"ON THE TEST OF A HYPOTHESIS CONCERNING TWO INDEPENDENT FREQUENCY DISTRIBUTIONS."

by

DIP NARAYAN LAL, M.A. (PATNA)

THESIS SUBMITTED FOR DEGREE OF Ph.D.

EDINBURGH UNIVERSITY,

1948.

CHAPTER I.

Introduction.

1.01. One of the chief uses of statistical analysis of given samples, which are of course assumed to be representative in the statistical sense, is to draw correct inferences regarding their populations.

Those cases in which we know a priori the exact distribution of their populations are evidently trivial and do not present any such statistical problem.

In some cases, however, knowing nothing about the nature of the populations, we may attempt to obtain a hypothetical population from which the two given samples may reasonably be assumed to have been drawn.

But quite often we may try to know whether or not the two given samples can arise from the same population when we know only the nature of the distribution of their populations without our knowing them exactly, for some of the parameters which specify them completely may be unknown; that is, when our hypothesis regarding the populations is a "composite" one. Thus, in any particular case we may know that the given samples belong to a normal population (say) without their means or variances or both - the two paramenters which completely specify a Normal Universe - being known.

1.02./

1.02. It is evident that whether the nature of the distribution of the populations is known or not, in order to get information about them, we shall have to estimate some unknown parameters or their functions which would specify them (populations) completely, and obtain a test criterion which will enable us to say whether or not the two given samples belong statistically to the same population.

Fundamentally, therefore, the problem is one of estimation of the unknown parameters (or their functions) of the populations from the given samples and, according to the established statistical practice, we shall assert that the two samples belong to the same population when the estimated values of the parameters of the population from the given samples do not differ significantly at pre-assigned levels of significance. These levels are in general determined by the amount of risks we are prepared to take.

1.03. The importance of this type of problem cannot be over-estimated. It may be used to study a variety of problems of great practical value, e.g. "the qualities and quantities of manufactured products, yield of agricultural techniques, results of different medical treatments, effects of suggested educational methods and/

and the like". Thus, to take a concrete case, we may have two samples of finished goods of same kind classified into the same different groups according to certain characteristics, which can be measured numerically; the question arises whether or not the two samples are from an identical source of production, i.e. whether or not the processes of manufacture of both the samples can be assumed to be identical.

1.04. In view of the importance of the above types of problems, we discuss here an equally useful and important problem of allied nature, namely :- *Given two independent sets of frequencies classified into the same K frequency classes. To develop a test of the hypothesis that the two samples may be said to belong to the same population, it being assumed that the samples are large and the law of distribution of the population is known except for certain unspecified parameters.

It is of course inherent in the above problem that even when the samples belong to different populations the nature of their distributions i.e. their mathematical form remains the same e.g. if the law of distribution is known to be Poissonian (say) we assume that both the samples come from Poissonian populations.

1.05./

1.05. The problem we have stated above appears at first to be one due to Karl Pearson in 1911 but the brief outlines of some of the problems of allied nature including K.Pearson's, given below, will clearly show the distinctive character of our problem. We shall further show (which is evident otherwise) that Pearson's result is a particular case of our general result.

1.06. K. Pearson's problem of 1911:-

In a memoir contributed to the Philosophical Magazine K. Pearson (1900) dealt with the problem of the probability that a given frequency distribution was a sample from a known population. That investigation was the basis of his treatment of "goodness of fit" of samples.

In Biometrika he (1911) stated a problem of somewhat different kinds but essentially as important in character.

The problem is as follows:-

"We have two samples and a priori they may be of the same population or of different populations, we desire to find out what is the probability that they are random samples of the same population. This population is one however of which we have no a priori experience".

Pearson proceeds as follows:-

Let the population from which the two

samples/

samples, if undifferentiated are supposed to be drawn be given by the class frequencies, $\mu_1, \mu_2 \ldots, \mu_p, \mu_q, \ldots \mu_s$, the total population being M.

Let the two samples be given by the frequencies in the same class as follows:-

									Total
1st.Sample:-	f_1	f_2	...	f_p	f_q	...	f_s	=	N
2nd.Sample:-	f_1'	f_2'	...	f_p'	f_q'	...	f_s'	=	N'

The totals N and N' may be equal or unequal. Then by detailed analysis which is equivalent to reducing the table to a single series of s cells he suggested that

$$\chi^2 = \sum_{r=1}^{s} \left\{ \frac{NN'\left(\frac{f_r}{N} - \frac{f_r'}{N'}\right)^2}{f_r + f_r'} \right\}$$

With (S-1) degrees of freedom should be our criterion to test whether or not the two given samples belong to the same population. He also suggested that the unknown p_r which measures the probability of an observation falling in the r^{th} class be replaced by $f_r + f_r' / N+N'$.

Later on he (1932) modifies this result and indicates the error in his assumption of the value of p_r cited above. He writes, "If we try to think over what the 'best hypothesis' means in this matter, ought we not to interpret them as signifying that hypothesis as to p_r's which will give the highest probability of

the/

the two samples being drawn from the same population? Surely if we are asking whether the two samples are likely to have been drawn from some unknown parent population we ought to choose for that unknown parent population the one that makes the probability P of their common sampledom a maximum or the value of χ^2 as small as possible".

Then he tries to find the value of p_r's by minimising

$$\chi^2 = v \sum_{r=1}^{s} \left\{ \left(\frac{f_r}{N} - \frac{f_r'}{N'} \right)^2 \frac{1}{p_r} \right\}$$

where $\sum_{r=1}^{s} p_r = 1$, and $v = \frac{NN'}{N+N'}$

and obtains

$$p_r = \frac{\frac{f_r}{N} \sim \frac{f_r'}{N'}}{\sum_{r=1}^{s} \left(\frac{f_r}{N} \sim \frac{f_r'}{N'} \right)}$$

and $$\chi^2_{min} = v \left\{ \sum_{r=1}^{s} \left(\frac{f_r}{N} \sim \frac{f_r'}{N'} \right) \right\}^2$$

Thus this new solution contradicts his results of 1911 which was also obtained by E.C.Rhodes in 1924 and J.Neyman and E.S. Pearson (jointly) in 1928 by different methods.

The incompatibility of the two sets of solutions given by K.Pearson can be explained if we note that the result of 1932 is based on/

on minimising χ^2 his suggested test criterion only. In order to get the best values of p_r's we must minimise the whole χ^2 and not only its component which is a measure of the test criterion of the hypothesis in question, and this gives the same value as suggested by him in 1911. His modification of his previous result, therefore, does not seem to meet the requirements of the problem in view.

1.07. E.C.Rhodes (1924) discusses the same problem and by his "doublet method" arrives at the same result as obtained by K.Pearson in 1911.

In this method he considers a multivariate normally correlated surface in (2s) variates for each of his two samples consist of the same $(s+1)$ categories and he finds the chance of obtaining the required samples as the volume of the surface outside the limiting contour. This volume measures the chance that in two samples of totals of n and m observations respectively, classified into the same $(s+1)$ categories we should get the two observed samples or other distributions occurring together which are less likely.

After certain transformation the integral measure/

measure of the volume reduces to

$$\frac{\int_{\chi}^{\infty} e^{-\frac{1}{2}\chi^2}\chi^{2s-1}d\chi}{\int_{0}^{\infty} e^{-\frac{1}{2}\chi^2}\chi^{2s-1}d\chi}$$

where the limiting contour reduces to

$$\chi^2 = \sum_{r=1}^{s+1}\left\{\frac{(f_r - np_r)^2}{np_r} + \frac{(f_r' - mp_r)^2}{mp_r}\right\}$$

On the basis that the two samples are from the same population.

Thus he too obtains χ^2 with s degrees of freedom as the test criterion and

$$p_r = \frac{f_r + f_r'}{n+m}, \ (r = 1, 2, \cdots \overline{s+1})$$

as the values of p_r's which minimise the total χ^2.

1.08. J.Neyman and E.S.Pearson in their joint paper (1928) take up the above problem of K. Pearson as an illustration of the applicability of the method of likelihood im drawing statistical inferences.

They assumed two samples classified into the same K Categories with n_{1r} and n_{2r} observations in the r^{th} category of the first and the second sample respectively. Then after some analysis they arrive at the same test criterion

and/the

and the same values of μ_r''s except for notational differences.

1.09. In the problem discussed above the population is assumed to be a discrete one but a similar problem has been discussed when the population is supposed to be continuous.

William R.Thompson (1938) in the second part of his paper discusses a criterion for testing whether or not the two samples belong to the same population when nothing is known about the distribution functions of the variates except that they are continuous.

A. Wald and J. Wolfowitz (1940) in their joint paper discuss the problem of W.R.Thompson cited above. They improve the result of Thompson and suggest a test criterion which not only enables us to discriminate whether or not both the samples belong to the same population but is also consistent in the sense that the probability of rejecting the null hypothesis if it is false approaches unity as the number of observations in the sample tends to infinity.

W.J.Dixon (1940) discusses the same problem. The assumption about the distribution functions of the variates involved is the same, namely that they are continuous. He obtains a different criterion/

criterion and indicates its relation to the criterion of Wald and Wolfowitz.

1.10. In all these solutions stated above it has been assumed that the nature of the distribution is either not known or we are not interested in it. But when the nature of the distribution is known the above solutions for discrete and continuous cases will not meet the requirements of the problem.

In practice generally the discrete case is more important. In the solution of K. Pearson an attempt is made to classify the population by means of class categories and thus the p_r's $(r = 1, 2, \ldots, s)$ may be regarded as the s parameters which would specify a population. Of these s parameters only (s - 1) are independent as $\sum_{r=1}^{s} p_r = 1$ The significant difference between the two estimated values of p_r's $(r = 1, 2, \ldots, s)$ from the two samples will mean that the two samples do not belong to the same population. Hence when no other relevant information is available this method may be looked upon as a good approximation but quite often from _a priori_ or other considerations we may be able to specify the nature of the distribution i.e. the/

the mathematical form of the population and in such cases it will not be correct to say that the insignificant difference between the estimated values of p_r's $(r = 1, 2, \ldots, s)$ from the two samples means that the samples come from the same population.

For let the first sample be from the population whose distribution law is

$$f(x;\ \theta_{11}, \theta_{12}, \cdots \theta_{1m}) \quad \text{(say)}$$

and that the second is from the population with distribution law as

$$f(x;\ \theta_{21}, \theta_{22}, \cdots \theta_{2m}) \quad \text{(say)}$$

then we have

$$p_{jr} = \int_{l_r}^{l_{r+1}} f(x;\ \theta_{j1}, \theta_{j2}, \cdots \theta_{jm})\, dx \quad (j = 1, 2;\ r = 1, 2, \cdots s)$$

l's being the limits of different groups and p_{jr} denotes the probability of an observation of the j^{th} sample falling in the r^{th} group.

It is thus clear that in this case it is not correct to assert that the two populations are identical only if $p_{1r} = p_{2r}$ for all r's. The populations can be said to be identical only when the respective estimated θ's are equal or their differences are insignificant.

CHAPTER II.

2.01. Some digression on the χ^2 distribution :-

The χ^2 distribution plays an important role in the theory of statistical inferences, as a large number of test criteria are finally made to depend on it. We shall therefore refer to some of its salient features and its relation to "Likelihood".

Prior to 1922 a good deal of confusion prevailed in the application of χ^2 test criterion on the goodness of fit. R.A. Fisher (1922) showed that for a correct application of the formula of χ^2 as a measure of goodness of fit, the term n' which occurs in it in the general case must not be confused with n', the number of cells of the table but that it must be taken as one more than the degrees of freedom of the table. Thus if we have a contingency table with r rows and c columns, then he suggested that our n' must be equal to $(r - 1)\ (c - 1) + 1$ and not equal to rc, the number of cells in the given table.

He also showed that the χ^2 test criterion suggested by Pearson and stated above (§1.06) could be easily obtained by treating the table as a contingency table having (s - 1) degrees of freedom.

This/

This modification in n' solved much of the confusion then prevalent and it also solved the so called failure of χ^2 test on the data of Yule and Greenwood given in their joint paper in 1915.

About the same time Fisher (1922) also put forward his technique of maximal likelihood in a form which can be readily utilised in statistical analysis. Notwithstanding its own limitations the method provides more powerful tools for statistical analysis than the method of moments.

He showed there that

$$L = \frac{1}{2}\chi^2$$

as a first approximation under certain conditions, namely, when the deviation of an observed frequency in any class from its expected value is small compared to its expected value and L differs by a constant from the logarithm of the likelihood with sign changed.

"In these cases, therefore, where χ^2 is a valid measure of departure of the sample from expectation it is equal to 2L, in all other cases the approximation fails and L itself must be used". It also follows from this that the maximum likelihood estimates of the parameters minimise/

minimise χ^2. Thus the method of maximum likelihood not only covers cases where χ^2 is applicable but also those where χ^2 is not applicable.

Fisher (1923-25) further showed that the method of optimum likelihood yields the estimates of parameters which are consistent, efficient and also sufficient if sufficient solutions exist.

Quite often the expression of χ^2 contains the unknown parameters of the law of distribution or their functions and the result remains indeterminate unless these unknowns can be replaced by their estimated values in terms of the observations at our disposal. Fisher (1924) showed that if we take an efficient estimate of a parameter and substitute it in the calculated expression of χ^2 we get back a χ^2 distribution but with a loss of one degree of freedom. Thus if the expression of χ^2 contains many parameters and if we substitute in it their efficient estimates from the data we shall get back χ^2 distribution with loss of degrees of freedom equal to the number of parameters estimated.

This result is very fundamental indeed, especially in problems of significance tests.

He has also shown in the same paper that

if/

if inefficient estimates of parameters are used in the expression of χ^2 then in general the new calculated value of χ^2 does not follow the usual χ^2 distribution and hence will not enable us to draw readily any useful statistical inferences.

This result thus brings into great relief the utility of the method of maximum likelihood in providing estimates of parameters which satisfy the criterion of efficiency.

2.02. The Use of "Likelihood" method in the field of Statistical inference:-

The method of testing statistical hypotheses was first developed by J.Neyman and E.S.Pearson (1928,1932-33).

The method essentially consists in selecting a rule of rejecting the hypothesis in question whenever the sample point E lies within a certain region ω (say), called "critical region" of the sample space W. The sample point E is assumed to be denoted in the n - dimensional sample space W by its n co-ordinates the data of observations $x_1, x_2, \ldots, x_n$.

The probability $P\{E \varepsilon \omega / H_0\} = \alpha$

of rejecting the hypothesis H_0 when it is true is called the size of the corresponding critical region/

region ω on which the test is based. The expression $P\{E \,\varepsilon\, \omega | H_0\}$ means the probability of the point E lying in the region ω on the hypothesis H_0.

It is clear that in all statistical inferences two types of errors are involved.

(i) The error of the first kind i.e. rejecting the hypothesis when it is true.

(ii) The error of the second kind which consists in accepting the hypothesis when it is false.

The error of the first kind can always be controlled and the error of the second kind must be minimum i.e. our failure to reject the hypothesis H_0 when an alternative hypothesis H is true should be minimum.

The probability $P\{E \,\varepsilon\, \omega | H\}$ of rejecting the hypothesis H_0 when an alternative H is true is called the power of the test with respect to H and it must be maximum if we want to minimise the error of the second kind. It has been shown that in general $p_0 = Kp$ on the boundary of the region of rejection (critical region) where $p_0 = P\{E \,\varepsilon\, \omega | H_0\}$ and $p = P\{E \,\varepsilon\, \omega | H\}$ and K is a positive constant.

This region within the boundary of ω controls error of first kind and minimises that of/

of the second.

It has been further proved that when the maximum likelihood estimate of the parameter is substituted in the relation $p_0 = k p$ we get the best critical regions, if they exist, for testing hypotheses.

In all these above discussions we assume that there is only one parameter in the law of distribution i.e. the hypothesis is a "composite" one with "one degree of freedom".

