Nonparametric Tests of Trend for Proportions

Mayer Alvo

Department of Mathematics and Statistics, University of Ottawa

585 King Edward, Ottawa K1N 6N5, Canada

Tel: 1-613-562-5800 ext. 3534 E-mail: malvo@uottawa.ca

Marie-Pascal Berthelot

Department of Mathematics and Statistics, University of Ottawa

585 King Edward, Ottawa K1N 6N5, Canada

E-mail: mbert074@uottawa.ca

Received: February 16, 2012 Accepted: February 29, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p92 URL: http://dx.doi.org/10.5539/ijsp.v1n1p92

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

Abstract

A general method is proposed for constructing nonparametric tests of trend for proportions. Such alternatives arise in situations where it is of interest to test for monotonicity in rates of growth. The class of tests is based on the ranks of the observations. The general approach consists of defining two sets of rankings: the first describes the time and the other the binary data itself. The test statistics measures the similarity between the two sets. The asymptotic null distributions are determined for similarity measures due to Spearman, Kendall and Hamming. A limited simulation study shows that the Spearman test has greater power.

Keywords: Tests of trend for proportions, Ranks, Spearman, Kendall, Hamming similarity measure, Asymptotic Distributions

1. Introduction

There are several instances in practice when one is interested in testing for a trend in proportions. For instance, one may be interested in the trend in birth rates, in mortality rates or in the incidence of a certain disease. As an example, we will consider the mortality statistics in South Africa during the period 2000-2008 in Table 1. One may ask if there is an increasing trend in the mortality rates. We refer the reader for other examples of applications to Chen et al. (1997), to Arase et al. (2001) for controlled clinical trials and to Ku et al. (2001) for community based surveys.

In such problems, one usually observes at time values $t_1 < t_2 < ... < t_k$ a random sample of n_i binary variables $\left\{y_{ij}\right\}$ with y_{ij} taking values 1 or 0 with unknown probabilities p_i and $1 - p_i$ respectively. The observed data may be viewed as :

	t_1	t_2	$\cdots$	t_k	$Total$
	y_1	y_2	$\cdots$	y_k	y
	$n_1 - y_1$	$n_2 - y_2$	$\cdots$	$n_k - y_k$	$n - y$
$Total$	n_1	n_2	$\cdots$	n_k	n

where $y_i = \sum_j y_{ij}$, $y = \sum_i y_i$. Let $\bar{p}_i = y_i/n_i$ and $\bar{p} = y/n$. Without loss in generality, we shall be interested in detecting a monotone increasing trend in p_i. There are a number of different approaches to the problem as discussed in Terpestra (1952) and Armitage (1955). One approach founded on regression and analysis of variance and discussed below leads to the Cochran-Armitage test (Cochran 1954; Armitage 1955). Williams (1988) and Chen et al. (1997) have noted that when the expected values $n_i p_i$, or $n_i (1 - p_i)$ are small, the normal approximation may become unreliable. As a consequence the Cochran-Armitage test becomes conservative and may lead to a type I error rate greater than to the prescribed significance level Portier and Hoel (1984), Hothorn and Bretz (2000). Corcoran et al. (2000) have noted that the Cochran-Armitage test is very sensitive to the choice of scores and conclude that this "makes its general use suspect". Williams (1988) has proposed an exact conditional test whereby the test statistic is calculated for all possible 2 x k tables with identical marginal totals. This is the randomization model in which the groups or treatments have the experimental units assigned to them at random. Randomization models are predominantly chosen in biomedical research. Neuhäuser (2006) proposed

a modification of the Baumgartner-Weiβ-Schindler (1998) two-sample statistic which utilizes ranks instead of scores. A simulation study is inconclusive and does not reveal a clear winner. In section 2, we consider the problem using ranking methods.

Let $\{x_i\}$ represent arbitrary pre-selected scores, $x_1 < x_2 < ... < x_k$ which mimic a monotone increasing time trend. The linear model regression of y_{ij} expressed by

$$p_i = \alpha + \beta (x_i - \bar{x}_k) \tag{1}$$

and subject to the constraint that $\sum_i n_i (x_i - \bar{x}_k) = 0$ where $\bar{x}_k = \sum n_i x_i / n$, yields estimates

$$\hat{\alpha} = \frac{\sum_i n_i \bar{p}_i}{\sum_i n_i} = \bar{p}, \ \ \hat{\beta} = \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sum_i n_i (x_i - \bar{x}_k)^2}.$$

The hypothesis of homogeneity is

$$H_0 : p_i = p, i = 1, ..k$$

(or equivalently $\beta = 0$). Possible alternatives are

$$\begin{aligned} H_1 &: \ p_1 \le p_2 \le ... \le p_k . \text{with at least one strict inequality} \\ H_2 &: \ p_i \ne p_j, \text{ for at least one pair } (i, j) \end{aligned}$$

Hypothesis H_2 can also be expressed as

$$\sum_{i<j} \left(p_i - p_j\right) = \sum_i (k + 1 - 2i) \, p_i \ne 0 \tag{2}$$

(Van Eeden & Hemelrijk, 1955).

