The Annals of Statistics
1989, Vol. 17, No. 4, 1709–1722

BOUNDS ON THE SIZE OF THE χ^2-TEST OF INDEPENDENCE IN A CONTINGENCY TABLE[1]

BY WEI-YIN LOH

University of Wisconsin-Madison

Bounds are obtained on the limiting size of the level-α χ^2-test of independence in a contingency table, as the sample size increases. The situations considered include (a) sampling with one or both sets of marginal totals random, (b) performing the test with or without the continuity correction and (c) with or without conditioning on the event $\mathscr{E}_k$ that the minimum estimated expected cell count is greater than a given $k \geq 0$. Bounds for both the unconditional and conditional (on $\mathscr{E}_k$) size are derived. It is shown, for example, that the limiting conditional size of the test is unity for all α if the continuity correction is used with $k = 0$ and sampling is done with both margins random. The same conclusion holds if sampling is done with one set of margins fixed and the dimensions of the table are not too small.

1. Introduction. Let Y_{ij}, $i = 1, \ldots, r$; $j = 1, \ldots, c$, be the observed counts in the (i, j)-cell of a contingency table with r rows and c columns. Let p_{ij} denote the probability that an observation belongs to the (i, j)-cell, let

$$p_{i\cdot} = \sum_j p_{ij}, \qquad p_{\cdot j} = \sum_i p_{ij}$$

and let

$$R_i = \sum_{j=1}^{c} Y_{ij}, \qquad C_j = \sum_{i=1}^{r} Y_{ij}, \qquad n = \sum_i R_i = \sum_j C_j$$

denote the ith row, jth column and grand totals, respectively. Consider the hypothesis H_0 that the rows and columns of the contingency table are independent, and let $E_{ij} = R_i C_j n^{-1}$ denote the estimated expected count in the (i, j)-cell under H_0. When $E_{ij} > 0$ for every cell, a standard test of H_0 at nominal level α is the Pearson χ^2-test which rejects H_0 if $X^2 > \chi^2_{\nu, \alpha}$, where $\chi^2_{\nu, \alpha}$ is the upper α-quantile of the χ^2_ν-distribution, $\nu = (r-1)(c-1)$ and

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} (Y_{ij} - E_{ij})^2 / E_{ij}. \tag{1}$$

There has been much controversy about this test. One issue is whether the "continuity corrected" statistic [Yates (1934)]

$$X_c^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left(|Y_{ij} - E_{ij}| - \tfrac{1}{2}\right)^2 / E_{ij} \tag{2}$$

Received July 1987; revised July 1988.

[1]Research supported by NSF Grants DMS-85-02303 and DMS-88-03271, AFOSR Grant ISSA-86-0068, and an IBM Junior Faculty Research Fellowship at the T. J. Watson Research Center.

AMS 1980 *subject classifications*. Primary 62F05; secondary 62H17.

Key words and phrases. Continuity correction, multinomial sampling.

should be used in place of (1). Pearson (1947), Plackett (1964) and Grizzle (1967) criticize the use of X_c^2 on the grounds that it tends to yield an overly conservative test. Mantel and Greenhouse (1968) and Yates (1984) defend the continuity correction with the argument that probabilities should be computed conditional on the marginal totals, and that X^2 can greatly exaggerate conditional significance, especially for tables with one or more small marginal values. Some computer-generated finite-sample results are reported in Conover (1974), Garside and Mack (1976) and Haber (1980). All these papers are chiefly concerned with the 2×2 table.

The values of X^2 and X_c^2 are undefined if one or more of the row or column totals is 0, or equivalently, if $\min E_{ij} = 0$. To improve the accuracy of the χ^2 approximation, it is usually recommended that either test be carried out only if the event

$$\mathscr{E}_k = \left\{ \min_{ij} E_{ij} > k \right\}$$

occurs, where k is a nonnegative number. Opinions differ on the smallest permissible value of k. Values such as $k = 5$ [Fisher (1925)] and $k = 1$ [Snedecor and Cochran (1980), page 77] have been proposed.

The purpose of this paper is to obtain bounds on the limiting size of the two tests under the null hypothesis for arbitrary values of k, r and c. For any $k \geq 0$, we define the *conditional size* of the two tests as

$$\alpha_n(k) = \sup_{H_0} \Pr\left(X^2 > \chi^2_{\nu,\alpha} | \mathscr{E}_k \right),$$

$$\alpha_n^{(c)}(k) = \sup_{H_0} \Pr\left(X_c^2 > \chi^2_{\nu,\alpha} | \mathscr{E}_k \right)$$

and the *unconditional size* as

$$\bar{\alpha}_n(k) = \sup_{H_0} \Pr\left(\left\{ X^2 > \chi^2_{\nu,\alpha} \right\} \cap \mathscr{E}_k \right),$$

$$\bar{\alpha}_n^{(c)}(k) = \sup_{H_0} \Pr\left(\left\{ X_c^2 > \chi^2_{\nu,\alpha} \right\} \cap \mathscr{E}_k \right).$$

Two sampling models are considered, namely:

1. One set of marginal totals fixed, the other random.
2. Both sets of marginal totals random (multinomial sampling).

Lower bounds for the limiting size are obtained by computing rejection probabilities under a sequence of null hypothesis distributions and using the resulting Poisson behavior of the cell counts. One conclusion is that under model 2, $\lim_{n \to \infty} \alpha_n^{(c)}(0) = 1$ for all $0 < \alpha < 1$. The same property holds under model 1 if the dimensions of the table are sufficiently large. Upper bounds are derived for finite-sample sizes through a Chebyshev-type argument. Practical implications of the results are discussed in the last section of the paper.