Further by applying the method of maximum likelihood Neyman and Pearson have suggested a method for obtaining functions of observations for testing what are called composite statistical hypotheses with various degrees of freedom.

The procedure is as follows:-
A population K is assumed in which a variate x (x may be a vector with each component representing variate) has a distribution function $f(x; \theta_1, \theta_2, \cdots, \theta_s)$ which depends on s parameters θ's.

A simple hypothesis is one in which the θ's are specified. Considering x's as fixed we examine the variation in p_0 according to the unspecified parameters $\theta_1, \theta_2, \cdots, \theta_r$ (say) which form a set ω(say). Let $p_0(\omega \max)$ be the maximum value of p_0 for such variations.
Similarly/

Similarly if Ω is the class of admissible alternatives, H, let $p(\Omega \max)$ be the maximum value of the likelihood for variations in all the values of θ_r's, $(r=1,2,\cdots s)$.

Let
$$\lambda = \frac{p_0(\omega_{max})}{p(\Omega_{max})}$$

then this λ is called the likelihood ratio and it obviously lies between o and 1.

A possible criterion for accepting H_0 is to take the critical region as consisting of those points for which $\lambda \leq C$ (say) a constant and C is determined by relation to a probability level α from the sampling distribution of λ which evidently is independent of the unknown parameters.

It also follows from above that if all the parameters are specified by a certain hypothesis and the alternative hypothesis is that they (parameters) are different then $p_0(\omega \max) = p_0$ on the first hypothesis and $p(\Omega \max)$ will be the maximum value of the likelihood obtained by maximising it for variations in all parameters involved.

This method is of great value and this further shows the importance of the method of likelihood in the field of statistical inference.

It is quite interesting to visualise

"composite"/

"composite" and "simple" hypothesis geometrically.

It is evident that Ω may be represented as a region in S dimensional space of the θ's in particular it may be the whole of S dimensional space. For any subset w of Ω we denote by H_w the hypothesis that the parameter point lies in w. If w consists of a single point H_w is called the simple hypothesis otherwise H_w is a "composite" hypothesis.

(For further details reference may be made to the paper of A.Wald (1943).).

Incidentally it may be mentioned that the problem of testing hypothesis or of estimation are particular cases of the more general problem of statistical inference which, following A. Wald (1942), may be put as follows:-

Let S be a system of sub-classes of the class Ω of distribution functions. For each element $\mathcal{s}$ of S consider the hypothesis $H_{\mathcal{s}}$ which states that the unknown distribution F is an element of $\mathcal{s}$; denote by H_s the system of all such hypotheses, the problem is/

is to decide by means of a sample which element of H_s should be accepted. If S consists of two exclusive sub-classes only which together cover the whole Ω then the general problem becomes a problem of testing hypothesis. If S consists of all elements of Ω then the problem reduces to estimation of parameters involved for the hypothesis corresponding to each element of Ω means specifying a set of values of the parameters and that thus is the problem of estimation.

On the other hand if S consists of three sub-classes then the problem reduces to the "trilemma" of Wald and is a problem distinct from estimation or testing the hypothesis. Such a situation may arise in practice in the case of a manufacturer who wants to keep the quality of his product between two limits and wants to test it by sampling. It is of course assumed that the/

the quality in question is measurable and can be represented by a real number.

CHAPTER III.

3.01 In this Chapter we shall indicate the possibility of application of a method due to S.S.Wilks to discuss our problem as enunciated in § 1.04 namely:-

Given two independent samples classified according to the same K categories and the nature of the distribution law, to obtain the criterion for testing whether the samples may be assumed to belong to the same population.

S.S. Wilks (1938) discusses the large sample distribution of the likelihood ratio for testing composite hypotheses and proves that if a population with a variate x is distributed according to the probability function $f(x; \theta_1, \theta_2, \ldots \theta_h)$ such that the optimum estimates $\bar{\theta}_i$ of θ_i, $(i = 1,2, \ldots h)$ exist which are distributed in large samples according to multivariate normally correlated surface then when the hypothesis H is true that $\theta_i = \theta_{oi}$, $(i = m+1, m+2, \ldots h)$, the distribution of $-2 \log \lambda$ is except for terms of order $\frac{1}{\sqrt{n}}$ distributed like χ^2 with $(h - m)$ degrees of freedom; n being the number of observations in the sample.

Here $\lambda = \dfrac{P_\omega(O_n)}{P_\Omega(O_n)}$, where $P_\omega(O_n)$

is/

is the maximum value of the probability of obtaining the sample point O_n on the basis of given hypothesis and $P_\Omega(O_n)$ is its maximum value when there is no restrictions on the values of θ's. When the maximum values do not exist we replace them by their least upper bounds.

3·02 Let our two independent samples be as follows:-

Samples	Limits of class frequencies						Total
	$l_1 - l_2$	$l_2 - l_3$	. . .	$l_i - l_{i+1}$	. . .	$l_k - l_{k+1}$	
First Sample	n_{11}	n_{12}	. . .	n_{1i}	. . .	n_{1k}	N_1
Sec.Sample	n_{21}	n_{22}	. . .	n_{2i}	. . .	n_{2k}	N_2
Total	n_1	n_2	. . .	n_i	. . .	n_k	N

From the table it is evident that

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

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

and
$$\sum_{i=1}^{k} n_i = N = N_1 + N_2$$

Let $\phi(x; \theta_{11}, \theta_{12}, \cdots \theta_{1s})$ and $\phi(x; \theta_{21}, \theta_{22}, \cdots \theta_{2s})$

be the laws of distribution and p_{1i} and p_{2i} be the probabilities of an observation falling in the

ith/

i^{th} category of the first and the second sample respectively.

Thus
$$p_{1i} = \int_{l_i}^{l_{i+1}} \phi(x; \theta_{11}, \theta_{12}, \cdots \theta_{1s})\, dx$$
$$= f(l_{i+1}, l_i; \theta_{11}, \theta_{12}, \cdots \theta_{1s}) \quad \text{(say)}$$
$$= f_{1i} \text{ (say)}$$

and similarly $p_{2i} = f_{2i}$ (say), for all i's.

Then P the probability of obtaining the samples is given by

$$P = \frac{N_1!\; N_2!}{\prod_{i=1}^{k}\{(n_{1i}!)(n_{2i}!)\}} \prod_{i=1}^{k}\left\{p_{1i}^{n_{1i}}\, p_{2i}^{n_{2i}}\right\} \cdots\cdots (1)$$

Hence
$$L \equiv \log P = \text{const} + \sum_{i=1}^{k}\left(n_{1i}\log p_{1i} + n_{2i}\log p_{2i}\right) \cdots\cdots (2)$$

To estimate the parameters of the two samples we have the following set of equations,

$$\left.\begin{aligned} &\frac{\partial L}{\partial \theta_{1j}} = 0, \; (j = 1, 2, \cdots s) \\ \text{and} \quad &\frac{\partial L}{\partial \theta_{2j}} = 0, \; (j = 1, 2, \cdots s) \end{aligned}\right\} \cdots\cdots (3)$$

The simultaneous solutions of each set of s equations in (3) will give the optimum values $\hat{\theta}_{1j}$ and $\hat{\theta}_{2j}$, $(j = 1, 2, \ldots s)$ respectively.

Assuming/

Assuming that $\log p_{1i}$ and $\log p_{2i}$, $(i = 1,2, \ldots k)$ can be expanded by Taylor's theorem, we have

$$\log p_{1i} = \text{Const} + \sum_{j=1}^{s}(\theta_{1j}-\hat{\theta}_{1j})\left(\frac{\partial \log f_{1i}}{\partial \theta_{1j}}\right)_{\hat{\theta}_{1j}}$$

$$+\frac{1}{2}\sum_{j=1}^{s}\sum_{h=1}^{s}(\theta_{1j}-\hat{\theta}_{1j})(\theta_{1h}-\hat{\theta}_{1h})\left(\frac{\partial^2 \log f_{1i}}{\partial \theta_{1j}\,\partial \theta_{1h}}\right)_{\hat{\theta}_{1j},\hat{\theta}_{1h}}$$

(approximately)
(since $p_{1i} = f_{1i}$)

Similarly

$$\log p_{2i} = \text{Const} + \sum_{j=1}^{s}(\theta_{2j}-\hat{\theta}_{2j})\left(\frac{\partial \log f_{2i}}{\partial \theta_{2j}}\right)_{\hat{\theta}_{2j}}$$

$$+\frac{1}{2}\sum_{j=1}^{s}\sum_{h=1}^{s}(\theta_{2j}-\hat{\theta}_{2j})(\theta_{2h}-\hat{\theta}_{2h})\left(\frac{\partial^2 \log f_{2i}}{\partial \theta_{2j}\,\partial \theta_{2h}}\right)_{\hat{\theta}_{2j},\hat{\theta}_{2h}}$$

(approximately)
(since $p_{2i} = f_{2i}$)

Here the expression $\left(\frac{\partial \log f_{1i}}{\partial \theta_{1j}}\right)_{\hat{\theta}_{1j}}$ means:

we differentiate $\log f_{1i}$ with respect to θ_{1j} and then replace all the parameters by their optimum estimates. A similar meaning holds for other such expressions.

Therefore

$$L \doteq \text{Const} - \frac{1}{2}\left\{\sum_{j=1}^{s}\sum_{h=1}^{s}\left(c_{1jh} z_{1j} z_{1h} + c_{2jh} z_{2j} z_{2h}\right)\right\}$$

? taking Expectation

(approximately) --- (4)

where/

where

$$C_{1jh} = -E\left(\frac{\partial^2 \log f_{1i}}{\partial\theta_{1j}\,\partial\theta_{1h}}\right),\quad C_{2jh} = -E\left(\frac{\partial^2 \log f_{2i}}{\partial\theta_{2j}\,\partial\theta_{2h}}\right)$$

E denoting mathematical expectation, and

$Z_{1j} = (\hat{\theta}_{1j} - \theta_{1j})\sqrt{N_1}$

and $Z_{2j} = (\hat{\theta}_{2j} - \theta_{2j})\sqrt{N_2}$, $(j, h = 1, 2, \ldots s)$.

Thus

$$P \doteq \text{Const}\; e^{-\frac{1}{2}\left\{\sum_j\sum_h \left(C_{1jh} Z_{1j} Z_{1h} + C_{2jh} Z_{2j} Z_{2h}\right)\right\}}$$

The value of the constant can be readily seen to be equal to

$$\frac{|C_{1jh}|^{\frac{1}{2}}\,|C_{2jh}|^{\frac{1}{2}}}{(2\pi)^s}$$

Hence

$$P \doteq \frac{|C_{1jh}|^{\frac{1}{2}}\,|C_{2jh}|^{\frac{1}{2}}}{(2\pi)^s}\; e^{-\frac{1}{2}\left\{\sum_j\sum_h \left(C_{1jh} Z_{1j} Z_{1h} + C_{2jh} Z_{2j} Z_{2h}\right)\right\}} \quad \text{-----}(5)$$

It may be noted that $[C_{1jh}]$ and $[C_{2jh}]$ are both positive definite matrices.

Thus we assume the existence of functions $\hat{\theta}_{1j}\,(n_{11}, n_{12}, \cdots, n_{1k})$ and $\hat{\theta}_{2j}\,(n_{21}, \cdots, n_{2k})$, $(j = 1, 2, \ldots s)$ such that their joint distribution is given by (5). For detailed conditions under which $\hat{\theta}$'s exist which are distributed according to (5) reference may be made to a paper by J.L. Doob (1934).

Now/

Now

$$P_{\Omega}(O_N) = \frac{|c_{1jh}|^{\frac{1}{2}} |c_{2jh}|^{\frac{1}{2}}}{(2\pi)^s} \left\{1 + O\left(\frac{1}{\sqrt{N}}\right)\right\} \cdots \quad (6)$$

where it is assumed that N_1 and N_2 both are of the same order as N and the values of θ_{1j} and θ_{2j} which maximise (5) are obtained independently from the two sets of equations namely:-

$$-\frac{1}{2}\left\{\frac{1}{|c_{1jh}|} \frac{\partial |c_{1jh}|}{\partial \theta_{1j}} - \sum_j \sum_h \frac{\partial c_{1jh}}{\partial \theta_{1j}} z_{1j} z_{1h} + \sqrt{N_1} \sum_h c_{1jh} z_{1h}\right\} P = 0$$

and

$$-\frac{1}{2}\left\{\frac{1}{|c_{2jh}|} \frac{\partial |c_{2jh}|}{\partial \theta_{2j}} - \sum_j \sum_h \frac{\partial c_{2jh}}{\partial \theta_{2j}} z_{2j} z_{2h} + \sqrt{N_2} \sum_h c_{2jh} z_{2h}\right\} P = 0$$

respectively, $(j = 1,2, \ldots s)$ .... (A)

Since $c_{1jh} = O(1)$ and $c_{2jh} = O(1)$ and since $|c_{1jh}| \neq 0$ and $|c_{2jh}| \neq 0$, we find from (A) that the values of θ_{1j} and θ_{2j}, $(j = 1,2, \ldots s)$ which maximise P differ from $\hat{\theta}_{1j}$ and $\hat{\theta}_{2j}$, $(j = 1,2, \ldots s)$ respectively by terms of order $\frac{1}{\sqrt{N}}$.

(N_1, N_2 being of the same order as N).

Again under the hypothesis that the two populations are identical i.e. when $\theta_{1j} = \theta_{2j} = \theta_j$ (say), $(j = 1,2, \ldots s)$

We/

We get

$$C_{1jh} = C_{2jh}, \quad (j,h = 1,2, \ldots s)$$

and

$$P_\omega = \frac{|C_{1jh}|}{(2\pi)^s} e^{-\frac{1}{2}\sum_j \sum_h C_{1jh}(Z_{1j}Z_{1h} + Z_{2j}Z_{2h})} \quad \ldots.(7)$$

Now we must find θ_j's which maximise (7)

For this we have

$$\frac{\partial P_\omega}{\partial \theta_j} = 0, \ (j = 1, 2, \cdots s)$$

This gives

$$\sum_h C_{1jh}\left(\sqrt{N_1}\, Z_{1h} + \sqrt{N_2}\, Z_{2h}\right) = 0, \ (j = 1, 2, \cdots s) \quad \text{---- (B)}$$

neglecting terms of the type $\dfrac{\partial C_{1jh}}{\partial \theta_j}$.

Hence

$$P_\omega(O_N) = \frac{|C_{1jh}|}{(2\pi)^s} e^{-\frac{1}{2}\sum_j \sum_h C_{1jh}(Z_{1j}Z_{1h} + Z_{2j}Z_{2h})}$$

where Z's are subject to restrictions (B)

Then

$$\lambda = \frac{P_\omega(O_N)}{P_\Omega(O_N)} = K e^{-\frac{1}{2}\sum_j \sum_h C_{1jh}(Z_{1j}Z_{1h} + Z_{2j}Z_{2h})}$$

where $K = |C_{1jh}|^{\frac{1}{2}} \Big/ |C_{2jk}|^{\frac{1}{2}}$ and when N is large $K = 1 + O\left(\frac{1}{\sqrt{N}}\right)$.

Therefore/

Therefore

$$-2\log\lambda = \sum_j \sum_h c_{1jh}\left(z_{1j} z_{1h} + z_{2j} z_{2h}\right) \cdots \quad (9)$$

But the expression on the right of (9) is a quadratic form in 2s variates subject to s linear constraints given by (B); hence it can be expressed as χ^2 with s degrees of freedom. The θ_j's ($j = 1,2, \ldots s$) can be calculated in terms of known quantities from (B). Thus theoretically, $-2\log\lambda$ will behave as χ^2 with s d.f. (degrees of freedom) and will serve as our criterion to test whether or not the two samples belong to the same population. In practice this method seems to be difficult, and in the following pages alternative methods of approach are discussed. They appear to be simple and easy to apply within their own limitations.

