TESTING FOR HOMOGENEITY IN BINOMIAL,
~~MULTINOMIAL~~, AND POISSON DISTRIBUTIONS

R. F. Potthoff and Maurice Whittinghill

Institute of Statistics
Mimeograph Series No. 395
June 1964

UNIVERSITY OF NORTH CAROLINA

Department of Statistics

Chapel Hill, N. C.

TESTING FOR HOMOGENEITY IN BINOMIAL,
MULTINOMIAL, AND POISSON DISTRIBUTIONS

by

Richard F. Potthoff and Maurice Whittinghill

June 1964

A common problem is to test whether k samples (representing k families or k individuals, e. g.) could have come from the same binomial, multinomial, or Poisson distribution, the parameter(s) of which may be either known or unknown. This paper develops some new tests for these problems and also discusses the usual χ^2 tests; the new tests have certain optimal power properties for the case where the parameter(s) are known. A number of examples and illustrations, mainly for the binomial problem and mainly biological, are presented. For the binomial problem for the case where the parameter is known, the results of a diverse group of examples indicate that, in practical situations, the sensitivity of the new homogeneity test compares favorably with that of the usual χ^2 test, thereby corroborating the effects of the optimal theoretical power property of this new test; also, this new test is shown to be easier to compute than the χ^2 test.

This research was supported in part by National Institutes of Health Research Grant RG-9358 and in part by the Mathematics Division of the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 395

UNIVERSITY OF NORTH CAROLINA

Department of Statistics

Chapel Hill, N. C.

TESTING FOR HOMOGENEITY IN BINOMIAL,
MULTINOMIAL, AND POISSON DISTRIBUTIONS

by

Richard F. Potthoff and Maurice Whittinghill

June 1964

A common problem is to test whether k samples (representing k families or k individuals, e. g.) could have come from the same binomial, multinomial, or Poisson distribution, the parameter(s) of which may be either known or unknown. This paper develops some new tests for these problems and also discusses the usual χ^2 tests; the new tests have certain optimal power properties for the case where the parameter(s) are known. A number of examples and illustrations, mainly for the binomial problem and mainly biological, are presented. For the binomial problem for the case where the parameter is known, the results of a diverse group of examples indicate that, in practical situations, the sensitivity of the new homogeneity test compares favorably with that of the usual χ^2 test, thereby corroborating the effects of the optimal theoretical power property of this new test; also, this new test is shown to be easier to compute than the χ^2 test.

This research was supported in part by National Institutes of Health Research Grant RG-9358 and in part by the Mathematics Division of the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 395

TESTING FOR HOMOGENEITY IN BINOMIAL, MULTINOMIAL, AND POISSON DISTRIBUTIONS

by

Richard F. Potthoff* and Maurice Whittinghill**
Department of Statistics and Department of Zoology
University of North Carolina
Chapel Hill, North Carolina

SUMMARY

If we have k binomial samples of different sizes, we may sometimes be interested in the question of homogeneity, i.e., we may want to know whether the k samples all came from binomial distributions with the same parameter p . A similar question of homogeneity may arise if we have k samples from multinomial or Poisson distributions.

Homogeneity tests for these situations already exist, but these existing tests apparently were not constructed with any optimal power properties explicitly in mind. This paper approaches the problem of homogeneity testing by attempting to construct tests having maximal power against certain reasonable alternative hypotheses.

Some new tests result from this approach; these tests will be described and numerical illustrations will be presented. The traditional tests we also will discuss. For the binomial problem, all tests which are considered are applicable to the situation (frequently arising in genetics) in which some or all of the k sample sizes are small numbers (even as small as 2 or 3).

*Research supported in part by National Institutes of Health Research Grant RG-9358 and in part by the Mathematics Division of the Air Force Office of Scientific Research.

**Research supported by National Institutes of Health Research Grant RG-9358.

Part I. of the paper is concerned with testing for homogeneity for the binomial case; a number of applications are presented. In Part II. we show briefly how the new test which is introduced in Part I. may be generalized to the multinomial case. Part III. then deals with the Poisson case. All of the more technical details have been relegated to the Mathematical Appendix, which forms the last part of the paper.

I. THE BINOMIAL PROBLEM

We have k independent samples, of sizes $n_1, n_2, \dots, n_k$. In the i -th sample, we observe x_i individuals of one type and $y_i = n_i - x_i$ of the other type. If the null hypothesis of homogeneity is true, then each x_i is binomially distributed with the same parameter p , i.e., x_i has probability function

$$f_o(x_i) = \binom{n_i}{x_i} p^{x_i} q^{n_i - x_i} \quad (i = 1, 2, \dots, k) \quad , \quad (1)$$

where $q = 1 - p$. In some problems p is known or specified; in other problems p is unknown. [Note: We shall speak of homogeneity testing in both situations, even though some writers seem to use this terminology only when p is unknown.]

The alternative hypothesis, against which a homogeneity test is intended to be sensitive, is typically stated only in rough terms. In general, though, the alternative hypothesis boils down to some statement to the effect that the distribution of the x_i 's is more "spread out" than under the null hypothesis (1). Such spreading out could be the result of lack of independence (and, more specifically, positive correlation) among the n_i elements of an individual sample; it could also result from different values of p in the different samples.

In order to formulate an alternative hypothesis in more explicit terms, it would seem to be not unreasonable to work with a model in which ¹, for each i ,

a random variable p_{*i} is drawn from a beta distribution with mean p and then x_i is drawn from a binomial distribution with parameter p_{*i} . For the case where p is known, it is then possible, by a simple application of the Neyman-Pearson Lemma, to construct a test of the null hypothesis (1) which is locally most powerful against the alternative which we just specified ("locally" in the sense that we take the variance of the beta distribution to be close to zero). It appears to us that the resulting test, whose critical region is based² on large values of the statistic

$$V = (X./p) + (Y./q) \quad , \quad (2)$$

where

$$X. = \sum_i x_i(x_i-1) = \sum_i x_i^2 - \sum_i x_i \text{ and } Y. = \sum_i y_i(y_i-1) = \sum_i y_i^2 - \sum_i y_i \quad , \quad (3)$$

will often be the best homogeneity test available. Further details about how to use the test will be given shortly, along with examples of applications.

Previous approaches to the problem of testing binomial data for homogeneity, with the possible exceptions of Smith's (1951) approach and of some specialized tests against linear trends [see Apmitage(1955)], have not been deliberately designed to produce tests which have any optimal power properties or which are directed against specific alternatives. In particular, the commonly-employed χ^2 homogeneity test is at present the standard test to use, but, al-³ though reasonably powerful, it appears not to have any optimal property analogous to that possessed by our test based on V (2).

For known p , this standard χ^2 test has critical region based on large values of the statistic

$$G_1 = \sum_i \frac{(x_i - n_i p)^2}{n_i p q} = (1/pq) \sum_i (x_i^2/n_i) - (2/q)x. + (p/q)n. \quad , \quad (4)$$

where we define $x. = \sum_i x_i$ and $n. = \sum_i n_i$; for unknown p , the χ^2 homogeneity test rejects for large values of

$$G_2 = (n.^2 / x.y.) \sum_i (x_i^2 / n_i) - (n.x./y.) \quad , \quad (5)$$

where $y. = \sum_i y_i$. Under the null hypothesis of homogeneity, G_1 and G_2 are distributed approximately as χ^2 with d.f. = k and $(k-1)$ respectively. However, these approximations are not good when the n_i 's are small. Haldane [see Haldane (1937), Gruneberg and Haldane (1937), and Haldane (1940)], faced with the problem of testing for homogeneity in genetic data with small n_i 's but large k , initiated the use of a normal approximation which [according to Cochran (1952)] is valid for small n_i 's and $k > 60$, but this still did not solve the problem of what to do for small n_i 's and small k . Finally, though, Nass (1959) found a technique which seems to handle even this latter problem and thereby provides the most accurate available way of utilizing G_1 and G_2 : Nass's refinement consists in multiplying G (i.e., either G_1 or G_2) by a constant c , so chosen that (cG) has the same first two moments as a certain χ^2 distribution, and then approximating the null distribution of (cG) by the appropriate χ^2 distribution.

We return now to our new test based on V (2). Although the exact distribution of V under the null hypothesis (1) is (like that of G_1) rather complicated, we can still determine the critical region for the test via an approximation. We use a χ^2 approximation similar to Nass's (1959); but instead of equating only the first two moments as Nass did, we choose a linear function of V whose first three moments are equal to those of a χ^2 distribution. More specifically, the null distribution of the statistic

$$eV + f \quad , \quad (6)$$

where

$$e = \frac{N.}{2\gamma N. + \sum_i n_i (n_i - 1)(n_i - 2)} , \quad (7)$$

$$f = e(e-1) N. , \quad (8)$$

$$N. = \sum_i n_i (n_i - 1) = \sum_i n_i^2 - \sum_i n_i , \quad (9)$$

$$\gamma = (1/4 pq) - 1 , \quad (10)$$

is assumed to be approximately that of χ^2 with

$$v = e^2 N. \quad (11)$$

degrees of freedom. On both theoretical and empirical grounds this χ^2 approximation seems to be satisfactorily accurate ⁷. [Incidentally, it is helpful to remember that $z = (2\chi_v^2)^{\frac{1}{2}} - (2v - 1)^{\frac{1}{2}}$ may be referred to the normal curve table if v is large enough.]

For v (11) sufficiently large, one can reduce computational labor by using a less refined approximation, the normal approximation to the (null) distribution of V . Now

$$E(V) = N. \quad \text{and} \quad \sigma^2(V) = 2N. \quad (12)$$

Our desired normal approximation [which is analogous to (A6) of Note 4] is thus based on

$$z = (V - N.) / (2N.)^{\frac{1}{2}} . \quad (13)$$

Our test based on V (2) was designed specifically for the case of known p , and so far its use has been described only for that case. However, it is natural to inquire whether V might somehow be used also when p is unknown. Now if V (2) is considered as a function of p , then the value of p which minimizes (2) is

$$p = X_{.}^{\frac{1}{2}} / (X_{.}^{\frac{1}{2}} + Y_{.}^{\frac{1}{2}}) , \quad (14)$$

for which

$$V = V_{\min} = (X_{.}^{\frac{1}{2}} + Y_{.}^{\frac{1}{2}})^2 ; \quad (15)$$

in other words, V (2) cannot be less than (15) no matter what the value of the unknown p . Hence this suggests that we will have a conservative (and therefore valid) test based on V if we just use (14) for p [and (15) for V]: such a test is certainly conservative if the normal approximation (13) can be used, and one would reasonably suspect it to generally be conservative even when the χ^2 approximation is used. However, this test based on (14-15) no longer has the optimal power property [against the alternative (A 2) of Note 1] which held in the case of known p . The test loses power as a result of being conservative. This power loss is apparently more severe in smaller experiments, so that the test based on G_2 (5) may well be more powerful [against (A2)] than our test based on (14-15) when the experiment is "small". But in some larger experiments, our test based on (14-15) may have favorable power in spite of its being conservative. Where the line should be drawn, though, is not easy to say.

