji} = \frac{x_{ji}}{\sqrt{p_{ji}}} \qquad (j = 1, 2;\ i = 1, 2, \ldots k) \qquad \text{(iii)}$$

it is clear that z_{ji} $(j = 1, 2,;\ i = 1, 2, \ldots k)$ are independent normal variates with zero means and variances N_j $(j = 1, 2)$ for all i's.

With the help of (iii), (ii) can be written as follows:

$$\left.\begin{aligned} &\Sigma_i \frac{1}{\sqrt{p_{ji}}}\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right) z_{ji} = 0 \\ \text{and} \quad & \\ &\Sigma_i \sqrt{p_{ji}}\ z_{ji} = 0 \end{aligned}\right\} \qquad \text{(iv)}$$

and (i) becomes

$$2L \doteqdot \text{const} - \Sigma_j \Sigma_i \frac{{z_{ji}}^2}{N_j} \qquad \text{(v)}$$

Further, let

$$\left.\begin{aligned} X_{jr} &= \Sigma_i \frac{1}{\sqrt{p_{ji}}}\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right) z_{ji} \\ t_j &= \Sigma_i \sqrt{p_{ji}}\ z_{ji} \end{aligned}\right\} (j = 1, 2;\ r = 1, 2, \ldots s) \qquad \text{(vi)}$$

It is clear that X_{jr}'s $(r = 1, 2, \ldots s)$ are orthogonal to t_j $(j = 1, 2)$. The quadratic form for X_{jr} and t_j $(j = 1, 2;\ r = 1, 2, \ldots s)$ is

$$[t_j\ x_{j1}\ \ldots\ x_{js}]\ V_j^{-1} \begin{bmatrix} t_j \\ x_{j1} \\ \cdot \\ \cdot \\ \cdot \\ x_{js} \end{bmatrix} \qquad (j = 1, 2), \qquad \text{(vii)}$$

where V_j^{-1} is the reciprocal matrix of the variance matrix V_j of the variates involved. But

$$\text{Var}\ (x_{jr}) = N_j\ \Sigma_i \frac{1}{p_{ji}}\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right)^2$$

$$\text{Var}\ (t_j) = N_j$$

$$\text{Cov}\ (x_{jr}, t_j) = 0$$

$$\text{Cov}\ (x_{jm}\ x_{jr}) = N_j \Sigma_i \frac{1}{p_{ji}}\left(\frac{\partial p_{ji}}{\partial \theta_{jm}}\right)\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right)$$

$$(j = 1, 2;\ m, r = 1, 2, \ldots s;\ m \neq r)$$

Thus

$$V_j = N_j \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \sum\limits_i \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_{j1}}\right)^2 & \cdots & \sum \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_j}\right)\left(\frac{\partial p_{ji}}{\partial \theta_{js}}\right) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 0 & \sum\limits_i \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_{js}}\right)\left(\frac{\partial p_{ji}}{\partial \theta_{ji}}\right) & \cdots & \sum \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_{js}}\right)^2 \end{bmatrix}$$

Hence (vii) can be written as

$$\frac{t^2_j}{N_j} + \frac{1}{N_j} [x_{j1} \ldots x_{js}] \, P_j^{-1} \begin{bmatrix} x_{j1} \\ \cdot \\ \cdot \\ \cdot \\ x_{js} \end{bmatrix} \quad (j = 1, 2), \qquad \text{(viii)}$$

where P_j is the same as in § 3 (*a*) (8).

In view of the linear constraints on z_{ji}'s given by (iv) we can write (v) with the help of (viii) as

$$2L_{H_1} : \text{ const.} - \sum_{j=1}^{2} \sum_{i=1}^{k} \frac{z^2_{ji}}{N_j} + \sum_{j=1}^{2} \frac{t_j^2}{N_j} + \sum_{j=1}^{2} \frac{1}{N_j} [x_{jr}]' \, P_j^{-1} \, [x_{jr}]$$

$$(r = 1, 2, \ldots s) \qquad \text{(ix)}$$

As before on the hypothesis H_0 we have $\theta_{1r} = \theta_{2r} = \theta_r$ (say) for all r ($r = 1, 2, \ldots s$) and also $p_{1i} = p_{2i} = p_i$ (say) for all i ($i = 1, 2, \ldots k$) and hence $P_1 = P_2 = P$ (say), where P is the same as in § 3 (*a*) (11).

Also instead of $2s$ equations

$$\frac{\partial L}{\partial \theta_{1r}} = 0$$

and

$$\frac{\partial L}{\partial \theta_{2r}} = 0 \qquad (r = 1, 2, \ldots s)$$

we get only s equations

$$\frac{\partial L}{\partial \theta_r} = 0 \qquad (r = 1, 2, \ldots s) \qquad \text{(x)}$$

in addition to two equations.

$$_1 = 0 = t_2.$$

Equation (x) can be written as

$$X_{1r} + X_{2r} = 0 \qquad (r = 1, 2, \ldots s).$$

Now the quadratic form for t_1, t_2 and $(X_{1r} + X_{2r})$ $(r = 1, 2, \ldots s)$ can be written as follows:

$$\sum_{j=1}^{2} \frac{t_j^2}{N_j} + [x_{1r} + x_{2r})]' \, V^{-1} \, [x_{1r} + x_{2r}] \qquad (r = 1, 2, \ldots s) \qquad \text{(xi)}$$

where

V^{-1} = reciprocal of the variance matirx V of the variates $x_{1r} + x_{2r}$ for all r.

