

Partition of the Chi-Squared Statistic in a Contingency Table

by

Jo Ann Colas

Thesis submitted submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.Sc. degree in
Mathematics and Statistics

Department of Mathematics and Statistics
Faculty of Sciences
University of Ottawa

© Jo Ann Colas, Ottawa, Canada, 2014

Abstract

The Pearson statistic, a well-known goodness-of fit test in the analysis of contingency tables, gives little guidance as to why a null hypothesis is rejected. One approach to determine the source(s) of deviation from the null is the decomposition of a chi-squared statistic. This allows writing the statistic as the sum of independent chi-squared statistics.

First, three major types of contingency tables and the usual chi-squared tests are reviewed. Three types of decompositions are presented and applied: one based on the partition of the contingency table into independent subtables; one derived from smooth models and one from the eigendecomposition of the central matrix defining the statistics. A comparison of some of the omnibus statistics decomposed above to a $\chi^2(1)$ -distributed statistic shows that the omnibus statistics lack power compared to this statistic for testing hypothesis of equal success probabilities against monotonic trend in the success probabilities in a column-binomial contingency table.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors Dr. M. Alvo and Dr. P.-J. Bergeron for their continuous support of my Master's study and research, for their patience, motivation and knowledge. Their guidance helped me in all the time of researching and writing this thesis. I would also like to thank my committee members, Dr. M. Zarepour and Dr. S Sinha, for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions.

I thank my fellow labmates at the Department of Mathematics and Statistics: Maryam, Blanche Nadege, Nada, Cheikh, Farid, Hicham, Rachid, Ewa, Farnoosh, Jack and Gael for the stimulating discussions, for the sleepless nights we worked together before deadlines and for all the fun we have had in the last two years.

Last but not least, I would like to thank my family: my mother Rosie as my editor, my father as my chauffeur and my brother as my motivator.

Contents

1	Introduction	1
2	Description of the Chi-Squared Test for a Contingency Table	3
2.1	The Chi-Squared Distribution	3
2.2	Contingency Tables	5
2.3	The Three Models Associated to Contingency Tables	6
2.3.1	The unrestricted bivariate sampling model	6
2.3.2	The product-multinomial model	8
2.3.3	The permutation model	10
2.3.4	The link between the three models	11
2.4	Chi-Squared Hypothesis Tests	12
2.4.1	The chi-squared goodness-of-fit test	13
2.4.2	The chi-squared test for homogeneity	16
2.4.3	The chi-squared test for independence	19
3	Tools for the Decomposition of the Pearson Statistic	23
3.1	Ordinal Categories and Testing for Trends	23
3.2	Ranked Data and Contingency Tables	25
3.2.1	Ranked data and their properties	25
3.2.2	Advantages and disadvantages of working with ranked data	26
3.3	Smooth Goodness-of-Fit Models	28
3.3.1	Neyman-type smooth models	28
3.3.2	Barton-type smooth models	29
3.3.3	Advantages and disadvantages of smooth models	30
3.4	Orthonormal Polynomials Defined From Distributional Moments	30
4	Decompositions of Some Chi-Squared Statistics	33
4.1	The Decomposition Into Subtables	33
4.2	The Decomposition of the Pearson Statistic	37
4.2.1	Decomposition of the Pearson statistic in Two-Way Tables	37
4.2.2	Under the unrestricted bivariate sampling model	39
4.2.3	Under the multinomial model	44

4.2.4	Under the binomial model	52
4.2.5	Under the randomised design model	53
4.2.6	Decomposition of the Pearson Statistic in Three-Way Tables	58
4.3	The Decomposition of the Cumulative Chi-Squared Statistic	61
4.3.1	General form of the decomposition of the cumulative chi-squared statistic	62
4.3.2	The distribution of the cumulative chi-squared statistic	64
4.3.3	Link to smooth models	65
4.3.4	Explicit solutions when the column are equiprobables	66
4.4	The Decomposition of the Chi-Squared Likelihood-Ratio Test Statistic	73
4.4.1	Log-linear models	73
4.4.2	Decomposition of the chi-squared likelihood-ratio test statistic	74
4.4.3	Extension of the decomposition	76
5	Comparisons of the CCS Statistics and a Correlation Based Statistic	77
5.1	The Alvo & Berthelot Spearman Test	78
5.2	The Comparisons	79
5.2.1	The distributions under the null hypothesis	80
5.2.2	Comparison of the significance levels	81
5.2.3	Comparison of the powers	84
6	Conclusion	89
A	Obtaining the first two orthogonal polynomials from the Emerson recursion	91
B	Proof of symmetry of $(AA')^{-1}$ when the columns are equiprobable	96

