# A LOCATION-DISPERSION TEST FOR $ 2 ¥times k $ TABLES

Jayasekara, Leslie
Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University

Yanagawa, Takashi
Department of Mathematics, Kyushu University

Tsujitani, Masaaki
Kobe Women's University

https://doi.org/10.5109/13438

出版情報：Bulletin of informatics and cybernetics. 26 (1/2), pp.125-139, 1994-03. Research Association of Statistical Sciences

*Bulletin of Informatics and Cybernetics, Vol. 26, No. 1~2, 1994*

# A LOCATION-DISPERSION TEST FOR 2×$k$ TABLES

By

**Leslie JAYASEKARA*, Takashi YANAGAWA† and Masaaki TSUJITANI‡**

## Abstract

A location-dispersion test for $2 \times k$ contingency tables is proposed. The asymptotic distributions of the proposed test are obtained both under the null and alternative hypotheses, and also its power and efficiency are studied. The proposed test is compared with several other chi-squared tests by Monte Carlo studies and it is shown that the test is superior to Pearson's chi-squared test, Nair's location and dispersion tests and to the cumulative chi-squared test in many cases.

*Key Words and Phrases: contingency table; asymptotic distribution; two-sample test; Wilcoxon test; Mood statistic; Pearson's chi-squared test; Nair's location and dispersion tests; cumulative chi-squared test.*

## 1. Introduction

Nonparametric two-sample tests for testing identity of distributions versus alternatives containing both location and scale parameters have been proposed by Lepage [3, 4], Duran, Tsai and Lewis [1] and others. The purpose of this paper is to explore an analogous test which is applicable for categorical data.

Table 1: $2 \times k$ contingency table

| | *Categories* | | | | | Total |
|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | . | . | $X_k$ | $n_1$ |
| | $Y_1$ | $Y_2$ | . | . | $Y_k$ | $n_2$ |
| Total | $\tau_1$ | $\tau_2$ | . | . | $\tau_k$ | $N$ |

* Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University 39, Kasuga 816, Japan

† Department of Mathematics, Kyushu University 33, Fukuoka 812, Japan

‡ Kobe Women's University, Kobe 654, Japan

Consider $2 \times k$ table given in Table 1. We suppose that $X = (X_1, \cdots, X_k)$ and $Y = (Y_1, \cdots, Y_k)$ are independent multinomial random vectors and consider the following null hypothesis:

$H_0$ : $X$ and $Y$ are identically distributed.

We propose a location-dispersion test for this hypothesis which is sensitive for both location shift and scale shift which seem common, for example, in clinical trial data. The proposed statistic will be expressed as functions of the Wilcoxon test and test whose score is orthonormal to the Wilcoxon score.

In section 2 we propose the statistic. The asymptotic distributions of the proposed statistic under the hypothesis and contiguous alternatives are studied in section 3. Its power and efficiency are studied in section 4, and finally in section 5, Monte Carlo studies are conducted to show the behaviors of the test. It is shown that the proposed test is superior to Pearson's chi-squared test, Nair's location and dispersion tests, and to the cumulative chi-squared test in many cases.

## 2. The Test Statistic

Let $c_i$, $i = 1, 2, \cdots, k$, be the Wilcoxon score defined by $c_i = \sum_{j=1}^{i-1} \tau_j + (\tau_i - N)/2$ for $i = 1, 2, \cdots, k$, so that $\sum_{i=1}^{k} \tau_i c_i = 0$. We define the inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$ by $(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{k} \tau_i a_i b_i$ and $||\mathbf{a}||^2 = (\mathbf{a}, \mathbf{a})$. For the normalized score,

$$\mathbf{a} = \frac{\mathbf{c}}{||\mathbf{c}||}, \tag{1}$$

where $\mathbf{c} = (c_1, c_2, \cdots, c_k)'$, we consider the Wilcoxon test statistic which is defined by:

$$S = \sum_{i=1}^{k} a_i Y_i.$$

Next we consider a score orthonormal to the Wilcoxon score. Let $d_i^*$, $i = 1, 2, \cdots, k$, be the score defined by $d_i^* = c_i^2$, and put $\mathbf{d}^* = (d_1^*, d_2^*, \cdots, d_k^*)'$ and $d_i = d_i^* - (\mathbf{d}^*, \mathbf{a})a_i$. For the normalized score,

$$\mathbf{b} = \frac{\mathbf{d}}{||\mathbf{d}||}, \tag{2}$$

where $\mathbf{d} = (d_1, d_2, \cdots, d_k)'$, we consider the test statistic defined by:

$$T = \sum_{i=1}^{k} b_i Y_i.$$

We call $\mathbf{b}$ the dispersion score and $T$ the dispersion test statistic. We consider statistics $S$ and $T$ conditioned on $C = \{n_1, n_2, \tau_1, \tau_2, \cdots, \tau_k\}$. The test statistic for $H_0$ we propose is

$$Q = \frac{S^2}{V_0[S|C]} + \frac{(T - E_0[T|C])^2}{V_0[T|C]},$$

where $E_0[\cdot|C]$ and $V_0[\cdot|C]$ are the conditional expectation and variance conditioned on $C$, respectively, under $H_0$. These expectation and variance are given by

$$E_0[S|C] = 0, \qquad V_0[S|C] = \tfrac{n_1 n_2}{N(N-1)},$$
$$E_0[T|C] = n_2\bar{b}, \quad V_0[T|C] = \tfrac{n_1 n_2}{N(N-1)}(1 - N\bar{b}^2),$$

