CHAPTER V

CLOSER APPROXIMATION TO THE MOMENTS AND THE DISTRIBUTION OF THE
CHI-SQUARE TYPE STATISTICS

5.1 Introduction

In Chapter I we have explained the fact that all members of the chi-square type statistics are asymptotically equivalent to the Pearson chi-square statistic and asymptotically follow a central chi-square distribution under the null hypothesis and a non-central chi-square distribution under the contiguous alternative hypothesis. In this chapter we obtain more accurate approximations to the first and second moments of the chi-square type statistics. For this purpose we consider the simple hypothesis \( H : \pi = \pi_0 \) and obtain corrections for the chi-square type statistics \( D_n(\rho, \pi_0) \). The objective of such an approximation is to check the adequacy of the chi-square approximation to the asymptotic distribution of \( D_n(\rho, \pi_0) \) and possibly to improve the approximation. The need for such an investigation has been emphasized in the literature on the chi-square tests [Bartlett, 1938, 1947; Cox and Hinkley, 1974; Cressie and Read, 1984]. In section 2 we derive expressions for first and second moments of the chi-square type statistics to \( O(\frac{1}{n}) \). We discuss the adequacy of chi-square approximation and necessary corrections to the statistics in section 3. Small sample comparisons of exact levels of tests are also worked out in this section. In section 4 we have obtained the first two moments of the statistics under contiguous alternative hypothesis. In section 5 we have discussed the asymptotic distribution under the condition that the number of cells is increased
at a faster rate than the sample size. Appendix at the end of the chapter provides the necessary details of the results obtained in section 2 and 4.

5.2 Moments Under the Simple Null Hypothesis

The chi-squared type statistics under the simple null hypothesis \( \pi = \pi_0 \) is

\[
D_n(\pi, \pi_0) = n \sum_{\pi_0} \frac{[h(p_1) - h(\pi_0)]^2}{[h'(\pi_0)]^2} 
\]  

(5.2.1)

Using the Taylor expansion we have approximately

\[
D_n(\pi, \pi_0) = \sum \frac{w_i^2}{\pi_0} + \frac{1}{n} \sum \frac{h''(\pi_0)}{h'(\pi_0)} \frac{w_i^3}{\pi_0} + \frac{1}{n} \left[ \frac{1}{4} \sum \left( \frac{h''(\pi_0)}{h'(\pi_0)} \right)^2 \frac{w_i^4}{\pi_0} + \frac{1}{3} \sum \frac{h''(\pi_0)}{h'(\pi_0)} \frac{w_i^5}{\pi_0} \right] + O(n^{-3/2}) 
\]  

(5.2.2)

where \( w_i = \sqrt{n}(p_i - \pi_0) \), \( i = 1, ..., k \). To compute the expected value and the second moment of \( D_n(\pi, \pi_0) \), we need expected values of the powers of \( w_i \) and \( w_j \). The necessary derivations are given in Appendix. Using the first three moments we get

\[
E[D_n(\pi, \pi_0)] = \sum (1 - \pi_0) \left[ \frac{1}{4} \sum \frac{h''(\pi_0)}{h'(\pi_0)} (1 - \pi_0)(1 - 2\pi_0) + \frac{3}{4} \sum \left( \frac{h''(\pi_0)}{h'(\pi_0)} \right)^2 \pi_0(1 - \pi_0)^2 + \sum \frac{h''(\pi_0)}{h'(\pi_0)} \pi_0(1 - \pi_0)^2 \right] + O(n^{-3/2}) 
\]  

(5.2.3)
\[
= (k-1) + \frac{1}{n} \left[ \sum_{i} \frac{h''(\pi_{01})}{h'(\pi_{01})} (1 - \pi_{01})(1 - 2\pi_{01}) \right.
\]
\[
+ \frac{3}{4} \sum_{i} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^2 \pi_{01}(1 - \pi_{01})^2 + \sum_{i} \frac{h''(\pi_{01})}{h'(\pi_{01})} \pi_{01}(1 - \pi_{01})^2 \right]
\]
\[
+ O(n^{-3/2}).
\]

Squaring \( D_n^2(\rho, \pi_0) \) and retaining terms of \( O(\frac{1}{n^2}) \) we get

\[
D_n^2(\rho, \pi_0) = \sum_{i} \frac{w_i^4}{\pi_{01}} + \sum_{i \neq j} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right) \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right) \frac{w_i^3}{\pi_{01}} \frac{w_j^2}{\pi_{01}} \frac{1}{n} \left[ \sum_{i} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^2 \frac{w_i^6}{\pi_{01}^2} \right]
\]
\[
+ \sum_{i \neq j} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right) \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right) \frac{w_i^3}{\pi_{01}} \frac{w_j^3}{\pi_{01}} \frac{1}{2} \sum_{i} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^2 \frac{w_i^6}{\pi_{01}^2}
\]
\[
+ \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^2 \frac{w_i^4}{\pi_{01}} \frac{w_j^2}{\pi_{01}} \frac{2}{3} \sum_{i} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^3 \frac{w_i^6}{\pi_{01}^2}
\]
\[
+ \frac{2}{3} \sum_{i \neq j} \left( \frac{h''(\pi_{01})}{h'(\pi_{01})} \right)^2 \frac{w_i^4}{\pi_{01}} \frac{w_j^2}{\pi_{01}} \right] + O(n^{-3/2}). \quad (5.2.5)
\]

The general expression for the \( E[D_n^2(\rho, \pi_0)] \) can be obtained by substituting the value of the moments (Appendix). We now consider the particular case of \( D_n(\rho, \pi_0) \) when \( h(\pi_i) = \pi_i^n \), \( \eta \in \mathbb{R} \) which is given by
\[ D_n(n) (p, \pi_0) = n \sum \frac{(p^\eta - \pi_0^\eta)^2}{\pi^\eta_0 2^{n-1}}. \] (5.2.6)

We shall denote \( D_n(n) (p, \pi_0) \) by \( D_n(n) \) for convenience. The expressions in (5.2.4) and (5.2.5), after simplification, reduce to

\[ E[D_n(n)] = (k-1) + \frac{1}{n} [(n-1)(S - 3k + 2) + \frac{3}{4} (n-1)^2(S-2k+1) + (n-1)(n-2)(S - 3k + 2)] + O(n^{-3/2}), \] (5.2.7)

and

\[ E[D_n(n)]^2 = k^2 - 1 + \frac{1}{n} [S - k^2 - 2k+2 + 2(n-1)(10 - 13k - 6k^2 + S(k+8)) + 3/2(n-1)^2(S(k+13) + 11 - 17k - 8k^2) + 2(n-1)(n-2)(S(k+3) + 3-5k - 2k^2)] + O(n^{-3/2}), \] (5.2.8)

where

\[ S = \sum \frac{1}{\pi_0}. \]

The above expressions give first and second moments of \( D_n(n) \) for different values of \( n \). Table 5.2.1 gives these expressions for some selected values of \( n \). Note that expressions for the Pearson chi-square statistic are well known [Kendall and Stuart, 1973]. Cressie and Read (1984) have obtained expressions for the first two moments of power divergence statistic. Freeman-Tukey chi-square statistic is a member of this family. Freeman-Tukey chi-square statistic is \( D_n(1/2) \) and the expression given in Table 5.2.1 agrees with the expressions given by Cressie and Read (1984).
Table 5.2.1  Second order moments of members of $D_n^{(s)}(p, \pi_o)$ under the null hypothesis

| Value of $n$ | Name of the test statistic | $E[ D_n^{(s)}(p, \pi_o)| H_o ]$ | $E[ ( D_n^{(s)}(p, \pi_o) )^2 | H_o ]$ |
|--------------|----------------------------|---------------------------------|---------------------------------|
| $-2 \times$  |                            | $\sqrt{k-1} + \frac{1}{n} \frac{[63 S - 57k + 51]}{4}$ | $k^2 - 1 + \frac{1}{2n} \frac{[(75k + 401)S - 242k^2 - 607k + 229]}{2}$ |
| $-1 \times$  |                            | $x_k - 1 + \frac{1}{n} \frac{[7S - 12k + 5]}{2}$ | $k^2 - 1 + \frac{1}{n} \frac{[(14k + 83)S - 49k^2 - 112k + 64]}{2}$ |
| $0 \times$   |                            | $k - 1 + \frac{1}{4n} \frac{[7S - 10k + 3]}{2}$ | $k^2 - 1 + \frac{1}{n} \frac{[(7k + 33)S - 18k^2 - 43k + 21]}{2}$ |
| $1/2 \times$ | Freeman-Tukey              | $1 + \frac{1}{16n} \frac{[7S - 6k - 1]}{2}$ | $k^2 - 1 + \frac{1}{8n} \frac{[(7k + 19)S - k^2 - 23k + 5]}{2}$ |
| $2/3 \times$ |                            | $k - 1 + \frac{1}{36n} \frac{[-5S + 34k - 29]}{2}$ | $k^2 - 1 + \frac{1}{18n} \frac{[(7k + 9)S - 2k^2 - 33k - 3]}{2}$ |
| $1 \checkmark$ | Pearson chi-square          | $k - 1$                            | $k^2 - 1 + \frac{1}{n} \frac{[S - k^2 - 2k + 2]}{2}$ |
| $2 \checkmark$ |                            | $k - 1 + \frac{1}{4n} \frac{[7S - 18k + 11]}{2}$ | $k^2 - 1 + \frac{1}{2n} \frac{[(7k + 73)S - 50k^2 - 107k + 77]}{2}$ |