List of Tables

2.1	An $I \times J$ contingency table	5
2.2	An $I \times J$ contingency table under the product-multinomial model	8
2.3	An $I \times J$ contingency table under the permutation model	10
4.1	The s -th 2×2 subtable	34
4.2	A rank-by-rank table (without ties)	43
4.3	A treatment-by-rank table (without ties)	49
4.4	The $I \times 2$ table around the overall sample median	53
4.5	A treatment-by-rank table (without ties)	55
4.6	The s -th collapsed $I \times 2$ contingency table	62
4.7	The $2 \times J$ collapsed table around the i -th row	68
4.8	Hierarchical log-linear models for an $I \times J$ contingency table	76
4.9	Hierarchical log-linear models for an $I \times J \times K$ contingency table	76
5.1	Minimum number of simulations	83
5.2	Significance levels for the Taguchi and Nair CCS statistics	83
5.3	Significance levels for the Pearson and Spearman statistics	83
5.4	Comparison of the powers – strictly increasing success probabilities	86
5.5	Comparison of the powers – increasing success probabilities with repeated values	87
5.6	Comparison of the powers – non-increasing and non-decreasing success probabilities	88

Chapter 1

Introduction

Karl Pearson's chi-square goodness-of-fit test was introduced at the end of the development of several important concepts in the preceding centuries. In 1733, de Moivre[19] established the asymptotic normality of the variable

$$X = \frac{Y - n\pi}{\sqrt{n\pi(1 - \pi)}},$$

where $Y \sim \text{Binomial}(n, \pi)$. Consequently, X^2 is asymptotically distributed as the square of the $N(0, 1)$ distribution. Bienaymé[12] published the distribution of the sum of m independently distributed $N(0, 1)$ variables in the gamma function form. Bravais[15], Schols[42] and Edgeworth[20] developed the joint multivariate normal distribution. A contemporary of Pearson, Sheppard[43, 44] considered possible tests of goodness-of-fit by comparing observed frequencies to expected frequencies for each cell of a contingency table. In a good fit, the difference would be small. In particular, Sheppard looked at 2×2 tables as a dichotomy of a bivariate normal distribution. However, he could not find a generalization for $I \times J$ tables due to the awkward form of the covariance matrix. By considering a multinomial distribution instead, Pearson[32] found a more tractable solution of the covariance matrix and provided the widely used chi-squared test of goodness-of-fit.

The advantages of the Pearson test include: easy to compute; used on categorical variables; used to compare two or more populations/samples; no assumption on the distribution of the population(s). However, it has some shortcomings. Since it is based on categorical variables, there will be some loss of information if it is used with samples based on continuous distributions. Also, since the distribution of the test statistic is obtained asymptotically, the Pearson test is sensitive to sample size.

In this thesis, we focus on the fact that the Pearson test is an omnibus test. When it rejects the null hypothesis, it tells there is enough evidence to suggest a relationship, but it is not clear about the strength or the type of this relationship. One approach to overcome this drawback is to use the additive property of the chi-squared distribution that allows writing a chi-squared variable as the sum of asymptotically independent chi-squared variables. Such a breakdown of a chi-squared statistic is called a *partition* or a *decomposition* of the statistic. Many decompositions of the Pearson statistic have been proposed, allowing for the detection of different deviations from

the null distribution.

Agresti[1] and Iversen[25] present a decomposition of the Pearson statistic based on the decomposition of the $I \times J$ contingency table into subtables. If the data is in the form of ranks, Rayner & Best provide a partition of the Pearson statistic based on a set of polynomials orthonormal with respect to distribution. Depending on the model of the contingency table, this approach results in the extension of widely-used tests on two-way contingency table, such as the Pearson product-moment correlation coefficient[36], the Kruskal-Wallis test[37] and the Friedman test[16]. If the alternative hypothesis is directional, other chi-squared test statistics have been proposed and decomposed, in particular, the class of cumulative chi-squared tests based on the weighted sum of increasing sums of squares initially proposed by Taguchi[45, 46]. If the data is binary and the alternative is that of monotone trend in the proportions, Alvo-Berthelot[2] proposed a test statistic motivated by rankings. This thesis reviews the above-mentioned decompositions.

In Chapter 2, we first review the types of contingency tables and then go over the usual chi-squared test based on the Pearson statistic. In Chapter 3, we review the necessary tools for the decompositions, including properties of ranked data, a brief review of smooth models and the derivation of polynomials orthonormal with respect to a distribution. Chapter 4 discusses the different decompositions. Finally, in Chapter 5 we study the power simulations of the Pearson test, the cumulative chi-squared tests and the Spearman statistic from Alvo-Berthelot[2] to show that when we want to test homogeneity against monotonicity in one parameter, alternative and simpler tests are more powerful.

Chapter 2

Description of the Chi-Squared Test for a Contingency Table

In this chapter, we first define the chi-squared distribution and describe some of its properties. We follow with describing the three major types of contingency tables and giving the distribution of the cell counts. We end the chapter with a review of the usual chi-squared hypothesis tests and the associated Pearson and likelihood ratio test statistics.