For results on the behavior of the χ^2 for testing goodness-of-fit for multinomial distributions, the reader is referred to Cressie and Read (1984), Kallenberg

(1985), Larntz (1978) and Yarnold (1970). Haberman (1988), Oosterhoff (1985) and Zelterman (1987) examine the power of the test in large sparse multinomial distributions.

2. One set of marginal totals fixed. We first consider the case of sampling with one set of marginal totals fixed. Without loss of generality, suppose that $C_j = f_j n$, $j = 1, \dots, c$, for some positive constants f_j's summing to 1. We wish to test the hypothesis

$$H_0\colon p_{i1} = p_{i2} = \cdots = p_{ic} = \pi_i, \qquad i = 1, \dots, r,$$

for some unknown $\{\pi_i\}$ such that $0 < \pi_i < 1$, $\sum_i \pi_i = 1$. Suppose that the test is only carried out if the event $\mathscr{E}_k$ occurs. This implies that $\min R_i \geq 1$.

2.1. *Without continuity correction.* Because the conditional distribution of Y_{ij}, given $R_i = m \geq 1$ and H_0, is hypergeometric with

$$E(Y_{ij} | R_i = m) = f_j m,$$

$$\operatorname{var}(Y_{ij} | R_i = m) = f_j(1 - f_j) m (n - m)(n - 1)^{-1},$$

we have for any set $\{m_i \geq 1\}$ such that $\sum_{i=1}^r m_i = n$,

$$\begin{aligned}
&\Pr\{X^2 > \chi^2_{\nu,\alpha} | R_i = m_i, i = 1, \dots, r\} \\
&\quad = \Pr\left\{ \sum_{i=1}^r \sum_{j=1}^c (f_j m_i)^{-1} (Y_{ij} - f_j m_i)^2 > \chi^2_{\nu,\alpha} \middle| R_i = m_i, i = 1, \dots, r \right\} \\
&\quad \leq \chi^{-2}_{\nu,\alpha} E\left\{ \sum_{i=1}^r \sum_{j=1}^c (f_j m_i)^{-1} (Y_{ij} - f_j m_i)^2 \middle| R_i = m_i, i = 1, \dots, r \right\} \\
&\quad = (n-1)^{-1} \chi^{-2}_{\nu,\alpha} \sum_{i=1}^r \sum_{j=1}^c (1 - f_j)(n - m_i) \\
&\quad = n(n-1)^{-1} \nu \chi^{-2}_{\nu,\alpha}.
\end{aligned}$$

Therefore

$$\Pr\left(\{\chi^2 > \chi^2_{\nu,\alpha}\} \cap \mathscr{E}_k\right) \leq n(n-1)^{-1} \nu \chi^{-2}_{\nu,\alpha} \Pr\{\mathscr{E}_k\}$$

and we obtain the following theorem.

THEOREM 1. *Under sampling model* 1,

$$\Pr\left(\{X^2 > \chi^2_{\nu,\alpha}\} \cap \mathscr{E}_k\right) \leq \Pr\left(X^2 > \chi^2_{\nu,\alpha} | \mathscr{E}_k\right) \leq n(n-1)^{-1} \nu \chi^{-2}_{\nu,\alpha}$$

for all $0 < \alpha < 1$, $k \geq 0$ *and all* $\{f_j\}$.

COROLLARY 1. *Under the assumptions of Theorem* 1,

$$\limsup_{n\to\infty} \alpha_n(k) \leq \nu \chi^{-2}_{\nu,\alpha} \quad \textit{and} \quad \limsup_{n\to\infty} \bar{\alpha}_n(k) \leq \nu \chi^{-2}_{\nu,\alpha} \tag{3}$$

for all $0 < \alpha < 1$ *and* $k \geq 0$.

When ν is large and $0 < \alpha < \frac{1}{2}$, the upper bound (3) is close to 1 because

$$\chi^2_{\nu,\alpha} \doteq \nu\left\{1 + z_\alpha\sqrt{2/(9\nu)} - 2(9\nu)^{-1}\right\}^3 \quad \text{as } \nu \to \infty,$$

where z_α is the upper α-quantile of the standard normal distribution. [This is the Wilson–Hilferty approximation; see, e.g., Johnson and Kotz (1970), pages 176–177.]

To obtain a lower bound, suppose that H_0 holds with

$$(4) \quad \pi_i = \omega n^{-1}, \qquad i = 1,\dots, r-1; \qquad \pi_r = 1 - (r-1)\omega n^{-1}, \qquad 0 < \omega < n.$$

Then $Y_{ij} \to_D S_{ij}$ as $n \to \infty$ for $i = 1,\dots, r-1$ and $j = 1,\dots, c$, where $\{S_{ij}\}$ is a set of $c(r-1)$ independent Poisson variables, with S_{ij} having mean $f_j\omega$ [see, e.g., Johnson and Kotz (1969), page 297]. (As usual, "$\to_D$" and "$\to_P$" denote convergence in distribution and convergence in probability, respectively.) Let

$$f_* = \min_{1\le j\le c} f_j \quad \text{and} \quad T_i = \sum_{j=1}^{c} S_{ij}, \qquad i = 1,\dots, r-1.$$

