

A consideration of the chi-square test of Hardy-Weinberg equilibrium in a non-multinomial situation

BY CHARLES F. SING AND EDWARD D. ROTHMAN

*Departments of Human Genetics and Statistics, University of Michigan,
Ann Arbor, Michigan 48104, U.S.A.*

Deviation from the Hardy-Weinberg proportion of each genotype frequency may be attributable to a number of forces (selection, inbreeding, mutation, migration, subdivision) operating in the population sampled. Smith (1970) has derived estimators of these deviations and discussed the difficulty of sorting out the effects of inbreeding, population subdivision, and silent alleles from the effects selection may have on the frequencies of genotypes. An additional condition which may contribute to deviations from Hardy-Weinberg that has not received adequate attention is non-independent sampling of individuals collected to detect the operation of such forces in a population. For example, the detection of an increase in the frequency of homozygotes and a decrease of heterozygotes in small populations of humans due to inbreeding is an obvious first consideration of the geneticist who is interested in studying the effects of selection on genotype frequencies. However, the correct statement of probability of observing a test statistic computed from the sample usually depends on the assumption that members of the sample are chosen independently of one another. In this report we investigate the effect of non-independence of observations, characterized by the multinomial Dirichlet, on the level of an $\alpha = 0.05$ size (significance level) χ^2 goodness-of-fit test of Hardy-Weinberg. A modified test is presented which gives the expected type I error when the null hypothesis of Hardy-Weinberg is true and there is sampling correlation.

We begin by defining the relative frequencies of genotypes for a locus with two codominant alleles in a very large reference population of diploid individuals. Letting p be the frequency of one allele A and $q (= 1 - p)$ be the frequency of the other, a , the expected relative frequencies in this population of the AA , Aa and aa genotypes are p^2 , $2pq$, and q^2 , respectively.

The test most often utilized on a sample of size N to detect inbreeding in this reference population is based on the value of X^2 :

$$X^2 = \frac{(T_1 - p^2N)^2}{p^2N} + \frac{(T_2 - 2pqN)^2}{2pqN} + \frac{(T_3 - q^2N)^2}{q^2N}, \quad (1)$$

where T_1 , T_2 and T_3 are the observed number of AA , Aa and aa , respectively, and $T_1 + T_2 + T_3 = N$ individuals randomly drawn from the population. (Significantly large values of X^2 are taken to indicate inbreeding.) The statistic, X^2 , is distributed as a χ^2 with one degree of freedom when p is estimated $\hat{p} = (T_1 + \frac{1}{2}T_2)/N$, the population is large compared to the sample, the expected number of each genotype is at least 5, and the individuals in the sample are unrelated, i.e. stochastically independent. Ward & Sing (1970) considered the statistical power of this goodness-of-fit test to detect inbreeding. It was shown that very large samples are required to be confident that χ^2 will reject $F = 0$ when in fact F in the population being sampled takes on a value which is reasonable for humans (i.e. $F < 0.10$). For example, a sample of 105,000 is required to be 90% confident of detecting $F = 0.01$ when a 5% test

criterion is chosen (10,500,000 when $F = 0.001$). Such samples are unreasonably large in view of the above conditions stating the relationship of population size to sample size.

