⟨Research Paper⟩

Journal of the Korean
Statistical Society
Vol. 13, No. 1, 1984

Detection of Random Effects in a Random Effects Model of a One-way Layout Contingency Table

Byung Soo Kim*

ABSTRACT

A random effects model of a one-way layout contingency table is developed using a Dirichlet-multinomial distribution. A test statistic, say T_k, is suggested for detecting Dirichlet-multinomial departure from a multinomial distribution. It is shown that the T_k test is asymptotically superior to the classical chi-square test based on the asymptotic relative efficiency. This superiority is further evidenced by a Monte Carlo simulation.

1. Introduction

Using an analogy to fixed effects and random effects in linear models, it appears that almost all the methods for the analysis of multi-dimensional contingency tables have focused on fixed-effects models. Fienberg (1975) points this out and lists the development of a discrete analog to the nested and random effects (Model Ⅱ) ANOVA models among the unsclved problems in the analysis of multi-dimensional contingency tables.

We develop a random effects model for the one-way layout contingency table. Define

$$S_{\underset{\sim}{p}}^{0}=\{(p_1,\dots,p_{I-1});\ 0<p_i<1,\ \sum_{i=1}^{I-1}p_i<1\}, \tag{1.1}$$

$$S_{\underset{\sim}{x}}=\{(x_1,\dots,x_{I-1});\ 0\le x_i\le n,\ \sum_{i=1}^{I-1}x_i\le n\}. \tag{1.2}$$

For our development we need the following definitions.

Definition 1.1. A random vector $\underset{\sim}{X}'=(X_1,\dots,X_{I-1})$ has a multinomial distribution

* Department of Applied Statistics, Yonsei University, Seoul 120, Korea

with n and $\underset{\sim}{p}'=(p_1, \ldots, p_{I-1})$, denoted by $X \sim M(n, \underset{\sim}{p})$ if

$$\Pr(\underset{\sim}{X}=\underset{\sim}{x})=\binom{n}{x_1, \ldots, x_I}\prod_{i=1}^{I} p_i^{x_i} \tag{1.3}$$

for $\underset{\sim}{x}\in S_{\underset{\sim}{x}}$, $\underset{\sim}{p}\in S_{\underset{\sim}{p}}^0$, where $x_I=n-\sum_{i=1}^{I-1} x_i$ and $p_I=1-\sum_{i=1}^{I-1} p_i$.

Notationally, $m(\underset{\sim}{x}\ ;\ n,\ \underset{\sim}{p})$ and $M(\underset{\sim}{x}\ ;\ n,\ \underset{\sim}{p})$, denote the multinomial mass function (1.3) and corresponding distribution function, respectively.

Definition 1.2. A random vector $\underset{\sim}{U}'=(U_1, \ldots, U_{I-1})$ has a Dirichlet distribution with $\underset{\sim}{\beta}'=(\beta_1, \ldots, \beta_I)$ denoted by $\underset{\sim}{U}\sim D(\underset{\sim}{\beta})$ if it has a probability density function $(p.\ d.\ f)$ given by

$$f(\underset{\sim}{u})=\frac{\Gamma(B)}{\prod_{i=1}^{I}\Gamma(\beta_i)}\left(\prod_{i=1}^{I-1} u_i^{\beta_i-1}\right)\left(1-\sum_{i=1}^{I-1} u_i\right)^{\beta_I-1} \tag{1.4}$$

for $\underset{\sim}{u}\in S_{\underset{\sim}{u}}^0$, where $\beta_i>0$ for $i=1, \ldots, I$, $B=\sum_{i=1}^{I}\beta_i$ and $\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}\,dx$ for $t>0$. The Dirichlet distribution can be reparametrized so that it can be denoted by $D(\underset{\sim}{\pi}\ ;\ \theta)$, where $\pi_i=\beta_i/B$ and $\theta=1/B$ for $\underset{\sim}{\pi}'=(\pi_1, \ldots, \pi_{I-1})\in S_{\underset{\sim}{\pi}}^0$ A Dirichlet mixture of multinomial distributions is called the Dirichlet-multinomial distribution, and denoted by $DM(n,\ \pi,\ \theta)$. Following Johnson and Kotz (1969) $DM(n,\ \underset{\sim}{\pi},\ \theta)$ can be expressed as

$$DM(n,\ \underset{\sim}{\pi},\ \theta)=M(n,\ \underset{\sim}{p})\underset{\underset{\sim}{p}}{\wedge}D(\underset{\sim}{\pi},\ \theta). \tag{1.5}$$

Mosimann (1962) provides an extensive study of the Dirichlet-multinomial distribution, thereby extending Skellam's work on the beta-binomial distribution (Skellam, 1948). Brier (1980) investigates the effects of the Dirichlet-multinomial distribution on the chi-square test of a general hypothesis in the one-way layout contingency table and shows that Pearson's chi-square statistic is in fact asymptotically a constant multiple of a chi-square random variable when the hypothesis is true.

Thus it follows that in a contingency table with I response categories and G groups the Dirichlet-multinomial distribution $DM(n_{+j}\ ;\ \underset{\sim}{\pi},\ \theta)$ can introduce random group effects, since the j-th group probability vector, say $\underset{\sim}{p}_j$ is now randomly generated from $D(\underset{\sim}{\pi},\ \theta)$ and, conditional on the observed $\underset{\sim}{p}_j$, the j-th group response vector, say $\underset{\sim}{n}_j$, has a multinomial distribution $M(n_{+j},\ \underset{\sim}{p}_j)$, where n_{+j} is the j-th group size. In a handy notation this is described as

$$\begin{aligned}&\underset{\sim}{p}_j \overset{iid}{\sim} D(\underset{\sim}{\pi},\ \theta)\\&(\underset{\sim}{n}_j \mid n_{+j},\ \underset{\sim}{p}_j)\ \sim\ M(n_{+j},\ \underset{\sim}{p}_j)\end{aligned} \tag{1.6}$$

for $j=1, 2, \dots, G$.

An example of a random group effects model for discrete data may be obtained by modifying a worker-day example in Scheffé (1959, p. 221). Suppose that an experiment is performed in a factory with G workers and a machine run by a single worker, which produces a certain item, say a metal sheet. We shall assume that day-to-day variation is negligible during the experiment. We observe the degree of defectiveness in each metal sheet and classify it into one of the I response categories. Then n_{ij} represents the number of metal sheets produced by the j-th worker with the i-th degree of defectiveness.

The primary concern of this paper is hypothesis testing of the presence of random group effects, which can be formulated as

$$H_0 : \theta=0 \quad \text{vs.} \quad H_a : \theta>0. \tag{1.7}$$

For testing (1.7) we find that Neyman's $C(\alpha)$ procedure yields a new test statistic, denoted by T_k. The asymptotic relative effiency $e(X_p{}^2 | T_k)$ of the classical chi-square statistic satisfies

$$e(X_p{}^2 | T_k) \leq 1, \tag{1.8}$$

where the equality holds iff group sizes $\{n_{+j}\}_{j=1}^{G}$ are asymptotically balanced or $G=2$. The superiority of T_k test to chi-square test based on (1.8) is further evidenced by a Monte Carlo simulation that compares the actual performances of those two statistics in terms of their sizes and powers.