where $\bar{b} = (1/N)\sum_{i=1}^{k} b_i\tau_i$. It is straightforward to show these formulae since we have

$$(\mathbf{a}, \mathbf{1}) = (\mathbf{a}, \mathbf{b}) = 0, \qquad (\mathbf{1} = (1, 1, \cdots, 1)') \tag{3}$$

$$||\mathbf{a}|| = ||\mathbf{b}|| = 1, \tag{4}$$

and furthermore the conditional distribution of $Y$ conditioned on $C$ is given by

$$\Pr[(Y_1, \cdots, Y_k) = (y_1, \cdots, y_k)|C] = \frac{\begin{pmatrix} \tau_1 \\ y_1 \end{pmatrix} \cdots \begin{pmatrix} \tau_k \\ y_k \end{pmatrix}}{\begin{pmatrix} N \\ n_2 \end{pmatrix}}.$$

Treating all of the observations in a certain category as being tied and assigning them a score equal to the midrank for that category, Nair [5] introduced statistics $\mathrm{SS}_A(l)$ and $\mathrm{SS}_A(d)$ from the Wilcoxon and Mood rank statistics, respectively. It is easy to see that $\mathrm{SS}_A(l)$ is equivalent to $S^2/V_0[S|C]$, and $\mathrm{SS}_A(d)$ is almost equivalent to $(T - E_0[T|C])^2/V_0[T|C]$. In this paper we call the tests based on $\mathrm{SS}_A(l)$ and $\mathrm{SS}_A(d)$ the Nair's location test and dispersion test, respectively.

## 3. Asymptotic Distributions

### 3.1. The Multiple Non-Central Hypergeometric Distribution

When $X$ and $Y$ are independently distributed multinomial random variables with the parameters $n_1, \pi_1 = (\pi_{11}, \cdots, \pi_{1k})$ and $n_2, \pi_2 = (\pi_{21}, \cdots, \pi_{2k})$, respectively, we have,

$$\Pr[(Y_2, \cdots, Y_k) = (y_2, \cdots, y_k)|C] = \frac{g(\mathbf{y})\psi_1^{y_1} \cdots \psi_k^{y_k}}{\sum_{j_1+\cdots+j_k=n_2} g(\mathbf{j})\psi_1^{j_1} \cdots \psi_k^{j_k}}, \tag{5}$$

where $g(\mathbf{y}) = n_1!n_2!/[y_1!\cdots y_k!(\tau_1 - y_1)!\cdots(\tau_k - y_k)!]$ and the odds-ratio parameters $\psi_j$, relative to category 1, for $j = 1, 2, \cdots, k$, are defined by $\psi_j = \pi_{11}\pi_{2j}/\pi_{21}\pi_{1j}$, so that $\psi_1 \equiv 1$.

We use the following assumptions:

(A1) $\lim_{N\to\infty} \frac{\tau_j}{N} = 0$, for $j = 2, 3, \cdots, k$.

(A2) $\frac{n_2}{N} = p$ for some given $p$ such that $0 < p < 1$.

THEOREM 3.1. *Suppose that* (A1) *and* (A2) *are satisfied, then*

$$\lim_{N\to\infty} \frac{g(\mathbf{y})\tau_1!\cdots\tau_k!}{N!\prod_{j=2}^{k}\begin{pmatrix} \tau_j \\ y_j \end{pmatrix} p^{y_j}(1-p)^{\tau_j - y_j}} = 1.$$

PROOF. Since

$$g(\mathbf{y})\frac{\tau_1!\cdots\tau_k!}{N!} = \frac{\frac{n_2!}{(n_2-y_2-\cdots-y_k)!}\frac{n_1!}{(n_1-\tau_2-\cdots-\tau_k+y_2+\cdots+y_k)!}}{\frac{N!}{(N-\tau_2-\cdots-\tau_k)!}}\begin{pmatrix}\tau_2\\ y_2\end{pmatrix}\cdots\begin{pmatrix}\tau_k\\ y_k\end{pmatrix},$$

from (A1) and (A2) we have

$$\lim_{N\to\infty} g(\mathbf{y})\frac{\tau_1!\cdots\tau_k!}{N!} = \prod_{j=2}^{k}\begin{pmatrix}\tau_j\\ y_j\end{pmatrix}p^{y_j}(1-p)^{\tau_j-y_j}$$

. This completes the proof of the theorem. □

From the theorem it follows that the numerator of (5) is approximated by

$$\prod_{j=2}^{k}\begin{pmatrix}\tau_j\\ y_j\end{pmatrix}(p\psi_j)^{y_j}(1-p)^{\tau_j-y_j}\frac{N!}{\tau_1!\cdots\tau_k!}.$$

Therefore by normalizing this we may approximate the distribution (5) by

$$\Pr[(Y_2,\cdots,Y_k)=(y_2,\cdots,y_k)|C] = \prod_{j=2}^{k}\begin{pmatrix}\tau_j\\ y_j\end{pmatrix}\left(\frac{p\psi_j}{p\psi_j+1-p}\right)^{y_j}\left(\frac{1-p}{p\psi_j+1-p}\right)^{\tau_j-y_j}. \tag{6}$$

Thus according to this approximation $Y_j$ is binomialy distributed with parameters $\tau_j$ and $p\psi_j/(p\psi_j+1-p)$, and $Y_2, Y_3, \cdots, Y_k$ are independent.