Because the conditional distribution of $\{S_{i1},\dots, S_{ic}\}$ given $T_i = t_i$ is multinomial with t_i trials and success probabilities $\mathbf{f} = (f_1,\dots, f_c)$, we have

$$\begin{aligned}
&\Pr\left(\{X^2 > \chi^2_{\nu,\alpha}\} \cap \mathscr{E}_k\right)\\
&\quad\ge \Pr\left(\left\{\sum_{i=1}^{r-1}\sum_{j=1}^{c}(Y_{ij} - f_jR_i)^2/(f_jR_i) > \chi^2_{\nu,\alpha}\right\} \cap \left\{\min_{1\le i\le r} R_i > kf_*^{-1}\right\}\right)\\
&\quad\to \Pr\left(\left\{\sum_{i=1}^{r-1}\sum_{j=1}^{c}(S_{ij} - f_jT_i)^2/(f_jT_i) > \chi^2_{\nu,\alpha}\right\} \cap \left\{\min_{1\le i< r} T_i > kf_*^{-1}\right\}\right)\\
&\quad= \sum_{\min t_i > kf_*^{-1}} \Pr\left\{\sum_{i=1}^{r-1}\sum_{j=1}^{c}(S_{ij} - f_jt_i)^2/(f_jt_i) > \chi^2_{\nu,\alpha} \,\middle|\, T_1 = t_1,\dots, T_{r-1} = t_{r-1}\right\}\\
&\qquad\qquad \times \Pr\{T_1 = t_1,\dots, T_{r-1} = t_{r-1}\}\\
&\quad= H(\alpha, \mathbf{f}, \omega, k), \quad \text{say},
\end{aligned}$$

where $H(\alpha, \mathbf{f}, \omega, k)$ may be written explicitly as

$$(5) \qquad H(\alpha, \mathbf{f}, \omega, k) = e^{-(r-1)\omega} \sum_{\min t_i > kf_*^{-1}} \omega^{t.} \sum_{s\in\mathscr{A}(t)} \prod_{j=1}^{c} \frac{f_j^{s_{.j}}}{\prod_{i=1}^{r-1}(s_{ij}!)},$$

$$\mathbf{t} = (t_1,\dots, t_{r-1}), \qquad \mathbf{s} = \{s_{ij}: i = 1,\dots, r-1;\ j = 1,\dots, c\},$$

$$t. = \sum_{i=1}^{r-1} t_i, \qquad s_{.j} = \sum_{i=1}^{r-1} s_{ij}$$

and

$$\mathscr{A}(\mathbf{t}) = \left\{\mathbf{s}: \sum_{i=1}^{r-1}\sum_{j=1}^{c}(s_{ij} - f_jt_i)^2/(f_jt_i) > \chi^2_{\nu,\alpha}\right\}.$$

THEOREM 2. *Under sampling model* 1,

$$\liminf_{n\to\infty} \bar{\alpha}_n(k) \geq \sup_{\omega>0} H(\alpha, \mathbf{f}, \omega, k) \geq \alpha \tag{6}$$

and

$$\liminf_{n\to\infty} \alpha_n(k) \geq \sup_{\omega>0} G(\alpha, \mathbf{f}, \omega, k) \geq \alpha \tag{7}$$

for any $0 < \alpha < 1$ *and* $k \geq 0$, *where*

$$G(\alpha, \mathbf{f}, \omega, k) = \left(\sum_{v > kf_*^{-1}} \frac{\omega^v}{v!} \right)^{-(r-1)} \sum_{\min t_i > kf_*^{-1}} \omega^{t.} \sum_{\mathbf{s}\in\mathscr{A}(t)} \prod_{j=1}^{c} \frac{f_j^{s_{.j}}}{\prod_{i=1}^{r-1}(s_{ij}!)}.$$

PROOF. The leftmost inequality of (6) is immediate, while that of (7) follows from (5) and

$$\Pr\left(\min_{1\leq i<r} T_i > kf_*^{-1} \right) = \left(e^{-\omega} \sum_{v > kf_*^{-1}} \frac{\omega^v}{v!} \right)^{(r-1)}.$$

The rightmost inequality in (6) follows from

$$\sup_{\omega>0} H(\alpha, \mathbf{f}, \omega, k) \geq \lim_{\omega\to\infty} H(\alpha, \mathbf{f}, \omega, k) = \alpha,$$

which is a consequence of the conditional distribution of

$$\sum_{i=1}^{r-1} \sum_{j=1}^{c} (S_{ij} - f_j t_i)^2 / (f_j t_i)$$

given $\{T_1 = t_1, \dots, T_{r-1} = t_{r-1}\}$ being asymptotically χ_ν^2 as $\min t_i \to \infty$. The same argument together with the fact

$$\lim_{\omega\to\infty} \Pr\left(\min_{1\leq i<r} T_i > kf_*^{-1} \right) \to 1$$

proves the rightmost inequality in (7). □

The quantity $H(\alpha, \mathbf{f}, \omega, k)$ is in general difficult to compute except when $r = c = 2$, for which it simplifies to

$$H(\alpha, \mathbf{f}, \omega, k) = \Pr\left[\left\{ |B - f_1 T|^2 > T f_1 (1 - f_1) \chi_{1,\alpha}^2 \right\} \cap \left\{ T > kf_*^{-1} \right\} \right],$$

where T is a Poisson variable with mean ω and the conditional distribution of B given $T = t$ is binomial with success probability f_1. Table 1 gives some values of $H(\alpha, \mathbf{f}, \omega, k)$ for this case.

