On the minimum expected quantity for the validity of the chi-squared test in $2 \times 2$ tables

A. MARTÍN ANDRÉS$^1$ & I. HERRANZ TEJEDOR$^2$, $^1$ Bioestadística, Facultad de Medicina, Universidad de Granada, Spain and $^2$ Bioestadística, Facultad de Medicina, Universidad Complutense, Madrid, Spain

ABSTRACT A $2 \times 2$ contingency table can often be analysed in an exact fashion by using Fisher’s exact test and in an approximate fashion by using the chi-squared test with Yates’ continuity correction, and it is traditionally held that the approximation is valid when the minimum expected quantity $E$ is $E \geqslant 5$. Unfortunately, little research has been carried out into this belief, other than that it is necessary to establish a bound $E > E^*$, that the condition $E \geqslant 5$ may not be the most appropriate (Martín Andrés et al., 1992) and that $E^*$ is not a constant, but usually increasing with the growth of the sample size (Martín Andrés & Herranz Tejedor, 1997). In this paper, the authors conduct a theoretical-experimental study from which they ascertain that $E^*$ value (which is very variable and frequently quite a lot greater than 5) is strongly related to the magnitude of the skewness of the underlying hypergeometric distribution, and that bounding the skewness is equivalent to bounding $E$ (which is the best control procedure). The study enables estimating the expression for the above-mentioned $E^*$ (which in turn depends on the number of tails in the test, the $\alpha$ error used, the total sample size, and the minimum marginal imbalance) to be estimated. Also the authors show that $E^*$ increases generally with the sample size and with the marginal imbalance, although it does reach a maximum. Some general and very conservative validity conditions are $E \geqslant 35.53$ (one-tailed test) and $E \geqslant 7.45$ (two-tailed test) for $\alpha$ nominal errors in $1\% \leqslant \alpha \leqslant 10\%$. The traditional condition $E \geqslant 5$ is only valid when the samples are small and one of the marginals is very balanced; alternatively, the condition $E \geqslant 5.5$ is valid for small samples or a very balanced marginal. Finally, it is proved that the chi-squared test is always valid in tables where both marginals are balanced, and that the maximum skewness permitted is related to the maximum value of the bound $E^*$, to its value for tables with at least one balanced marginal and to the minimum value that those marginals must have (in non-balanced tables) for the chi-squared test to be valid.

Correspondence: A. Martín Andrés, Bioestadística, Facultad de Medicina, Universidad de Granada, 18071 Granada, Spain. E-mail: amartina@ugr.es

ISSN 0266-4763 print; 1360-0532 online/00/070807-14 © 2000 Taylor & Francis Ltd
1 Introduction

By a $2 \times 2$ table we mean a layout of data as shown in Table 1. The data can arise from three types of sampling: zero fixed marginals or the problem of the association of two dichotomic qualities; one fixed marginal or the problem of comparison of two independent proportions; and two fixed marginals or sampling of hypergeometric distribution. However, the analysis, from the conditional standpoint (two fixed marginals), is carried out by means of a single method: Fisher's exact test (Fisher, 1935) for the exact solution; or the chi-squared test with Yates' correction (Yates, 1934) for the approximate solution. The same description, with appropriate changes, is valid for more general contingency tables ($r \times s$ tables). Generally speaking, the aim here is to contrast the null hypothesis $H_0$, which supposes the characteristics A and B to be independent. The theoretical convenience of dealing with the three samplings as a single model (the conditional one) may be found in Yates (1984).

For many years, the complexity of the computation for Fisher's exact test has resulted in the traditional application of the chi-squared test as an approximate solution to the problem, but nowadays there exists abundant software (for example, StatXact) to solve it in an exact and easy way (especially in the case of $2 \times 2$ tables). Nevertheless, teaching, tradition and convenience have meant that textbooks and research publications continue to make use of the chi-squared test as a solution to the problem, while, paradoxically, the validity conditions of the test (v.c. in the following) are still unknown. This paper studies what takes place in $2 \times 2$ tables (which are the most frequent), and its findings may serve as a working model for the study of other similar or more complex situations. In particular it is experimentally confirmed that the validity of the approximate test is strongly conditioned by the skewness of the base hypergeometric distribution. This suggests that the v.c. for any other test in which a distribution approximates via the normal one would be greatly conditioned by the skewness of the distribution of the statistic of the test, and thus, a combination of the formula for skewness with an experimental study similar to the following, would produce the desired v.c.'s.

It is common to state that the chi-squared test is valid if the minimum expected quantity is sufficiently large ($E > E^*$), but it is still not known what is understood by 'sufficiently large'. It is traditional to accept Fisher's advice (1925) that it should be $E \geq 5$, but Brownlee (1967, p. 153) advises $E > 3.5$; Cochran (1954) states that E should be smaller than 2 or 5; Haber (1980) that $E \geq \text{Max}(5; n/10)$ and, finally, Martín Andrés et al. (1992) found that the rule $E \geq 5$ oversimplifies (especially in one-tailed tests or in low $p$-values), since the performance of the test does not depend on $E$ alone. In fact, Martín Andrés & Herranz Tejedor (1997) proved
experimentally that it is sometimes necessary for \( E > 20 \) and that \( E^* \) is not a constant, but usually increases with \( n \) and depends on the error \( \alpha \) used and on whether the test is one- or two-tailed. The aim of this paper is to prove experimentally that the value of \( E^* \) (for fixed values of nominal error \( \alpha \) and the number of tails of the test) depends not only on \( n \), but also on the marginal imbalance; that \( E^* \) is related to the skewness of the hypergeometric distribution which produces it (and whose control is linked to the control of \( E \)). Other aims include determining the form of the function \( E^* \) and estimating the implied constants within it. All this contributes to the aim of obtaining a universal formula for the validity of the chi-squared test in \( 2 \times 2 \) tables.