2. A Radom Effects Model of a One-way Layout

We consider a product of Dirichlet-multinomial distributions so as to implement random effects in a one-way layout contingency table. Experimentally this can arise from a situation in which we have G unordered experimental groups and I unordered response

Table 2.1 $I \times G$ Contingency Table

Response \ Group	1	2	…	G	Response Total
1	n_{11}	n_{12}	…	n_{1G}	n_{1+}
2	n_{21}	n_{22}	…	n_{2G}	n_{2+}
⋮	⋮	⋮		⋮	⋮
I	n_{I1}	n_{I2}	…	n_{IG}	n_{I+}
Group Total	n_{+1}	n_{+2}	…	n_{+G}	n_{++}

categories with n_{+j} observations taken in group j for $j=1, 2, \ldots, G$. Data from such sampling can be represented in the previous contingency table.

Let the j-th group response vector, given the group total be denoted by $\underset{\sim}{n}_j'=(n_{1j}, \ldots, n_{I-1j})$. One natural way of imposing random group effects on the j-th group response vector is to generalize the multinomial distribution by allowing the group probability vector itself to have a Dirichlet distribution. Thus we have

$$(\underset{\sim}{n}_j|\underset{\sim}{u}_j, n_{+j}) \overset{ind}{\sim} M(n_{+j}, \underset{\sim}{u}_j) \tag{2.1}$$

for $j=1, 2, \ldots, G$ and

$$\underset{\sim}{U}_1, \underset{\sim}{U}_2, \ldots, \underset{\sim}{U}_G \overset{iid}{\sim} D(\beta). \tag{2.2}$$

From (2.2) it can be easily shown that the means, variances, and covariances of $\underset{\sim}{U}_1'=\underset{\sim}{U}'=(U_1, \ldots, U_{I-1})$ are

$$\begin{aligned} &E(U_i)=\beta_i/B \\ &\mathrm{Var}(U_i)=\beta_i(B-\beta_i)/B^2(B+1) \\ &\mathrm{Cov}(U_i, U_j)=-\beta_i\,\beta_j/B^2(B+1), \quad i\neq j \end{aligned} \tag{2.3}$$

for $i, j=1, 2, \ldots, I-1$, where $B=\sum_{i=1}^{I}\beta_i$.

It is useful to change the parameters by putting

$$\begin{aligned} &p_i=\beta_i/B, \ i=1, \ldots, I-1 \\ &\theta=1/B. \end{aligned} \tag{2.4}$$

Then it is an easy exercise to show that the marginal distribution of $\underset{\sim}{n}_j$, which is a Dirichlet mixture of multinomial distributions, has a probability mass function ($p.\ m.\ f$)

$$h(\underset{\sim}{n}_j\,;\underset{\sim}{p}_j, \theta)=\binom{n_{+j}}{n_{1j}, \ldots, n_{Ij}}\left[\prod_{i=1}^{I}\prod_{r=0}^{n_{ij}-1}(p_i+r\theta)\right]\Big/\left[\prod_{r=1}^{n_{+j}-1}(1+r\theta)\right], \tag{2.5}$$

where $\quad n_{Ij}=n_{+j}-\sum_{i=1}^{I-1}n_{ij}$ for $j=1, \ldots, G$.

We refer to a Dirichlet mixture of multinomials (2.5) as a Dirichlet multinomial distribution and denote it by

$$\underset{\sim}{n}_j \sim DM(n_{+j}, \underset{\sim}{p}, \theta), \tag{2.6}$$

where $\underset{\sim}{p}'=(p_1, \ldots, p_{I-1})$, which is symbolically described as

$$\underset{\sim}{n}_j \sim M(n_{+j}, \underset{\sim}{U}_j)\underset{\underset{\sim}{U}_j}{\wedge}D(\underset{\sim}{p}, \theta).$$

Thus as an extension of a product multinomial distribution the joint distribution of $(\underset{\sim}{n}_1, \ldots, \underset{\sim}{n}_G)$ becomes a product Dirichlet-multinomial distribution, i.e.,

$$\underset{\sim}{n}_j \overset{ind}{\sim} DM(n_{+j}, \underset{\sim}{p}, \theta), \ j=1, \ldots, G. \tag{2.7}$$

3. Detection of the Random Effects

In the product Dirichlet-multinomial model (2.7) θ becomes the parameter of interest for testing the existence of random group effects, becaus if $\theta=0$ the model reduces to a product of multinomials; this is a device we and others have employed to allow a single parameter to introduce random effects. Thus the null hypothesis H_0 of no random effects and the alternative hypothesis H_a of the existence of the random effects can be expressed as

$$H_0 : \theta=0,$$
$$H_a : \theta>0. \tag{3.1}$$

Based on the one-way layout contingency table in Table (2.1) the loglikelihood function of θ, apart from the additive constant, is given by

$$\ell(\theta)=\sum_{j=1}^{G}\left\{\sum_{i=1}^{I}\sum_{r=0}^{n_{ij}-1}\log(p_i+r\theta)-\sum_{r=0}^{n_{+j}-1}\log(1+r\theta)\right\}, \tag{3.2}$$

where $\quad p_I=1-\sum_{i=1}^{I-1}p_i.$

3.1 Case of $\underset{\sim}{p}$ Known

It is easy to show that the uniformly most powerful (UMP) test for H_0 versus H_a does not exist in this case. However, the LMP test of Potthoff and Whittinghill (1966) rejects H_0 for large values of

$$\left.\frac{\partial \ell(\theta)}{\partial\theta}\right|_{\theta=0}=\frac{1}{2}\sum_{j=1}^{G}\left\{\sum_{i=1}^{I}\frac{n_{ij}(n_{ij}-1)}{p_i}-n_{+j}(n_{+j}-1)\right\} \tag{3.3}$$

$$\propto\sum_{j=1}^{G}\left\{\sum_{i=1}^{I}\frac{n_{ij}(n_{ij}-1)}{p_i}\right\}\equiv T_1,$$

where $\quad n_{Ij}=n_{+j}-\sum_{i=1}^{I-1}n_{ij}$ and $p_I=1-\sum_{i=1}^{I-1}p_i.$

Potthoff and Whittinghill(1966) proposed a method of moment approximation to the null distribution of T_1 by finding constants e, f and g that satisfied

$$eT_1+f \;\dot{\sim}\; \chi^2(g), \tag{3.4}$$

where $\chi^2(g)$ refers to a chi-square random variable with g degrees of freedom and '$\dot{\sim}$' is for 'approximately distributed as'. However, by expressing T_1 in (3.3) in terms of a quadratic form we can suggest another approximation of the null distribution of T_1. To aid in the development, we introduce some useful results without proofs. Proofs can be found in Ronning (1982).