CHAPTER IV.

4.01. In this Chapter we discuss our problem as stated in the preceding chapter. We start with the case when the law of distribution of the populations contains only one parameter and then extend it to the case of several parameters.

4.02. Case of one parameter:-

Let our two independent samples classified into the same k different categories be the ones assumed in § 3.02.

To simplify writing let us put

n_{ji} = Number of observations of the j^{th} sample belonging to the i^{th} class frequency

N_j = total number of observations in the j^{th} sample

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

$\phi(x;\theta_j)$ = the assumed law of distribution of the j^{th} sample, θ_j being the unknown parameter.

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

Thus we have

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

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

$$N = \sum_{j=1}^{2} N_j = \sum_{j=1}^{2} \sum_{i=1}^{k} n_{ji} = \sum_{i=1}^{k} n_i$$

and $$p_{ji} = \int_{l_i}^{l_{i+1}} \phi(x; \theta_j)\, dx$$

$$= p_{ji}(\theta_j) \quad \text{(say)} \quad (j = 1,2 \quad i = 1,2, \ldots k) \qquad \ldots\ldots\ldots\ldots(1)$$

Hence $p_{1i} = p_{1i}(\theta_1)$, $p_{2i} = p_{2i}(\theta_2)$ $(i = 1,2, \ldots k)$

It is clear that

$$\sum_{i=1}^{k} p_{ji} = 1, \quad (j = 1, 2).$$

The probability P of drawing the two samples is given by

$$P = \prod_{j=1}^{2} \left\{ \frac{(N_j)!}{\prod_{i=1}^{k} (n_{ji}!)} \prod_{i=1}^{k} p_{ji}^{n_{ji}} \right\} \quad \ldots \quad (2)$$

If N_j and n_{ji} $(j = 1,2 \quad i = 1,2, \ldots k)$ be both large i.e. when our samples are large we make use of Stirling approximation to factorials namely

$$n! \simeq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Making use of this approximation in (2) we get

$$P \doteqdot \prod_{j=1}^{2} \left\{ \frac{N_j^{N_j + \frac{1}{2}} e^{-N_j} \sqrt{2\pi}}{\prod_{i=1}^{k} \left(n_{ji}^{n_{ji} + \frac{1}{2}} e^{-n_{ji}} \sqrt{2\pi} \right)} \prod_{i=1}^{k} p_{ji}^{n_{ji}} \right\}$$

i.e.

$$P \doteq \text{Const} \prod_{j=1}^{2} \left\{ \prod_{i=1}^{k} \left(\frac{N_j p_{ji}}{n_{ji}} \right)^{n_{ji}+\frac{1}{2}} \right\} \cdots (3)$$

Let

$$N_j \, p_{ji} = \xi_{ji}$$

and

$$\eta_{ji} = \frac{n_{ji} - N_j p_{ji}}{\sqrt{N_j p_{ji}}} = \frac{n_{ji} - \xi_{ji}}{\sqrt{\xi_{ji}}} \cdots (4)$$

$$(j = 1, 2;\ i = 1, 2, \cdots k).$$

From (3) with the help of (4) we get

log P - constant

$$\doteq \sum_{j=1}^{2} \sum_{i=1}^{k} (n_{ji} + \tfrac{1}{2}) \log \left(\frac{\xi_{ji}}{n_{ji}} \right)$$

$$= - \sum_j \sum_i \left\{ \left(\xi_{ji} + \eta_{ji} \sqrt{\xi_{ji}} + \tfrac{1}{2} \right) \log \left(1 + \frac{\eta_{ji}}{\sqrt{\xi_{ji}}} \right) \right\}$$

If ξ_{ji} be large then η_{ji} will be small compared with this for different values of i and j: and so expanding the logarithm in the above expression we get

log P - const.

$$\doteq - \sum_j \sum_i \left(\xi_{ji} + \eta_{ji} \sqrt{\xi_{ji}} + \tfrac{1}{2} \right) \left(\frac{\eta_{ji}}{\sqrt{\xi_{ji}}} - \frac{1}{2} \frac{\eta_{ji}^2}{\xi_{ji}} \right) \cdots (\text{approx.})$$

$$= - \sum_j \sum_i \left\{ \tfrac{1}{2} \eta_{ji}^2 + \eta_{ji} \sqrt{\xi_{ji}} + O\left(\xi_{ji}^{-\frac{1}{2}} \right) \right\} \cdots\cdots (5)$$

But/

But $$\sum_{i=1}^{k} \eta_{ji}\sqrt{\xi_{ji}} = \sum_{i}(n_{ji} - N_j p_{ji}) = 0, (j=1,2)$$

Hence the expression (5) becomes

$$\log P - \text{Const} \doteq -\tfrac{1}{2}\sum_{j}\sum_{i} \eta_{ji}^2, \text{ to the order of } \xi_{ji}^{-\frac{1}{2}}, (j=1,2)$$

It is of course assumed that the order of N_j $(j = 1,2)$ is the same.

Thus to this approximation we have

$$P \doteq \text{Const}.\, e^{-\frac{1}{2}\sum_j \sum_i \eta_{ji}^2} \quad \ldots. (6)$$

Therefore from (6) it follows that correct to the order $\frac{1}{\sqrt{N_j}}$ $(j = 1,2)$, $\eta_{ji}, (j=1,2,\ i=1,2,\ldots k)$ are independent normal variates each with zero mean and unit variance. (subject to a linear constraint)

Let $x_{ji} = n_{ji} - N_j p_{ji}, (j = 1,2,\ i = 1,2,\ \ldots\ k)$ $\quad \ldots (7)$

then it follows from the above that the x_{ji}'s are independent normal variates with zero means and variances $N_j p_{ji}$'s

i.e. $\frac{x_{ji}^2}{N_j p_{ji}}$, $(j = 1,2,\ i = 1,2,\ \ldots\ k)$ are the ~~sum of~~ squares of independent standard normal variates.

Thus finally we have

$$P \doteq \text{Const}\quad e^{-\frac{1}{2}\sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}}}$$

Therefore/

Therefore

$$2L \equiv 2\log P \doteq Const - \sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}} \quad \ldots\ldots \quad (8)$$

Now $\frac{\partial L}{\partial \theta_j} = 0$ (j = 1,2) give the optimum solutions of θ_j ($j = 1,2$) respectively i.e. $\frac{\partial L}{\partial \theta_1} = 0$ gives the solution of θ_1 and $\frac{\partial L}{\partial \theta_2} = 0$ give the solution of θ_2. Let these equations when solved independently give $\check{\theta}_1$ and $\check{\theta}_2$ as the values of θ_1 and θ_2 respectively.

Then

$$\left(\frac{\partial L}{\partial \theta_j}\right)_{\check{\theta}_j} = 0 \quad (j = 1,2) \quad \ldots\ldots \quad (9)$$

But by Taylor's theorem

$$\left(\frac{\partial L}{\partial \theta_j}\right)_{\check{\theta}_j} \doteq \left(\frac{\partial L}{\partial \theta_j}\right) + (\check{\theta}_j - \theta_j)\frac{\partial^2 L}{\partial \theta_j^2}$$

approximately (j = 1,2) .... (10).

From (10) with the help of (9) we get

$$\check{\theta}_j - \theta_j \doteq - \frac{\partial L}{\partial \theta_j} \Big/ \frac{\partial^2 L}{\partial \theta_j^2} \quad (j = 1,2) \quad ..(11)$$

But from (8) neglecting terms proportional to $O\left(\frac{1}{\sqrt{N_j}}\right)$, ($j = 1,2$) we get

$$\left.\begin{aligned} \frac{\partial L}{\partial \theta_j} &\doteq \sum_i \left(\frac{p'_{ji}}{p_{ji}} x_{ji}\right) \\ \frac{\partial^2 L}{\partial \theta_j^2} &\doteq -N_j \sum_i \frac{p'^2_{ji}}{p_{ji}} \end{aligned}\right\} \quad (j = 1,2) \quad \ldots.(12)$$

where/

where $p'_{ji} = \frac{dp_{ji}}{d\theta_j}$

Substituting (12) in (11) we have

$$\check{\theta}_j - \theta_j \doteq \frac{1}{N_j} \sum_i \omega_{ji} x_{ji} \qquad (j = 1,2) \quad \ldots (13)$$

where $\omega_{ji} = \frac{p'_{ji}}{p_{ji}} \Big/ \sum_i \left(\frac{p'^2_{ji}}{p_{ji}} \right)$

From the two equations of (13) we get by subtraction

$$(\check{\theta}_1 - \check{\theta}_2) - (\theta_1 - \theta_2) \doteq \sum_{i=1}^{k} \left(\omega_{1i} \frac{x_{1i}}{N_1} - \omega_{2i} \frac{x_{2i}}{N_2} \right) \quad \ldots\ldots (14)$$

If $\theta_1 = \theta_2 = \theta$ (say), then $p_{1i} = p_{2i} = p_i$ (say) $(i = 1, 2, \ldots k)$.

and hence $\omega_{1i} = \omega_{2i} = \omega_i$ (say) for all i.

Under this hypothesis H_o i.e. when the samples belong to the same population, (14) becomes

$$\check{\theta}_1 - \check{\theta}_2 = \sum_i \omega_i \left(\frac{x_{1i}}{N_1} - \frac{x_{2i}}{N_2} \right)$$

$$= \sum_i \omega_i \left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2} \right) \qquad \text{(using (7))} \ldots (15)$$

Evidently $(\check{\theta}_1 - \check{\theta}_2)$ is a normal variate with zero mean and variance $\mathrm{Var}\,(\check{\theta}_1 - \check{\theta}_2) = \mathrm{Var}\,(\check{\theta}_1) + \mathrm{Var}\,(\check{\theta}_2)$.

But $\mathrm{Var}\,(\check{\theta}_j) = - \frac{1}{\frac{\partial^2 L}{\partial \theta_j^2}} \qquad (j = 1,2) \ldots (16)$

Therefore under the hypothesis H_o, using (12)

we/

we obtain from (16)

$$\mathrm{Var}(\check{\theta}_j) = 1\Big/N_j \sum_i \left(\frac{p_i'^2}{p_i}\right), \quad (j = 1,2) \quad \ldots(17)$$

also

$$\omega_i = \frac{p_i'}{p_i}\Big/\sum_i\left(\frac{p_i'^2}{p_i}\right) \quad (i = 1,2, \ldots K)\ldots(17a)$$

It is thus clear that $\dfrac{(\check{\theta}_1 - \check{\theta}_2)^2}{\mathrm{Var}(\check{\theta}_1)+\mathrm{Var}(\check{\theta}_2)}$ behaves as χ^2

with 1 d.f on the hypothesis Ho.

From (15) and (17a) we have

$$(\check{\theta}_1 - \check{\theta}_2) = \sum_i \frac{p_i'}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\Big/\sum_i\left(\frac{p_i'^2}{p_i}\right)$$

also from (17) we get

$$\mathrm{Var}(\check{\theta}_1)+\mathrm{Var}(\check{\theta}_2) = \left(\frac{1}{N_1}+\frac{1}{N_2}\right)\Big/\sum_i\left(\frac{p_i'^2}{p_i}\right)$$

Therefore

$$\frac{(\check{\theta}_1 - \check{\theta}_2)^2}{\mathrm{Var}(\check{\theta}_1)+\mathrm{Var}(\check{\theta}_2)} = \frac{\left\{\sum_i \frac{p_i'}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right\}^2}{\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\left(\sum_i \frac{p_i'^2}{p_i}\right)} \quad \ldots(18)$$

and hence our criterion to test whether or not the samples belong to the same population is given by the expression (18) which behaves as $\chi^2[1]$ where $\chi^2[1]$ means χ^2 with one degree of freedom.

4.03./

4.03. The same criterion is obtained by the method of likelihood.

Proceeding as before we have from (6) above

$$2L \doteq \text{Const} - \sum_j \sum_i (\eta_{ji}^2) \quad (j = 1,2,\ i = 1,2,\ldots k) \qquad \ldots.(A)$$

As usual to estimate θ_1 and θ_2 we have $\frac{\partial L}{\partial \theta_1} = 0$ and $\frac{\partial L}{\partial \theta_2} = 0$ respectively,

From these two equations we obtain to the same approximation as before the following two conditions respectively

$$\sum_i \frac{h'_{ji}}{\sqrt{h_{ji}}} \eta_{ji} = 0 \quad (j = 1, 2)$$

also identically we have

$$\sum_i \sqrt{h_{ji}}\, \eta_{ji} = 0, \quad (j = 1,2)$$

Thus $\sum_i \eta_{ji}^2$ $(j = 1,2)$ are two independent quadratic forms each subject to two independent linear constraints.

The linear constraints for the two quadratic forms can be written as

$$\left.\begin{aligned} X_j &\equiv \sum_i \sqrt{h_{ji}}\, \eta_{ji} = 0 \\ \text{and} \quad y_j &\equiv \sum_i \left(\frac{h'_{ji}}{\sqrt{h_{ji}}} \Big/ \sqrt{\left(\sum_i \frac{h'^2_{ji}}{h_{ji}}\right)} \right) \eta_{ji} = 0 \end{aligned}\right\} (j = 1,2) \qquad \ldots(B)$$

In/

In view of these restrictions

$\sum_i \eta_{ji}^2$, (j = 1,2) become

$$\sum_i \eta_{ji}^2 - x_j^2 - y_j^2 \quad , \quad (j = 1,2) \; \ldots\ldots \; \ldots \text{(C)}$$

Thus

$$2L_{H_1} \doteqdot \text{Const} - \sum_j \sum_i \eta_{ji}^2 + \sum_j (x_j^2 + y_j^2) \ldots \text{(D)}$$

where L_{H_1} stands for likelihood on the hypothesis H_1 that the two samples do not belong to the same population. On the hypothesis H_0 the quadratic expression $\sum_j \sum_i \eta_{ji}^2$ is subject to the linear constraints given by

$$X_j = 0 \qquad (j = 1,2)$$

and $\frac{dL}{d\theta} = 0$ i.e. $\sqrt{N_1}\, y_1 + \sqrt{N_2}\, y_2 = 0$ ...(E)

The last condition can be written in the standard form as

$$\sum_{j=1}^{2} \sqrt{N_j}\, y_j \Big/ \sqrt{\sum_{j=1}^{2} N_j} = 0$$

Hence under the hypothesis H_0, $2L_{H_0}$ can be written as

$$2L_{H_0} \doteqdot \text{Const} - \sum_j \sum_i \eta_{ji}^2 + \sum_j x_j^2 + \left(\frac{\sum_j \sqrt{N_j}\, y_j}{\sqrt{\sum_j N_j}} \right)^2 \ldots\ldots \text{(F)}$$

Thus from (D) and (F) by subtraction we get

$$-2\left(L_{H_0} - L_{H_1}\right) \doteqdot \sum_j y_j^2 - \frac{1}{N_1 + N_2} \left(\sum_j \sqrt{N_j}\, y_j \right)^2$$

$$\doteqdot \frac{N_1 N_2}{N_1 + N_2} \left(\frac{y_1}{\sqrt{N_1}} - \frac{y_2}{\sqrt{N_2}} \right)^2$$

[Since
$$y_1^2 + y_2^2 = \frac{1}{N_1 + N_2} \left\{ \left(\sqrt{N_1}\, y_1 + \sqrt{N_2}\, y_2 \right)^2 + N_1 N_2 \left(\frac{y_1}{\sqrt{N_1}} - \frac{y_2}{\sqrt{N_2}} \right)^2 \right\}$$
]

i.e./

i.e. - 2 log (likelihood ratio)

$$= \frac{1}{\left(\frac{1}{N_1}+\frac{1}{N_2}\right)} \left(\frac{Y_1}{\sqrt{N_1}} - \frac{Y_2}{\sqrt{N_2}}\right)^2 \cdots \text{(G)}$$

But on the hypothesis H_0 we have

$$\frac{Y_1}{\sqrt{N_1}} - \frac{Y_2}{\sqrt{N_2}} = \sum_i \left(\frac{p_i'}{\sqrt{p_i}} \Big/ \sqrt{\sum_i\left(\frac{p_i'^2}{p_i}\right)}\right)\left(\frac{\eta_{1i}}{\sqrt{N_1}} - \frac{\eta_{2i}}{\sqrt{N_2}}\right)$$

$$= \sum_i \left(\frac{p_i'}{p_i} \Big/ \sqrt{\sum_i\left(\frac{p_i'^2}{p_i}\right)}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)$$

(using (4) and (B)).

and since Y_j $(j = 1,2)$ are standard normal variates it follows that

$$- 2 \log(\text{likelihood ratio}) = \frac{\left\{\sum_i \frac{p_i'}{p_i}\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right\}^2}{\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\left(\sum_i \frac{p_i'^2}{p_i}\right)}$$

behaves as χ^2 with 1 d.f., a result which agrees with one obtained before in § 4.02.