In the following, the statement that 'the chi-squared test is valid' is to be understood in the sense given by Cochran (1954): the \( P \)-value obtained by the chi-squared test \( (P_X) \) is approximately equal to the exact \( P \)-value \( (P_F) \) obtained from Fisher's exact test, that is, \( P_X \approx P_F \) (although Cochran proposed it for real and objective \( \alpha \) errors). The magnitude allowed for the difference \( |P_X - P_F| \) may, according to Cochran, be 20% of \( P_F \) for \( P_F = 5\% \) or 50% of \( P_F \) for \( P_F = 1\% \), which implies, for example, that the approximate \( P \)-values \( P_X \) of \( 4\% \leq P_X \leq 6\% \) are acceptable for a \( P_F = 5\% \). Since the values of \( P_F \) will rarely be exactly 5% or 1%, Martín Andrés et al. (1992) and Martín Andrés & Herranz Tejedor (1997) extended the previous criterion to the following one: the chi-squared test is valid in a given table if it is verified that:

\[
|P_X - P_F| \leq \delta P_F \quad \text{with} \quad \delta = \begin{cases} 
0.5 & \text{if } P_F \leq 1\% \\
0.575 - 7.5 P_F & \text{if } 1\% < P_F < 5\% \\
0.2 & \text{if } P_F \geq 5\%
\end{cases} \quad (1)
\]

(where the case of \( 1\% < P_F < 5\% \) has been obtained by interpolation) and this criterion is the one that will be assumed in what follows.

### 2 Experimental values of \( E^* \)

#### 2.1 On the tests to be used

Under \( H_0 \), and assuming that the two marginals \( a_i \) and \( n_i \) are fixed previously (conditional method), the probability of results like those in Table 1 are given by the hypergeometric distribution

\[
P(x_i) = \binom{n_i}{x_i} \binom{n - n_i}{x_i} / \binom{n}{a_i} \quad (2)
\]

where \( r = \max (0; a_i - n_i) \leq x_i \leq \min (a_i; n_i) = s \). If the alternative is a left-handed tail (negative association), Fisher's \( p \)-value will be \( P_F = P(r) + \ldots + P(x_i) \), where \( P(\cdot) \) is obtained by \( (2) \). If the alternative is two-tailed and supposing (without loss of generality) that it is \( x_i < E_{11} = a_i n_i / n \), Fisher's \( p \)-value will be \( P_F = \{P(r) + \ldots + P(x_i)\} + \{P(x'_i) + \ldots + P(s)\} \), with \( x'_i \) the first value of \( x_i > E_{11} \) as extreme or more so than \( x_1 \), which depends on the arrangement criterion employed (Martín Andrés & Herranz Tejedor, 1995). In order to make the results comparable, it is advisable to take the arrangement criterion of the chi-squared statistics, hence \( x'_i = [2E_{11} - x]^+ \), where \([x]^+\) refers to the first integer larger or equal to \( x \) (Martín Andrés & Herranz Tejedor, 1995). When \( x_1 = E_{11} \), then \( P_F = 1 \).
The chi-squared test can be used as an approximate method with the appropriate continuity correction (c.c. in the following). Although the best known c.c. is that of Yates (1934), which gives rise to the statistic:

\[
\chi^2 = \frac{\left( |x_1y_2 - x_2y_1| - \frac{n}{2} \right)^2}{a_1a_2n_1n_2n}
\]  

(3)

Martín Andrés et al. (1992) describe other possibilities and Martín Andrés & Herranz Tejedor (1997) select the optimal. For the one-tailed tests, the optimal c.c. is that of Martín Andrés et al. (1992), derived from Conover’s (1974) criterion, which yields the statistic: \( \chi^2_C = \chi^2 + n^2 / (4a_1a_2n_1n_2) \). So, if \( \chi^2 \) is a random variable \( \chi^2 \) with one degree of freedom, \( \chi^2_C \) the experimental value of the statistic that is being used (\( \chi^2 \) or \( \chi^2_C \)) and \( P(\chi^2_C) = P(\chi^2 > \chi^2_C) \), then the \( p \)-value for the one-tailed test will be \( P_Y = P(\chi^2_Y) / 2 \) or \( P_C = P(\chi^2_C) / 2 \). For the two-tailed tests, that of Yates is also the optimal c.c., taking the caution added by Mantel (1974): the \( p \)-value for two-tailed test is \( P_M = P(\chi^2_Y) + P(\chi^2_C) / 2 \), where \( \chi^2_C \) refers to the value of (3) in a table where \( x_1 = x'_1 \). The values \( P_Y, P_C \) or \( P_M \) will be approximations of the exact \( P_F \) values already mentioned.

Finally, it should be pointed out that the c.c.’s described are optimals for \( \alpha \) nominal errors between 1% and 10% (and not in any other case), which this paper concentrates on (as being the most usual and because the v.c.’s depend on \( \alpha \), so that it is desirable to have \( \alpha \) relatively fixed).

2.2 On the validity of the chi-squared test and the tables to be considered

Under the conditional model, the sample space for the marginals in Table 1 are all the values of \( x_1 \) compatible with the triplet \( (a_1, n_1, n) \): \( r \leq x_1 \leq s \). Let us say that the chi-squared test ‘does not fail’ for a given value of \( x_1 \) if expression (1) is verified for it. On the contrary, we shall say that it ‘fails’ if \( |P_X - P_F| > \delta P_F \). Finally, we shall say that the chi-squared test ‘is valid’ for the whole sample space if it does not fail in any of its \( x_1 \) values. In this way, there will be sample spaces based on the triplet \( (a_1, n_1, n) \) in which the chi-squared test is valid and others in which it will not be so (when at least one of its \( x_1 \) values fails the criterion). In all the above, because of what was stated at the end of Section 2.1, we restrict ourselves to the \( x_1 \) values which yield a \( P_X \) such that \( 1% \leq P_X \leq 10% \).

According to the published articles on the subject, the chi-squared is valid in the triplets \( (a_1, n_1, n) \) which verify \( E > E^* \), where \( E = a_1n_1/n \) if Table 1 is rearranged, without loss of generality, so that \( a_1 = \min(a_1, a_2, n_1, n_2) \) and \( n_1 = \min(n_1, n_2) \), in which case \( 0 \leq x_1 \leq a_1 \). This implies that for a given pair \( (n_1, n) \), the sample spaces in which the chi-squared test is valid are those where \( a_1 > nE^*/n_1 = a_1^* \), meaning that \( a_1^* \) is a value which must be determined experimentally for each pair \( (n_1, n) \) or, by making \( K = n_2/n_1 \) (\( \geq 1 \)), for each pair \( (n_1, K) \). Thus: \( a_1^* \) is the first value of \( a_1 \) (for given \( K \) and \( n_1 \)) for which the chi-squared test is not valid.