* Defined by continuity.
From the Table 5.2.1 it is clear that the first term of the expressions correspond to the first two moments of a chi-square random variable with k-1 df. Also, the first moment of the Pearson chi-square statistic coincides with the first moment of the chi-square random variable and the second moment differs least as compared to other members of D_n^{(n)}. To find out whether there is any other member of D_n^{(n)} which has the same performance as the Pearson chi-square statistic, we have solved the quadratic equation in \eta for which terms of O(\frac{1}{n}) vanish, as done in Cressie and Read (1984). The computations are tabulated in Table 5.2.2. It is clear that for the first moment \eta = 1 is a root irrespective of the value of S. The other root converges to 1 as k increases. Even for the second moment, both the roots converge to 1 as k increases. The case S = k^2 is interesting as it corresponds to the case of symmetric null hypothesis \pi_i = 1/k, i = 1, ..., k. In Table 5.2.3 we have computed the difference between the cumulative distribution of a chi-square random variable with k-1 df. and the exact distribution of D_n^{(n)} for the particular situation in which number of classes is 3 and the sample size is 9. The empirical results also show that the Pearson chi-square statistic can be better approximated by the chi-square random variable as compared to other members of D_n^{(n)}. The figure 5.1 shows exact cumulative distributions of D_n^{(n)} for selected values of \eta.
Table 5.2.2: Entries give the two roots $\eta_1$, $\eta_2$ where the second order correction factors of the first two moments $M_1 = E[(D_n(\eta, P, \pi_0))^i \mid H_0]$ $i = 1, 2$ are zero.

<table>
<thead>
<tr>
<th>$S = k^2$</th>
<th>$M_1$</th>
<th>$M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.29</td>
<td>1.14</td>
</tr>
<tr>
<td>$S = 5k^2$</td>
<td>$M_1$</td>
<td>$M_2$</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>1.03</td>
<td>1.04</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>1.60</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Note - A dash indicates that no roots exist.
Table 5.2.3: Difference between the cumulative distribution of a chi-square random variable with k-1 df and cumulative distribution of some members of $D_n(p, l/k)$ at selected values of the variable when $k = 3$ and $N = 9$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\eta = 1/2$</th>
<th>$\eta = 2/3$</th>
<th>$\eta = 1$</th>
<th>$\eta = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.1474</td>
<td>0.0475</td>
<td>0.1626</td>
<td>0.0666</td>
</tr>
<tr>
<td>4</td>
<td>0.0304</td>
<td>0.0305</td>
<td>0.0304</td>
<td>-0.0080</td>
</tr>
<tr>
<td>6</td>
<td>0.0391</td>
<td>0.0392</td>
<td>0.0007</td>
<td>0.0775</td>
</tr>
<tr>
<td>8</td>
<td>0.0595</td>
<td>0.0597</td>
<td>0.0066</td>
<td>0.0066</td>
</tr>
<tr>
<td>10</td>
<td>0.0711</td>
<td>0.0073</td>
<td>-0.0038</td>
<td>0.0182</td>
</tr>
<tr>
<td>12</td>
<td>0.0753</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.0224</td>
</tr>
<tr>
<td>14</td>
<td>0.0385</td>
<td>0.0021</td>
<td>0.0017</td>
<td>0.0240</td>
</tr>
</tbody>
</table>

Note - $F_D = \text{Cumulative distribution of } D_n(p, l/k)$

$F_{\chi^2} = \text{Cumulative distribution of chi-square random variable with } k-1 \text{ d.f.}$
FIG 51: EXACT CUMULATIVE DISTRIBUTION OF $D_n(\frac{1}{k})$ FOR SELECTED VALUE OF $\eta$, $k=3$ $n=9$
5.3 Closer Approximations to the Asymptotic Distributions of $D_n(\eta)$

Our analysis in section 2 shows that for moderate sample size $D_n(\eta)$ cannot be approximated by a chi-square distribution. Corrections have been suggested to improve the approximation. The Bartlett's correction, based on the first moment, is obtained by multiplying $D_n(\eta)$ by

$$\left[1 + \frac{q(n)}{n}\right]^{-1}$$

where $q(n)$ corresponds to the coefficient of $\frac{1}{n}$ in the expected value of $D_n(\eta)$ and from (5.2.7) it is given by

$$q(n) = \frac{1}{(k-1)} \left[ (n-1)(S-3k+2) + \frac{3}{4}(n-1)^2(S-2k+1)\right]$$

$$+ (n-1)(n-2)(S-3k+2))$$

(5.3.1)

The corrected statistic is

$$D_n(\eta)^* = \frac{D_n(\eta)}{1 + \frac{q(n)}{n}}.$$  (5.3.2)

The expected value of $D_n(\eta)^*$ is $k-1$ to $O(\frac{1}{n^2})$. Such corrections have been suggested by Cox and Hinkley (1974), Williams (1976); and Smith, Rae, Manderscheid and Silbergeld (1981). Lawley (1956) has shown that in the case of the likelihood ratio statistic all moments of the corrected statistic agree with the moments of the approximating chi-square distribution to $O(\frac{1}{n^2})$. For the score test, Cox and Hinkley (1974) observed that different correction factors may be required for the moments of different order. Harris (1984) proved that there does not exist a single multiplicative correction factor for the simultaneous removal of the $O(\frac{1}{n})$ terms for the score test (chi-square test). We feel that such a result is true for the family $D_n(\eta)$ also.
The second correction for \( D_n^{(n)} \) is based on the first two moments [Cressie and Read, 1984]. From (5.2.7) and (5.2.8) we have,

\[
E[ D_n^{(n)} ] = (k - 1) + \frac{a_n}{n} + O(\frac{1}{n^2}) ,
\]

\[
E[ D_n^{(n)} ]^2 = (k^2 - 1) + \frac{b_n}{n} + O(\frac{1}{n^2}) ,
\]

and

\[
\text{Var}[ D_n^{(n)} ] = 2(k-1) + \frac{c_n}{n} + O(\frac{1}{n^2}) ,
\]

where

\[
a_n = (n - 1)(S - 3k + 2) + \frac{3}{4}(n - 1)^2(S - 2k + 1) + (n - 1)(n - 2)(S - 3k + 2),
\]

\[
b_n = S - k^2 - 2k + 2 + 2(n - 1)(10 - 13k - 6k^2 + S(k + 8)) + 3/2(n - 1)^2(S(k+13) + 11 - 17k - 8k^2) + 2(n - 1)(n - 2)(S(k + 3) + 3 - 5k - 2k^2),
\]

and

\[
c_n = b_n - 2(k-1)a_n.
\]

Let \( d_n = 1 + c_n / 2n(k-1) \) and \( \ell_n = (k - 1)(1 - \sqrt{d_n}) + \frac{a_n}{n} \).

We consider the connected statistic

\[
D_n^{(n)*} = \frac{D_n^{(n)} - \ell_n}{\sqrt{d_n}} .
\]

The statistic \( D_n^{(n)*} \) has mean \( k-1 \) and variance \( 2(k-1) \) to \( O(\frac{1}{n^2}) \). As the correction is based on the first two moments \( D_n^{(n)*} \) is expected to provide better approximation to the distribution.