### 3.2. Asymptotic Distribution of $Q$ Under $H_0$

Now we consider an approximation of $\Pr[(Y_2,\cdots,Y_k)=(y_2,\cdots,y_k)|C]$ for large $\tau_j$, $j=2,3,\cdots,k$. To begin with we shall make the following assumption.

(A3) $N^{-\epsilon}\tau_j = O(1)$, $j=2,3,\cdots,k$, for some $\epsilon$ such that $0<\epsilon<\frac{1}{2}$.

The notation $N^{-\epsilon}\tau_j = O(1)$ which is used in this paper, means $N^{-\epsilon}\tau_j$ tends to a constant as $N\to\infty$. Note that (A3) includes (A1).

LEMMA 3.2. *If* (A3) *is satisfied, then*

(i) $N^{1-\frac{\epsilon}{2}}a_1 = O(1)$, $N^{\frac{\epsilon}{2}}a_i = O(1)$, $i=2,3,\cdots,k$.

(ii) $N^{1-\frac{\epsilon}{2}}b_1 = O(1)$, $N^{\frac{\epsilon}{2}}b_i = O(1)$, $i=2,3,\cdots,k$.

PROOF. From (A3) and using the definitions of $c_i$ and $d_i$ we have

$N^{-\epsilon}c_1 = O(1)$, $N^{-1}c_i = O(1)$,

$N^{-1-\epsilon}d_1 = O(1)$, and $N^{-2}d_i = O(1)$ for $i=2,3,\cdots,k$.

Thus the proof is completed from the definitions of $a_i$ and $b_i$ which are given in (1) and (2). □

Denote by $E_{A0}[\cdot]$ and $V_{A0}[\cdot]$ the expectation and variance, respectively, under $H_0$ when $Y_j$, $j = 2, 3, \cdots, k$, follow distribution given in (6). We have the following lemma.

LEMMA 3.3. *If* (A2) *and* (A3) *are satisfied, then*

(i) $\mathrm{E}_{A0}[S] = 0$,

(ii) $\lim_{N\to\infty} \frac{V_0[S|C]}{V_{A0}[S]} = 1$.

PROOF. Since $\sum_{i=1}^{k} Y_i = n_2$, we have

$$S = a_1 n_2 + \sum_{i=2}^{k} (a_i - a_1) Y_i. \tag{7}$$

Thus

$$\begin{aligned} E_{A0}[S] &= a_1 n_2 + \sum_{i=2}^{k} (a_i - a_1) E_{A0}[Y_i], \qquad (8) \\ &= a_1 n_2 + \sum_{i=2}^{k} (a_i - a_1) \tau_i p, \\ &= a_1 N \left( \frac{n_2}{N} - p \right), \end{aligned}$$

since $\sum_{i=1}^{k} a_i \tau_i = 0$. Thus from (A2) we have (i). Next we prove (ii). From (7) we have

$$\begin{aligned} V_{A0}[S] &= \sum_{i=2}^{k} (a_i - a_1)^2 V_{A0}[Y_i], \\ &= \sum_{i=2}^{k} (a_i - a_1)^2 \tau_i p(1-p), \\ &= p(1-p)(1 + a_1^2 N). \end{aligned}$$

Therefore using (A2) and Lemma 3.2 we have

$$\frac{V_0[S|C]}{V_{A0}[S]} = \frac{n_1 n_2 / N(N-1)}{p(1-p)(1+a_1^2 N)} \to 1 \quad as \quad N \to \infty. \qquad \square$$

THEOREM 3.4. *If* (A2) *and* (A3) *are satisfied, then under* $H_0$*, the distribution of* $S/\sqrt{V_0[S|C]}$ *may be approximated by a standard normal distribution for large* $N$.

PROOF. We can write

$$\Pr\left[\frac{S}{\sqrt{V_0[S|C]}} \le x\right] = \Pr\left[\frac{S - E_{A0}[S]}{\sqrt{V_{A0}[S]}} \le \sqrt{\frac{V_0[S|C]}{V_{A0}[S]}} \left\{ x - \frac{E_{A0}[S]}{\sqrt{V_0[S|C]}} \right\} \right].$$

When $N \to \infty$, (A3) $\Rightarrow$ (A1), and the distribution of $(Y_2, Y_3, \cdots, Y_k)$ may be approximated by the multiple of the independent binomial distributions. Furthermore

$\tau_j \to \infty$, $j = 2, 3, \cdots, k$ when $N \to \infty$ from (A3). Therefore the distribution of $(S - E_{A0}[S])/\sqrt{V_{A0}[S]}$ may be approximated by $N(0,1)$ for large $N$. Then the result follows from Lemma 3.3 and by Slutsky theorem (See Lehmann [2] Appendix, Sec. 3, Corollary 2). □

We next consider statistic $T$. We first show:

LEMMA 3.5. *If* (A2) *and* (A3) *are satisfied, then*

(i) $E_{A0}[T] - E_0[T|C] = 0$,

(ii) $\lim_{N\to\infty} \frac{V_0[T|C]}{V_{A0}[T]} = 1$.