Lemma 3.1. Under H_0 the convariance matrix of $\underset{\sim}{n}_j'=(n_{1j},\ldots,n_{I-1j})$ is given by

$$\begin{aligned}\mathrm{Cov}(\underset{\sim}{n}_j)&=n_{+j}[D_{p_i}-\underset{\sim}{p}\underset{\sim}{p}']\\&=n_{+j}V,\end{aligned}\tag{3.5}$$

where $D_{p_i}=\mathrm{diag}(p_1,\ldots,p_{I-1})$, and $V=D_{p_i}-\underset{\sim}{p}\underset{\sim}{p}'$.

Lemma 3.2. Let V and D_{p_i} be defined as in (3.5). Then

$$V^{-1}=D_{p_i}^{-1}+(1/p_I)E,\tag{3.6}$$

where E is an $(I-1)\times(I-1)$ matrix consisting of one's only.

Lemma 3.3. Let $\underset{\sim}{Z}_j$ be an $(I-1)\times 1$ vector with entries $z_{ij}=\sqrt{n_{+j}}\left(\frac{n_{ij}}{n_{+j}}-p_i\right)$ for $i=1,\ldots,I-1$, $j=1,\ldots,G$, then

$$\underset{\sim}{Z}_j'V^{-1}\underset{\sim}{Z}_j=\sum_{i=1}^{I}(n_{ij}-n_{+j}p_i)^2/n_{+j}p_i\tag{3.7}$$

is Pearson's chi-square statistic for goodness of fit in the j-th group. Hence $\underset{\sim}{Z}_j'V^{-1}\underset{\sim}{Z}_j$ has an asymptotic chi-square distribution with I degrees of freedom under H_0.

Simple calculation can show that the test based on T_1 in (3.3) is equivalent to the test based on

$$T_1^*=[\underset{\sim}{n}_j-n_{+j}\underset{\sim}{p}+\frac{I}{2}(\underset{\sim}{p}-\frac{1}{I}\underset{\sim}{1})'V^{-1}[\underset{\sim}{n}_j-\underset{\sim}{n}_{+j}\underset{\sim}{p}+\frac{I}{2}(\underset{\sim}{p}-\frac{1}{I}\underset{\sim}{1})],\tag{3.8}$$

where $\underset{\sim}{1}$ is a $(I-1)\times 1$ vector of one's only and $I\geq 2$.

Then by use of Lemma 3.3 we can derive the following results;

(1) When the n_{+j}'s are all equal and $\underset{\sim}{p}=(1/I)\underset{\sim}{1}$, T_1^* is equivalent to Pearson's chi-square statistic.

(2) If we assume that there exist α_j, $0<\alpha_j<1$ for $j=1,\ldots,G$ such that $\alpha_j=\lim_{n_{++}\to\infty}\frac{n_{+j}}{n_{++}}$, then the limiting distribution of $(1/n_{++})T_1^*$ is of the form

$$\frac{1}{n_{++}}T_1^*\ \underset{H_0}{\overset{D}{\longrightarrow}}\ \sum_{j=1}^{G}\alpha_j\chi_j^2(I-1,\delta),\tag{3.9}$$

where $\{\chi_j^2(I-1,\delta);\ j=1,\ldots,G\}$ is a set of independent noncentral chi-square random variables with $I-1$ degrees of freedom and noncentrality parameter $\delta=\frac{I^2}{4}\sum_{i=1}^{I-1}\left(p_i-\frac{1}{I}\right)^2$ and 'D' implies convergence in distribution.

(3) In the special case of equal n_{+j}'s we have

$$\frac{1}{n_{++}}T_1^*\ \underset{H_0}{\overset{D}{\longrightarrow}}\ \chi^2(G(I-1),\,G\delta),$$

where δ is defined in (3.9).

3.2 Case of $\underset{\sim}{p}$ Unknown

The case of unknown $\underset{\sim}{p}$ is far more interesting, especially in terms of applicability to real problems. It can be shown that a locally optimal test $H_0 : \theta=0$ versus $H_a : \theta>0$ does not exist. However, a $C(\alpha)$ test is readily available.

In order to derive the $C(\alpha)$ test statistic we need the following partial derivatives of the log-likelihood $\ell(\theta)$ in (3.2) evaluated at $\theta=0$.

$$\phi_1(\underset{\sim}{p})\equiv\frac{\partial\ell(\theta)}{\partial\theta}\bigg|_{\theta=0}=\sum_{j=1}^{G}\left\{\sum_{i=1}^{I}\frac{n_{ij}(n_{ij}-1)}{2p_i}-\frac{n_{+j}(n_{+j}-1)}{2}\right\} \tag{3.10}$$

$$\phi_{2i}(\underset{\sim}{p})\equiv\frac{\partial^2\ell(\theta)}{\partial p_i\,\partial\theta}\bigg|_{\theta=0}=\sum_{j=1}^{G}\left\{\frac{-n_{ij}(n_{ij}-1)}{2p_i^2}+\frac{(n_{+j}-\sum_{i=1}^{I-1}n_{ij}-1)\,(n_{+j}-\sum_{i=1}^{I-1}n_{ij}-1)}{2p_I^2}\right\} \tag{3.11}$$

for $i=1, 2, \ldots\ldots, I-1$.

$$\phi_3(\underset{\sim}{p})\equiv\frac{\partial^2\ell(\theta)}{\partial\theta^2}\bigg|_{\theta=0}=-\frac{1}{6}\sum_{j=1}^{G}\left\{\sum_{i=1}^{I}\frac{n_{ij}(n_{ij}-1)(2n_{ij}-1)}{p_i^2}-n_{+j}(n_{+j}-1)(2n_{+j}-1)\right\} \tag{3.12}$$

Under H_0 $n_{ij}\sim B(n_{+j}, p_i)$, hence

$$E_0[\phi_{2i}(\underset{\sim}{p})]=0 \text{ for } i=1, \ldots, I-1,$$

where E_0 implies that the expectation is taken under $\theta=0$.

Neyman (1959) (see also Moran, 1970) has shown that when $E_0[\phi_{2i}(\underset{\sim}{p})]=0$, the null hypothesis can be tested using the statistic $\phi_1(\tilde{\underset{\sim}{p}})$, where $\tilde{\underset{\sim}{p}}$ is a root-n_{++} consistent estimator of $\underset{\sim}{p}$. An obvious choice of $\tilde{\underset{\sim}{p}}$ is the *MLE* $\hat{\underset{\sim}{p}}=\frac{1}{n_{++}}\sum_{j=1}^{G}\underset{\sim}{n}_j$ under H_0. Substituting the *MLE* $\hat{\underset{\sim}{p}}$ in (3.10) we obtain

$$\begin{aligned}2\phi_1(\hat{\underset{\sim}{p}})&=\sum_{j=1}^{G}[\underset{\sim}{n}_j-n_{+j}\hat{\underset{\sim}{p}}]'\hat{V}^{-1}[\underset{\sim}{n}_j-n_{+j}\hat{\underset{\sim}{p}}]-(I-1)n_{++} \qquad (3.13)\\ &=n_{++}\sum_{j=1}^{G}\sum_{i=1}^{I}\frac{n_{+j}}{n_{i+}}\left[\frac{n_{ij}}{\sqrt{n_{+j}}}-\frac{n_{i+}}{n_{++}}\sqrt{n_{+j}}\right]^2-(I-1)n_{++},\end{aligned}$$

where $\hat{V}=D_{\hat{p}_i}-\hat{\underset{\sim}{p}}\hat{\underset{\sim}{p}}'$.

