TESTING THE HOMOGENEITY OF POISSON FREQUENCIES

By Paul G. Hoel

University of California at Los Angeles

1. Introduction. The standard procedure for testing the homogeneity of a set of k Poisson frequencies seems to be to apply the Poisson index of dispersion to those frequencies. The originators of this procedure [1] pointed out that this procedure may be regarded as a χ^2 test of goodness of fit in which the Poisson frequencies constitute observed frequencies corresponding to k cells with equal expected values. Somewhat later it was shown [2] that the corresponding likelihood ratio test was approximately equivalent to the index of dispersion test. Then the problem was approached from the viewpoint of conditional variation [3], [4]. This approach permitted exact tests to be studied in some detail for small samples. A few years later an exact test for the special case of $k = 2$ was introduced and studied [5]. In this investigation consideration was given for the first time to the efficiency of the proposed test. Tables of critical regions for the test and tables for computing the power of the test corresponding to certain alternatives were made available.

In spite of the desirable features of this last test, it still possesses certain drawbacks. First, this test, as well as the others referred to, did not consider the problem in which the rate of occurrence of a rare event is constant but for which the sampling units differ in size. For example, these methods were not designed to enable one to test whether a factory's accident rate had remained unchanged during the past month as compared with the preceding three months. Second, in order to use this test it is necessary to possess the special tables or charts of critical regions constructed for the test.

In this paper a method which does not require special tables is considered for dealing with these more general situations. In the course of the development it is shown that this method is, in a certain sense, the best method possible for testing the hypothesis of homogeneity against one sided alternatives. Since this paper is principally concerned with removing the undesirable features of the method advocated in the last mentioned paper, it is advisable to read that paper in conjunction with this one. The procedure to be followed here will be to derive a uniformly most powerful test, show that it is equivalent to a χ^2 test, and then compare it with the previously mentioned test.

2. Similar regions. In the following two sections a study will be made of the efficiency of a generalization of the critical region proposed in [5]. For this purpose let x and y represent sample frequencies from two independent Poisson distributions with means m_x and m_y. The probability of obtaining this sample is given by

$$P(x, y) = \frac{e^{-m_x} m_x^x}{x!} \cdot \frac{e^{-m_y} m_y^y}{y!}. \tag{1}$$

Following the notation and procedure given in [5], let

$$\mu = m_x + m_y, \qquad p = \frac{m_x}{m_x + m_y}, \qquad n = x + y. \tag{2}$$

Then algebraic manipulation will show that $P(x, y)$ reduces to

$$P(x, y) = \frac{e^{-\mu}\mu^n}{n!} \cdot \frac{n!}{x!(n - x)!} p^x(1 - p)^{n-x}. \tag{3}$$

The hypothesis which it is desired to test is that

$$\frac{m_y}{m_x} = r, \tag{4}$$

where r has been specified. The value of r will often be the ratio of the sizes of the two populations under consideration or the ratio of the time units of the two samples. In many situations the alternatives to (4) which are of interest will be one-sided. For example, after a factory has instituted a safety campaign, it would be of interest to see if the rate was unaffected as against the possibility of the rate having decreased; hence the alternatives to (4) would be

$$\frac{m_y}{m_x} < r. \tag{5}$$

In terms of the parameters introduced in (2), the hypothesis (4) and its alternatives (5) become

$$p = \frac{1}{1 + r} \quad \text{and} \quad p > \frac{1}{1 + r}. \tag{6}$$

Consider the probability given by (3) in much the same manner as was done in [5]. This probability depends upon two parameters, μ and p, only the latter of which is specified by the hypothesis; consequently if critical regions independent of μ are desired, it will be necessary to find similar regions [6] with respect to μ. Since x and y are discrete variables, it is not possible to find similar regions of arbitrary size; consequently it will be necessary to introduce continuous approximating functions if such regions are desired and if best critical regions are to be found. Toward this end consider the expression for $P(x, y)$ in (3). It states that the probability that x and y will take on specified values is the Poisson probability that the sample point will fall on the line $x + y = n$, multiplied by the binomial conditional probability that the point will have the specified x coordinate when the point is known to lie on this line. If p and n are not small, this binomial function could be approximated well by means of a normal function. Or, if desired, factorials could be replaced by corresponding gamma functions and the necessary normalizing factor introduced. Regardless of what continuous function is chosen, a region on each line $x + y = n$ ($n = 0, 1, 2, \cdots$) can be selected such that the conditional probability for this approximating function is α that a point on that line will lie in that region. Most natural approximating functions would become trivial for $n = 0$; therefore it may be