As our last topic before turning to practical examples, we call attention to a specialized homogeneity test which is applicable just in the special case⁸ where p is known to be $= \frac{1}{2}$. We are describing it briefly in the appendix because in rare situations it could turn out to be the most powerful test available. [For known $p (= \frac{1}{2})$, our test based on V (2) is (locally) most powerful against the alternative (A2), which we feel to be a highly appropriate model for the alternative hypothesis; but one can obviously construct other alternatives (which might or might not arise in practice) against which the G_1 test (4) or the test described in Note 8 would have better power.]

We present a number of illustrations of applications of homogeneity testing in the binomial distribution, covering the cases of both known p and unknown p :

Example 1 (crossovers in fruit flies). The extensive *Drosophila melanogaster* crossover data of Gowen (1919) provides an interesting application. His article reported great variability in the total crossover rates obtained from combined complementary crossover classes. It has been suggested that certain processes such as gonial crossing over [see, e.g., Whittinghill (1950)] may lead to agglutination (i.e., heterogeneity) also with respect to the numbers of offspring in complementary crossover classes. Therefore re-examination of Gowen's data for such heterogeneity seems natural. Table 1 presents a small part of his data [taken from Gowen (1919), Table A, between pp. 242 and 243], which we use to illustrate numerically our test based on $V(2)$.

Offspring of 37 families were classified according to several characteristics on the third chromosome. The results of this classification for the two characteristics *Dichaete* bristle (D') and curled wing (cu) are indicated in Table 1 for the two complementary crossover classes $++$ and $D' cu$ (the other two categories, $+ cu$ and $D' +$, are non-crossovers, which are not shown since they do not interest us). If classical genetic theory holds strictly, the crossover data of Table 1 will observe the binomial distribution (1); but if a process such as gonial crossing over is occurring, then the data would have greater dispersion than under (1).

TABLE 1

Numbers of the complementary crossover offspring, + + and D' cu, in each of 37 Drosophila melanogaster families [from Gowen (1919), Table A]

Family reference number	Number of + + offspring x_1	Number of D' cu offspring y_1	Total crossover offspring n_1	Family reference number	Number of + + offspring x_1	Number of D' cu offspring y_1	Total crossover offspring n_1
1348	7	4	11	1390	1	3	4
1349	1	0	1	1391	2	10	12
1350	4	2	6	1392	3	5	8
1351	3	4	7	1393	1	3	4
1352	5	3	8	1395	1	4	5
1363	3	3	6	1396	4	1	5
1364	0	2	2	1397	5	1	6
1365	11	8	19	1415	3	1	4
1366	3	1	4	1417	3	7	10
1368	0	2	2	1420	5	3	8
1369	10	5	15	1422	1	10	11
1372	3	3	6	1426	1	3	4
1382	0	6	6	1427	3	1	4
1383	4	6	10	1428	0	0	0
1385	2	6	8	1429	4	0	4
1386	2	2	4	1431	0	2	2
1387	3	2	5	1432	1	1	2
1388	5	1	6	1473	2	1	3
1389	2	4	6				
Totals					108	120	228

The data of Table 1 will illustrate the use of $V(2)$ in testing for homogeneity. We find $x. = 108$, $y. = 120$, $n. = 228$, $\Sigma x_i^2 = 542$, $\Sigma y_i^2 = 630$, $\Sigma n_i^2 = 1958$, $X. = 434$, $Y. = 510$, $N. = 1730$, and $\Sigma n_i^3 = 21474$. Now if we are willing to assume that $++$ and D^+cu are equally likely products (which is theoretically the case, if we disregard such possible complications as differential viability), then we consider that p is known to be $= \frac{1}{2}$, so that (2) becomes

$$V = (434/.5) + (510/.5) = 1888$$

From (7-11) we obtain $\gamma = 0$, $e = 1730 / [0 + (21474 - 3 \cdot 1958 + 2 \cdot 228)] = .107748$, $f = -(.107748)(.892252)(1730) = -166.32$, and $v = (.107748)^2(1730) = 20.08$. Hence we refer $eV + f = (.107748)(1888) - 166.32 = 37.11$ to the χ^2 distribution with $v = 20.08$ d.f., and find that V is significant at the level $P = .011$.

If the cruder but quicker normal approximation (13) is used for testing the significance of V , we get

$$z = (1888 - 1730) / (2 \cdot 1730)^{1/2} = +2.686$$

for which $P = .0036$.

The traditional homogeneity test based on $G_1(4)$ gives $G_1 = 4(64.775) - 4(108) + 228 = 55.10$. From (A4) and (A7) we have $E(G_1) = 36$, $\sigma^2(G_1) = 2.36 + (4 - 6)(8.327) = 55.346$, and $c = 2.36/55.346 = 1.3009$. Hence, using Nass's refinement, we refer $cG_1 = 71.68$ to the χ^2 distribution with d.f. $= cE(G_1) = 46.83$, and we find $P = .011$ (which, coincidentally, is the same P -value that we got with the V test).

Note that $V(2)$ is easier to calculate than $G_1(4)$.

Now suppose that we are not willing to assume $p = \frac{1}{2}$, but rather prefer to consider p unknown. Then (14) and (15) become $p = 20.833 / (20.833 + 22.583) = .4798$ and $V = (43.416)^2 = 1884.9$. Thus $z(13)$ cannot be less than $(1884.9 - 1730) / (2 \cdot 1730)^{1/2} = +2.633$ (for which $P = .0042$), no matter what the value of

p; however, we saw already that the normal approximation is somewhat inaccurate for this data. If we re-calculate (6 - 8, 10, 11) with $p = .4798$, we get $\gamma = .00163$, $e = .107710$, $f = -166.27$, $eV + f = 36.76$, and $v = 20.07$, so that $P = .013$.

The traditional homogeneity test when p is unknown uses $G_2(5)$, and we find $G_2 = (51984/12960)(64.775) - (24624/120) = 54.62$. From (A5) and (A7) we get $E(G_2) = 35.154$, $\sigma^2(G_2) = 54.379$, and $c = 1.2929$, so that Nass's refinement gives $cG_2 = 70.62$ and d.f. = $cE(G_2) = 45.45$. Thus $P = .010$.

The data of Table 1 was selected so as to provide a suitable detailed illustration; not all of Gowen's (1919) data evinces such heterogeneity with respect to complementary crossover classes. Our preliminary calculations suggest, though, that more such heterogeneity is present in his Table A, but that there is little if any in his Tables B, C, and D.

Example 2 (mouse genetics). An early use of homogeneity testing in dichotomous genetical data was exhibited by Grüneberg and Haldane (1937) in their examination of four experiments with mice. They tested the litters for abnormal deviations from an expected Mendelian ratio by calculating $G_1(4)$ and then applying the normal approximation according to (A4) and (A6). They assumed p to be known: the expected Mendelian ratio was always 1:1 (i.e., $p = \frac{1}{2}$) except for Experiment 3, where it was 3:1 (i.e., $p = 3/4$, $q = 1/4$). Although Experiments 2 and 4 showed no departure from homogeneity, Experiment 1 displayed definite heterogeneity and Experiment 3 was not quite significant. Grüneberg and Haldane felt that the two most likely explanations of the heterogeneity in Experiment 1 were "the presence of recessive lethal or sub-lethal genes linked with those segregating" and/or "abnormalities of meiosis leading to production of gametes in unequal numbers where equality was expected".

With their test based on (A6), Grüneberg and Haldane obtained $z = +2.83$ ($P = .002$) for Experiment 1 (all five groups combined) and $z = +1.27$ ($P = .10$)

TABLE 2

Numbers of litters with x dominant and y recessive mice [from Grüneberg and Haldane (1937), Experiment 3; ratio is 3:1]

$\begin{smallmatrix} x \\ y \end{smallmatrix}$	0	1	2	3	4	5	6	7	8	9	Total
0		11	10	17	13	9	6	8	3	0	77
1	3	9	14	14	12	18	10	3	0	1	84
2	2	1	7	7	7	9	7	1	1		42
3	1	1	8	7	9	4	3	0			33
4	1	0	0	3	0	0	0				4
5	0	0	1	0	0	1					2
6	0	0	0	1	0						1
Total	7	22	40	49	41	41	26	12	4	1	243

n	1	2	3	4	5	6	7	8	9	10	Total
Litters of size n	14	21	33	36	36	38	40	17	5	3	243

in Experiment 3. For purposes of comparison, we ran our test based on $V(2)$, using the normal approximation (13) (which is adequate for such large experiments), and found $z = +3.18$ ($P = .001$) for Experiment 1 and $z = +1.49$ ($P = .07$) for Experiment 3. (Both z -tests for Experiment 3 neglect the minor effect of selection in that experiment.)

In order to illustrate some computational short cuts for large experiments, we present the calculations for $V(2)$ for Experiment 3. We have reproduced the data, but re-arranged it to suit our computational purposes; see Table 2 (the entry for $x = 5$ and $y = 3$, e.g., tells us that there were 4 litters which had 5 dominant mice and 3 recessives). From Table 2 we get

$$X. = 40(2) + 49(6) + 41(12) + 41(20) + 26(30) + 12(42) + 4(56) + 1(72) = 3266,$$

$$Y. = 42(2) + 33(6) + 4(12) + 2(20) + 1(30) = 400,$$

and

$$N. = 21(2) + 33(6) + \dots + 3(90) = 5794.$$

Hence

$$V = (3266/.75) + (400/.25) = 5954.67,$$

so that (13) becomes

$$z = (5954.67 - 5794)/(11588)^{\frac{1}{2}} = +1.49.$$

Example 3 (diseased trees). Pielou (1963, p. 456, Table 1) tabulated the number of diseased (x) and healthy (y) trees in each of 800 "randomly placed" quadrats. This data may be tested for homogeneity to see whether the diseased trees are distributed randomly or patchily. (Pielou himself felt that the patchiness was obvious, but he verified it with a special test of his own.) We will assume that the 800 quadrats may be regarded as independent for practical purposes; also, p is known to be equal to $453/2110 = .21469$ because the quadrats were taken from a forest known to contain 453 diseased and 1657 healthy trees.

The V test based on (2) and (13) gives $z = +8.73$ when applied to Pielou's data; the G_1 test based on (4), (A4), and (A6) gives $z = +4.48$. Thus it is re-confirmed that the forest has a patchy distribution of infected trees.