PROBABILITY MODEL

Consider a parameter ρ defined as the average correlation between the genotypes of any two individuals sampled from a population with Hardy-Weinberg genotype frequencies. This correlation could be attributable to co-ancestry of the individuals forming the sample, finiteness of the population being sampled, or any other bias in the sampling procedure which leads to non-independence of the observations to be used for analysis. In a previous paper (Rothman, Sing & Templeton, 1974) we suggested the multinomial Dirichlet (MD) as a simple model for the analysis of measures of correlation between pairs of alleles within and between individuals in a population divided into subpopulations. Here we apply the multinomial Dirichlet distribution to describe the sampling process whereby there is a correlation between the genotypes of any pair of individuals in a sample drawn from a single unsubdivided population. Even though it may be argued that the underlying set of biological assumptions necessary to derive such a model are not met, experience has shown that in practice a useful distribution need only accurately describe the measurable outcome of the process being modelled. The multinomial Dirichlet can be derived by mixing the multinomial with a Dirichlet distribution (see Johnson & Kotz, 1969). Briefly, if (T_1, T_2, T_3) conditional on $(\theta_1, \theta_2, \theta_3)$ is multinomial $(N; \theta_1, \theta_2, \theta_3)$ and $(\theta_1, \theta_2, \theta_3)$ is Dirichlet $(\phi P_1, \phi P_2, \phi P_3)$, then the unconditional distribution of (T_1, T_2, T_3) is multinomial Dirichlet $(N; \rho, P_1, P_2, P_3)$, where $\phi = (1-\rho)/\rho$. In forming a sample, when the Hardy-Weinberg law is true and $\rho \neq 0.0$, on the first draw the probability of drawing each genotype is $P_1 = p^2$, $P_2 = 2pq$ and $P_3 = q^2$. However on the j th succeeding draw these probabilities conditional on the past are

$$P_{1j} = \frac{\rho T_{1(j-1)} + p^2(1-\rho)}{(j-1)\rho + (1-\rho)},$$

$$P_{2j} = \frac{\rho T_{2(j-1)} + 2pq(1-\rho)}{(j-1)\rho + (1-\rho)}$$

and

$$P_{3j} = 1 - P_{1j} - P_{2j}.$$

Alternatively the probability mass function of the multinomial Dirichlet may be written

$$\Pr (T_1, T_2, T_3 | P_1, P_2, \phi) = \frac{\binom{T_1 + P_1\phi - 1}{T_1} \binom{T_2 + P_2\phi - 1}{T_2} \binom{T_3 + P_3\phi - 1}{T_3}}{\binom{N + \phi - 1}{N}}, \quad (2)$$

where $\phi = (1-\rho)/\rho$. For this parameterization of the sampling scheme as $\phi \rightarrow \infty$ ($\rho \rightarrow 0$) equation (2) reduces to the multinomial $(N; P_1, P_2, P_3)$:

$$\Pr (T_1, T_2, T_3 | P_1, P_2) = \frac{N!}{T_1! T_2! T_3!} P_1^{T_1} P_2^{T_2} P_3^{T_3}.$$

We note that the expected T_i , $i = 1, 2, 3$ is the same under either model while the variance of T_i when the MD is true is $(1+\rho(N-1))$ multiplied by the variance of T_i when the multi-

nomial is true. The latter result forms the basis of the modified χ^2 test given below. We will use the multinomial Dirichlet distribution to evaluate the effect of correlation between observations forming a sample on the type I error of the χ^2 goodness-of-fit test given by equation (1).

CHI-SQUARE RESULTS

The type I error for the χ^2 goodness-of-fit test of the Hardy-Weinberg hypothesis when sampling follows the multinomial Dirichlet distribution was computed for each combination of three allele frequencies (0.05, 0.25 and 0.50) with four values of ρ (0.0, 0.001, 0.01 and 0.1) and two sample sizes (50 and 100). The exact level of the test based on large values of X^2 when samples are drawn from the multinomial Dirichlet with $(P_1, P_2, P_3) = (p^2, 2pq, q^2)$ is computed as

$$L = \sum_{S_c} \text{MD} (N; \rho, p^2, 2pq, q^2),$$

where

$$S_c = \left\{ (T_1, T_2, T_3) : \sum_{i=1}^3 T_i = N, T_i \geq 0, X^2 > c \right\}$$

and $c =$ the 95th percentile of a $\chi^2_{(D.F.)}$ distribution. In addition, rather than computing the values of X^2 using known values of allele frequency, in which case the expected X^2 has 2 degrees of freedom and the critical value will be 5.991, we considered the X^2 statistic based on the estimate of $p = (T_1 + \frac{1}{2}T_2)/N$. In the latter case X^2 is compared to the 95th percentile of a χ^2 with 1 D.F., namely 3.84. In cases where expected numbers in any class did not exceed 5, a Yates correction was applied.