Under the null hypothesis of homogeneity, the estimate of variance of $\hat{\beta}$ is given by

$$V\left(\hat{\beta}\right) = \frac{\bar{p}(1 - \bar{p})}{\sum_i n_i (x_i - \bar{x}_k)^2}$$

and consequently we reject H_0 in favor of H_1 (or equivalently $\beta > 0$) for large values of the statistic

$$\hat{\beta} / \sqrt{V\left(\hat{\beta}\right)} = \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\bar{p}(1 - \bar{p})} \sqrt{\sum_i n_i (x_i - \bar{x}_k)^2}}$$

which for large samples has a standard normal. If we suppose further that $\bar{p}$ is small (of the order of $1\% - 2\%$) so that $\bar{p}^2$ is negligible, then the statistic becomes

$$\frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\sum_i n_i \bar{p} (x_i - \bar{x}_k)^2}}.$$

The difference between observed and expected frequencies $n_i (\bar{p}_i - \bar{p})$ are multiplied by the score effect $(x_i - \bar{x}_k)$ in the numerator whereas the square of the score effects are weighted by the expected frequencies in the denominator. The statistic provides a test of trend in frequencies as opposed to a test of trend on the proportions.

Since the sample correlation between $\{\bar{p}_i - \bar{p}\}$ and $\{x_i - \bar{x}_k\}$ is given by

$$\begin{aligned} & \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\sum_i n_i (\bar{p}_i - \bar{p})^2 \sum_i n_i (x_i - \bar{x}_k)^2}} \\ = \ & \hat{\beta} \sqrt{\frac{\sum_i n_i (x_i - \bar{x}_k)^2}{\sum_i n_i (\bar{p}_i - \bar{p})^2}} \end{aligned}$$

we may view the test of monotonicity as equivalent to a test that the correlation is 0.

Alternatively, the test of homogeneity may be conducted by treating the data as coming from a 2 x k contingency table. That tests rejects H_0 in favor of H_2 for large values of the statistic

$$\begin{aligned} & \sum_i \frac{(n_i \bar{p}_i - n_i \bar{p})^2}{n_i \bar{p}} + \sum_i \frac{(n_i (1 - \bar{p}_i) - n_i (1 - \bar{p}))^2}{n_i (1 - \bar{p})} \\ = \ & \frac{\sum_i n_i (\bar{p}_i - \bar{p})^2}{\bar{p}(1 - \bar{p})} \end{aligned}$$

which for large samples $n_i \to \infty$, has a chi square distribution with $(k-1)$ degrees of freedom. In that case there is no need to define scores. We note that even though H_1 is included in H_2, "one-sided" tests which focus strictly on H_1 will in general be more powerful.

The regression model may be extended to apply to two or more groups of individuals. For example, to model the birth rates of men and women in the population, the functional regression model becomes

$$p_i = \alpha + \beta (x_i - \bar{x}_k) + \gamma (\alpha' + \beta' (x_i - \bar{x}_k)) \tag{3}$$

where γ takes value 1 for the first group and 0 otherwise.

In section two, we propose a general approach for constructing nonparametric test statistics based on the ranks of the observations to test H_0 against H_2. We obtain the asymptotic null distributions for these test statistics whenever either the sample sizes gets large or the number of time points k gets large. In section three, we report on the results of some simulation studies and note that the test statistics perform well under a variety of different underlying patterns. We present an application and conclude with some final remarks.

2. The Construction of the Test Statistics

In the previous section we considered the problem of testing for trend in proportions when scores provide a proxy for time. In this section we consider the problem from an entirely new perspective using methods based on the ranks of the data. The approach by-passes the use of scores. We first provide an introduction to statistical methods based on ranks.

A complete ranking of n objects is a permutation of the integers $(1, ..., n)$. For any two rankings $\mu = (\mu(1), ..., \mu(n))'$, $\nu = (\nu(1), ..., \nu(n))'$, we may define the following measures of similarities due to Spearman, Kendall and Hamming respectively:

$$\begin{aligned}
\mathcal{A}_S(\mu, \nu) &= \sum_{i=1}^{n} \left(\mu(i) - \frac{n+1}{2}\right)\left(\nu(i) - \frac{n+1}{2}\right) \\
\mathcal{A}_K(\mu, \nu) &= \sum_{i<j} sgn\,(\mu(j) - \mu(i))\, sgn\,(\nu(j) - \nu(i)) \\
\mathcal{A}_H(\mu, \nu) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(I\,[\mu(i) = j] - \frac{1}{n}\right)\left(I\,[\nu(i) = j] - \frac{1}{n}\right)
\end{aligned}$$

where $sgn(x)$ is either 1 or -1 depending on whether $x > 0$ or $x < 0$ and where $I[\cdot]$ is the indicator function which is 1 or 0 depending on whether the statement in brackets holds or not. Measures of similarity can be used to define rank correlations to provide tests of trend and of independence. A review of related results may be found in Alvo and Cabilio (1992).

In what follows, we make use of the notion of tie compatibility (see Alvo & Cabilio, 1999) to extend the measures of similarity defined above to the case where ties occur in the data.

Definition 1. A tied ordering of n objects is a partition into e sets, $1 \le e \le n$, each of which contains d_i objects, $d_1 + d_2 + ... + d_e = n$, so that the d_i objects in each set share the rank i, $1 \le i \le e$. Such a tie pattern is denoted by $\delta = (d_1, d_2, ..., d_e)$ .The ranking denoted by $\nu_\delta = (\nu_\delta(1), \nu_\delta(2), ..., \nu_\delta(n))$, resulting from such an ordering, is a tied ranking, and is one of $n!/(d_1!d_2!...d_e!)$ possible permutations.

Example 1. Suppose that $n = 7$ objects have the ranking (3172465), that is object 1 is ranked 3, object 2 is ranked 1, object 3 is ranked 7, and so on. Suppose that ties are allowed and that the ordering assumes the form $\langle(24)(157)(6)(3)\rangle$, where now the parentheses indicate members of the same tie class, so that objects 2 and 4 receive rank 1, objects 1, 5, and 7 receive rank 2, and objects 6 and 3 receive ranks 3 and 4 respectively. The tied ranking then becomes (2141232). In this case $e = 4, d_1 = 2, d_2 = 3, d_3 = d_4 = 1$.