In order to determine \( a_1^* (n_1; K) \) it is necessary to take a wide range of values of \( (n_1, K) \), given by \( n_1 = 40, 60, 100, (50, 250) \) and \( K = 1(0.1)2(0.5)4, 5, 6 \) (with a very exhaustive number of small values of \( K \), as these are the most frequent in practice). In continuation, for each pair \( (n_1, K) \), its associated \( a_1^* \) value will be determined experimentally in the following way: (1) fix the value of \( a_1 \), with \( 1 \leq a_1 \leq n_1 \), starting at \( a_1 = n_1 \) and descending one unit at each step; (2) obtain all the tables where
0 \leq x_1 \leq a_1$, calculate their approximate $p$-value ($P_0$) and retain only the tables that give a $P_0$ between 1% and 10%; (3) calculate the exact $p$-value ($P_1$) for these tables and check whether or not the chi-squared method 'fails' (according to the method described above); (4) if any failure is obtained, the process is halted; if not, we return to Step 1.

2.3 Results

Table 2 contains the values of $a^*_k$ for the one-tailed test $\chi^2_k$ and the two-tailed test (which do the job of the $X$ methods). They automatically determine value $E^* = a^*_k/(K+1)$. The v.c.'s $E > E^*$ and $a_i > a^*_k$ are therefore equivalents. For the case of one tail it is seen that:

1. The values of $E^*$ are frequently appreciably larger than 5, only being lower when they are near $K = 1$ (a balanced marginal). This indicates how inappropriate is the habitual common rule that the v.c. is $E > 5$.

2. For constant $n_1$, $a^*_k$ increases with $K$ (and also, generally, with $E^*$), which is one proof that $E^*$ varies with $K$ (and with $n_1$, as can be shown). The growth of $a^*_k$ cannot be maintained indefinitely, as, by construction, $a_i \leq n_i$, meaning that when $a^*_k$ approaches $n_1$ the growth slows down.

3. For $K = 1$, it is always the case that $a^*_k = 3$ (and $E^* = 1.5$), indicating that in no circumstance can the v.c. be as liberal as $E > 0$.

Table 2 also indicates the values of $a^*_k$ for the two-tailed test $\chi^2_k$ with Mantel's caution. Here it can be observed how $a^*_k$ also increases with $K$, but more slowly and with more irregularities than in the case of the one-tailed test, without managing to approach the maximum possible (which is $n_1$), although starting from higher values ($a^*_k = 7$ or 8 compared with the value 3 for the one-tailed test). This indicates that now the v.c.'s will be less strict in general (except for small values of $K$), as is seen from the experimental values of $E^*$. On the other hand, abrupt falls in the value of $a^*_k$ can be observed, first for $K$ between 1.3 and 1.6, then in the neighbourhood of $K = 3$ and finally, in the neighbourhood of $K = 5$. The first also occurred (although much more smoothly) in the one-tailed test, but not the second. All this points to something that is well known: the two-tailed chi-squared test performs in a different way from the one-tailed test (which is a reflection of the erratic performance of Fisher's two-tailed exact test: Cornfield, 1966; Yates, 1984; Dupont, 1986; Cormack, 1986; Martin Andrés et al., 1992). Here the extra values of $K = 2.6(0.1)2.9$ have now been included for the reason given below. Now, the old 'rule of 5' is generally conservative.

In this way, and in the case of both one- and two-tailed tests, it has been experimentally observed that neither $a^*_k$ nor $E^*$ are a constant, but rather that they vary with $K$ and with $n_1$, and these should be understood as a function of both: $a^*_k(n_1; K)$ and $E^*(n_1; K)$. What form will these functions take?

3 Estimated values of $E^*$

3.1 The condition of bounded asymmetry and its consequences

Let

$$f(x) = \frac{(n - 2x)^2}{x(n - x)}$$

(4)
TABLE 2. Real minimum values ($a_\gamma^*$ and $E^*$) of the smaller marginal ($a_\gamma$) and the minimum expected quantity ($E$) from where the chi-squared test is valid (as an approximation to Fisher's exact test) and conservative predictions $\hat{d}_\gamma^*$ of $a_\gamma^*$. It is taken as given that $K = n_1/n_2 \geq 1.$