There is a third method of approximating the distribution by matching the first two moments and the range to the corresponding parameters of a beta distribution. Smith, Rae, Manderscheid and Silbergeld
(1981) used this approach to approximate the distribution of the likelihood ratio statistic. The approximation can be obtained for the family of $D_n(\eta)$, for $\eta > 0$; but we consider the approximation $D_n^{(\eta)^*}$ and $D_n^{(\eta)**}$ for our empirical calculations.

Finite sample comparisons

Although it is interesting to compare the exact distribution of $D_n(\eta)$ in the entire range of the variate, in practical situations of testing, we are concerned with only the level of the test. In this empirical study, we restrict ourselves to the finite sample comparisons of the attained level of the test for different members of $D_n(\eta)$. We consider the symmetric null hypothesis $\pi_i = 1/k$, $i = 1, \ldots, k$. There have been various studies published, indicating that equiprobable class intervals produce the most sensitive tests [Cohen and Sackrowitz, 1975; Spruill, 1977]. Also $D_n(\eta)$ is invariant to permutation in the observed frequencies under the equiprobable case; this greatly reduces our computations for calculating the test size of the exact distribution of $D_n(\eta)$. The Table 5.3.1 indicates the different values of sample size ($n$), number of classes ($k$) and level of significance ($\alpha$).

Table 5.3.1: Cases considered

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 6, 9, 12, 15</td>
<td>3</td>
</tr>
<tr>
<td>4, 8, 12, 16, 20</td>
<td>4</td>
</tr>
<tr>
<td>5, 10</td>
<td>5</td>
</tr>
<tr>
<td>6, 12</td>
<td>6</td>
</tr>
<tr>
<td>$n = -2, -1, -0.5, 0, 0.5, 2/3, 1, 2$</td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.10$ and $0.01$</td>
<td></td>
</tr>
</tbody>
</table>
The values of n and k selected are similar to those of Chapman (1976). The nominal levels 0.1 and 0.01 have been considered by Larntz in an empirical study. For any given n, k, it is necessary to enumerate all possible \( \binom{n+k-1}{n} \) partitions of cell frequencies into k parts. With such frequencies \((n_1, \ldots, n_k)\), there is an associated value of the statistic

\[
D_n(n) (p_{i/k}) = \frac{\sum [(n p_i)^{\eta} - (n/k)^{\eta}]^2}{\eta^2 (n/k)^{2\eta-1}},
\]

and a multinomial probability

\[
P(n_1, \ldots, n_k) = \frac{n!}{k \prod_{i=1}^{k} n_i!}
\]

These values are computed for different values of \( \eta, k \) and \( n \) along with \( a_\eta, b_\eta, c_\eta, d_\eta \) and \( e_\eta \). For assessing the accuracy of the approximations in evaluating the size of the test, we have considered the chi-squared cut-off point for the significance level assumed. The Table 5.3.2 to 5.3.5 summarize our computations. Figures 5.2 and 5.3 illustrate the errors incurred by the two approximations for \( n = 16 \), \( k = 4 \) and \( \alpha = 0.10 \) and 0.01.

Based on our computations we make the following observations:

1. For \( \eta < 0 \), the attained level of significance is considerably larger than the nominal level of significance. The two statistics \( D_n(n)^* \) and \( D_n(n)^{**} \) improve the approximations but still the attained level of significance is higher than the nominal significance level. This is due to the fact that for \( \eta < 0 \), for all partitions with zero observations in one or more cells, \( D_n(n) \) is infinite, and the exact critical region
Table 5.3.2: True and approximate significance levels under symmetric hypothesis at nominal level $\alpha = 0.10$, $k = 3$ and different values of $n$; for various values of $n$

<table>
<thead>
<tr>
<th>value of $n$</th>
<th>$n = 3$</th>
<th>$n = 6$</th>
<th>$n = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
</tr>
<tr>
<td>-2</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
</tr>
<tr>
<td>-1</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
</tr>
<tr>
<td>0</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7778</td>
<td>0.1111</td>
<td>0.1111</td>
</tr>
<tr>
<td>2/3</td>
<td>0.1111</td>
<td>0.1111</td>
<td>0.1111</td>
</tr>
<tr>
<td>1</td>
<td>0.1111</td>
<td>0.1111</td>
<td>0.1111</td>
</tr>
<tr>
<td>2</td>
<td>0.1111</td>
<td>0.1111</td>
<td>0.1111</td>
</tr>
</tbody>
</table>

appr.1 = approximation 1

appr.2 = approximation 2

see text for details.
<table>
<thead>
<tr>
<th>Value of n</th>
<th>n = 12</th>
<th></th>
<th>n = 15</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
</tr>
<tr>
<td>-2</td>
<td>0.5175</td>
<td>0.1777</td>
<td>0.1777</td>
<td>0.3507</td>
<td>0.2356</td>
<td>0.2356</td>
<td>0.5874</td>
</tr>
<tr>
<td>-1</td>
<td>0.2672</td>
<td>0.1610</td>
<td>0.1610</td>
<td>0.2566</td>
<td>0.2356</td>
<td>0.2356</td>
<td>0.3952</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.1777</td>
<td>0.1610</td>
<td>0.1610</td>
<td>0.2566</td>
<td>0.1603</td>
<td>0.1038</td>
<td>0.3773</td>
</tr>
<tr>
<td>0</td>
<td>0.1777</td>
<td>0.1610</td>
<td>0.1610</td>
<td>0.2356</td>
<td>0.0917</td>
<td>0.0603</td>
<td>0.3232</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1777</td>
<td>0.1151</td>
<td>0.1151</td>
<td>0.1038</td>
<td>0.1038</td>
<td>0.1038</td>
<td>0.1356</td>
</tr>
<tr>
<td>2/3</td>
<td>0.1151</td>
<td>0.1151</td>
<td>0.1151</td>
<td>0.1038</td>
<td>0.1038</td>
<td>0.1038</td>
<td>0.1356</td>
</tr>
<tr>
<td>1</td>
<td>0.0704</td>
<td>0.0704</td>
<td>0.1151</td>
<td>0.1247</td>
<td>0.1247</td>
<td>0.1247</td>
<td>0.1051</td>
</tr>
<tr>
<td>2</td>
<td>0.1986</td>
<td>0.0645</td>
<td>0.0555</td>
<td>0.0951</td>
<td>0.0951</td>
<td>0.0925</td>
<td>0.1665</td>
</tr>
</tbody>
</table>
Table 5.3.3: True and approximate significance levels under symmetric hypothesis at nominal level $\alpha = 0.01$, $k = 3$ and different values of $n$; for various values of $\eta$

<table>
<thead>
<tr>
<th>Value of $\eta$</th>
<th>$n = 3$</th>
<th></th>
<th></th>
<th>$n = 6$</th>
<th></th>
<th></th>
<th>$n = 9$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
</tr>
<tr>
<td>-2</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.3827</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.4153</td>
<td>0.4153</td>
<td>0.4153</td>
</tr>
<tr>
<td>-1</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.4153</td>
<td>0.0779</td>
<td>0.0779</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.0889</td>
<td>0.0779</td>
<td>0.0779</td>
</tr>
<tr>
<td>0</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.7778</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.2593</td>
<td>0.0889</td>
<td>0.0779</td>
<td>0.0779</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1111</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1770</td>
<td>0.0635</td>
<td>0.0041</td>
<td>0.0779</td>
<td>0.0779</td>
<td>0.0779</td>
</tr>
<tr>
<td>$2/3$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0041</td>
<td>0.0041</td>
<td>0.0041</td>
<td>0.0139</td>
<td>0.0139</td>
<td>0.0139</td>
</tr>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0041</td>
<td>0.0041</td>
<td>0.0041</td>
<td>0.0029</td>
<td>0.0029</td>
<td>0.0029</td>
</tr>
<tr>
<td>2</td>
<td>0.1111</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0535</td>
<td>0.0535</td>
<td>0.0041</td>
<td>0.0248</td>
<td>0.0248</td>
<td>0.0248</td>
</tr>
</tbody>
</table>