Example 4 (diseased tomato plants). Along a similar line, Cochran (1954, pp. 427-429) examined data on 1440 tomato plants, each classified as diseased or healthy; the plants were divided into 160 groups of 9 (neighboring) plants each so that a homogeneity test could be run to determine if the infection was random or patchy. In this example p must be regarded as unknown (it is about .2). Using the G_2 test based essentially on (5, A5, A6), Cochran found $z = +3.96$. Our V test does not do quite as well: from (15) and (13) we obtain $z = +3.76$. {Note: In this example as in others, the P-values for both G and V will differ slightly if the more accurate χ^2 approximation [Note 5; (6-11)] is used rather than the simpler normal approximation [(A6); (13)]: in many cases a less small P results. However, in our Examples 2-6 the difference is not of importance, either in absolute terms, or with respect to comparisons between the P-values for G and V.}

Example 5 (sex ratio in humans). Are the numbers of boys and girls in human families binomially distributed, or are certain families pre-disposed to favor production of offspring of one sex or the other? This question [for lists of references, see Gini (1951) and Edwards and Fraccaro (1958)] has been examined by a number of investigators using a variety of statistical techniques. We have considered p , the proportion of boys at birth, to be known [Scheinfeld (1961, p.24) states that "year after year ... in the United States and most other countries, there is an excess of boys by almost exactly the same percentage: For every 100 girls born, there will be close to 106 boys"], and with $p = .514$ we have calculated both the V test (2, 13) and the G_1 test (4, A4, A6) for family sex distribution data from each of four investigations:

(i) Because of the way in which Geissler's (1889) extensive data on German births was collected, families of different sizes cannot be grouped together. [This data is reproduced, incidentally, in Table 1 of Harris and Gunstad (1930), p. 459.] We applied the V test and the G_1 test to all of Geissler's families of size 12, and obtained respectively $z = +8.77$ and $z = +8.75$.

(ii) Rife and Snyder's (1937) data, obtained from a city in Ohio, yields $z = +0.66$ with the V test and $z = +0.08$ with the G_1 test.

(iii) For the data of Slater (1943, p. 123, Table 2), which was taken in Britain, we find $z = +1.86$ ($P = .03$) using the V test and $z = +1.30$ ($P = .10$) using the G_1 test.

(iv) Edwards and Fraccaro (1958) presented data from Swedish ministers' families covering several centuries. We calculated $z = -0.75$ with the V test and $z = -0.05$ with the G_1 test.

[Note: If the various tests are made assuming p unknown rather than $p = .514$, then the resulting P 's are substantially the same, except in (iii), where P rises from .03 to .06 for the V test and from .10 to .13 for the G test.]

Example 6 (sex ratio in cattle). Although certain sociological factors may possibly cause heterogeneity in data on sex distribution within human families, such factors would not affect animal families. It is therefore of interest to examine Johansson's (1932) data covering thousands of births in cattle families, and test it for homogeneity. Treating p as unknown, we found $z = +1.45$ ($P = .07$) with the conservative V test (15, 13) and $z = +1.40$ ($P = .08$) with the G_2 test (5, A5, A6). [We made our calculations from the summary of Johansson's data which appears in Table 4 of Robertson (1951), p.8.]

Example 7 (number of toes on guinea pigs). Wright (1934, p. 518, Table 5)

presented data on litters of guinea pigs in which each pig was classified as four-toed or three-toed, and he easily established that the data was not homogeneous. We compared the G_2 test and our V test on his 659 litters of Beltsville stock (p is unknown, but about .3); with the refined approximation [i.e., $z = (2 \chi^2)^{\frac{1}{2}} - (2v - 1)^{\frac{1}{2}}$] we obtained respectively $z = +14.8$ and $z = +14.4$.

Conclusion. Examples 1-7 serve not only to illustrate biological applications of homogeneity testing for the binomial problem, but also provide empirical comparisons between our V test and the widely-used G tests. In the situations where p is known, the results are highly favorable to our V test (since, among the 6 cases where one or both of the tests yield $P < .10$, the V test yields the smaller P-value in 5 cases and an equal P-value in the 6th case); this reinforces the theoretical argument concerning the optimal power property of the V test. Even in the group of illustrations where p is unknown, the overall results for the V test do not compare unfavorably with those for the G_2 test.

II. THE MULTINOMIAL PROBLEM

Although we dealt at length with the more elemental binomial problem, the theoretical results of Part I. do have a straightforward generalization to the multinomial case. Let the i -th of k samples ($k \geq 1$) be of size n_i , and let there be $x_{i1}, x_{i2}, \dots, x_{im}$ individuals observed respectively in m classes

($\sum_{j=1}^m x_{ij} = n_i$). The null hypothesis [analogous to (1)] is that, for each i ,

($x_{i1}, x_{i2}, \dots, x_{im}$) follows the multinomial distribution with sample size n_i

and with respective probabilities $p_1, p_2, \dots, p_m$ for the m classes

($\sum_{j=1}^m p_j = 1$). For the alternative hypothesis of non-homogeneity, we shall work

with a model in which ¹⁰, for each i , a random vector $(p_{1*i}, p_{2*i}, \dots, p_{m*i})$ is drawn from a Dirichlet distribution with mean vector $(p_1, p_2, \dots, p_m)$, and then $(x_{i1}, x_{i2}, \dots, x_{im})$ is drawn from a multinomial distribution with the parameters $(p_{1*i}, p_{2*i}, \dots, p_{m*i})$. [The Dirichlet distribution is the multivariate generalization of the beta distribution (A1).] It turns out that, for known (specified) p_j 's, the locally most powerful test of the multinomial null hypothesis against the Dirichlet-multinomial alternative has critical region based ¹⁰ on large values of the statistic

$$V = \sum_{j=1}^m (X_{.j}/p_j) \quad , \quad (16)$$

where

$$X_{.j} = \sum_i x_{ij}(x_{ij} - 1) = \sum_i x_{ij}^2 - \sum_i x_{ij} \quad . \quad (17)$$

Note the similarity between (16-17) and (2-3).

As a generalization of (12), we obtain for V (16) the results ¹¹

$$E(V) = N. \quad \text{and} \quad \sigma^2(V) = 2(m-1)N. \quad , \quad (18)$$

where $N.$ is the same as (9). Thus, under the null hypothesis,

$$z = (V - N.) / [2(m-1)N.]^{\frac{1}{2}} \quad (19)$$

is [like (13)] approximately $N(0,1)$ for a sufficiently large experiment.

Alternatively, we may utilize a more refined approximation to the distribution of V (16), a χ^2 approximation similar to that given by (6-11). The null distribution of

$$eV + f \quad [V \text{ given by (16)}] \quad , \quad (20)$$

where

$$e = \frac{N.}{\frac{1}{2(m-1)} \left(\sum_{j=1}^m \frac{1}{p_j} - 3m+2 \right) N. + \sum_i n_i (n_i - 1)(n_i - 2)} \quad (21)$$

and

$$f = e[(m-1)e - 1]N. \quad , \quad (22)$$

is approximately that of χ^2 with

$$v = e^2(m-1)N. \quad (23)$$

degrees of freedom. The constants e and f (21-22) were determined¹¹ so that $(eV + f)$ (20) has the same first three moments as the χ^2 distribution with v (23) d.f.

For the situation where the p_j 's are unknown, (14) and (15) have a simple generalization, (24-25). The values of the p_j 's which minimize V (16) are clearly

$$p_j = X_{\cdot j}^{\frac{1}{2}} / \sum_{J=1}^m X_{\cdot J}^{\frac{1}{2}} \quad , \quad (24)$$

for which

$$V = V_{\min} = \left(\sum_{j=1}^m X_{\cdot j}^{\frac{1}{2}} \right)^2 \quad . \quad (25)$$

Thus V (16) can be no smaller than (25) no matter what the values of the unknown p_j 's.

III. THE POISSON PROBLEM

We suppose that we observe k independent random variables $x_1, x_2, \dots, x_k$, where x_i comes from a Poisson distribution with mean $b_i \lambda$ under the null hypothesis, i.e.,

$$f_0(x_i) = e^{-b_i \lambda} (b_i \lambda)^{x_i} / x_i! \quad (26)$$

The b_i 's are (given) constants: e.g., if x_i is the number of accidents incurred by the i -th of k truck-drivers, then b_i could be the number of miles he drove (and λ the average accident rate per mile); or, if x_i is the number of particles in the i -th of k samples of a liquid, then b_i could be the volume of that sample (and λ the average number of particles per unit volume). An important special case occurs when all b_i 's are 1. [In fact, it is difficult to find any Poisson data examples in the literature other than those where all b_i 's are 1. One wonders whether this is because the data really always naturally occur in this way (with all b_i 's = 1), or whether it is due to reluctance to become involved with unequal b_i 's; analyses of data on accidents of factory workers, e.g.,, seem never to take account of varying numbers of days worked (b_i 's), but it would appear to be more realistic if they did so.]

We wish to test the null hypothesis (26) against an alternative hypothesis of non-homogeneity. By non-homogeneity we mean (as before) that, roughly speaking, the x_i 's are more "spread out" than under the null hypothesis, either as a result of λ being different for different i or else as a result of some kind of non-independence of events.

First we consider the case where λ in (26) is known (or specified). One possible homogeneity test rejects the null hypothesis (26) for large values of the statistic

$$G_3 = \sum_i \frac{(x_i - b_i \lambda)^2}{b_i \lambda} = \frac{1}{\lambda} \sum_i \frac{x_i^2}{b_i} - 2 \sum_i x_i + \lambda \sum_i b_i, \quad (27)$$

which [under (26)] is distributed approximately as χ^2 with k d.f. A more refined approximation to the distribution of G_3 (27) is available via Nass's (1959) technique.

The G_3 (27) test, although probably reasonably powerful against non-homogeneous alternatives, was not specifically devised to attain some optimal power property. We will now utilize methods like those of Parts I. and II. in order to construct a different homogeneity test, one which will be locally most powerful against a selected alternative. For this alternative hypothesis, we specify a model in which ¹³, for each i , a random variable λ_{*i} is drawn from a gamma distribution with mean λ and then x_i is drawn from a Poisson distribution with mean $b_i \lambda_{*i}$. If the variance of the gamma distribution is taken to be arbitrarily close to 0, then it is not hard to show that the most powerful ¹⁴ test of (26) against the alternative (A20) has critical region based on large values of the statistic

$$U = \sum_i x_i^2 - \sum_i x_i - 2 \lambda \sum_i b_i x_i \quad (28)$$

It is easily established that

$$E(U) = -\lambda^2 \sum_i b_i^2 \quad \text{and} \quad \sigma^2(U) = 2 \lambda^2 \sum_i b_i^2 = -2E(U) \quad (29)$$

under the null hypothesis. Hence, if we use the normal approximation to the distribution of U (28), we calculate

$$z = (U + \lambda^2 \sum_i b_i^2) / (2 \lambda^2 \sum_i b_i^2)^{1/2} \quad (30)$$

But a more refined (χ^2) approximation is available if we refer

$$gU + h, \quad (31)$$

where

$$g = \sum_i b_i^2 / \left(\frac{1}{2} \sum_i b_i^2 + \lambda \sum_i b_i^3 \right) \quad (32)$$

and

$$h = g(g+1) \lambda^2 \sum_i b_i^2, \quad (33)$$

to the χ^2 distribution with

$$v = g^2 \lambda^2 \sum_i b_i^2 \quad (34)$$

degrees of freedom. The constants g and h (32-33) were chosen¹⁵ so that $(gU + h)$ (31) has the same first three moments as the χ^2 distribution with v (34) d.f. [For the special but important case where all b_i 's are 1, we might point out that formulas (32, 34, 33) reduce respectively to

$$g = \frac{1}{\lambda + \frac{1}{2}}, \quad v = \left(\frac{\lambda}{\lambda + \frac{1}{2}} \right)^2 k, \quad h = \left(\lambda + \frac{3}{2} \right) v.]$$

Although in practice one frequently encounters cases both of known p and unknown p in homogeneity testing for the binomial problem (Part I.), a similar situation does not seem to prevail for the Poisson problem: in the practical illustrations in the literature, the case of known λ (just discussed) seems to be far less frequent than the case where λ is unknown. It is to this latter case that we now turn our attention.

