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} $(i=1, \cdots, r;$ $j=1, \cdots, s;$ $k=1, \cdots, t)$ 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 $P_k = \Sigma_{i,j} P_{ijk}$ assumed known and $\Sigma_k P_k = 1$. The hypothesis of interest is:

$$H_o: P_{ij.} = P_{i..} P_{.j.} \quad (i=1, \cdots, r;\ j=1, \cdots, s). \tag{2.1}$$

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 kth 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 $\Sigma_{i,j} n_{ijk} = n_k$). The joint probability function of $\{n_{ijk}\}$, given the values of $\{P_{ijk}\}$, is then:

$$\psi_A[\{n_{ijk}\} \mid \{P_{ijk}\}] = \Pi_k \left[\frac{n_k!}{\prod_{i,j} n_{ijk}!} \Pi_{i,j} \left(\frac{P_{ijk}}{P_k} \right)^{n_{ijk}} \right]. \tag{2.2}$$

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

$$G = 2\Sigma_{ijk} n_{ijk} \ln[\hat{P}_{ijk}/\hat{\hat{P}}_{ijk}(N)], \tag{2.3}$$

to test the hypothesis (2.1), where:

$$\hat{P}_{ijk} = (n_{ijk}/n_k) P_k \tag{2.4}$$

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

* Gad Nathan is visiting associate professor, Department of Biostatistics, University of North Carolina at Chapel Hill, on leave of absence from the Hebrew University, Jerusalem and the Central Bureau of Statistics, Israel. The research was supported by National Center for Health Statistics Grant NCHS-IS-1. This article has benefited from several helpful suggestions made by Professor K. R. Gabriel and E. Samuel-Cahn. Thanks are due to Israel Sauerbrun for his Fortran programming and execution of the simulation and iteration programmes on the Hebrew University's CDC 6400 computer.

© Journal of the American Statistical Association
December 1972, Volume 67, Number 340
Theory and Methods Section

$$\chi_1^2 = \Sigma_k(n_k/P_k)\Sigma_{i,j}[(\hat{P}_{ijk} - P_{ijk})^2/\hat{P}_{ijk}], \quad (2.5)$$

subject to a linearization of the null hypothesis (2.1), as proposed by Bhapkar [1].

If we define:

$$\boldsymbol{P}' = (P_{111}, \cdots, P_{rst}); \quad (2.6)$$

$$\hat{\boldsymbol{P}}' = (\hat{P}_{111}, \cdots, \hat{P}_{rst}); \quad (2.7)$$

$$\boldsymbol{h}(\boldsymbol{P}) = [h_{11}(\boldsymbol{P}), \cdots, h_{r-1,s-1}(\boldsymbol{P})]'; \quad (2.8)$$

where

$$h_{ij}(\boldsymbol{P}) = P_{ij.} - P_{i..}P_{.j.}; \quad (2.9)$$

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

$$T = [\boldsymbol{h}(\hat{\boldsymbol{P}})]'\boldsymbol{C}^{-1}[\boldsymbol{h}(\hat{\boldsymbol{P}})], \quad (2.10)$$

where:

$$\boldsymbol{C} = n^{-1}[\boldsymbol{H}(\hat{\boldsymbol{P}})\boldsymbol{B}^{-1}(\hat{\boldsymbol{P}})\boldsymbol{H}'(\hat{\boldsymbol{P}})], \quad (2.11)$$

$$\boldsymbol{B}(\boldsymbol{P}) = \frac{1}{n}\left\| E_{\boldsymbol{P}}\left(\frac{\partial^2 \ln \psi}{\partial P_{ijk}\partial P_{i'j'k'}}\right)\right\|_{(rs-1)t\times(rs-1)t} \quad (2.12)$$

and

$$\boldsymbol{H}(\boldsymbol{P}) = \left\|\frac{\partial h_{ij}(\boldsymbol{P})}{\partial P_{xyk}}\right\|_{(r-1)(s-1)\times(rs-1)t}. \quad (2.13)$$