In both the methods discussed above the p_i's can be calculated by substituting the estimated value of θ obtained on the basis of H_0 i.e. when $\theta_1 = \theta_2 = \theta$.

In this case the equation for estimating θ is given by

$$\frac{dL}{d\theta} = 0$$

i.e./

i.e.
$$\sum_i \frac{p_i'}{p_i}(x_{1i}+x_{2i}) = 0$$

i.e.
$$\sum_i \frac{p_i'}{p_i}\left\{(n_{1i}+n_{2i})-(N_1+N_2)p_i\right\}=0$$

i.e.
$$\sum_i \frac{p_i'}{p_i}(n_{1i}+n_{2i}) = 0 \quad \ldots\ldots\ldots \mathrm{I}$$

because $\sum_i p_i' = 0$.

This equation gives the optimum value $\check{\theta}$ of θ which is to be substituted in the expressions for p_i and p_i', $(i = 1,2, \ldots K)$ to obtain the values of p_i's and their differential coefficients.

4.04. <u>Aliter (for the second method)</u>:-

As before

$$2L \doteq Const - \sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}} \quad \ldots (i)$$

To get the optimum estimate of θ_1 and θ_2 we have the equations

$$\frac{\partial L}{\partial \theta_j} = 0 \quad \text{i.e.} \quad \sum_i \frac{p_{ji}'}{p_{ji}} x_{ji} = 0, \quad (j=1,2)$$

also

$$\sum_i x_{ji} = 0, \quad (j=1,2)$$

Let
$$z_{ji} = \frac{x_{ji}}{\sqrt{p_{ji}}}, \quad (j=1,2;\ i=1,2,\cdots K)$$

then the above equations become

become

$$\left.\begin{aligned} \sum_i \frac{h'_{ji}}{\sqrt{h_{ji}}}\, z_{ji} &= 0 \\ \sum_i \sqrt{h_{ji}}\; z_{ji} &= 0 \end{aligned}\right\} (j=1,2) \ldots\ldots (ii)$$

Again put

$$\left.\begin{aligned} x_j &= \sum_i \frac{h'_{ji}}{\sqrt{h_{ji}}}\, z_{ji} \\ t_j &= \sum_i \sqrt{h_{ji}}\; z_{ji} \end{aligned}\right\} (j=1,2) \ldots\ldots (iii)$$

Since Var $(z_{ji}) = N_j$, $(j=1,2;\ i=1,2,\cdots k)$

it follows that

$$\left.\begin{aligned} \mathrm{Var}\,(X_j) &= N_j \sum_i \frac{h'^{2}_{ji}}{h_{ji}} \\ \mathrm{Var}\,(t_j) &= N_j \\ \mathrm{Cov}\,(t_j, x_j) &= 0 \end{aligned}\right\} (j=1,2)$$

The quadratic form for $t_j,\ x_j$ is

$$\begin{bmatrix} t_j & x_j \end{bmatrix} \underline{V}_j^{-1} \begin{bmatrix} t_j \\ x_j \end{bmatrix}, (j=1,2) \ldots\ldots (iv)$$ Matrix Not.

where $\underline{V}_j = \underline{N}_j\, \underline{P}_j$ and $\underline{P}_j = \begin{bmatrix} 1 & 0 \\ 0 & \sum_i \left(\frac{h'^{2}_{ji}}{h_{ji}}\right) \end{bmatrix}$

It follows that $V_j^{-1} = \frac{1}{N_j} P_j^{-1}$

The/

The quadratic forms (iv) reduce to

$$\frac{t_j^2}{N_j} + \frac{x_j^2}{N_j \sum_i \left(\frac{k_{ji}'^2}{k_{ji}}\right)} \qquad (j = 1,2) \quad \ldots.(V)$$

Therefore

$$2L_{H_1} \doteqdot Const - \sum_j \sum_i \frac{z_{ji}^2}{N_j} + \sum_j \frac{t_j^2}{N_j} + \sum_j \left\{ \frac{x_j^2}{N_j \sum (k_{ji}'^2/k_{ji})} \right\} \quad \ldots.(VI)$$

When the samples come from the same population i.e. under the hypothesis H_0 we have

$P_1 = P_2 = P$ (say)

where
$$P = \begin{bmatrix} 1 & 0 \\ 0 & \sum_i \left(\frac{k_i'^2}{k_i}\right) \end{bmatrix}$$

and in addition to $t_j = 0$, $(j = 1,2)$ we have the condition $\frac{dL}{d\theta} = 0$ i.e. $X_1 + X_2 = 0$

Also $Var\,(X_1+X_2) = Var\,(X_1) + Var\,(X_2)$

$$= (N_1+N_2) \sum_i \left(\frac{k_i'^2}{k_i}\right)$$

Therefore as before

$$2L_{H_0} \doteqdot Const - \sum_j \sum_i \frac{z_{ji}^2}{N_j} + \sum_j \frac{t_j^2}{N_j} + \frac{(x_1+x_2)^2}{(N_1+N_2) \sum_i \left(\frac{k_i'^2}{k_i}\right)} \quad \ldots(vii)$$

Subtracting (vii) from (vi) we get

$$-2\left(L_{H_0} - L_{H_1}\right) \doteqdot \frac{1}{\sum_i \left(\frac{k_i'^2}{k_i}\right)} \left\{ \frac{x_1^2}{N_1} + \frac{x_2^2}{N_2} - \frac{(x_1+x_2)^2}{N_1+N_2} \right\}$$

$$= \frac{\left(\frac{x_1}{N_1} - \frac{x_2}{N_2}\right)^2}{\left\{\sum_i \left(\frac{k_i'^2}{k_i}\right)\right\} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \quad \ldots.(viii)$$

Thus

- 2 log (likelihood ratio) =

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

and hence our test criterion is the same as before.

4.05. Let us apply the foregoing result to the Binomial, Poissonian, Exponential and Normal cases.

1. Binomial law:-

This evidently is a case of one parameter and is a discrete distribution. Let the Observations in the (k+1) class categories of two, independent samples, be as follows:-

Samples	Class Categories						Total
	No success	1 success	...	r successes	...	k successes	
First Sample	n_{10}	n_{11}	...	n_{1r}	...	n_{1k}	N_1
Second Sample	n_{20}	n_{21}	...	n_{2r}	...	n_{2k}	N_2

The probability of an observation falling in the $(r+1)^{th}$ group (i.e. group of r successes) is given by ${}^{K}C_{r}\, p^{r} q^{K-r}$, where p is the probability of a success and $p+q = 1$. Thus in this case, referring to our results of one parameter case discussed above, we find that/

that $p_{jr} = {}^{K}C_r\, p_j^r\, q_j^{K-r}$, $(j = 1,2)$, p_j being the probability of success in the j^{th} sample and $q_j = 1 - p_j$.

Our equations determining p_j, $(j = 1,2)$ are given by

$$\sum_{r=0}^{K} \frac{p'_{jr}}{p_{jr}}\, x_{jr} = 0 \;, (j=1,2) \cdots\cdots (A)$$

where

$$x_{jr} = n_{jr} - N_j\, {}^{K}C_r\, p_j^r\, q_j^{K-r}$$

But

$$\frac{p'_{jr}}{p_{jr}} = \frac{d}{dp_{jr}} \log p_{jr} = \frac{r - Kp_j}{p_j q_j} \;, (j=1,2)$$

Therefore equations (A) become

$$\sum_{r=0}^{K} \left(\frac{r-Kp_j}{p_j q_j}\right)\left(n_{jr} - N_j\, {}^{K}C_r\, p_j^r\, q_j^{K-r}\right) = 0 \qquad (j = 1,2)$$

i.e.

$$\sum_r \left(\frac{r\, n_{jr}}{p_j q_j}\right) - \frac{N_j}{p_j q_j} \sum_r r\, {}^{K}C_r\, p_j^r\, q_j^{K-r} = 0 \;, (j=1,2) \cdots (A')$$

because $\sum_{r=0}^{K} \left(n_{jr} - N_j\, {}^{K}C_r\, p_j^r\, q_j^{K-r}\right) = N_j - N_j = 0$

Since $\sum_r r\, {}^{K}C_r\, p_j^r\, q_j^{K-r} \quad = Kp_j$

the equations (A') reduces to

$$\sum_r r\, n_{jr} - N_j K p_j = 0 \;, (j = 1,2)$$

Therefore/

Therefore

$$\overset{v}{p}_j = \frac{\sum_{r=0}^{k} r n_{jr}}{k N_j} \qquad (j = 1,2) \ldots (B)$$

we have seen that

$$\frac{p'_{jr}}{p_{jr}} = \frac{r - k p_j}{p_j q_j}$$

Therefore

$$\sum_{r=0}^{k} \frac{p'^2_{jr}}{p_{jr}} = \sum_{r=0}^{k} \left(\frac{p'_{jr}}{p_{jr}}\right)^2 p_{jr}$$

$$= \sum_{r=0}^{k} \left(\frac{r - k p_j}{p_j q_j}\right)^2 {}^kC_r\, p_j^r\, q_j^{k-r}$$

$$= \frac{1}{p_j^2 q_j^2}\left\{ \sum_r r^2\, {}^kC_r\, p_j^r q_j^{k-r} - 2 k p_j \sum_r r\, {}^kC_r\, p_j^r q_j^{k-r} + k^2 p_j^2 \sum_r {}^kC_r\, p_j^r q_j^{k-r} \right\}$$

$$= \frac{1}{p_j^2 q_j^2}\left\{ k p_j + k(k-1) p_j^2 - 2 k p_j (k p_j) + k^2 p_j^2 \right\}$$

$$= \frac{k}{p_j q_j}, \quad (j = 1,2) \cdots (C)$$

$$\left[\text{Since } \sum_{r=0}^{k} r^2\, {}^kC_r\, p_j^r q_j^{k-r} = k p_j + k(k-1) p_j^2\right]$$

The best estimated value of p on the hypothesis H_0 is given by

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

because/

because in this case $p_{1r} = p_{2r} = {}^{k}C_r\, p^r q^{k-r}$,

*

Since

$p_1 = p_2 = p$ (say)

Thus (D) becomes

$$\sum_r r(n_{1r}+n_{2r}) - (N_1+N_2)kp = 0$$

Therefore $$p^* = \frac{\sum_r r(n_{1r}+n_{2r})}{(N_1+N_2)k} \quad \ldots (D')$$

According to our final result obtained in § 4.02 (18) our test criterion for the hypothesis H_0 in this case is given by the expression

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

i.e.

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

χ^2 with 1 d.f. .....(E)

We/

We note that k in the denominator is one less than the number of categories and $\check{p}$ and $\check{q}$ $(= 1 - \check{p})$ are given by (D').

Hence if this is significant at 5% or 1% level of significance we reject our hypothesis.

This result (E) appears to be new and may be looked upon as a generalisation of the well known result for a 2 x 2 contingency table which follows as a particular case.

For let us put k = 1, then the table reduces to the form

	no success	~~one~~ success	Total
I Sample	n_{10}	n_{11}	N_1
II Sample	n_{20}	n_{21}	N_2

and the expression (E) becomes

$$\frac{(n_{11}n_{20} - n_{10}n_{21})^2 (n_{10} + n_{11} + n_{20} + n_{21})}{(n_{10}+n_{11})(n_{20}+n_{21})(n_{10}+n_{20})(n_{11}+n_{21})}$$

and it behaves as χ^2 with 1 d.f. - a familiar form.

2. Poissonian Law (case of one parameter):-

Let the observations of two independent samples in the same (k+1) categories be the same as assumed in the Binominal case. In this case the $(k+1)^{th}$ category may be looked upon as/

as including **the** case of k or more successes. The probability of an observation of the j^{th} sample falling in the $(r+1)^{th}$ group i.e. group of r successes is given by $e^{-\theta_j}\dfrac{\theta_j^r}{r!}$, $(j=1,2)$

where θ_j is the parameter of the j^{th} sample.

Thus $p_{jr} = e^{-\theta_j}\dfrac{\theta_j^r}{r!}$, $(j=1,2)$

When we assume the hypothesis H_o we have

$$p_{1r} = p_{2r} = p_r \text{ (say)}$$

where $p_r = e^{-\theta}\dfrac{\theta^r}{r!}$, $\quad$ if $\theta_1 = \theta_2 = \theta$ (say)

Now $\log p_r = -\theta + r\log\theta + \text{Const.}$

Therefore $\dfrac{p_r'}{p_r} = -1 + \dfrac{r}{\theta}$

Hence

$$\sum_{r=0}^{k}\frac{p_r'}{p_r}\left(\frac{n_{1r}}{N_1}-\frac{n_{2r}}{N_2}\right) = \sum_{r=0}^{k}\left(\frac{r}{\theta}-1\right)\left(\frac{n_{1r}}{N_1}-\frac{n_{2r}}{N_2}\right)$$

$$= \frac{1}{\theta}\sum_r r\left(\frac{n_{1r}}{N_1}-\frac{n_{2r}}{N_2}\right) \quad \text{---- (A)}$$

Also

$$\sum_{r=0}^{k}\frac{p_r'^2}{p_r} = \sum_r\left(\frac{p_r'}{p_r}\right)^2 p_r$$

$$= \sum_r\left(\frac{r}{\theta}-1\right)^2 e^{-\theta}\frac{\theta^r}{r!} = \frac{1}{\theta} \quad \text{---- (B)}$$

remembering that

$$\sum_{r=0}^{k} e^{-\theta}\frac{\theta^r}{r!} = 1, \quad \sum_{r=0}^{k} r e^{-\theta}\frac{\theta^r}{r!} = \theta$$

and

$$\sum_{r=0}^{k} r^2 e^{-\theta}\frac{\theta^r}{r!} = \theta + \theta^2$$

making/

Making the necessary substitutions in § 4.02 (18), our test criterion in this case reduces to

$$\left\{ \sum_{r=0}^{K} r\left(\frac{n_{1r}}{N_1} - \frac{n_{2r}}{N_2}\right)\right\}^2 \Big/ \theta\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \quad \ldots . (C)$$

and this follows the χ^2 distribution with 1 d.f. In the expression (C) above θ is to be replaced by its optimum estimate $\check{\theta}$ given by the equation § 4.03 (I) which in this case reduces to

$$\sum_r \left(\frac{r}{\theta} - 1\right)(n_{1r} + n_{2r}) = 0$$

i.e. $$\check{\theta} = \sum_r r(n_{1r} + n_{2r}) \Big/ N_1 + N_2 \quad \ldots . (D)$$

Substituting this value of θ in (C) we have the expression

$$= N_1 N_2 \left\{ \sum_{r=0}^{K} r\left(\frac{n_{1r}}{N_1} - \frac{n_{2r}}{N_2}\right)\right\}^2 \Big/ \sum_{r=0}^{K} r(n_{1r} + n_{2r}) \quad \ldots (E)$$

and it behaves as χ^2 with 1 d.f.

