Spring 2010

Analysis of Discrete Data under Order Restrictions

Jeff Campbell

ANALYSIS OF OF DISCRETE DATA UNDER ORDER RESTRICTIONS

by

JEFF CAMPBELL

(Under the Direction of Dr. Broderick O. Oluyede)

ABSTRACT

Strategies for the analysis of discrete data under order restrictions are discussed. We consider inference for sequences of binomial populations, and the corresponding risk difference, relative risk and odds ratios. These concepts are extended to deal with independent multinomial populations. Natural orderings such as stochastic ordering and cumulative ratio probability ordering are discussed. Methods are developed for the estimation and testing of differences between binomial as well as multinomial populations under order restrictions. In particular, inference for sequences of ordered binomial probabilities and cumulative probability ratios in multinomial populations are presented. Closed-form estimates of the multinomial parameters under order restrictions and test procedures for testing equality of two multinomial populations against the notion of cumulative probability ratio ordering which is stronger than stochastic ordering of the distributions are presented. Numerical examples are given to illustrate the techniques developed.

Key Words: Binomial distributions; Restricted estimates; Multinomial distribution; Test procedures

2009 *Mathematics Subject Classification*: 62H05, 62H17

ANALYSIS OF OF DISCRETE DATA UNDER ORDER RESTRICTIONS

by

JEFF CAMPBELL

B.S. in Mathematics

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirement for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2010

ANALYSIS OF OF DISCRETE DATA UNDER ORDER RESTRICTIONS

by

JEFF CAMPBELL

Major Professor: Dr. Broderick O. Oluyede

Committee: Dr. Charles Champ

Dr. Hani Samawi

Electronic Version Approved:

May 2010

ACKNOWLEDGMENTS

I wish to thank Dr. Broderick O. Oluyede for his hard work in helping me put this document together, the other members of the committee, Dr. Charles Champ and Dr. Hani Samawi, and the faculty that I have worked with at Georgia Southern University for their dedication to mentoring students.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Many recent papers have been concerned with estimation and testing of binomial and multinomial populations under order restrictions (see for example Chacko (1966), Robertson (1978), Robertson and Wright (1981), Lee (1987), Robertson, Wright and Dykstra (1988), Dykstra, Kochar and Robertson (1991), and references therein). This thesis develops tests for the equality several binomial populations as well tests of homogeneity of two multinomial populations against order-restricted alternatives. When a study is concerned with situations in which the categories are ordered, a test of homogeneity of the two distributions against a stochastically ordered alternative, with treatment distribution being stochastically larger than control, will be of interest.

It is often the case that disease exposure in one group may exceed that for another group. If this is the case, it may be appropriate to consider estimation of the parameters of interest and development of test procedures that take into consideration, the appropriate order restriction with respect to the exposure probabilities. These estimates can then be used to construct estimates of the risk difference, relative risk and odds ratio, where they are appropriate.

The parameter p, $0 < p < 1$, is often used to represent one population proportion and q, $0 < q < 1$, is used to represent the other population. There are a few ways to compare these two parameters. The relative risk is the ratio of p to q, or $\frac{p}{q}$, and the odds ratio (OR) is given by $OR = \frac{p(1-q)}{q(1-p)}$. The odds ratio is a value used to compare two proportions. The odds ratio can be computed from values in the contingency table. In this equation p represents the percentage chance of the affliction occurring in the first group, and q represents the percentage chance of the affliction occurring

in the second group.

1.1 Gathering Data: Prospective and Retrospective Studies

In a retrospective study, a researcher gathers a numbers of individuals with and without the affliction and then counts the numbers in each group who possess and do not possess the proposed risk factor. In this case, having or not having the affliction is the independent variable and having or not having the proposed risk factor is the dependent variable.

In a prospective study, a researcher gathers two cohort of individuals with and without the proposed risk factor. In each group, they will count the number of individuals who have (or who will develop) the affliction. In this case, having or not having the proposed risk is the independent variable. The dependent variable is having or not having the affliction.

In prospective studies, relative risk is used. In a retrospective study, the odds ratio is used.

Risk factors are identified by determining whether they significantly increase or decrease the risk of developing a disease. The magnitude of increased/decreased risk is expressed as a relative risk or odds ratio. To determine the rates of disease by person, place and time we use absolute risk (incidence, prevalence), to identify the risk factors for the disease we use relative risk (or odds ratio) and to develop approaches for disease prevention we employ attributable risk/fraction. Incidence is the number of new cases of a disease occurring in during a given period of time divided by the number of individuals at risk of developing the disease during the same time.

Prevalence is the total number of existing cases of disease at a given point in time divided by the number of individuals in the population at the time.

Relative risk (RR) is ratio of incidence of disease in exposed individuals to the incidence of disease in non-exposed individuals (from a cohort/prospective study). If $RR > 1$, there is a positive association, and if $RR < 1$, there is a negative association. Similarly, odds ratio (OR) is the ratio of the odds that cases were exposed to the odds that the controls were exposed (from a case control/retrospective study). Odds ratios are only estimates of relative risks, since true incidence rates cannot be determined from case control studies.

In developing approaches to disease prevention, one should consider attributable risk (AR)/fraction (AF), where AR is the amount of disease incidence that can be attributed to a specific exposure. Note that AF is the proportion of disease incidence that can be attributed to a specific exposure (among those who were exposed) and AF is AR divided by incidence in the exposed X 100percent.

Indeed, all individuals, whether they have or have not been exposed to a risk factor, have some chance of developing a disease if no prevention measures have been taken. AR/AF estimates the risk above and beyond this baseline risk that all people have.

In epidemiology and statistics, the major study designs include case control (retrospective), cohort (prospective) and cross sectional (one point in time). In case control studies we identify affected and unaffected individuals and the risk factor data is collected retrospectively. In cohort (prospective) studies we identify unaffected individuals, the risk factor data collected at baseline and individuals are followed until occurrence of the disease. In this setting, the measures of the risk are absolute risk

(incidence), relative risk and attributable risk. The disadvantages in this case include lengthy follow-up and criteria/methods may change over time.

In cross sectional studies, the measures of risk include absolute risk (prevalence), odds ratio and attributable risk (if incidence is known). The disadvantages include biased assessment of exposure, and often cause-effect assessment cannot be established.

Often within clinical trials and the social sciences, one will have discrete data with ordinal categories. In clinical trial setting, researchers often cannot measure a physical quantity which would be a measure of the effectiveness of a treatment. Often they are judged upon manifestations of the drugs effectiveness instead of performing a measurement of a continuous variable.

Let Y be the response variable. In a clinical trial setting, the explanatory variable X might represent the treatment number. In a social science setting it will represent particular subsets of the populations in which the response variables are compared. The sample sizes for each group is fixed. The parameters for the contingency table are as follows:

$$\pi_{i,j} = P(Y = j|X = i) \tag{1.1}$$

and

$$\gamma_{i,j} = \pi_{i,1} + \pi_{i,2} + \pi_{i,3} + ... + \pi_{i,j}. \tag{1.2}$$

For 2 x c tables, the local odds ratio θ_j^L and cumulative odds ratio θ_j^C are defined as follows:

$$\theta_j^L = \frac{\pi_{1,j}\pi_{2,j+1}}{\pi_{1,j+1}\pi_{2,j}}, \tag{1.3}$$

and

$$\theta_j^C = \frac{\gamma_{1,j}(1-\gamma_{2,j})}{(1-\gamma_{1,j})\gamma_{2,j}} \tag{1.4}$$

respectively.

The null hypothesis is that treatment 1 is no better than treatment 2 corresponds to θ_j^C=1. The alternative hypothesis that treatment 2 is better than treatment 1 is $\theta_j^C > 1$. Note that the cumulative odds ratio θ_j^C is equivalent to merging the first j categories into one category and categories j+1 to c into another and computing the odds ratio of the resulting 2x2 contingency table as demonstrated above.

Another parameter of interest is the continuation odds ratio, which is the odds ratio of a 2x2 table formed when comparing the populations at one level to those above it. This is denoted as θ_j^{CO} and is calculated as follows.

$$\theta_j^{CO} = \frac{\pi_{1,j} * (\pi_{2,j+1} + \pi_{3,j+1} + ...\pi_{2,c})}{\pi_{2,j} * (\pi_{1,j+1} + \pi_{1,j+1} + ...\pi_{1,c})} \tag{1.5}$$

The continuation odds ratio being greater than or equal to 1 at all levels is equivalent of the cumulative odds ratio being greater than or equal to 1 at all levels. Similarly, the condition that the continuation odds ratio is greater than or equal to 1 at all levels with a the condition of being strictly greater than 1 at at least one level is equivalent for the same in the cumulative odds ratio. These conditions carry over for the local odds ratio as well.