Also

$$V = N_1 P_1 + N_2 P_2 \qquad P\text{'s being the same as in § 5 } (a)\ (8)$$

$$= (N_1 + N_2)\, P \text{ on the hypothesis } H_0.$$

In view of (xi) we write (v) as follows:

$$2L_{H0} \doteqdot \text{const.} - \sum_{j=1}^{2} \sum_{i=1}^{k} \frac{z_{ji}^2}{N_j} + \sum_{j=1}^{2} \frac{t_j^2}{N_j} + \frac{1}{N_1 + N_2} [x_{1r} + x_{2r}]' \, P^{-1} \, [x_{1r} + x_{2r}] \qquad (r = 1, 2, \ldots s) \qquad \text{(xii)}$$

From (ix) and (xii) we have by subtraction

$$-2\,(L_{H0} - L_{H1}) \doteqdot \sum_{j=1}^{2} \frac{1}{N_j} [x_{jr}]' \, P^{-1} \, [x_{jr}] - \frac{1}{N_1 + N_2} [x_{1r} + x_{2r}]' \, P^{-1} \, [x_{1r} + x_{2r}] \quad (r = 1, 2, \ldots s).$$

i.e.,

$$-2\,(L_{H0} - L_{H1}) \doteqdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} \left[\frac{x_{1r}}{N_1} - \frac{x_{2r}}{N_2}\right]' P^{-1} \left[\frac{x_{1r}}{N_1} - \frac{x_{2r}}{N_2}\right] \qquad (r = 1, 2, \ldots s) \qquad \text{(xiii)}$$

Therefore

$$-2 \log (\text{likelihood ratio}) = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} Q, \qquad \text{(xiv)}$$

where

$$Q = \left[\sum_i \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \theta_r}\right) \left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right]' P^{-1} \left[\sum_i \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \theta_r}\right) \left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right] \qquad (r = 1, 2, \ldots s)$$

and this behaves as x^2 with s degrees of freedom and this result which is our test criterion agrees with the result of § 3 (a) (17) as was expected.

4. Karl Pearson's result of 1911 referred to in the Introduction follows as a particular case of our general result. Pearson's result follows if we put

$$\theta_i = p_i \qquad (i = 1, 2, \ldots k).$$

Thus

$$\Sigma_i \frac{1}{p_i} \frac{\partial p_i}{\partial \theta_i}\left(\frac{n_{1i}}{N_1} + \frac{n_{2i}}{N_2}\right) = \frac{1}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)$$

and in this case

$$P = \begin{pmatrix} \frac{1}{p_1} & 0 & 0 & . & . & . & 0 \\ 0 & \frac{1}{p_2} & 0 & . & . & . & 0 \\ 0 & 0 & \frac{1}{p_3} & . & . & . & 0 \\ . & . & . & & & & \\ . & . & . & & & & \\ . & . & . & & & & \\ 0 & 0 & 0 & . & . & & \frac{1}{p_s} \end{pmatrix}$$

a diagonal matrix. Therefore

$$P^{-1} = \begin{bmatrix} p_1 & 0 & 0 & . & . & . & 0 \\ 0 & p_2 & 0 & . & . & . & 0 \\ . & . & . & & & & \\ . & . & . & & & & \\ . & . & . & & & & \\ 0 & 0 & 0 & . & . & . & p_s \end{bmatrix}$$

Our test criterion thus reduces to

$$\frac{N_1 N_2}{N_1 + N_2} \Sigma_i \, p_i \left\{\frac{1}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right\}^2$$

i.e.,

$$= \frac{N_1 N_2}{N_1 + N_2} \Sigma \frac{1}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)^2 \qquad (1)$$

and since $\Sigma_i p_i = 1$ we have only $(k - 1)$ independent p_i's and thus (1) behaves as x^2 with $(k - 1)$ d.f. The optimum values of p_i's follow from the equations

$$\frac{\partial L}{\partial p_i} = 0 \qquad (i = 1, 2, \ldots k - 1), \qquad (2)$$

where

$$p_k = 1 - \sum_{i=1}^{k-1} p_i$$

i.e.,

$$\frac{x_{1i} + x_{2i}}{p_i} - \frac{x_{1k} + x_{2k}}{p_k} = 0 \qquad (i = 1, 2, \ldots k - 1)$$

i.e.,

$$p_k\,(n_{1i} + n_{2i}) = p_i\,(n_{1k} + n_{2k})\,, \qquad (i = 1, 2 \ldots k-1). \qquad (3)$$

Adding (3) for all i's we get

$$p_k = \frac{n_{1k} + n_{2k}}{N_1 + N_2}$$

Therefore from (3)

$$p_i = \frac{n_{1i} + n_{2i}}{N_1 + N_2}. \qquad (4)$$

thus the suitable values for p_i's are given by (4) for all i's including $i = k$.