necessary to choose an artificial function for this case or to adopt a convention of letting the origin be the critical region for this case but accepting only 100α percent of samples for which $n = 0$ as belonging to this critical region. The totality of such α regions will constitute a critical region of size α which is independent of μ because from (3) the probability of a point lying in this critical region would now be given by

$$\sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!} \cdot \alpha = \alpha \sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!} = \alpha.$$

Thus, similar regions with respect to μ of size α can be obtained by selecting regions of size α on each line $x + y = n$.

The preceding method for obtaining similar regions is the only method for doing so if such regions are restricted to be found on the lines $x + y = n$, because if a region of size α_n were selected on each line $x + y = n$, it would be necessary that

$$\sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!} \cdot \alpha_n = \alpha$$

independent of μ. This is equivalent to requiring that

$$e^{\mu} \equiv \sum_{n=0}^{\infty} \frac{\alpha_n}{\alpha} \frac{\mu^n}{n!};$$

but since the power series for e^{μ} is unique, it follows that $\alpha_n = \alpha$.

3. Common best critical region. Among these similar regions there will exist a best critical region for testing the hypothesis $p = p_0$ against the single alternative $p = p_1$ if there exist best critical regions on each line $x + y = n$. From (6) it will be observed that this formulation is equivalent to testing the hypothesis $r = r_0$ against the single alternative $r = r_1$. The best critical region [6] on such a line, if it exists, will be that region which satisfies the inequality

$$\frac{f(x; p_0)}{f(x; p_1)} \leq k, \tag{7}$$

where f denotes the continuous function selected to approximate the binomial distribution on this line and k is a constant determined so that the probability, under the hypothesis $p = p_0$, will be α that a point on this line will lie in this region. If the normal approximating function with $m = np$ and $\sigma^2 = npq$ is used, (7) becomes

$$\sqrt{\frac{p_1 q_1}{p_0 q_0}}\, e^{\frac{1}{2}\left[\frac{(x-np_1)^2}{np_1q_1} - \frac{(x-np_0)^2}{np_0q_0}\right]} \leq k. \tag{8}$$

After completing the square in x, it will be found that this inequality reduces to

$$e^{\frac{1}{n}[1/p_1q_1 - 1/p_0q_0]\left[x - \frac{n(1/q_1 - 1/q_0)}{1/p_1q_1 - 1/p_0q_0}\right]^2} \leq c, \tag{9}$$

where c is independent of x.

If x_0 is a value of x such that

$$P[x > x_0 \mid p = p_0] = \alpha, \tag{10}$$

then (9) will hold for $x > x_0$ provided that $p_1 > p_0$. To demonstrate this fact, it is convenient to consider the three cases $p_0 + p_1 \gtrless 1$ separately. If $p_0 + p_1 > 1$,

$$\frac{1}{q_1} - \frac{1}{q_0} > 0, \qquad \frac{1}{p_1 q_1} - \frac{1}{p_0 q_0} > 0, \qquad \frac{1}{q_1} - \frac{1}{q_0} > \frac{1}{p_1 q_1} - \frac{1}{p_0 q_0},$$

and therefore $x \leq n \leq n\left(\frac{1}{q_1} - \frac{1}{q_0}\right) \Big/ \left(\frac{1}{p_1 q_1} - \frac{1}{p_0 q_0}\right)$. Since the coefficient of the brackets in (9) which involves x is positive, increasing x will reduce the left side of (9). If $p_0 + p_1 < 1$,

$$\frac{1}{p_1 q_1} - \frac{1}{p_0 q_0} < 0$$

and

$$\frac{n(1/q_1 - 1/q_0)}{1/p_1 q_1 - 1/p_0 q_0} < 0.$$

Since the coefficient is now negative, increasing x will reduce the left side of (9). Finally, if $p_0 + p_1 = 1$, (9) will reduce to

$$e\left[\frac{1}{p_1} - \frac{1}{p_0}\right]\left[x - \frac{n}{2}\right] \leq k.$$

Since $1/p_1 - 1/p_0 < 0$, increasing x will decrease the left side of this inequality. It therefore follows that the region defined by (10) is a best critical region for every alternative of the form $p_1 > p_0$ on the line $x + y = n$. The totality of such regions for $n > 0$, together with the previously mentioned convention for $n = 0$, then constitutes a common best critical region among all possible similar regions for testing the hypothesis (4) against the set of alternatives (5).