An indication of the magnitude of the lower bound (7) for the conditional size can be obtained by noting that

$$\lim_{\omega\to 0} G(\alpha, \mathbf{f}, \omega, k) = \Pr\left\{ \mathbf{N} \in \mathscr{A}\left(\lfloor kf_*^{-1} + 1 \rfloor, \dots, \lfloor kf_*^{-1} + 1 \rfloor \right) \right\}, \tag{8}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x, and $\mathbf{N} = (N_{ij})$ is an $(r - 1) \times c$ random matrix with independent and identically distributed rows

TABLE 1
Lower and upper bounds for the limiting unconditional size $\lim \bar{\alpha}_n(k)$; *one margin fixed*; $r = c = 2$

α	(f_1, f_2)	k	ω	Lower bound $H(\alpha, \mathbf{f}, \omega, k)$	Upper bound $\nu\chi_{\nu,\alpha}^{-2}$
0.05	(0.01, 0.99)	0	13.0	0.104	0.260
0.10	(0.01, 0.99)	0	17.0	0.138	0.370
0.10	(0.50, 0.50)	0	4.7	0.100	0.370
0.10	(0.50, 0.50)	1	4.7	0.100	0.370
0.32	(0.10, 0.90)	0	12.0	0.391	1.001
0.32	(0.50, 0.50)	0	1.8	0.529	1.001

such that $(N_{i1}, \dots, N_{ic})$ is a multinomial vector with $\lfloor kf_*^{-1} + 1 \rfloor$ trials and probability vector $(f_1, \dots, f_c)$. The value of (8) can be evaluated simply when $k = 0$. For example, suppose $\mathbf{f} = \{f, \dots, f, 1 - (c-1)f\}$, with $f < c^{-1}$. Then because $kf_*^{-1} = 0$, we have $\mathbf{N} \in \mathscr{A}(1, \dots, 1)$ if and only if

$$\sum_{i=1}^{r-1} \sum_{j=1}^{c} (N_{ij} - f_j)^2 / f_j > \chi_{\nu,\alpha}^2.$$

The left side takes on its maximum value of

$$(r-1)\{f^{-1}(1-f)^2 + 1 - f\} = (r-1)f^{-1}(1-f),$$

when the event

$$\mathscr{B} = \bigcap_{i=1}^{r-1} \bigcup_{j=1}^{c-1} \{N_{ij} = 1, N_{ij'} = 0, j' \neq j\}$$

occurs. The maximum value exceeds $\chi_{\nu,\alpha}^2$ if and only if

$$f < (r-1)\left(\chi_{\nu,\alpha}^2 + r - 1\right)^{-1}.$$

Because $\Pr(\mathscr{B}) = \{(c-1)f\}^{r-1}$, it follows that for f sufficiently close to (but less than) $(r-1)(\chi_{\nu,\alpha}^2 + r - 1)^{-1}$,

$$\lim_{\omega \to 0} G(\alpha, \mathbf{f}, \omega, 0) = \{(c-1)f\}^{r-1}.$$

A similar calculation shows that if $\nu \leq \chi_{\nu,\alpha}^2$ and $\mathbf{f} = (c^{-1}, \dots, c^{-1})$, then

$$\lim_{\omega \to 0} G(\alpha, \mathbf{f}, \omega, 0) = 0.$$

Some values of $G(\alpha, \mathbf{f}, 0+, k)$ for $\alpha = 0.05$ are given in Table 2. We see that collapsing the rows of a table does not always lead to a decrease of the lower bound $G(\alpha, \mathbf{f}, 0+, k)$. For example, collapsing a 3×2 table with $\mathbf{f} = (0.20, 0.80)$ into a 2×2 table with the same $\mathbf{f}$ increases $G(\alpha, \mathbf{f}, 0+, 0)$ from 0.04 to 0.20. Another example is collapsing a 3×5 table with $\mathbf{f} = (0.08, 0.08, 0.08, 0.08, 0.68)$ into a 2×5 table with the same $\mathbf{f}$. On the other hand, collapsing columns to make the f_j's less extreme does not appear to be harmful. An example is

TABLE 2
Lower and upper bounds for the limiting conditional size $\lim \alpha_n(k)$;
one margin fixed; $\alpha = 0.05$

ν	r	c	$(f_1, \dots, f_c)$	k	Lower bound $G(\alpha, \mathbf{f}, 0+, k)$	Upper bound $\nu\chi_{\nu,\alpha}^{-2}$
1	2	2	(0.50, 0.50)	0	0	0.260
1	2	2	(0.50, 0.50)	5	0.065	0.260
1	2	2	(0.30, 0.70)	1	0.084	0.260
1	2	2	(0.30, 0.70)	5	0.060	0.260
1	2	2	(0.20, 0.80)	0	0.200	0.260
2	3	2	(0.20, 0.80)	0	0.040	0.334
4	3	3	(0.16, 0.16, 0.68)	0	0.102	0.421
4	2	5	(0.08, 0.08, 0.08, 0.08, 0.68)	0	0.320	0.421
8	3	5	(0.08, 0.08, 0.08, 0.08, 0.68)	0	0.102	0.516
9	4	4	(0.15, 0.15, 0.15, 0.55)	0	0.091	0.533