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$K$</th>
<th>$a_\gamma^*$</th>
<th>$\hat{d}_\gamma^*$</th>
<th>$E^*$</th>
<th>$n_1$</th>
<th>$K$</th>
<th>$a_\gamma^*$</th>
<th>$\hat{d}_\gamma^*$</th>
<th>$E^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1.5</td>
<td>150</td>
<td>3</td>
<td>10</td>
<td>1.5</td>
<td>8</td>
</tr>
<tr>
<td>1.1</td>
<td>3</td>
<td>8</td>
<td>1.4</td>
<td>3.3</td>
<td>150</td>
<td>1.1</td>
<td>3</td>
<td>1.4</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>3</td>
<td>9</td>
<td>1.4</td>
<td>3.6</td>
<td>150</td>
<td>1.2</td>
<td>8</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>1.3</td>
<td>3</td>
<td>9</td>
<td>1.3</td>
<td>3.6</td>
<td>150</td>
<td>1.3</td>
<td>12</td>
<td>1.3</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>3</td>
<td>10</td>
<td>1.2</td>
<td>3.8</td>
<td>150</td>
<td>1.4</td>
<td>9</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>9</td>
<td>11</td>
<td>3.6</td>
<td>3.6</td>
<td>150</td>
<td>1.5</td>
<td>10</td>
<td>4.0</td>
<td>9</td>
</tr>
<tr>
<td>1.6</td>
<td>10</td>
<td>12</td>
<td>3.8</td>
<td>3.6</td>
<td>150</td>
<td>1.6</td>
<td>16</td>
<td>3.8</td>
<td>11</td>
</tr>
<tr>
<td>1.7</td>
<td>10</td>
<td>13</td>
<td>3.7</td>
<td>4.1</td>
<td>150</td>
<td>1.7</td>
<td>15</td>
<td>5.6</td>
<td>11</td>
</tr>
<tr>
<td>1.8</td>
<td>11</td>
<td>15</td>
<td>3.9</td>
<td>3.6</td>
<td>150</td>
<td>1.8</td>
<td>16</td>
<td>5.7</td>
<td>10</td>
</tr>
<tr>
<td>1.9</td>
<td>16</td>
<td>16</td>
<td>5.5</td>
<td>4.1</td>
<td>150</td>
<td>1.9</td>
<td>17</td>
<td>5.9</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>17</td>
<td>4.0</td>
<td>1.0</td>
<td>150</td>
<td>2</td>
<td>18</td>
<td>6.0</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>20</td>
<td>23</td>
<td>5.7</td>
<td>4.6</td>
<td>150</td>
<td>2.5</td>
<td>32</td>
<td>9.1</td>
<td>16</td>
</tr>
<tr>
<td>2.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.6</td>
<td>150</td>
<td>2.6</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>2.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.7</td>
<td>150</td>
<td>2.7</td>
<td>—</td>
<td>—</td>
<td>17</td>
</tr>
<tr>
<td>2.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.5</td>
<td>150</td>
<td>2.8</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>2.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.5</td>
<td>150</td>
<td>2.9</td>
<td>—</td>
<td>—</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>29</td>
<td>6.0</td>
<td>2.0</td>
<td>150</td>
<td>3</td>
<td>43</td>
<td>10.8</td>
<td>2</td>
</tr>
<tr>
<td>3.5</td>
<td>33</td>
<td>35</td>
<td>7.3</td>
<td>0.4</td>
<td>150</td>
<td>3.5</td>
<td>49</td>
<td>10.9</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>38</td>
<td>40</td>
<td>7.6</td>
<td>2.3</td>
<td>150</td>
<td>4</td>
<td>62</td>
<td>12.4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>40</td>
<td>6.3</td>
<td>19</td>
<td>150</td>
<td>5</td>
<td>84</td>
<td>14.0</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>40</td>
<td>5.7</td>
<td>32</td>
<td>150</td>
<td>6</td>
<td>108</td>
<td>15.4</td>
<td>39</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1.5</td>
<td>8</td>
<td>200</td>
<td>1</td>
<td>3</td>
<td>1.5</td>
<td>8</td>
</tr>
<tr>
<td>1.1</td>
<td>3</td>
<td>9</td>
<td>1.4</td>
<td>8</td>
<td>200</td>
<td>1.1</td>
<td>3</td>
<td>1.4</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>3</td>
<td>9</td>
<td>1.4</td>
<td>8</td>
<td>200</td>
<td>1.2</td>
<td>8</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>1.3</td>
<td>3</td>
<td>10</td>
<td>1.3</td>
<td>9</td>
<td>200</td>
<td>1.3</td>
<td>3</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>1.4</td>
<td>9</td>
<td>11</td>
<td>3.8</td>
<td>9</td>
<td>200</td>
<td>1.4</td>
<td>9</td>
<td>3.8</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>9</td>
<td>13</td>
<td>3.6</td>
<td>9</td>
<td>200</td>
<td>1.5</td>
<td>10</td>
<td>4.0</td>
<td>9</td>
</tr>
<tr>
<td>1.6</td>
<td>10</td>
<td>14</td>
<td>3.8</td>
<td>11</td>
<td>200</td>
<td>1.6</td>
<td>15</td>
<td>5.8</td>
<td>11</td>
</tr>
<tr>
<td>1.7</td>
<td>10</td>
<td>15</td>
<td>3.7</td>
<td>11</td>
<td>200</td>
<td>1.7</td>
<td>15</td>
<td>5.6</td>
<td>11</td>
</tr>
<tr>
<td>1.8</td>
<td>11</td>
<td>17</td>
<td>3.9</td>
<td>10</td>
<td>200</td>
<td>1.8</td>
<td>16</td>
<td>5.7</td>
<td>10</td>
</tr>
<tr>
<td>1.9</td>
<td>16</td>
<td>18</td>
<td>5.5</td>
<td>12</td>
<td>200</td>
<td>1.9</td>
<td>17</td>
<td>5.9</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>20</td>
<td>5.7</td>
<td>3</td>
<td>200</td>
<td>2</td>
<td>18</td>
<td>6.0</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>26</td>
<td>27</td>
<td>7.4</td>
<td>16</td>
<td>200</td>
<td>2.5</td>
<td>32</td>
<td>9.1</td>
<td>16</td>
</tr>
<tr>
<td>2.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>200</td>
<td>2.6</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>2.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>17</td>
<td>200</td>
<td>2.7</td>
<td>—</td>
<td>—</td>
<td>19</td>
</tr>
<tr>
<td>2.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>200</td>
<td>2.8</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>2.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>18</td>
<td>200</td>
<td>2.9</td>
<td>—</td>
<td>—</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>34</td>
<td>7.5</td>
<td>2</td>
<td>200</td>
<td>3</td>
<td>43</td>
<td>10.8</td>
<td>2</td>
</tr>
<tr>
<td>3.5</td>
<td>41</td>
<td>42</td>
<td>9.1</td>
<td>23</td>
<td>200</td>
<td>3.5</td>
<td>56</td>
<td>12.4</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>46</td>
<td>49</td>
<td>9.2</td>
<td>23</td>
<td>200</td>
<td>4</td>
<td>63</td>
<td>12.6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>60</td>
<td>9.3</td>
<td>9</td>
<td>200</td>
<td>5</td>
<td>93</td>
<td>15.5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>60</td>
<td>8.1</td>
<td>32</td>
<td>200</td>
<td>6</td>
<td>109</td>
<td>15.6</td>
<td>39</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>8</td>
<td>250</td>
<td>1</td>
<td>3</td>
<td>1.5</td>
<td>8</td>
</tr>
<tr>
<td>1.1</td>
<td>3</td>
<td>10</td>
<td>1.4</td>
<td>8</td>
<td>250</td>
<td>1.1</td>
<td>3</td>
<td>1.4</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>8</td>
<td>10</td>
<td>3.6</td>
<td>8</td>
<td>250</td>
<td>1.2</td>
<td>8</td>
<td>3.6</td>
<td>10</td>
</tr>
<tr>
<td>1.3</td>
<td>3</td>
<td>11</td>
<td>1.3</td>
<td>3</td>
<td>250</td>
<td>1.3</td>
<td>8</td>
<td>3.5</td>
<td>10</td>
</tr>
<tr>
<td>1.4</td>
<td>9</td>
<td>12</td>
<td>3.8</td>
<td>10</td>
<td>250</td>
<td>1.4</td>
<td>9</td>
<td>3.8</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>9</td>
<td>14</td>
<td>3.6</td>
<td>9</td>
<td>250</td>
<td>1.5</td>
<td>10</td>
<td>4.0</td>
<td>9</td>
</tr>
<tr>
<td>1.6</td>
<td>10</td>
<td>15</td>
<td>3.8</td>
<td>11</td>
<td>250</td>
<td>1.6</td>
<td>15</td>
<td>5.8</td>
<td>11</td>
</tr>
<tr>
<td>1.7</td>
<td>15</td>
<td>17</td>
<td>5.6</td>
<td>11</td>
<td>250</td>
<td>1.7</td>
<td>15</td>
<td>5.6</td>
<td>11</td>
</tr>
<tr>
<td>1.8</td>
<td>16</td>
<td>19</td>
<td>5.7</td>
<td>10</td>
<td>250</td>
<td>1.8</td>
<td>21</td>
<td>7.5</td>
<td>13</td>
</tr>
<tr>
<td>1.9</td>
<td>17</td>
<td>20</td>
<td>5.9</td>
<td>12</td>
<td>250</td>
<td>1.9</td>
<td>17</td>
<td>5.9</td>
<td>12</td>
</tr>
</tbody>
</table>
be a function measuring the imbalance in the marginal of values \(x\) and \((n-x)\): 
\(f(x) = 0\) for \(x = n/2\) (total balance) and 
\(f(x) = (n-2)^2/(n-1)\) for \(x = 1\) (maximum imbalance). 
Thus, if 
\[\beta_1 = (n_2 - n_1)^2/(2(n-1)/\{a_1 a_2 n_1 n_2 (n-2)^2\}\] 
is the square of the skewness coefficient of the hypergeometric distribution 
(Johnson & Kotz, 1969, Chapter 6), then 
\[\beta_1 = f(a_1) \times f(n_1)/f(1),\]
in other words, \(\beta_1\) is the product of the marginal imbalances in relation to the maximum. 
Given that the chi-squared test performs by approaching the hypergeometric distribution through the normal distribution, it seems reasonable to assume that the approximation will go badly when its skewness \(\beta_1\) is too far from the skewness of the normal distribution (0), 
or in other words when \(\beta_1 \geq B\), with \(B\) a positive constant to be determined experimentally. 
However, it seems clear that the magnitude of \(B\) will depend at least on the type of test used (one- or two-tailed) and of the \(\alpha\) error used. 
In an attempt to minimize its incidence, until now the authors have only dealt with those tables that are significant for \(\alpha\) nominal errors between 1% and 10%, separating the study of the one-tailed from the two-tail tests.