In testing for homogeneity, it might occur to us to try to utilize U (28) somehow even when λ is unknown. The value of λ which minimizes z (30) is

$$\lambda = \left[\left(\sum_i x_i^2 - \sum_i x_i \right) / \sum_i b_i^2 \right]^{\frac{1}{2}}, \quad (35)$$

for which

$$z = z_{\min} = \left[2 \left(\sum_i x_i^2 - \sum_i x_i \right) \right]^{\frac{1}{2}} - \frac{\sum_i b_i x_i}{\left(\frac{1}{2} \sum_i b_i^2 \right)^{\frac{1}{2}}}; \quad (36)$$

in other words, z (30) cannot be smaller than (36), no matter what the value of the unknown λ . Hence, if k and the $(b_i \lambda)$'s are large enough so that the normal

approximation can be used, then a test based on (36) will clearly be conservative (but at the same time will no longer have the optimal power property which held in the case of known λ). If the normal approximation cannot be used, then one might be inclined to calculate a test based on (28, 31-34) with λ as given by (35); such a test, one would suspect, would generally be conservative.

A second type of Poisson homogeneity test for the case of unknown λ [see, e.g., Rao (1952), pp. 205-206] has critical region based on large values of the statistic

$$G_h = \sum_i \frac{[x_i - (b_i / \sum_I b_I) \sum_I x_I]^2}{(b_i / \sum_I b_I) \sum_I x_I} = \left(\sum_i \frac{x_i^2}{b_i} \right) \left(\frac{\sum_I b_I}{\sum_I x_I} \right) - \sum_i x_i \quad (37)$$

Under the null hypothesis, $G_h(37)$ is distributed approximately as χ^2 with d.f. $k-1$; the approximation can be improved (if necessary) via Nass's refinement. Now (37) reduces to

$$G_h = \sum_i \frac{(x_i - \bar{x})^2}{\bar{x}} = \frac{\sum_i x_i^2}{\bar{x}} - \sum_i x_i \quad (38),$$

when all b_i 's = 1 [we are using $\bar{x}$ to denote $\sum_i x_i / k$]. Sometimes called the "variance test", the test based on (38) is an old and well-known one which was introduced by R. A. Fisher [see Fisher, Thornton, and Mackenzie (1922), Section 5]. Fisher (1950) compares the G_h (38) ^{test} with two other tests, but these latter two are goodness-of-fit tests which do not seem to be directed so specifically against alternatives of the type that we are here considering.

Now it turns out, interestingly enough, that the commonly-used variance test based on (38) possesses a certain optimal power property against the negative binomial alternative (A20): for the case where all b_i 's are 1 and λ is unknown, we prove in the appendix ¹⁷ that, among all (locally) unbiased tests of

the hypothesis (26) against the alternative (A20), the test based on (38) is locally most powerful with respect to σ^2 (i.e., most powerful for σ^2 arbitrarily close to 0), irrespective of the value of λ . Thus this property furnishes a cogent reason for preferring the G_4 (38) variance test to any other test, when all b_i 's are 1.

When the b_i 's are not all equal, though, the picture is less clear: in some cases, the G_4 (37) test might not be preferable to the test which is based on U (28) and which uses the value (35) for the unknown λ . Examples 8 and 9 below may shed a scant amount of light on this question.

We may now summarize our results for the Poisson problem as follows:

- (a) If λ is known (and regardless of whether the b_i 's are equal or unequal), then the test based on U (28) possesses an optimal power property against the alternative (A20).
- (b) If λ is unknown and all b_i 's are 1, then the variance test based on G_4 (38) possesses an optimal power property against the alternative (A20).
- (c) For the case where λ is unknown and the b_i 's are not all equal, neither the U (28) test [with (35) used for λ] nor the G_4 (37) test has been shown to possess any optimal power property (nor has a third possible test which is described in the appendix¹⁸).

We present a couple of ~~ex~~amples, the first of which illustrates numerically the computation of the U (28) test:

Example 8 (printing errors). In order to have an example with unequal b_i 's, we are using some data (see Table 3) on the numbers of misprints in newspaper articles of (necessarily) varying lengths (b_i 's). The number (x_i) of certain types of printing errors was counted for each of $k = 112$ articles which

comprised most (but not all) of the articles that appeared in a recent Saturday issue of a large North Carolina daily newspaper. The length (b_i) of each article was measured by counting the number of lines in the article. For each of the 112 articles, Table 3 indicates the section and page of the newspaper on which the article started, and then shows the number of errors (x_i) followed by the length of the article (b_i).

TABLE 3

- 24

Number of printing errors (x_i) and number of lines (b_i) in each of 112 newspaper articles.

Page	Errors x_i	Length b_i	Page	Errors x_i	Length b_i	Page	Errors x_i	Length b_i
A1	0	48	A3	0	95	B1	0	15
A1	0	70	A3	2	37	B1	0	20
A1	3	149	A3	0	23	B1	0	58
A1	0	48	A3	0	17	B1	0	28
A1	0	80	A5	0	53	B1	0	47
A1	3	150	A5	0	24	B1	1	63
A1	0	61	A5	0	10	B1	0	160
A1	1	32	A5	2	93	B1	0	49
A1	1	14	A5	1	83	B2	0	35
A1	1	96	A5	1	48	B2	0	96
A1	1	79	A5	1	73	B2	1	74
A1	1	42	A5	0	50	B2	0	119
A2	0	69	A5	0	18	B2	1	111
A2	0	44	A5	0	23	B3	0	14
A2	1	77	A6	1	52	B3	1	34
A2	0	49	A6	2	53	B3	1	19
A2	0	13	A6	2	124	B3	1	20
A2	4	101	A6	0	23	B3	0	38
A2	1	65	A6	1	35	B3	0	36
A2	0	42	A6	0	19	B3	1	28
A2	6	97	A6	0	16	B3	0	39
A2	0	10	A7	0	26	B3	0	20
A2	0	50	A8	2	82	B4	2	103
A2	0	56	A8	1	31	B4	0	43
A2	0	27	A8	0	18	B4	0	33
A2	2	61	A8	0	11	B4	0	35
A2	1	56	A8	0	18	B4	0	17
A2	1	33	A8	0	10	B4	0	13
A3	1	110	A8	0	20	B4	0	35
A3	0	29	A8	0	22	B4	1	21
A3	0	67	A8	0	11	B4	0	21
A3	0	56	A8	2	72	B4	3	162
A3	0	38	A8	0	41	B4	1	22
A3	1	18	B1	0	42	B5	1	69
A3	0	10	B1	0	42	B5	0	13
A3	1	16	B1	0	31	B5	1	35
A3	0	13	B1	0	87			
A3	1	34	B1	0	41			
						Totals	65	5429

Under the null hypothesis, the expected error rate per line is equal to λ homogeneously throughout the newspaper, and x_i is taken to follow a Poisson distribution with mean $b_i \lambda$. Under the alternative hypothesis, the expected error rate is different for different articles; this could result, e.g., from heterogeneous conditions of typesetting or proof-reading.

From Table 3 we find $\sum_i x_i = 65$, $\sum_i b_i = 5429$, $\sum_i x_i^2 = 141$, $\sum_i b_i^2 = 396441$, $\sum_i b_i x_i = 5109$, and $\sum_i b_i^3 = 38188685$. If λ were considered to be known (which could be the case if extensive studies of printing errors had been made), then we would use the known value of λ with either (28, 31-34) or (28, 30). As it is, with λ not known, we calculate (36) to get a conservative test (i.e., conservative assuming that the normal approximation is adequate), and we obtain

$$z_{\min} = [2(76)]^{\frac{1}{2}} - \frac{5109}{(198220.5)^{\frac{1}{2}}} = +.854,$$

for which $P = .20$. On the other hand, if we calculate (28, 31-34) using (35) for λ , then we first get $\lambda = [76/396441]^{\frac{1}{2}} = .0138458$ from (35), after which we compute $U = 76 - 2(.0138458)(5109) = -65.476$ from (28), $g = 396441/[198220.5 + (.0138458)(38188685)] = .545331$ from (32), $h = (.545331)(1.545331)(76) = 64.05$ from (33) [and (35)], $v = (.545331)^2(76) = 22.60$ from (34) [and (35)], and finally $gU + h = (.545331)(-65.476) + 64.05 = 28.34$ from (31). Referring 28.34 to the χ^2 distribution with $v = 22.60$ d.f., we find $P = .19$ (which thus turns out to be in close agreement with the figure $P = .20$ given above).

We may consider also the G_k (37) test. Table 3 gives us $\sum_i (x_i^2/b_i) = 2.02469$ and $\sum_i (1/b_i) = 3.751955$. From (37) we then obtain $G_k = (2.02469)(5429/65) - 65 = 104.11$, which we refer to the χ^2 distribution with $k - 1 = 111$ d.f., finding $P = .67$. If we use Nass's refinement, (A25) becomes $E(G_k) = 111$ and $\sigma^2(G_k) = 2(111) + (1/65) [(5429)(3.751955) - 111^2 - 2(112)*2] = 338.975$, so that c (A7) is $2(111)/338.975 = .65492$; referring

$c G_4 = (.65492)(104.11) = 68.18$ to the χ^2 distribution with d.f. = $c(k-1)$
 $= (.65492)(111) = 72.70$, we find $P = .63$.

Thus, in this particular example, the test based on U (28 ff.) comes much closer to rejecting the null hypothesis of homogeneity than does the G_4 (37) test ($P = .19$ versus $P = .63$); and it appears that the null hypothesis really is false (i.e., there really is non-homogeneity), since, if we run a homogeneity test just on the total figures of Sections A and B, then we are able to reject at about the 5% level [specifically, we note from Table 3 that Section A (pages A1-A8) has 49 errors in 3441 lines and Section B (pages B1-B5) has 16 errors in 1988 lines, and we then find $P = .054$ upon applying (37) to these figures].

Example 9 (raisins). A box of each of three different brands of raisin bran breakfast cereal was poured out successively into portions of varying volumes (b_i 's), and the number of raisins (x_i) in each portion was counted. If raisins are distributed homogeneously throughout a box, then we would expect each x_i to follow a Poisson distribution with mean $b_i \lambda$, where λ (which we treat as unknown) is the mean number of raisins per unit volume. In this example (unlike Example 8), the use of differing b_i 's (which was effected by using measuring cups of different sizes) was perhaps somewhat artificial, in that it might have been more natural to make all portions the same volume.

For the three experiments (boxes), the number of raisins ($\sum_1 x_i$) ranged from 208 to 286, and the number of portions (k) from 17 to 24; the coefficients of variation of the b_i 's were smaller than in Example 8. Brand A showed no departure from homogeneity using either the U test or the G_4 test (P was $> .5$ with both tests). For Brand B, both tests gave highly significant P -values (much less than .0001), apparently as a result of a concentration of raisins at the bottom of the box. The results for Brand C were the most interesting, and

avored the G_4 test; the U (28-35) test yielded $P = .027$, while the G_4 test (37, A25, A7) yielded $P = .004$.