In a similar manner it will be found that if the inequality in (10) is reversed, the critical region so defined, together with the convention, will constitute a common best critical region for every alternative of the form $p_1 < p_0$. If the alternative hypotheses consist of $p \neq p_0$, there will not exist a common best critical region using these approximating functions.

The critical region proposed in [5] is that for the special hypothesis $p_0 = \frac{1}{2}$ and the set of alternatives $p \neq p_0$. It will be found that the lower half of this critical region for $P = 2\alpha$ will differ little, except for very small samples, from that given by (10) for this special case; however, it possesses the disadvantage of being numerical and therefore of requiring a special table. The critical region given by (10) does not possess this disadvantage. This fact will be demonstrated in the next section.

4. Chi-square test. Consider the problem of testing compatibility between observed and expected frequencies in two cells. Let x and y represent the ob-

served frequencies and e_x and e_y the expected frequencies in a sample of size n. If the probability that an observation will fall in the first cell is, as in (6), $p = \frac{1}{1+r}$, then

$$e_1 = np = \frac{x+y}{1+r}$$

and

$$e_2 = n(1-p) = \frac{r(x+y)}{1+r}.$$

The chi-square function for testing compatibility then reduces to

$$\chi^2 = \sum_{i=1}^{2} \frac{(o_i - e_i)^2}{e_i} = \frac{(y-rx)^2}{r(y+x)}. \tag{11}$$

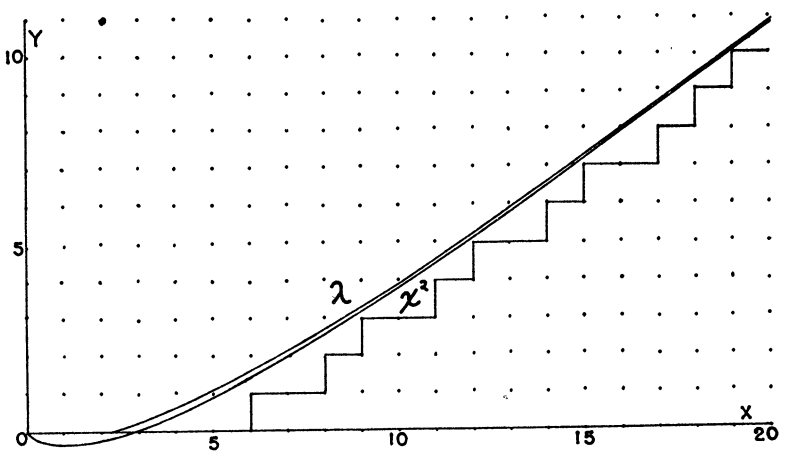

FIGURE 1.

Let χ_0^2 be the value of χ^2 such that $P[\chi^2 > \chi_0^2] = 2\alpha$ for one degree of freedom. With χ^2 replaced by χ_0^2 in (11), this equation determines a parabola in the x, y plane. If $x + y = n$ is not small, the probability of a point on the line $x + y = n$ lying outside of this parabola will be approximately 2α, the accuracy depending on the accuracy of the χ^2 approximation, and hence the probability of a point lying outside of and below this parabola will be approximately α. Thus, a critical region for testing $p = p_0$ against $p > p_0$ will be given by that part of the positive x, y plane which lies below this parabola. In Figure 1 the lower half of this parabola for the special case of $p_0 = \frac{1}{2}$ is indicated by the symbol χ^2. The critical region for the alternatives $p < p_0$ would be the region lying above the upper half of this same parabola, while the critical region for the alternatives $p \neq p_0$ would consist of both of these regions at the 2α level. For one degree of freedom, χ has a standard normal distribution; consequently the critical region given by (11) is the same as that given by (10) in which a normal approximation is used

on each line $x + y = n$. This equivalence is easily verified by replacing y by $n - x$ and r by q/p in (11).