In rxc tables, the values can be defined as follows.

$$\theta_{i,j}^L = \frac{P(Y = j|X = i)P(Y = j + 1|X = i + 1)}{P(Y = j|X = i + 1)P(Y = j + 1|X = i)}, \quad \text{Local Odds Ratio.} \tag{1.6}$$

$$\theta_{i,j}^C = \frac{P(Y \leq j|X = i)P(Y > j|X = i + 1)}{P(Y \leq j|X = i + 1)P(Y > j|X = i)}, \quad \text{Cumulative Odds Ratio.} \tag{1.7}$$

$$\theta_{i,j}^G = \frac{P(Y \leq j|X \leq i)P(Y > j|X > i)}{P(Y \leq j|X \geq i)P(Y > j|X < i)}, \quad \text{Global Odds Ratio.} \tag{1.8}$$

and

$$\theta_{i,j}^{CO} = \frac{P(Y = j|X = i)P(Y > j|X = i + 1)}{P(Y = j|X = i + 1)P(Y > |X = i)}, \quad \text{Continuation Odds Ratio.} \tag{1.9}$$

For rxc tables, the global odds ratio is formed by making a 2x2 table by merging all the cells into four groups depending on whether or not the level for Y (or column number) is greater than i and whether or not the group number (row number) is greater than j. The other three values are computed by making a 2xc table from two adjacent rows and computing them as you would in a 2xc table.

1.2 Outline of Results

The outline of this thesis is as follows. In chapter 1 background information is given about the collection and processing of categorical data. Chapter 2 explains the model used in procedures that involve a two-way contingency tables and deals with the parameter estimates under certain order restrictions. Chapter 3 introduces the truncated normal variable, and the various tests procedures are developed under some important order restrictions. Chapter 4 involves comparing several populations proportions to a standard when the standard proportion is either known or unknown. Chapter 5 contains results on procedures for comparing multinomial populations. Alternative procedures in which both test statistics have in the limit the chi-square distribution with k degrees of freedom are developed and presented.

CHAPTER 2

RESTRICTED ESTIMATES AND PROPERTIES

2.1 Introduction

In this chapter, the basic models for contingency tables are introduced. We present the setting, derive the maximum likelihood estimate (MLE) of the parameters under order restrictions, herein referred to as the restricted estimates of the parameter. We also give the restricted estimates of the risk difference, relative risk and odds ratio. Consistency of the restricted estimates is established in Theorem 2.1.

2.2 Sampling Distributions

In this section we present some of the sampling distributions are useful in the analysis of discrete data as far as the results presented in this thesis are concerned. The most common sampling distribution is the binomial, or the distribution independent sums of the Bernoulli random variables. The binomial probability mass function (pmf) is as follows

$$P(X=k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \quad \text{for } 0<p<1 \text{ and } k=0,1,2,3,....,n. \quad (2.1)$$

For trials with more than two outcomes we have the multinomial probability mass function(pmf) given by

$$p(n_1,n_2,...,n_k)=\frac{n!}{\prod_{i=1}^{k} n_i!}\prod_{i=1}^{k}\pi_i^{n_i}, \quad \text{for } n_i>0, \sum_{i=1}^{k} n_i=1, \text{ and } \sum_{i=1}^{k}\pi_i=1. \quad (2.2)$$

For the multinomial distribution, the mean and variance are given by

$$E(n_i)=n\pi_i, \quad \text{and} \quad var(n_i)=n\pi_i(1-\pi_i) \quad (2.3)$$

respectively. The covariance is $cov(n_i, n_j) = -n\pi_i\pi_j$.

When counts do not result from a fixed number of trials the Poisson Distribution comes into play. The Poisson distribution is a special case of the Binomial distribution. Each time subinterval within a given time interval T has a Bernoulli trial and the parameter for the probability is adjusted so the expected value over T is constant. The limit as the subintervals of time go to zero is the Poisson distribution. The probability mass function (pmf) is given by

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, ..., \cdot. \tag{2.4}$$

where λ is the expected number of occurrences and k the number of occurrences. A Poisson sampling distribution can be useful when the sample size is random rather than fixed, but often one does not have the luxury. To adapt to this, the Poisson can be conditioned on a fixed sample size. Let $Y_1, Y_2, ..., Y_k$ be independent Poisson variables with means $\lambda_1, \lambda_2, ..., \lambda_k$. Then conditional distribution of $Y_1, Y_2, ..., Y_k$ given $Y_1 + Y_2 + ... + Y_k = n$ is

$$\begin{aligned} P(Y_1 = n_1, Y_2 = n_2, ..., Y_k = n_k | \sum_{i=1}^{k} Y_i = n) &= \frac{\prod_{i=1}^{k} e^{-\lambda_i} \lambda_i^{n_i} / n_i!}{e^{-\sum_{i=1}^{k} \lambda_i} (\sum_{i=1}^{k} \lambda_i)^n / n!} \\ &= \frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} \pi_i^{n_i}, \end{aligned} \tag{2.5}$$

where $\pi_i = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i}$. The resulting pmf is that of the multinomial distribution. The multinomial distribution is an extension of the binomial distribution. The multinomial distribution plays very important role in the analysis of discrete data.

2.3 The Setting

In this chapter, some preliminary results are presented. Consider the following setting. We would like to test for evidence that one group (group I or group II) is more or less likely to have the exposure (or the affliction or any other characteristic). The data can be placed in the table of marginal information as follows.

Table 2.1: Table of Marginal Information

Groups	Exposure	Non-exposure	total
Group I	a_1	a_0	m_1
Group II	b_1	b_0	n_2
Total	$a_1 + b_1$	$a_0 + b_0$	$m_1 + n_1$

Note that $a_1 \sim$ BINOMIAL(m_1,p_1) and $b_1 \sim$ BINOMIAL(n_1,q_1) respectively. The exposure variable represents whether or not a person has a proposed risk factor. The response variable would indicate whether the person had the affliction or the characteristic of the proposed risk factor. We wish to carry out a test in which the null hypothesis is that p_1 is equal to q_1. In this table, the estimated odds ratio is given by

$$\hat{\theta} = \frac{a_1 b_0}{b_1 a_0}, \tag{2.6}$$

and the asymptotic standard error $ASE(\hat{\theta})$ is given by

$$ASE(\hat{\theta}) = \sqrt{\frac{1}{a_1} + \frac{1}{a_0} + \frac{1}{b_1} + \frac{1}{b_0}}. \tag{2.7}$$

A large sample confidence interval for the odds ratio can be easily obtained as is given by:

2.4 Restricted Estimates

Let X_j and Y_j be the number of patients that survive beyond a year in each group of hospital j, $j = 1, 2, ..., k$. Clearly, X_j and Y_j have binomial (m_j, p_j) and binomial (n_j, q_j) distributions respectively. Also, let x_j, y_j and y^* denote the number of successes for the j^{th} experimental treatment in each group and from the control group. The MLE of p_j, q_j and p^* are $\hat{p}_j = x_j/m_j$, $\hat{q}_j = y_j/n_j$ and $p^* = y^*/n^*$, where p^* is the response rate for the control experiment and n^* is the total number trials in the control group.

First, we give the maximum likelihood estimator of (p, q) subject to $p \geq q$, where p and q are parameters of the binomial (m, p) and binomial (n, q) distributions, with observed values x and y respectively. We state a lemma (see Oluyede (1994)) that establishes the existence of the maximum likelihood estimates (MLE) under the order restriction. Consistency of the restricted estimates is established in Theorem 2.1.

Lemma 2.1. If $p > q$ and $x/m < y/n$, then these exists $\bar{p}$, $0 < \bar{p} < 1$ for which $L(\bar{p}, \bar{p}, x, y) > L(p, q; x, y)$, where $L(p, q; x, y) = \binom{m}{x}\binom{n}{y} p^x (1-p)^{m-x} q^y (1-q)^{n-y}$.

As a consequence of the lemma, if $x/m < y/n$, then the likelihood function $L(p, q; x, y)$ subject to $p \geq q$ is maximized when $p = q$. The restricted maximum likelihood estimate of (p, q) subject to $p \geq q$ is

$$(p^*, q^*) = \begin{cases} ((x+y)/(m+n), (x+y)/(m+n)) & \text{if } x/m < y/n, \\ (x/m, y/n) & \text{if } x/m \geq y/n. \end{cases}$$

The restricted maximum likelihood estimate of the risk difference $(p - q)$ subject to

$p \geq q$ is

$$(p^* - q^*) = \begin{cases} 0 & \text{if } x/m < y/n, \\ (x/m - y/n) & \text{if } x/m \geq y/n. \end{cases}$$

If q is known, the maximum likelihood estimator of p subject to $p \geq q$ is

$$p^* = \begin{cases} x/m & \text{if } q \leq x/m < 1, \\ q & \text{if } 0 < x/m < q. \end{cases}$$

If q is known, the maximum likelihood estimator of the risk difference $p - q$ subject to $p \geq q$ is

$$p^* - q = \begin{cases} x/m - q & \text{if } q \leq x/m < 1, \\ 0 & \text{if } 0 < x/m < q. \end{cases}$$

The restricted estimate of the relative risk $\frac{p}{q}$ subject to $p \geq q$, when q is unknown is

$$\frac{p^*}{q^*} = \begin{cases} \frac{nx}{my} & \text{if } q \leq x/m \geq y/m, \\ 1 & \text{if } x/m < y/n. \end{cases}$$

If q is known, the restricted maximum likelihood estimator of the relative risk p/q subject to $p \geq q$ is

$$\frac{p^*}{q} = \begin{cases} \frac{x}{mq} & \text{if } q \leq x/m < 1, \\ 0 & \text{if } 0 < x < mq. \end{cases}$$

The restricted estimates of the odds ratio can be readily obtained.