Hence we see that the $C(\alpha)$ test is based on

$$T_k=\sum_{j=1}^{G}\sum_{i=1}^{I}\frac{n_{+j}}{n_{i+}}\left(\frac{n_{ij}}{\sqrt{n_{+j}}}-\frac{n_{i+}}{n_{++}}\sqrt{n_{+j}}\right)^2. \tag{3.14}$$

In determining the approximate null distribution of T_k two limiting results are available. One uses the central limit theorem (*CLT*) on the *iid* multinomial random vectors as the sample size G tends to infinity. In this limiting argument, T_k, properly normalized, has an asymptotic $N(0, 1)$ distribution by the result of Neyman's $C(\alpha)$ procedure (Neyman, 1959).

Since $E_0[\phi_{2i}(\underset{\sim}{p})]=0$ for $i=1, 2, \ldots, I-1$, the variance of $\phi_1(\underset{\sim}{\hat{p}})$ is estimated by $-E_0[\phi_3(\underset{\sim}{\hat{p}})]$. From (3.12) it follows that

$$-E_0[\phi_3(\underset{\sim}{\hat{p}})]=\frac{1}{2}(I-1)\sum_{j=1}^{G} n_{+j}(n_{+j}-1). \tag{3.15}$$

Since $T_k=\frac{1}{n_{++}}2\phi_1(\underset{\sim}{\hat{p}})+(I-1)$, by normalizing T_k, we find that under $H_0: \theta=0$ the statistic

$$X_G^2=\frac{[T_k-(I-1)]^2}{\frac{2(I-1)}{n^2_{++}}\sum_{j=1}^{G} n_{+j}(n_{+j}-1)} \tag{3.16}$$

has an asymptotic chi-square distribution with 1 degree of freedom.

We may consider another limiting argument that uses the multivariate normal approximation of the multinomial distribution when the number of groups, G, is held fixed and the group sizes $\{n_{+j}\}_{j=1}^{G}$ tend to infinity in such a manner that $n_{+j}/n_{++}\longrightarrow\alpha_j$, $0<\alpha_j<1$, for $j=1, \ldots, G$. In the following discussion, the approximate null and alternative distributions are based on this limiting arguments, which may better reflect practical experimental considerations where the number of groups is fixed; we conjecture that these results will provide a better sampling approximation for finite sample sizes.

The hypothesis test $H_0: \theta=0$ versus $H_a: \theta>0$ has been described as detection of a Dirichlet-multinomial departure from the multinomial distribution. For this purpose Pearson's chi-square statistic X_p^2 that has been proposed for fixed effects problem is worthy of consideration:

$$X_p^2=\sum_{j=1}^{G}\sum_{i=1}^{I}\frac{n_{++}}{n_{i+}}\left[\frac{n_{ij}}{\sqrt{n_{+j}}}-\frac{n_{i+}}{n_{++}}\sqrt{n_{+j}}\right]^2. \tag{3.17}$$

For the relation between these two statistics T_k and X_p^2, we observe that when $n_{+j}=n$ for $j=1, \ldots, G$, the test based on T_k is identical to the test based on X_p^2. For the comparison of these two statistics in terms of large sample behavior, we obtain the asymptotic relative effiency (ARE) of X_p^2 relative to T_k.

4. Approximate Null and Alternative Distributions

4.1 Approximate Null Distributions

We define the following notations for $j=1, 2, \ldots, G$;

$$\underset{\sim}{Z}_j'=\underset{\sim}{Z}_j'(\underset{\sim}{p})=\frac{1}{\sqrt{n_{+j}}}(n_{1j}-n_{+j}p_1, \ldots, n_{I-1j}-n_{+j}p_{I-1}) \tag{4.1}$$

$$\hat{\underset{\sim}{p}}'=(\hat{p}_1, \ldots, \hat{p}_{I-1})=\left(\frac{n_{1+}}{n_{++}}, \ldots, \frac{n_{I-1+}}{n_{++}}\right) \tag{4.2}$$

$$\hat{\underset{\sim}{Z}}_j=\underset{\sim}{Z}_j(\hat{\underset{\sim}{p}}) \tag{4.3}$$

$$\underset{\sim}{Z}=(\underset{\sim}{Z}_1', \underset{\sim}{Z}_2', \ldots, \underset{\sim}{Z}_G')' \tag{4.4}$$

$$\hat{\underset{\sim}{Z}}=(\hat{\underset{\sim}{Z}}_1', \hat{\underset{\sim}{Z}}_2', \ldots, \hat{\underset{\sim}{Z}}_G')' \tag{4.5}$$

$$\sqrt{M}=\left(\sqrt{\frac{n_{+1}}{n_{++}}}, \sqrt{\frac{n_{+2}}{n_{++}}}, \ldots, \sqrt{\frac{n_{+G}}{n_{++}}}\right)' \tag{4.6}$$

$$M=\left(\frac{n_{+1}}{n_{++}}, \frac{n_{+2}}{n_{++}}, \ldots, \frac{n_{+G}}{n_{++}}\right)' \tag{4.7}$$

$$\alpha_j=\lim_{\substack{n_{+j}\to\infty \\ n_{++}\to\infty}} \frac{n_{+j}}{n_{++}} \tag{4.8}$$

$$\sqrt{A}=(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_G})' \tag{4.9}$$

$$A=(\alpha_1, \ldots, \alpha_G)' \tag{4.10}$$

We may express $\sqrt{A}=\lim_{n_{++}\to\infty}\sqrt{M}$ and $A=\lim_{n_{++}\to\infty} M$.