The ranking which describes the time points may be viewed as a tied ranking with tie pattern

$$\delta_1 = (n_1, n_2, ..., n_k) \tag{4}$$

and with ordering

$$\left\langle (1, ..., n_1)(n_1 + 1, ..., n_1 + n_2) ... \left(\sum_{i=1}^{k-1} n_i + 1, ..., \sum_{i=1}^{k} n_i\right)\right\rangle \tag{5}$$

On the other hand, the binary variables y_{ij} have with $e = 2$ the simple tie pattern

$$\delta_2 = (y, n - y) \tag{6}$$

We now define tie compatibility whereby a tie ranking could be conceived as having arisen from a complete ranking in which some objects are grouped as being of equivalent standing. A re-ranking will then produce the tied ranking. All complete rankings which could give rise in this way to the specified tied ranking are then said to be compatible to it. More precisely, we have the following.

Definition 2. A complete ranking of n objects is compatible with a tied ranking of these objects with tie pattern $\delta = (d_1, d_2, ..., d_e)$, if every pair of objects which receive distinct ranks is given the same relative ranking in both rankings. We shall denote by $C(\mu)$ and $C(\nu)$ the class of complete rankings compatible to μ, ν respectively.

Example 2. Suppose that $k = 2, n_1 = 3, n_2 = 2$, and we observe 2 successes 1 failure at time t_1, and 1 success 1 failure at time t_2. The compatibility class corresponding to the time ranking ν contains the 12 permutations obtained by permuting ranks $1, 2, 3$ among themselves and ranks $4, 5$ among themselves

$$C(\nu) = \left\{ \begin{array}{c} (123|45), (132|45), (213|45), (231|45), (321|45), (312|45), \\ (123|54), (132|54), (213|54), (231|54), (321|54), (312|54) \end{array} \right\}$$

On the other hand, there are a total of 6 patterns of rankings with (2 $successes$, 1 $failure$) at time t_1, and (1 $success$, 1 $failure$) at time t_2 as follows:

$$(341|52), (351|42), (451|32), (342|51), (352|41), (452|31).$$

Permuting the entries in blocks 1 and 2 respectively, we obtain a total of 72 compatible rankings in the class $C(\mu)$.

Definition 3. The measure of similarity for the case of tied rankings $\mu_{\delta_1}, \nu_{\delta_2}$ is defined to be the conditional expectation

$$A(\mu_{\delta_1}, \nu_{\delta_2}) = E[\mathcal{A}(\mu, \nu)|C(\mu), C(\nu)]$$

where the expectation is computed by averaging over the complete rankings compatible to $\mu_{\delta_1}, \nu_{\delta_2}$.

In the next section, we compute the statistics corresponding to tied rankings for each of the similarity measures defined above.

2.1 The Test Statistic Corresponding to Spearman Similarity

To that end we note that

$$\begin{aligned} & \mathcal{A}_S(\mu_{\delta_1}, \nu_{\delta_2}) \\ = & \sum_{i=1}^{n} E\left[\left(\mu(i) - \frac{n+1}{2}\right)|C(\mu)\right] E\left[\left(\nu(i) - \frac{n+1}{2}\right)|C(\nu)\right] \end{aligned}$$

The average of the compatible ranks at time t_1 is $g_1 = \left(\frac{n_1+1}{2}\right)$, at time t_2 it is $g_2 = n_1 + \left(\frac{n_2+1}{2}\right)$ and so on. In general at time t_l the average rank is $g_l = \sum_{i=1}^{l-1} n_i + \left(\frac{n_l+1}{2}\right), l = 1, .., k$. Clearly, $\sum_{i=1}^{k} n_i g_i = \frac{n(n+1)}{2}$. Hence, the conditional expectation

$$E[\nu(l)|C(\nu)] = g_l, l = 1, ..., n_i$$

Turning attention now to the binary observations, the average rank for the $n - y$ which take value 0 is $l_1 = \frac{n-y+1}{2}$ whereas the average rank for the y observations which take value 1 is $l_2 = n - y + \left(\frac{y+1}{2}\right) = n - \frac{y}{2} + \frac{1}{2}$ It follows that at time t_i,

$$E[\mu(l)|C(\mu)] = \left\{ \begin{array}{ll} l_1 & l = 1, ..., n_i - y_i, \\ l_2 & l = n_i - y_i + 1, ..., n_i, \end{array} \right.$$

Since $l_2 - l_1 = \frac{n}{2}$ and $\sum_{i=1}^{k} n_i g_i = \frac{n(n+1)}{2}$, we have that

$$\begin{aligned} \mathcal{A}_S(\mu_{\delta_1}, \nu_{\delta_2}) &= \sum_{i=1}^{k} \left[g_i - \frac{n+1}{2}\right] [l_2 y_i + (n_i - y_i) l_1] \\ &= \frac{n}{2} \sum_{i=1}^{k} \left[g_i - \frac{n+1}{2}\right] y_i \end{aligned}$$

We shall define the Spearman statistic to be

$$\begin{aligned} S &= \sum_{i=1}^{k}\left[g_i - \frac{n+1}{2}\right]y_i \\ &= \sum_{i=1}^{k} n_i c_i \left(\overline{p_i} - \overline{p}\right) \end{aligned} \tag{7}$$

where $c_i = g_i - \frac{n+1}{2}$ and $\sum_{i=1}^{k} n_i c_i = 0$.