The solid lines in Fig. 1 give the probability of rejecting the Hardy-Weinberg expectation given the values of allele frequency, N , and ρ considered. For the range of allele frequencies considered, the type I error increases as ρ increases. There is a substantial deviation from the expected α level of the test due to sample correlation. As expected, the larger sample size rejects the hypothesis of Hardy-Weinberg when ρ is not 0 more often than for the smaller sample size. The type I error when $\rho = 0$ is a measure of the error of an $\alpha = 0.05$ level test when sampling is multinomial. Again, as expected, this error increases as sample size decreases and allele frequency deviates from 0.5. The peculiar result for $p = 0.05$ based on samples of size 50 reflects the fact that the χ^2 distribution does not approximate the distribution of X^2 even when $\rho = 0$. The χ^2 goodness-of-fit for Hardy-Weinberg equilibrium is very sensitive to non-independent sampling.

These results are not unexpected since, if we consider the testing problem

$$H_0: \text{multinomial} (N; p^2, 2pq, q^2) \text{ versus } H_A: \text{MD} (N; \rho, p^2, 2pq, q^2),$$

the locally ($\rho \rightarrow 0$) most powerful test is based on large values of X^2 ! This result obtains from the expansion of the likelihood ratio test about $\rho = 0$. Since H_A is the null hypothesis in our investigation, the observed level of the χ^2 test as a function of ρ is in fact the power of this test in the problem H_0 versus H_A .

MODIFICATION OF THE χ^2 TEST

Consideration of the expectation of the X^2 statistic when there is correlation among individuals implies that