PROOF. Since $\sum_{i=1}^{k} Y_i = n_2$, we have

$$T = b_1 n_2 + \sum_{i=2}^{k} (b_i - b_1) Y_i, \tag{9}$$

and

$$\begin{aligned} E_{A0}[T] &= b_1 n_2 + \sum_{i=2}^{k} (b_i - b_1) E_{A0}[Y_i], \\ &= b_1 n_2 + \sum_{i=2}^{k} (b_i - b_1) \tau_i p, \\ &= b_1 n_2 + pN\bar{b} - b_1 pN. \end{aligned}$$

Now since

$$E_{A0}[T] - E_0[T|C] = b_1 N \left(\frac{n_2}{N} - p\right) - \bar{b} N \left(\frac{n_2}{N} - p\right),$$

the proof of (i) is immediate from (A2). Next,

$$\begin{aligned} V_{A0}[T] &= \sum_{i=2}^{k} (b_i - b_1)^2 V_{A0}[Y_i], \\ &= \sum_{i=2}^{k} (b_i - b_1)^2 \tau_i p(1-p), \\ &= p(1-p)(1 - 2N b_1 \bar{b} + b_1^2 N). \end{aligned}$$

Thus it follows that

$$\frac{V_0[T|C]}{V_{A0}[T]} = \frac{n_1 n_2 / N(N-1)}{p(1-p)} A,$$

where $A = (1 - N\bar{b}^2)/(1 - 2Nb_1\bar{b} + b_1^2 N)$. Using Lemma 3.2 we can show $A$ tends to 1 as $N$ tends to infinity. So from (A2) the result follows. □

THEOREM 3.6. *If* (A2) *and* (A3) *are satisfied, then under* $H_0$, *the distribution of* $(T - E_0[T|C])/\sqrt{V_0[T|C]}$ *may be approximated by a standard normal distribution for large* $N$.

PROOF. Similarly as the proof of Theorem 3.4, the distribution of $(T - E_{A0}[T])/\sqrt{V_{A0}[T]}$ may be approximated by $N(0,1)$ for large $N$, under (A3), and

$$\Pr\left[\frac{T - E_0[T|C]}{\sqrt{V_0[T|C]}} \leq x\right] = \Pr\left[\frac{T - E_{A0}[T]}{\sqrt{V_{A0}[T]}} \leq \sqrt{\frac{V_0[T|C]}{V_{A0}[T]}}\left\{x - \frac{E_{A0}[T] - E_0[T|C]}{\sqrt{V_0[T|C]}}\right\}\right].$$

Thus it is straightforward to show the theorem from Lemma 3.5 and by Slutsky theorem. □

THEOREM 3.7. *If* (A2) *and* (A3) *are satisfied then the distribution of $Q$ under $H_0$ may be approximated by a chi-squared distribution with 2 degrees of freedom for large $N$.*

PROOF. It is sufficient to show that the conditional covariance of $S$ and $T$ conditioned on $C$ under $H_0$, which is denoted by $Cov_0[S,T|C]$, is 0. From (7) and (9) we have

$$\begin{aligned} Cov_0[S,T|C] &= \sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)V_0[Y_i|C] \\ &\qquad + \sum_{i=2}^{k}\sum_{j=2, i\neq j}^{k} (a_i - a_1)(b_j - b_1)Cov_0[Y_i, Y_j|C], \\ &= \sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\frac{\tau_i(N-\tau_i)n_1 n_2}{N^2(N-1)} \\ &\qquad - \sum_{i=2}^{k}\sum_{j=2, i\neq j}^{k} (a_i - a_1)(b_j - b_1)\frac{\tau_i \tau_j n_1 n_2}{N^2(N-1)}, \\ &= \sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\frac{\tau_i n_1 n_2}{N(N-1)} \\ &\qquad - \sum_{i=2}^{k}\sum_{j=2}^{k}(a_i - a_1)(b_j - b_1)\frac{\tau_i \tau_j n_1 n_2}{N^2(N-1)}. \end{aligned}$$

Now

$$\sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\tau_i = a_1(b_1 - \bar{b})N,$$

$$\begin{aligned} \sum_{i=2}^{k}\sum_{j=2}^{k}(a_i - a_1)(b_j - b_1)\tau_i\tau_j &= \left[\sum_{i=2}^{k}(a_i - a_1)\tau_i\right]\left[\sum_{j=2}^{k}(b_j - b_1)\tau_j\right], \\ &= a_1(b_1 - \bar{b})N^2. \end{aligned}$$

Thus substituting these equalities into the above formula, we have the desired result. □

### 3.3. Asymptotic Distribution of $Q$ Under the Alternative

In this section we obtain asymptotic distribution of $Q$ under the alternative hypothesis $H_1 : \psi_j = 1 + A_j/N^{\epsilon}$, $j = 1, 2, \cdots, k$, where $A_j$ is a constant and $0 < \epsilon < 1/2$. We denote by $E_A[\cdot]$ and $V_A[\cdot]$ the expectation and variance, respectively, under $H_1$ when $Y_j$, $j = 2, 3, \cdots, k$, follow distribution given in (6).

LEMMA 3.8. *If* (A2) *and* (A3) *are satisfied, then*

(i) $\lim_{N\to\infty} \frac{E_A[S]}{\sqrt{V_0[S|C]}} = l_S$,

(ii) $\lim_{N\to\infty} \frac{V_0[S|C]}{V_A[S]} = 1$,

*where* $l_S = \sqrt{p(1-p)} \sum_{j=2}^{k} (a_j - a_1)\tau_j(\psi_j - 1)$.