The expression (E) can be written as $\dfrac{N_1 N_2 (M_1 - M_2)^2}{N_1 M_1 + N_2 M_2}$ ... (E')

where $M_j = \dfrac{\sum_{r=0}^{K} r n_{jr}}{N_j}$, (j = 1,2).

This result (E) or (E') is evident otherwise, for the best estimate of parameters for the two samples are their respective means M_1 and M_2 and as such their difference can certainly be a measure of the validity or otherwise of the hypothesis/

hypothesis H_0.

3. Exponential Law:- (case of one parameter)

The law of distribution of x given by

$$f(x) = \frac{1}{\alpha}\, e^{-\frac{x}{\alpha}}$$

is called the Exponential law, where $x > 0$ and α is an unknown parameter (positive of course), to be determined from the observations at our disposal. It is clear that a variate under this law may be regarded as being distributed in a χ^2 population with two degrees of freedom.

Let the two independent samples classified into the same k different categories be the same as in § 3.02.

Let the interval of the i^{th} category be from $l_i - \frac{h}{2}$ to $l_i + \frac{h}{2}$, h being small $(i = 1, 2, \ldots K)$

Thus in this case we have

$$p_{ji} = \int_{l_i - \frac{h}{2}}^{l_i + \frac{h}{2}} \frac{1}{\alpha_j}\, e^{-\frac{x}{\alpha_j}}\, dx$$

$$\doteqdot \frac{h}{\alpha_j}\, e^{-\frac{l_i}{\alpha_j}}; \left(j = 1, 2; \; i = 1, 2, \cdots K\right)$$

Where α_j is the parameter of the j^{th} sample.

But on the hypothesis H_0 let $\alpha_1 = \alpha_2 = \alpha$ (say) and therefore $p_{1i} = p_{2i} = p_i$ (say) where

$$p_i \doteqdot \frac{h}{\alpha}\, e^{-\frac{l_i}{\alpha}}, \qquad \text{for all } i.$$

Thence/

Thence $\dfrac{p_i'}{p_i} = \dfrac{l_i}{\alpha^2} - \dfrac{1}{\alpha}$, (i = 1,2, ...k)

and $\sum_i \dfrac{p_i'^2}{p_i} = \dfrac{1}{\alpha^2}$ $\Bigg\}$ ... (i)

remembering that $\dfrac{h}{\alpha} \sum_{i=1}^{K} e^{-\frac{l_i}{\alpha}} = 1$

Following § 4.03 (I), the equation for the estimation of α is given by

$$\sum_i \left(\frac{l_i}{\alpha^2} - \frac{1}{\alpha}\right)\left(n_{1i} + n_{2i}\right) = 0$$

i.e.

$$\alpha = \frac{\sum_i l_i (n_{1i} + n_{2i})}{N_1 + N_2}$$

$$= \frac{N_1 M_1 + N_2 M_2}{N_1 + N_2} = M \text{ (say)} \text{ ... (ii)}$$

where $M_j = \dfrac{\sum_{i=1}^{K} l_i n_{ji}}{N_j}$, (j = 1,2)

and M is thus the mean of the combined samples. Making use of (i) and (ii) in our test criterion § 4.02 (18), it reduces to

$$\frac{N_1 N_2}{N_1 + N_2} \left(\frac{M_1 - M_2}{M}\right)^2 \text{ ....(iii)}$$

and this is $\chi^2_{[1]}$.

Thus expression (iii) is our test criterion under the assumed exponential law.

4./

4. Normal Law:- (case of one parameter).

Let the nature of the law of distribution be

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}},$$ where θ is an unknown parameter to be determined from the samples in question.

Let the two independent samples classified into the same k different categories be the same as those assumed in the preceding case.

Let the intervals of different categories be also identical.

Then in this case we have

$$p_{ji} = \frac{1}{\sqrt{2\pi}} \int_{l_i-\frac{h}{2}}^{l_i+\frac{h}{2}} e^{-\frac{(x-\theta_j)^2}{2}} dx$$

$$\doteqdot \frac{h}{\sqrt{2\pi}} e^{-\frac{(l_i-\theta_j)^2}{2}}, \quad (j=1,2;\ i=1,2,\cdots k)$$

Where θ_j is the parameter of the j^{th} sample.

But if they come from the same population let $\theta_1 = \theta_2 = \theta$ (say), then $p_{1i} = p_{2i} = p_i$ (say) and $p_i \doteqdot \frac{h}{\sqrt{2\pi}} e^{-(l_i-\theta)^2/2}$, for all i.

From this it follows that

$$\left.\begin{aligned} \frac{p_i'}{p_i} &= l_i - \theta, \quad (i=1,2,\cdots k) \\ \text{and} \quad \sum_i \frac{p_i'^2}{p_i} &= 1 \end{aligned}\right\} \cdots (i)$$

because

$$\frac{h}{\sqrt{2\pi}} \sum_{i=1}^{k} e^{-\frac{(l_i-\theta)^2}{2}} = 1$$

Also the equation for the optimum estimate of θ is given by

$$\sum_i (l_i - \theta)(n_{1i} + n_{2i}) = 0$$

Hence $\hat{\theta} = \sum_i l_i (n_{1i} + n_{2i}) / N_1 + N_2$

$$= \frac{N_1 M_1 + N_2 M_2}{N_1 + N_2} \text{ ....(ii)}$$

where $M_j = \sum_i l_i n_{ji} / N_j$, (j = 1,2).

With proper substitutions our test criterion reduces to

$$\frac{\left\{ \sum_i l_i \left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2} \right) \right\}^2}{\frac{1}{N_1} + \frac{1}{N_2}}$$

i.e.

$$= \frac{N_1 N_2}{N_1 + N_2} (M_1 - M_2)^2 \text{ ..... (iii)}$$

and this behaves as $\chi^2_{[1]}$.

CHAPTER V.

I. Case of two parameters:-

Let us take the case of two parameters, i.e. it is assumed that the minimum number of parameters necessary to specify the distribution law is two. Also let our two given independent samples be the same as in § 3.02.

Let $\phi(x; \theta_j, \phi_j)$ be the assumed law of distribution of the jth sample, θ_j, ϕ_j being the two unknown parameters involved.

As before let

$$\left.\begin{aligned} x_{ji} &= n_{ji} - N_j p_{ji} \\ \text{where} \quad p_{ji} &= \int_{l_i}^{l_{i+1}} \phi(x; \theta_j, \phi_j)\, dx \\ &= r_{ji}(\theta_j, \phi_j) \quad (say) \end{aligned}\right\} \quad \cdots\cdots (1)$$

$$(j = 1, 2;\ i = 1, 2, \cdots k)$$

the notations having the same meaning as in § 3.02 and in § 4.02.

5.01. First Method:-

As in § 4.02 (8) we have

$$2L \doteqdot \text{Const.} - \sum_j \sum_i \frac{x_{ji}^2}{N_j r_{ji}} \quad \cdots\cdots (2)$$

Now/

Now,

$$\left.\begin{aligned}
\frac{\partial L}{\partial \theta_j} &\doteq \sum_i \frac{1}{r_{ji}}\left(\frac{\partial r_{ji}}{\partial \theta_j}\right) x_{ji} \\
\frac{\partial L}{\partial \phi_j} &\doteq \sum_i \frac{1}{r_{ji}}\left(\frac{\partial r_{ji}}{\partial \phi_j}\right) x_{ji} \\
\frac{\partial^2 L}{\partial \theta_j^2} &\doteq -N_j \sum_i \frac{1}{r_{ji}}\left(\frac{\partial r_{ji}}{\partial \theta_j}\right)^2 \\
\frac{\partial^2 L}{\partial \theta_j \partial \phi_j} &\doteq -N_j \sum_i \frac{1}{r_{ji}}\left(\frac{\partial r_{ji}}{\partial \theta_j}\right)\left(\frac{\partial r_{ji}}{\partial \phi_j}\right) \\
\frac{\partial^2 L}{\partial \phi_j^2} &\doteq -N_j \sum_i \frac{1}{r_{ji}}\left(\frac{\partial r_{ji}}{\partial \phi_j}\right)^2
\end{aligned}\right\} \cdots (3) \quad (j=1,2)$$

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

Thus to get the optimum values of θ_j and ϕ_j we must have

$$\frac{\partial L}{\partial \theta_j} = 0, \quad \frac{\partial L}{\partial \phi_j} = 0 \quad (j = 1,2)$$

and their values $\check{\theta}_j$ and $\check{\phi}_j$ must satisfy these equations simultaneously i.e.

$$\left(\frac{\partial L}{\partial \theta_j}\right)_{\check{\theta}_j, \check{\phi}_j} = 0 = \left(\frac{\partial L}{\partial \phi_j}\right)_{\check{\theta}_j, \check{\phi}_j} \quad \cdots\cdots (4) \quad (j=1,2)$$

But $\frac{\partial L}{\partial \theta_j}$ and $\frac{\partial L}{\partial \phi_j}$ are functions of θ_j and ϕ_j

and so /

and so

$$\left.\begin{aligned}
\left(\frac{\partial L}{\partial \theta_j}\right)_{\check{\theta}_j, \check{\phi}_j} &= \left(\frac{\partial L}{\partial \theta_j}\right) + (\check{\theta}_j - \theta_j)\frac{\partial^2 L}{\partial \theta_j^2} + (\check{\phi}_j - \phi_j)\frac{\partial^2 L}{\partial \theta_j \partial \phi_j} \quad \text{(approx).} \\
&\text{and} \\
\left(\frac{\partial L}{\partial \phi_j}\right)_{\check{\theta}_j, \check{\phi}_j} &= \left(\frac{\partial L}{\partial \phi_j}\right) + (\check{\theta}_j - \theta_j)\frac{\partial^2 L}{\partial \theta_j \partial \phi_j} + (\check{\phi}_j - \phi_j)\frac{\partial^2 L}{\partial \phi_j^2} \quad \text{(approx.)}
\end{aligned}\right\} \cdots (5)$$

From (5) in view of (4) we get

$$\left.\begin{aligned}
(\check{\theta}_j - \theta_j)\frac{\partial^2 L}{\partial \theta_j^2} + (\check{\phi}_j - \phi_j)\frac{\partial^2 L}{\partial \phi_j \partial \theta_j} &= -\frac{\partial L}{\partial \theta_j} \\
(\check{\theta}_j - \theta_j)\frac{\partial^2 L}{\partial \theta_j \partial \phi_j} + (\check{\phi}_j - \phi_j)\frac{\partial^2 L}{\partial \phi_j^2} &= -\frac{\partial L}{\partial \phi_j}
\end{aligned}\right\} \quad (j = 1, 2) \quad \ldots\ldots(6)$$

From (6) we get

$$\begin{bmatrix} \check{\theta}_j - \theta_j \\ \check{\phi}_j - \phi_j \end{bmatrix} = A_j^{-1} \begin{bmatrix} \frac{\partial L}{\partial \theta_j} \\ \frac{\partial L}{\partial \phi_j} \end{bmatrix} \quad \ldots\ldots (7) \quad (j = 1, 2)$$

where A_j^{-1} is the reciprocal matrix of A_j and

A_j/

$$A_j = \begin{bmatrix} -\dfrac{\partial^2 L}{\partial \theta_j^2} & -\dfrac{\partial^2 L}{\partial \phi_j \partial \theta_j} \\ -\dfrac{\partial^2 L}{\partial \theta_j \partial \phi_j} & -\dfrac{\partial^2 L}{\partial \phi_j^2} \end{bmatrix}$$

$$= N_j \begin{bmatrix} \sum_i \dfrac{1}{h_{ji}} \left(\dfrac{\partial h_{ji}}{\partial \theta_j}\right)^2 & \sum_i \dfrac{1}{h_{ji}} \left(\dfrac{\partial h_{ji}}{\partial \theta_j}\right)\left(\dfrac{\partial h_{ji}}{\partial \phi_j}\right) \\ \sum_i \dfrac{1}{h_{ji}} \left(\dfrac{\partial h_{ji}}{\partial \theta_j}\right)\left(\dfrac{\partial h_{ji}}{\partial \phi_j}\right) & \sum_i \dfrac{1}{h_{ji}} \left(\dfrac{\partial h_{ji}}{\partial \phi_j}\right)^2 \end{bmatrix}$$

(using (3)).

$= N_j \; P_j$ (say) .... (8)

In the above solution we assume that A_j ($j = 1,2$) is non-singular. We also assume that

$\dfrac{\partial^2 L}{\partial \theta_j \partial \phi_j} = \dfrac{\partial^2 L}{\partial \phi_j \partial \theta_j}$, for all j and hence P_j is symmetrical i.e. $P_j' = P_j$, ($j = 1,2$). It is evident that N_j is scalar.

From two equations of (7) we have by subtraction

$$\begin{bmatrix} (\check{\theta}_1 - \check{\theta}_2) - (\theta_1 - \theta_2) \\ (\check{\phi}_1 - \check{\phi}_2) - (\phi_1 - \phi_2) \end{bmatrix} = A_1^{-1} \begin{bmatrix} \dfrac{\partial L}{\partial \theta_1} \\ \dfrac{\partial L}{\partial \phi_1} \end{bmatrix} - A_2^{-1} \begin{bmatrix} \dfrac{\partial L}{\partial \theta_2} \\ \dfrac{\partial L}{\partial \phi_2} \end{bmatrix}$$

$$= P_1^{-1} \begin{bmatrix} \dfrac{1}{N_1} \dfrac{\partial L}{\partial \theta_1} \\ \dfrac{1}{N_1} \dfrac{\partial L}{\partial \phi_1} \end{bmatrix} - P_2^{-1} \begin{bmatrix} \dfrac{1}{N_2} \dfrac{\partial L}{\partial \theta_2} \\ \dfrac{1}{N_2} \dfrac{\partial L}{\partial \phi_2} \end{bmatrix} \cdots (9)$$

for $A_j^{-1} = N_j^{-1} P_j^{-1}$, ($j = 1,2$)

Now on the hypothesis H_0 we have $\theta_1 = \theta_2 = \theta$ (say) and $\phi_1 = \phi_2 = \phi$ (say) and hence $p_{1i} = p_{2i} = p_i$ (say)

Therefore

$$P_1 = P_2 = \begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)^2 & \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(\frac{\partial p_i}{\partial \phi}\right) \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(\frac{\partial p_i}{\partial \phi}\right) & \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)^2 \end{bmatrix}$$

$$= \quad P \text{ (say)} \quad \ldots\ldots\ldots\ldots(10)$$

Also

$$\begin{bmatrix} \frac{1}{N_j}\frac{\partial L}{\partial \theta_j} \\ \frac{1}{N_j}\frac{\partial L}{\partial \phi_j} \end{bmatrix} = \begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\frac{x_{ji}}{N_j} \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\frac{x_{ji}}{N_j} \end{bmatrix} \ldots\ldots (11)$$

$$(j = 1, 2)$$

(using (3)).