Theorem 2.1.

(a) If q is known, p^* converges in probability to p when $m \to \infty$, provided $p \geq q$.

(b) (p^*, q^*) converges in probability to (p, q) when $m, n \to \infty$ provided $p \geq q$.

Note that the proof of part (a) will follow quite easily once the proof of part (b) is established, thus the statement and proof of the part (b) of the theorem is presented below, that is, we prove the result:

If $p \geq q$, then (p^*, q^*) converges in probability to (p, q), as $m, n \to \infty$.

Proof: If $p = q$, then for $\epsilon > 0$,

$$\begin{aligned} P(||(p^*, q^*) &- (p, q)|| > \epsilon) \\ &\leq P(|p^* - p| + |q^* - q| > \epsilon) \\ &\leq P(|p^* - p| > \epsilon/2) + P(|q^* - q| > \epsilon/2) \\ &+ 2P(|(p^* + q^*)/(m+n) - q| > \epsilon/2) \\ &< 4q(1-q)[1/m\epsilon^2 + 1/n\epsilon^2 + 1/(m+n)\epsilon^2]; \end{aligned} \tag{2.8}$$

by Chebychev's inequality, where $||Z|| = (\sum_{i=1}^{k} Z_i^2)^{1/2}$.

If $p > q$, then for $\epsilon > 0$,

$$\begin{aligned}
P(||(p^*, q^*) &- (p, q)|| > \epsilon) \\
&\leq P(|p^* - p| + |q^* - q| > \epsilon) \qquad (2.9) \\
&\leq P(|p^* - p| > \epsilon/2) + P(|q^* - q| > \epsilon/2) \\
&= P(|p^* - p| > \epsilon/2, p^* \geq q^*) + P(|p^* - p| > \epsilon/2, p^* < q^*) \\
&+ P(|q^* - q| > \epsilon/2, p^* \geq q^*) + P(|q^* - q| > \epsilon/2, p^* < q^*) \\
&\leq P(|p^* - p| > \epsilon/2) + P(|q^* - q| > \epsilon/2) \\
&+ 2P(|(p^* - q^*) - (p - q)| > (p - q)) \\
&\leq 4p(1-p)/m\epsilon^2 + 4q(1-q)/n\epsilon^2 \\
&+ 2[p(1-p)/m + q(1-q)/n]/(p-q)^2,
\end{aligned}$$

by Chebychev's inequality. Therefore if $p \geq q$, (p^*, q^*) converges in probability to (p, q) whenever m, $n \to \infty$ and the proof is complete. $\square$

2.5 Concluding Remarks

In this chapter, we have presented the estimates of the population parameters including the risk difference and relative risk under order restrictions. These estimates are also shown to be consistent. Theorem 2.1 shows that under the stated order restriction, the estimates converges in probability to the true population parameters for large m and n.

CHAPTER 3

TEST PROCEDURES AND DISTRIBUTIONS

3.1 Introduction

In this chapter, procedures are developed for comparing sequences of binomial populations under order restriction. Alternative procedures are also presented. First, we present some results on the truncated normal probability density function. This result is useful in the establishment of the asymptotic distribution of some of the test statistics considered in this chapter.

First we introduce the probability density function (pdf) of a truncated standard normal random variable. Let the probability density function of the truncated standard normal random variable R_j, $j = 1, 2, ..., k$, be given by

$$\begin{aligned} f(r_j) &= \frac{\phi_{0,1}(r_j) I_{(0,\infty)}(r_j)}{1 - \Phi_{0,1}(0)} \\ &= \frac{2exp(-r_j^2/2)}{\sqrt{2\pi}}, \end{aligned} \tag{3.1}$$

for $r_j > 0$, where $\phi_{0,1}(r_j)$, and $\Phi_{0,1}(r_j) = \int_{-\infty}^{r_j} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ are the standard normal probability density function and cumulative distribution function of R_j respectively.

The mean and variance of R_j are given by

$$E(R_j) = \sqrt{\frac{2}{\pi}} \quad \text{and} \quad Var(R_j) = 1 - \frac{2}{\pi}, \tag{3.2}$$

respectively.

If R_j, $j = 1, 2, ...k$, are independent, then $Cov(R_i, R_j) = 0$.

3.2 Test Statistics and Distributions

We now turn to the problem of comparing sequences of population parameters under order restrictions in the binomial data setting. Consider the problem of testing the hypothesis:

$H_0 : p_j - q_j = 0$ against $H_1 : p_j - q_j \geq 0$, $j = 1, 2, ..., k$ and $p_j - q_j > 0$ for at least one j.

When q_j, $j = 1, 2, ..., k$ are known, we reject H_0 in favor of H_1 if

$$T_A = \sum_{j=1}^{k} S_j \geq C, \tag{3.3}$$

where

$$S_j = 2\{x_j[ln(x_j/m_j) - lnq_j] + (m_j - x_j)[ln(1 - x_j/m_j) - ln(1 - q_j)]\}$$

if $x_j/m_j \geq q_j$ and 0 otherwise.

For large m_j, the log likelihood statistic can be approximated by

$$T_B = \sum_{j=1}^{k} (x_j/m_j - q_j)^2/[q_j(1 - q_j)/m_j] \quad \text{if} \quad x_j/m_j \geq q_j \quad \text{and 0 otherwise.} \tag{3.4}$$

Alternatively, it is well known that

$$R_j = \sin^{-1}\{(\hat{p}_j)^{1/2}\}, \tag{3.5}$$

$j = 1, 2, ..., k$ are approximately normally distributed with mean

$$\mu_j = \sin^{-1}\{(p_j)\}^{1/2}\} \quad \text{and variance} \quad \sigma_j^2 = 1/(4m_j). \tag{3.6}$$

Hence an approximate test is to reject H_0 in favor of H_1 if

$$T = \sum_{j=1}^{k} 4m_j(R_j^* - \bar{R})^2 \geq C^*, \tag{3.7}$$

where R_j^* is the restricted of R_j,

$$\bar{R} = \sum_{i=1}^{k} n_i R_i / \sum_{i=1}^{k} n_i \tag{3.8}$$

and

$$P(T \geq C^*) = \sum_{i=2}^{k} P(i,k;n) P_r(\chi^2_{i-1} \geq C^*), \tag{3.9}$$

where $P(i,k;n)$ are the level probabilities determined by $(n_1, n_2, ..., n_k)$, and χ^2_{i-1} is a chi-square random variable with $i-1$ degrees of freedom.

When q_j are unknown, the testing procedure is: Reject H_0 in favor of H_1 if

$$T_C = \sum_{j=1}^{k} T_j \geq C_B, \tag{3.10}$$

where C_B is chosen so that the test is an asymptotically α-level test, and T_j is given by

$$\begin{aligned} T_j &= 2\{x_j[ln(x_j/m_j) - ln((x_j+y_j)/(m_j+n_j))] \\ &\quad +(m_j - x_j)[ln(1 - x_j/m_j) - ln(1-(x_j+y_j)/(m_j+n_j))] \\ &\quad +y_j[ln(y_j/n_j) - ln((x_j+y_j)/(m_j+n_j))] \\ &\quad +(n_j - y_j)[ln(1 - y_j/n_j) - ln(1-(x_j+y_j)/(m_j+n_j)], \end{aligned} \tag{3.11}$$

if $x_j/m_j - y_j/n_j > 0$, and 0 otherwise.

For large m_j and n_j, $j = 1, 2, ..., k$, the test statistic T_C can be approximated by

$$T_D = \sum_{j=1}^{k} U_j, \tag{3.12}$$

where

$$U_j = (x_j/m_j - y_j/m_j)^2 / \{[(x_j+y_j)/(m_j+n_j)] \cdot [(1-(x_j+y_j)/(m_j+n_j)(1/m_j+1/n_j)]\} \tag{3.13}$$

if $x_j/m_j \geq y_j/m_j$ and 0 otherwise.

For large m_j and n_j, $j = 1, 2, ..., k$, T_C and T_D are equivalent and

$$\begin{aligned}
\lim_{m_j, n_j \to \infty} P_{H_0}(T_C > t) &= \lim_{m_j, n_j \to \infty} P_{H_0}(T_D > t) \\
&= \begin{cases} P(\sum_{j=1}^{k} Z_j^2 I_{[0,\infty)}(Z_j) > t) & \text{if } t > 0 \\ 1 & \text{if } t \geq 0, \end{cases} \\
&= \begin{cases} \sum_{j=1}^{k} \begin{pmatrix} k \\ j \end{pmatrix} (\frac{1}{2})^k P(\chi^2_{(j)} > t) & \text{if } t > 0 \\ 1 & \text{if } t \leq 0, \end{cases}
\end{aligned}$$

where $\chi^2_{(j)}$ is a random variable having a chi-square distribution with j degrees of freedom. See Oluyede (1993) for details.