PROOF. (i) From (7) we have

$$\begin{aligned} E_A[S] &= a_1 n_2 + \sum_{j=2}^{k} (a_j - a_1) E_A[Y_j], \\ &= a_1 n_2 + \sum_{j=2}^{k} (a_j - a_1) \frac{\tau_j p \psi_j}{p\psi_j + 1 - p}. \end{aligned}$$

Substituting (8), we have

$$\begin{aligned} E_A[S] &= E_A[S] - E_{A0}[S] + E_{A0}[S], \\ &= \sum_{j=2}^{k} (a_j - a_1)\tau_j \left[\frac{p(1-p)(\psi_j - 1)}{p\psi_j + 1 - p}\right] + E_{A0}[S]. \end{aligned}$$

Now consider $E_A[S]/\sqrt{V_0[S|C]}$ and using (A2) the result follows.
(ii) From (7) we have

$$\begin{aligned} V_A[S] &= \sum_{j=2}^{k} (a_j - a_1)^2 V_A[Y_j], \\ &= \sum_{j=2}^{k} (a_j - a_1)^2 \frac{\tau_j p(1-p)\psi_j}{(p\psi_j + 1 - p)^2}. \end{aligned}$$

Thus it follows that

$$\frac{V_0[S|C]}{V_A[S]} = \frac{n_1 n_2 / N(N-1)}{p(1-p)A},$$

where $A = \sum_{j=2}^{k} (a_j - a_1)^2 \tau_j \psi_j / (p\psi_j + 1 - p)^2$. Furthermore since

$$\sum_{j=2}^{k} (a_j - a_1)^2 \tau_j = 1 + a_1^2 N,$$

we can show by using Lemma 3.2 that $A \to 1$ as $N \to \infty$ under $H_1$. So from (A2) the proof is completed. □

THEOREM 3.9. *If* (A2) *and* (A3) *are satisfied, then under* $H_1$ *the distribution of* $S/\sqrt{V_0[S|C]}$ *may be approximated by* $N(l_S, 1)$ *for large* $N$.

PROOF. Similarly as the proof of Theorem 3.4, the distribution of $(S - E_A[S])/\sqrt{V_A[S]}$ may be approximated by $N(0,1)$, for large $N$, under (A3) and

$$\Pr\left[\frac{S}{\sqrt{V_0[S|C]}} \leq x\right] = \Pr\left[\frac{S - E_A[S]}{\sqrt{V_A[S]}} \leq \sqrt{\frac{V_0[S|C]}{V_A[S]}}\left\{x - \frac{E_A[S]}{\sqrt{V_0[S|C]}}\right\}\right].$$

Thus it is straightforward to show the theorem from Lemma 3.8 and by Slutsky theorem. □

LEMMA 3.10. *If* (A2) *and* (A3) *are satisfied, then*

(i) $\lim_{N\to\infty} \frac{E_A[T] - E_0[T|C]}{\sqrt{V_0[T|C]}} = l_T$,

(ii) $\lim_{N\to\infty} \frac{V_0[T|C]}{V_A[T]} = 1$,

*where* $l_T = \sqrt{p(1-p)}\sum_{j=2}^{k}(b_j - b_1)\tau_j(\psi_j - 1)$.

PROOF. (i) From (9) we have

$$\begin{aligned} E_A[T] &= b_1 n_2 + \sum_{j=2}^{k}(b_j - b_1)E_A[Y_j], \\ &= b_1 n_2 + \sum_{j=2}^{k}(b_j - b_1)\tau_j \frac{p\psi_j}{1 + p(\psi_j - 1)}, \end{aligned}$$

and

$$E_A[T] - E_0[T|C] = \sum_{j=2}^{k}(b_j - b_1)\tau_j \frac{p(1-p)(\psi_j - 1)}{1 + p(\psi_j - 1)} + E_{A0}[T] - E_0[T|C].$$

Dividing this by $\sqrt{V_0[T|C]}$ and using Lemma 3.5, the result follows. (ii) From (9) we have

$$\begin{aligned} V_A[T] &= \sum_{j=2}^{k}(b_j - b_1)^2 V_A[Y_j], \\ &= \sum_{j=2}^{k}(b_j - b_1)^2 \frac{\tau_j p(1-p)\psi_j}{(p\psi_j + 1 - p)^2}. \end{aligned}$$

Thus it follows that

$$\frac{V_0[T|C]}{V_A[T]} = \frac{n_1 n_2/N(N-1)}{p(1-p)}B,$$

where $B = (1 - N\bar{b}^2)/\sum_{j=2}^{k}(b_j - b_1)^2 \frac{\tau_j\psi_j}{(p\psi_j+1-p)^2}$. From (A3), $B \to 1$ as $N \to \infty$. So using (A2) the result follows. □

THEOREM 3.11. *If* (A2) *and* (A3) *are satisfied, then under* $H_1$ *the distribution of* $(T - E_0[T|C])/\sqrt{V_0[T|C]}$ *may be approximated by* $N(l_T, 1)$ *for large* $N$.

PROOF. Similarly as the proof of Theorem 3.4, the distribution of $(T - E_A[T])/\sqrt{V_A[T]}$ may be approximated by $N(0,1)$ for large $N$, under (A3) and

$$\Pr\left[\frac{T - E_0[T|C]}{\sqrt{V_0[T|C]}} \le x\right] = \Pr\left[\frac{T - E_A[T]}{\sqrt{V_A[T]}} \le \sqrt{\frac{V_0[T|C]}{V_A[T]}}\left\{x - \frac{E_A[T] - E_0[T|C]}{\sqrt{V_0[T|C]}}\right\}\right].$$

Thus it is straightforward to show the theorem from Lemma 3.10 and by Slutsky theorem. □

LEMMA 3.12. *Suppose that* $Y_j$, $j = 2, 3, \cdots, k$, *follow the distribution given in* (6), *and that* (A2) *and* (A3) *are satisfied, then under* $H_1$ *the correlation of* $S$ *and* $T$ *tends to* $0$ *as* $N \to \infty$.