Thus subject to hypothesis H_0 (9) can be written because of (10) and (11) as

$$\begin{bmatrix} \check{\theta}_1 - \check{\theta}_2 \\ \check{\phi}_1 - \check{\phi}_2 \end{bmatrix} = P^{-1}\begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(\frac{x_{1i}}{N_1} - \frac{x_{2i}}{N_2}\right) \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\left(\frac{x_{1i}}{N_1} - \frac{x_{2i}}{N_2}\right) \end{bmatrix}$$

$$= P^{-1}\begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right) \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\phi}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right) \end{bmatrix}$$

$\left(\text{since } \frac{x_{1i}}{N_1}-\frac{x_{2i}}{N_2}=\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)$

Transposing (12) and remembering that $P' = P$, we get

$$\left[\check{\theta}_1-\check{\theta}_2 \quad \check{\phi}_1-\check{\phi}_2\right] =$$

$$\left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right) \quad \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\phi}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right] P^{-1} \quad \ldots\ldots (13)$$

Since $(\check{\theta}_1 - \check{\theta}_2)$ and $(\check{\phi}_1 - \check{\phi}_2)$ are normal variates to discuss their joint distribution we take the quadratic form

$$\left[\check{\theta}_1-\check{\theta}_2 \quad \check{\phi}_1-\check{\phi}_2\right] V^{-1} \begin{bmatrix} \check{\theta}_1-\check{\theta}_2 \\ \check{\phi}_1-\check{\phi}_2 \end{bmatrix} \quad \ldots\ldots (14)$$

Where V = variance matrix of $(\check{\theta}_1 - \check{\theta}_2)$ and $(\check{\phi}_1 - \check{\phi}_2)$

$$= \begin{bmatrix} \mathrm{Var}\,(\check{\theta}_1-\check{\theta}_2) & \mathrm{Cov}(\check{\theta}_1-\check{\theta}_2)\,(\check{\phi}_1-\check{\phi}_2) \\ \mathrm{Cov}(\check{\theta}_1-\check{\theta}_2)\,(\check{\phi}_1-\check{\phi}_2) & \mathrm{Var}\,(\check{\phi}_1-\check{\phi}_2) \end{bmatrix} \quad (15)$$

But $\mathrm{Var}\,(\check{\theta}_1 - \check{\theta}_2) = \mathrm{Var}\,(\check{\theta}_1) + \mathrm{Var}\,(\check{\theta}_2)$

$$\mathrm{Cov}\left\{(\check{\theta}_1 - \check{\theta}_2), (\check{\phi}_1 - \check{\phi}_2)\right\} = \mathrm{cov}\,(\check{\theta}_1, \check{\phi}_1) + \mathrm{Cov}\,(\check{\theta}_2, \check{\phi}_2)$$

$$\mathrm{Var}(\check{\phi}_1 - \check{\phi}_2) = \mathrm{Var}\,(\check{\phi}_1) + \mathrm{Var}\,(\check{\phi}_2)$$

Thus (15) can be written as

$$V = V_1 + V_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

Where V_j is the variance matrix of $\check{\theta}_j$, $\check{\phi}_j$. $(j = 1,2)$

Also

$$V_j = \begin{bmatrix} -E\left(\frac{\partial^2 L}{\partial \theta_j^2}\right) & -E\left(\frac{\partial^2 L}{\partial \theta_j \partial \phi_j}\right) \\ -E\left(\frac{\partial^2 L}{\partial \theta_j \partial \phi_j}\right) & -E\left(\frac{\partial^2 L}{\partial \phi_j^2}\right) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} N_j \sum_i \frac{1}{h_{ji}}\left(\frac{\partial h_{ji}}{\partial \theta_j}\right)^2 & N_j \sum_i \frac{1}{h_{ji}}\left(\frac{\partial h_{ji}}{\partial \theta_j}\right)\left(\frac{\partial h_{ji}}{\partial \phi_j}\right) \\ N_j \sum_i \frac{1}{h_{ji}}\left(\frac{\partial h_{ji}}{\partial \theta_j}\right)\left(\frac{\partial h_{ji}}{\partial \phi_j}\right) & N_j \sum_i \frac{1}{h_{ji}}\left(\frac{\partial h_{ji}}{\partial \phi_j}\right)^2 \end{bmatrix}^{-1}$$

$$= \frac{1}{N_j} P_j^{-1}, \quad (j=1,2) \quad \ldots (17)$$

Therefore (16) becomes $V = \sum_{j=1}^{2} \frac{1}{N_j} P_j^{-1}$ .. (18)

And/

And on the hypothesis H_0 this (18) can be written as

$$V = \left(\frac{1}{N_1} + \frac{1}{N_2}\right) P^{-1}$$

because then we have $P_1^{-1} = P_2^{-1} = P^{-1}$

Therefore

$$V^{-1} = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} P \quad \text{.........(19)}$$

In view of (12), (13) and (19) we can write the expression (14) as

$$\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} \left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \quad \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right] P^{-1} \begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \end{bmatrix} \quad \text{--- (20)}$$

Thus our test criterion in this case reduces to (20) where P is given by (10) and this clearly behaves as χ^2 with 2 d.f. and may be written as $\chi^2_{[2]}$

5.02. Second Method:-

Let us approach the same problem by the method of likelihood. We shall see that the result given by this method agrees with the one obtained by the first method.

As/

As before we have

$$2L \doteqdot \text{Const} - \sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}} \qquad \ldots(i)$$

where the symbols have the same meaning as in the preceding discussion.

Thus for obtaining the optimum estimate of $\theta_j, \phi_j \quad (j = 1,2)$ we have the two sets of equations.

$$\frac{\partial L}{\partial \theta_j} = 0 \text{ and } \qquad \frac{\partial L}{\partial \phi_j} = 0 \quad (j = 1,2)$$

These reduce to

$$\left.\begin{aligned} &\sum_i \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \theta_j}\right) x_{ji} = 0 \\ &\sum_i \frac{1}{p_{ji}} \left(\frac{\partial p_{ji}}{\partial \phi_j}\right) x_{ji} = 0 \end{aligned}\right\} \quad (j = 1,2)..(ii)$$

and in addition we have identically

$$\sum_i x_{ji} = 0 \qquad (j = 1,2)..(ii!)$$

Let us put $\quad Z_{ji} = x_{ji}/\sqrt{p_{ji}}$

$$(j = 1,2\,;\ i = 1,2,\ \ldots k) \ \ldots..(iii)$$

Then/

Then (i), (ii) and (ii')* can be written as

$$2\,L \doteqdot \text{Const} - \sum_j \sum_i \frac{z_{ji}^2}{N_j} \qquad \ldots\ldots(iv)$$

$$\left.\begin{aligned} &\sum_i \frac{1}{\sqrt{r_{ji}}}\left(\frac{\partial r_{ji}}{\partial \theta_j}\right) z_{ji} = 0 \\ &\sum_i \frac{1}{\sqrt{r_{ji}}}\left(\frac{\partial r_{ji}}{\partial \phi_j}\right) z_{ji} = 0 \\ \text{and} \quad &\sum_i \sqrt{r_{ji}}\, z_{ji} = 0 \end{aligned}\right\} \quad (j = 1,2)\ldots(v)$$

Further let

$$\left.\begin{aligned} X_j &= \sum_i \frac{1}{\sqrt{r_{ji}}}\left(\frac{\partial r_{ji}}{\partial \theta_j}\right) z_{ji} \\ Y_j &= \sum_i \frac{1}{\sqrt{r_{ji}}}\left(\frac{\partial r_{ji}}{\partial \phi_j}\right) z_{ji} \\ t_j &= \sum_i \sqrt{r_{ji}}\, z_{ji} \end{aligned}\right\} \quad (j = 1,2)\ldots(vi)$$

It is clear from the above relations that t_j is orthogonal to X_j and Y_j and that X_j, Y_j, t_j are normal variates for all values of j.

The/

The quadratic form for t_j, X_j, Y_j is given by

$$\begin{bmatrix} t_j & X_j & Y_j \end{bmatrix} \quad U_j^{-1} \begin{bmatrix} t_j \\ x_j \\ y_j \end{bmatrix} \qquad (j = 1,2)..(vii)$$

Where U_j = variance matrix of t_j, X_j, Y_j

$$= \begin{bmatrix} N_j & 0 & 0 \\ 0 & N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_j}\right)^2 & N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_j}\right)\left(\frac{\partial h_{ji}}{\partial \phi_j}\right) \\ 0 & N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_j}\right)\left(\frac{\partial h_{ji}}{\partial \phi_j}\right) & N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \phi_j}\right)^2 \end{bmatrix}$$

$(j = 1, 2)$.

Therefore expression (vii) can be written as

$$\frac{t_j^2}{N_j} + \frac{1}{N_j} \quad \begin{bmatrix} X_j & Y_j \end{bmatrix} \quad P_j^{-1} \begin{bmatrix} X_j \\ Y_j \end{bmatrix}, \quad (j = 1,2)..(viii)$$

Where P_j is the same as in (8) of the preceding Section.

Thus in view of the restrictions on Z_{ji} given by (V), we can write $\sum_j \sum_i \frac{Z_{ji}^2}{N_j}$ as

$$\sum_j \sum_i \frac{Z_{ji}^2}{N_j} - \sum_j \frac{t_j^2}{N_j} - \sum_j \frac{1}{N_j} \begin{bmatrix} x_j & y_j \end{bmatrix} P_j^{-1} \begin{bmatrix} x_j \\ y_j \end{bmatrix}$$

and/

and therefore

$$2L_{H_1} \doteqdot \text{Const} - \sum_j \sum_i \frac{z_{ji}^2}{N_j} + \sum_j \frac{t_j^2}{N_j} + \sum_j \frac{1}{N_j}\begin{bmatrix} x_j & y_j \end{bmatrix} P_j^{-1} \begin{bmatrix} x_j \\ y_j \end{bmatrix} \qquad \text{....(ix)}$$

But when the samples are assumed to come from the same population (i.e. on the hypothesis H_o), we must have $\theta_1 = \theta_2 = \theta$ (say) and $\phi_1 = \phi_2 = \phi$ (say) and therefore $p_{1i} = p_{2i} = p_i$ (say) and $P_1 = P_2 = P$ (say); p's and P's having the same meaning as in the preceding section.

Thus on the hypothesis H_o we get instead of two sets of equations as in (ii) only one set, namely

$$\left.\begin{aligned} \sum_i \frac{1}{p_{1i}}\left(\frac{\partial p_{1i}}{\partial \theta_1}\right) x_{1i} + \sum_i \frac{1}{p_{2i}}\left(\frac{\partial p_{2i}}{\partial \theta_2}\right) x_{2i} = 0 \\ \sum_i \frac{1}{p_{1i}}\left(\frac{\partial p_{1i}}{\partial \phi_1}\right) x_{1i} + \sum_i \frac{1}{p_{2i}}\left(\frac{\partial p_{2i}}{\partial \phi_2}\right) x_{2i} = 0 \end{aligned}\right\} \text{...(x)}$$

i.e. $X_1 + X_2 = 0$ and $Y_1 + Y_2 = 0$ (because of (vi)).

Of/

Of course in addition to these two we have the conditions-

$$t_1 = 0, \text{ and } t_2 = 0$$

Thus on the hypothesis H_0 we have

$$2L_{H_0} \doteq Const - \sum_j \sum_i \frac{z_{ji}^2}{N_j} + \sum_j \frac{t_j^2}{N_j}$$

$$+ \frac{1}{N_1+N_2} \begin{bmatrix} x_1+x_2 & y_1+y_2 \end{bmatrix} P^{-1} \begin{bmatrix} x_1+x_2 \\ y_1+y_2 \end{bmatrix} \quad \text{---(xi)}$$

for in this case the variance matrix of x_1+x_2 and y_1+y_2 is $(N_1+N_2)P$

From (ix) and (xi) by Subtraction we get

$$-2\left(L_{H_0} - L_{H_1}\right) \doteq \frac{1}{N_1} \begin{bmatrix} x_1 & y_1 \end{bmatrix} P^{-1} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \frac{1}{N_2} \begin{bmatrix} x_2 & y_2 \end{bmatrix} P^{-1} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$- \frac{1}{N_1+N_2} \begin{bmatrix} x_1+x_2 & y_1+y_2 \end{bmatrix} P^{-1} \begin{bmatrix} x_1+x_2 \\ y_1+y_2 \end{bmatrix} \quad \text{---(xii)}$$

i.e./

i.e. $-2 \log$ (likelihood ratio) =

$$\left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1}\left[\left(\frac{x_1}{N_1}-\frac{x_2}{N_2}\right)\quad\left(\frac{y_1}{N_1}-\frac{y_2}{N_2}\right)\right]P^{-1}\begin{bmatrix}\left(\frac{x_1}{N_1}-\frac{x_2}{N_2}\right)\\ \left(\frac{y_1}{N_1}-\frac{y_2}{N_2}\right)\end{bmatrix}\quad\text{....(xiii)}$$

But on the hypothesis H_o we have

$$\frac{x_1}{N_1}-\frac{x_2}{N_2}=\sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)$$

and

$$\frac{y_1}{N_1}-\frac{y_2}{N_2}=\sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\phi}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)$$

Making use of these relations the expression on the right hand side of (Xiii) becomes

$$\left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1}\left[\sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\quad\sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\phi}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right]P^{-1}\begin{bmatrix}\sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\\ \sum_i\frac{1}{p_i}\left(\frac{\partial p_i}{\partial\phi}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\end{bmatrix}\quad\text{---(xiv)}$$

and/

and this behaves as χ^2 with 2 d.f. i.e. as $\chi^2_{[2]}$ and this agrees with the final result obtained by the first method.

In the final result we have to substitute the values of p_i's and their differential co-efficients in terms of observations at our disposal.

Here the values of p_i and its d.c. for all i's are determined by estimating the values of θ and ϕ on the hypothesis H_0.

The equations for estimating θ and ϕ are

$$\left.\begin{aligned}&\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(x_{1i}+x_{2i}\right)=0\\ \text{and}\quad&\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\left(x_{1i}+x_{2i}\right)=0\end{aligned}\right\}$$

i.e.

$$\left.\begin{aligned}&\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)\left(n_{1i}+n_{2i}\right)=0\\ \text{and}\quad&\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \phi}\right)\left(n_{1i}+n_{2i}\right)=0\end{aligned}\right\}\text{---(xv)}$$

and/

and their estimated values $\check{\theta}$ and $\check{\phi}$ must satisfy the two equations in (XV) simultaneously.

Thus all the quantities involved in (xiv) can be determined in terms of known quantities i.e. data of the two samples.

Of course in any particular case it may not be easy and straightforward. As a matter of fact it may be a very difficult affair indeed. In these cases the method of approximation available in the solution of maximal likelihood equations will enable us to approximate to true value. This method is indicated by R.S.Koshal(1933) and F. Garwood (1941).

It is thus clear that the method is available in all cases where optimum solutions of the parameters can be obtained

II. <u>Case of Several Parameters:-</u>

The case of two parameters discussed above can easily be extended to one involving several parameters. In all that follows we assume that the minimum number of independent parameters necessary to specify the law of distribution is s where $s \leqslant k-1$; k being the number of different categories into which each of the two given samples is classified.

The/

The significance of this restriction on the number of independent parameters will become clear later on when we discuss the various degrees of freedom of our test criterion involved in the solution of the problem.

Let $\phi(x; \theta_{j1}, \theta_{j2}, \ldots \theta_{js})$ be the assumed law of distribution of the j^{th} Sample; θ_{jr} $(r = 1,2, \ldots s)$ being its s unknown parameters.

5.03. <u>First Method</u>:-

As before we have

$$2L \doteqdot \text{Const.} - \sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}} \qquad \ldots(1)$$

the various symbols involved having the same meaning.

Now

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

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

neglecting/

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

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

$$\text{i.e.} \quad \frac{\partial L}{\partial \theta_{jr}} = 0 \quad (j = 1,2;\ r = 1,2 \ldots s) \ldots (3)$$

Their estimated value $\check{\theta}_{jr}$ will satisfy the equations (3) i.e.

$$\left(\frac{\partial L}{\partial \theta_{jr}}\right)_{\check{\theta}_{jr}} = 0 \text{ for all } j \text{ and } r \ \ldots. \ (4).$$

Thus $\check{\theta}_{1r}$ and $\check{\theta}_{2r}$ for all r will be the solutions of the set of equations

$\frac{\partial L}{\partial \theta_{1r}} = 0$ and of the set $\frac{\partial L}{\partial \theta_{2r}} = 0$

(r = 1,2, ...s) respectively.