$$E (X^2 | \rho) = \text{D.F.} / \beta,$$

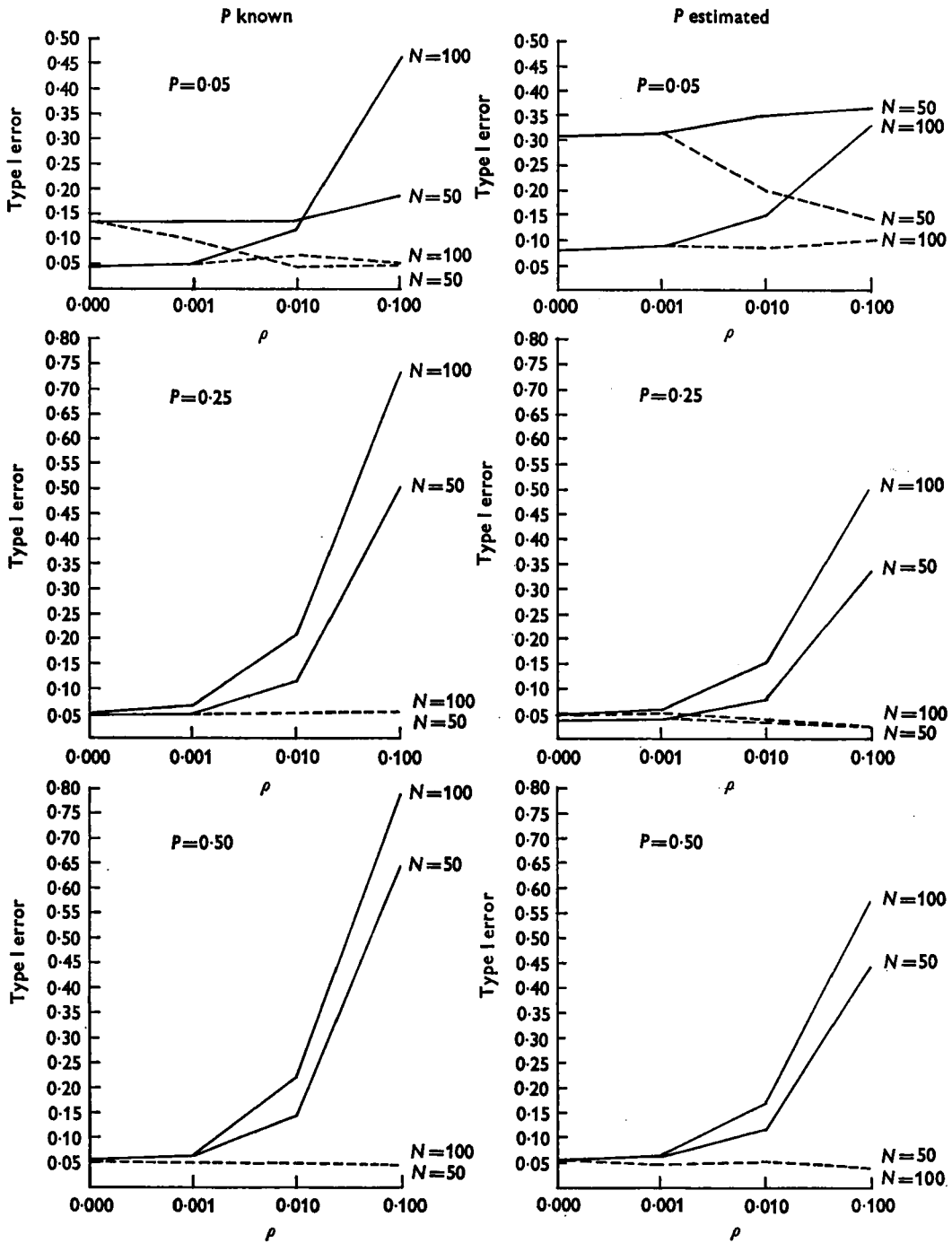


Fig. 1. A comparison of the goodness-of-fit X^2 (solid line) with βX^2 (dashed line) for two sample sizes, three allele frequencies and four values of ρ when sampling is according to the multinomial Dirichlet.

where $\beta = 1/[1 + \rho(N - 1)]$. This result does not require the assumption of the multinomial Dirichlet. For, if we set

$$T_i = \sum_{j=1}^N \delta_{ij} \quad (i = 1, 2, 3),$$

where

$$\delta_{ij} = 1 \text{ if } j\text{th observation is in cell } i, 0 \text{ otherwise}$$

and

$$\Pr (\delta_{ij} = 1, \delta_{i'j'} = 1) = (1 - \rho)P_i^2 + \rho P_i \quad (j \neq j'),$$

the expected X^2 is as stated when P_i are known.

The type I error for the test statistic βX^2 was computed for the same combinations of p , ρ , and N in the manner described above. The results are also given in Fig. 1 (dashed line). Essentially, the corrected X^2 gives an approximate 5% test when H-W is true regardless of the level of ρ considered here (on the order of $1/N$). We note that, in practice, choice of a larger value of ρ than is correct for a population will lead to a conservative test.

SUMMARY

The multinomial Dirichlet distribution was used to study the effect of correlation between observations in a sample on the frequency of rejection of the Hardy-Weinberg Law when in fact it was true in the population being sampled. It was shown that the usual χ^2 goodness-of-fit test of Hardy-Weinberg is very sensitive to non-multinomial sampling. In view of the lack of statistical power of the test to detect deviations due to inbreeding, it is likely that whenever H-W is rejected using samples of size 100 or less, the underlying causation is sample correlation rather than failure of the H-W law to be true. Related to these findings is, of course, the effect of pooling heterogeneous frequencies or, in the case of contingency tables, Simpson's paradox (see Simpson, 1951).

This work was supported by U.S. Atomic Energy Commission Contract AT(11-1)-1552.

The authors wish to thank Professor C. A. B. Smith for his kind advice during the preparation of this manuscript.

REFERENCES

- JOHNSON, N. L. & KOTZ, S. (1969). *Discrete Distributions*. Somerset, N.J.: Wiley.
- ROTHMAN, E. D., SING, C. F. & TEMPLETON, A. R. (1974). A model for analysis of population structure. *Genetics* **78**, 943-60.
- SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B* **13** (2), 238-41.
- SMITH, C. A. B. (1970). A note on testing the Hardy-Weinberg law. *Annals of Human Genetics* **33**, 377-83.
- WARD, R. H. & SING, C. F. (1970). A consideration of the power of the χ^2 test to detect inbreeding effects in natural populations. *American Naturalist* **104**, 355-63.