collapsing the above 3×5 table into a 3×3 table with $\mathbf{f} = (0.16, 0.16, 0.68)$, where we see that $G(\alpha, \mathbf{f}, 0+, 0)$ remains unchanged. The most improvement occurs when collapse of rows and columns is accompanied by an increase in k. For example, collapsing a 4×4 table with $k = 0$ and $\mathbf{f} = (0.15, 0.15, 0.15, 0.55)$ into a 2×2 table with $k = 1$ and $\mathbf{f} = (0.30, 0.70)$ reduces $G(\alpha, \mathbf{f}, 0+, k)$ from 0.091 to 0.084. Increasing k to 5 for this 2×2 table leads to further reduction of $G(\alpha, \mathbf{f}, 0+, k)$ to 0.060.

The next corollary shows that the bound (8) converges to α for large values of k.

COROLLARY 2.

$$\lim_{\omega \to 0} G(\alpha, \mathbf{f}, \omega, k) \to \alpha \quad \text{as } k \to \infty. \tag{9}$$

PROOF. From (8), the left side of (9) is just the probability of a type I error of a χ^2 goodness-of-fit test applied to the random matrix $\mathbf{N}$ consisting of $(r-1)$ i.i.d. rows of a multinomial vector with $\lfloor k f_*^{-1} + 1 \rfloor$ samples each. Letting k grow without limit implies letting the total sample size of $\mathbf{N}$ tend to ∞, and the result follows from the standard χ^2 asymptotics. □

2.2. *With continuity correction.* Consider now the continuity corrected statistic (2). A proof parallel to that of Theorem 2 yields the following theorem.

THEOREM 3. *Under sampling model* 1,

$$\liminf_{n \to \infty} \alpha_n^{(c)}(k) \geq \sup_{\omega > 0} G_c(\alpha, \mathbf{f}, \omega, k) \geq \alpha$$

and

$$\liminf_{n \to \infty} \bar{\alpha}_n^{(c)}(k) \geq \sup_{\omega > 0} H_c(\alpha, \mathbf{f}, \omega, k) \geq \alpha$$

for any $0 < \alpha < 1$ *and* $k \geq 0$, *where*

$$G_c(\alpha, \mathbf{f}, \omega, k) = \left(\sum_{v > kf_*^{-1}} \frac{\omega^v}{v!} \right)^{-(r-1)} \sum_{\min t_i > kf_*^{-1}} \omega^{t.} \sum_{\mathbf{s} \in \mathscr{A}_c(\mathbf{t})} \prod_{j=1}^{c} \frac{f_j^{s_{\cdot j}}}{\prod_{i=1}^{r-1}(s_{ij}!)},$$

$$H_c(\alpha, \mathbf{f}, \omega, k) = e^{-(r-1)\omega} \sum_{\min t_i > kf_*^{-1}} \omega^{t.} \sum_{\mathbf{s} \in \mathscr{A}_c(\mathbf{t})} \prod_{j=1}^{c} \frac{f_j^{s_{\cdot j}}}{\prod_{i=1}^{r-1}(s_{ij}!)}$$

and

$$\mathscr{A}_c(\mathbf{t}) = \left\{ \mathbf{s}: \sum_{i=1}^{r-1} \sum_{j=1}^{c} \left(|s_{ij} - f_j t_i| - \tfrac{1}{2} \right)^2 / (f_j t_i) > \chi^2_{\nu, \alpha} \right\}.$$

It turns out that when $k = 0$ and the dimensions of the table are sufficiently large, the limiting conditional size of the test is 1.

THEOREM 4. *Suppose that*

$$(r-1)(c-2)^2 > 4\chi^2_{\nu,\alpha}. \tag{10}$$

Then under sampling model 1, $\lim_{n \to \infty} \alpha_n^{(c)}(0) = 1$ *for all* $0 < \alpha < 1$ *and all* $\mathbf{f}$.

PROOF. Suppose that H_0 holds with π_i given in (4). Then using the notation there

$$\begin{aligned}
&\Pr\left(X_c^2 > \chi^2_{\nu,\alpha} | \mathscr{E}_0 \right) \\
&\quad \geq \Pr\left\{ \sum_{i=1}^{r-1} \sum_{j=1}^{c} \left(|Y_{ij} - E_{ij}| - \tfrac{1}{2} \right)^2 / E_{ij} > \chi^2_{\nu,\alpha} \,\middle|\, R_i \geq 1 \,\forall\, i \right\} \\
&\quad \to_{n \to \infty} \sum_{t_i \geq 1} \Pr\left[\left\{ \sum_{i=1}^{r-1} \sum_{j=1}^{c} \left(|S_{ij} - f_j T_i| - \tfrac{1}{2} \right)^2 / f_j T_i > \chi^2_{\nu,\alpha} \right\} \bigcap_{i=1}^{r-1} \{ T_i = t_i \} \right] \qquad (11)\\
&\qquad\qquad \div \Pr(\min T_i \geq 1) \\
&\quad \geq \Pr\left\{ \sum_{i=1}^{r-1} \sum_{j=1}^{c} \left(S_{ij} - f_j| - \tfrac{1}{2} \right)^2 / f_j > \chi^2_{\nu,\alpha} \,\middle|\, T_1 = \cdots = T_{r-1} = 1 \right\} \\
&\qquad \times \Pr(T_1 = \cdots = T_{r-1} = 1) / \Pr(\min T_i \geq 1) \\
&\quad = \Pr(T_1 = \cdots = T_{r-1} = 1) / \Pr(\min T_i \geq 1) \to_{\omega \to 0} 1.
\end{aligned}$$