These are the values suggested by K. Pearson and this result was confirmed by E. C. Rhods (1924) and J. Neyman and E. S. Pearson (1928) by other methods.

5. It may be noted that in the general case we can write

$$-2L_{H_0} + \text{const.} \doteqdot \left\{ \sum_i \frac{x_{1i}^2}{N_1 p_i} - \frac{t_1^2}{N_1} + \frac{1}{N_1}\,[x_{1r}]'\,P^{-1}\,[x_{1r}] \right\}$$
$$+ \left\{ \sum_i \frac{x_{2i}^2}{N_2 p_i} - \frac{t_2^2}{N_2} - \frac{1}{N_2}\,[x_{2r}]'\,P^{-1}\,[x_{2r}] \right\}$$
$$+ \chi_{[s]}^2, \qquad (1)$$

where

$$\chi_{[s]}^2 = \frac{1}{N_1}\,[x_{1r}]'\,P^{-1}\,[x_{1r}] + \frac{1}{N_2}\,[x_{2r}]'\,P^{-1}\,[x_{2r}] - \frac{1}{N_1 + N_2}$$
$$[x_{1r} + x_{2r}]'\,P^{-1}\,[x_{1r} + x_{2r}] \quad (r = 1, 2, \ldots s)$$

Thus it may be observed that the first expression within curled brackets on the right-hand side of (1) gives a measure of goodness of fit of the assumed law of distribution to the first sample and behaves as χ^2 with $(k - s - 1)$ d.f. Similarly the second expression on the right-hand side of (1) within curled bracket measures the goodness of fit of the assumed law to the second sample while $\chi_{[s]}^2$ is our test criterion which tests whether or not the two samples belong to the same population. Thus none of these χ^2 should be significant at assigned probability levels. If we want to have a criterion for the goodness of fit then $k - s - 1 \geqslant 0$, *i.e.*, $s \leqslant k - 1$ and this explains the restriction we put on the values of s in the beginning of § 3.

6. In the case of single parameter the test criterion (for large samples of course) behaves as χ^2 with 1 d.f. and the criterion is given by the expression

$$\frac{\left\{\sum_i \frac{p_i'}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right\}}{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)\sum_i\left(\frac{p_i'^2}{p_i}\right)} \tag{A}$$

where

$$p_i' = \frac{\partial p_i}{\partial \theta}$$

The above result when applied to the case of Binomial Law gives us the result

$$\frac{\left\{\sum_{r=0}^{k-1} r\left(\frac{n_{1r}}{N_1} - \frac{n_{2r}}{N_2}\right)\right\}^2}{(k-1)\,\check{p}\,\check{q}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}, \tag{B}$$

where $\check{p}$ is maximum likelihood estimate when H_0 is true and $\check{q} = 1 - \check{p}$, and $\check{p}$ is given by

$$\check{p} = \frac{\sum_{r=0}^{k-1} r \quad (n_{1r} + n_{2r})}{(N_1 + N_2)(k - 1)}$$

when $k = 2$, we get the ordinary contingency table in the case of Binomial Law and the expression (B) reduced to the familiar result

$$\frac{(n_{12}n_{21} - n_{11}n_{22})^2\,(n_{11} + n_{12} + n_{21} + n_{22})}{(n_{11} + n_{12})\,(n_{21} + n_{22})\,(n_{11} + n_{21})\,(n_{12} + n_{22})}$$

SUMMARY

A method to test the hypothesis that two independent samples classified into same k frequency classes belong to the same population involving s parameters has been developed in this paper. The criterion reduces to a χ^2 test with s d.f. The results given by Karl Pearson (1911) are shown to be special cases of this result.

I am indebted to Dr. D. N. Lawley of Edinburgh University for suggesting the problem and for his valuable criticism and advice in the preparation of this paper.

REFERENCES

1. Cramer, H. .. *Mathematical Methods of Statistics*, Princeton University Press, 1946, 425–35.

2. Dixon, W. J. .. "A critesion for testing the hypothesis that two samples are from the same population," *Ann. Math. Stats.*, 1940, **11**, 199–204.

3. Fisher, R. A. .. *Theory of Statistical Estimates*, 1923–25.

4. Neyman, J. and Pearson, E. S. "On the use and interpretation of certain criteria for the purpose of statistical inference", *Biom.*, 1928, **20 A**, 283–89.

5. Pearson, K. .. "On the probability that two independent distributions of frequency are really samples from the same population," *ibid.*, 1911, **8**, 250–52.

6. Wald, A. and Wolfowitz, J. "On a test whether two independent samples are from the same population," *Ann. Math. Stats.*, 1940, **11**, 149–62.

7. Wilks, S. S. "The large sample distribution of likelihood ratio for testing composite hypotheses," *ibid.*, 1938, **9**, 60–62.