2.2 The Test Statistic Corresponding to Kendall Similarity

We now consider the test statistic when using Kendall's similarity measure. Note that within a given time period, the difference between ties is zero and hence there is no contribution to the distance. Between different time periods we have

$$\begin{aligned} & \sum_{i<j}^{n} E\left[sgn\left(\mu(j)-\mu(i)\right)|C\left(\mu\right)\right] E\left[sgn\left(\nu(j)-\nu(i)\right)|C\left(\nu\right)\right] \\ = & \sum_{i<j}^{k}\left[\left(n_i - y_i\right)y_j - \left(n_j - y_j\right)y_i\right] \\ = & \sum_{i<j}^{k}\left[n_i y_j - n_j y_i\right] \end{aligned}$$

Now

$$\begin{aligned} \sum_{j=i+1}^{k} n_j &= \left[n - n_i - \sum_{j=1}^{i-1} n_j\right] \\ &= \left[n - g_i - \frac{n_i - 1}{2}\right] \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{A}_K\left(\mu_{\delta_1}, \nu_{\delta_2}\right) &= \sum_{i=1}^{k} y_i\left[g_i - \frac{n_i+1}{2}\right] - \sum_{i=1}^{k} y_i\left[n - g_i - \frac{n_i - 1}{2}\right] \\ &= 2\sum_{i=1}^{k}\left[g_i - \frac{n+1}{2}\right]y_i \end{aligned}$$

Hence, the Kendall and Spearman statistics are equivalent.

2.3 The Test Statistic Corresponding to Hamming Similarity

In this section, we consider a test statistic based on the Hamming similarity. Returning to Example 2, we may define a matrix of scores given by

$$\begin{array}{c} t_1 \\ t_1 \\ t_1 \\ t_2 \\ t_2 \end{array} \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

where rows indicate time periods and columns indicate rank. At time t_1, there are 3 observations and hence ranks $1, 2, 3$ will occur $1/3$ of the time. Similarly, at time t_2, ranks $4, 5$ occur with probability $1/2$. Entries outside the block diagonals are 0.

Hence, in general we have for time t_i, $l = \sum_{q=1}^{i-1} n_q + 1, ..., \sum_{q=1}^{i} n_q$

$$E\left[I\left(\nu(l) = j\right)|C\left(\nu\right)\right] = \left\{ \begin{array}{l} \frac{1}{n_i},\ \sum_{q=1}^{i-1} n_q + 1 \le j \le \sum_{q=1}^{i} n_q \\ \\ 0,\ \text{otherwise} \end{array} \right\}$$

We now consider the data ranking in accordance with the model. First the complete rankings corresponding to successes (and correspondingly failures) can be permuted among themselves in

$$\frac{y!}{\Pi y_i!}\frac{(n-y)!}{\Pi\,(n_i-y_i)!}$$

ways. Since entries in time blocks can be permuted in $\Pi n_i!$ ways, we have that the total number of compatible rankings is given by the product

$$\frac{y!\,(n-y)!}{\Pi y_i!\,(n_i-y_i)!}\Pi n_i!$$

Define the set of integers

$$\begin{aligned} B_i &= \left\{\sum_{q=1}^{i-1} n_q+1,...,\sum_{q=1}^{i} n_q\right\}, i=1,...,k \\ A_{0y} &= \{1,2,...,n-y\} \\ A_{1y} &= \{n-y+1,...,n\} \end{aligned}$$

and let

$$\begin{aligned} w_{0i} &= Card\ (A_{0Y}\cap B_i) \\ w_{1i} &= Card\ (A_{1Y}\cap B_i) \end{aligned}$$

where $w_{0i}=0$ if $y=n$ and $w_{1i}=0$ if $y=0$.

It follows that at time t_i and $l=\sum_{q=1}^{i-1} n_q+1,...,\sum_{q=1}^{i} n_q$,

$$\begin{aligned} E\left[I\left(\mu(l)=j\right)|C\left(\mu\right)\right] &= P\left(\mu\left(l\right)=j|C\left(\mu\right)\right) \\ &= \begin{cases} \frac{y_i}{n_i y}, & j\epsilon A_{1y} \\ \frac{(n_i-y_i)}{n_i(n-y)}, & j\epsilon Ay \end{cases} \end{aligned}$$

In fact, at time t_i, $j\epsilon A_{0y}$, we have

$$P\left(\mu\left(l\right)=j|C\left(\mu\right)\right)=\frac{\frac{y!(n-y-1)!(n_i-1)!}{\Pi y_i!\Pi_{j\neq i}\left(n_j-y_j\right)!(n_i-y_i-1)!}\Pi_{j\neq i}n_j!}{\frac{y!(n-y)!}{\Pi y_i!(n_i-y_i)!}\Pi n_i!}$$

In Example 2, we obtain a total of 72 compatible rankings. Averaging the frequency of occurrences of each ranking leads to the following matrix of scores:

$$\begin{array}{c} t_1 \\ t_1 \\ t_1 \\ t_2 \\ t_2 \end{array}\begin{pmatrix} 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/4 & 1/4 & 1/6 & 1/6 & 1/6 \\ 1/4 & 1/4 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Noting that each row at time t_i is repeated n_i times, and since $w_{0i}+w_{1i}=n_i$, the Hamming similarity measure for tied data becomes

