ASSESSING THE DISTRIBUTION OF THE GNVQ GRADES

D.S.G. Pollock: January 8th 1996

The Multinomial Distribution

Imagine a circumstance in which there are k possible outcomes of a given trial or experiment which may be denoted by $A_1, \ldots, A_k$. An example of such a trial might be the examination result of a single candidate on a GNVQ course selected at random from a large population. Then the outcomes in question would be the categories of $A_1 =$ *Fail*, $A_2 =$ *Pass*, $A_3 =$ *Merit* and $A_4 =$ *Distinction*.

Imagine that there are n trials altogether, which correspond to the examination results of n candidates selected at random, and let the number of outcomes in the ith category A_i be denoted by x_i. Then $n = x_1 + \cdots + x_k$ and, if n were sufficiently large, we should expect the proportions $r_i = x_i/n; i = 1, \ldots, k$ to give a good indication of the corresponding parameters $p_1, \ldots, p_k$ of an underlying probability distribution. The parameter p_i is simply the probability that the result of any single trial will be an outcome in the the ith category; and we may denote this by writing $p_i = P(A_i)$.

Our object is to discover the probability that, from n trials, we shall obtain a specific distribution of outcomes denoted by the vector $(x_1, \ldots, x_k)$. To obtain this probability, we may begin by imagining a somewhat remarkable sequence of outcomes in which x_1 observed instances of A_1 are followed by x_2 instances of A_2 and so on down to a succession of x_k instances in the final category A_k. The probability of this event is given by

$$p = p_1^{x_1} \times p_2^{x_2} \times \cdots \times p_k^{x_k}. \tag{1}$$

A less remarkable event, which, nevertheless, has the same probability, arises when the same number of outcomes from each category are interspersed in a different, but precisely specified, sequence.

Every sequence of this nature, obtained by permuting the elements of the original sequence, represents an event of the same probability. Together, the set of all such sequences represents the variety of mutually exclusive ways of obtaining the same distribution of frequencies $(x_1, \ldots, x_k)$ over the set of categories $A_1, \ldots, A_k$. There are altogether

$$\frac{n!}{x_1!x_2!\cdots x_k!} \tag{2}$$

distinct permutations of the original sequence. Therefore the probability of obtaining the distribution vector $(x_1, \ldots, x_k)$ in any manner is just

$$(3) \qquad P(x_1, \ldots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

This function constitutes the multinominal probability distribution.

The Multinomial and the Binomial Distributions

The multinominal distribution is a generalisation of the binomial distribution

$$(4) \qquad b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

which expresses the probability of obtaining x successes in n trials. In order to assimilate the binomial distribution to the multinominal distribution, one might change the notation of equation (4) by setting $x = x_1$ and $n - x = x_2$ and by setting $p = p_1$ and $1 - p = p_2$. Then the binomial function could be written in the form of

$$(5) \qquad P(x_1, x_2) = \frac{n!}{x_1!x_2!} p_1^{x_1} p_2^{x_2}.$$

Conversely, the multinominal distribution can be reduced to a binomial distribution by dividing its categories into two. One such division is to take $A = A_1$ and $B = \{A_2, A_3, \ldots, A_k\}$. This partitioning serves to show that x_1 is a binomial variate with a probability distribution which is given by the function $b(x_1; n, p_1)$. The remaining multinominal frequencies $x_2, \ldots, x_k$ are, likewise, binomial variates.

If x_1 is any of the multinominal frequencies, then it follows, from a familiar result concerning the binomial distribution, that the random variable

$$(6) \qquad z = \frac{x_1 - np_1}{\sqrt{np_1(1-p_1)}}$$

has a distribution which tends asymptotically to the standard normal distribution $N(0, 1)$ as n increases. By the same token, the limiting distribution of z^2 is the chi-square $\chi^2(1)$ distribution of one degree of freedom.

As a preliminary step in constructing tests of certain statistical hypotheses, it is helpful to consider writing the expression for z^2 in the form of

$$(7)\qquad \begin{aligned} z^2 &= \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} \\ &= \frac{(x_1 - np_1)^2 p_2 + (\{x_2 - n\} - n\{1 - p_2\})^2 p_1}{np_1 p_2} \\ &= \frac{(x_1 - np_1)^2}{np_1(1 - p_1)}. \end{aligned}$$

The first expression on the RHS is in the form of Peason's goodness-of-fit statistic. The final expression on the RHS is the square of the expression under (6).