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:

$$\chi^2 = \Sigma_k(n_k/P_k)\Sigma_{i,j}[\hat{P}_{ijk} - \hat{\hat{P}}_{ijk}(N)]^2/\hat{\hat{P}}_{ijk}(N) \quad (2.14)$$

$$\chi_1^2 = \Sigma_k(n_k/P_k)\Sigma_{i,j}[\hat{P}_{ijk} - \hat{\hat{P}}_{ijk}(N)]^2/\hat{P}_{ijk}. \quad (2.15)$$

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

$$G_1 = (\hat{\boldsymbol{q}}'\boldsymbol{\Gamma})(\boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma})^{-1}(\hat{\boldsymbol{q}}'\boldsymbol{\Gamma})' \quad (2.16)$$

where $\hat{\boldsymbol{q}}' = (\hat{q}_{11}, \cdots, \hat{q}_{r,s-1})$, with $\hat{q}_{ij} = \hat{P}_{ij.}/\hat{P}_{i..}$, $\boldsymbol{\Gamma}$ is an $r(s-1)\times(r-1)(s-1)$ matrix such that $\boldsymbol{q}'\boldsymbol{\Gamma} = 0$ is equivalent to the null hypothesis (2.1) and $\boldsymbol{\Sigma}$ is an estimate of the variance-covariance matrix of $\hat{\boldsymbol{q}}$. 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 ($\Sigma_k r_k = 1$). Given the stratum assignment k, the sample unit is selected at random (with replacement) from stratum k.

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

$$\psi_B[\{n_{ijk}\} \mid \{P_{ijk}\}] = \frac{n!}{\prod_{i,j,k} n_{ijk}!}\prod_{i,j,k}\left(\frac{P_{ijk}}{P_k}r_k\right)^{n_{ijk}}. \quad (3.1)$$

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:

$$\frac{\partial^2 \ln \psi_A}{\partial P_{ijk}\partial P_{i'j'k'}} = \frac{\partial^2 \ln \psi_B}{\partial P_{ijk}\partial P_{i'j'k'}} = -\delta_k^{k'}\left[\frac{n_{rsk}}{P_{rsk}^2} + \delta_i^{i'}\delta_j^{j'}\frac{n_{ijk}}{P_{ijk}^2}\right]. \quad (3.2)$$

Thus:

$$E_{\boldsymbol{P}}\left[\frac{\partial^2 \ln \psi_A}{\partial P_{ijk}\partial P_{i'j'k'}}\right] = -\frac{n_k}{P_k}\delta_k^{k'}\left[\frac{1}{P_{rsk}} + \delta_i^{i'}\delta_j^{j'}\frac{1}{P_{ijk}}\right] \quad (3.3)$$

and

$$E_{\boldsymbol{P}}\left[\frac{\partial^2 \ln \psi_B}{\partial P_{ijk}\partial P_{i'j'k'}}\right] = -\frac{nr_k}{P_k}\delta_k^{k'}\left[\frac{1}{P_{rsk}} + \delta_i^{i'}\delta_j^{j'}\frac{1}{P_{rsk}}\right]. \quad (3.4)$$

If $n_k = nr_k$ 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 $n_k = nr_k$, 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 $\{\hat{P}_{ijk}\} = \{(n_{ijk}/n_k)P_k\}$. 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 $\{\sqrt{n}[\hat{P}_{ijk} - P_{ijk}]\}$, given $Z_k = z_k$, is asymptotically multinormal with zero means and variance-covariance matrix:

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

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 $\{\sqrt{n}[\hat{P}_{ijk} - P_{ijk}]\}$ is asymptotically equal to their conditional distribution, given $Z_k = r_k$ (i.e., $n_k = nr_k$). 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

$$\phi[\{x_{ijk}^{(\ell)}\} \mid \{P_{ijk}\}] = \prod_{\ell=1}^{n}\left[\prod_{i,j,k}\left(\frac{P_{ijk}}{P_k}r_k\right)^{x_{ijk}^{(\ell)}}\right], \quad (3.5)$$

where $x_{ijk}^{(\ell)}=1$ if the ℓth sample item is in stratum k and in the (i, j)th cell $(\ell=1, \cdots, n)$, and zero otherwise. Thus the random variables:

$\mathbf{x}^{(\ell)}=(x_{111}^{(\ell)}, \cdots, x_{rst}^{(\ell)})$ are independently and identically distributed. But for this case Wald [8] has shown that both T and G are asymptotically distributed noncentral chi square, with $(r-1)(s-1)$ degrees of freedom and with noncentrality parameter:

$$\lambda = n[\boldsymbol{h}(\boldsymbol{P})]'[\boldsymbol{H}(\boldsymbol{P})\boldsymbol{B}^{-1}(\boldsymbol{P})\boldsymbol{H}'(\boldsymbol{P})]^{-1}[\boldsymbol{h}(\boldsymbol{P})]. \quad (3.6)$$

(Note that λ is independent of n, since $\boldsymbol{B}(\boldsymbol{P})$ has a factor of n^{-1}). This is the value of T, given by (2.10), with $\hat{\boldsymbol{P}}$ replaced by $\boldsymbol{P}$, the true parameter vector.

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 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 2×3 contingency table were used $(r=3, s=2, t=3)$—all with the same marginal probabilities $P_{i.k}$, $P_{.jk}$. The first set, which fulfilled the null hypothesis—H_o: $P_{ij.}=P_{i..}P_{.j.}$, 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 $P_{i.k}$, $P_{.jk}$ fixed, so as to obtain for H_1, H_2, H_3, respectively, values of $H=.0025$, .0053, .0253 for the information measure:

$$H = -2[\textstyle\sum_{i,j} P_{ij.}\ln P_{ij.} - \sum_i P_{i..}\ln P_{i..} - \sum_j P_{.j.}\ln P_{.j.}].$$

1. VALUES OF THE PARAMETERS—P_{ijk}—FOR EACH OF THE ALTERNATIVES

Stratum	j	Total[a]	H_0		H_1		H_2		H_3	
			i=1	i=2	i=1	i=2	i=1	i=2	i=1	i=2
	Total*	1.00000	.32740	.67260	.32740	.67260	.32740	.67260	.32470	.67260
Total	1	.29640	.09700	.19940	.08683	.20957	.08172	.21468	.06512	.23128
	2	.52360	.17140	.35220	.18157	.34203	.18157	.34203	.20328	.32032
	3	.1800	.05900	.12100	.05900	.12100	.06411	.11589	.05900	.12100
	Total*	.17080	.03360	.13720	.03360	.13720	.03360	.13720	.03360	.13720
K=1	1	.05210	.01870	.03340	.01696	.03514	.01609	.03601	.01325	.03885
	2	.09560	.01160	.08400	.01334	.08226	.01334	.08226	.01705	.07855
	3	.02310	.00330	.01980	.00330	.01980	.00417	.01893	.00330	.01980
	Total*	.34640	.16840	.17800	.16840	.17800	.16840	.17800	.16840	.17800
K=2	1	.10780	.05280	.05500	.04928	.05852	.04751	.06029	.04176	.06604
	2	.18750	.09520	.09230	.09872	.08878	.09872	.08878	.10624	.08126
	3	.05110	.02040	.03070	.02040	.03070	.02217	.02893	.02040	.03070
	Total*	.48280	.12540	.35740	.12540	.35740	.12540	.35740	.12540	.35740
K=3	1	.13650	.02550	.11100	.02059	.11591	.01812	.11838	.01011	.12639
	2	.24050	.06460	.17590	.02951	.17099	.06951	.17099	.07999	.16051
	3	.10580	.03530	.07050	.03530	.07050	.03777	.06803	.03530	.07050

[a] Marginal probabilities are equal for all alternatives.