appr.1 = approximation 1
appr.2 = approximation 2

See text for details.
<table>
<thead>
<tr>
<th>Value of n</th>
<th>n = 12</th>
<th>n = 15</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
</tr>
<tr>
<td>-2</td>
<td>0.6175</td>
<td>0.1610</td>
<td>0.1610</td>
</tr>
<tr>
<td>-1</td>
<td>0.1610</td>
<td>0.1610</td>
<td>0.1610</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.1610</td>
<td>0.1610</td>
<td>0.0313</td>
</tr>
<tr>
<td>0</td>
<td>0.0537</td>
<td>0.0239</td>
<td>0.0239</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0313</td>
<td>0.0239</td>
<td>0.0239</td>
</tr>
<tr>
<td>2/3</td>
<td>0.0239</td>
<td>0.0239</td>
<td>0.0239</td>
</tr>
<tr>
<td>1</td>
<td>0.0108</td>
<td>0.0108</td>
<td>0.0108</td>
</tr>
<tr>
<td>2</td>
<td>0.0555</td>
<td>0.0108</td>
<td>0.0108</td>
</tr>
</tbody>
</table>
Table 5.3.4: True and approximate significance levels under symmetric hypothesis at nominal level $\alpha = 0.10$, $k = 4$ and different values of $n$; for various values of $\eta$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.6539</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.5666</td>
<td>0.5666</td>
<td>0.5666</td>
</tr>
<tr>
<td>-1</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.3976</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.5666</td>
<td>0.1729</td>
<td>0.1729</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.5666</td>
<td>0.1729</td>
<td>0.1412</td>
</tr>
<tr>
<td>0</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.9063</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.3771</td>
<td>0.1899</td>
<td>0.1299</td>
<td>0.1299</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3438</td>
<td>0.3438</td>
<td>0.2031</td>
<td>0.3771</td>
<td>0.2746</td>
<td>0.1977</td>
<td>0.1899</td>
<td>0.1412</td>
<td>0.1412</td>
</tr>
<tr>
<td>$\frac{2}{3}$</td>
<td>0.2031</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.1977</td>
<td>0.0951</td>
<td>0.0951</td>
<td>0.1780</td>
<td>0.1412</td>
<td>0.1412</td>
</tr>
<tr>
<td>1</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0951</td>
<td>0.0951</td>
<td>0.1156</td>
<td>0.0851</td>
<td>0.0851</td>
<td>0.0851</td>
</tr>
<tr>
<td>2</td>
<td>0.2031</td>
<td>0.2031</td>
<td>0.2031</td>
<td>0.1156</td>
<td>0.1092</td>
<td>0.0336</td>
<td>0.1381</td>
<td>0.0654</td>
<td>0.0654</td>
</tr>
<tr>
<td>value of n</td>
<td>( n = 16 )</td>
<td>( n = 20 )</td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
<td>True</td>
<td>appr. 1</td>
<td>appr. 2</td>
</tr>
<tr>
<td>-2</td>
<td>0.6761</td>
<td>0.2613</td>
<td>0.2475</td>
<td>0.4149</td>
<td>0.3486</td>
<td>0.3486</td>
<td>0.6436</td>
<td>0.4920</td>
<td>0.4892</td>
</tr>
<tr>
<td>-1</td>
<td>0.3076</td>
<td>0.2475</td>
<td>0.2468</td>
<td>0.3497</td>
<td>0.1676</td>
<td>0.1225</td>
<td>0.5056</td>
<td>0.3743</td>
<td>0.3651</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.3076</td>
<td>0.2475</td>
<td>0.2468</td>
<td>0.3497</td>
<td>0.1124</td>
<td>0.1078</td>
<td>0.5015</td>
<td>0.3632</td>
<td>0.3558</td>
</tr>
<tr>
<td>0</td>
<td>0.1221</td>
<td>0.0477</td>
<td>0.0421</td>
<td>0.1686</td>
<td>0.1124</td>
<td>0.1078</td>
<td>0.3528</td>
<td>0.3147</td>
<td>0.3126</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1154</td>
<td>0.1154</td>
<td>0.1154</td>
<td>0.1299</td>
<td>0.1236</td>
<td>0.1236</td>
<td>0.2312</td>
<td>0.1997</td>
<td>0.1562</td>
</tr>
<tr>
<td>2/3</td>
<td>0.1154</td>
<td>0.1155</td>
<td>0.1155</td>
<td>0.1198</td>
<td>0.1110</td>
<td>0.1110</td>
<td>0.1628</td>
<td>0.0957</td>
<td>0.0957</td>
</tr>
<tr>
<td>1</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.0934</td>
<td>0.0934</td>
<td>0.0934</td>
<td>0.0789</td>
<td>0.0789</td>
<td>0.0830</td>
</tr>
<tr>
<td>2</td>
<td>0.1323</td>
<td>0.1323</td>
<td>0.1108</td>
<td>0.1689</td>
<td>0.0997</td>
<td>0.0678</td>
<td>0.1516</td>
<td>0.1219</td>
<td>0.0961</td>
</tr>
</tbody>
</table>
Table 5.3.5: True and approximate significance levels under symmetric hypothesis at nominal level $\alpha = 0.01$, $k = 4$ and different values of $n$; for various values of $n$.

<table>
<thead>
<tr>
<th>Value of $n$</th>
<th>True $r_H$</th>
<th>Approx. $r_H$</th>
<th>True $p_H$</th>
<th>Approx. $p_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 4$</td>
<td>0.9063</td>
<td>0.0000</td>
<td>0.3771</td>
<td>0.3771</td>
</tr>
<tr>
<td>$n = 8$</td>
<td>0.9063</td>
<td>0.0000</td>
<td>0.3771</td>
<td>0.3771</td>
</tr>
<tr>
<td>$n = 12$</td>
<td>0.9063</td>
<td>0.0000</td>
<td>0.3771</td>
<td>0.3771</td>
</tr>
<tr>
<td>$n = 16$</td>
<td>0.9063</td>
<td>0.0000</td>
<td>0.3771</td>
<td>0.3771</td>
</tr>
</tbody>
</table>

-2 0.9063 0.0000 0.3771 0.3771
-1 0.9063 0.0000 0.3771 0.3771
0 0.9063 0.0000 0.3771 0.3771
0.5 0.9063 0.0000 0.3771 0.3771
1 0.9063 0.0000 0.3771 0.3771
2/3 0.9063 0.0000 0.3771 0.3771
2 0.9063 0.0000 0.3771 0.3771
Table 5.3.5  Continued

<table>
<thead>
<tr>
<th>Value of n</th>
<th>True</th>
<th>$n = 16$ appr. 1</th>
<th>appr. 2</th>
<th>$n = 20$ appr. 1</th>
<th>appr. 2</th>
<th>True</th>
<th>appr. 1</th>
<th>appr. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>0.3076</td>
<td>0.2717</td>
<td>0.2468</td>
<td>0.3574</td>
<td>0.1077</td>
<td>0.1078</td>
<td>0.5071</td>
<td>0.4459</td>
</tr>
<tr>
<td>-1</td>
<td>0.2475</td>
<td>0.2468</td>
<td>0.2468</td>
<td>0.3486</td>
<td>0.0966</td>
<td>0.0966</td>
<td>0.4892</td>
<td>0.4387</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.2468</td>
<td>0.0496</td>
<td>0.0458</td>
<td>0.1138</td>
<td>0.0132</td>
<td>0.0966</td>
<td>0.3634</td>
<td>0.3038</td>
</tr>
<tr>
<td>0</td>
<td>0.0697</td>
<td>0.0318</td>
<td>0.0318</td>
<td>0.0988</td>
<td>0.0465</td>
<td>0.0243</td>
<td>0.3155</td>
<td>0.2975</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0434</td>
<td>0.0389</td>
<td>0.0374</td>
<td>0.0216</td>
<td>0.0135</td>
<td>0.0174</td>
<td>0.0603</td>
<td>0.0403</td>
</tr>
<tr>
<td>2/3</td>
<td>0.0243</td>
<td>0.0176</td>
<td>0.0176</td>
<td>0.0180</td>
<td>0.0165</td>
<td>0.0165</td>
<td>0.0205</td>
<td>0.0124</td>
</tr>
<tr>
<td>1</td>
<td>0.0090</td>
<td>0.0090</td>
<td>0.0103</td>
<td>0.0068</td>
<td>0.0068</td>
<td>0.0068</td>
<td>0.0092</td>
<td>0.0092</td>
</tr>
<tr>
<td>2</td>
<td>0.0345</td>
<td>0.0305</td>
<td>0.0299</td>
<td>0.0577</td>
<td>0.0183</td>
<td>0.0162</td>
<td>0.0924</td>
<td>0.0277</td>
</tr>
</tbody>
</table>
Figure 52: True and Approximate Significance Level - Symmetric Hypothesis at Nominal Chi-Square Level \( \alpha = 0.04 \), \( n = 16 \), \( k = 4 \).
FIG. 5.3 - TRUE AND APPROXIMATE SIGNIFICANCE LEVEL - SYMMETRIC HYPOTHESIS
AT NOMINAL CHI-SQUARE LEVEL $\alpha = 0.01, n = 16, k = 4$
will always contain all such partitions. Since \( D_n^{(\eta)} \) and \( D_n^{(\eta)**} \) are infinite whenever \( D_n^{(\eta)} \) is infinite, the approximations are not suitable. Therefore, for \( \eta < 0 \) and for moderate sample size, the use of \( D_n^{(\eta)} \) is not appropriate.