Conclusion. Here in Part III. we have made no attempt to present a large group of practical examples as we did in Part I.; in fact, the many Poisson data examples available in the literature apply to little else besides the case where λ is unknown and all b_i 's = 1. The two examples we just presented both pertain to the case of unknown λ and unequal b_i 's, which seems to be the case where there would be the greatest doubt as to what test to use. Of course, very little can really be concluded just from these two examples, particularly since Example 8 shows the U (28 ff.) test doing better than the G_4 (37) test while Example 9 shows the reverse. However, a tentative hypothesis might be suggested: the closer the b_i 's are to being all equal (roughly speaking), the more favorably we might expect the G_4 test to compare with the U test [this would be not inconsistent either with the results of Examples 8 and 9 (since the coefficient of variation of the b_i 's is .715 in the former against only .308 in the latter for Brand C) or with theoretical considerations]. Another thing we might suspect is that, the smaller the sampling variability of the statistic on the right-hand side of (35), the better the U test would be. Finally, for this case of unknown λ and unequal b_i 's, we should remark that the possibility of using the third test, based on V (A31), should not be overlooked .

MATHEMATICAL APPENDIX

Note 1

With this model for the alternative hypothesis, p_{*i} follows a beta distribution [see, e.g., Cramér (1946), pp. 243-244] which may be written in the form

$$\phi(p_{*i}) = \frac{\Gamma\left(\frac{pq}{\sigma^2} - 1\right)}{\Gamma\left(\frac{pq}{\sigma^2} - p\right)\Gamma\left(\frac{pq}{\sigma^2} - q\right)} p_{*i}^{\frac{pq}{\sigma^2} - p - 1} (1 - p_{*i})^{\frac{pq}{\sigma^2} - q - 1}, \quad (A1)$$

for which $E(p_{*i}) = p$ and $\text{var}(p_{*i}) = \sigma^2$. The probability function for x_i under the alternative hypothesis is then

$$\begin{aligned} f_1(x_i) &= \int_0^1 \phi(p_{*i}) \binom{n_i}{x_i} p_{*i}^{x_i} (1 - p_{*i})^{n_i - x_i} dp_{*i} \\ &= \frac{\Gamma(a)}{\Gamma(pa)\Gamma(qa)} \binom{n_i}{x_i} \frac{\Gamma(pa + x_i)\Gamma(qa + n_i - x_i)}{\Gamma(a + n_i)}, \end{aligned} \quad (A2)$$

where we define $a = (pq/\sigma^2) - 1$.

The choice of a specific model for the alternative hypothesis is necessarily somewhat arbitrary. However, the use of the beta distribution (A1) provides a model in which the p_{*i} 's have the proper range (from 0 to 1), and for which all of the theoretical manipulations are mathematically tractable. Other authors have also suggested models in which a binomial parameter p is drawn from a beta distribution [see, e.g., Skellam (1948); Lindley (1957); Horsnell (1957), p. 155; Gurland (1958), p. 231; Darwin (1960); and Raiffa and Schlaifer (1961), Chapter 3]. The fact that the alternative hypothesis of the form (A2) leads to a test which seems to be discriminating well in a variety of practical situations (see examples of applications in Part. I.) lends support to the belief that the use of the beta distribution (A1) was a wise choice to begin with.

Note 2

If σ^2 is arbitrarily close to 0 and p is known or specified, then, according to the fundamental lemma of Neyman and Pearson, the most powerful test of the simple hypothesis (1) against the simple alternative (A2) has critical region based on large values of

$$\begin{aligned} \frac{\prod_{i=1}^k f_1(x_i)}{\prod_{i=1}^k f_0(x_i)} &= \prod_{i=1}^k \frac{a^{n_i} \Gamma(a)}{\Gamma(a+n_i)} \frac{\Gamma(pa+x_i)}{(pa)^{x_i} \Gamma(pa)} \frac{\Gamma(qa+y_i)}{(qa)^{y_i} \Gamma(qa)} \\ &= \prod_{i=1}^k C_i \left(1 + \frac{1}{pa}\right) \left(1 + \frac{2}{pa}\right) \dots \left(1 + \frac{x_i-1}{pa}\right) \left(1 + \frac{1}{qa}\right) \dots \left(1 + \frac{y_i-1}{qa}\right) \\ &= \prod_{i=1}^k C_i \left[1 + \left\{ \frac{\binom{x_i}{2}}{p} + \frac{\binom{y_i}{2}}{q} \right\} a^{-1} + O(\sigma^4) \right] \\ &= C \left[1 + \frac{1}{2} \left\{ \frac{X_i}{p} + \frac{Y_i}{q} \right\} \frac{\sigma^2}{pq - \sigma^2} + O(\sigma^4) \right] \quad (A3) \end{aligned}$$

i.e., on large values of $V(2)$. [C and the C_i 's in (A3) are constants free of the x_i 's.]

Note 3

In order to obtain an intuitive feeling as to how the tests based on $V(2)$ and $G_1(4)$ compare, it is instructive to consider simple cases. For instance, suppose $p = \frac{1}{8}$ (known) and $n_1 = 2$. Then $x_1 = 1$ offers stronger evidence of heterogeneity than $x_1 = 0$ if we use (4), whereas $x_1 = 0$ provides stronger evidence than $x_1 = 1$ according to (2). The latter would appear more sensible.

$V(2)$ also appears to give relatively more weight to larger n_i 's than does $G_1(4)$. Note in particular that all samples with $n_1 = 1$ are dis-

regarded completely with V (as they should be) but not with G_1 .

For the special case where $p = \frac{1}{2}$ and all n_i 's are the same, the test based on $V(2)$ is equivalent to the test based on $G_1(4)$.

Observe that, if $k = 1$, then the test based on $V(2)$ is a two-tailed test for the binomial parameter p which is different from the usual two-tailed test of p .

Incidentally, the correlation coefficient (under the null hypothesis) between $V(2)$ and $G_1(4)$ is

$$\rho(V, G_1) = 2(n. - k) / [\sigma(V) \sigma(G_1)] ,$$

where $\sigma^2(V)$ and $\sigma^2(G_1)$ are given by (12) and (A4b) respectively.

Note 4

The mean and variance of $G_1(4)$ and $G_2(5)$ (when the null hypothesis is true) were obtained by Haldane and may be written [see, e.g., Nass (1959), formulas (7-9)] as

$$E(G_1) = k , \quad (A4a)$$

$$\sigma^2(G_1) = 2k + \left(\frac{1}{pq} - 6\right) \sum_i \frac{1}{n_i} , \quad (A4b)$$

$$E(G_2) = (k-1) n. / (n.-1) , \quad (A5a)$$

and

$$\sigma^2(G_2) = \frac{2n.}{n.-3} (\rho - \pi)(\mu - \tau) + \frac{n.^2}{n.-1} \pi \tau , \quad (A5b)$$

where

$$\rho = (k-1)(n.-k)/(n.-1), \quad \pi = (n. \sum_i \frac{1}{n_i} - k^2)/(n.-2) ,$$

$$\mu = (n. - 2)/(n. - 1) , \quad \tau = \left(\frac{n.^2}{x. y.} - 4\right)/(n. - 2)$$

[note that $\sigma^2(G_2)$ (A5b) is a conditional variance, conditional on x .]

Under the null hypothesis ,

$$z = [G - E(G)] / \sigma(G) \quad (A6)$$

is assumed to be approximately $N(0, 1)$ for sufficiently large k .

Note 5

Nass sets

$$c = 2 E(G) / \sigma^2(G) \quad , \quad (A7)$$

where $E(G)$ and $\sigma^2(G)$ are obtained from (A4) or (A5). Then, under the null hypothesis, the statistic (cG) is assumed to follow approximately a χ^2 distribution with d.f. = $c E(G) = 2[E(G)]^2 / \sigma^2(G)$ (the degrees of freedom will of course not usually be a whole number). Note that c (A7) is chosen so that $\text{var}(cG) = 2E(cG)$.

Note 6

The mean and variance of $V(2)$ are given by (12). [Incidentally, both $E(V)$ and $\sigma^2(V)$ are free of p , conveniently enough.] For use in proving formulas (12), we define $X_i = x_i(x_i-1)$, $Y_i = y_i(y_i-1)$, $N_i = n_i(n_i-1)$, and we denote $(m)_r = m(m-1) \dots (m-r+1)$. Then, under the null hypothesis (1),

$$\begin{aligned} E(V) &= E\left[\sum_i \left(\frac{X_i}{p} + \frac{Y_i}{q} \right) \right] = \sum_i E\left[\frac{(x_i)_2}{p} + \frac{(y_i)_2}{q} \right] \\ &= \sum_i \left[\frac{(n_i)_2 p^2}{p} + \frac{(n_i)_2 q^2}{q} \right] = \sum_i N_i = N. \end{aligned}$$

and

$$\begin{aligned}
 \sigma^2(V) &= \sum_i \text{var}\left(\frac{X_i}{p} + \frac{Y_i}{q}\right) = \sum_i \left[E\left(\left(\frac{(x_i)_2}{p} + \frac{(y_i)_2}{q}\right)^2\right) - N_i^2 \right] \\
 &= \sum_i \left[E\left(\frac{(x_i)_4 + 4(x_i)_3 + 2(x_i)_2}{p^2} + \frac{2(x_i)_2(y_i)_2}{pq} \right. \right. \\
 &\quad \left. \left. + \frac{(y_i)_4 + 4(y_i)_3 + 2(y_i)_2}{q^2} \right) - N_i^2 \right] \\
 &= \sum_i \left[\frac{(n_i)_4 p^4 + 4(n_i)_3 p^3 + 2(n_i)_2 p^2}{p^2} + \frac{2(n_i)_4 p^2 q^2}{pq} \right. \\
 &\quad \left. + \frac{(n_i)_4 q^4 + 4(n_i)_3 q^3 + 2(n_i)_2 q^2}{q^2} - (n_i)_4 - 4(n_i)_3 - 2(n_i)_2 \right] \\
 &= 2 \sum_i (n_i)_2 = 2N.
 \end{aligned}$$

Using similar techniques, we obtain

$$\begin{aligned}
 E[(V - N.)^3] &= \sum_i E\left[\left(\frac{X_i}{p} + \frac{Y_i}{q} - N_i\right)^3\right] \\
 &= 16\gamma N. + 8 \sum_i (n_i)_3 \quad (A8)
 \end{aligned}$$

for the third central moment of V . [In (A8) we have omitted the detailed algebra.]

Now the mean, variance, and third central moment of a χ^2 distribution with ν degrees of freedom [see, e.g., Cramér (1946), p. 234] are respectively ν , 2ν , and 8ν . It is then easily verified [from (12) and (A8)] that our statistic $(eV + f)$ [see (6-10)] also has mean ν [as given by (11)], variance 2ν , and third central moment 8ν (and there is no other linear function of V for which such a property holds).

Incidentally, we should forewarn users that the statistic $(eV + f)$ (6), unlike χ^2 itself, may in rare instances assume negative values.

We remark finally that the asymptotic normality of V [see (13)] is easily established via the Liapounoff version of the central limit theorem.