On the other hand, if \(f(n_1) = 0\) (\(K = 1\): a balanced marginal), it makes \(\beta_1 = 0 < B\), 
and so the chi-squared test would always be valid, which goes counter to the experimental data given in Table 2. 
For this not to occur, it is necessary to increase \(f(n_1)\) by a constant \(A\) (also to be determined), 
and so the approximate test would perform badly if 
\[\{f(n_1) + A\}f(a_1)/f(1) \geq B,\]
or reorganizing the terms:

\[\beta_1 = f(n_1)f(a_1)/f(1) \geq B - Af(a_1)/f(1)\]

(5)

thus the maximum allowed skewness is not a constant, but rather depends on 
\(f(a_1)/f(1)\): the relative imbalance of the most imbalanced marginal. 
The precaution of including the constant \(A\) is in keeping with the fact that when \(a_1\) is very small, 
it is well-nigh impossible for there to be a significant table for an \(\alpha\) nominal error between 1% and 10%. 
Finally, by making 
\(y = f(1)/f(a_1),\ x = f(n_1),\ a = A/B\) and 
\(b = 1/B\), expression (5) indicates that the chi-squared test is not valid if 
\(y < y^* = a + bx\). As condition \(y < y^*\) can be written as 
\(a_1 \leq a_1^{**}\), where:

\[a_1^{**} = \frac{n}{2} \left\{ 1 - \left( 1 + \frac{4(n-1)y^*}{(n-2)^2} \right)^{-0.5} \right\}\]

(6)

then expression (5) is compatible with what has been stated in the previous section.
and allows us to deduce that \( y^* = f(1)/f(a_1^*) \), since \( a_1^* \) was the first value of \( a_1 \) in which the chi-squared test is valid (and so \( y^* \) is the first value of \( y \) in which the same occurs). At this moment \( a_1^* = a_1^{**} \), but when predictions for \( a_1^* \) must be made, as in the case below, it is not so: hence the different symbols used.

In this fashion, conditions \( E < E^* \), \( a_1 < a_1^* \) or \( y < y^* \) are all equivalents and indicate that the chi-squared test is not valid, but now has a functional form for \( y^* \) (and consequently for \( a_1^* \) and \( E^* \)). In order to estimate the unknown parameters \( a \) and \( b \) it is sufficient to perform a linear regression:

\[
y^* = a + bx \quad \text{with} \quad x = f(n_1) = \frac{(K-1)^2}{K} \quad \text{and} \quad y^* = \frac{f(1)}{f(a_1^*)}
\]

based on the pairs \((x, y^*)\) which are deduced from Table 2. Once \( a \) and \( b \) are known, the predictions for \( y^* \) will be \( y^* = a + bx \) and the v.c. for the chi-squared test will be \( y > y^* \) or, equivalently, \( a_1 > [a_1^{**}]^- = a_1^* \), where \([x]^\) refers to the integral part of \( x \), \( a_1^{**} \) obtained from expression (6) and \([\]^-\) because \( a_1 \) is an integer number. From condition \( a_1 > a_1^* \), the classic v.c. of \( E > E^* = a_1^*/(K + 1) \) can immediately be deduced. One way of simplifying (6), which is the basic expression from which everything else that follows, is to consider that \((n - 2)^2/(n - 1) = f(1) \approx (n - 3) = (K + 1)n_1 - 3 \), to which it converges quite quickly, and so

\[
a_1^{**} \approx \frac{(K + 1)n_1}{2} \left\{ 1 - \frac{K(K + 1)n_1 - 3K}{\sqrt{K(K + 1)n_1 + (4a_3 - 3)K + 4b(K - 1)^2}} \right\}
\]