$$\begin{aligned} &\mathcal{A}_H\left(\mu_{\delta_1},\nu_{\delta_2}\right) \\ = &\sum_{i,j}^{n}\left(E\left[I\left(\mu(i)=j\right)|C\left(\mu\right)\right]E\left[I\left(\nu(i)=j\right)|C\left(\nu\right)\right]\right)-1 \\ = &\sum_{i=1}^{k}\left[w_{1i}\frac{y_i}{n_i y}+w_{0i}\frac{(n_i-y_i)}{n_i\left(n-y\right)}\right]-1 \\ = &\sum_{i=1}^{k} w_{1i}\left[\frac{\bar{p}_i}{y}-\frac{(1-\bar{p}_i)}{(n-y)}\right] \\ = &\frac{1}{n\bar{p}\left(1-\bar{p}\right)}\sum_{i=1}^{k}\left[n_i w_i\left(\bar{p}_i-\bar{p}\right)\right] \end{aligned}$$

where $w_i = \frac{w_{1i}}{n_i}$. We may define the Hamming statistic as

$$H = \frac{1}{\bar{p}(1-\bar{p})} \sum_{i=1}^{k} \left[n_i w_i (\bar{p}_i - \bar{p}) \right] \tag{8}$$

The weights $\{w_i\}$ which depend on y may also be expressed as

$$w_i = \begin{cases} 0 & n - y + 1 > \sum_{q=1}^{i} n_q \\ \frac{1}{n_i}\left(\sum_{q=1}^{i} n_q - (n-y)\right) & \sum_{q=1}^{i-1} n_q + 1 \le n - y + 1 \le \sum_{q=1}^{i} n_q \\ 1 & n - y + 1 < \sum_{q=1}^{i-1} n_q + 1 \end{cases} \tag{9}$$

and in this form we see that most of the weight is assigned to later time points. In the next section, we determine the asymptotic null distribution of the Spearman and Hamming statistics.

3. The Asymptotic Distribution of the Test Statistics Under the Null Hypothesis

In this section, we consider the asymptotic distribution of the Spearman and Hamming test statistics under the null hypothesis. Let $\{y_i\}$ be k independent binomials (n_i, p_i) and suppose we would like to test H_0 vs H_1. In most applications the asymptotic situation of interest occurs when

$$n_i \to \infty, with \frac{n_i}{n} \to \lambda_i > 0, i = 1, ..., k. \tag{10}$$

In Theorem 1, we shall show that the Spearman statistic has an asymptotic normal distribution under either (10) or under the condition that the $\{n_i\}$ are bounded while $k \to \infty$.

Theorem 1. *Suppose that both $y \to \infty$ and $n - y \to \infty$ as $n \to \infty$. Under H_o the Spearman test statistic has asymptotically a standard normal distribution, i.e.*

$$\frac{\sum_{i=1}^{k} n_i c_i (\bar{p}_i - \bar{p})}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i p_i (1-p_i)}} \to_d N(0,1)$$

under either i) (10) or ii) the $\{n_i\}$ are bounded and $k \to \infty$.

We may estimate p_i either by $\bar{p}_i$ or by $\bar{p}$. In the first case, the test rejects whenever

$$\frac{\sum_{i=1}^{k} c_i y_i}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i \bar{p}_i (1-\bar{p}_i)}} \ge z_\alpha$$

where z_α is the upper $100(1-\alpha)\%$ percentage point from a standard normal distribution. The expression for the estimate of the asymptotic power becomes

$$\begin{aligned} & 1 - \Phi\left(z_\alpha - \frac{\sum_{i=1}^{k} n_i c_i p_i}{\sqrt{\sum c_i^2 n_i p_i (1-p_i)}} \right) \\ = \quad & \Phi\left(\frac{\sum_{i=1}^{k} n_i c_i p_i}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i p_i (1-p_i)}} - z_\alpha \right) \end{aligned}$$

It is seen that the power converges to 1 with increasing n.

Alternatively, we may use the statistic

$$\frac{\sum_{i=1}^{k} n_i c_i (\bar{p}_i - \bar{p})}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\sum_{i=1}^{k} n_i c_i^2}} \tag{11}$$

which in the simulation studies reported appears to more closely attain the prescribed significance level. In the case of equal sample sizes, $n_i = n_0$ say, the test statistic (11) takes the simpler form

$$\frac{\sum_{i=1}^{k} \left(i - \frac{k+1}{2}\right)(\bar{p}_i - \bar{p})}{\sqrt{n_0}\sqrt{\bar{p}(1-\bar{p})}\sqrt{\frac{k(k^2-1)}{12}}}$$

Theorem 2. *Under the null hypothesis and under* (10), *the conditional mean and variance of* $\sum_{i=1}^{k} w_i y_i$ given *that* $\sum_{i=1}^{k} y_i = y$, *are given respectively by* $y\bar{p}$ *and*

$$\bar{p}\,(1-\bar{p})\sum_{i=1}^{k} n_i\,(w_i-\bar{p})^2$$

In the next theorem we demonstrate that asymptotically the Hamming statistic converges to a normal distribution.

Theorem 3. *Suppose that both* $y \to \infty$ *and* $n-y \to \infty$ *as* $n \to \infty$. *Under* H_o, *the Hamming test statistic has asymptotically a standard normal distribution, i.e.*

$$\frac{\sum_{i=1}^{k}\left[n_i w_i\,(\bar{p}_i-\bar{p})\right]}{\sqrt{\bar{p}\,(1-\bar{p})\sum_{i=1}^{k} n_i\,(w_i-\bar{p})^2}} \to_d N(0,1)$$

under either i) (10) *or ii*) *the* $\{n_i\}$ *are bounded and* $k \to \infty$.

Both test statistics share the same general form as the regression statistic, namely

$$\frac{\sum_i n_i\,(x_i-\bar{x}_k)\,(\bar{p}_i-\bar{p})}{\sqrt{\sum_i n_i\,(x_i-\bar{x}_k)^2}\,\sqrt{\bar{p}\,(1-\bar{p})}},\ \bar{x}_k = \sum n_i x_i/n \tag{12}$$

where for Spearman, $x_i = g_i$ and for Hamming $x_i = w_i$.