Equality (11) is a consequence of the joint action of assumption (10), the fact that $(|y - f| - \frac{1}{2})^2$ is constant when $y = 0, 1$ for all $0 < f < 1$ and the inequality

$$\sum_{j=1}^{c} \left(f_j - \tfrac{1}{2} \right)^2 / f_j \geq (c-2)^2 / 4, \tag{12}$$

which holds because the left side of (12) is minimized when all the f_j are equal to c^{-1}. □

The next theorem gives an upper bound on the limiting size when $k > 0$.

THEOREM 5. *Under sampling model* 1,

$$\limsup_{n\to\infty} \alpha_n^{(c)}(k) \le \chi_{\nu,\alpha}^{-2}\{\nu + rc/(4k)\}$$

and

$$\limsup_{n\to\infty} \bar{\alpha}_n^{(c)}(k) \le \chi_{\nu,\alpha}^{-2}\{\nu + rc/(4k)\}$$

for all $0 < \alpha < 1$, $k > 0$ *and all* $\mathbf{f}$.

PROOF. Since $E_{ij} > k$ for each i and j, we have

$$X_c^2 \le X^2 + 4^{-1} \sum_{i=1}^{r} \sum_{j=1}^{c} E_{ij}^{-1} \le X^2 + rc/(4k).$$

The proof of Theorem 1 shows that for any set of integers $\{m_i\}$ such that

$$m_i \ge k(\min f_j)^{-1} \quad \text{and} \quad \sum_{i=1}^{r} m_i = n,$$

we have

$$\begin{aligned}
&\Pr\{X_c^2 > \chi_{\nu,\alpha}^2 | R_i = m_i, i = 1,\dots,r\}\\
&\quad\le \chi_{\nu,\alpha}^{-2} E\left\{ \sum_{i=1}^{r} \sum_{j=1}^{c} (f_j m_i)^{-1}(|Y_{ij} - f_j m_i| - \tfrac{1}{2})^2 \,\middle|\, R_i = m_i, i = 1,\dots,r \right\}\\
&\quad\le \chi_{\nu,\alpha}^{-2} E\left\{ \sum_{i=1}^{r} \sum_{j=1}^{c} (f_j m_i)^{-1}(Y_{ij} - f_j m_i)^2 + rc/(4k) \,\middle|\, R_i = m_i, i = 1,\dots,r \right\}\\
&\quad= \chi_{\nu,\alpha}^{-2}\{n(n-1)^{-1}\nu + rc/(4k)\}.
\end{aligned}$$

Thus

$$\Pr(\{X_c^2 > \chi_{\nu,\alpha}^2\} \cap \mathscr{E}_k) \le \Pr(X_c^2 > \chi_{\nu,\alpha}^2 | \mathscr{E}_k) \le \chi_{\nu,\alpha}^{-2}\{n(n-1)^{-1}\nu + rc/(4k)\}$$

and the proof is completed by taking limits. □

3. Both sets of marginal totals random. Suppose now that sampling is performed such that the marginal totals $\{R_i\}$ and $\{C_j\}$ are all random. The hypothesis to be tested is

$$H_0\colon p_{ij} = p_{i\cdot}p_{\cdot j}, \qquad i = 1,\dots,r;\ j = 1,\dots,c.$$

3.1. *Without continuity correction.* Because conditioning on the column totals reduces the problem to that of sampling with one set of marginal totals fixed, the next result follows from Theorem 1 by taking expectations.

THEOREM 6. *Under sampling model* 2,

$$\bar{\alpha}_n(k) \le \alpha_n(k) \le n(n-1)^{-1}\nu\chi_{\nu,\alpha}^{-2}$$

for all $0 < \alpha < 1$ *and* $k \ge 0$.

To obtain a lower bound, assume H_0 holds with

(13) $\quad p_{i\cdot} = p_{\cdot j} = \sqrt{\lambda/n} \quad$ for $i = 1, \dots, r-1$; $j = 1, \dots, c-1$; $\lambda > 0$.

Then as $n \to \infty$, we have

(14) $\quad Y_{ij} \to_D T_{ij}(\lambda), \qquad E_{ij} \to_P \lambda, \qquad \forall\, i = 1, \dots, r-1;\ j = 1, \dots, c-1,$

where $\{T_{ij}(\lambda): i = 1, \dots, r-1;\ j = 1, \dots, c-1\}$ is a set of mutually independent and identically distributed Poisson random variables with mean λ. Therefore

$$\Pr(X^2 > \chi_{\nu,\alpha}^2) \ge \Pr\left\{\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}(Y_{ij} - E_{ij})^2/E_{ij} > \chi_{\nu,\alpha}^2\right\}$$

$$\to \Pr\left\{\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\lambda^{-1}[T_{ij}(\lambda) - \lambda]^2 > \chi_{\nu,\alpha}^2\right\}.$$

Because $\lambda^{-1/2}\{T_{ij}(\lambda) - \lambda\}$ is asymptotically standard normal as $\lambda \to \infty$, we have the following theorem.