(2) For values of \( \eta \) somewhere between 0.5 and 1.5, the exact level of significance is very close to nominal level of significance. The chi-square approximation is quite adequate to the asymptotic distribution of \( D_n^{(\eta)} \).

(3) The chi-square approximation is quite satisfactory for the Pearson chi-square statistic. Similar conclusion is also found in Good, Gover and Mitchell (1970), Chapman (1976) and Larntz (1978). Larntz (1978) used extensive simulation to compare the errors in approximating the Pearson chi-square statistic, likelihood ratio statistic and Freeman-Tukey statistic with a chi-square random variable. He concludes that the approximate tests based on the likelihood ratio and Freeman-Tuky statistics yield exact levels that are typically in excess of the nominal levels for moderate cell frequencies. On the other hand the Pearson chi-square statistic attains exact levels that are quite close to the nominal levels.

(4) The two approximations produce significance levels that are quite close to the nominal levels, especially in the range [0.5 to 2.0]. Among the two approximations, the second approximation produce significance levels much closer to the nominal level than the first one. Thus Freeman-Tuky chi-square statistic can be used with the second approximation.

(5) We have not investigated the case of non-symmetric null hypothesis. However the study of Larntz (1978) points out that the conclusions are not much different from the equiprobable hypothesis. Hence, we expect
similar conclusions in this case also.

5.4 Moments Under the Contiguous Alternative Hypothesis

It is well known that under the sequence of contiguous alternatives

\[ H_1 : \pi_i = \pi_{01} + \frac{C_i}{\sqrt{n}}, \quad i = 1, \ldots, k, \quad \sum C_i = 0 \]

the asymptotic distribution of \( D_n(p, \pi_0) \) is a non-central chi-square distribution with degrees of freedom \( k - 1 \) and non-centrality parameter

\[ \delta = \sum \frac{C_i^2}{\pi_{01}}. \]

We, now, obtain the moments of \( D_n(p, \pi_0) \) up to \( O(\frac{1}{\sqrt{n}}) \).

We have to compute the expected value and the second moment of the approximate expression for \( D_n(p, \pi_0) \) given in (5.2.2) under the alternative hypothesis. Note that

\[
E_1 \left( \frac{w_i^2}{\pi_{01}} \right) = (k - 1) + \frac{1}{\sqrt{n}} \sum \frac{C_i}{\pi_{01}} + \delta + O(n^{-1}),
\]

and

\[
E_1 \left( \frac{1}{\sqrt{n}} \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \frac{w_i^3}{\pi_{01}} \right) = \frac{1}{\sqrt{n}} \left[ 3 \left( \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \right) C_i - \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \pi_{01} C_i \right]
+ \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \frac{C_i^3}{\pi_{01}} + O(n^{-1}),
\]

where \( E_1(.) \) denotes the expected value under the alternative. The above expressions have been obtained after simplifications and the derivations are given in Appendix D. It follows from (5.4.2) and (5.4.3) that
\[ E_1 [ D_n(p, \pi_0)] = (k - 1) + \delta + \frac{1}{\sqrt{n}} \left[ \sum \frac{C_1}{\pi_{01}} + 3 \left( \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \frac{C_1}{\pi_{01}} \right) - \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \pi_{01} C_1 \right] + \frac{1}{\sqrt{n}} \left( \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \pi_{01} C_1 \right) \] + \frac{1}{\sqrt{n}} \left( \sum \frac{h''(\pi_{01})}{h'(\pi_{01})} \pi_{01} C_1 \right) \] + O(n^{-1}) \quad (5.4.4) \\

and
\[ E_1 [ D_n(\eta)] = (k - 1) + \frac{1}{\sqrt{n}} \left[ \sum \frac{C_1}{\pi_{01}} + (\eta - 1)(3 \sum \frac{C_1}{\pi_{01}} + \right] + O(n^{-1}) \quad (5.4.5) \]

The expression for \( E_1 [(D_n(p, \pi_0))^2] \) is unwieldy and hence we obtain expression for \( E_1 [(D_n(\eta))^2] \). The expression (5.2.5) contains first two terms of the \( O(1) \) and the next two terms to the \( O(\frac{1}{\sqrt{n}}) \). To obtain \( E_1 [(D_n(\eta))^2] \), we need the expected values of these four terms to the \( O(\frac{1}{\sqrt{n}}) \). Computations are quite lengthy and are presented in Appendix D. The expression for second moment of \( D_n(\eta) \) is
\[ E_1 [(D_n(\eta))^2] = k^2 - 1 + 2(k+1)\delta + \delta^2 + \frac{1}{\sqrt{n}} \left[ 2(k+3+\delta) \sum \frac{C_1}{\pi_{01}} \right] + 4 \sum \frac{C_1^3}{\pi_{01}^2} + 2(\eta - 1)(3(k+3+\delta) \sum \frac{C_1}{\pi_{01}} \right] + (k+5+\delta) \sum \frac{C_1^3}{\pi_{01}^2} \] + O(n^{-1}) \quad (5.4.6) 

The Table 5.4.1 gives the expressions for the expected value and the second moment of \( D_n(\eta) \) for different values of \( \eta \). Our computations agree with the results obtained by Cressie and Read (1984) for the Pearson chi-square statistic and Freeman-Tukey statistic.
Table 5.4.1 Second order moments of various members of $D_n^{(n)}(P, \pi_0)$ under the contiguous alternative hypothesis

| value of $n$ | Name of the test statistic | $E[D_n^{(n)}(P, \pi_0)|H_1]$ | $E[(D_n^{(n)}(P, \pi_0))^2 | H_1]$ |
|--------------|-----------------------------|--------------------------------|---------------------------------|
| -2           |                             | $(k-1) + \delta - \frac{1}{\sqrt{n}} \left[ 5 \sum \frac{C_1}{\pi_0} + 2 \sum \frac{C_1^3}{\pi_0} \right]$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 - \frac{1}{\sqrt{n}} \left[ 16(k+3+6) \sum \frac{C_1}{\pi_0} \right]$ |
| -1           |                             | $(k-1) + \delta - \frac{1}{\sqrt{n}} \left[ 2 \sum \frac{C_1}{\pi_0} + \sum \frac{C_1^3}{\pi_0} \right]$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 - \frac{1}{\sqrt{n}} \left[ 10(k+3+6) \sum \frac{C_1}{\pi_0} \right]$ |
| 0            |                             | $(k-1) + \delta - \frac{1}{\sqrt{n}} \left[ 2 \sum \frac{C_1}{\pi_0} + \sum \frac{C_1^3}{\pi_0} \right]$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 - \frac{1}{\sqrt{n}} \left[ 2(k+3+6) \sum \frac{C_1}{\pi_0} \right]$ |
| 1/2          | Freeman-Tukey               | $(k-1) + \delta - \frac{2}{n} \left[ \sum \frac{C_1}{\pi_0} + \sum \frac{C_1^3}{\pi_0} \right]$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 - \frac{1}{\sqrt{n}} \left[ (k+3+6) \sum \frac{C_1}{\pi_0} \right]$ |