It is well known that as $n_{+j}\to\infty$

$$\underset{\sim}{Z}_j \xrightarrow[H_0]{D} N(\underset{\sim}{O}, V) \tag{4.11}$$

for $j=1, \ldots, G$, where

$$V=V(\underset{\sim}{p})=D_{p_i}-\underset{\sim}{p}\underset{\sim}{p}'.$$

Also we note for $j=1, \ldots, G$

$$\hat{\underset{\sim}{Z}}_j=\underset{\sim}{Z}_j-\sqrt{n_{+j}}(\hat{\underset{\sim}{p}}-\underset{\sim}{p})$$

$$=\underset{\sim}{Z}_j-\left(\frac{n_{+j}}{n_{++}}\right)^{1/2}\sum_{k=1}^{G}\left(\frac{n_{+k}}{n_{++}}\right)^{1/2}\underset{\sim}{Z}_k. \tag{4.12}$$

By using the above we can express $\hat{\underset{\sim}{Z}}$ in terms of $\hat{\underset{\sim}{Z}}$ as

$$\hat{\underset{\sim}{Z}}=[(I_G-\sqrt{M}\sqrt{M'})\otimes I_{I-1}]\underset{\sim}{Z}, \tag{4.13}$$

where I_k is a $k\times k$ identity matrix and $\otimes$ stands for the Kronecker product. The asymptotic distribution of $\underset{\sim}{Z}$ can be obtained by using (4.11) and the independence of $\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_G$:

$$\underset{\sim}{Z} \xrightarrow[H_0]{D} N(\underset{\sim}{O}, I_G\otimes V). \tag{4.14}$$

Hence by using (4.13) and the idempotency of $(I_G-\sqrt{A}\sqrt{A'})$ we obtain

$$\hat{\underset{\sim}{Z}} \xrightarrow[H_0]{D} N(\underset{\sim}{O}, (I_G - \sqrt{A}\sqrt{A}') \otimes V). \tag{4.15}$$

Now, Pearson's chi-square statistic X_p^2 can be expressed as

$$\begin{aligned} X_p^2 &= \sum_{j=1}^{G} \hat{Z}_j' \hat{V}^{-1} \hat{Z}_j \\ &= \hat{\underset{\sim}{Z}}'(I_G \otimes \hat{V}^{-1})\hat{\underset{\sim}{Z}}. \end{aligned} \tag{4.16}$$

For further discussion, the following Lemma is useful.

Lemma 4.1. Under H_0 $\hat{V} = V + o_p(1)$.

Proof. Using maximum absolute column sum $\|\cdot\|_1$ for the matrix norm we have

$$\|\hat{V} - V\|_1 = \max_{1 \le i \le I-1} \sum_{i=1}^{I-1} |\hat{p}_i \hat{p}_I - p_i p_I| = \sum_{i=1}^{I-1} |\hat{p}_i \hat{p}_I - p_i p_I|,$$

where $p_I = 1 - \sum_{i=1}^{I-1} p_i$ and $\hat{p}$ is accordingly defined. Since $n_{i+} \underset{H_0}{\sim} B(n_{++}, p_i)$, $\hat{p}_i = p_i + o_p(1)$ as $n_{++} \to \infty$ for $i = 1, \dots, I-1$. Thus by the continuity the result follows.

Thus by Lemma 4.1 (4.16) can be written as

$$X_p^2 = \hat{\underset{\sim}{Z}}'(I_G \otimes V^{-1})\hat{\underset{\sim}{Z}}(1 + o_p(1)). \tag{4.17}$$

By invoking a theorem in quadratic forms it can be seen that X_p^2 is asymptotically distributed as

$$X_p^2 \xrightarrow[H_0]{D} \sum_{i=1}^{(I-1)} \lambda_i^* \chi_i^2(1), \tag{4.18}$$

where $\{\lambda_i^*;\ i = 1, \dots, (I-1)G\}$ is the set of eigenvalues of

$$(I_G \otimes V^{-1})[(I_G - \sqrt{A}\sqrt{A}') \otimes V] = (I_G - \sqrt{A}\sqrt{A}') \otimes I_{I-1} \tag{4.19}$$

and $\{\chi_i^2(1);\ i = 1, \dots, (I-1)G\}$ are *iid* chi-square random variables with 1 degree of freendom. The eigenvalues of $(I_G - \sqrt{A}\sqrt{A}') \otimes I_{I-1}$ are cross products of eigenvalues of $(I_G - \sqrt{A}\sqrt{A}')$ and those of I_{I-1}. Since I_{I-1} has an eigenvalue 1 with multiplicity $I-1$, (4.18) is equivalent to

$$X_p^2 \xrightarrow[H_0]{D} \sum_{i=1}^{G} \rho_i \chi_i^2(I-1), \tag{4.20}$$

where ρ_i's are eigenvalues of $(I_G - \sqrt{A}\sqrt{A}')$.

Since $I_G - \sqrt{A}\sqrt{A}'$ is idempotent and of rank $G-1$ we have $(G-1)$ one's and one zero for its eigenvalues. Thus (4.20) becomes

$$X_p^2 \xrightarrow[H_0]{D} \chi^2((I-1)(G-1)), \tag{4.21}$$

a well known result.

We now consider the null distribution of the $C(\alpha)$ statistic T_k, which can be expressed as

$$T_k=(n_{++})^{-1}\sum_{j=1}^{G} n_{+j}\,\hat{\underset{\sim}{Z}}_j'\,\hat{V}^{-1}\,\hat{\underset{\sim}{Z}}_j$$

$$=\sum_{j=1}^{G}\left(\frac{n_{+j}}{n_{++}}\right)^{1/2}\hat{\underset{\sim}{Z}}_j'\,\hat{V}^{-1}\left(\frac{n_{+j}}{n_{++}}\right)^{1/2}\hat{\underset{\sim}{Z}}_j. \quad (4.22)$$

For notational convenience we define

$$\hat{\underset{\sim}{Z}}_j{}^*=\left(\frac{n_{+j}}{n_{++}}\right)^{1/2}\hat{\underset{\sim}{Z}}_j \quad (4.23)$$

and

$$\hat{\underset{\sim}{Z}}{}^*=(\hat{\underset{\sim}{Z}}_1{}^{*\prime},\ldots,\hat{\underset{\sim}{Z}}_G{}^{*\prime}) \quad (4.24)$$

By the same arguments for obtaining the distribution of $\hat{\underset{\sim}{Z}}$ in (4.13) we obtain:

$$\hat{\underset{\sim}{Z}}{}^* \xrightarrow[H_0]{D} N(\underset{\sim}{O},(D_{\alpha_i}-AA')\otimes V),$$

where

$$D_{\alpha_i}=\mathrm{diag}(\alpha_1,\ldots,\alpha_G).$$

Now, using $\hat{\underset{\sim}{Z}}{}^*$ and Lemma 4.1, we can express T_k as

$$T_k=\hat{\underset{\sim}{Z}}{}^{*\prime}(I_G\otimes V^{-1})\hat{\underset{\sim}{Z}}{}^*(1+o_p(1)). \quad (4.26)$$

Thus, using the same arguments employed in (4.18)—(4.20) the asymptotic distribution of T_k under H_0 is obtained as

$$T_k \xrightarrow[H_0]{D} \sum_{j=1}^{G}\lambda_j\chi_j^2(I-1), \quad (4.27)$$

where $\{\lambda_j\,;\,j=1,\ldots,G\}$ is the set of eigenvalues of $(D_{\alpha_i}-AA')$.
We may note here that $n(D_{\alpha_i}-AA')$ is the singular covariance matrix of a multinomial distribution $M(n,\underset{\sim}{A})$.