PROOF. We denote by $Cov_A[S,T]$, the covariance of $S$ and $T$, under $H_1$ and when $Y_j$, $j = 2, 3, \cdots, k$ are supposed to follow the distribution given in (6), and we have

$$Cov_A[S,T] = \sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\tau_i q_i(1 - q_i),$$

where $q_i = p\psi_i/(p\psi_i + 1 - p)$ and $\psi_i = 1 + A_i/N^{\epsilon}$. Thus

$$|Cov_A[S,T]| \le (1/4)\sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\tau_i.$$

Now from Lemma 3.2 and (A3) we may show that

$$\begin{aligned}\sum_{i=2}^{k}(a_i - a_1)(b_i - b_1)\tau_i &= -a_1\sum_{i=1}^{k}(b_i - b_1)\tau_i,\\ &= \frac{c}{N^{1-\epsilon}} + O\left(\frac{1}{N^{1-\epsilon}}\right),\end{aligned}$$

for some constant $c$. Thus we have $Cov_A[S,T] \to 0$ as $N \to \infty$.
Employing again Lemma 3.2 and (A3) we may show that

$$V_A[S] = \frac{p}{1-p}\left[1 + \frac{c'}{N^{1-\epsilon}} + O\left(\frac{1}{N^{1-\epsilon}}\right)\right],$$

for some constant $c'$. Thus $V_A[S] \to \frac{p}{1-p}$ as $N \to \infty$.
Similarly we have $V_A[T] \to \frac{p}{1-p}$ as $N \to \infty$. Therefore we have the desired result. □

Summarizing Theorem 3.9, 3.11 and Lemma 3.12 we have the following theorem.

THEOREM 3.13. *If* (A2) *and* (A3) *are satisfied, then the distribution of* $Q$ *under* $H_1$ *may be approximated by a noncentral chi-squared distribution with 2 degrees of freedom and noncentrality parameter* $l_S^2 + l_T^2$, *when* $N$ *is large.*

## 4. Power and Efficiency

The power of the location-dispersion test is given by

$$\Pi_Q = \Pr_{H_1}\left[Q > \chi^2_{2,\alpha}\right],$$

where $\chi^2_{2,\alpha}$ is the upper $100\alpha$ percentage point of the chi-squared distribution with 2 degrees of freedom. From Theorem 3.13 we have

$$\lim_{N\to\infty} \Pi_Q = \Pr\left[{\chi'}^2_2 > \chi^2_{2,\alpha}\right],$$

where ${\chi'}^2_2$ is the density function of the noncentral chi-squared distribution with 2 degrees of freedom and the noncentrality parameter $\lambda = l_S^2 + l_T^2$. It is well known that the power is the increasing function of $\lambda$. Now using Schwarz inequality we have

$$\begin{aligned} l_S^2 &\le N^2 p(1-p)\sum_{i=2}^{k}(a_i - a_1)^2\left(\frac{\tau_i}{N}\right)\sum_{j=2}^{k}(\psi_j - 1)^2\left(\frac{\tau_j}{N}\right), \\ l_T^2 &\le N^2 p(1-p)\sum_{i=2}^{k}(b_i - b_1)^2\left(\frac{\tau_i}{N}\right)\sum_{j=2}^{k}(\psi_j - 1)^2\left(\frac{\tau_j}{N}\right). \end{aligned}$$

Thus $l_S^2$ is maximized when $\psi_j - 1 = \beta(a_j - a_1)$, that is $log\psi_j = \beta(a_j - a_1)$ approximately. Similarly $l_T^2$ is maximized when $log\psi_j = \beta(b_j - b_1)$ approximately. These findings characterize the optimality of the location score test and dispersion score test. Unfortunately, however, it is difficult to find the alternative hypothesis which maximize the power of the location-dispersion test. As will be seen in the simulation in the next section there exists a wide family of alternative hypotheses which provides larger powers to the location-dispersion test than the location score or dispersion score tests.

## 5. Monte Carlo Studies

Monte Carlo studies were conducted to assess the accuracy of the chi-squared approximation to the nominal test size, and to compare the power of the proposed test with several other chi-squared tests. Those chi-squared tests considered, except for the proposed test based on $Q$, are (i) Pearson's chi-squared test [6], (ii) Nair's location test [5], (iii) Nair's dispersion test [5], and (iv) The cumulative chi-squared test [7].

Table 2: The cell probabilities for $2 \times 5$ table

| | | | | |
|---|---|---|---|---|
| 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

To assess the accuracy of the chi-squared approximation to the nominal test size, we considered $2 \times 5$ table and used the cell probabilities given in Table 2. Generating random digits from the product-multinomial distribution with these cell probabilities, we run ten thousand trials. The sample size employed was $n_i = 10(10)100$. The nominal significance level was taken as 0.05. The results of the simulation are given in Figure 1. The inspection of the figure shows that the chi-squared approximation of the $Q$ test and the other tests provide the values that are quite close to the nominal level $\alpha = 0.05$ except for Pearson's test in small sample sizes.