3.3 Alternative Procedures

Let

$$W_j = \frac{(x_j/m_j - y_j/n_j)}{\{\sqrt{[(x_j+y_j)/(m_j+n_j)]\cdot[(1-(x_j+y_j)/(m_j+n_j))(1/m_j+1/n_j)]}\}} \quad (3.14)$$

if $x_j/m_j \geq y_j/n_j$ and 0 otherwise and define the vector of statistics W by

$$W = (W_1, W_2, ..., W_k)^T.$$

For testing the hypothesis $H_0 : p_j - q_j = 0$ against $H_1 : p_j - q_j \geq 0$, $j = 1, 2, ..., k$ and $p_j - q_j > 0$ for at least one j, the following test statistic is proposed

$$Y^2_{m,n} = W^T \Psi^{-1} W, \quad (3.15)$$

where Ψ is the covariance matrix of the random vector W, and is independent of the parameters. We reject H_0 for large values of $Y^2_{m,n}$.

Alternatively, for testing the hypothesis, reject H_0 for large values of

$$T = \sum_{j=1}^{k} W_j, \quad (3.16)$$

where W_j is given above.

These test statistics both characterizes the null and alternative hypotheses. They are small when H_0 is true and large under H_1.

Note that since each W_j, $j = 1, 2, ..., k$, has a standard normal distribution, asymptotically under H_1 the random vector W has in the limit as $n \to \infty$ the k dimensional normal distribution, $N_k(\lambda, \Psi)$, and hence the statistic

$$Y^2_{m,n} = W^T \Psi^{-1} W, \quad (3.17)$$

where

$$\Psi = diag((1 - \frac{2}{\pi}))\delta_{ij}, \tag{3.18}$$

$\delta_{ij} = I_{(i=j)}$, has as $m, n \to \infty$, the non-central chi-square distribution with k degrees of freedom and non-centrality parameter $\nu = \sum_{i=1}^{k} \lambda_i$, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k)^T$, and rank $(\Psi) = k$. Under H_0, $\lambda_i = 0$, $j = 1, 2, .., k$ and hence $Y_{m,n}^2$ has in the limit as $m, n \to \infty$ the chi-square distribution with k degrees of freedom.

3.4 Concluding Remarks

We have presented tests procedures concerning the risk difference and well as the relative risk under order restrictions. Several procedures are presented including alternative procedures that uses arcsine transformation leading to the asymptotic distributions of the test statistics including the truncated normal distribution, the chi-bar type distributions, and chi-square distributions for the hypotheses considered.

CHAPTER 4

COMPARISON WITH CONTROL AND OTHER RESULTS

4.1 Introduction

This chapter deals with comparison with a known standard or control. Also, presented are other alternative procedures for comparisons of binomial populations under order restrictions.

First consider a sequence of binomial populations denoted by $B(m_j, p_j)$, $j = 1, 2, ..., k$ and a standard $B(n, p^*)$, respectively.

4.2 Comparison with Standard

In this section, we consider testing the hypothesis:

$$H_0 : p_1 = p_2 = ... = p_k = p*$$

against

$$H_1 : p_j > p^* \text{for at least one } j,\ j = 1, 2, ..., k. \tag{4.1}$$

First consider testing

$$H_0 : p = p^* \text{against } H_1 : p > p^*.$$

The log likelihood ratio statistic is given by

$$2\{x[ln(x/m) - lnp*] + (m - x)[ln(1 - x/m) - ln(1 - p*)]\}$$

if $x/m \geq p^*$ and 0 otherwise, where x is the observed value of X, a random variable having a binomial distribution with parameters (m, p) and $p*$ is the response rate for the control treatment.

For testing (4.1) when $p*$ is known, we reject H_0 in favor of H_1 if

$$\max\{L_j, j = 1, 2, ..., k\} \geq C^*,$$

where

$$L_j = [(x_j/m_j - p^*)/\sqrt{p^*(1-p^*)/m_j}]I_{(0,\infty)}(x_j/m_j - p^*)] \tag{4.2}$$

and C^* is chosen such that

$$P(\max\{L_j, j = 1, 2, ..., k\} \geq C^*) = \alpha.$$

The critical values C^* can be determined from the asymptotic distribution of $\max\{L_j, j = 1, 2, ..., k\}$ under H_0. See Gupta et al (1985).

If p^* is unknown, reject H_0 whenever $\max\{W_j^*, j = 1, 2, ..., k\} \geq C^*$, where

$$W_j^* = \frac{(x_j/m_j - y^*/n^*)}{\sqrt{[((x_j + y^*)/(m_j + n^*))(1 - (x_j + y^*)/(m_j + n^*))(1/m_j + 1/n^*)]}}$$

if $x_j/m_j \geq y^*/n^*$ and 0 otherwise. (4.3)

The asymptotic distribution of $W^* = (W_1^*, W_2^*, ..., W_k^*)^T$ is k-dimensional normal $N_k(\underline{b}, B)$ under H_1. It is evident that if H_0 is true, then $\underline{b} = (0, 0...0)^T$ and $B = \sigma^2 I$, where $\sigma^2 = 1 - 2/\pi$.

4.3 Alternative Procedure

Alternatively, consider the following test statistic $L = (L_1, L_2, ..., L_k)^T$ for $p*$ known and $W^* = (W_1^*, W_2^*, ..., W_k^*)^T$ for $p*$ unknown. Asymptotically under H_1, the statistics have in the limit as $m_j \to \infty$, and $m_j, n_j \to \infty$, the k-dimensional normal distributions and hence the statistics

$$L_n^2 = L^T A^{-1} L \tag{4.4}$$

and

$$W_{m,n}^2 = W^{*T} B^{-1} W^* \tag{4.5}$$

have in the limit, the non central chi-square distribution with rank $(B) = k$ and non centrality parameter

$$\delta = \sum_{j=1}^{k} b_j, \underline{b} = (b_1, b_2, ..., b_k)^T.$$

Consequently, when H_0 is true, both statistics have in the limit the chi-square distribution with k degrees of freedom.

However, the test statistics

$$T_L = \sum_{j=1}^{k} L_j \quad \text{and} \quad T_W = \sum_{j=1}^{k} W_j^* \tag{4.6}$$

are stochastically no larger than L_n^2 and $W_{m,n}^2$. This follows from the fact that for every $t > 0$,

$$\begin{aligned} P(T_L > t) &= \sum_{j=1}^{k} \binom{k}{j} \left(\frac{1}{2}\right)^k P(\chi_{(j)}^2 > t) \\ &< \sum_{j=0}^{k} \binom{k}{j} \left(\frac{1}{2}\right)^k P(\chi_{(k)}^2 > t) \\ &= P(\chi_{(k)}^2 > t), \end{aligned}$$

since

$$P(\chi^2_{(k)} \geq t) \leq P(\chi^2_{(k+1)} \geq t) \tag{4.7}$$

for $k = 0, 1, 2, ...$ and for all t.

It follows that the tests based on T_L and T_W rejects H_0 more often than the tests based on L^2_n and $W^2_{m,n}$.

4.4 Concluding Remark

We have developed and presented procedures for comparing a sequence of binomial populations with a standard. Alternative procedures in which both test statistics (under the null hypothesis) have in the limit the chi-square distribution with k degrees of freedom are developed and presented. The distributions of the considered test statistics (under the alternative hypothesis) were shown to be the non-central chi-square distribution.

CHAPTER 5

COMPARISON OF MULTINOMIAL POPULATIONS

5.1 Introduction

This chapter deals with extensions to multinomial populations. These extensions provide local comparisons of multinomial populations via certain probability ratio and hazard probability orderings. Estimation of the multinomial probabilities under the orderings and appropriate tests procedures are developed.

5.2 Stochastic Domination in Multinomial Populations

Let U and V be two discrete random variables with $x_1 < x_2 < ... < x_k$ as possible values. Then, the random variable U is stochastically no larger than V, denote by $U \overset{st}{\prec} V$, if

$$\sum_{i=1}^{j} p_i \geq \sum_{i=1}^{j} q_i, \qquad \text{where } p_j = P(U = x_j) \text{ and } q_j = P(V = x_j), j = 1, 2, ..., k. \quad (5.1)$$

The next two definitions are due to Oluyede (1993).

Definition 5.2.1. *The random variable U is no larger than V in local probability ratio ordering, denoted by $U \overset{lpr}{\prec} V$, if and only if*

$$\frac{\sum_{i=1}^{j} p_i}{\sum_{i=1}^{j+1} p_i} \geq \frac{\sum_{i=1}^{j} q_i}{\sum_{i=1}^{j+1} q_i}, \quad (5.2)$$

with a strict inequality for at least one $j, j = 1, 2, ..., k-1$.

Similarly, the random variable U is locally no larger than V in probability ratio

star order, denoted by $U \overset{lpr^*}{\prec} V$, if and only if

$$\frac{q_{j+1}}{\sum_{i=1}^{j} q_i} \geq \frac{p_{j+1}}{\sum_{i=1}^{j} p_i}, \qquad \text{with a strict inequality for at least one } j, j = 1, 2, ..., k-1. \tag{5.3}$$

Suppose two medical treatments that lead to two multinomial populations can be characterized by P and Q. Assuming that the higher category corresponds to the best response, a reasonable criteria is to prefer treatment Q if (5.2) holds, or equivalently if $P \overset{lpr}{\prec} Q$, where $P = (p_1, p_2, ..., p_k)$ and $Q = (q_1, q_2, ..., q_k)$. In clinical trials for example, tumor response may be classified as stable disease, partial response and complete response.

