

Note on the derivation and distribution of Pearson's χ^2

By A. C. AITKEN.

(Received 28th January, 1939. Read 3rd February, 1939.)

The object of this note is to deduce from first principles the probability differential of Pearson's χ^2 , and to account for "degrees of freedom."

1. *Probability Differential of Class Deviations in a Sample.*

We consider a given aggregate or population, of metrical or qualitative or mixed character, univariate or multivariate, and we suppose that in it m mutually exclusive classes are distinguished, their probabilities being p_1, p_2, \dots, p_m , summing to 1. This aggregate is now sampled (with replacement each time if it is finite) n times. Let $n_i, i = 1, 2, \dots, m$ be the numbers of individuals found in the respective classes of the sample, and let $n_i - np_i = n\epsilon_i$. The ϵ_i are thus the deviations of relative class frequencies from expectation, and $|\epsilon_i| < 1$. Also

$$\epsilon_1 + \epsilon_2 + \dots + \epsilon_m = 0, \tag{1}$$

a relation in virtue of which only $m - 1$ of the ϵ_i are independent.

The generating function of probability of the n_i , the last class frequency n_m being suppressed (as "tails" are when we record "heads" in coin throws), is

$$(p_1 t_1 + p_2 t_2 + \dots + p_{m-1} t_{m-1} + p_m)^n. \tag{2}$$

Putting $t_i = \exp(a_i/n)$ and transferring to the means, we have the moment generating function of the ϵ_i ,

$$\exp(-\sum p_i a_i) \{\sum' p_i \exp(a_i/n) + p_m\}^n, \tag{3}$$

where Σ, Σ' indicate summation from 1 to m and from 1 to $m - 1$ respectively. Provided that none of the p_i is $O(n^{-1})$, in which case the deviations in the class in question would follow the Poisson distribution, this m. g. f. may be expanded as

$$\{1 + \frac{1}{2}n^{-2}(\sum' p_i(1 - p_i) a_i^2 - 2 \sum' p_i p_j a_i a_j + O(n^{-1}))\}^n,$$

which tends with increasing n to

$$\exp\{\frac{1}{2}n^{-1}(\sum' p_i(1 - p_i) a_i^2 - 2 \sum' p_i p_j a_i a_j)\}. \tag{4}$$

The m. g. f. thus tends towards that of a normal correlated distribution in $m - 1$ variables $\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1}$. We may write it in matrix notation $\exp \frac{1}{2} \alpha' V \alpha$. Then by the known reciprocity of quadratic forms in m. g. f. and normal probability ordinate we obtain the probability differential as $c \exp \frac{1}{2} \epsilon' V^{-1} \epsilon d\epsilon$, where ϵ is the vector $\{\epsilon_1 \epsilon_2 \dots \epsilon_{m-1}\}$, ϵ' is ϵ transposed, $d\epsilon$ stands for $d\epsilon_1 d\epsilon_2 \dots d\epsilon_{m-1}$, and

$$V = n^{-1} \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_{m-1} \\ -p_1 p_2 & p_2(1-p_2) & \dots & -p_2 p_{m-1} \\ \dots & \dots & \dots & \dots \\ -p_1 p_{m-1} & -p_2 p_{m-1} & \dots & p_{m-1}(1-p_{m-1}) \end{bmatrix}. \tag{5}$$

Now $V = n^{-1} P(I - Q)$, where

$$P = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \dots & \\ & & & p_{m-1} \end{bmatrix}, \quad Q = \begin{bmatrix} p_1 & p_2 & \dots & p_{m-1} \\ p_1 & p_2 & \dots & p_{m-1} \\ \dots & \dots & \dots & \dots \\ p_1 & p_2 & \dots & p_{m-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix} P. \tag{6}$$

By inspection, since $p_1 + p_2 + \dots + p_{m-1} = 1 - p_m$, we note that $Q^2 = (1 - p_m) Q$, and so $Q^3 = (1 - p_m)^2 Q$, and so on. Hence

$$\begin{aligned} V^{-1} &= n(I - Q)^{-1} P^{-1} \\ &= n(I + Q + Q^2 + Q^3 + \dots) P^{-1} \\ &= n\{I + Q(1 + (1 - p_m) + (1 - p_m)^2 + \dots)\} P^{-1} \\ &= n(I + p_m^{-1} Q) P^{-1} \\ &= n \begin{bmatrix} p_1^{-1} + p_m^{-1} & p_m^{-1} & \dots & p_m^{-1} \\ p_m^{-1} & p_2^{-1} + p_m^{-1} & \dots & p_m^{-1} \\ \dots & \dots & \dots & \dots \\ p_m^{-1} & p_m^{-1} & \dots & p_{m-1}^{-1} + p_m^{-1} \end{bmatrix}. \end{aligned} \tag{7}$$

