On the Asymptotic Power of Tests for Independence in Contingency Tables from Stratified Samples

Gad Nathan


Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197212%2967%3A340%3C917%3AOTAPOT%3E2.0.CO%3B2-D

On the Asymptotic Power of Tests for Independence in Contingency Tables from Stratified Samples

GAD NATHAN

The statistics of several alternative procedures, proposed to test for independence in contingency tables derived from stratified samples, all have the same asymptotic chi-square distribution, under the null hypothesis. By considering a modified sampling procedure, it is shown here that they also have the same asymptotic distribution under any alternative hypothesis—the noncentral chi square, with the noncentrality parameter of Wald's test for the case of independently and identically distributed variables. Simulation, used to compare the finite-sample powers of the tests, shows small differences between the various methods and close approximation to the asymptotic results.

1. INTRODUCTION

Several alternative approximative methods have been proposed to test for independence in contingency tables which are derived from a stratified sample, rather than from a simple random sample as in the classical case.

The author [6] has proposed an iterative procedure which approximates, as closely as necessary, the exact maximum likelihood estimates of the cell-probabilities, under the nonlinear null hypothesis. The log-likelihood ratio, the chi-square or chi-one square statistic based on these estimates can then be used. The first iteration estimates of the cell probabilities under this procedure are exactly those which minimize the chi-one square statistic, subject to a linearization of the null hypothesis. This was proposed by Bhapkar [1], who showed [2] that this minimal chi-one square statistic is algebraically equivalent to the large sample (Wald) test statistic, based on the asymptotic normality of unbiased estimators of the linearized hypothesis function or of "natural" estimators of the nonlinear hypothesis function. Garza-Hernandez and McCarthy [4] have proposed using the likelihood ratio test, based on the assumption of asymptotic normality of ratio estimators of the conditional cell probabilities, given one of the marginal classifications. This is again equivalent to a large sample (Wald) statistic based on the ratio estimators of the hypothesis function, which is linear in terms of the conditional probabilities.

The asymptotic distribution of all statistics proposed is chi-square under the null hypothesis. By considering a modified sampling procedure, it is shown here that Bhapkar's statistic [1] and the three statistics based on the author's approximations to maximum likelihood estimators under the null hypothesis [6], all have the same asymptotic distribution under any hypothesis. Simulation is used to compare finite sample size powers of these statistics with the asymptotic powers for a certain set of parameters.

2. THE STATISTICS COMPARED

The population is assumed to be divided into t strata and, within each stratum, classified by a double classification with r and s categories. Let \( P_{ijk} \) (\( i=1, \ldots, r; j=1, \ldots, s; k=1, \ldots, t \)) be the joint probability that an item is in stratum \( k \) and is classified in the \((i,j)\)th cell. Marginal probabilities are denoted \( P_{i..}, P_{.j..}, P_{ij..} \) etc., with \( P_k = \sum_i \sum_j P_{ijk} \) assumed known and \( \Sigma_k P_k = 1 \). The hypothesis of interest is:

\[ H_0: P_{ij..} = P_{.i..} P_{.j..} \quad (i=1, \ldots, r; j=1, \ldots, s). \]  

The basic sampling procedure considered in [6] (referred to in the following as "Procedure A") is simple stratified sampling (with replacement), of size \( n \), with a fixed number of units, \( n_k \), selected within the \( k \)th stratum, \( (\Sigma_k n_k = n) \). Let \( n_{ijk} \) be the number of sample items from stratum \( k \) in the \((i,j)\)th cell (with \( \Sigma_i \Sigma_j n_{ijk} = n_k \)). The joint probability function of \( \{n_{ijk}\} \), given the values of \( \{P_{ijk}\} \), is then:

\[ \psi_A([n_{ijk}] | \{P_{ijk}\}) = \prod_k \left[ \frac{n_k!}{\prod_{i,j} n_{ijk}} \prod_{i,j} \left( \frac{P_{ijk}}{P_k} \right)^{n_{ijk}} \right]. \]  

The author [6] has proposed the log-likelihood ratio statistic:

\[ G = 2 \sum_i n_{ijk} \ln \left( \frac{\hat{P}_{ijk}}{\hat{P}_{ijk}(N)} \right), \]  

to test the hypothesis (2.1), where:

\[ \hat{P}_{ijk} = (n_{ijk}/n_k) P_k \]  

are the unrestricted maximum likelihood estimates of \( P_{ijk} \), and \( \hat{P}_{ijk}(N) \) are obtained at the \( N \)th step of an iteration procedure to approximate as closely as necessary the maximum likelihood estimates under the null hypothesis. The first step estimates \( \hat{P}_{ijk} \) (2.1) were shown [6] to be those which minimize the chi-one square statistic:
subject to a linearization of the null hypothesis (2.1), as proposed by Bhapkar [1].