Note 7

It is natural to try to devise some sort of χ^2 approximation for the distribution of V , since $V(2)$ is the sum of quadratic functions of the x_i 's. A rather severe theoretical criterion for judging the accuracy of our approximation is to compare the fourth moment of (6) with the fourth moment of its approximating χ^2 . We can obtain the result

$$\begin{aligned} E[(V-N.)^4] &= \sum_i E\left[\left(\frac{X_i}{p} + \frac{Y_i}{q} - N_i\right)^4\right] + 6 \sum_{i < I} (2N_i)(2N_I) \\ &= 48 \sum_i (n_i)_4 + 576\gamma \sum_i (n_i)_3 + (128\gamma^2 + 64\gamma - 16) \sum_i (n_i)_2 + 12N^2 \end{aligned}$$

after some lengthy but straightforward calculations. Hence the β_2 -coefficient (fourth central moment divided by the square of the variance) for $(eV + f)$ (6) is

$$\beta_2(eV+f) = 3 + (4/N^2) \left[3 \sum_i (n_i)_4 + 36\gamma \sum_i (n_i)_3 + (8\gamma^2 + 4\gamma - 1) \sum_i (n_i)_2 \right]. \quad (A9)$$

For the χ^2 distribution with d.f. = ν (11), we have

$$\beta_2(\chi_\nu^2) = 3 + \frac{12}{\nu} = 3 + (4/N^2) \left[3 \frac{\left\{ \sum_i (n_i)_3 \right\}^2}{\sum_i (n_i)_2} + 12\gamma \sum_i (n_i)_3 + 12\gamma^2 \sum_i (n_i)_2 \right]. \quad (A10)$$

It looks as though (A9) and (A10) will ordinarily be close to each other.

(A10) can never be < 3 , and (A9) can be < 3 only in an unusual case with very small n_i 's and small γ . The leading term inside the square brackets (which is the most important term except when n_i 's are very small or γ is large) is ordinarily greater in (A10) than in (A9), but the second term is at least as great in (A9) as in (A10); hence these first two terms tend to counter-

balance each other. It may be instructive to compare (A9) and (A10) for some simple specific examples, in which all $k n_i$'s are equal: if $p = \frac{1}{2}$ and $n_i = 3$, then (A9) is $3 - (2/3k)$ and (A10) is $3 + (2/k)$; if $p = \frac{1}{2}$ and $n_i = 4$, then (A9) is $3 + (5/3k)$ and (A10) is $3 + (4/k)$; if $p = \frac{1}{4}$ and $n_i = 3$, then (A9) is $3 + (238/27k)$ and (A10) is $3 + (150/27k)$; if $p = 1/3$ and $n_i = 12$, then (A9) is $3 + (1678/176k)$ and (A10) is $3 + (1681/176k)$; if $p = 1/6$ and $n_i = 25$, then (A9) is $3 + (14.5848/k)$ and (A10) is $3 + (12.1032/k)$.

We can also assess the accuracy of the approximation on a more empirical basis, by comparing the exact and approximating distributions of (6) for a special case. The results of such a comparison for the case where $p = \frac{1}{4}$ and $n_1 = 2$, $n_2 = 2$, $n_3 = 4$, $n_4 = 5$ are presented in Table 4. We necessarily had to choose a case with small n_i 's and small k , in order for the calculations to be manageable; but, since the approximation is so good (note particularly the extreme right-hand tail) even for such a severe case, this seems to assure sufficient accuracy for more typical cases. The first column in Table 4 lists all possible values of $(e V + f)$ [where e (7) is $1/3$ and f (8) is -8]; the second column gives the exact probability of obtaining a value of $(e V + f)$ which is $\geq$ that shown in the first column; and the third column gives the approximate probability of the same event, figured according to a χ^2 distribution with $\nu = 4$ d.f. (11) and using a continuity correction of half the distance to the previous value.

T A B L E 4Exact and approximate distributions of $(e V + f)$ for the case where $p = \frac{1}{4}$ and $n_1 = 2, n_2 = 2, n_3 = 4, n_4 = 5$

$(e V + f)$	Exact distribution	Approximate distribution	$(e V + f)$	Exact distribution	Approximate distribution
0	1.000	1.000	168/9	$.0^2_{17}$	$.0^2_{11}$
8/9	.961	.979	176/9	$.0^2_{12}$	$.0^3_{75}$
16/9	.824	.856	184/9	$.0^2_{11}$	$.0^3_{50}$
24/9	.677	.695	192/9	$.0^2_{10}$	$.0^3_{33}$
32/9	.591	.540	200/9	$.0^3_{98}$	$.0^3_{22}$
40/9	.458	.406	208/9	$.0^3_{77}$	$.0^3_{15}$
48/9	.315	.299	216/9	$.0^3_{56}$	$.0^4_{98}$
56/9	.227	.216	224/9	$.0^3_{42}$	$.0^4_{65}$
64/9	.156	.155	232/9	$.0^3_{23}$	$.0^4_{43}$
72/9	.090	.109	240/9	$.0^4_{96}$	$.0^4_{29}$
80/9	.050	.077	248/9	$.0^4_{71}$	$.0^4_{19}$
88/9	.041	.053	256/9	$.0^4_{25}$	$.0^4_{12}$
96/9	.027	.037	264/9	$.0^4_{11}$	$.0^5_{82}$
104/9	.021	.025	272/9	$.0^5_{72}$	$.0^5_{54}$
112/9	.017	.017	288/9	$.0^5_{40}$	$.0^5_{29}$
120/9	.012	.012	312/9	$.0^5_{38}$	$.0^5_{10}$
128/9	.010	$.0^2_{80}$	320/9	$.0^5_{33}$	$.0^6_{44}$
136/9	$.0^2_{64}$	$.0^2_{54}$	328/9	$.0^5_{17}$	$.0^6_{29}$
144/9	$.0^2_{39}$	$.0^2_{37}$	336/9	$.0^6_{46}$	$.0^6_{19}$
152/9	$.0^2_{34}$	$.0^2_{25}$	344/9	$.0^6_{28}$	$.0^6_{12}$
160/9	$.0^2_{25}$	$.0^2_{17}$	360/9	$.0^7_{15}$	$.0^7_{66}$

Note 8

For this other homogeneity test (which is only for the special case where p is known to be $\frac{1}{2}$), we define $r_i = \min(x_i, y_i)$ and

$$\begin{aligned} u_i = u_{n_i r_i} &= 2^{-n_i} \left[2 \sum_{\ell=0}^{r_i-1} \binom{n_i}{\ell} + \binom{n_i}{r_i} \right] & \text{if } r_i \neq \frac{1}{2} n_i \\ &= 2^{-n_i} \left[2 \sum_{\ell=0}^{r_i-1} \binom{n_i}{\ell} + \frac{1}{2} \binom{n_i}{r_i} \right] & \text{if } r_i = \frac{1}{2} n_i . \end{aligned}$$

It can be shown that, under (1) with $p = \frac{1}{2}$, $u_{n_i r_i}$ has expectation

$$E(u_{n_i r_i}) = \frac{1}{2}$$

and variance

$$v(n_i) = \text{var}(u_{n_i r_i}) = (1/12) - (1/3) 2^{-3n_i} S_{n_i} + 2^{-3n_i-2} \left(\frac{1}{2} n_i \right)^3 \delta_{n_i} ,$$

where δ_n is 0 or 1 according as n is odd or even, and where

$$S_n = \sum_{\ell=0}^n \binom{n}{\ell}^3 . \quad \text{Thus the statistic}$$

$$\frac{\sum_i w(n_i) u_{n_i r_i} - \frac{1}{2} \sum_i w(n_i)}{\left\{ \sum_i [w(n_i)]^2 v(n_i) \right\}^{\frac{1}{2}}} , \quad (A11)$$

where $w(n)$ denotes any (arbitrary) weight function, is approximately $N(0,1)$ under the null hypothesis. Now if the distribution of x_i is (speaking roughly) more "spread out" than (1) but has the same mean ($\frac{1}{2} n_i$), then we would naturally expect $E(u_{n_i r_i})$ to be $> \frac{1}{2}$; hence we use a test which rejects the null hypothesis of homogeneity for large (positive) values of the statistic (A11).

The $u_{n_i r_i}$'s needed in (A11) are most easily calculated via tables of the cumulative binomial distribution for $p = \frac{1}{2}$. The values of $v(n_i)$ can be obtained from the following listing if $n_i \leq 20$:

v(1) = .000000	v(11) = .072529
v(2) = .062500	v(12) = .076274
v(3) = .046875	v(13) = .074148
v(4) = .068359	v(14) = .077087
v(5) = .060425	v(15) = .075346
v(6) = .071655	v(16) = .077728
v(7) = .066650	v(17) = .076267
v(8) = .073759	v(18) = .078247
v(9) = .070218	v(19) = .076997
v(10) = .075211	v(20) = .078677

The weight function $w(n)$ in (All) serves to provide flexibility but at the same time forces a choice: if d_n denotes the amount (except for a multiplicative constant) by which $E(u_{nr})$ exceeds $\frac{1}{2}$ under an alternative hypothesis, then setting $w(n) = d_n/v(n)$ results in favorable power against that alternative; but different alternatives have different d_n 's, so that choices such as (e.g.) $w(n) = 1/v(n)$, $n^{\frac{1}{2}}/v(n)$, or $n/v(n)$ are all reasonable possibilities.

Full details of the (All) test appear elsewhere [Potthoff and Whittinghill (1963)]; here we only presented the practical essentials, in order to make the test available. The test is likely to find only occasional use. It could have favorable power against certain peculiar alternatives (such as might arise, e.g., in testing to detect cheating or fabrication of data).

Note 9

For Experiment 3, Gruneberg and Haldane (1937, p. 150) actually report the results $G_1 = 269.77$, $E(G_1) = 243$, and $\sigma(G_1) = 18.89$, which would give $z = +1.42$. However, a re-calculation shows that the correct value of $\sigma(G_1)$ is 21.01, which leads to $z = +1.27$.

Note 10

With p_{m*1} omitted, the $(m-1)$ -variate Dirichlet distribution [see, e.g., Wilks (1962), pp. 177-179] for $p_{*1}(\overline{m-1} \times 1) = (p_{1*1}, p_{2*1}, \dots, p_{(m-1)*1})'$ may be written in the form

$$\phi(p_{*1}) = \frac{\Gamma(a)}{\prod_{j=1}^m \Gamma(p_j a)} \left(\prod_{j=1}^{m-1} p_j^{a-1} \right) \left(1 - \sum_{j=1}^{m-1} p_{j*1} \right)^{p_m a - 1} \quad (A12)$$

and has $E(p_{j*1}) = p_j$, $\text{var}(p_{j*1}) = p_j(1-p_j)/(a+1)$, and (for $j \neq J$) $\text{cov}(p_{j*1}, p_{J*1}) = -p_j p_J / (a+1)$. Note how (A12) is a multivariate generalization of (A1). The probability function for $x_1(m \times 1) = (x_{11}, x_{12}, \dots, x_{1m})'$ under the alternative hypothesis is

$$\begin{aligned} f_1(x_1) &= \int \phi(p_{*1}) \frac{n_i!}{\prod_{j=1}^m x_{ij}!} \left(\prod_{j=1}^{m-1} p_{j*1}^{x_{ij}} \right) \left(1 - \sum_{j=1}^{m-1} p_{j*1} \right)^{x_{im}} \prod_{j=1}^{m-1} dp_{j*1} \\ &= \frac{\Gamma(a)}{\prod_{j=1}^m \Gamma(p_j a)} \frac{n_i!}{\prod_{j=1}^m x_{ij}!} \frac{\prod_{j=1}^m \Gamma(p_j a + x_{ij})}{\Gamma(a + n_i)} , \quad (A13) \end{aligned}$$

which is a generalization of (A2). We note that Mosimann (1962, p. 66) presented this same distribution (A13).