THEOREM 7. *Under sampling model* 2,

$$\liminf_{n\to\infty} \alpha_n(k) \ge J(k), \qquad \liminf_{n\to\infty} \bar{\alpha}_n(k) \ge J(k),$$

where

$$J(k) = \sup_{\lambda > k} \Pr\left[\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\lambda^{-1}\{T_{ij}(\lambda) - \lambda\}^2 > \chi_{\nu,\alpha}^2\right] \ge \alpha$$

for all $0 < \alpha < 1$ *and* $k \ge 0$.

Observe that $J(k)$ is a decreasing function of k. A lower bound for $J(k)$ may be obtained as follows. Suppose that $u - \frac{1}{2} \le \lambda < u + \frac{1}{2}$, and that m of the T_{ij} takes value u, for some integers $0 \le m \le \nu$ and $u \ge k$. Then

$$\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\{T_{ij}(\lambda) - \lambda\}^2 \ge m(\lambda - u)^2 + (\nu - m)(1 - |\lambda - u|)^2$$

and hence

(15) $\quad J(k) \ge \sup\{K(\nu, \alpha, \lambda, u): \lambda > k, u \ge k, |\lambda - u| < \frac{1}{2}\},$

TABLE 3

Lower and upper bounds for the limiting conditional and unconditional size without continuity correction; $k = 0$; both margins random

ν	α	λ	Lower bound $K(\nu, \alpha, \lambda, 0)$	Upper bound $\nu\chi_{\nu,\alpha}^{-2}$
1	0.01	0.117	0.110	0.151
	0.05	0.176	0.161	0.260
5	0.01	0.059	0.255	0.331
	0.05	0.078	0.323	0.452
10	0.01	0.040	0.330	0.431
	0.05	0.050	0.393	0.546
50	0.01	0.012	0.451	0.657
	0.05	0.014	0.503	0.741
100	0.01	0.007	0.503	0.736
	0.05	0.007	0.503	0.804
∞	(0, 1)	$\approx \nu^{-1}$	0.632	1

where

$$K(\nu, \alpha, \lambda, u) = \sum_{m \in \mathscr{M}} \frac{\nu!}{m!(\nu - m)!} \theta^m (1 - \theta)^{\nu - m},$$

$$(16) \quad \mathscr{M} = \left\{m: 0 \le m \le \nu \text{ and } m(\lambda - u)^2 + (\nu - m)(1 - |\lambda - u|)^2 > \lambda\chi_{\nu,\alpha}^2\right\}$$

and

$$\theta = e^{-\lambda}\lambda^u / u!.$$

Thus $K(\nu, \alpha, \lambda, u)$ is a lower bound for $J(k)$ for appropriate values of λ and u. Table 3 gives some values of $K(\nu, \alpha, \lambda, 0)$ which are lower bounds for $J(0)$. They are the maximum values of $K(\nu, \alpha, \lambda, u)$ over a grid $\{(u, \lambda): u = 0, 1, \dots, 5;\ \lambda = u \pm i/1000,\ i = 0, 1, \dots, 500\}$. It happens that the maximum is consistently attained at $u = 0$. This observation leads to the following corollary.

COROLLARY 3. *Under sampling model* 2,

$$\liminf_{n \to \infty} \alpha_n(0) \ge 0.632, \qquad \liminf_{n \to \infty} \bar{\alpha}_n(0) \ge 0.632$$

for all $0 < \alpha < 1$ *and* ν *sufficiently large.*

PROOF. Put $u = 0$ and $\lambda = \eta\nu^{-1}$ in (16), where $0 < \eta < 1$. Because $\chi_{\nu,\alpha}^2 \approx \nu + z_\alpha\sqrt{2\nu}$, we have $\mathscr{M} = \{0, 1, \dots, \nu - 1\}$ for all large ν. Let V denote a binomial variable with ν trials and success probability $e^{-\eta/\nu}$, and let $T(\eta)$ denote a

Poisson variable with mean η. Then for each $\eta < 1$ and ν sufficiently large,

$$K(\nu, \alpha, \eta\nu^{-1}, 0) = \Pr(0 \le V \le \nu - 1) \to \Pr\{T(\eta) \ge 1\} \quad \text{as } \nu \to \infty.$$

The corollary now follows from (15), Theorem 7 and the fact that $\Pr\{T(1) \ge 1\} = 0.632$. □

Notice that, unlike the lower bounds in Table 2, which are not monotonic in ν because they are functions of the column probabilities, those in Table 3 are increasing in ν.

3.2. *With continuity correction.* The following analog of Theorem 7 with the continuity correction is similarly proved.

THEOREM 8. *Under sampling model* 2,

$$\liminf_{n\to\infty} \alpha_n^{(c)}(k) \ge J_c(k), \qquad \liminf_{n\to\infty} \bar{\alpha}_n^{(c)}(k) \ge J_c(k),$$

where

$$J_c(k) = \sup_{\lambda > k} \Pr\left[\sum_{i=1}^{r-1}\sum_{j=1}^{c-1} \lambda^{-1}\{|T_{ij}(\lambda) - \lambda| - \tfrac{1}{2}\}^2 > \chi^2_{\nu,\alpha}\right] \ge \alpha$$

for all $0 < \alpha < 1$ *and* $k \ge 0$.