In the comparison of two treatments, say old and new treatments, the appropriate null hypothesis specifies equal outcome probabilities across treatments against a one-sided alternative that patients receiving the new treatment have a larger probability of a favorable outcome than patients receiving the old treatment, whether or not partial response is considered favorable. In this case complete response is favorable and stable disease is not. Since there is a natural ordering in the tumor response the alternative hypothesis (5.2) and/or (5.1) provides a reasonable criteria for specifying the effectiveness of one treatment over the other.

There are several other examples that include global assessment (much improved, improved, no change, worse or much worse; Mehta, Patel, and Tsiatis, 1984, page 824), degree of toxicity (mild, moderate, severe, life-threatening, or drug death; Mehta, Patel, and Senchaudhuri, 1988, page 1002), clinical change (marked improvement, moderate improvement, slight improvement, stationary, or worse; Rahlfs and Zimmerman, 1993, page 1228), adverse event intensity (mild, moderate, severe, or intolerable; Chuang-Stein and Mohberg, 1993, page 246) and pneumonia evaluation (cure,

improvement, or failure; Spilker and Schoenfelder, 1991, page 581). A common problem in these examples is to determine whether or not one treatment is "better" than another on the basis of such ordered categories.

Local stochastic domination (5.2) or (5.3) is then an appropriate alternative, especially in situations where stochastic domination does not clearly reveal preference between the treatments. Models involving orderings of these types are easy to interpret and explain in real-world situations.

An alternative reparametrization using appropriate terminology for failure time data is as follows. If we suppose that an individual may fail in any one of $k+1$ time intervals $J_i = [c_{i-1}, c_i)$, where $0 < c_1 < c_2 < c_3 < .... < c_{k+1} = \infty$ and let p_i=P(an individual fails in J_i), then the number of individuals failing in $J_1, J_2, ..., J_{k+1}$ have a multinomial distribution with parameters n and $p_1, p_2, ..., p_{k+1}$. Noting the P(an individual fails in J_i given that they have not failed by c_{i-1}) is

$$\delta_i = \frac{p_i}{\sum_{j=i}^{k+1} p_j}. \tag{5.4}$$

It is clear that comparisons can be done via the multinomial hazard functions. Now, Let

$$\gamma_i = \frac{q_i}{\sum_{j=i}^{k+1} q_j}, \quad \text{with a strict inequality for at least one } j, j = 1, 2, ..., k. \tag{5.5}$$

Definition 5.2.2. *We say the random variable U is locally no larger than V in hazard order, denoted by $U \overset{lhr}{\prec} V$, if and only if*

$$\gamma_j \geq \delta_j, \qquad \textit{with a strict inequality for at least one } j, j = 1, 2, ..., k. \tag{5.6}$$

5.3 Restricted Estimates

Note that for a given $x > 0$, and $y > 0$, the expression $\alpha^x(1-\alpha)^y$ with the constraint that $0 \leq t \leq \alpha \leq 1$ is maximized at $\alpha^* = \max\left(\frac{x}{x+y}, t\right)$.

Now differentiating the expression $\alpha^x(1-\alpha)^y$ with respect to α gives

$$x\alpha^{x-1}(1-\alpha)^y - y(1-\alpha)^{y-1}\alpha^x = \alpha^{x-1}(1-\alpha)^{y-1}[x(1-\alpha) - \alpha y]. \tag{5.7}$$

The critical value or point is $\alpha = \frac{x}{x+y}$. This is the only critical point and it can be shown that this is a local maximum. Thus, if $t < \frac{x}{y+x}$, then this critical point is in the interval $[t, 1]$, where α is permitted to be and thus the max is at the critical point. If $t \geq \frac{x}{x+y}$, then because the function is decreasing on the interval $(\frac{x}{x+y}, 1]$, the maximum is attained at $\alpha = t$. Therefore we have the following lemma. See Oluyede (1994) for additional details on lemmas 5.1 and 5.2.

Lemma 5.1. The maximum of the function $\alpha^x(1-\alpha)^y$ subject to $0 \leq t \leq \alpha \leq 1$ is attained at

$$\alpha^* = max\{x/(x+y), t\}. \tag{5.8}$$

Let $\alpha_j = \frac{\sum_{i=1}^{j} p_i}{\sum_{i=1}^{j+1} p_i}$, $j = 1, 2, ..., k-1$, then $p_1 = \prod_{j=1}^{k-1} \alpha_j$, $p_i = (1-\alpha_{i-1}) \prod_{j=i}^{k-1} \alpha_j$, $i = 1, 2, ..., k-1$, and $p_k = 1 - \alpha_{k-1}$. The MLE of p_j, $j = 1, 2, ..., k$ is obtained by maximizing the likelihood function

$$\prod_{j=1}^{k-1} \alpha_j^{\sum_{i=1}^{j} x_i} (1-\alpha_j)^{x_{j+1}}, \qquad \text{for } 0 \leq \alpha_j \leq 1, \tag{5.9}$$

subject to $\alpha_j \geq \beta_j$, where

$$\beta_j = \frac{\sum_{i=1}^{j} q_i}{\sum_{i=1}^{j+1} q_i}, \qquad j = 1, 2, ..., k-1, \text{ is known.} \tag{5.10}$$

Applying lemma 5.1, we obtain the restricted maximum likelihood estimate given by

$$\alpha_j^* = max\{\sum_{i=1}^{j} x_i / \sum_{i=1}^{j+1} x_i, \beta_j\}, \text{ for } j = 1, 2, ..., k-1. \tag{5.11}$$

Lemma 5.2. The maximum of the function $\alpha^x(1-\alpha)^y\beta^r(1-\beta)^s$ subject to $0 \leq \beta \leq \alpha \leq 1$ is attained at

$$(\alpha^*, \beta^*) = \begin{cases} \left(\frac{x}{x+y}, \frac{r}{r+s}\right) & \text{if } \frac{x}{x+y} \geq \frac{r}{r+s}, \\ \left(\frac{x+r}{x+y+r+s}, \frac{x+r}{x+y+r+s}\right) & \text{if } \frac{x}{x+y} < \frac{r}{r+s}. \end{cases}$$

Applying lemma 5.2 to the following likelihood function

$$L = \prod_{j=1}^{k-1} \alpha_j^{\sum_{i=1}^{j} x_i} (1-\alpha_j)^{x_{j+1}} \beta_j^{\sum_{i=1}^{j} y_i} (1-\beta_j)^{y_{j+1}}, \tag{5.12}$$

where $\alpha_j = \frac{\sum_{i=1}^{j} p_i}{\sum_{i=1}^{j+1} p_i}$ and $\beta_j = \frac{\sum_{i=1}^{j} q_i}{\sum_{i=1}^{j+1} q_i}$, $j = 1, 2, ..., k-1$, we obtain the restricted maximum likelihood estimate of (α_j, β_j) given by

$$(\alpha_j^*, \beta_j^*) = \begin{cases} \left(\overline{\alpha}_j, \overline{\beta}_j\right) & \text{if } \overline{\alpha}_j \geq \overline{\beta}_j, \\ \left(\frac{\sum_{i=1}^{j} x_i + y_i}{\sum_{i=1}^{j+1} x_i + y_i}, \frac{\sum_{i=1}^{j} x_i + y_i}{\sum_{i=1}^{j+1} x_i + y_i}\right) & \text{if } \overline{\alpha}_j < \overline{\beta}_j, \end{cases}$$

where $\overline{\alpha}_j = \sum_{i=1}^{j} x_i / \sum_{i=1}^{j+1} x_i$ and $\overline{\beta}_j = \sum_{i=1}^{j} y_i / \sum_{i=1}^{j+1} y_i$ from which the MLE of (p_i, q_i), $i = 1, 2, ..., k$ can be readily obtained.

5.4 Estimates Under Stochastic Order

We present some basic results and closed-form estimates of P and Q subject to stochastic ordering. The test procedures in one and two-samples are developed in subsequent sections.

The estimation technique involves a reduction of the parameter space to a subspace containing the restricted estimates and provides an alternative method to the isotonic regression technique. Let $\Delta = \{(p_1, p_2, ..., p_k, q_1, q_2, ..., q_k) : p_i \geq 0, q_i \geq 0, i = 1, 2, ..., k, \sum_{i=1}^{k} p_i = \sum_{i=1}^{k} q_i = 1\}$ denote the parameter space.