Freeman-Tukey
| Value of $\eta$ | Name of the test statistic | $E[D_n^{(n)}(P, \pi_0) | H_1]$ | $E[(D_n^{(n)}(P, \pi_0))^2 | H_1]$ |
|----------------|-----------------------------|--------------------------------|----------------------------------|
| 2/3            |                              | $(k-1) + \delta - \frac{1}{3\sqrt{n}} \sum \frac{C_1^3}{\pi_{01}^2}$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 - \frac{1}{\sqrt{n}} \left[ (k+3+\delta) \sum \frac{C_1}{\pi_{01}} + \frac{2}{3}(k+1+\delta) \sum \frac{C_1^3}{\pi_{01}^2} \right]$ |
| 1              | Pearson chi-square           | $(k-1) + \delta + \frac{1}{\sqrt{n}} \sum \frac{C_1}{\pi_{01}}$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 + \frac{2}{\sqrt{n}} \left[ (k+3+\delta) \sum \frac{C_1}{\pi_{01}} + 4 \sum \frac{C_1^3}{\pi_{01}^2} \right]$ |
| 2              |                              | $(k-1) + \delta + \frac{1}{\sqrt{n}} \left[ 4 \sum \frac{C_1}{\pi_{01}} + \sum \frac{C_1^3}{\pi_{01}^2} \right]$ | $k^2 - 1 + 2(k+1)\delta + \delta^2 + \frac{1}{\sqrt{n}} \left[ 8(k+3+\delta) \sum \frac{C_1}{\pi_{01}} + 2(k+5+7) \sum \frac{C_1^3}{\pi_{01}^2} \right]$ |
Looking at expressions (5.4.5) and (5.4.6) we notice that the second order terms are non-zero for all members of $D_n(n)$. For the symmetric hypothesis, the second order term in the first moment vanishes for $n=1$ and for

$$\eta = \frac{1}{1+ \frac{2}{k+3+k^2} \sum \xi_i^3}$$

in the second moment. Note that $\eta \to 1$ as $k$ becomes large. This indicates that under the contiguous alternatives also, the Pearson chi-square statistic provides closer approximation to the non-central chi-square distribution. We have obtained a closer approximation to the asymptotic distribution of these statistics in Chapter III.

5.5 Limiting Normal Distribution

The chi-square approximation for $D_n(P, \pi_0)$ is derived under the condition that the number of cells $k$ is fixed and the sample size tends to infinity. It is interesting to investigate the limiting distribution under the condition that the number of cells $k$ tends to infinity with $n$. Intuitively it should be the normal because as degrees of freedom increases chi-square distribution must tend to the normal distribution. We consider the situation in which $n$ and $k$ simultaneously increase in such a way that $n/k \to \alpha$. Our result is a direct application of the theorem proved by Holst (1972).

Theorem 5.1. Suppose $n \to \infty$ and $k \to \infty$ in such a way that $n/k \to \alpha (0 < \alpha < \infty)$, and suppose $k \pi_i \leq d < \infty$, $i = 1, \ldots, k$, for all $k$. Define

$$S_k = \sum_{i=1}^{k} f_i(X_i)$$

(5.5.1)
The functions $f_i(.)$ are assumed real valued functions such that

$$
\sigma_n^2 = \sum_{i=1}^{k} \text{Var}(f_i(Y_i)) - \frac{1}{n} \left[ \sum_{i=1}^{k} \text{Cov}(Y_i, f_i(Y_i)) \right]^2
$$

satisfies

$$
0 < \lim \inf_{n \to \infty} \frac{\sigma_n^2}{n} \leq \lim \sup_{n \to \infty} \frac{\sigma_n^2}{n} < \infty,
$$

where $Y_1, \ldots, Y_k$ are independent Poisson random variables with means $\pi_1; \ldots; \pi_k$, respectively. Moreover, assume the $f_i$ satisfy $|f_i(x)| \leq c e^{b x}$, $i = 1, \ldots, k$, where $c > 0$, $b > 0$. Suppose

$$
\mu_n = \sum_{i=1}^{k} E(f_i(Y_i)).
$$

Then $(S_k - \mu_n)/\sigma_n$ converges in distribution to the standard normal variable.

**Corollary 5.1.** Suppose $n$ and $k \to \infty$ such that $n/k \to \alpha$ ($0 < \alpha < 1$) and $\pi_{i0} = 1/k$ for $i = 1, \ldots, k$. Then

$$
D_n \left( \frac{p}{\alpha k}, \frac{1}{k} \right) - \frac{\mu_n}{\sigma_n}, \quad \eta > 0
$$

converges in distribution to the standard normal random variable, where

$$
\mu_n = \frac{n}{n^2 \alpha^2} \left[ E(\gamma^{2n}) + a^n - 2a^n E(\gamma^n) \right], \quad (5.5.2)
$$

and

$$
\sigma_n^2 = \frac{k}{n} \left[ E(\gamma^{4n}) + 4a^{2n} E(\gamma^{2n}) - 4a^n E(\gamma^{3n}) - (E(\gamma^{2n}))^2 \right. \\
- 4a^n (E(\gamma^{2n}))^2 + 4a^n E(\gamma^{2n}) E(\gamma^n) - \frac{1}{\alpha} \left\{ E(\gamma^{2n+1}) \right. \\
- 2a^n E(\gamma^{2n+1}) - \alpha E(\gamma^{2n}) + 2a^{n+1} E(\gamma^n) \right\}^2 \right], \quad (5.5.3)
$$
where \( Y \) is a Poisson random variable with mean \( \alpha \).

Proof. Note that

\[
D_n(T_1) ( \frac{1}{\alpha}, 1/k ) = \frac{\sum t (n \pi_1^n - \alpha^n)^2}{n^2 \alpha^n - 1}
\]

Therefore \( f_1(x) \) can be identified as

\[
f_1(x) = \frac{(x^n - \alpha^n)^2}{n^2 \alpha^n - 1}
\]

Therefore

\[
\mu_n = \sum_{i=1}^{K} E(f(Y_i))
\]

\[
= \frac{n}{n^2 \alpha^n} \left[ E(Y^{2n}) + \alpha^n - 2\alpha E(Y^n) \right].
\]

Also we obtain

\[
\text{Var}(f(Y)) = \frac{1}{n^4 \alpha^n - 2} \left[ E(Y^{2n}) + 4\alpha^n E(Y^{2n}) - 4\alpha E(Y^{3n})
\right.
\]

\[
- (E(Y^{2n}))^2 - 4\alpha^n (E(Y^{2n}))^2 + 4\alpha^n (E(Y^{2n}))(E(Y^n)) \right],
\]

\[\ldots \ (5.5.4)\]

and

\[
\text{Cov}(Y, f(Y)) = \frac{1}{n^2 \alpha^n - 1} \left[ E(Y^{2n+1}) - 2\alpha E(Y^{n+1}) - \alpha E(Y^{2n})
\right.
\]

\[
+ 2\alpha E(Y^{n+1}) E(Y^n) \right] \ (5.5.5)
\]

Using (5.5.4) and (5.5.5) after simplification we arrive at the expression for \( \sigma_n^2 \) as given in (5.3.3).

Holst (1972) and Ivchenko and Medvedev (1978) have also considered the distribution under the alternative hypothesis

\[
H_1: \pi_1 = 1/k + \frac{C_1}{n^{1/4}}, \quad \sum C_1 = 0 \quad i = 1, \ldots, k. \quad (5.5.6)
\]
Using the Theorem 5.1 it follows that under the alternative hypothesis (5.5.6) also, $D_n(\eta) (p, 1/k)$ has got an asymptotic normal distribution as $n \to \infty$, $k \to \infty$ in such a way that $n/k = \alpha$. From this result it follows that the asymptotic power of the test is monotonic in

$$e_{\eta} = \lim_{n \to \infty, k \to \infty} \frac{\mu_{n,1} - \mu_{n,0}}{\sigma_{n,1}},$$

(5.5.7)

where $\mu_{n,1}$ and $\mu_{n,0}$ denote the mean of $D_n(\eta) (p, 1/k)$ under alternative and null hypothesis respectively and $\sigma_{n,1}$ denotes the standard deviation under $H_1$. From the results of Ivchenko and Medvedev (1978) it follows that

$$e_{\eta} = \sqrt{\frac{\alpha}{2}} g^2 \text{corr}(f(Y) - aY, Y^2 - (2\alpha + 1)Y)$$

(5.5.8)

where

$$g^2 = \frac{1}{n} \sum C_i^2$$

and

$$a = \frac{1}{\alpha} \text{cov}(f(Y), Y).$$

From (5.5.8) it follows that the power is maximum when $|\text{corr}(f(Y) - aY, Y^2 - (2\alpha + 1)Y)| = 1$. From the properties of the correlation coefficient it follows that such a function is $f(X) = X^2$, which corresponds to Pearson statistic. Thus it follows that for any other member of $D_n(\eta) (p, 1/k)$ the Pitman efficiency of the test is less than 1, compared to Pearson statistic.
APPENDIX D