If we define:

\[ P' = (P_{11}, \ldots, P_{rs}); \]

\[ \hat{P}' = (\hat{P}_{11}, \ldots, \hat{P}_{rs}); \]

\[ h(P) = [h_{11}(P), \ldots, h_{r-1,s-1}(P)]; \]

then Bhapkar [2] has shown that the minimum of (2.5) subject to a linearization of (2.1) is given by:

\[ T = [h(\hat{P})]C^{-1}[h(\hat{P})], \]  

(2.10)

where:

\[ C = n^{-1}[H(\hat{P})B^{-1}(\hat{P})H'(\hat{P})], \]  

(2.11)

\[ B(P) = \frac{1}{n} \left\| E_p \left( \frac{\partial^2 \ln \psi}{\partial P_{ijk} \partial P'_{i'j'k'}} \right) \right\|_{(r-1)\times(r-1)} \]  

(2.12)

and

\[ H(P) = \left\| \frac{\partial h_{ijk}(P)}{\partial P_{x_{ik}}} \right\|_{(r-1)\times(r-1)} \]  

(2.13)

Other statistics which can be used to test the hypothesis are the chi-square and the chi-one square statistics, based on the final step iteration estimates \( \hat{P}_{ijk}(N) \), defined as follows:

\[ \chi^2 = \sum (n_{ijk}/P_k) \sum_i \left( \hat{P}_{ijk} - \hat{P}_{ijk}(N) \right)^2 / \hat{P}_{ijk}(N) \]  

(2.14)

\[ \chi^2_1 = \sum (n_{ijk}/P_k) \sum_i \left( \hat{P}_{ijk} - \hat{P}_{ijk}(N) \right)^2 / \hat{P}_{ijk}. \]  

(2.15)

Garza-Hernandez and McCarthy [4] have proposed a maximum likelihood ratio statistic:

\[ G_1 = (\hat{q}' \Gamma)(\Gamma \Sigma \Gamma)^{-1}(\hat{q}' \Gamma)' \]  

(2.16)

where \( \hat{q}' = (\hat{q}_{11}, \ldots, \hat{q}_{r-1,s-1}) \), with \( \hat{q}_{ij} = \hat{P}_{ij} - \hat{P}_{ijk}(N) \). \( \Gamma \) is an \( r(s-1) \times (r-1)(s-1) \) matrix such that \( \Gamma = 0 \) is equivalent to the null hypothesis (2.1) and \( \Sigma \) is an estimate of the variance-covariance matrix of \( \hat{q} \). It should be noted that this statistic is not invariant under transposition of the contingency table (exchanging rows and columns), as are the other statistics.

All the foregoing statistics are asymptotically distributed chi square with \( (r-1)(s-1) \) degrees of freedom, under the null hypothesis.

3. THE ASYMPTOTIC POWER

Bhapkar [2] has pointed out that (2.10) is algebraically identical with Wald’s [8] large sample statistic, (if all \( n_{ijk} > 0 \), and conjectured that Wald’s asymptotic results for independent identically distributed variables hold also for the case of several different independent multinomial distributions. This conjecture can be shown to hold for the situation considered here by considering the following modified sampling procedure (“Procedure B”): For each of the \( n \) sample units, a stratum \( k \) is assigned independently with probability \( r_k \) (\( \sum r_k = 1 \)). Given the stratum assignments, the sample unit is selected at random (with replacement) from stratum \( k \).

The joint probability function of \( \{n_{ijk}\} \) is then:

\[
\psi_B(\{n_{ijk}\} | \{P_{ijk}\}) = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} \left( \frac{P_{ijk}}{P_k} r_k \right)^{n_{ijk}}. 
\]  

(3.1)

It is easy to see that the unrestricted maximum likelihood estimates of \( P_{ijk} \) for Procedure B are the same as for Procedure A, as given by (2.4), except that the \( n_k \) are random variables.