Hence

$$\epsilon' V^{-1} \epsilon = \sum' n p_i^{-1} \epsilon_i^2 + n p_m^{-1} (\epsilon_1 + \epsilon_2 + \dots + \epsilon_{m-1})^2 = \sum n p_i^{-1} \epsilon_i^2, \tag{8}$$

since $(\epsilon_1 + \epsilon_2 + \dots + \epsilon_{m-1})^2 = \epsilon_m^2$. In fact $\epsilon' V^{-1} \epsilon$ is Pearson's χ^2 . In its more usual form we use absolute class deviations $\eta_i = n\epsilon_i$ and $\epsilon' V^{-1} \epsilon$ is then $\sum \eta_i^2 / n p_i$.

We must next express the probability differential of χ^2 in terms of $d\chi^2$, that is, we must transform $c \exp(-\frac{1}{2}\chi^2) d\epsilon$ from $d\epsilon$ to $d\chi^2$. We use the result that any positive definite quadratic form $x' A x = \theta^2$ of rank r can be reduced by a real congruent linear transformation $x = H y$ to a sum of r squares $y'y$. Then by the ordinary spherical

polar transformation the differential element $cf(\theta^2) dy$ becomes $c_1 f(\theta^2) \theta^{r-1} d\theta^2$, where θ is the radius. Hence, in the present case, $r = m - 1$, the probability differential is $c_1 \exp(-\frac{1}{2}\chi^2) \chi^{m-2} d\chi^2$, and c_1 is determined by the condition that the integral over 0 to ∞ is 1.

2. *Additional Linear Restrictions: Degrees of Freedom.*

The question of degrees of freedom has already arisen because of the condition (1), which reduces the number of independent variables from m to $m - 1$, causing χ^2 to be a quadratic form of rank $m - 1$. In the common practice of curve fitting further restraints may be imposed. For example if curve fitting is carried out by equating moments of the fitted curve to moments of the observed frequencies, the class deviations ϵ_i will be subjected to a number of restrictive relations of the type

$$\begin{aligned} \epsilon_1 + \epsilon_2 + \dots + \epsilon_m &= 0, \\ \lambda_1 \epsilon_1 + \lambda_2 \epsilon_2 + \dots + \lambda_m \epsilon_m &= 0, \\ \lambda_1^2 \epsilon_1 + \lambda_2^2 \epsilon_2 + \dots + \lambda_m^2 \epsilon_m &= 0, \\ \dots\dots\dots \end{aligned} \tag{9}$$

Now it would be unfair and mathematically unjustifiable to assess the probability of a χ^2 , obtained from such restricted deviations, by referring to the probability table of a presumptive χ^2 which suffers no such restrictions, or rather suffers only the first of them. In a just comparison we must assess the probability of the observed χ^2 against that of a presumptive χ^2 evaluated under the same restrictions; in other words, the probability table of χ^2 must be obtained by integrating the probability differential not over the whole field of the ϵ_i , from large domains of which the observed ϵ_i are debarred, but over the restricted field.

Let there be k such independent linear relations, that is, their matrix is of rank k . In virtue of these at least one set of k variables ϵ_i can be eliminated from χ^2 , so that χ^2 , though always preserving the unchanged value $\sum n\epsilon_i^2/p_i$, has its rank reduced to $m - k$, and is thus reducible by congruent transformation to a sum of squares of $m - k$ independent variables. The remaining non-independent ϵ_i are irrelevant, and their corresponding $d\epsilon_i$ may be integrated away. The restricted differential element is $c \exp(-\frac{1}{2}\chi^2) d\epsilon_1 d\epsilon_2 \dots d\epsilon_{m-k}$, if we take the remaining effective variables as $\epsilon_1, \epsilon_2, \dots, \epsilon_{m-k}$. This

is transformable in the same manner as before to $c_1 \exp(-\frac{1}{2}\chi^2) \chi^{m-k-1} d\chi^2$, which is the form of the probability differential of χ^2 for $m - k$ degrees of freedom.

It is necessary to insist on the fact that the restrictive relations must be linear. If they are not linear, then on elimination of the redundant variables, χ^2 is no longer a quadratic form in the remaining variables, and the spherical polar transformation is inapplicable.

REFERENCE.

Pearson, K., *Phil. Mag.* (5), 50 (1900), 157-175.

THE MATHEMATICAL INSTITUTE,
16 CHAMBERS STREET, EDINBURGH, 1.