Even though some computer subroutines can readily provide the eigenvalues of $(D_{\alpha_i}-AA')$, the determination of the eigenvalues appears to be an algebraically unsolved problem except that one of the eigenvalues is known to be zero. (Roy et al, 1960, Light and Margolin, 1971, and Ronning, 1982). Since the α_i's are known, however, we may approximate the distribution of T_k by $g\chi^2(h)$, where the constants g and h are chosen so that $g\chi^2(h)$ has the same first two moments as those of T_k. In doing this we use following results on $D_{\alpha_i}-AA'$;

$$\mathrm{trace}(D_{\alpha_i}-AA')=\lambda_1+\ldots+\lambda_{G-1}=1-\sum_{j=1}^{G}\alpha_j^2 \quad (4.28)$$

$$\mathrm{trace}(D_{\alpha_i}-AA')^2=\lambda_1^2++\lambda_{G-1}^2=\sum_{j=1}^{G}\alpha_j^2-2\sum_{j=1}^{G}\alpha_j^3+(\sum_{j=1}^{G}\alpha_j^2)^2. \quad (4.29)$$

Thus the asymptotic distribution of T_k can be approximated as

$$g^{-1}T_k \underset{H_0}{\sim} \chi^2(h),$$

where

$$g^{-1}=\frac{1-\sum_{j=1}^{G}\alpha_j^2}{\sum_{j=1}^{G}\alpha_j^2-2\sum_{j=1}^{G}\alpha_j^3+(\sum_{j=1}^{G}\alpha_j^2)^2}$$

and

$$h=\frac{(I-1)(1-\sum_{j=1}^{G}\alpha_j^2)^2}{\sum_{j=1}^{G}\alpha_j^2-2\sum_{j=1}^{G}\alpha_j^3+(\sum_{j=1}^{G}\alpha_j^2)^2}.$$

4.2 Approximate Alternative Distributions

We next derive the asymptotic distribution of X_p^2 and T_k under H_a. Here we use the remarkable resemblance of the mean and covariance matrix of the Dirichlet-multinomial to those of the multinomial distribution (Mosimann, 1962);

$$E_\theta(\underset{\sim}{n}_j)=n_{+j}\underset{\sim}{p},\ \ j=1,\dots,G \tag{4.30}$$

$$\mathrm{Cov}_\theta(\underset{\sim}{n}_j)=\left(\frac{n_{+j}\theta+1}{\theta+1}\right)n_{+j}V,\ \ j=1,\dots,G, \tag{4.31}$$

where the subscript θ indicates that the underlying distribution is the Dirichletmultinomial.

It has been observed that there are four different asymptotic forms of the Dirichlet-multinomial distribution (Paul and Plackett, 1978). Among them, one is of particular relevance to our development.

Theorem 4.1 (Paul and Plackett, 1978).
Let

$$\underset{\sim}{n}_j \sim M(n_{+j},\ \underset{\sim}{u})\underset{\underset{\sim}{u}}{\wedge}D(\underset{\sim}{\beta})=M(n_{+j},\ \underset{\sim}{u})\underset{\underset{\sim}{u}}{\wedge}D(\underset{\sim}{p},\ \theta), \tag{4.32}$$

where

$$\underset{\sim}{\beta}=(\beta_1,\dots,\beta_I)',\ \ \underset{\sim}{p}=(p_1,\dots,p_{I-1})'$$

and the p_i's and θ are defined in (2.4).
Write $\beta_i=n_{++}\phi_i$ for all i, where ϕ_i's are fixed quantities and let $n_{++}\to\infty$. Then

$$n_{+j}^{-\frac{1}{2}}(\underset{\sim}{n}_j-n_{+j}\underset{\sim}{p})\ \xrightarrow[H_a]{D}\ N(\underset{\sim}{O},\ \gamma_j(\theta)V), \tag{4.33}$$

where

$$\gamma_j(\theta)=\lim_{n_{++}\to\infty}(n_{+j}\theta+1).$$

We may note that by the construction of $\beta_i=n_{++}\phi_i$ for all i we assume that

$$\theta=\left(\sum_{i=1}^{I}\beta_i\right)^{-1}=\theta_{n_{++}}=O(1/n_{++}).$$

Hence using this result of Paul and Plackett, it is easy to see that

$$\underset{\sim}{Z} \xrightarrow[H_a]{D} N(\underset{\sim}{O},\ \gamma_i(\theta) V), \tag{4.34}$$

and

$$\underset{\sim}{Z} \xrightarrow[H_a]{D} N(\underset{\sim}{O},\ D_{\gamma_i} \otimes V), \tag{4.35}$$

where

$$D_{\gamma_i} = \text{diag}(\gamma_1(\theta), \ldots\ldots, \gamma_G(\theta)).$$

Thus by using (4.13) and (4.35) we obtain

$$\hat{\underset{\sim}{Z}} \xrightarrow[H_a]{D} N(\underset{\sim}{O},\ Q \otimes V), \tag{4.36}$$

where

$$Q = D_{\gamma_i} - \sqrt{A}\sqrt{A}' D_{\gamma_i} - D_{\gamma_i}\sqrt{A}\sqrt{A}' + \sqrt{A}\sqrt{A}' D_{\gamma_i}\sqrt{A}\sqrt{A}'$$

i.e.,

$$Q = \begin{pmatrix} \gamma_1 - 2\alpha_1\gamma_1 + \alpha_1\sum_{k=1}^{G}\alpha_k\gamma_k & & & & \\ & \gamma_2 - 2\alpha_2\gamma_2 + \alpha_2\sum_{k=1}^{G}\alpha_k\gamma_k & & & \\ & \cdot & & & \\ & & \cdot & & \text{sym.} \\ -\sqrt{\alpha_i\alpha_j}(\gamma_i + \gamma_j - \sum_{k=1}^{G}\alpha_k\gamma_k) & & & \cdot & \\ & & & & \gamma_G - 2\alpha_G\gamma_G + \alpha_G\sum_{k=1}^{G}\alpha_k\gamma_k \end{pmatrix}$$

Now it becomes straightforward to show that

$$X_p^2 \xrightarrow[H_a]{D} \textstyle\sum_{i=1}^{G} \delta_i \chi_i^2(I-1), \tag{4.37}$$

where $\{\delta_i : i=1, \ldots, G\}$ is a set of eigenvalues of Q, and

$$T_k \xrightarrow[H_a]{D} \sum_{i=1}^{G} \delta_i^* \chi_i^2(I-1), \tag{4.38}$$

where $\{\delta_i^* : i=1, \ldots, G\}$ is a set of eigenvalues of $D_{\sqrt{\alpha_i}} Q D_{\sqrt{\alpha_i}}$.

4.3 *ARE* of X^2 Relative to T_k

To summarize the relevant distribution results, we have derived the following:

(a) $X_p^2 \xrightarrow[H_0]{D} \chi^2[(I-1)(G-1)]$

(b) $T_k \xrightarrow[H_0]{D} \sum_{i=1}^{G} \lambda_i \chi_i^2(I-1)$,

where λ_i's are eigenvalues of $D_{\alpha_i} - AA'$.

$$\text{(c)}\quad X_p^2 \xrightarrow[H_a]{D} \sum_{i=1}^{G} \delta_i \chi_i^2(I-1),$$

where δ_i's are eigenvalues of Q.

$$\text{(d)}\quad T_k \xrightarrow[H_a]{D} \sum_{i=1}^{G} \delta_i^* \chi_i^2(I-1),$$

where δ_i^*'s are eigenvalues of $D_{\sqrt{\alpha_i}} Q D_{\sqrt{\alpha_i}}$.