Furthermore:

\[
\frac{\partial^2 \ln \psi_A}{\partial P_{ijk} \partial P_{i'j'k'}} = - \delta_k^{i'j'} \left[ \frac{n_{ijk}}{P_k} + \delta_i \delta_j \frac{1}{P_{ijk}} \right]. 
\]  

(3.2)

Thus:

\[
E_p \left[ \frac{\partial^2 \ln \psi_A}{\partial P_{ijk} \partial P_{i'j'k'}} \right] = - \frac{n_{ijk}}{P_k} \delta_k^{i'j'} \left[ \frac{1}{P_k} + \delta_i \delta_j \right] \frac{1}{P_{ijk}}. 
\]  

(3.3)

and

\[
E_p \left[ \frac{\partial^2 \ln \psi_B}{\partial P_{ijk} \partial P_{i'j'k'}} \right] = - \frac{n_{ijk}}{P_k} \delta_k^{i'j'} \left[ \frac{1}{P_k} + \delta_i \delta_j \right] \frac{1}{P_{ijk}}. 
\]  

(3.4)

If \( n_k = n r_k \) for Procedure A it follows that (2.12), and thus also the test statistic (2.10), are algebraically identical for the two sampling procedures. The sample distribution of the statistic (2.10) is however in general not identical for the two sampling procedures. In fact, the distribution of \( T \) under Procedure A is the conditional distribution of \( T \) under Procedure B, given the values of \( n_k \).

However, if \( n_k = n r_k \), the limit as \( n \to \infty \), of the conditional distribution is the same as the limit of the unconditional distribution of \( T \) under Procedure B.

This can be seen by noting that \( T \) depends on \( \{n_{ijk}\} \) only through the values of \( \{\hat{P}_{ijk} - P_{ijk}\} \). The conditional distribution of \( \{n_{ijk}\} \), given \( n_k \), is that of the product of \( t \) independent multinomial distributions with parameters \( \{n_k; P_{ijk}/P_k\} \). Let \( Z_k = n_k/n \). Then the conditional distribution of \( \{\sqrt{n_k} (\hat{P}_{ijk} - P_{ijk})\} \), given \( Z_k = z_k \), is asymptotically multinormal with zero means and variance-covariance matrix:

\[
\left[ (\delta_k^{i'j'}/z_k) [\delta_{ij} P_{ijk} - P_{ijk} P_{i'j'k'}] \right]. 
\]  

Since the vector random variables \( \{Z_k\} \) converge in probability to \( \{r_k\} \) as \( n \to \infty \), it is easy to see that the unconditional distribution of \( \{\sqrt{n_k} (\hat{P}_{ijk} - P_{ijk})\} \) is asymptotically equal to their conditional distribution, given \( Z_k = z_k \) (i.e., \( n_k = n r_k \)). Thus the distribution of \( T \) for Procedure A is asymptotically equal to the distribution of \( T \) for Procedure B.
The log-likelihood ratio statistic $G$, given by (2.3), can easily be seen to be algebraically equivalent under both sampling procedures (in the same way as was shown for $T$). Moreover Neyman [7] has shown that the probability that they contradict each other tends to zero as $n\to \infty$, irrespective of the true parameters, both under sampling Procedure A and B. This implies that $G$ and $T$ have the same asymptotic distribution under either sampling Procedure A or B, and, according to the preceding statements, that they have the same asymptotic distribution for both sampling procedures, for any set of parameters.

Sampling Procedure B fulfills the conditions required for Wald’s results [8], since the likelihood (3.1) can be expressed as

$$\phi([x_{ik}] | \{P_{ijk}\}) = \prod_{k=1}^{K} \prod_{i,j,k} \left( P_{ijk} \right)^{(t)} / P_{k}^{(t)} ,$$

where $x_{ik}^{(t)} = 1$ if the $i$th sample item is in stratum $k$ and in the $(i,j)$th cell ($t = 1, \ldots, n$), and zero otherwise. Then any random variable $X_{ijk}$ is independently and identically distributed. But for this case Wald [8] has shown that they have the same asymptotic distribution for both $T$ and $G$.

Neyman [7] has also shown that the chi-square and chi-one square tests defined by (2.14) and (2.15) are asymptotically equivalent to the log likelihood ratio test of $n-I$. This is the value of $T$, given by (2.10), with $P$ replaced by $P$, the true parameter vector.

Neyman [7] has shown that the chi-square and chi-one square tests defined by (2.14) and (2.15) are asymptotically equivalent to the log likelihood ratio test defined by (2.3), so that they too have the same asymptotic power as $G$ or $T$.