3.2 Estimating the parameters: predictions

In order to check the experimental validity of the model of bounded asymmetry and to estimate the parameters \( a \) and \( b \) in the said model, Figure 1 gives the scatter plot of points obtained by using the data from Table 2 (for one- and two-tailed tests), which represents the pairs \((x, y^*)\). It has already been seen that the growth in \( a_1^* \) must decrease when \( a_1^* \) comes close to \( n_1 \), so that these points will no longer fit properly into the line to be adjusted from \( y^* \) versus \( x \). This occurs with the four points of values \( n_1 = 40 \) or \( n_1 = 60 \) and \( K = 5 \) or \( K = 6 \), as is shown by the scatter plot in Fig. 1(a) corresponding to the case of the one-tailed test. The remaining 98 points adjust themselves quite well to the regression line \( y^* = 4.23 + 34.773x \), with a determination coefficient of \( r^2 = 0.99 \). However, of even greater relevance is the fact that the predictions \( a_1^* \) obtained by (8) fit well. For the 102 initial points (the four outliers are included), the quantity \( \text{DIF}(a_1) = a_1^* - a_1^* \) goes from \(-6 \) (in \( a_1^* = 21 \)) to \(+5 \) (in \( a_1^* = 109 \)), with 93% of the cases within \( \pm 3 \), 20% of the cases with \( \text{DIF}(a_1) = 0 \) and a mean of \( |\text{DIF}(a_1)| \) of 1.63. Excluding the low values \( a_1^* = 3 \), the value \( \text{RE}(a_1) = |a_1^* - a_1^*|/a_1^* \) (relative error) has a mean of 9.7%. For the 102 points, the determination coefficient between \( a_1^* \) and \( a_1^* \) is \( r^2(a_1) = 0.99 \) and the correlation ratio is \( r^2(a_1) = 0.99 \). All the above information is a guarantee that the proposed model of controlled skewness is quite appropriate and the condition of bounded skewness is strongly linked to the validity of the chi-squared test.

However, in practice, the researcher does not require a v.c. that performs well on average, but a v.c. that guarantees that s/he can apply the chi-squared test to one particular table with a degree of certainty, which means adjusting a conservative line like that drawn in Figure 1(a), guaranteeing that \( a_1^* \geq a_1^* \) (at least for the
observed data). This line is the one defined by the points where \( n_1 = 100 \) and \( K = 2.5 \) or \( 6 \), since it puts all the rest to one side and gives priority to the adjustment in small values of \( K \). A more reasonable possibility in principle would be to adjust a line parallel to the regression line (\( b = 34.773 \)) through the most divergent point from above (\( n_1 = 100; \ K = 6 \)), but these and other possibilities give worse predictions of \( a_k^* \). The optimal conservative line (for the one-tailed test) is thus given by:

\[
\delta^* = 11.194594 + 35.5282x
\]

which, naturally, plays a crucial role in what follows. For this, DIF(\( a_1 \)) goes from 0 (in \( a_k^* = 97 \)) to +10 (in \( a_k^* = 109 \)), with a mean absolute value of 4.4, so that RE(\( a_1 \)) has a mean of 22.15% in the values of \( a_k^* \neq 3 \). For all the points, \( r^2 (a_1) = 0.99 \) and \( r^2 (a_1) = 0.97 \). In the same Table 2, the predictions of \( \delta_k^* \) of \( a_k^* \) are shown. Expression (9) has been tested for new data obtained from both within and outside the original ranges of \( n_1 \) and \( K \), which always give \( \delta_k^* \geq a_k^* \) and are not too divergent. For example, for \( n_1 = 200 \) and \( K = 8 \) is \( a_k^* = 166 \) and \( \delta_k^* = 167 \); for \( n_1 = 300 \) and \( K = 1 \) is \( a_k^* = 3 \) and \( \delta_k^* = 10 \); for \( n_1 = 300 \) and \( K = 8 \) is \( a_k^* = 179 \) and \( \delta_k^* = 183 \); for \( n_1 = 120 \) and \( K = 3 \) is \( a_k^* = 91 \) and \( \delta_k^* = 93 \). The adjustment that yields (9) has also been carried out with other triplet sets (\( n_1, K, a_k^* \)), and it was noted that the obtained slope was quite stable (in fact the slope in (9) can be quite steep), while the intercept on the \( y \) axis is rather unstable (in fact its value in (9) can be somewhat low).

Another question is whether the previous methodology also works with other c.c.'s. Expression (9) gives the v.c.'s for \( \chi^2 \). For the same data of (\( n_1, K \)), the conservative line for \( \chi^2 \) test (which is a more traditional test) is \( \delta^* = 21.7733 + 36.2373x \), which yields stricter v.c.'s (since its intercept on the \( y \) axis and slope is greater than those in (9)), and some worse predictions, and this confirms the previous statement that \( \chi^2 \) is a better statistic than \( \delta^* \).

In the case of the two-tailed test, Figure 1(b) presents the scatter plot of \( y^* \) versus \( x \), showing various peculiarities. In the first place there are now more points that do not fit in well with the rest, although in a conservative way. In the second
place, two lines are observed, with a point of change in the neighbourhood of 
\( K = 3 \) and the first line having a greater slope than the second. Given the possible 
interest of the points with \( K = 2.6(0.1)2.9 \), these have been added to the points 
used for the one-tailed test. Finally, it can be seen that the adjustment to the model 
of bounded symmetry is not so good as for the one-tailed test. Omitting the 
outliers, there are 78 points remaining where \( K < 3 \) which give a \( r^2 = 0.90 \) and 13 
points with \( K \geq 3.5 \) which give a \( r^2 = 0.92 \) (lower than 0.99 for the case of the one-
tailed test). The conservative lines are (for the two-tailed test):

\[
\begin{align*}
\hat{y}^* &= 11.307150 + 13.3303x & \text{if } K < 3 \\
\hat{y}^* &= 17.064775 + 7.4455x & \text{if } K \geq 3
\end{align*}
\]

and the first is obtained by using the points where \( n_1 = 40 \) and \( K = 1.7 \) or 2.5 (thus 
giving priority to the adjustment in small values of \( K \)). The second one is obtained 
parallel to the line of regression and through the point where \( n_1 = 40 \) and \( K = 4 \) 
(since the large values of \( K \) are now of interest and the dispersion of the scatter 
plot is homogeneous). Table 2 indicates the \( \hat{a}^* \) predictions thus obtained. As 
far as predictions are concerned, the points with \( K = 3 \) have been included in the 
second equation since this gives more liberal results than the first (and the tables 
with \( K = 3 \) require very low values of \( a_i^* \)). The two lines intersect at \( K \approx 2.6 \), so 
probably the point of change from one rule of prediction to the other is located in 
this neighbourhood (as the graph seems to indicate). However, in the following the 
criterion of (10), which is a more conservative one, is maintained. By comparing 
(10) and (9) it can be seen that somewhat higher initial values of \( a_i^* \) are now 
demanded, but their growth is slower. The moment at which the \( a_i^* \) for one- and 
two-tailed tests coincide, is at \( K \approx 1.1 \).

4 Consequences

In the following, and where is there no ground for confusion, no distinction will 
be made between observed values (\( y^* \), \( a_i^* \) and \( E^* \)) and estimated ones (\( \hat{y}^* \), \( \hat{a}^* \) and 
\( \hat{E}^* \)), assuming that (9) and (10) describe reality. Now that the values of \( a \) and \( b \) 
are known, it is possible to extract some general and particular conclusions from 
the model such as those shown below.