(a) Moments of the Multinomial Distribution

We assume that \( \{ n_1, \ldots, n_k \} \sim \text{M}(n; \pi_1, \ldots, \pi_k) \). We list below the moments of \( n_1 \)'s:

\[
\begin{align*}
E(n_1) &= n \pi_{01}, \\
E(n_1^2) &= n(2) \pi_{01}^2 + n \pi_{01}, \\
E(n_1 n_2) &= n(2) \pi_{01} \pi_{02}, \\
E(n_1^3) &= n(3) \pi_{01}^3 + 3n(2) \pi_{01}^2 \pi_{02}^{+} + n \pi_{01}, \\
E(n_1^2 n_2) &= n(3) \pi_{01}^2 \pi_{02} + n(2) \pi_{01} \pi_{02}, \\
E(n_1^4) &= n(4) \pi_{01}^4 + 6n(3) \pi_{01}^3 \pi_{02}^{+} + 7n(2) \pi_{01}^2 \pi_{02}^{+} + n \pi_{01}, \\
E(n_1^2 n_2^2) &= n(4) \pi_{01}^2 \pi_{02}^2 + n(3) \pi_{01}^2 \pi_{02} + n(3) \pi_{01} \pi_{02}^2 + n(2) \pi_{01} \pi_{02}, \\
E(n_1^3 n_2) &= n(4) \pi_{01}^3 \pi_{02} + 3n(3) \pi_{01}^2 \pi_{02}^2 + n(2) \pi_{01} \pi_{02}, \\
E(n_1^5) &= n(5) \pi_{01}^5 + 10n(4) \pi_{01}^4 \pi_{02}^{+} + 25n(3) \pi_{01}^3 \pi_{02}^{+} + 15n(2) \pi_{01}^2 \pi_{02}^{+} + n \pi_{01}, \\
E(n_1^4 n_2) &= n(5) \pi_{01}^4 \pi_{02} + 6n(4) \pi_{01}^3 \pi_{02}^2 + 7n(3) \pi_{01}^2 \pi_{02}^2 + n(2) \pi_{01} \pi_{02}, \\
E(n_1^3 n_2^2) &= n(5) \pi_{01}^3 \pi_{02}^2 + 3n(4) \pi_{01}^2 \pi_{02} + n(4) \pi_{01} \pi_{02}^3 + n(2) \pi_{01} \pi_{02}, \\
E(n_1^6) &= n(6) \pi_{01}^6 + 15n(5) \pi_{01}^5 \pi_{02}^{+} + 65n(4) \pi_{01}^4 \pi_{02}^{+} + 90n(3) \pi_{01}^3 \pi_{02}^{+}
+ 31n(2) \pi_{01}^2 + n \pi_{01},
\end{align*}
\]
\[
E(n_1^5 n_j^3) = n(6) \pi_{01}^5 \pi_{0j}^3 + 10n(5) \pi_{01}^4 \pi_{0j}^2 + 25n(4) \pi_{01}^3 \pi_{0j}^2 \\
+ 15n(3) \pi_{01}^2 \pi_{0j}^3 + n(2) \pi_{01} \pi_{0j}^4 ,
\]
\[
E(n_1^4 n_j^2) = n(6) \pi_{01}^4 \pi_{0j}^2 + 6n(5) \pi_{01}^3 \pi_{0j}^2 + n(5) \pi_{01}^2 \pi_{0j}^4 \\
+ 7n(4) \pi_{01}^2 \pi_{0j}^2 + 6n(4) \pi_{01} \pi_{0j}^3 + 7n(3) \pi_{01} \pi_{0j}^2 \\
+ n(3) \pi_{01} \pi_{0j}^3 + n(2) \pi_{01} \pi_{0j}^4 ,
\]
\[
E(n_1^3 n_j^3) = n(6) \pi_{01}^3 \pi_{0j}^3 + 3n(5) \pi_{01}^3 \pi_{0j}^2 + 3n(5) \pi_{01}^2 \pi_{0j}^3 \\
+ 9n(4) \pi_{01}^2 \pi_{0j}^2 + n(4) \pi_{01} \pi_{0j}^3 + n(4) \pi_{01} \pi_{0j}^4 \\
+ 3n(3) \pi_{01} \pi_{0j}^3 + 3n(3) \pi_{01} \pi_{0j}^2 + n(2) \pi_{01} \pi_{0j}^4 , \quad (D.1)
\]
where \( n(r) = n(n-1) \ldots (n-r+1) \).

(b) Moments of \( w_i \)'s when \( \pi_i = \pi_{01} \), \( i = 1, \ldots, k \).

Using the raw moments, we now compute the moments of
\( w_i = \sqrt{n(p_i - \pi_{01})} \), \( i = 1, \ldots, k \). We have computed the relevent part of the moments ignoring the lower order terms. They are listed below
\[
E(w_1^2) = \pi_{01} (1 - \pi_{01}) ,
\]
\[
E(w_1^3) = n^{-1/2} \pi_{01} (1 - \pi_{01}) (1 - 2\pi_{01}) ,
\]
\[
E(w_1^4) = 3\pi_{01}^2 (1 - \pi_{01})^2 + O(n^{-1}) ,
\]
\[
E(w_1^5) = n^{-1/2} (-20 \pi_{01}^5 + 50 \pi_{01}^4 - 40 \pi_{01}^3 + 10 \pi_{01}^2 ) + O(n^{-1}) ,
\]

\[ E(w^3_1 w^2_j) = n^{-1/2} \left( -20 \pi_{o1}^3 \pi_{o3}^2 + 5 \pi_{o1}^3 \pi_{oJ} + 15 \pi_{o1}^2 \pi_{oJ}^2 \right. \\
- 6 \pi_{o1}^2 \pi_{oJ} + \pi_{o1} \pi_{oJ} - \pi_{o1} \pi_{oJ}^2 \bigg) + O(n^{-1}), \quad i \neq j \]

\[ E(w^6_1) = -15 \pi_{o1}^6 + 45 \pi_{o1}^5 - 45 \pi_{o1}^4 + 15 \pi_{o1}^3 + O(n^{-1}), \]

\[ E(w^3_1 w^3_j) = -15 \pi_{o1}^3 \pi_{oJ}^3 + 9 \pi_{o1}^3 \pi_{oJ}^2 + 9 \pi_{o1}^2 \pi_{oJ}^3 - 9 \pi_{o1}^2 \pi_{oJ}^2 \]

\[ + O(n^{-1}), \quad i \neq j \]

and

\[ E(w^4_1 w^2_j) = -15 \pi_{o1}^4 \pi_{oJ}^2 + 3 \pi_{o1}^4 \pi_{oJ} + 18 \pi_{o1}^3 \pi_{oJ} - 6 \pi_{o1}^3 \pi_{oJ} \]

\[ -3 \pi_{o1}^2 \pi_{oJ}^2 + 3 \pi_{o1}^2 \pi_{oJ} + O(n^{-1}), \quad i \neq j. \]  

We illustrate the computation of one of them namely \( E(w^3_1 w^2_j) \).

\[ E(w^3_1 w^2_j) = \frac{1}{n^{5/2}} E(n_1 - n \pi_{o1})^3 (n_j - \pi_{oJ})^2 \]

and

\[ (n_1 - n \pi_{o1})^3 (n_j - n \pi_{oJ})^2 = n_1^3 n_j^2 - 2n_1^3 n_j (n \pi_{oJ}) + n_1^3 (n \pi_{oJ})^2 \]

\[ - 3n_1^2 n_j^2 (n \pi_{o1}) + 6n_1^2 n_j (n \pi_{o1}) (n \pi_{oJ}) - 3n_1^2 (n \pi_{o1}) (n \pi_{oJ})^2 \]

\[ - 3n_1 (n \pi_{o1})^2 (n \pi_{oJ})^2 - n_j^2 (n \pi_{o1})^2 + 2 n_j (n \pi_{o1})^3 (n \pi_{oJ}) \]

\[ - (n \pi_{o1})^3 (n \pi_{oJ})^2 + 3n_1 n_j^2 (n \pi_{o1})^2 - 6n_1 n_j (n \pi_{o1})^2 (n \pi_{oJ}) \]  