As in model 1, the lower bound for the limiting conditional size is too conservative when $k = 0$.

THEOREM 9. *Under sampling model* 2, $\lim_{n\to\infty} \alpha_n^{(c)}(0) = 1$ *for all* $0 < \alpha < 1$.

PROOF. Let H_0 hold with the probabilities given in (13) and suppose that $k = 0$. Then (14) implies that

$$\begin{aligned}\Pr(X_c^2 > \chi^2_{\nu,\alpha}) &\ge \Pr(\{|Y_{11} - E_{11}| - \tfrac{1}{2}\}^2/E_{11} > \chi^2_{\nu,\alpha}) \\ &\to_{n\to\infty} \Pr(\{|T_{11}(\lambda) - \lambda| - \tfrac{1}{2}\}^2/\lambda > \chi^2_{\nu,\alpha}) \\ &\to_{\lambda\to 0} 1.\end{aligned}$$

The theorem follows because $\Pr(\mathscr{E}_0) \to 1$ as $n \to \infty$. □

Our last result gives an upper bound for the case $k > 0$. It is proved by taking expectations in Theorem 5.

THEOREM 10. *Under sampling model* 2,

$$\limsup_{n\to\infty} \alpha_n^{(c)}(k) \le \chi^{-2}_{\nu,\alpha}\{\nu + rc/(4k)\}$$

and

$$\limsup_{n\to\infty} \bar{\alpha}_n^{(c)}(k) \le \chi^{-2}_{\nu,\alpha}\{\nu + rc/(4k)\}$$

for all $0 < \alpha < 1$ *and* $k > 0$.

4. Concluding remarks. Although the bounds we derived may sometimes be conservative, they suggest the following conclusions.

1. Whether the continuity correction is used or not, and whether sampling is performed with one set of margins fixed or with both sets random, the asymptotic size of the χ^2-test, as the sample size increases, can be many times larger than its nominal level.
2. The numbers in Tables 1 and 2 suggest that when sampling is performed with one set of margins fixed, the best choice of $\mathbf{f}$ is one with equal components.
3. If the continuity correction is used, the test should be carried out conditional on the event $\mathscr{E}_k$ for some $k > 0$. Otherwise the asymptotic conditional size of the test is one for all α.
4. Larger values of k yield tests with size closer to the nominal level than do smaller values of k.
5. The upper bounds suggest that another way to improve the accuracy of the size of a test is to collapse rows and columns to reduce degrees of freedom.
6. Interestingly, both an increase in k and a decrease in degrees of freedom can be achieved by collapse of the rows and columns of a table. This explains why the usual practice of collapsing a table to ensure that k is sufficiently large tends to make the test more effective.

Acknowledgments. I am grateful to an Associate Editor and three referees for their useful comments.

REFERENCES

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco.

CONOVER, W. J. (1974). Some reasons for not using the Yates continuity correction on 2 × 2 contingency tables. *J. Amer. Statist. Assoc.* **69** 374–382.

CRESSIE, N. and READ, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*, 5th ed. Oliver and Boyd, Edinburgh.

GARSIDE, G. R. and MACK, C. (1976). Actual type 1 error probabilities for various tests in the homogeneity case of the 2 × 2 contingency table. *Amer. Statist.* **30** 18–21.

GRIZZLE, J. E. (1967). Continuity correction in the χ^2-test for the 2 × 2 tables. *Amer. Statist.* **21** 28–32.

HABER, M. (1980). A comparison of some continuity corrections for the chi-squared test on 2 × 2 tables. *J. Amer. Statist. Assoc.* **75** 510–515.

HABERMAN, S. J. (1988). A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Amer. Statist. Assoc.* **83** 555–560.

JOHNSON, N. L. and KOTZ, S. (1969). *Distributions in Statistics. Discrete Distributions.* Wiley, New York.

JOHNSON, N. L. and KOTZ, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions* **1**. Wiley, New York.

KALLENBERG, W. C. M. (1985). On moderate and large deviations in multinomial distributions. *Ann. Statist.* **13** 1554–1580.

LARNTZ, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *J. Amer. Statist. Assoc.* **73** 253–263.

MANTEL, N. and GREENHOUSE, S. W. (1968). What is the continuity correction? *Amer. Statist.* **22** 27–30.

OOSTERHOFF, J. (1985). The choice of cells in chi-square tests. *Statist. Neerlandica* **39** 115–127.

PEARSON, E. S. (1947). The choice of a statistical test illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika* **34** 139–167.

PLACKETT, R. L. (1964). The continuity correction in 2 × 2 tables. *Biometrika* **51** 327–337.

SNEDECOR, G. W. and COCHRAN, W. G. (1980). *Statistical Methods*, 7th ed. Iowa State Univ. Press, Ames, Ia.

YARNOLD, J. K. (1970). The minimum expectation in χ^2 goodness-of-fit tests and the accuracy of approximations for the null distribution. *J. Amer. Statist. Assoc.* **65** 864–886.

YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. *J. Roy. Statist. Soc. Suppl.* **1** 217–235.

YATES, F. (1984). Tests of significance for 2 × 2 contingency tables (with discussion). *J. Roy. Statist. Soc. Ser. A* **147** 426–463.

ZELTERMAN, D. (1987). Goodness-of-fit tests for large sparse multinomial distributions. *J. Amer. Statist. Assoc.* **82** 624–629.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706