When the model has been checked, expression (5) indicates that the smaller the 
most frequent marginal is—and the larger \( f(a_i) \) is—the more symmetry is 
required. Examples of extreme cases: (1) when \( f(a_i) = 0 \) (which occurs when all 
the marginals are balanced: \( a_i = n_i = n/2 \) the maximum possible (B) is allowed 
and, as \( \beta_1 = 0 \), the chi-squared test will always be valid; (2) when \( B - A \times f(a_i)/ 
\) \( f(1) \leq 0 \), equation (5) is always verified and the chi-squared test will not be 
valid. The previous inequality occurs (approximately) when \( a_i \leq n[1 - \{(n - 3)/ 
(4a + n - 3)\}^{0.5}] / 2 \), a quantity increasing to \( a \) when \( n \to \infty \), which indicates that a 
conservative rule is the fact that the non-balanced tables with \( a_i \leq [a]^- \) cannot be 
analysed by chi-squared (\( a_i \leq 11 \) for one-tailed test or two-tailed test with \( K < 3 \); 
\( a_i \leq 17 \) for two-tailed test with \( K \geq 3 \)).

The way to see if a particular table can be analysed by chi-squared test consists of 
checking that (5) is not verified: the chi-squared test will be valid when 
\( \{b(n_2 - n_1)^2 + an_1n_2\} \{a_2 - a\}^2(n - 1)/\{a_2a_1n_1n_2(n - 2)\} < 1 \). Alternatively and 
equivalently, the v.c. is \( y > y^* \), or \( a_i > a_i^* = [a_i^{**}]^- \), or, more suitably, that
Validity of the chi-squared test in $2 \times 2$ tables

$E = a_1 n_1 / n > E^* = a_1^* n_1 / n = a_1^*/(K + 1) = [a_1^*]^{-} / (K + 1)$ (the traditional one); that is:

$$E > E^*(n_1; K) = \left[ \frac{n}{2} \left( 1 - \left( 1 + \frac{4(n - 1)\{aK + b(K - 1)^2\}}{K(n - 2)^2} \right)^{-0.5} \right) \right]^{-}$$

(11)

In the following there is a study of the function $E^*(n_1; K)$.

How does $E^*$ change when $n_1$ or $K$ increase? As $a_1^*$ is the integral part of $a_1^{**}$, then $a_1^{**} - 1 < a_1^* < a_1^{**}$ and $(a_1^{**} - 1)/(K + 1) < E^* < a_1^{**}/(K + 1)$. So, for a fixed value of $K$, the evolution of $E^*$ with respect to $n_1$ is the same as that of $a_1^{**}/(K + 1)$ to be obtained from (20) omitting the symbol $[\cdot]^*$. Differentiating this expression in $n_1$ we obtain 'for $K$ constant, $E^*(n_1; K)$ increases with $n_1$, faster as $K$ increases'.

For the evolution of $E^*$ in $K$ (by fixed $n_1$) we have to work with the two expressions containing it and the result is not so clear as before, but broadly speaking it can be stated that 'for constant $n_1$, $E^*(n_1; K)$ decreases with $K$ as far as the proximity of 1.1; it then increases until it approaches the maximum and, finally, decreases to zero'. The two previous affirmations are of a 'macroscopic' type—Figure 2(a)—since the discrete type of $a_1^*$ values means that, in the first case, $E^*$ does not decrease with the increase of $n_1$, while, in the second case, $E^*$ decreases slowly and transitionally with the increase of $K$ from when an $a_1^*$ is obtained until the following $a_1^* + 1$ (since in this case $E^* \propto 1/(K + 1)$ the derivative of which is negative), and this would be a 'microscopic' description (Fig. 2(b)). In the case of two-tailed test, the results are similar. Now, if $K$ is constant, $E^*(n_1; K)$ increases with $n_1$ (the more quickly the larger the values of $K$), in each one of the two groups of $K$ values. Macroscopically and for $n_1$ constant, $E^*(n_1; K)$ decreases with $K$ until $K \approx 1.3$, then increases until it reaches a maximum and finally decreases until zero.

What happen when $n_1 \to \infty$ (for fixed value of $K$)? If equation (8) is multiplied and divided by the conjugate of the term between brace brackets, some mathematical operations are performed and the limit for $n_1 \to \infty$ is obtained, then $a_1^* \to [a + b(K - 1)^2/K]^{-}$ and $E^* \to \lim a_1^*/(K + 1)$. Consequently, since that $E^*(n_1; K)$

---

Fig. 2. Graph representing the minimum expected quantity $E^*(n_1, K)$ for the one-tailed chi-squared test to be valid. On the vertical axis $E^*$; on the horizontal plane $K$ and $n_1$. 
increases with \( n_1 \) (for fixed value of \( K \)), its maximum will be found by

\[
E^*(n_1 = \infty; K) = \frac{a + b(K-1)^2}{K+1} \tag{12}
\]

This function is always increasing with respect to \( K \), in such a way that the absolute maximum of \( E^*(n_1; K) \) is \( E^*(n_1 = \infty; K = \infty) = b \). This implies that a universal and conservative rule for the validity of the chi-squared test is that \( E > 3.55282 \) (one-tailed test), guaranteeing that the v.c. for such a test does not tend to infinity with the increase of \( n_1 \) and/or \( K \). For the two-tailed test, the maximum possible value of \( E^*(n_1; K) \) is 7.27 for the first line—\( K = 3 \) in (12)—and 7.4455 in the second (its slope) so that for two tails, a general and conservative v.c. is \( E \geq 7.45 \).

A particular case of the above is that of \( E^*(n_1 = \infty; K = 1) = [a]^{-1} = 2.55 \) (one- or two-tailed test). As \( E^*(n_1; K = 1) \leq E^*(n_1 = \infty, K = 1) \), then \( E > 5.5 \) is the v.c. for any table with \( n_1 = n_2 \), which is surprisingly close to the classic value of 5. By working this out with the function \( E^*(n_1; K) \) it can be proved that, for the one-tailed test, \( E > 5.5 \) is a valid v.c. for \( K \leq 1.16 \) (\( n_1 \leq 60 \) or \( K \leq 1.68 \) in two-tailed), and that the classic v.c. of \( E \geq 5 \) is valid for \( K \leq 1.19 \) and \( n_1 \leq 126 \) (\( K \leq 1.14 \) and \( n_1 \leq 300 \) in two-tailed): in other words, in small and well balanced samples.