We may also consider the asymptotic null distribution of the Spearman and Hamming test statistics under the condition that $k \to \infty$.in lieu of (10).

4. Simulation Study

Comparing Table 2 with Table 3, it is seen that the significance level actually attained by the Spearman similarity measure when the $\{p_i\}$ in the variance expressions are estimated by $\overline{p}$ are closer to the prescribed 5% level than when $\{\overline{p_i}\}$ are used. From Table 4 we note that the level actually attained by the Hamming similarity measure is closer to the prescribed 5% level than Spearman's.

In Table 5 we report on simulations for the power when $k = 5$. Three cases were considered: proportions which are strictly increasing, proportions which are non-decreasing and some which have no particular pattern. It can be seen that the Spearman measure is clearly only superior in the first two cases. Predictably the power is smaller when the $\{p_i\}$ are closer together than when they are further apart.

5. Applications

Returning to the example on mortality rates in South Africa, we calculated values of 260.1 and 70.3 for the Spearman and Hamming similarity measures respectively. These yielded p-values $< 10^{-4}$.

Another application deals with water quality at Hong Kong beaches. Table 6 shows the geometric E. Coli count for each of 6 beaches in the Sai Kung district of Hong Kong during the period 1986-2009. A beach is classified as good if the count is at most 24. The Spearman test statistic yielded a value of 22.98 which points to strong evidence of an upward trend in the annual proportion of good beaches. The Hamming test statistic on the other hand yielded a value of 2.90 which has a p-value equal to 0.0018.

6. Conclusion

An approach has been proposed for constructing nonparametric tests of trend in proportions. Similarity measures due to Spearman, Kendall and Hamming led to new test statistics. It was shown that the Spearman and Kendall measures led to identical test statistics. The asymptotic null distributions for the Spearman and Hamming test statistics were shown to be normal. Simulations were performed to check on the significance level attained. It was seen that the Hamming measure was closer to the prescribed level. On the other hand, the Spearman similarity measure attained greater power under a variety of alternatives. Two applications were considered.

Acknowledgment

The authors are grateful to Professor Philip Yu for suggesting the application to water quality data.

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix

Proof of Theorem 1.

i) The Spearman statistic (7) is expressible as a linear combination of independent binomials, $\sum_{i=1}^{k} c_i y_i$ where $c_i = g_i - \frac{n+1}{2}$ and $g_i = \sum_{i=1}^{i-1} n_i + \left(\frac{n_i+1}{2}\right)$. Hence, under (10), we have approximately $y_i \approx_d N(n_i p_i, n_i p_i (1-p_i))$. In view of the independence of the $\{y_i\}$,

$$\sum c_i y_i \approx_d N\left(\sum c_i n_i p_i, \sum c_i^2 n_i p_i (1-p_i)\right)$$

and hence

$$\frac{(\sum c_i y_i - \sum c_i n_i p_i)}{\sqrt{\sum c_i^2 n_i p_i (1-p_i)}} \to_d N(0,1).$$

Under H_0, $\sum c_i n_i p_i = 0$ and hence

$$\frac{\sum c_i y_i}{\sqrt{\sum c_i^2 n_i p_i (1-p_i)}} \to_d N(0,1).$$

The theorem follows.

ii) For this part, note that the Spearman statistic is expressible as

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i y_{ij}$$

where the $\left\{y_{ij}\right\}$ are Bernoulli (p_i). We need to show that

$$\frac{\max_{1\le i\le k} c_i^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2} \to 0, as\ k \to \infty$$

Note that if $n_i \le M$, for all i,

$$\begin{aligned} c_i^2 &\le g_i^2 + \frac{(n+1)^2}{4} \\ &\le M^2\left(i - \frac{1}{2}\right)^2 + \frac{1}{4} + \frac{(n+1)^2}{4} = O\left(k^2\right) \end{aligned}$$

Moreover, since $n = \sum_{i=1}^{k} n_i \le Mk$

$$\begin{aligned} \sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2 &= \sum_{i=1}^{k} n_i g_i^2 - \frac{n(n+1)^2}{4} \\ &\ge \sum_{i=1}^{k} n_i g_i^2 - \frac{M^3 k^3}{4} \\ &= \sum_{i=1}^{k} n_i \left(\sum_{j=1}^{i-1} n_j\right)^2 + \sum_{i=1}^{k} (n_i+1)\left(\sum_{j=1}^{i-1} n_j\right) + \sum_{i=1}^{k} n_i \left(\frac{n_i+1}{2}\right)^2 - \frac{M^3 k^3}{4} \\ &\ge \sum_{i=1}^{k} n_i (i-1)^2 + 2\sum_{i=1}^{k} (i-1) = O\left(k^3\right) \end{aligned}$$

It follows that

$$\frac{\max_{1\le i\le k} c_i^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2} = O(\frac{1}{k}) \to 0, as\ k \to \infty$$

and the theorem is proved.

Proof of Theorem 2.