Proceeding analogously to (A3), we obtain

$$\begin{aligned} \frac{\prod_{i=1}^k f_1(x_i)}{\prod_{i=1}^k f_0(x_i)} &= \prod_{i=1}^k \frac{a_i \Gamma(a)}{\Gamma(a + n_i)} \prod_{j=1}^m \frac{\Gamma(p_j a + x_{ij})}{(p_j a)^{x_{ij}} \Gamma(p_j a)} \\ &= C \left[1 + \frac{1}{2} \left(\sum_{j=1}^m \frac{X_{\cdot j}}{p_j} \right) a^{-1} + o(a^{-2}) \right] . \quad (A14) \end{aligned}$$

Now if we consider a to be arbitrarily large [which makes the variances and covariances of the Dirichlet distribution (A12) arbitrarily close to 0], then (A14) verifies that the locally most powerful test is based on V (16).

Note 11

To prove formulas (18) for the mean and variance of V (16), one proceeds just as in Note 6.

Generalizing (A8), we find the third central moment of V (16) to be

$$E[(V-N.)^3] = 4 \left(\sum_{j=1}^m \frac{1}{p_j} - 3m + 2 \right) N. + 8(m-1) \sum_i (n_i)_3. \quad (A15)$$

Using (A15) and (18), we easily confirm that $(eV + f)$ [see (20-23)] has the correct first three moments.

If we generalize (A9) and (A10), we obtain respectively

$$\beta_2(eV+f)=3+ \frac{4}{(m-1)N.} \left[3 \sum_i (n_i)_4 + (9\gamma_1+6m-12) \sum_i (n_i)_3 + \left(\frac{1}{2} \gamma_2 - 2\gamma_1 + \frac{1}{2} m^2 - 3m + 3 \right) \sum_i (n_i)_2 \right] \quad (A16)$$

and

$$\beta_2(\chi_v^2)=3+ \frac{4}{(m-1)N.} \left[3 \frac{\{\sum_i (n_i)_3\}^2}{\sum_i (n_i)_2} + (3\gamma_1 + 3m-6) \sum_i (n_i)_3 + \frac{3}{4} (\gamma_1+m-2)^2 \sum_i (n_i)_2 \right] \quad (A17)$$

as the β_2 -coefficients of the exact and approximating distributions of $(eV+f)$ for the general multinomial case, γ_1 and γ_2 in (A16-A17) being defined by

$$\gamma_1 = \frac{\sum_{j=1}^m \frac{1}{p_j} - m^2}{m-1}, \quad \gamma_2 = \frac{\sum_{j=1}^m \frac{1}{p_j^2} - m^3}{m-1}.$$

Note 12

It is not difficult to show that

$$E(G_3) = k \quad \text{and} \quad \sigma^2(G_3) = 2k + (1/\lambda) \sum_i (1/b_i) . \quad (A18)$$

Using Nass's method, we calculate c (A7) via formulas (A18), and then refer $(c G_3)$ to the χ^2 distribution with d.f. = ck .

The statistic G_3 (27), incidentally, is a generalization of the χ^2 statistic mentioned by Kendall and Stuart (1961, p. 579, equation 33.123).

Note 13

If λ_{*i} follows a gamma distribution [see, e.g., Mood (1950), pp. 112-115] with mean λ and variance σ^2 , then the density is

$$\phi(\lambda_{*i}) = \frac{1}{(\sigma^2/\lambda) \lambda^2/\sigma^2 \Gamma(\lambda^2/\sigma^2)} \lambda_{*i}^{(\lambda^2/\sigma^2)-1} e^{-(\lambda/\sigma^2)\lambda_{*i}} . \quad (A19)$$

Thus x_i follows the negative binomial distribution

$$\begin{aligned} f_1(x_i) &= \int_0^\infty \phi(\lambda_{*i}) \left[e^{-b_i \lambda_{*i}} (b_i \lambda_{*i})^{x_i} / x_i! \right] d\lambda_{*i} \\ &= \frac{\Gamma(x_i + \frac{\lambda^2}{\sigma^2})}{x_i! \Gamma(\frac{\lambda^2}{\sigma^2})} \left[\frac{\frac{\lambda}{\sigma^2}}{b_i + \frac{\lambda}{\sigma^2}} \right] \left[\frac{b_i}{b_i + \frac{\lambda}{\sigma^2}} \right]^{x_i} \end{aligned} \quad (A20)$$

under the alternative hypothesis.

The choice of the gamma distribution (A19) gives us a model which is not only tractable but also has the proper range (0 to ∞) for the λ_{*i} 's. Numerous authors, particularly those dealing with accident distributions, have worked with models in which a Poisson parameter is drawn from a gamma distribution. An early article by Greenwood and Yule (1920, see especially pp. 273-276) suggested that, if individuals vary in their pre-dispositions to incur accidents, then these varying pre-dispositions (as represented by the

λ_{*i} 's) might be assumed to follow a gamma distribution and the number of accidents of the i -th individual would follow a Poisson distribution with mean λ_{*i} . [Still earlier, Pearson (1915) had suggested that a negative binomial distribution could result from a mixture of Poisson distributions.] Cochran (1954, pp. 419-422) not only mentions the negative binomial distribution (Poisson compounded by gamma) as a common alternative against which to test a Poisson null hypothesis, but he even presents a specific numerical example (p. 419) to illustrate the comparative power of different tests against a negative binomial alternative. In a Bayesian context, Raiffa and Schlaifer (1961, Chapter 3) combine the Poisson distribution with a gamma "conjugate prior" distribution.

Note 14

We use (26) and (A20), and write

$$\begin{aligned}
 \frac{\prod_{i=1}^k f_1(x_i)}{\prod_{i=1}^k f_0(x_i)} &= \prod_{i=1}^k C_i \frac{\Gamma(x_i + \frac{\lambda^2}{\sigma^2}) / \Gamma(\frac{\lambda^2}{\sigma^2})}{x_i^{x_i} (b_i \lambda + \frac{\lambda^2}{\sigma^2})} \\
 &= \prod_{i=1}^k C_i \frac{1 + (1/\lambda^2) \binom{x_i}{2} \sigma^2 + o(\sigma^4)}{1 + (1/\lambda^2) \lambda b_i x_i \sigma^2 + o(\sigma^4)} \\
 &= C [1 + (1/2\lambda^2) \{ \sum_i x_i(x_i-1) - 2\lambda \sum_i b_i x_i \} \sigma^2 + o(\sigma^4)]. \quad (A21)
 \end{aligned}$$

Taking σ^2 arbitrarily small in (A21), we verify that the locally most powerful test is based on U (28).

Note 15

It is not hard to show that

$$E[\{U - E(U)\}^3] = 4 \lambda^2 \sum_i b_i^2 + 8 \lambda^3 \sum_i b_i^3 \quad (A22)$$

is the third central moment of U (28). By utilizing (29) and (A22), we easily verify that $(gU + h)$ [see (31-34)] has the correct first three moments.

For purposes of judging the accuracy of the χ^2 approximation to the distribution of $(gU + h)$ (31), it is helpful to know the β_2 -coefficients of the exact and approximating distributions. We find respectively

$$\beta_2(gU + h) = 3 + \frac{12 \sum_i b_i^4}{(\sum_i b_i^2)^2} + \frac{36 \sum_i b_i^3}{\lambda (\sum_i b_i^2)^2} + \frac{2}{\lambda^2 \sum_i b_i^2} \quad (A23)$$

and

$$\beta_2(\chi^2_\nu) = 3 + \frac{12 (\sum_i b_i^3)^2}{(\sum_i b_i^2)^3} + \frac{12 \sum_i b_i^3}{\lambda (\sum_i b_i^2)^2} + \frac{3}{\lambda^2 \sum_i b_i^2} \quad (A24)$$

Both (A23) and (A24) will always be > 3 , and would appear to be satisfactorily close to one another for most cases.

Note 16

Utilizing (e.g.) formulas (10) given by Nass (1959, p. 368), we can write

$$E(G_k) = k-1 \text{ and } \sigma^2(G_k) = 2(k-1) + \frac{1}{\sum_i x_i} \left[\left(\sum_i b_i \right) \left(\sum_i \frac{1}{b_i} \right) - k^2 - 2k + 2 \right] \quad (A25)$$

for the G_k (37) statistic; $\sigma^2(G_k)$ in (A25) is a conditional variance, conditional on $\sum_i x_i$. To employ Nass's refinement, we calculate c (A7) using formulas (A 25), and then refer $(c G_k)$ to the χ^2 distribution with d.f. = $c(k-1)$.

For the case where all b_i 's are 1, Nass (1959, pp. 368-369) covers the G_4 (38) test in some detail, and suggests a continuity correction.

Note 17

This property is proved via the methods indicated (e.g.) by Lehmann (1959, Chapter 4) or Fraser (1957, pp. 87-89, 93). The argument contains several steps:

(i) If $\phi(x)$ [where $x(k \times 1) = (x_1, x_2, \dots, x_k)'$] denotes any test of f_0 (26) against f_1 (A20) [remember that all b_i 's = 1], then

$$\sum_{s=0}^{\infty} \left[\sum_{x: \sum_{i=1}^k x_i = s} \phi(x) \prod_{i=1}^k f_1(x_i) \right] \quad (A26)$$

is its power function. Using the fact that $\lim_{\sigma^2 \rightarrow 0} f_1(x_i) = f_0(x_i)$, we can easily show that each term in the series (A26) is a continuous function of σ^2 for fixed λ . Hence, to show that the sum (A26) is a continuous function of σ^2 for fixed λ and $0 \leq \sigma^2 \leq E$ (say), it is sufficient to establish that the series (A26) is uniformly convergent for fixed λ and $0 \leq \sigma^2 \leq E$. Now (A26) clearly converges, say to $\psi(\lambda, \sigma^2)$. To show that this convergence is uniform in $0 \leq \sigma^2 \leq E$, we use A_s to denote the expression inside the square brackets in (A26) and note that

$$\begin{aligned} |\psi(\lambda, \sigma^2) - \sum_{s=0}^M A_s| &= \sum_{s=M+1}^{\infty} A_s \leq \sum_{s=M+1}^{\infty} \left[\sum_{x: \sum_{i=1}^k x_i = s} \prod_{i=1}^k f_1(x_i) \right] \\ &= \Pr \left\{ \sum_{i=1}^k x_i > M \right\} \leq k(\lambda + \sigma^2)/(M - k\lambda)^2 \leq k(\lambda + E)/(M - k\lambda)^2, \end{aligned} \quad (A27)$$

where the first inequality follows from the fact that $0 \leq \phi(x) \leq 1$, and the second inequality (which is valid for $M > k\lambda$) is obtained by applying Tchebycheff's inequality using the relations $E(\sum_{i=1}^k x_i) = k\lambda$,

$\text{var} \left(\sum_i x_i \right) = k(\lambda + \sigma^2)$ [see Feller (1957), p. 253]. Since the final expression in (A27) will be $< \epsilon$ if M is $> k\lambda + [k(\lambda + E)/\epsilon]^{\frac{1}{2}}$ (an expression free of σ^2), it follows that the convergence is uniform.