5. Likelihood ratio test. The chi-square test of the preceding section yields a common best critical region for testing (4) against (5) for the normal approximation. It is interesting to compare this critical region with that obtained by the maximum likelihood principle, which requires no such approximations. Consider, therefore, the two dimensional parameter space

$$\Omega: \qquad m_x > 0, \qquad m_y > 0,$$

and the subspace

$$\omega: \qquad \frac{m_y}{m_x} = r.$$

Maximizing P in (1) over Ω yields $\hat{m}_x = x$ and $\hat{m}_y = y$. Maximizing P over ω, treating P as a function of m_x, yields $\hat{m}_x = x + y/1 + r$. Then the maximum likelihood ratio becomes

$$\lambda = \frac{\max P\omega}{\max P\Omega} = \frac{e^{-(x+y)}\left(\frac{x+y}{1+r}\right)^{x+y} r^y}{x!y!} \div \frac{e^{-(x+y)}x^x y^y}{x!y!}.$$

This reduces to

$$\lambda = \left(\frac{x+y}{1+r}\right)^{x+y} \cdot \frac{r^y}{x^x y^y}. \tag{12}$$

For a fixed value of λ, this equation determines a curve in the x, y plane which may be used to determine a critical region. Since $-2 \log \lambda$ is known to possess an asymptotic chi-square distribution under certain conditions [7], choose as critical region that part of the positive x, y plane lying below the curve determined by (12) when λ has been replaced by λ_0, where λ_0 is determined from $-2 \log \lambda_0 = \chi_0^2$. This curve may be plotted by reducing it to the parametric form

$$x = \frac{\log \lambda_0}{(1+v)\log\frac{1+v}{1+r} + v \log\frac{r}{v}}, \qquad y = vx.$$

A comparison of the critical regions corresponding to (11), (12), and a slight modification of [5] for the special case of $p_0 = \frac{1}{2}$ and $\alpha = .05$ is given in the accompanying sketch. The modification of [5] consists in choosing x_0 to be that integer which most nearly satisfies (10), rather than to be the smallest integer for which the left side of (10) does not exceed α. The latter method of choosing x_0 has a tendency to make the first type of error considerably smaller than α for small values of n. It will be observed that there are no appreciable differences between the maximum likelihood and chi-square critical regions. Furthermore, it will be found that there are only two values of n, namely $n = 3$ and $n = 9$, for $n \leq 30$

for which the chi-square test and the modification of [5] might yield different decisions at this significance level.

The preceding sections show that the chi-square test is highly satisfactory for testing the homogeneity of two Poisson frequencies, except possibly for very small frequencies, and that therefore special numerical tables are not necessary.

6. Several Poisson frequencies. The generalization of (11) for a set of k frequencies is, of course, the ordinary chi-square function

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}, \tag{13}$$

where $n = \sum_{i=1}^{k} x_i$, p_i is proportional to the sampling unit from which x_i was obtained, and $\sum_{i=1}^{k} p_i = 1$. The Poisson index of dispersion is merely a special case of (13) when $p_i = 1/k$. The adequacy of (13) for this special case has been studied elsewhere [3], [8], while studies of (13) in general are numerous and well known.

REFERENCES

[1] R. A. Fisher, H. G. Thornton, and W. A. Mackenzie, "The accuracy of the plating method of estimating the density of bacterial populations," *Annals of Applied Biology*, Vol. 9 (1922), pp. 325–359.

[2] P. V. Sukhatme, "The problem of k samples for Poisson population", *Proceedings of the National Institute of Sciences of India*, Vol. 3 (1937), pp. 297–305.

[3] W. G. Cochran, "The chi-square distribution for the binomial and Poisson series with small expectations", *Annals of Eugenics*, Vol. 7 (1936), pp. 207–217.

[4] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Roy. Soc. Proc.*, Series A, Vol. 160 (1937), pp 268–282.

[5] J. Przyborowski and H. Wilenski, "Homogeneity of results in testing samples from Poisson series," *Biometrika*, Vol. 31 (1939), pp. 313–323.

[6] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. Phil. Trans.*, Vol. 231 (1933), pp. 289–337.

[7] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 60–62.

[8] P. G. Hoel, "On indices of dispersion," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 155–162.