Turning attention to the Hamming statistic (8), we note that under H_0 and conditional on $\sum_{i=1}^{k} y_i = y$, the joint probability distribution of the $\{y_i\}$ is the multivariate hypergeometric

$$\frac{\Pi_{i=1}^{k}\binom{n_i}{y_i}}{\binom{n}{y}} \tag{13}$$

It follows that

$$\begin{aligned} E\left(y_i|H_0, y\right) &= \frac{n_i}{n}y \\ Var\left(y_i|H_0, y\right) &= \frac{n_i\left(n-n_i\right)}{n^2\left(n-1\right)}y\left(n-y\right) \\ Cov\left(y_i, y_j|H_0, y\right) &= -\frac{n_i n_j}{n^2\left(n-1\right)}y\left(n-y\right), i \neq j \end{aligned}$$

Hence,

$$\begin{aligned} E\left(\sum_{i=1}^{k} w_i y_i|H_0, y\right) &= \left(\sum_{i=1}^{k} w_i \frac{n_i}{n} y\right) \\ &= y\bar{p} \end{aligned}$$

and

$$\begin{aligned} Var\left(\sum_{i=1}^{k} w_i y_i|H_0, y\right) &= \sum_i w_i^2 Var\left(y_i|H_0, y\right) + \sum_{i\neq j} w_i w_j Cov\left(y_i, y_j|H_0, y\right) \\ &= \frac{y\left(n-y\right)}{n^2\left(n-1\right)}\left[\sum_i n_i\left(n-n_i\right)w_i^2 - \sum_{i\neq j} n_i n_j w_i w_j\right] \\ &= \frac{y\left(n-y\right)}{n^2\left(n-1\right)}\left[n\sum_i n_i w_i^2 - \left(\sum n_i w_i\right)^2\right] \\ &= \bar{p}\left(1-\bar{p}\right)\sum_{i=1}^{k} n_i\left(w_i-\bar{p}\right)^2 \end{aligned}$$

Proof of Theorem 3.

Given (10), it follows that (13) converges to a multinomial distribution independent of p given by

$$\binom{y}{y_1 y_2 ... y_k}\lambda_1^{y_1}\lambda_2^{y_2}...\lambda_k^{y_k}$$

Suppose now that the observed weights $\{w_i\}$ are the values obtained from an i.i.d. sample from a population with mean μ and variance σ^2. Consider a bootstrap random sample of size y, say $W_1^*, ..., W_y^*$ obtained with replacement from that population of observed weights $\{w_i\}$ in accordance with the distribution

$$P\left(W^* = w_i\right) = \lambda_i, i = 1, ..., k$$

The central limit theorem then asserts that for large y, the sum of the bootstrap sample

$$S^* = \sum_{i=1}^{y} W_i^*$$

is asymptotically normal. Now y will get large a.s. since $n \to \infty$ for otherwise $y/n \nrightarrow p$ a.s. and thus contradict the strong law of large numbers.

To relate the Hamming distribution to that of the bootstrap, let x_i denote the number of times w_i is selected in the resampling. It follows that the vector $(x_1, ..., x_k)$ has a multinomial distribution $(y; \lambda_1, ..., \lambda_k)$ and S^* can be equivalently written as

$$S^* = \sum_{i=1}^{k} w_i x_i =_d \sum_{i=1}^{k} w_i y_i.$$

Consequently, under the null hypothesis for large n and conditional on y

$$\frac{\sum_{i=1}^{k}\left[n_i w_i\left(\bar{p}_i-\bar{p}\right)\right]}{\sqrt{\bar{p}\left(1-\bar{p}\right)\sum_{i=1}^{k} n_i\left(w_i-\bar{p}\right)^2}} \rightarrow_d N\left(0,1\right)$$

and hence this is also true unconditionally. Note that

$$\sum_{i=1}^{k} n_i\left(w_i-\bar{p}\right)=\left(\sum_{i=1}^{k} w_{1i}\right)-n\bar{p}=0.$$

ii) For this part, it follows that since $0 \leq w_i \leq 1$, for all i and

$$\sum_{i=1}^{k} n_i w_i^2 \geq y \tag{14}$$

$$\frac{\max_{1\leq i\leq k} w_i^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} w_i^2}=O(\frac{1}{y})\rightarrow 0$$

and the theorem is proved.

References

Alvo, M., & Cabilio, P. (1995). Testing ordered alternatives in the presence of incomplete data. *Journal of the American Statistical Association, 90*, 1015-1024. http://dx.doi.org/10.2307/2291338

Alvo, M., & Cabilio, P. (1999). A general rank based approach to the analysis of block data. *Commun. Statist.-Theory Meth, 28*(1), 197-215. http://dx.doi.org/10.1080/03610929908832291

Alvo, M., & Pan, J. (1997). A general theory of hypothesis testing based on rankings. *Journal of Statistical Planning and Inference, 61*, 219-248. http://dx.doi.org/10.1016/S0378-3758(97)89436-0

Arase, Y., Chayama, K., & Ikeda, K. (2001). Randomized controlled clinical trial of lymphoblastoid interferon-alpha for chronic hepatitis C. *Hepathology, 21*, 55-66.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics, 11*(3), 375-386. http://dx.doi.org/10.2307/3001775

Baumgartner, W., Weiβ, P., & Schindler, H. (1998). A nonparametric test for the general two sample problem. *Biometrics, 54*, 1129-1135. http://dx.doi.org/10.2307/2533862

Chen, J. J., Kodell, R. L., & Pearce, B. A. (1997). Significance levels of randomization trend tests in the event of rare occurrences. *Biometrical Journal, 39*, 327-337. http://dx.doi.org/10.1002/bimj.4710390307

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics, 10*, 417-451. http://dx.doi.org/10.2307/3001616

Corcoran, C., Mehta, C., & Senchaudhuri, P. (2000). Power comparisons for tests of trend in dose-response studies. *Statistics in Medicine, 19*, 3037-3050. http://dx.doi.org/10.1002/1097-0258(20001130)19:22<3037::AID-SIM601>3.0.CO;2-7