Inasmuch as the power function (A26) is continuous in σ^2 for fixed λ and $0 \leq \sigma^2 \leq E$, we may conclude that any unbiased test of size α must necessarily be a similar test of size α .

(ii) It is well-known that the statistic $\sum_i x_i$ is both sufficient and boundedly complete for f_0 (26). Hence any similar test must be a test of Neyman structure with respect to $\sum_i x_i$ (i.e., its conditional size, given $\sum_i x_i$, is equal to α for all values of $\sum_i x_i$).

(iii) The frequency function of x given $\sum_i x_i$ is

$$\frac{(\sum_i x_i)!}{k \prod_{i=1}^k x_i!} \left(\frac{1}{k} \right)^{\sum_i x_i} \quad (\text{A28})$$

under the null hypothesis, and is

$$\frac{\prod_{i=1}^k f_1(x_i)}{\frac{\Gamma(\sum_i x_i + \frac{k\lambda^2}{\sigma^2})}{(\sum_i x_i)! \Gamma(\frac{k\lambda^2}{\sigma^2})} \left[\frac{\frac{\lambda}{\sigma^2}}{1 + \frac{\lambda}{\sigma^2}} \right]^{k\lambda^2/\sigma^2} \left[\frac{1}{1 + \frac{\lambda}{\sigma^2}} \right]^{\sum_i x_i}} \quad (\text{A29})$$

under the alternative hypothesis. [To obtain the denominator of (A29), we use Feller (1957), p. 253, equation (2.16).] The most powerful test of Neyman structure thus has critical region based on large values of the ratio of (A29) to (A28), which is

$$\frac{\prod_{i=1}^k [\Gamma(x_i + \frac{\lambda^2}{\sigma^2}) / \Gamma(\frac{\lambda^2}{\sigma^2})]}{\Gamma(\sum x_i + \frac{k\lambda^2}{\sigma^2}) / \prod_{i=1}^k \Gamma(\frac{\lambda^2}{\sigma^2})} = 1 + \frac{\bar{x}}{2\lambda^2} (G_4 - k + 1) \sigma^2 + o(\sigma^4), \quad (A30)$$

where G_4 is given by (38). If we assume a small positive lower bound for λ ($\lambda \geq \delta > 0$, say), then (remembering that $\bar{x}$ is treated as a constant) we may conclude from (A30) that the G_4 (38) test is the most powerful test of Neyman structure when σ^2 is arbitrarily close to 0. [An alternative interpretation from (A30) is also possible: the right-hand side of (A30) may be written as a series of terms in powers of σ^2/λ^2 (rather than σ^2), from which we may then conclude that the G_4 (38) test is most powerful for σ^2/λ^2 (rather than σ^2) arbitrarily close to 0 (no lower bound δ now being assumed for λ).]

(iv) Since a most powerful test of Neyman structure is necessarily unbiased, it follows finally that the G_4 (38) test is (locally) a most powerful unbiased test, irrespective of the value of λ .

Note 18

We might note still a third possible test for the case where λ is unknown and the b_i 's are not all equal. Under the null hypothesis of homogeneity, the conditional distribution of $(x_1, x_2, \dots, x_k)$ given $\sum_i x_i$ is multinomial with parameters $(b_1/\sum_i b_i, b_2/\sum_i b_i, \dots, b_k/\sum_i b_i)$. Hence the test based on V (16) is applicable. Formulas (16) and (21) become

$$V = \left(\sum_{i=1}^k b_i \right) \left[\sum_{i=1}^k \frac{x_i(x_i-1)}{b_i} \right] \quad (A31)$$

and

$$e = \frac{2(k-1)}{\left(\sum_{i=1}^k b_i \right) \left[\sum_{i=1}^k \frac{1}{b_i} \right] - 3k+2+2(k-1) \left(\sum_{i=1}^k x_i - 2 \right)} \quad (A32)$$

respectively; in using (22) and (23), we set $m = k$ and $N. = \left(\sum_{i=1}^k x_i \right) \left(\sum_{i=1}^k x_i - 1 \right)$.

This third test is exact rather than conservative. When all b_i 's are 1, the test reduces essentially to the G_4 (38) test.

Note 19

We might mention that application of the V (A31, A32, 22-23, 20) test gives $P = .43$ for the data of Example 8 (better than the G_4 test but not as good as the U test), and gives $P = .004$ for the data for Brand C in Example 9 (better than the U test and almost as good as the G_4 test).

ACKNOWLEDGMENTS

The authors are indebted to Professor N. L. Johnson and to Professor Wassily Hoeffding for their helpful comments and suggestions.

REFERENCES

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. Biometrics, 11, pp. 375 - 386.

COCHRAN, WILLIAM G. (1952). The χ^2 test of goodness of fit. Ann. Math Statist., 23, pp. 315 - 345.

COCHRAN, WILLIAM G. (1954) Some methods for strengthening the common χ^2 tests. Biometrics, 10, pp. 417 - 451.

CRAMER, HARALD (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

DARWIN, J. H. (1960). An ecological distribution akin to Fisher's logarithmic distribution. Biometrics, 16, pp. 51 - 60.

EDWARDS, A. W. F., and FRACCARO, M. (1958). The sex distribution ratio in the offspring of 5477 Swedish ministers of religion, 1585 - 1920. Hereditas, 44, pp. 447 - 450.

FELLER, WILLIAM (1957). An Introduction to Probability Theory and Its Applications, Volume 1 (2nd edition). John Wiley and Sons, Inc., New York.

FISHER, R. A. (1950). The significance of deviations from expectation in a Poisson series. Biometrics, 6, pp. 17 - 24 .

FISHER, R. A., THORNTON, H. G., and MACKENZIE, W. A. (1922). The accuracy of the plating method of estimating the density of bacterial populations. Annals of Applied Biology, 9, pp. 325 - 359. (Reprinted as Paper 4 in Fisher, R. A., Contributions to Mathematical Statistics, John Wiley and Sons, Inc., New York, 1950.)

FRASER, D. A. S. (1957). Nonparametric Methods in Statistics. John Wiley and Sons, Inc., New York.

GEISSLER, ARTHUR (1889). Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. Zeitschrift des Königlichen Sächsischen Statistischen Bureaus, 35, pp. 1 - 24.

GINI, CORRADO (1951). Combinations and sequences of sexes in human families and mammal litters. Acta Genetica et Statistica Medica, 2, pp. 220 - 244.

GOWEN, JOHN WHITTEMORE (1919). A biometrical study of crossing over. On the mechanism of crossing over in the third chromosome of Drosophila melanogaster. Genetics, 4, pp. 205 - 250.

GREENWOOD, MAJOR, and YULE, G. UDNY (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. A.R. Statist. Soc., 83, pp. 255 - 279.

GRUNEBERG, HANS, and HALDANE, J. B. S. (1937). Tests of goodness of fit applied to records of Mendelian segregation in mice. Biometrika, 29, pp. 144 - 153.

GURLAND, JOHN (1958). A generalized class of contagious distributions. Biometrics, 14, pp. 229 - 249.

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2 , used as a test of goodness of fit, when expectations are small. Biometrika, 29, pp. 133 - 143.

HALDANE, J. B. S. (1940). The mean and variance of χ^2 , when used as a test of homogeneity, when expectations are small. Biometrika, 31, pp. 346 - 355.

HARRIS, J. ARTHUR, and GUNSTAD, BORGHILD (1930). The correlation between the sex of human siblings. I. The correlation in the general population. Genetics, 15, pp. 445 - 461.

HORSNELL, G. (1957). Economical acceptance sampling schemes. J. R. Statist. Soc. A, 120, pp. 148 - 201.

JOHANSSON, I. (1932). The sex ratio and multiple births in cattle. Zeitschrift für Züchtung, Reihe B, 24, p. 183.

KENDALL, MAURICE G., and STUART, ALAN (1961). The Advanced Theory of Statistics, vol. 2. Charles Griffin and Company Limited, London.

LEHMANN, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, Inc., New York.

LINDLEY, D. V. (1957). Binomial sampling schemes and the concept of information. Biometrika, 44, pp. 179 - 186.

MOOD, ALEXANDER MCFARLANE (1950). Introduction to the Theory of Statistics. McGraw-Hill Book Company, Inc., New York.

MOSIMANN, JAMES E. (1962). On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. Biometrika, 49, pp. 65 - 82.

NASS, C. A. G. (1959). The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika, 46, pp. 365 - 385.

PEARSON, KARL (1915). On certain types of compound frequency distributions in which the components can be individually described by binomial series. Biometrika, 11, pp. 139 - 144.

PIELOU, E. C. (1963). The distribution of diseased trees with respect to healthy ones in a patchily infected forest. Biometrics, 19, pp. 450 - 459.

POTTHOFF, RICHARD F., and WHITTINGHILL, MAURICE (1963). On testing for independence of unbiased coin tosses lumped in groups too small to use χ^2 . Institute of Statistics Mimeo Series No. 347, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

RAIFFA, HOWARD, and SCHLAIFER, ROBERT (1961). Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration, Harvard University, Boston, Massachusetts.

RAO, C. RADHAKRISHNA (1952). Advanced Statistical Methods in Biometric Research. John Wiley and Sons, Inc., New York.

RIFE, D. CECIL, and SNYDER, L. H. (1937). The distribution of sex ratios within families in an Ohio city. Studies in human inheritance, XVI. Human Biology, 9, pp. 99-103.

ROBERTSON, A. (1951). The analysis of heterogeneity in the binomial distribution. Ann. Eugen., Lond., 16, pp. 1-15.

SCHEINFELD, AMRAM (1961). The Basic Facts of Human Heredity. Washington Square Press, Inc., New York.

SKELLAM, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J. R. Statist. Soc. B, 10, pp. 257-261.

SLATER, ELIOT (1943). A demographic study of a psychopathic population. Ann. Eugen., Lond., 12, pp. 121-137.

SMITH, C. A. B. (1951). A test for heterogeneity of proportions. Ann. Eugen., Lond., 16, pp. 16-25.

WHITTINGHILL, MAURICE (1950). Consequences of crossing over in oögonial cells. Genetics, 35, pp. 38-43.

WILKS, SAMUEL S. (1962). Mathematical Statistics. John Wiley and Sons, Inc., New York.

WRIGHT, SEWALL (1934). An analysis of variability in number of digits in an inbred strain of guinea pigs. Genetics, 19, pp. 506-536.