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} $(i=1, \cdots, r;$ $j=1, \cdots, s)$ were generated within each stratum $(k=1, \cdots, t)$ from the multinomial distribution with the parameters $(n_k, \{P_{ijk}/P_k\})$, where the values of n_k were fixed $(n_1=92, n_2=112, n_3=78$, 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_o 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

2. RELATIVE FREQUENCIES OF TIMES NOMINAL SIGNIFICANCE LEVELS EXCEEDED

Hypothesis	α	Simulation - estimated power: Bhapkar T	Maximum likelihood χ^2	Maximum likelihood χ^2	Maximum likelihood G	Garza-Hernandez G_1	Asymptotic power	Standard error[a]
H_0	.01	.005	.005	.011	.005	.006	.010	.033
(1,000 iterations)	.05	.044	.036	.058	.036	.047	.050	.007
H = .0000	.10	.092	.090	.105	.092	.097	.100	.009
H_1	.01	.036	.016	.044	.028	.038	.027	.007
(500 iterations)	.05	.096	.080	.116	.092	.100	.098	.013
H = .0025	.10	.187	.166	.200	.174	.188	.170	.017
H_2	.01	.056	.048	.076	.048	.064	.053	.010
(500 iterations)	.05	.174	.154	.200	.160	.190	.157	.016
H = .0053	.10	.266	.248	.288	.250	.276	.251	.019
H_3	.01	.378	.326	.418	.332	.388	.396	.022
(500 iterations)	.05	.638	.592	.674	.596	.654	.647	.021
H = .0253	.10	.762	.742	.782	.742	.774	.766	.019

[a] Computed as $\sqrt{P(1-P)/N}$, where P is the asymptotic power and N the number of simulations.

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_o 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.

3. ESTIMATES OF EXPECTED SIGNIFICANCE LEVELS (E.S.L.)

Statistic	H_1 (H = .0025)	H_2 (H = .0053)	H_3 (H = .0253)
Minimal Chi-one-square (Bhapkar) - T	.5472	.6446	.9099
Max. Likelihood - chi-square - χ^2	.5424	.6417	.9034
Max. Likelihood - chi-one-square - χ_1^2	.5472	.6451	.9065
Max. Likelihood - log-likelihood ratio - G	.5425	.6394	.9049
Garza-Hernandez - log-likelihood ratio - G_1	.5460	.6445	.9099

[*Received June 1971. Revised April 1972.*]

REFERENCES

[1] Bhapkar, V. P., "Some Tests for Categorical Data," *Annals of Mathematical Statistics*, 32 (February 1961), 72–83.

[2] Bhapkar, V. P., "A Note on the Equivalence of Two Test Criteria for Hypotheses in Categorical Data," *Journal of the American Statistical Association*, 61 (March 1966), 228–35.

[3] Dempster, Arthur P. and Schatzoff, Martin, "Expected Significance Level as a Sensitivity Index for Test Statistics," *Journal of the American Statistical Association*, 60 (June 1965), 420–36.

[4] Garza-Hernandez, Tomas and McCarthy, Philip J., "A Test of Homogeneity for a Stratified Sample," *American Statistical Association—Proceedings of the Social Statistics Section*, 1962, 200–2.

[5] Gabriel, K. Ruben, "Simultaneous Test Procedures—Some Theory of Multiple Comparisons," *Annals of Mathematical Statistics*, 40 (April 1969), 224–49.

[6] Nathan, Gad, "Tests of Independence in Contingency Tables from Stratified Samples," in N. L. Johnson and H. Smith, eds., *New Developments in Survey Sampling*, New York: John Wiley and Sons, Inc., 1969, 578–600.

[7] Neyman, Jerzy, "Contribution to the Theory of the χ^2 Test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1949, 239–73.

[8] Wald, Abraham, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54 (November 1943), 426–82.