Hong Kong water quality data from 1996 to 2000. [Online] Available: http://www.epd.gov.hk/epd/english/environmentinhk/water/beach_quality/bwq _report8600_appendixa.htm

Hong Kong water quality data from 1997 to 2009. [Online] Available: http://www.epd.gov.hk/epd/english/environmentinhk/water/beach _quality/files/bwq_report2009.pdf

Hothorn, L. A., & Bretz, F. (2000). Evaluation of animal carcinogenicity studies : Cochran-Armitage trend test vs multiple contrast tests. *Biometrical Journal, 42*, 553-567. http://dx.doi.org/10.1002/1521-4036(200009)42:5<553::AID-BIMJ553>3.0.CO;2-R

Ku, J. H., Kim, M. E., Lee, N. K., & Park, Y. H. (2001). The prevalence of chronic prostatitis-like symptoms in young men: a community-based survey. *Urological Research, 29*, 108-112. http://dx.doi.org/10.1007/s002400100172

Neuhäuser, M. (2006). An exact test for trend among binomial proportions based on a modified Baumgartner-Wei β-Schindler statistic. *Journal of Applied Statistics, 33*, 79-88.

Portier, C., & Hoel, D. (1984). Type I error of trend tests in proportions and the design of cancer screens. *Commun. Statist.-Theory Meth., 13*, 1-14. http://dx.doi.org/10.1080/03610928408828657

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present. *Indagationes Mathematicae, 14*, 327-333.

Van Eeden, C., & Hemelrijk, J. (1955). A test for the equality of probabilities against a class of specified alternative hypotheses including tesnt I. *Proc. Kon. Ned. Akad. van Wet., A*(58), 191-198.

Williams, D. A. (1998). Tests for differences between several small proportions. *Applied Statistics, 37*, 421-434. http://dx.doi.org/10.2307/2347316

Table 1. Mortality statistics for South Africa 2000-2008

Year	Number of deaths	Population size
2000	416, 155	43, 789, 115
2001	454, 882	43, 997, 828
2002	502, 050	44, 187, 637
2003	556, 779	44, 344, 136
2004	576, 709	42, 718, 530
2005	598, 131	42, 768, 678
2006	612, 778	43, 647, 658
2007	603, 094	43, 586, 097
2008	592, 073	43, 421, 021

Table 2. Significance level for Spearman's similarity when p_i is estimated by $\overline{p_i}$

p	10	20	30	50
0.1	0.018	0.035	0.045	0.053
0.2	0.048	0.057	0.055	0.055
0.3	0.062	0.060	0.058	0.054
0.4	0.067	0.057	0.057	0.055
0.5	0.071	0.055	0.059	0.053

Table 3. Significance level for Spearman's similarity when p_i is estimated by $\overline{p}$

p	10	20	30	50
0.1	0.044	0.045	0.050	0.053
0.2	0.051	0.054	0.052	0.053
0.3	0.052	0.053	0.052	0.052
0.4	0.052	0.049	0.052	0.052
0.5	0.048	0.052	0.049	0.051

Table 4. Significance level for Hamming's similarity

p	10	20	30	50
0.1	0.056	0.064	0.061	0.056
0.2	0.054	0.056	0.056	0.055
0.3	0.054	0.052	0.052	0.052
0.4	0.054	0.055	0.052	0.053
0.5	0.053	0.053	0.050	0.050

Table 5. Power for Spearman (S) and Hamming (H) similarity measures using a 5% significance level

					n_i = 10		n_i = 30	
p_1	p_2	p_3	p_4	p_5	H	S	H	S
0.1	0.2	0.4	0.5	0.6	0.77	0.89	0.99	1.00
0.3	0.6	0.7	0.8	0.9	0.88	0.94	1.00	1.00
0.5	0.6	0.7	0.8	0.9	0.58	0.72	0.94	0.99
0.2	0.25	0.3	0.35	0.4	0.25	0.29	0.50	0.60
0.1	0.3	0.3	0.4	0.5	0.51	0.64	0.88	0.97
0.3	0.5	0.5	0.5	0.9	0.58	0.81	0.94	1.00
0.4	0.5	0.6	0.8	0.8	0.62	0.75	0.95	0.99
0.2	0.5	0.3	0.8	0.4	0.51	0.39	0.98	0.80
0.6	0.2	0.4	0.8	0.5	0.46	0.18	0.88	0.39

Table 6. Annual Geometric Mean E. Coli level (per 100ml) in the Sai Kung District. Beaches: Clear Water Bay First (1), Clear Water Bay Second (2), Hap Mun Bay (3), Kiu Tsui (4), Silverstrand (5), Trio (6); number of good beaches (7)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1986	102	69	9	18	255	49	2
1987	133	52	6	9	62	32	2
1988	39	35	4	3	129	35	2
1989	80	38	3	5	192	23	3
1990	51	42	4	5	89	31	2
1991	30	14	2	4	106	14	4
1992	52	42	2	5	94	32	2
1993	31	16	3	4	56	20	4
1994	30	35	3	3	72	14	3
1995	55	39	6	3	226	16	3
1996	34	43	5	5	126	29	2
1997	62	66	3	5	148	30	2
1998	41	44	2	4	99	21	3
1999	11	12	2	4	32	17	5
2000	16	26	2	5	61	10	4
2001	28	22	1	5	100	12	4
2002	28	14	2	4	133	6	4
2003	17	21	4	5	97	10	5
2004	9	10	3	17	74	2	5
2005	16	19	4	14	67	6	5
2006	20	13	4	11	30	5	5
2007	14	9	3	6	33	2	5
2008	11	19	5	12	35	12	5
2009	15	27	3	19	31	5	4