Detecting Deviations from a National Norm

Now imagine that there is a well-established set of proportions or probabilities $p_1, \ldots, p_k$ associated with the categories $A_1, \ldots, A_k$ which constitute the possible outcomes of a single trial. If these categories are the classes of a GNVQ assessment, then the parameters $p_1, \ldots, p_k$ may represent the relative frequencies of the classes determined empirically at the national level from a large sample, or they may represent the proportions which are deemed to be desirable from the point of view of a nationally-determined educational policy.

In investigating the performance of candidates on a regional level or at the level of individual examination centres, the object might be to determine whether the frequency distribution of a randomly selected sample of results is diverging to a significant extent from the national distribution. Let n be the number of candidates in the local sample. Then the expected value of the number of outcomes in the category A_i is given by

$$(8)\qquad \xi_i = E(x_i) = np_i.$$

In this notation, the problem is to determine whether the local frequencies $x_1, \ldots, x_k$ are compatible with the expected frequencies $\xi_1, \ldots, \xi_k$.

Consider the following quantity which comprises a weighted sum of squares of the deviations of the frequencies from their expected values:

$$(9)\qquad \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}.$$

Let it be supposed that the local population from which the sample is taken is indeed characterised by the probabilities $p_1, \ldots, p_k$. Then it can be proved that, as n increases, the distribution of this quantity will tend to that of a chi-square variate $\chi^2(k-1)$ of $k-1$ degrees of freedom. The particular instance of the result in the case where $k = 2$ is given by the asymptotic distribution of the statistic under (6) which, as we have shown, is a chi-square $\chi^2(1)$ of one degree of freedom.

The statistic of (9) provides a means of testing the hypothesis that the local frequencies are compatible with the national probabilities or proportions. If the values of the statistic is significantly in excess of its expected value, which is judged in reference to the relevant critical values of the $\chi^2(k-1)$ distribution, then the null hypothesis that the local sample conforms to the national norms is liable to be rejected.

Testing the Mutual Consistency of Local Results

In the absence of any large nation-wide sample of examination results, it is difficult to determine whether there are significant local deviations from the norm. The most that one might expect to do is to determine whether or not the results from two localities are mutually consistent or to determine whether or not the results from two subject areas are consistent.

To understand how such pairwise comparisons may be conducted, let us consider two independent multinomial distributions, the first of which has the parameters $n_1, p_{11}, p_{21}, \ldots, p_{k1}$, and the second of which has the parameters $n_2, p_{12}, p_{22}, \ldots, p_{k2}$. The observed frequencies may be denoted likewise by $x_{11}, x_{21}, \ldots, x_{k1}$ and $x_{12}, x_{22}, \ldots, x_{k2}$. For each distribution, we may consider forming a statistic in the manner of equation (9); and these statistics would be treated as if they were $\chi^2(k-1)$ variates. In view of their mutual independence, their sum, which is denoted by

$$\sum_{j=1}^{2}\sum_{i=1}^{k}\frac{(x_{ij}-n_jp_{ij})^2}{n_jp_{ij}}, \tag{10}$$

would be treated as a chi-square $\chi^2(2k-2)$ variate of $2k-2$ degrees of freedom—which is the sum of the degrees of freedom of its constituent parts. It should be recognised, however, that these variables embody the probability parameters p_{ij} about which we have no direct knowledge.

Nevertheless, consider the hypothesis that $p_{i1} = p_{i2} = p_i$ for all categories $A_i; i = 1, \ldots, k$, which is the hypothesis that the two multinomial distributions are the same. This hypothesis allows estimates of the probabilities to be obtained according to the formula

$$\hat{p}_i = \frac{x_{i1}+x_{i2}}{n_1+n_2}. \tag{11}$$

Given the constraint that $\hat{p}_1 + \cdots + \hat{p}_k = 1$, it follows that there are only $k-1$ independent parameters to be estimated which consume altogether $k-1$ degrees of freedom. When the estimates are put in place of the unknown parameters, a statistic is derived in the form of

$$\sum_{j=1}^{2}\sum_{i=1}^{k}\frac{\Big[x_{ij} - n_j\big\{(x_{i1}+x_{i2})/(n_1+n_2)\big\}\Big]^2}{n_j\big\{(x_{i1}+x_{i2})/(n_1+n_2)\big\}} \tag{12}$$

which has an approximating $\chi^2(k-1)$ distribution. The statistic serves to test the hypothesis that the two multinomial distributions are the same, and the hypothesis will be rejected if the value of the statistic exceeds a critical number.