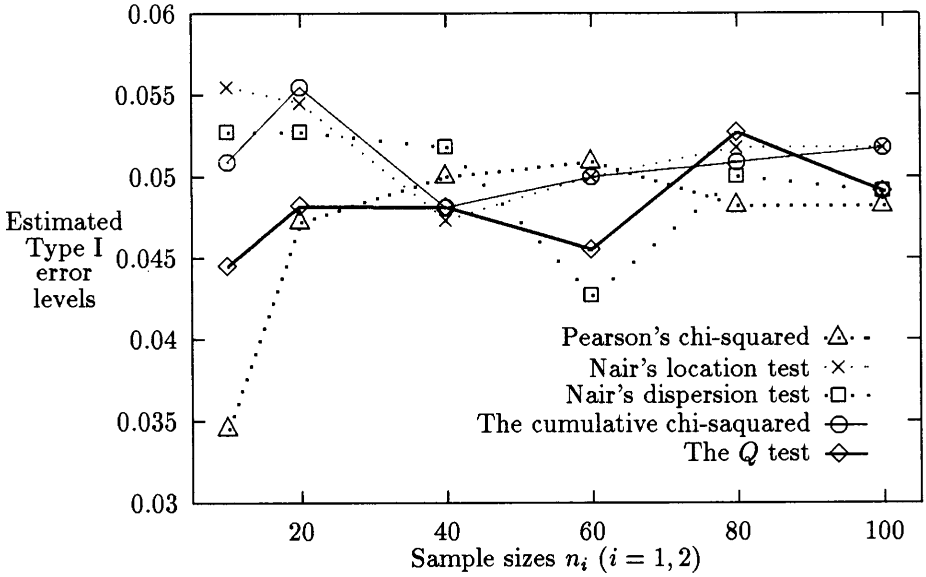


Figure 1: Estimated Type I error levels of the $Q$ test, Pearson's test, Nair's location and dispersion tests and the cumulative chi-squared test when the distributions in Table 2 are used for the population distributions.

Table 3: Uniform type vs. Convex type

| | | | | |
|---|---|---|---|---|
| 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| 0.2 | 0.3 | 0.25 | 0.15 | 0.1 |

Table 4: Uniform type vs. Concave type

| | | | | |
|---|---|---|---|---|
| 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| 0.25 | 0.1 | 0.15 | 0.2 | 0.3 |

To compare the powers of the $Q$ test with the other chi-squared tests, we conducted similar simulation using the cell probabilities given in Tables 3, 4 and 5. The cell probabilities in Tables 3 and 4 present the distributions of the uniform vs. convex type

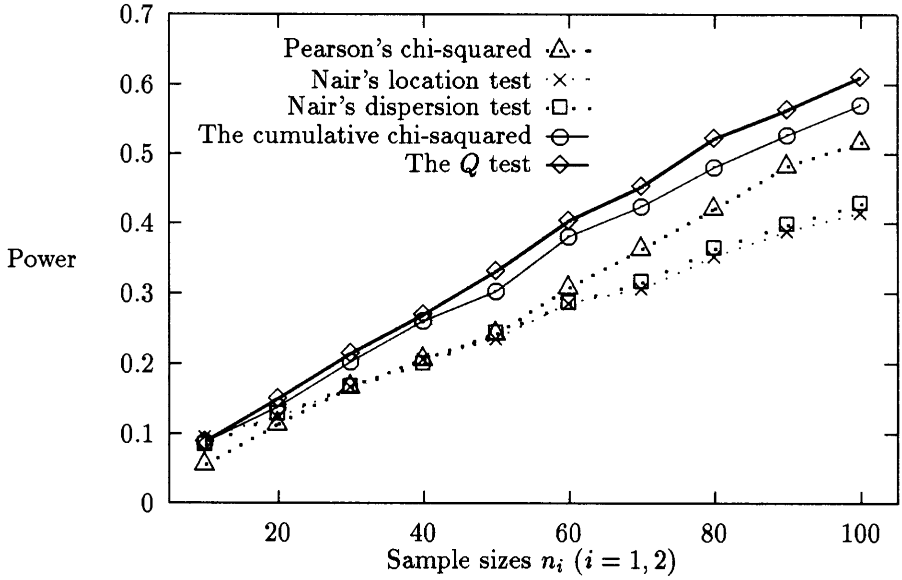


Figure 2: Estimated powers of the $Q$ test, Pearson's test, Nair's location and dispersion tests and the cumulative chi-squared test when the distributions in Table 3 are used for the population distributions.