By Taylor's theorem we have

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

(approximately) ---- (5)

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

Hence from (5) with the help of (3) we get

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

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

Before proceeding further we explain some matrix notations which will simplify writing lengthy and complicated expressions.
The matrix $[a_{jr}]$ $(r = 1,2, \ldots.s)$ for fixed j denotes a column matrix with elements obtained by giving different possible values to r in the expression a_{jr} and $[a_{jr}]'$ as usual denotes the transposed of the matrix $[a_{jr}]$ and is thus a row matrix. Different values to j will mean different such column and row matrices respectively.
Thus

$$[a_{jr}]\ (r=1,2,\cdots s) = \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{js} \end{bmatrix}$$

$$[a_{jr}]'\ (r=1,2,\cdots s) = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{js} \end{bmatrix}$$

and $[a_{jr}]$ $(j=1,2;\ r=1,2,\cdots s)$ stands for/

$\left[a_{1r}\right]$ and $\left[a_{2r}\right]$ $(r=1,2,\cdots s)$

Similar interpretations being for the transposed matrix.

Also $\left[a_{jr}\right] = \left[b_{jr}\right]$ $(j = 1,2;\ r = 1,2,\ldots .s)$

means $\left[a_{1r}\right] = \left[b_{1r}\right]$ and $\left[a_{2r}\right] = \left[b_{2r}\right]$

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

Similarly $\left[a_{jmr}\right]$ for fixed j and $(m,r = 1,2,\ ..s)$ denotes a square matrix of order s.

Thus $\left[a_{jmr}\right]$ $(m,r = 1,\ 2,\ \ldots\ s) =$

$$\begin{bmatrix} a_{j11} & a_{j12} & \cdots & a_{j1s} \\ a_{j21} & a_{j22} & \cdots & a_{j2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{js1} & a_{js2} & \cdots & a_{jss} \end{bmatrix}$$

Also $\left[a_{jmr}\right]$ $(j = 1,2;\ m,\ r = 1,2,\ \ldots .s)$ means

$\left[a_{1mr}\right]$ and $\left[a_{2mr}\right]$ $(m,\ r = 1,\ 2,\ \ldots .s)$

With/

With these explanations we have from (6)

$$\left[\overset{v}{\theta}_{jr} - \theta_{jr}\right] = A_j^{-1}\left[\frac{\partial L}{\partial \theta_{jr}}\right]$$

$$(j = 1,2;\ r = 1,2,\ldots s) \qquad \ldots .(7)$$

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

In the result (7) we assume that A_j is non-singular. 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 a symmetrical matrix.

With the help of (2) we have

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

where $P_j = \left[\sum_i \frac{1}{p_{ji}}\left(\frac{\partial p_{ji}}{\partial \theta_{jm}}\right)\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right)\right]$

$$(m, r = 1, 2, \cdots s)$$

Evidently P_j is also symmetrical i.e. $P_j' = P_j$

$(j = 1,2)$.

Thus/

Thus (7) can be written, in view of (8), as

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

From the two sets of equations of (9) we have by subtraction

$$\left[\left(\overset{v}{\theta}_{1r}-\overset{v}{\theta}_{2r}\right)-\left(\theta_{1r}-\theta_{2r}\right)\right] = \frac{1}{N_1} P_1^{-1}\left[\frac{\partial L}{\partial \theta_{1r}}\right] - \frac{1}{N_2} P_2^{-1}\left[\frac{\partial L}{\partial \theta_{2r}}\right] \quad (r=1,2,\cdots s)$$

$$= P_1^{-1}\left[\sum_i \frac{1}{p_{1i}}\left(\frac{\partial p_{1i}}{\partial \theta_{1r}}\right)\frac{x_{1i}}{N_1}\right] - P_2^{-1}\left[\sum_i \frac{1}{p_{2i}}\left(\frac{\partial p_{2i}}{\partial \theta_{2r}}\right)\frac{x_{2i}}{N_2}\right] \quad (r=1,2,\cdots s)$$
$$\cdots(10)$$

(using 2).

On the hypothesis H_o we have $\theta_{1r} = \theta_{2r} = \theta_r$ (say) $(r = 1,2, \ldots s)$

and/

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

and thence $P_1 = P_2 = P$ (say).

Thus on the hypothesis H_o, (10) becomes, using § 5 (1),

$$\left[\check{\theta}_{1r} - \check{\theta}_{2r}\right] = P^{-1}\left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right]$$

$$(r = 1, 2, \cdots s) \cdots (11)$$

where $P = \left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_m}\right)\left(\frac{\partial p_i}{\partial \theta_r}\right)\right], (m, r = 1, 2, \cdots s)$

To study the joint distribution of $\left(\check{\theta}_{1r} - \check{\theta}_{2r}\right)$ $(r = 1, 2, \ldots .s)$

we obtain the quadratic form in these s normal variates namely

$$\left[\check{\theta}_{1r} - \check{\theta}_{2r}\right]' V^{-1} \left[\check{\theta}_{1r} - \check{\theta}_{2r}\right]$$

$$(r = 1,2, \ldots .s) \ldots\ldots(12)$$

Where V is the variance matrix of the variates involved.

It/

It is clear that $V = V_1 + V_2$ ..... (13)

where V_j is the variance matrix of θ_{jr}

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

We also note that $V_j = \left[E(A_j)\right]^{-1}$ $(j = 1,2)$

$$= \frac{1}{N_j} P_j^{-1}$$

$$= \frac{1}{N_j} P^{-1} \quad \text{(on the hypothesis } H_o \text{)} \ldots (14)$$

Where $\left[E(A_j)\right]$ means the matrix obtained by replacing the elements of the matrix A_j by their expected values.

Thus from (13) and (14) we get

$$V = \sum_{j=1}^{2} \frac{1}{N_j} P_j^{-1} \quad \ldots (15)$$

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

Hence/

Hence $V^{-1} = \left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1} P$ on the hypothesis H_o .....(16)

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

$$\left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1}\left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right]' P^{-1}$$

$$\left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)\right]$$

$$(r=1,2,\cdots s)\cdots(17)$$

and this behaves as χ^2 with s.d.f. i.e. as $\chi^2_{[s]}$.

If this value of χ^2 is significant at the usual 5% or 10% level then our hypothesis is contradicted i.e. the samples do not belong to the same population and the parameters of the populations can be estimated from the appropriate sets of equations namely (3).

The values of p_i's and their differential coefficients involved in (17) can be obtained by estimating θ_r (r = 1, 2, ...s) from the following equations obtained on the hypothesis H_o namely/

namely

$$\sum_{i=1}^{k} \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(x_{1i}+x_{2i}\right)=0\,,\ (r=1,2,\cdots s)$$

i.e.

$$\sum_{i=1}^{k} \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(n_{1i}+n_{2i}\right)=0 \text{ for all } r \cdots (18)$$

These s equations (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 d.c.'s.

The same limitations exist as were stated previously in the two parameters case regarding the general applicability of the method.

5.04. Second Method:

As before

$$2\,L \doteqdot \text{Const.} - \sum_j \sum_i \frac{x_{ji}^2}{N_j p_{ji}} \cdots\cdots (i)$$

where the symbols have their former implications. To get the optimum estimates of θ_{1r} and θ_{2r} $(r = 1,2, \ldots s)$

i.e./

i.e. of θ_{jr} ($j = 1,2$; $r = 1,2 \ldots\ldots s$) we solve independently the two sets of s equations each namely

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

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

i.e. $\dfrac{\partial L}{\partial \theta_{jr}} = 0 \quad (j = 1,2;\ r = 1,2, \ \ldots . s)$

$$\left.\begin{aligned} &\text{i.e.}\quad \sum_i \frac{1}{p_{ji}}\left(\frac{\partial p_{ji}}{\partial \theta_{jr}}\right) x_{ji} = 0 \\ &\qquad\qquad (j = 1,2;\ r = 1,2, \ \ldots . s) \\ &\text{also}\quad \sum_i x_{ji} = 0 \quad (j = 1,2) \end{aligned}\right\} \ldots \text{(ii)}$$

Let us again put $Z_{ji} = \dfrac{x_{ji}}{\sqrt{p_{ji}}}$

$$(j = 1,2;\ i = 1,2, \ldots k) \qquad \ldots\ldots\ldots\text{(iii)}$$

then it is clear that Z_{ji} ($j = 1,2$) are independent normal variates with zero mean and variances N_j ($j = 1,2$) for all i's.

In/

In view of (iii), (ii) becomes

$$\left.\begin{aligned} &\sum_i \frac{1}{\sqrt{h_{ji}}}\left(\frac{\partial h_{ji}}{\partial \theta_{jr}}\right) z_{ji} = 0 \\ &\sum_i \sqrt{h_{ji}}\; z_{ji} = 0 \end{aligned}\right\} (j=1,2;\ r=1,2,\cdots s) \qquad \text{..(iv)}$$

and (i) becomes

$$2L \doteqdot \text{Const.} - \sum_j \sum_i \frac{z_{ji}^2}{N_j} \qquad \text{...(v)}$$

Further let

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

Evidently/

Evidently the X_{jr}'s $(r = 1,2 \ldots s)$ are each orthogonal to t_j for all j $(j = 1,2)$

The quadratic form for X_{jr} and t_j $(j = 1,2; r = 1,2,..s)$ is

$$\left[t_j \;\; X_{j1} \ldots X_{js}\right] V_j^{-1} \begin{bmatrix} t_j \\ X_{j1} \\ \vdots \\ X_{js} \end{bmatrix} \qquad (j = 1,2) \ldots \text{(vii)}$$

where V_j is the variance matrix of the variates involved.

$$\left.\begin{aligned}
\text{But Var}\,(X_{jr}) &= N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{jr}}\right)^2 \\
\text{Var}\,(t_j) &= N_j \\
\text{Cov}\,(X_{jr}, t_j) &= 0 \\
\text{Cov}\,(X_{jm}, X_{jr}) &= N_j \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{jm}}\right)\left(\frac{\partial h_{ji}}{\partial \theta_{jr}}\right)
\end{aligned}\right\}$$

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

Therefore

$$V_j = N_j \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{j1}}\right)^2 & \cdots & \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{j1}}\right)\left(\frac{\partial h_{ji}}{\partial \theta_{js}}\right) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{js}}\right)\left(\frac{\partial h_{ji}}{\partial \theta_{j1}}\right) & \cdots & \sum_i \frac{1}{h_{ji}} \left(\frac{\partial h_{ji}}{\partial \theta_{js}}\right)^2 \end{bmatrix}$$

Hence/

Hence (vii) can be written as

$$\frac{t_j^2}{N_j} + \frac{1}{N_j}\left[x_{j1} \cdots x_{js}\right] P_j^{-1} \begin{bmatrix} x_{j1} \\ \vdots \\ x_{js} \end{bmatrix}, \ (j=1,2) \ \text{---} \ (viii)$$

where P_j is the same as in §5.03 (8)

In view of the restrictions on Z_{ji}'s given by (iv), (v) can be written with the help of (viii) as

$$2L_{H_1} \doteq Const - \sum_{j=1}^{2}\sum_{i=1}^{k} \frac{z_{ji}^2}{N_j} + \sum_{j=1}^{2} \frac{t_j^2}{N_j}$$

$$+ \sum_{j=1}^{2} \frac{1}{N_j}\left[x_{j1} \cdots x_{js}\right] P_j^{-1} \begin{bmatrix} x_{j1} \\ \vdots \\ x_{js} \end{bmatrix} \cdots (ix)$$

On the hypothesis H_0 we have $\theta_{1r} = \theta_{2r} = \theta_r$ (say) $(r = 1,2, \ldots s)$

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

and/

and hence $P_1 = P_2 = P$ (say) where P is the same as in § 5.03 (11).

Also instead of $2s$ equations $\frac{\partial L}{\partial \theta_{1r}} = 0$

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

we get only s equations $\frac{\partial L}{\partial \theta_r} = 0$ $(r = 1, 2, \ldots s)$ ....(x)

in addition to two equations $t_1 = 0 = t_2$.

Equation (x) can be written as

$$\sum_i \frac{1}{p_{1i}} \left(\frac{\partial p_{1i}}{\partial \theta_{1r}} \right) x_{1i} + \sum_i \frac{1}{p_{2i}} \left(\frac{\partial p_{2i}}{\partial \theta_{2r}} \right) x_{2i} = 0$$
$$(r = 1, 2, \cdots s)$$

i.e. $X_{1r} + X_{2r} = 0$, $(r = 1, 2, \cdots s)$

Now the quadratic form for t_1, t_2 and $(X_{1r} + X_{2r})$

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

can be written as before as/

as

$$\sum_{j=1}^{2} \frac{t_j^2}{N_j} + \left[(x_{11}+x_{21}) \cdots (x_{1s}+x_{2s})\right] V^{-1} \begin{bmatrix} (x_{11}+x_{21}) \\ \vdots \\ (x_{1s}+x_{2s}) \end{bmatrix} \cdots (xi)$$

where V = variance matrix of the variates

$(x_{1r} + x_{2r})$ for all r

$= N_1 \; P_1 + N_2 \; P_2$, P's being the same as in § 5.03 (8)

$= (N_1 + N_2) \; P$ on the hypothesis H_0.

In view of (xi) we wrote (v) as follows:-

$$2L_{H_0} \doteq Const - \sum_{j=1}^{2}\sum_{i=1}^{K} \frac{z_{ji}^2}{N_j} + \sum_{j=1}^{2} \frac{t_j^2}{N_j}$$

$$+ \frac{1}{N_1+N_2}\left[(x_{11}+x_{21}) \cdots (x_{1s}+x_{2s})\right] P^{-1} \begin{bmatrix} (x_{11}+x_{21}) \\ \vdots \\ (x_{1s}+x_{2s}) \end{bmatrix} \cdots (xii)$$

From/

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

$$-2\left(L_{H_0}-L_{H_1}\right) \doteq \sum_{j=1}^{2} \frac{1}{N_j}\left[x_{j1} \cdots x_{js}\right] P^{-1}\begin{bmatrix} x_{j1} \\ \vdots \\ x_{js} \end{bmatrix}$$

$$-\frac{1}{N_1+N_2}\left[(x_{11}+x_{21}) \cdots (x_{1s}+x_{2s})\right] P^{-1}\begin{bmatrix} (x_{11}+x_{21}) \\ \vdots \\ (x_{1s}+x_{2s}) \end{bmatrix}$$

i.e.

$$-2\left(L_{H_0}-L_{H_1}\right) \doteq \left(\frac{1}{N_1}+\frac{1}{N_2}\right)^{-1}\left[\left(\frac{x_{11}}{N_1}-\frac{x_{21}}{N_2}\right) \cdots \left(\frac{x_{1s}}{N_1}-\frac{x_{2s}}{N_2}\right)\right] P^{-1}$$

$$\begin{bmatrix} \left(\frac{x_{11}}{N_1}-\frac{x_{21}}{N_2}\right) \\ \vdots \\ \left(\frac{x_{1s}}{N_1}-\frac{x_{2s}}{N_2}\right) \end{bmatrix}$$

But

$$\frac{x_{1r}}{N_1}-\frac{x_{2r}}{N_2} = \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_r}\right)\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right), \quad (r=1,2,\cdots s)$$

Therefore/

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

=/

$$Q = \left[\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_1}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \cdots \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_s}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right)\right] P^{-1} \begin{bmatrix} \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_1}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \\ \vdots \\ \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta_s}\right)\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2}\right) \end{bmatrix}$$

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

$$(r = 1, 2, \cdots s) \quad \cdots \text{(xiv)}$$

and this behaves as $\chi^2_{[s]}$ and thus this result, our test criterion, agrees with the result § 5.03 (17).

Here as before the s θ's will be determined from the same set of s equations as in § 5.03 (18) and substituted in p_i's and their differential coefficients to obtain their estimates from the two samples.

If this value of $\chi^2_{[s]}$ is significant at the usual levels of significance, our hypothesis H_0 is rejected and in this case the optimum estimates of $2s$ $\hat{\theta}$'s will be obtained from equations (ii) above.

5.05. On partitions and degrees of freedom of χ^2

Let us now analyse the degrees of freedom of χ^2 into separate components and account for the total degrees of freedom.