Thus it can be shown that

$$\mathrm{Var}(X_p^2 \mid H_0) \longrightarrow 2(I-1)(G-1) \tag{4.39}$$

$$\mathrm{Var}(T_k \mid H_0) \longrightarrow 2(I-1)\sum_{i=1}^{G} \lambda_i^2 = 2(I-1)\ \mathrm{trace}\ (D_{\alpha_i} - AA')^2$$

$$= 2(I-1)\Big[\sum_{j=1}^{G} \alpha_j^2 - \sum_{j=1}^{G} \alpha_j^3 + \Big(\sum_{j=1}^{G} \alpha_j^2\Big)^2\Big] \tag{4.40}$$

$$\frac{d}{d\theta} E_\theta[X_p^2 \mid H_a]\Big|_{\theta=0} = (I-1)\Big(1 - \sum_{j=1}^{G} \alpha_j^2\Big) = (I-1)\ \mathrm{trace}\ (D_{\alpha_i} - AA') \tag{4.41}$$

$$\frac{d}{d\theta} E_\theta[T_k \mid H_a]\Big|_{\theta=0} = (I-1)\Big[\sum_{j=1}^{G} \alpha_j^2 - 2\sum_{j=1}^{G} \alpha_j^3 + \Big(\sum_{j=1}^{G} \alpha_j^2\Big)^2\Big]$$

$$= (I-1)\ \mathrm{trace}(D_{\alpha_i} - AA')^2. \tag{4.42}$$

Hence the asymptotic relative efficiency (ARE) of the chi-square statistic X_p^2 relative to the $C(\alpha)$ statistic T_k is given by

$$e_{\mathrm{PIC}} = \frac{\Big(1 - \sum_{j=1}^{G} \alpha_j^2\Big)^2}{(G-1)\big[\sum \alpha_j^2 - 2\sum \alpha_j^3 + (\sum \alpha_j^2)^2\big]}$$

$$= \frac{(\lambda_1 + \lambda_2 + \dots + \lambda_{G-1})^2}{(G-1)(\lambda_1^2 + \dots + \lambda_{G-1}^2)}, \tag{4.43}$$

where under H_a $\theta = \theta_{n++} = 0(1/n_{++})$.

Interestingly Collings and Margolin (1983) obtained the same expression of an ARE as (4.43) when they compared a $C(\alpha)$ test with another test for detecting a negative binomial departure from a Poisson in the regression through the origin case. They proved the following:

Theorem 4.2 (Collings and Margolin, 1983).

$$\frac{1}{G-1} \le e_{\mathrm{PIC}} \le 1,$$

where the left equality holds if and only if $G=2$ and the right equality holds if and only if the group sizes $\{n_{+j}\}_{j=1}^{G}$ are asymptotically balanced.

Using the expression of ARE e_{PIC} in (4.43) we can prove

Lemma 4.2. The $C(\alpha)$ test is asymptotically equivalent to Pearson's chi-square test if and only if $G=2$ or all the group sizes $\{n_{+j}\}_{j=1}^{G}$ are asymptotically balanced.

Proof. We may express e_{PIC} as

$$e_{PIC}=\bar{\lambda}^2/(S_\lambda^2+\bar{\lambda}^2),$$

where

$$\bar{\lambda}=(G-1)^{-1}\sum_{i=1}^{G-1}\lambda_i \text{ and } S_\lambda^2=(G-1)\sum_{i=1}^{G-1}(\lambda_i-\bar{\lambda})^2.$$

Thus $e_{PIC}=1$ if and only if $S_\lambda^2=0$. But $S_\lambda^2=0$ if and only if $G=2$ or $\lambda_1=\ldots=\lambda_{G-1}$, which is equivalent to $\alpha_1=\ldots=\alpha_G$(Light and Margolin, 1971, and Ronning, 1982).

4.4 Monte Carlo Simulation: Power Comparison

As shown above, the test based on T_k is superior to Pearson's chi-square test based on considerations of asymptotic relative efficiency; however, the large sample properties do not necessarily hold for small samples, nor are the local properties of the asymptotic relative efficiency readily transferable to practical situations. Therefore, a Monte Carlo simulation was conducted to compare the performance of the two tests in terms of their sizes and powers.

The data for the Monte Carlo simulation were generated on the *VAX* 780 computer system at the National Institute of Environmental Health Sciences. The program was written in Fortan and used two *IMSL* subroutines: *GGAMR* and *GGMTN*.

The following Lemma is useful to generate random observations from a Dirichlet distribution, say $D(\underset{\sim}{p}, \theta)$.

Lemma 4.3 (Wilks, 1962). Let $X_1, X_2, \ldots, X_{k+1}$ be independent variables having gamma distributions $G(1, \beta_1)$, $G(1, \beta_2), \ldots, G(1, \beta_{k+1})$.

Define

$$Y_i=X_i/(\sum_{j=1}^{k+1}X_j)$$

for $i=1, \ldots, k$.

Then $\underset{\sim}{Y}=(Y_1, \ldots, Y_k)$ has a Dirichlet distribution $D(\underset{\sim}{\beta})$, where $\underset{\sim}{\beta}=(\beta_1, \ldots, \beta_{k+1})$, and $D(\underset{\sim}{\beta})$ is defined in (1.4).

The Dirichlet distribution $D(\underset{\sim}{\beta})$ can be reparametrized as $D(\underset{\sim}{p}, \theta)$ by (2.4). The Fortran program of the Monte Carlo simulation is outlined as follows;

(i) Set $\underset{\sim}{p}=\underset{\sim}{p}_0$, and $\theta=\theta_0$, and initialize $\{n_{+j}\}_{j=1}^{5}$ and the upper bound(upbound) of θ.

(ii) Generate a set of 5 independent probability vectors $\underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_5$ from a Dirichlet distribution $D(\underset{\sim}{p}, \theta)$ using *IMSL* subroutine *GGAMR* and Lemma 4.3.

(iii) Generate a contingency table from a product multinomial distributions

$\prod_{j=1}^{G} \underset{\sim}{M}(n_{+j}, \underset{\sim}{u}_j)$ using *IMSL* subroutine *GGMTN*.

(iv) Calculate T_k and X_p^2.

(v) Count the number of T_k and X_p^2 values exceeding their cut off values corresponding to $\alpha=0.05$.

(vi) Go to the step (ii) and repeat for 2,000 times.

(vii) Set $\theta=\theta_0+\Delta$ and go to the step (ii) until $\theta \geq$ upbound.