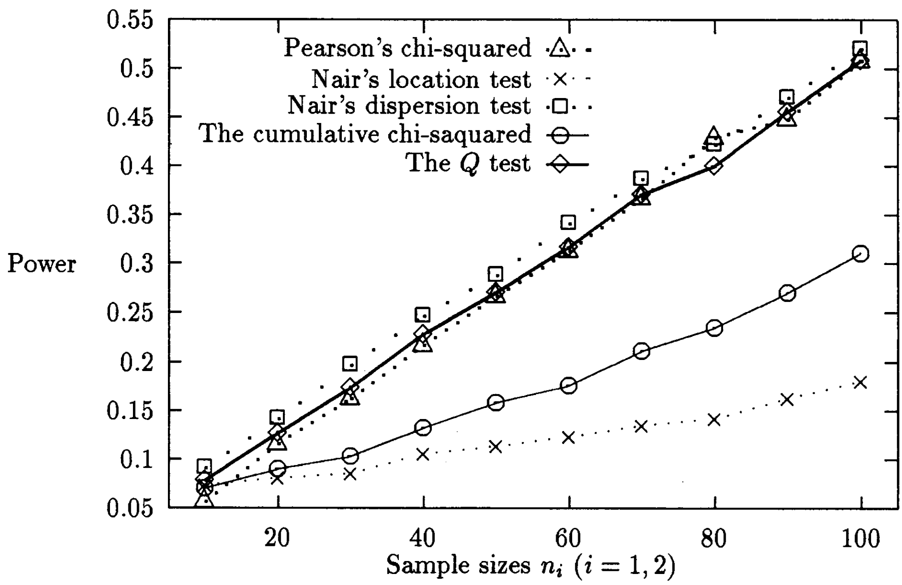


Figure 3: Estimated powers of the $Q$ test, Pearson's test, Nair's location and dispersion tests and the cumulative chi-squared test when the distributions in Table 4 are used for the population distributions.

and uniform vs. concave type, respectively, and also Table 5 presents the distributions of the convex vs. concave type. The powers assessed are plotted in Figures 2, 3 and 4. Figures 2 and 4 show that the $Q$ test is superior to Pearson's chi-squared test, Nair's location and dispersion tests and to the cumulative chi-squared test. Figure 3 shows that the $Q$ test is superior to Pearson's chi-squared test, Nair's location test and the cumulative chi-squared test, but it is inferior to the Nair's dispersion test.

Table 5: Convex type vs. Concave type

| | | | | |
|---|---|---|---|---|
| 0.2 | 0.3 | 0.25 | 0.15 | 0.1 |
| 0.25 | 0.1 | 0.15 | 0.2 | 0.3 |

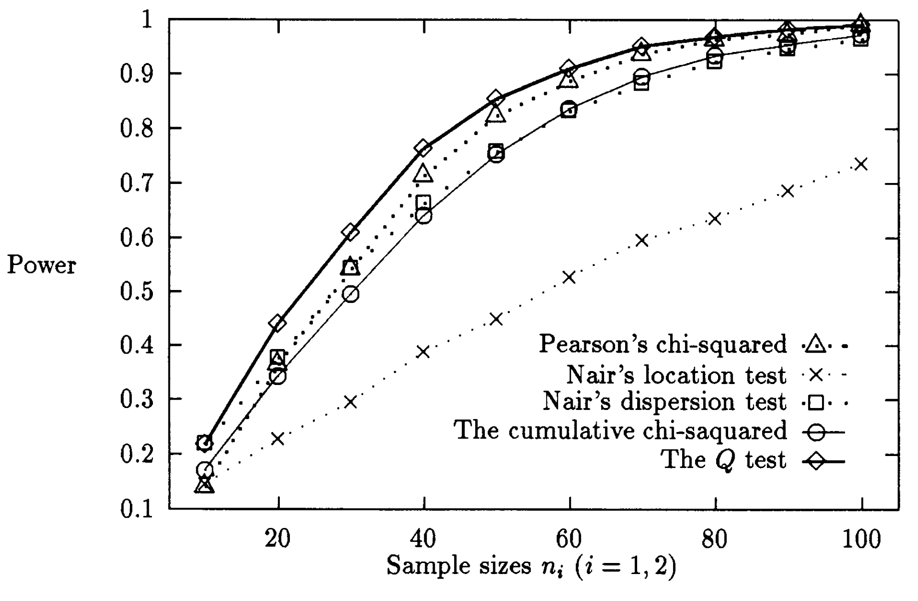


Figure 4: Estimated powers of the $Q$ test, Pearson's test, Nair's location and dispersion tests and the cumulative chi-squared test when the distributions in Table 5 are used for the population distributions.

We carried out many simulation studies using various cell probabilities other than those given in Tables 3, 4 and 5. General implication we obtained is that so long as testing the distributions of uniform vs. convex type, and convex vs. concave type are concerned, the $Q$ test is in many cases superior to the Pearson's test, Nair's location and dispersion tests, and to the cumulative chi-squared test. For testing the distributions of uniform vs. concave type the $Q$ test is also better in many cases than the other chi-squared tests.

## References

[1] DURAN, B. S., TSAI, W. S. and LEWIS, T. O.: *A class of location-scale nonparametric tests*, Biometrika, **63** (1976), 173-176.

[2] LEHMANN, E. L.: *NONPARAMETRICS: Statistical Methods Based on Ranks*, Holden Day, Inc., San Francisco, (1975).

[3] LEPAGE, Y.: *A combination of Wilcoxon's and Ansari-Bradley's statistics*, Biometrika,**58** (1971), 213-217.

[4] LEPAGE, Y.: *A class of nonparametric tests for location and scale parameters*, Commun. Statist.-Theor. Meth., **A6(7)** (1977), 649-659.

[5] NAIR, V. N.: *On Testing in Industrial Experiments with Ordered Categorical Data, (with discussion)*, Technometrics, **28** (1986), 283-311.

[6] PEARSON, K.: *On criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philos. Mag., series 5, **50** (1900) 157-175. (*Reprinted* 1948 *in Karl Pearson's Early statistical papers*, ed. by E. S. Pearson, Cambridge: Cambridge University Press.)

[7] TAKEUCHI, K. and HIROTSU, C.: *The Cumulative Chi-squares Method Against Ordered Alternatives in Two-way Contingency Tables*, Reports of Statistical Application Research, Japanese Union of Scientists and Engineers, **29** (1982), 1-13.

*Received November 25, 1993*
*Revised February 23, 1994*