Now, let $F_j = \sum_{i=1}^{j} p_i$ and $G_j = \sum_{i=1}^{j} q_j$. Then the likelihood function can be written as

$$L(p_1, p_2, ..., p_k, q_1, q_2, ..., q_k; x_1, x_2, ..., x_k, y_1, y_2, ..., y_k)$$
$$= \frac{m!}{\prod_{j=1}^{k} x_j!} F_1^{x_1} \prod_{j=2}^{k} (F_j - F_{j-1})^{x_j} \frac{n!}{\prod_{j=1}^{k} y_j!} G_1^{y_1} \prod_{j=2}^{k} (G_j - G_{j-1})^{y_j}. \tag{5.13}$$

The restricted MLE are given by

$$\begin{aligned} p^*_{h_r} &= (x_{h_r} / \sum_{i=j_r+1}^{j_{r+1}} x_i)(\sum_{i=j_r+1}^{j_{r+1}} (x_i + y_i)/(m+n)), \\ q^*_{h_r} &= (y_{h_r} / \sum_{i=j_r+1}^{j_{r+1}} y_i)(\sum_{i=j_r+1}^{j_{r+1}} (x_i + y_i)/(m+n)), \end{aligned} \tag{5.14}$$

$h_r = j_r + 1, \ldots, j_{r+1}, r = 0, 1, 2, ..., t.$

When $(q_1, q_2, \ldots, q_k)$ is known, the restricted MLE of $(p_1, p_2, ..., p_k)$ is given by

$$p^*_{h_r} = (x_{h_r} / \sum_{i=j_r+1}^{j_{r+1}} x_i)(\sum_{i=j_r+1}^{j_{r+1}} q_i),$$

for $h_r = j_r + 1, \ldots, j_{r+1}, r = 0, 1, 2, \ldots, t$. See Lee (1987) and Oluyede (1993, 2009) for details.

Note that if $x_i = 0$, for some i, place a small weight $\delta_m(w) < 1$, which may depend on m and w in each empty cell and compute estimate as in the case $x_i > 0$, so that p_i^* is the limit of these estimates as $\delta_m(w) \to 0$, (Lee (1987)). These weights may be chosen such that $(p_1^*, p_2^*, ..., p_k^*, q_1^*, q_2^*, ..., q_k^*)$ is strongly consistent for $P \overset{st}{\prec} Q$.

5.5 Test Procedures

5.5.1 One-Sample Procedure

We now construct a test for

$$H_0 : P = Q \text{ against } H_A : P \overset{lpr}{\prec} Q,$$

where $Q = (q_1, q_2, ..., q_k)$ is a known standard. (5.15)

To construct a test for (5.17), (see Oluyede (1994) for details), note that under $H_0 : P = Q$, $V \xrightarrow{d} N(0, \Gamma)$ as $m \to \infty$, where $\Gamma = B\Sigma B^T = (q_i(\delta_{ij} - q_j)/\sqrt{q_i q_j})$, and $\Sigma = (\sigma_{ij}) = q_i(\delta_{ij} - q_j)$, and $\xrightarrow{d}$ denotes convergence in distribution. The vector $V = UB$, where $U = \sqrt{m}(x_1/m - q_1, x_2/m - q_2, ..., x_k/m - q_k)^T$, $B = (1/\sqrt{q_i})\delta_{ij}$, and $\delta_{ij} = 1$, if $i = j$ and 0 otherwise. Let $H = (h_{ij})$ be an orthogonal matrix, then $S = HV \xrightarrow{d} N(0, H\Gamma H^T)$, where

$$S_j = \frac{m[q_{j+1}(\sum_{i=1}^{j} x_i) - x_{j+1}(\sum_{i=1}^{j} q_i)]}{\sqrt{mq_{j+1}(m\sum_{i=1}^{j} q_i)(m\sum_{i=1}^{j+1} q_i)}}, j = 1, 2, ..., k-1, \tag{5.16}$$

are asymptotically normally distributed and mutually independent with mean 0 and unit variance.

Let $S = (S_1 I_{(0,\infty)}(S_1), S_2 I_{(0,\infty)}(S_2), ..., S_{k-1} I_{(0,\infty)}(S_{k-1}))^T$ be a vector of the statistics, where S_j is given above. For testing the hypothesis (5.17), the following test statistic is proposed

$$Y_m^2 = S^T \Psi^{-1} S, \tag{5.17}$$

where Ψ is the covariance matrix of the random vector S, and is independent of the parameters. We reject H_0 for large values of Y_m^2.

Alternatively, for testing the hypothesis (5.17), reject H_0 for large values of

$$W = \sum_{j=1}^{k-1} S_j I_{(0,\infty)}(S_j) = \sum_{j=1}^{k-1} W_j. \tag{5.18}$$

Another related test statistic of interest is given by

$$T = \sum_{j=1}^{k-1} S_j^2 I_{(0,\infty)}(S_j), \tag{5.19}$$

where S_j is given by (5.18).

These test statistics both characterizes the null and alternative hypotheses. They are small when H_0 is true and large under H_A.

Note that since each S_j, $j = 1, 2, ..., k-1$, has a standard normal distribution, asymptotically under H_A the statistics

$$S = (S_1 I_{(0,\infty)}(S_1), S_2 I_{(0,\infty)}(S_2), ..., S_{k-1} I_{(0,\infty)}(S_{k-1}))^T, \tag{5.20}$$

has in the limit as $m \to \infty$ the $k-1$ dimensional normal distribution, $N_{k-1}(\lambda, \Psi)$, and hence the statistic $Y_m^2 = S^T \Psi^{-1} S$, where $\Psi = diag((1 - \frac{2}{\pi}))\delta_{ij}$, $\delta_{ij} = I_{(i=j)}$, has as $m \to \infty$, the non-central chi-square distribution with $k-1$ degrees of freedom and non-centrality parameter $\nu = \sum_{i=1}^{k-1} \lambda_i$, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_{k-1})^T$, and rank $(\Psi) = k-1$. Under H_0, $\lambda_i = 0$, $j = 1, 2, .., k-1$ and hence Y_m^2 has in the limit as $m \to \infty$ the chi-square distribution with $k-1$ degrees of freedom.

For the statistic W, note that since each S_j, $j = 1, 2, ..., k-1$, is normally distributed with mean 0 and unit variance, we see that each $W_j = S_j I_{(0,\infty)}(S_j)$ is normally distributed with mean $\sqrt{(2/\pi)}$ and variance $1 - 2/\pi$, and are independent. Consequently, the statistic W is normally distributed with mean $(k-1)\sqrt{(2/\pi)}$ and variance $\sum_{i=1}^{k-1} Var(W_j) = (k-1)(1-2/\pi)$, and covariance $Cov(S_j, S_j) = 0$.

As for the statistic $T = \sum_{j=1}^{k-1} S_j^2 I_{(0,\infty)}(S_j)$, note that under the null hypothesis, the asymptotic distribution of this proposed test statistic T is given by

$$\lim_{m\to\infty} P_{H_0}(T > C) = \begin{cases} 1 & \text{if } C \leq 0, \\ P(\sum_{j=1}^{k-1} Z_j^2 I_{(0,\infty)}(Z_j) > C) & \text{if } C > 0, \end{cases}$$

where Z_j, $j = 1, 2, ..., k-1$, are independent and identically distributed with a standard normal distribution. The asymptotic null distribution is given by Robertson, Wright and Dykstra (1988).

An α-level test based on T will reject H_0 if $T(x_1, x_2, ..., x_k) \geq C$, where $(x_1, x_2, ..., x_k)$ is the observed value of $X = (X_1, X_2, ..., X_k)$, and C is defined by

$$\lim_{m\to\infty} P_{H_0}(T \geq C) = \sum_{j=1}^{k-1} \binom{k-1}{j} (\frac{1}{2})^{k-1} P(\chi^2_{(j)} \geq C) = \alpha,$$

and $\chi^2_{(j)}$ denotes a random variable having a chi-square distribution with j degrees of freedom.

5.5.2 Two-Sample Procedure

For the two-sample procedure, the hypothesis $H_0 : P = Q$ against $H_A : P \overset{lpr}{\prec} Q$ can be reduced to the following:

$$H_0 : P = Q \text{ against } \ H_A : \bar{p}_j \geq \bar{q}_j, j = 2, 3, ..., k,$$

and $\bar{p}_j > \bar{q}_j$ for at least one j, where $\bar{p}_j = \sum_{i=1}^{j-1} p_i / \sum_{i=1}^{j} p_i$ and $\bar{q}_j = \sum_{i=1}^{j-1} q_i / \sum_{i=1}^{j} q_i$.

For testing the hypothesis $H_0 : P = Q$ against $H_A : P \overset{lpr}{\prec} Q$, the following test statistic is proposed

$$Y_{m,n}^{*2} = S^{*T} \Theta^{-1} S^*, \tag{5.21}$$

where Θ is the covariance matrix of the random vector

$$S^* = (S_1^* I_{(0,\infty)}(S_1^*), S_2^* I_{(0,\infty)}(S_2^*), ..., S_{k-1}^* I_{(0,\infty)}(S_{k-1}^*))^T.$$

The statistic S_j^* are given by

$$S_j^{*2} = \frac{(d_{j+1} b_{j+1})^{-2} [y_{j+1} (\sum_{i=1}^{j} x_i) - x_{j+1} (\sum_{i=1}^{j} y_i)]^2}{\bar{q}_{j+1}^* (1 - \bar{q}_{j+1}^*)(d_{j+1} + b_{j+1}) / d_{j+1} b_{j+1}}, \qquad j = 1, 2, ..., k-1, \tag{5.22}$$

and $\bar{q}_j^*$ is given by $\sum_{i=1}^{j-1}(x_i + y_i) / \sum_{i=1}^{j}(x_i + y_i)$. We have replaced $\bar{q}_j$ by $\bar{q}_j^*$ due to the fact that $\bar{q}_j$ is unknown, and under H_0, $p_j = q_j$ can be approximated by $(x_j + y_j)/(m + n)$, $j = 1, 2, ..., k$. The justification for the approximation process of $\bar{q}_j$ is given by the Theorem below.