(D.3)

In the above expression the relevent terms to be considered are \( n_1^3 n_j^2 \),

\[ -2n_1^3 n_j (n \pi_{oJ}) \] and \(-3n_1^2 n_j^2 (n \pi_{oJ})\). We have taken expectation of these terms and used expression for \( n(r) \) [Bishop, Fienberg and Holland, 1975] to simplify the expressions. After lengthy simplifications we get
The term we need is
\[
\frac{2}{\sqrt{n}} \sum_{i,j} \frac{E(w_{i}^3 w_{j}^2)}{\pi_{i0} \pi_{j0}} \quad \text{[Cf (5.2.5)]. From (D.4)}
\]

after simplifications we get

\[
2 \sum_{i,j \neq j} \frac{E(w_{i}^3 w_{j}^2)}{\pi_{i0} \pi_{j0}} = \frac{2}{\sqrt{n}} \left[ -40 + 20 \sum \pi_{0i}^2 + 27k - 6k^2 + S(k-2) \right] + O(n^{-1}).
\]

Similarly \(3/2 \sum \frac{E(w_{i}^6)}{\pi_{0i}^2} = \frac{45}{2} \left[ -\pi_{0i}^2 + 3 - 3k + S \right] + O(n^{-1})\),

\[
\sum \frac{E(w_{i}^4)}{\pi_{0i}^2} = 3k - 6 + 3 \sum \pi_{0i}^2 + \frac{1}{n} \left[ S - 7k + 12 - 6 \sum \pi_{0i}^2 \right] + O(n^{-3/2})
\]

\[
\sum_{i,j \neq j} \frac{E(w_{i}^2 w_{j}^2)}{\pi_{i0} \pi_{j0}} = -3 \sum \pi_{0i}^2 - 3k + 5 + k^2 + \frac{1}{n} \left[ 6 \sum \pi_{0i}^2 - k^2 + 5k - 10 \right] + O(n^{-3/2})
\]

\[
\sum_{i,j \neq j} \frac{E(w_{i}^3 w_{j}^3)}{\pi_{i0} \pi_{j0}^2} = 3 \left[ -11 + 9k - 3k^2 + 5 \sum \pi_{0i}^2 \right] + O(n^{-1})
\]

\[
\sum_{i,j \neq j} \frac{E(w_{i}^4 w_{j}^2)}{\pi_{i0} \pi_{j0}^3} = 3 \left[ -12 + 5 \sum \pi_{0i}^2 + 10k - 2k^2 + S(k-2) \right] + O(n^{-1})
\]

and

\[
\sum \frac{E(w_{i}^5)}{\pi_{0i}^3} = \frac{2}{\sqrt{n}} \left[ -20 \sum \pi_{0i}^2 + 50 - 40k + 10S \right] + O(n^{-1}). \quad (D.5)
\]
(c) Moments of \( w_i \)'s when \( \pi_1 = \pi_{11} \), where \( \pi_{11} = \pi_{01} + \frac{C_1}{\sqrt{n}} \)

\( w_i, \ i = 1, \ldots, k \) can be rewritten as

\[
\begin{align*}
\quad w_i &= \sqrt{n}(p_1 - \pi_{01}) \\
&= \sqrt{n}(p_1 - \pi_{11}) + \sqrt{n}(\pi_{11} - \pi_{01}) \\
&= (v_1 + C_1), \quad (D.6)
\end{align*}
\]

where \( v_1 = \sqrt{n}(p_1 - \pi_{11}) \). Since we are computing the expected values under the alternative hypothesis, we substitute for \( w_i, \ i = 1, \ldots, k \) from \( (D.6) \) and obtain the required moments. We shall illustrate this

\[
E_1 \left[ \frac{w_i^4}{\pi_{01}} \right] \text{ and list the other necessary expectations. From binomial expansion, we have}
\]

\[
w_i^4 = v_1^4 + 4v_1^3C_1 + 6v_1^2C_1^2 + 4v_1C_1^3 + C_1^4. \quad (D.7)
\]

Further we can show that

\[
E_1 \left[ \frac{v_1^4}{\pi_{01}} \right] = 3 \sum \frac{\pi_{11}^2(1 - \pi_1)^2}{\pi_{01}^2} + O(n^{-1})
\]

\[
= 3k - 6 + 3 \sum \pi_{01}^2 + \frac{6}{\sqrt{n}} \sum C_1 \pi_{01}^2
\]

\[
= \frac{12}{\sqrt{n}} \sum C_1 \pi_{01} + O(n^{-1}). \quad (D.8)
\]

\[
E_1 \left[ \frac{C_1v_1^3}{\pi_{01}} \right] = \frac{1}{\sqrt{n}} \sum \frac{C_1\pi_{11}(1 - \pi_{11})(1 - 2\pi_{11})}{\pi_{01}^2}
\]

\[
= \frac{2}{\sqrt{n}} \sum C_1 \pi_{01} + \frac{1}{\sqrt{n}} \sum \frac{C_1}{\pi_{01}} + O(n^{-1}). \quad (D.9)
\]
and

$$E_1 \left[ \Sigma \frac{C_1^2 \nu_1^2}{\pi_{01}} \right] = \Sigma \frac{C_1^2 \pi_{11}(1 - \pi_{k1})}{\pi_{01}}$$

$$= \delta + \frac{1}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}^2} - \Sigma \frac{C_1^2}{\pi_{01}}$$

$$- \frac{2}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}} + O(n^{-1}). \tag{D.10}$$

Combining (D.8), (D.9) and (D.10) we get

$$E_1 \left[ \Sigma \frac{w_1^4}{\pi_{01}^2} \right] = 3k - 6 + 3 \Sigma \frac{\pi_{01}^2}{\pi_{01}} + \frac{10}{\sqrt{n}} \Sigma \frac{C_1}{\pi_{01}}$$

$$+ \frac{20}{\sqrt{n}} \Sigma \frac{C_1}{\pi_{01}^2} + 6\delta + \frac{6}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}}$$

$$- 6 \Sigma \frac{C_1^2}{\pi_{01}} - \frac{12}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}^2} + \Sigma \frac{C_1^4}{\pi_{01}^2} + O(n^{-1}). \tag{D.11}$$

Similarly

$$E_1 \left[ \Sigma \frac{w_1^2 w_2}{\pi_{01}^3} \right] = -3 \Sigma \frac{\pi_{01}^2}{\pi_{01}} - 3k + 5 + k^2 - \frac{20}{\sqrt{n}} \Sigma \frac{C_1}{\pi_{01}}$$

$$+ \frac{2}{\sqrt{n}} (\delta + k - 2) \Sigma \frac{C_1}{\pi_{01}} + 2\delta(k - 2) + 6 \Sigma \frac{C_1^2}{\pi_{01}}$$

$$- \frac{2}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}^2} + \delta^2 - \Sigma \frac{C_1^4}{\pi_{01}^2} + \frac{12}{\sqrt{n}} \Sigma \frac{C_1^3}{\pi_{01}^2}$$

$$+ O(n^{-1}), \tag{D.12}$$

$$E_1 \left[ \Sigma \frac{w_1^5}{\pi_{01}^3} \right] = 15 \left[ \Sigma \frac{C_1}{\pi_{01}} + \Sigma \frac{C_1}{\pi_{01}^2} \right] + 10 \left[ \Sigma \frac{C_1^3}{\pi_{01}^2} - \Sigma \frac{C_1^3}{\pi_{01}} \right]$$

$$+ \Sigma \frac{C_1^5}{\pi_{01}^3} + O(n^{-1/2}). \tag{D.13}$$
\[ E_1 \left[ \sum_{i} \sum_{j} \frac{w_i^3 w_j^2}{\pi_{01}^2 \pi_{0j}^2} \right] = -15 \sum \frac{C_1}{\pi_{01}} - 3(1 - \delta) \sum \frac{C_1}{\pi_{01}} \\
+ 10 \sum \frac{C_1^3}{\pi_{01}^3} + (k - 5 + \delta) \sum \frac{C_1^3}{\pi_{01}^2} \\
- \sum \frac{C_1^5}{\pi_{01}^5} + 3(k - 1) \sum \frac{C_1}{\pi_{01}} + O(n^{-1/2}). \]  

... (D.14)