Another consequence of the fact that \( E^*(n_1; K) \) increases with \( n_1 \) is that by fixing \( n_1 = \infty \), conservative bounds of \( E \) for each value of \( K \) can be obtained and these are valid for any value of \( n_1 \). For the values of \( K = 1; 1.5; 2; 5; 10 \) the bounds in the one-tailed test are \( E > 5.5; 6.8; 9.33; 20.67 \) and 27.09 respectively, values which are obtained from (12).

What happen when \( K \to \infty \) (for fixed value of \( n_1 \))? If in (8) we make \( K \to \infty \), it is clear that \( a^* \to \infty \). However, as \( n_1 \) is fixed and \( a_1 \) cannot be higher than \( a^* \), what happens is that \( a^* \to n_1 \), which will occur for a \( \hat{K}(n_1) \) in which \( a^* = n_1 \) in (8), that is, for a \( K \) such that \( n_1 = (3K^2 + aK(K-1) + b(K-1)^4) / \{K^2(K+1)\} \). Thus \( E^* \to n_1 / \{K(n_1) + 1\} \leq n_1 \{1 - a(n_1 + 4b)^{0.5} \}/2 \), if \( K \to K(n_1) \), and from \( K(n_1) \) on, it decreases to 0. For \( n_1 = 60; 100 \) and 1000 these are \( K(n_1) = 4.74; 6.17 \) and 32.6, respectively, in the one-tailed test. For the two-tailed test, the value \( K(n_1) \) where the maximum is reached is always in the second line (\( K \geq 3 \)) and comes quite a lot later than in the one-tailed test: for example, here \( K(n_1 = 60) = 10.39 \) compared to 4.74 in the other. Both results indicate that, although the maximum is reached relatively soon for low values of \( n_1 \), it is not reached in the majority of practical cases (small \( K \)'s) for large values of \( n_1 \).

Finally, before we have described the irregularities that take place in the two-tailed test. The abrupt fall in some \( a^* \) values may be explained as follows: (1) when \( K \) is close to one, the skewness is almost null and the two-tailed test with \( 1% \leq P_\chi \leq 10% \) behaves like the one-tailed test with \( 0.5% \leq P_\chi \leq 5% \) (that is why the former is stricter than the latter); (2) as the skewness increases, the over-estimation of \( P_\chi \) by \( P_\chi \) in one of the tails is compensated by the underestimation of \( P_\chi \) in the other tail (that is why the slope is smaller in the two-tailed test than in the one-tailed test); (3) when the skewness is excessive, the \( P_\chi \) are now only in one of the tails and the situation is altered again (probably from \( K = 3 \)). But it is clear that this cannot explain everything that takes place. It is probable that the kurtosis \( \beta_2 \) of the hypergeometric distribution, which is barely involved in the one-tailed test, has some role in the two-tailed test. This would be a matter worthy of further investigation. From the actual data can be verified that high values of
$|\gamma_2| = |\beta_2 - 3|$— the excess coefficient— compensate for the magnitude of the skewness, favouring the validity of the chi-squared test.

5 Conclusions

In this paper, we have seen how the validity of the chi-squared test for $2 \times 2$ tables, as an approximation of Fisher's exact test, is fundamentally linked to the skewness of the underlying hypergeometric distribution, and that when the skewness coefficient is less than the given number, the chi-squared test is valid. In fact, a simple (and conservative) way of checking if the chi-squared test is valid in a particular table is through the verification of (5), or, in a classic way, through minimum expected quantity $E$, in which case the chi-squared statistic is valid if (11) is verified. The quantity $E^*(n_1; K)$ has been experimentally shown to be very variable, to depend on $n_1$ and on $K$, and very often to be a good deal larger than the classic value of 5 (especially in the one-tailed tests). Consequently, the minimum value $E^*(n_1; K)$ of $E$ for the approximate test to be valid, is not a constant as is traditionally supposed, but generally increases with $n_1$ and $K$ (although, fortunately, it reaches an absolute maximum). In particular, the classic condition $E \geq 5$ (or the new $E \geq 5.5$) is only valid for tables with small marginals and (or) with at least one of them balanced.

The only way of giving bounds which are generally valid is by making a very conservative statement: the approximate test is valid if $E \geq 35.53$ (one-tailed test) or $E \geq 7.45$ (two-tailed test). It has also been seen in a general (and very conservative) way that non-balanced tables whose minimum marginal is less or equal to 11 (one-tailed test) or 17 (two-tailed test) cannot be analysed by a chi-squared test, and that the chi-squared test is always valid when both marginals are balanced. A possibility of fixing intermediate validity conditions between demanding the no verification of (5) and giving a sole (very conservative) value of $E^*(n_1; K)$, is through the function $E^*(n_1 = \infty; K)$ given by (12), which is conservatively valid for any sample size.

The numerical results above refer to the optimal chi-squared test which, although in the case of two tails is Yates' classic test (with Mantel's caution), in the case of the one-tailed test is not Yates', but the test obtained by Conover's criterion. For Yates' one-tailed test, the general conclusions above are also valid, except in the implied constants, and the validity conditions are rather stricter (for example $E \geq 36.24$ is the general condition). Whatever the case, the validity conditions are stricter in the one-tailed than in the two-tailed test (except for tables with a well-balanced marginal).

To obtain the above results, the equation $y^* = a + bx$, in which $x$ and $y^*$ depend only on the marginals of the table, and the constants $a$ and $b$ have different physical significances, has been determined experimentally. On the one hand $[a]$ is the minimum value of the smallest marginal for the chi-squared test to be valid in a non-balanced table, while $[a]^-/2$ is the minimum value of $E$ for the validity of the test when one of the marginals is balanced. On the other hand, $b$ is the maximum possible value of $E^*$. At the same time, both are related to the maximum skewness coefficient allowed, since $A = a/b$ and $B = 1/b$, with $B$ being the maximum skewness allowed in a table for the chi-squared test to be valid, and $A$ being a reducing constant of $B$ (which acts proportionally on the relative skewness of the most asymmetric marginal). The values of $B$ are of the order of 0.03 in the one-tailed test and 0.10 in the two-tailed test.
Acknowledgements

This research was financed by the Dirección General de Enseñanza Superior (PB96–1443), Spain. The authors wish to thank David Prieto, University of Alcalá (Madrid, Spain), for several editorial comments and suggestions that have greatly improved the clarity and exposition of the paper.

REFERENCES