Theorem 5.5.1. *If $\bar{p}_j \geq \bar{q}_j$, $(\bar{p}_j^*, \bar{q}_j^*)$ converges in probability to $(\bar{p}_j, \bar{q}_j)$, $j = 2, 3, ..., k$ as $m, n \to \infty$.*

Proof : If $\bar{p}_j = \bar{q}_j$, then for $\epsilon > 0$,

$$\begin{aligned} P(||(\bar{p}_j^*, \bar{q}_j^*) &- (\bar{p}_j, \bar{q}_j)|| > \epsilon) \\ &\leq P(|\bar{p}_j^* - \bar{p}_j| + |\bar{q}_j^* - \bar{q}_j| > \epsilon) \\ &\leq P(|\bar{p}_j^* - \bar{p}_j| > \epsilon/2) + P(|\bar{q}_j^* - \bar{q}_j| > \epsilon/2) \qquad (5.23) \\ &+ 2P(|(\bar{p}_j^* + \bar{q}^*)/(m+n) - \bar{q}_j| > \epsilon/2) \\ &< 4\bar{q}_j(1-\bar{q}_j)[1/m\epsilon^2 + 1/n\epsilon^2 + 1/(m+n)\epsilon^2]; j = 2, 3, ..., k \end{aligned}$$

by Chebychev's inequality, where $||Z|| = (\sum_{i=1}^{k} Z_i^2)^{1/2}$.

If $\bar{p}_j > \bar{q}_j$, then for $\epsilon > 0$,

$$\begin{aligned} P(||(\bar{p}_j^*, \bar{q}_j^*) &- (\bar{p}_j, \bar{q}_j)|| > \epsilon) \\ &\leq P(|\bar{p}_j^* - \bar{p}_j| + |\bar{q}_j^* - \bar{q}_j| > \epsilon) \qquad (5.24) \\ &\leq P(|\bar{p}_j^* - \bar{p}_j| > \epsilon/2) + P(|\bar{q}_j^* - \bar{q}| > \epsilon/2) \\ &= P(|\bar{p}_j^* - \bar{p}_j| > \epsilon/2, \bar{p}_j^* \geq \bar{q}_j^*) + P(|\bar{p}_j^* - \bar{p}_j| > \epsilon/2, \bar{p}_j^* < \bar{q}_j^*) \\ &+ P(|\bar{q}_j^* - \bar{q}_j| > \epsilon/2, \bar{p}_j^* \geq \bar{q}_j^*) + P(|\bar{q}_j^* - \bar{q}_j| > \epsilon/2, \bar{p}_j^* < \bar{q}_j^*) \\ &\leq P(|\bar{p}_j^* - \bar{p}_j| > \epsilon/2) + P(|\bar{q}_j^* - \bar{q}_j| > \epsilon/2) \\ &+ 2P(|(\bar{p}_j^* - \bar{q}_j^*) - (\bar{p}_j - \bar{q}_j)| > (\bar{p}_j - \bar{q}_j)) \\ &\leq 4\bar{p}_j(1-\bar{p}_j)/m\epsilon^2 + 4\bar{q}_j(1-\bar{q}_j)/n\epsilon^2 \\ &+ 2[\bar{p}_j(1-\bar{p}_j)/m + \bar{q}_j(1-\bar{q}_j)/n]/(\bar{p}_j - \bar{q}_j)^2, \end{aligned}$$

$j = 2, 3, ..., k$ by Chebychev's inequality. Therefore if $\bar{p}_j \geq \bar{q}_j$, $(\bar{p}_j^*, \bar{q}_j^*)$ converges in probability to $(\bar{p}_j, \bar{q}_j)$ whenever m, $n \to \infty$ and the proof is complete. □

Alternatively, for testing the hypothesis $H_0 : P = Q$ against $H_A : P \overset{lpr}{\prec} Q$, reject H_0 for large values of

$$W^* = \sum_{j=1}^{k-1} S_j^* I_{(0,\infty)}(S_j^*). \tag{5.25}$$

Another related test statistic is given by

$$T^* = \sum_{j=1}^{k-1} S_j^{*2} I_{(0,\infty)}(S_j^*). \tag{5.26}$$

These test statistics both characterizes the null and alternative hypotheses. They are small when H_0 is true and large under H_A.

5.6 Tests For and Against Stochastic Ordering

5.6.1 Test Statistics and Procedures

Consider testing $H_0 : P = Q$ against $H_1 - H_0$, where $H_1 : P \overset{st}{\prec} Q$ and $Q = (q_1, q_2, .., q_k)$ is known. The log-likelihood ratio test rejects H_0 for large values of

$$T_{01} = -2ln\lambda_{01} = -2m \sum_{i=1}^{k} \hat{p}_i (lnq_i - lnp_i^*). \tag{5.27}$$

For the two-sample test, the log-likelihood ratio test of H_0 against $H_1 - H_0$ rejects H_0 for large values of

$$T_{02} = 2m \sum_{i=1}^{k} \hat{p}_i (lnp_i^* - ln\bar{p}_i) + 2n \sum_{i=1}^{k} \hat{q}_i (lnq_i^* - ln\bar{q}_i),$$

$$\text{where } \bar{p}_i = \bar{q}_i = (m\hat{p}_i + n\hat{q}_i)/(m+n). \tag{5.28}$$

The corresponding chi-square analogue (one-sample) test rejects H_0 for large values of

$$X_{01}^2 = m \sum_{i=1}^{k} [(p_i^* - q_i)^2 / q_i] \tag{5.29}$$

and the two-sample test rejects H_0 for large values of

$$X_{02}^2 = m \sum_{i=1}^{k} (p_i^* - \bar{p}_i)^2 / \bar{p}_i + n \sum_{i=1}^{k} (q_i^* - \bar{q}_i)^2 / \bar{q}_i. \tag{5.30}$$

The log-likelihood ratio test and the corresponding chi-square analogues for both the one- and two-sample cases for testing $H_1 : P \overset{st}{\prec} Q$ against $H_2 :\sim H_1$ are given below.

Reject H_1 for large values of

$$T_{12} = -2ln\lambda_{12} = 2m \sum_{i=1}^{k} \hat{p}_i (ln\hat{p}_i - lnp_i^*) \tag{5.31}$$

in the one-sample problem. For the two-sample problem reject H_1 for large values of

$$T_{22} = -2ln\lambda_{22} = 2m\sum_{i=1}^{k}\hat{p}_i(ln\hat{p}_i - lnp_i^*) + 2n\sum_{i=1}^{k}\hat{q}_i(ln\hat{q}_i - lnq_i^*). \tag{5.32}$$

Reject H_1 for large values of

$$X_{12}^2 = m\sum_{i=1}^{k}[(p_i^* - \hat{p}_i)^2/p_i^*] \tag{5.33}$$

for one-sample and for the two-sample test reject H_1 for large values of

$$X_{22}^2 = m\sum_{i=1}^{k}[(p_i^* - \hat{p}_i)^2/p_i^*] + n\sum_{i=1}^{k}[(q_i^* - \hat{q}_i)^2/q_i^*]. \tag{5.34}$$

The asymptotic distributions under the corresponding null hypothesis H_0 for T_{01} and (H_1 for T_{12}) are given by Robertson, Wright and Dykstra (1988).

5.6.2 Test Restricted to Stochastic Order

In this chapter, we established a test procedure for testing $H_0 : P = Q$ against $H_A : P \overset{lpr}{\prec} Q$. However, since we know that if $P \overset{lpr}{\prec} Q$ then $P \overset{st}{\prec} Q$, we can test the hypothesis $H_0 : P = Q$ against $H_1 : P \overset{st}{\prec} Q$ by using the test statistic developed in section 5. In the one-sample problem, for testing $H_0 : P = Q$ vs. $H_1 : P \overset{st}{\prec} Q$, reject H_0 at α-level if and only if $T \geq C$, where T is given by (5.21) and C is defined by $P_{H_0}(S \geq C) = \alpha$. For the two-sample procedure, one rejects H_0 if $T^* \geq C$, where T^* is given by (5.28) and C_1, is such that $P(T^* \geq C_1) = \alpha$. It can be verified that the test based on T, T^*, has the tendency to reject the null hypothesis more often than the test based on T_{01}, T_{02} or the chi-square analogues.

5.7 Examples

This section contains two examples which illustrate the estimation and testing procedures developed in earlier sections. The first set of data (Devore and Peck 1986, page 636) given in Table 1, compares opinion of smokers and nonsmokers on an anti smoking advertisement. The empirical distributions for smokers and nonsmokers are not stochastically ordered. Comparison in this case can be done with respect to cumulative probability ratio ordering. Also, the maximum likelihood estimates of P and Q restricted to $P \overset{st}{\prec} Q$ and the values of the test statistics T_{22} and T^* are computed.