For the calculation of sizes of T_k test and the X_p^2 test a subset consisting of (iii)–

Table 4.1 Two Sets of Input Values of the Program

	First Set	Second Set
$\underset{\sim}{P}_0$	(0.05, 0.1, 0.4, 0.45)	(0.1, 0.15, 0.3, 0.45)
θ_0	0.001	0.001
Δ	0.002	0.003
Upbound	0.031	0.025
Group sizes	20, 20, 20, 200, 400	Same

Table 4.2 Approximate Power of T_k and X_p^2 for 0.05 Size, $\underset{\sim}{p}_0=(.05, .1, .40, .45)$ and $\{n_{+j}\}_{j=1}^{5}=\{20, 20, 20, 200, 400\}$

	Approximate Power		
θ	T_k	X_p^2	Difference
0.000	0.0525	0.0505	0.0020
0.001	0.1060	0.0885	0.0175
0.003	0.2445	0.1700	0.0745
0.005	0.3545	0.2685	0.0860
0.007	0.4535	0.3680	0.0855
0.009	0.5255	0.4470	0.0785
0.011	0.5860	0.5175	0.0685
0.013	0.6385	0.5855	0.0530
0.015	0.7030	0.6480	0.0550
0.017	0.7165	0.6645	0.0520
0.019	0.7555	0.7270	0.0285
0.021	0.7805	0.7600	0.0205
0.023	0.7950	0.7925	0.0025
0.025	0.8110	0.8100	0.0010
0.027	0.8250	0.8175	0.0075
0.029	0.8465	0.8410	0.0055
0.031	0.8640	0.8610	0.0030

(vi) of the above program was employed, because putting $\theta=0$ in the step (i) involved division by zero in the step (iii).

The actual program was run for two sets of $\underset{\sim}{p}_0$ and θ ranges with the same group sizes, which are listed in Table 4. 1.

The asymptotic relative efficiency of X_p^2 to T_k is 0. 415 for these group sizes. Tables 4. 2 and 4. 3, respectively, display approximate power functions of T_k and X_p^2 tests for an 0. 05 level based on the first and the second sets of input values. Over the ranges of θ values considered the difference in powers can be as large as 0. 086 for the first set of input values and 0. 115 for the second set. The ratio of the power of the T_k test to that of the X_p^2 test falls as low as 0. 76 in both cases considered. Clearly, the T_k test can perforn better than the X_p^2 test.

Table 4.3 Approximate Power of T_k and $X_p{}^2$ for 0.05 size, $\underset{\sim}{p}_0=(.1,\ .15,\ .3,.45)$ and $\{n_{+j}\}^5_{j=1}=\{20,\ 20,\ 20,\ 200,\ 400\}$

	Approximate Power		
θ	T_k	$X_p{}^2$	Difference
0. 000	0. 0535	0. 0480	0. 0055
0. 001	0. 1050	0. 0780	0. 0270
0. 004	0. 2900	0. 2095	0. 0805
0. 007	0. 4790	0. 3640	0. 1150
0. 010	0. 5825	0. 4885	0. 0940
0. 013	0. 6460	0. 5840	0. 0620
0. 016	0. 7265	0. 6865	0. 0400
0. 019	0. 7640	0. 7390	0. 0250
0. 022	0. 7870	0. 7815	0. 0055
0. 025	0. 8440	0. 8335	0. 0005

5. Summary

This paper develops a random effects model of a one-way layout contingency table using Dirichlet-multinomial distributions $\{DM(n_{+j}\,;\underset{\sim}{p},\theta)\}^c_{j=1}$, and suggests a new test statistic, say T_k by applying Neyman's $C(\alpha)$ procedure for detecting random effects when $\underset{\sim}{p}$ is unknown. The T_k statistic is basically a weighted average of the Pearson's chi-square statistics. Comparison is made between the T_k test and the chi-square test in terms of the Pitman asymptotic relative efficiency. It turns out that the T_k test is in

general asymptotically superior to the chi-square test. This superiority is further evidenced by a Monte Carlo simulation which compares the performance of these two tests in terms of their sizes and powers.

6. Acknowledgements

This is a part of the author's Ph. D dissertation at the University of North Carolina at Chapel Hill. The author would like to thank Dr. B. H. Margolin for his guidance. He is also indebted to Dr. D. Hoel of the National Institute of Environmental Health Sciences for providing computer facilities.

REFERENCES

(1) Brier, Stephen S. (1980). Analysis of Contingency Table under Cluster Sampling, *Biometrika,* Vol. 67, 591—596.

(2) Collings, Bruce J. and Margolin, Barry H. (1983). Testing of Fit for the Poisson Assumption When Observations are not Identically Distributed. Submitted for *J. Amer. Statist. Assoc.*

(3) Fienberg, Stephen E. (1975). Comment on "The Observational Study—a Review" by Sonya Mckinlay, *J. Amer. Statist. Assoc.,* Vol. 70, 521—523.

(4) Johnson, Norman L. and Kotz, Samuel (1969). *Discrete Distributions,* Wiley, New York.

(5) Johnson, Norman L. and Kotz, Samuel (1977). *Urn Models and Their Application,* Wiley, New York.

(6) Light, Richard J. and Margolin, Barry H. (1971). An Analysis of Variance for Categorical Data, *J. Amer. Statist. Assoc.,* Vol. 66, 534—544.

(7) Mosimann, James E. (1962). On the Compound Multinomial Distributon, the Multivariate β-Distribution, and Correlations among Proportions, *Biometrika,* Vol. 49, 65—82.

(8) Moran, P. A. P. (1970). On Asymptotially Optimal Test of Composite Hypotheses, *Biometrika,* Vol. 57, 47—55.

(9) Neyman, Jerzy (1959). Optimal Asymptotic Tests of Composite Hypotheses, *Probability and Statistics: The Herald Cramer Volumn,* ed. Ulf Grenander, Wiley, New York.

(10) Paul, S.R. and Placktet, R.L. (1978). Inference Sensitivity for Poisson Mixtures. *Biometrika,* Vol. 65, 591—602.

(11) Potthoff, Richard F. and Whittinghill, Maurice (1969). Testing for Homogeneity I: The Binomial and Multinomial Distributions, *Biometrika,* Vol. 53, 167—182.

(12) Ronning, Gerd. (1982). Characteristic Values and Triangular Factorization of the Covariance Matrix for Multinomial, Dirichlet and Multivariate Hypergeometric Distributions and Some Related Results, *Statist. Hefte,* Vol. 23, 152—176.

(13) Roy, S.N., Greenberg, B.G. and Sarhan, A.E. (1960). Evaluation of Determinants,

Characteristic Equations, and Their Roots for a Class of Matrices, *J. Roy. Statist. Soc. Ser. B.*, Vol. 22, 349—359.

(14) Scheffé, Henry. (1959). *The Analysis of Variance*, Wiley, New York.

(15) Skellam, J.G. (1948). A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable between the Sets of Trials, *J. Roy. Statist. Soc. Ser. B.*, Vol. 10, 257—261.

(16) Wilks, Samuel S. (1962). *Mathematical Statistics*, Wiley, New York.