If the law of distribution assumed is definitely known, i.e. if all the parameters involved in the law of distribution are specified, in other words, if the hypothesis is a simple one, we can easily test whether or not the given law fits the data well. For this purpose it is merely necessary to calculate

$$\sum_i \frac{x_{1i}^2}{N_1 p_{1i}} \quad \text{and} \quad \sum_i \frac{x_{2i}^2}{N_2 p_{2i}}$$

where the symbols have the same meaning as above, subject to requisite linear constraints i.e. $\sum_i x_{1i} = 0$ and $\sum_i x_{2i} = 0$, one for each respectively.

Thus

$$\sum_i \frac{x_{1i}^2}{N_1 p_{1i}} - \frac{\left(\sum_i x_{1i}\right)^2}{N_1} \quad \text{and} \quad \sum_i \frac{x_{2i}^2}{N_2 p_{2i}} - \frac{\left(\sum_i x_{2i}\right)^2}{N_2}$$

will each behave as $\chi^2_{[k-1]}$.

Hence/

Hence for a proper fit these values must not be significant at the usual levels or any other assigned levels of significance of χ^2 with $(k-1)$ d.f.

If they are significant, the law assumed is a poor fit to the data and therefore the inferences drawn on this basis are bound to be unreliable. In general, however, the hypothesis is not a simple one, i.e. the nature of the law of distribution assumed is known except for certain unspecified parameters, which too are to be determined from the data of the two independent samples at our disposal. In other words our hypothesis is in general a composite one with various "degrees" of freedom".

In such cases the estimation of parameters from the first sample (say) puts further linear restrictions on x_{1i}'s which are already subject to one linear constraint namely $\sum_i x_{1i} = 0$.

A similar position exists for the x_{2i}'s.

In the case of one parameter we get the expression $\sum_i \frac{x_{1i}^2}{N_1 p_{1i}}$ subject to two linear restrictions, (i) $\sum_i x_{1i} = 0$.

and/

and (ii) $\sum_i \frac{p'_{1i}}{p_{1i}} x_{1i} = 0.$

Thus $\sum_i \frac{x_{1i}^2}{N_1 p_{1i}}$ can be expressed as

$\chi^2_{[k-2]}$. Similarly $\sum_i \frac{x_{2i}^2}{N_2 p_{2i}}$ can

be expressed as $\chi^2_{[k-2]}$. These values of $\chi^2_{[k-2]}$ must not be significant at the usual levels.

Assuming then that in a particular case it is so, the problem that naturally arises is whether or not they (the two independent samples) come from the same population, i.e. whether or not the two values of the parameters estimated from the two given samples are significantly different. This is tested by the expression obtained as the measure of our test criterion in the discussion of the case of one parameter and is $\chi^2_{[1]}$.

It is also clear that when the two samples emanate from the same population containing one parameter, we can express $-2L_{H_0}$ + Const. as $\chi^2_{[2k-3]}$.

This total degrees of freedom (2k-3) can be split up into (k-2) d.f. for the first sample, (k - 2) d.f. for the second sample as a measure

of/

of their goodness of fit to the law of distribution assumed and 1 d.f. for the comparison of the estimated values of the parameter from the two samples.

Thus $\chi^2_{[2k-3]} = {}_1\chi^2_{[k-2]} + {}_2\chi^2_{[k-2]} + \chi^2_{[1]}$

where ${}_1\chi^2_{[k-2]}$ means the component of $\chi^2_{[2k-3]}$ which measures the goodness of fit of the first sample to the hypothesis stipulated and ${}_2\chi^2_{[k-2]}$ is a similar measure for the second sample and $\chi^2_{[1]}$ measures the criterion whether or not the samples belong to the same population. Thus none of these components must be significant if our conclusions are to be valid.

Also

$$- 2L_{H_0} + \text{Const} \doteqdot \left\{ \sum_i \frac{x_{1i}^2}{N_1 h_i} - \frac{t_1^2}{N_1} - \frac{x_1^2}{N_1 \sum_i \frac{1}{h_i}\left(\frac{\partial h_i}{\partial \theta}\right)^2} \right\}$$

$$+ \left\{ \sum_i \frac{x_{2i}^2}{N_2 h_i} - \frac{t_2^2}{N_2} - \frac{x_2^2}{N_2 \sum_i \frac{1}{h_i}\left(\frac{\partial h_i}{\partial \theta}\right)^2} \right\}$$

$$+ \chi^2_{[1]} \quad --- \quad (A)$$

because we have shown in § 4.04(8) that

$\chi^2_{[1]} = /$

$$\chi^2_{[1]} = \sum_{j=1}^{2} \left(\frac{x_j^2}{N_j \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)^2} \right) - \frac{(x_1 + x_2)^2}{(N_1 + N_2) \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta}\right)^2} .$$

The first expression within brackets on the R.H.S. of (A) is ${}_1\chi^2_{[k-2]}$ and the second expression within bracket is ${}_2\chi^2_{[k-2]}$.

Similar partition of χ^2 into components which measure the goodness of fit as well as the criterion of the samples belonging to the same population is available even in the general case of s parameters.

In this general case we can express "
$-2L_{H_o}$ + Const as $\chi^2_{[2k-s-2]}$ as follows:-

$- 2L_{H_o}$ + Const.

$$\doteq \left\{ \sum_i \frac{x_{1i}^2}{N_1 p_i} - \frac{t_1^2}{N_1} - \frac{1}{N_1}\left[x_{11} \cdots x_{1s}\right] P^{-1} \begin{bmatrix} x_{11} \\ \vdots \\ x_{1s} \end{bmatrix} \right\}$$

$$+ \left\{ \sum_i \frac{x_{2i}^2}{N_2 p_i} - \frac{t_2^2}{N_2} - \frac{1}{N_2}\left[x_{21} \cdots x_{2s}\right] P^{-1} \begin{bmatrix} x_{21} \\ \vdots \\ x_{2s} \end{bmatrix} \right\}$$

$$+ \chi^2_{[s]} \quad \cdots \quad (B)$$

where/

where $\chi^2_{[s]}$ is the same as in § 5.03 (17).
This reduction follows from the fact that

$$\chi^2_{[s]} = \frac{1}{N_1}\left[x_{11} \cdots x_{1s}\right] P^{-1}\begin{bmatrix} x_{11} \\ \vdots \\ x_{1s}\end{bmatrix} + \frac{1}{N_2}\left[x_{21} \cdots x_{2s}\right] P^{-1}\begin{bmatrix} x_{21} \\ \vdots \\ x_{2s}\end{bmatrix}$$

$$-\frac{1}{N_1+N_2}\left[(x_{11}+x_{21}) \cdots (x_{1s}+x_{2s})\right] P^{-1}\begin{bmatrix} (x_{11}+x_{21}) \\ \vdots \\ (x_{1s}+x_{2s})\end{bmatrix}$$

As before the first expression within curled bracket on the R.H.S. of (B) measures goodness of fit of the law of distribution assumed to the first sample and is denoted by ${}_1\chi^2_{[k-s-1]}$.

Similarly the second expression within curled bracket of (B) measures goodness of fit to the second sample and is denoted by ${}_2\chi^2_{[k-s-1]}$.

Thus in the general case also we have

$$\chi^2_{[2k-s-2]} = {}_1\chi^2_{[k-s-1]} + {}_2\chi^2_{[k-s-1]} + \chi^2_{[s]} \qquad \ldots\ldots(C)$$

Of course all these χ^2 components must not be significant at the usual levels of significance if/

if the law of distribution assumed is to be valid and if the two samples belong to the same population.

Logically, therefore, we must first test the significance of ${}_1\chi^2_{[k-s-1]}$ and ${}_2\chi^2_{[k-s-1]}$ before we test $\chi^2_{[s]}$, for we must make sure that the law of distribution assumed is a reasonable fit to the data before we can say that the two given samples come from the same assumed population.

In practice we may calculate $\chi^2_{[s]}$ first and if it is significant we can assert that the two samples do not belong to the same population. If, however, it is not significant we may assume that the given samples emanate from the same population, but in this case we may further confirm it by calculating the necessary χ^2 to measure the goodness of fit of theory to data. If it seems to satisfy the criterion of good or reliable fit we may safely assert that the two given samples have come from the same population whose law of distribution is the one assumed.

It is also clear that (k-s-1) must be positive i.e. $k - s - 1 \geqslant 0$ i.e. $s \leqslant k - 1$ and this explains the restrictions imposed on the/

the number of unknown independent parameters in the law of distribution, if the samples are to be specified into the same k given categories.

If we further want to get a measure of goodness of fit of the observed data to the law of distribution from the same analysis we must have the necessary χ^2 with at least 1 d.f. i.e. in the general case

$$k-s-1 \geqslant 1 \quad \text{i.e. } s \leqslant k-2.$$

Thus if we do not want to have a criterion of goodness of fit we can take (k-1) independent unspecified parameters in the law of distribution, it being of course assumed that the samples have been classified into the same k given categories.

In the above discussion it is also assumed that N_1 and N_2 the total number of observations in the two samples are also fixed.

Hence from above considerations it is better to have at most (k-2) unknown independent parameters in the law of distribution if the number of class categories is k for then it will in addition enable us to test the goodness of fit of the data to the assumed law of distribution of the population.

5.06./

5.06. It was mentioned in §1.05 that the result of K. Pearson (1911) can be deduced as a particular case of our general result (17) in §5.03.

This we prove as follows:-

As stated in §1.10 if we regard p_i's $(i = 1,2,\ldots k)$ as our unknown parameters then the result of K. Pearson follows immediately from §5.03 (17).

For when $\theta_i = p_i$ $(i = 1,2, \ldots\ldots k)$ we get

$$\sum_i \frac{1}{p_i}\frac{\partial p_i}{\partial \theta_i}\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)=\frac{1}{p_i}\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right), \quad (i=1,2,\cdots k)$$

Also in this case

$$P = \begin{bmatrix} \frac{1}{p_1} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \frac{1}{p_2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{p_3} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \frac{1}{p_k} \end{bmatrix}$$

a diagonal matrix.

Therefore
$$P^{-1} = \begin{bmatrix} p_1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & p_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & p_3 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & p_k \end{bmatrix}$$

Thus/

Thus our criterion reduces to

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

$$\text{i.e.} = \frac{N_1 N_2}{N_1+N_2} \sum_{i=1}^{k} \frac{1}{p_i}\left(\frac{n_{1i}}{N_1}-\frac{n_{2i}}{N_2}\right)^2 \cdots (1)$$

and since $\sum_{i=1}^{k} p_i = 1$ There are only

(k-1) independent p_i's and thus (1) behaves as $\chi^2_{[k-1]}$ and the values of p_i's follow from the equations

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

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

i.e. $\frac{x_{1i}+x_{2i}}{p_i} - \frac{x_{1k}+x_{2k}}{p_k} = 0, \quad (i=1,2,\cdots k-1)$

$$\text{i.e. } p_k(n_{1i}+n_{2i}) - p_i(n_{1k}+n_{2k}) = 0, \quad (i = 1,2,\ldots k-1)$$

$$\ldots(2)$$

Adding/

Adding all such (k - 1) relations we get

$$p_K \sum_{i=1}^{K-1} (n_{1i}+n_{2i}) - (n_{1K}+n_{2K}) \sum_{i=1}^{K-1} p_i = 0$$

i.e. $$p_K \sum_{i=1}^{K} (n_{1i}+n_{2i}) - (n_{1K}+n_{2K}) \sum_{i=1}^{K} p_i = 0$$

i.e. $$p_k = \frac{n_{1K}+n_{2K}}{N_1+N_2}$$

Therefore from (2) we get $p_i = \dfrac{n_{1i}+n_{2i}}{N_1+N_2}$

(i = 1,2,...k-1)

and this is the value of p_i (i = 1,2,....k) which was suggested by K.Pearson and confirmed by E.C.Rhodes (1924) and J.Neyman and E.S. Pearson (1928) by other methods.

It is thus clear that Pearson's result is a particular case of our general result. Further Pearson's method cannot provide a criterion of goodness of fit as we take (k-1) parameters to specify the population and this we have already seen is the maximum number we can have if our population is classified into k categories.

REFERENCES

1. Dixon, W.J. (1940) - A criterion for testing the hypothesis that two samples are from the same population.
(Ann. Math. Stats.Vol 11 p.199-204).

2. Doob,J.L. (1934) - Probability and statistics
(Trans.Amer.Math.Soc.Vol.36, 759-75).

3. Fisher, R.A.(1922) - On the Mathematical foundations of Theoretical Statistics.
(Phil.Trans.Roy.Soc. Vol.222A,p.309-68).

4. Fisher, R.A.(1922a) - On the interpretation of χ^2 from contingency tables and calculation of P. (Journal R.S.S. Vol.85 p.88-90).

5. Fisher, R.A. (1923-25) - Theory of Statistical estimation.
(Proc.Camb. Phil.Soc.Vol.22 p.700-25).

6. Fisher, R.A. (1924) - The conditions under which χ^2 measures the discrepancy between observations and hypothesis.
(J.R.S.S.Vol.87 p.442-9).

7. Fisher,R.A.(1934) - Two New Properties of Mathematical likelihood.
(Proc.Roy.Soc.Vol.144A p.288-307).

8./

8. Fisher, R.A. (1938) - Statistical Theory of Estimation (Calcutta University Readership lectures).

9. Garwood, F. (1941) - The application of maximal likelihood, to dosage mortality curves. (Biom. Vol.32 p.46-58).

10. Greenwood, M and Yule G.U. (1915) - The statistics of anti-typhoid and anti-cholera inoculations and the interpretation of such statistics in general. (Proc. Roy. Soc. of Medicine Section of Epidemiology and state medicine (viii), p.113-90.)

11. Kendall, M.G. - The advanced Theory of Statistics Vol.I and Vol.II.

12. Koshal, R.S. (1933) - Application of the method of maximum likelihood in the improvement of curves fitted by the method of moments (J.R.S.S. Vol.96 p.303- 6.)

13. Neyman, J. and Pearson, E.S. (1928) - On the use and interpretation of certain criteria for the purpose of statistical inference. (Biom. Vol.20A p.283-9).

14. Neyman, J. and Pearson E.S. (1932-33) - On the problem of the most efficient tests of statistical hypotheses (Phil. Trans. Roy. Soc. Vol.231A, p.289-337).

15./

15. Pearson, K.(1900) - On the criterion that a given system of deviations from the Probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from sampling (Phil.Mag. Vol.50 p.157-75).

16. Pearson,K,(1911) - On the probability that two independent distributions of frequency are really samples from the same population. (Biom.Vol.8 p.250-2).

17. Pearson,K.(1932) - On the probability that two independent distributions of frequency are really samples from the same population. (Biom.Vol.24, p.457-9).

18. Przyborowski,J. and Wilénski,H.(1940) - Homogeneity of results in testing samples from Poisson series with an application to testing clover seeds for Dodder. (Biom. Vol. 31, p.313-23).

19. Rhodes,E.C.(1924) - On the problem whether two given samples can be supposed to have been drawn from the same population (Biom.Vol.16 p.239-45).

20. Thomson,W.R. (1938) - Biological applications of Normal range and associated significance tests in ignorance of the original distribution forms. (Ann.Math.Stats. Vol.9 p.281-7).

21./

21. Wald, A. and Wolfowitz, J. (1940) - On a test whether two independent samples are from the same population. (Ann. Math. Stats. Vol. 11 p.147-62).

22. Wald, A. - On the principles of statistical inference (Notre Dame Mathematical Lectures, number 1, 1942).

23. Wald, A. (1943) - Tests of statistical hypothesises concerning several parameters when the number of observations is large. (Trans.Amer.Math.Soc.Vol.54 p.426-82).

24. Wilks, S.S. (1938) - The large sample distribution of likelihood ratios for testing composite hypotheses (Ann.Math.Stats.Vol.9 p.60-2).