The second set of data (Agresti 1984, page 30) compares changes in size of ulcer crater under two treatments, A and B. The empirical distribution of crater size under treatment A is stochastically less than the empirical distribution of crater size under treatment B. To compare the effectiveness of the two treatments, the hypothesis of homogeneity against local probability ratio ordering is tested.

Table 5.1 Opinion of Anti smoking AD

	Opinions					
	Strongly Dislike	Dislike	Neutral	Like	Strongly Like	Total
Smoker	8	14	35	21	19	97
Nonsmoker	31	42	78	61	69	281

The MLE of $(p_1, p_2, p_3, p_4, p_5, q_1, q_2, q_3, q_4, q_5)$ restricted to $P \overset{st}{\prec} Q$ is obtained as follows: Compute $\hat{F}_j$ and $\hat{G}_j$, $j = 1, 2, 3, 4, 5$. Since $\hat{F}_1 < \hat{G}_1$ and $\hat{F}_2 < \hat{G}_2$, the MLE of $(p_1, p_2, p_3, p_4, p_5, q_1, q_2, q_3, q_4, q_5)$ is given below:

j	$\hat{F}_j$	$\hat{G}_j$	p_j^*		q_j^*	
1	$\frac{8}{97}$	$\frac{31}{281}$	$\frac{39}{378}$	$= 0.103$	$(\frac{39}{378})$	$= 0.103$
2	$\frac{22}{97}$	$\frac{73}{281}$	$\frac{56}{378}$	$= 0.148$	$(\frac{56}{378})$	$= 0.148$
3	$\frac{57}{97}$	$\frac{151}{281}$	$(\frac{35}{75})(\frac{283}{378})$	$= 0.349$	$(\frac{78}{208})(\frac{283}{378})$	$= 0.281$
4	$\frac{78}{97}$	$\frac{212}{281}$	$(\frac{21}{75})(\frac{283}{378})$	$= 0.210$	$(\frac{61}{208})(\frac{283}{378})$	$= 0.221$
5	1	1	$(\frac{19}{75})(\frac{283}{378})$	$= 0.190$	$(\frac{69}{208})(\frac{283}{378})$	$= 0.248$

The loglikelihood ratio statistic (see Dykstra et al (1988)for details) for testing

$$H_0 : P = Q \quad \text{against} \quad H_A : P \overset{st}{\prec} Q,$$

is given by

$$T_{22} = 2m \sum_{i=1}^{k} \hat{p}_i (ln\hat{p}_i - lnp_i^*) + 2n \sum_{i=1}^{k} \hat{q}_i (ln\hat{q}_i - lnq_i^*). \tag{5.35}$$

To compare smokers' and nonsmokers' opinion on the anti smoking advertisement, we demonstrate the tests for both cumulative probability ratio and stochastic orderings by testing the hypothesis of homogeneity against the alternative that the two multinomial distributions are local probability ratio ordered. The value of the test statistic T_{22} and T^* are given by

$$\begin{aligned} T_{22} &= 2m \sum_{i=1}^{k} \hat{p}_i (ln\hat{p}_i - lnp_i^*) + 2n \sum_{i=1}^{k} \hat{q}_i (ln\hat{q}_i - lnq_i^*) \\ &= 0.1970, \end{aligned} \tag{5.36}$$

and

$$\begin{aligned} T^* &= \sum_{j=1}^{k-1} S_j^{*2} I_{(0,\infty)}(S_j^*) \\ &= 0 + 0 + .0963 + .9963 \\ &= 1.0926. \end{aligned} \tag{5.37}$$

These values do not lead to the rejection of the null hypothesis.

Table 5.2 Change in Size of Ulcer Crater

		$< 2/3$	$\geq 2/3$		
Treatment	Larger	Healed	Healed	Healed	Total
A	12	10	4	6	32
B	5	8	8	11	32

The MLE of $(p_1, p_2, p_3, p_4, q_1, q_2, q_3, q_4)$ restricted $P \overset{st}{\prec} Q$ is given below.

j	$\hat{F}_j$	$\hat{G}_j$	p_j^*	q_j^*
1	$\frac{12}{32}$	$\frac{5}{32}$	0.3750	0.15625
2	$\frac{22}{32}$	$\frac{13}{32}$	0.3125	0.25000
3	$\frac{26}{32}$	$\frac{21}{32}$	0.1250	0.25000
4	1	1	0.1875	0.34375

The value of the test statistic T^* is given by

$$\begin{aligned} T^* &= \sum_{j=1}^{k-1} S_j^{*2} I_{(0,\infty)}(S_j^*) \\ &= 0.8463 + 3.152 + 2.003 \\ &= 6.0013. \end{aligned}$$

This value corresponds to attained significance level of between 0.025 and 0.05 which leads to the rejection of the null hypothesis. Clearly, this is an indication that treatment A is superior. Note that Pearson's chi-square statistic $X^2 = 5.34$ with a p-value greater than 0.10 does not lead to the rejection of the null hypothesis.

5.8 Concluding Remark

Restricted parameter estimates of the multinomial populations including those under local stochastic domination (local probability ratio order) and stochastic order are presented. We have also developed and presented procedures for comparing multinomial populations under these orderings. In particular, procedures for tests restricted to and against stochastic ordering are presented. Alternative procedures in which the test statistics have in the limit the chi-square distribution with k degrees of freedom are developed and presented. Numerical examples to illustrate the developed techniques were developed presented.

BIBLIOGRAPHY

[1] Agresti, A. and Min, Y., Unconditional small-sample confidence intervals for odds ratios, Biostatistics, 3(3)(2002), 379–386.

[2] Fletcher, R.H., Fletcher, S.W., Wagner, E.H., Clinical epidemiology: The essentials, Lippincott Williams and Wilkins, 1996.

[3] Agresti, A. (2002). Categorical Data Analysis. Second Edition, Wiley series, New York.

[4] Agresti, A. (1984). Analysis of Ordinal Categorical Data, New York: John Wiley.

[5] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*, New York: Wiley and Sons.

[6] Chacko, V. J. (1966). Modified Chi-square Test for Ordered Alternatives. *Sankhya*, Series B 28, 185-190.

[7] Chuang-Stein, C. and Mohberg, N. R. (1993). A Unified Approach to the Analysis of Safety Data in Clinical Trials. In *Drug Safety Assessment in Clinical Trials*, Sogliero - Gilbert, G. Editor, Marcel Dekker, New York.

[8] Devore, J. and Peck, R. (1986). Statistics, the Exploration and Analysis of Data. St. Paul, MN: West.

[9] Dykstra, R. L., Kochar, S., and Robertson, T. (1988). Statistical Inference for Uniform Stochastic Ordering in Several Populations. *Annals of Statistics*, 19, 870-888.

[10] Gupta, S. S., Panchapakesan, S., and John, J. K., T. (1985). On the Distribution of the Studentized Maximun of Equally Correlated Normal Random Variables. *Communications in Statistics, Simulation and Computation*, 14(1), 103-135.

[11] Hirji, K. F.,(2006). Exact Analysis of Discrete Data. London, Chapman and Hall/CRC.

[12] Lee, C. I. C., (1987) Maximum Likelihood Estimation for Stochastically Ordered Multinomial Populations with Fixed and Random Zeros. In *Proc. Symp. in Statist. and Festschrift in Honor of V. M. Joshi*, 173-189.

[13] Mehta, C. R., Patel, N. R., and Senchaudhuri, P. (1988). Importance Sampling for Estimating Exact Probabilities in Permutation Inference. *Journal of American Statistical Association*, 83, 404, 999-1005.

[14] Mehta, C. R., Patel, N. R., and Tsiatis, A. A. (1984). Exact Significance Testing to Establish Treatment Equivalence with Ordered Categorical Data. *Biometrics*, 40, 819-825.

[15] Oluyede, B.O., A modified chi-square test for testing equality or two multinomial populations against an order restricted alternative, *Communications in Statistics-Theory and Methods*, Vol. 22, No. 4 (1993) pp. 1133-1155.

[16] Oluyede, B.O., Estimation and testing of multinomial parameters under order restrictions, Proceedings of the ASA Section on Quality and Productivity, (1994) pp. 276-280.

[17] Oluyede, B.O., (2009). Constrained Analysis of Probability Ratios in Multinomial Populations, Submitted.

[18] Rahlfs, V. W. and Zimmerman, H. (1993). Scores: Ordinal Data with Few Categories, How Should They Be Analyzed? *DIA Journal*, 27, 1227-1240.

[19] Robertson, T. and Wright, F. T. (1981). Likelihood Ratio Tests for and Against a Stochastic Ordering Between Multinomial Populations. *Annals of Statistics*, 9, 1248-1257.

[20] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). Order Restricted Statistical Inference, New York: Wiley.

[21] Silvapulle, J. S., and Sen, P. K. (2005). Constrained Statistical Inference. *Inequality, Order and Shape Restrictions.* Hoboken, New Jersey, John Wiley and Sons.

[22] Spilker, B., and Schoenfelder, J. (1991). Data Collection Forms in Clinical Trials, New York: Raven Press.