4. The Simulations Performed

To compare the powers of the statistics with the asymptotic power for the case of a finite size sample, four sets of parameters $\{P_{ijk}\}$ for a three-strata $2 \times 3$ contingency table were used ($r = 3$, $s = 2$, $t = 3$)—all with the same marginal probabilities $P_{ik}, P_{jk}$, $P_{j}$. The first set, which fulfilled the null hypothesis—$H_0$: $P_{ijk} = P_{ik} P_{jk}$, was the final iteration solution for maximum likelihood estimates obtained from the observed frequencies of the example used by Garza-Hernandez [4] (and also in [6]). The other three sets, which represent alternative hypotheses, denoted by $H_1$, $H_2$, and $H_3$, were obtained from the null hypothesis by changing the values of $P_{ijk}$, keeping marginal probabilities $P_{ik}, P_{jk}$. The following values of $H = .0025, .0053, .0253$ for the information measure:

$$H = -2 \sum_{i,j} P_{ij} \ln P_{ij} - \sum_{i} P_{i.} \ln P_{i.} - \sum_{j} P_{.j} \ln P_{.j} .$$

The values of $P_{ijk}$ used for each of the hypotheses are given in Table 1. They cover the important range of “interesting” alternative hypotheses for this case, since more extreme hypotheses produce samples which in general would show obvious departures from independence, by inspection only.

In each simulation the values of $n_{ijk}$ ($i = 1, \ldots, r$; $j = 1, \ldots, s$) were generated within each stratum ($k = 1, \ldots, t$) from the multinomial distribution with the parameters ($n_k$, $\{P_{ijk} / P_k\}$), where the values of $n_k$ were fixed ($n_2 = 92$, $n_3 = 112$, $n_4 = 78$, as in the example data used in [4]) and values of $P_{ijk}$ were determined as previously mentioned for each hypothesis. One thousand simulations were made for $H_0$ and 500 for each of $H_1$, $H_2$, and $H_3$. For each set of $n_{ijk}$ values generated, the iteration procedure of [6] was executed and the five statistics defined in Section 2 by (2.3), (2.10), (2.14), (2.15) and (2.16) were computed. Each statistic was compared with the critical chi-square value for significance levels $\alpha_1 = .01$, $\alpha_2 = .05$ and $\alpha_3 = .10$.

For each of the four hypotheses, the relative frequencies of the number of times that each of the five statistics exceeded the critical chi-square values, for each of the three significance levels, were computed. The results are given in Table 2 and they estimate the probabilities of rejecting $H_0$.
the null hypothesis using each of the five statistics, for each of the three nominal significance levels and for each of the four hypotheses. In addition, the common asymptotic power of the first four statistics was computed from the results of the previous section and also noted in Table 2 together with the standard errors of estimates based on the given number of simulations.

5. ANALYSIS OF RESULTS

Table 2 shows that while the probabilities of rejecting $H_0$ using the chi-one square statistic (2.14) somewhat exceed, but are close to, the nominal significance levels, these probabilities are much lower than the nominal levels for the other statistics.

The estimates of power, obtained from the relative frequencies of exceeding the critical chi-square values for the alternative hypothesis, are obviously affected by the estimated true significance levels attained. Thus, the chi-one square statistic with a high estimated true significance level, retains apparent superiority of power under all the alternative hypotheses for all nominal levels of significance.

For all five statistics the asymptotic power (of the first four statistics) is well approximated by the empirical simulation power.

To eliminate the effect of the different true levels of significance of the statistics, unbiased estimates of the Expected Significance Levels (E.S.L.), proposed by Dempster and Schatzoff [3], were computed, for each alternative hypothesis. The estimated E.S.L. is the Mann-Whitney statistic based on the number of times each value of the statistic under the alternative hypothesis exceeds a value of the same statistic under the null hypothesis. The E.S.L. values measure the relative efficiencies of the statistics independently of the true significance levels attained. The estimated values, given in Table 3, show very small differences in the E.S.L. Simultaneous test procedures for testing differences between the E.S.L. values of any subset of the five statistics, based on tests proposed by Gabriel [5], were applied to 100 simulation runs for each of the three alternative hypotheses and no pair of statistics showed a significant difference (at the five percent level).

For the specific data used and under the specified alternative hypotheses for which simulations were made, it seems that no single statistic can be shown to be the best, as indicated also by the asymptotic result. Thus the chi-one square statistic of Bhapkar [1] could serve at least as well as any other, so that the additional iterations needed to approximate the maximum likelihood estimates, as proposed by the author in [6], are not really necessary.

[Received June 1971. Revised April 1972.]

REFERENCES