2.1 The Chi-Squared Distribution

Definition 2.1. A random variable W has a chi-squared distribution with ν degrees of freedom, denoted by $W \sim \chi^2(\nu)$, if its density function is given by

$$f(w|\nu) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} w^{\frac{\nu-2}{2}} e^{-\frac{w}{2}}, \quad 0 \leq w < \infty, \quad \nu > 0.$$

From the definition of the density function for the chi-squared distribution, we can see that this distribution is in fact a special case of the gamma distribution.

The mean, variance and characteristic function are respectively given by

$$E[W] = \nu, \quad Var[W] = 2\nu, \quad \phi_W(t) = \frac{1}{(1 - 2it)^{\frac{\nu}{2}}}$$

Theorem 2.1. If W_1 and W_2 are two independent variables such that $W_1 \sim \chi^2(\nu_1)$ and $W_2 \sim \chi^2(\nu_2)$, then the random variable $W_1 + W_2 \sim \chi^2(\nu_1 + \nu_2)$. Conversely, if a random variable $W \sim \chi^2(\nu)$, W can be always be expressed as the sum of two independent random variables W_1 and W_2 such that $W_1 \sim \chi^2(\nu_1)$, $W_2 \sim \chi^2(\nu_2)$ and $\nu = \nu_1 + \nu_2$.

Proof. Let

$$\phi_{W_k}(t) = (1 - 2it)^{-\nu_k/2}$$

be the characteristic function for W_k , $k = 1, 2$. Since W_1 and W_2 are independent, the characteristic function

of $W_1 + W_2$ is

$$\begin{aligned}\phi_{W_1 \cdot W_2}(t) &= \phi_{W_1}(t) \phi_{W_2}(t) \\ &= \frac{1}{(1 - 2it)^{\frac{\nu_1}{2}}} \frac{1}{(1 - 2it)^{\frac{\nu_2}{2}}} \\ &= \frac{1}{(1 - 2it)^{\frac{\nu_1 + \nu_2}{2}}}\end{aligned}$$

which is the the characteristic function for the $\chi^2(\nu_1 \cdot \nu_2)$ distribution. \square

Remark 2.2. By recursion, Theorem 2.1 is true for any finite number independent chi-squared variables such that the sum of their degrees of freedom is ν .

Lemma 2.3. Let Z be a standard normal variable. Then, Z^2 has a $\chi^2(1)$ distribution.

Proof. Let $W = Z^2$. Then, $dw = 2zdz$. Then, for all $w \in (0, \infty)$, the cdf for W is

$$\begin{aligned}F_W(w) &= P(W \leq w) \\ &= P(Z^2 \leq w) \\ &= P(-w^{1/2} \leq Z \leq +w^{1/2}) \\ &= \Phi(+w^{1/2}) - \Phi(-w^{1/2})\end{aligned}$$

where Φ is the cdf for the $N(0, 1)$ distribution which has pdf $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$ for $-\infty \leq z \leq +\infty$. Then, the pdf for $W = Z^2$ is

$$\begin{aligned}f_W(w) &= \frac{d}{dw} F_W(w) \left| \frac{dz}{dw} \right| \\ &= \frac{1}{2} \left[\phi(+w^{1/2}) + \phi(-w^{1/2}) \right] w^{-1/2} \\ &= \frac{1}{\sqrt{2\pi w}} e^{-\frac{w}{2}} \\ &= \frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{\frac{1}{2}}} w^{\frac{1-2}{2}} e^{-\frac{w}{2}}\end{aligned}$$

which is the density function for the $\chi^2(1)$ distribution. \square

Theorem 2.4. Let Z_1, \dots, Z_ν be ν independent and identically distributed (iid) variables with a $N(0, 1)$ distribution. Then, $W = Z_1^2 + \dots + Z_\nu^2 \sim \chi^2(\nu)$.

Proof. By Lemma 2.3, the Z_1^2, \dots, Z_ν^2 are iid with a $\chi^2(1)$ distribution and so by Remark 2.2, $W = Z_1^2 + \dots + Z_\nu^2 \sim \chi^2(\nu)$. \square

Alternatively, some authors define a chi-squared variable as the sum of ν independent squared standard normal variables and then show that it is in fact a special case of a gamma variable.

2.2 Contingency Tables

We recall the definition of a categorical variable, also known as a nominal variable.

Definition 2.2. A categorical variable is a variable that can take on one of a limited, and usually fixed, number of possible values which are defined in terms of categories.

A categorical variable can be represented on either a nominal or ordinal scale. For example, the variable *colour* has nominal scale where the categories *red*, *orange*, *yellow*, *green*, *blue*, *indigo* and *violet* have no particular order. On the other hand, the variable *work qualification*, with categories *qualified*, *semi-qualified* and *not qualified*, has an ordinal scale, where there is a decreasing level of qualification from the first to the last category.

Let X be a categorical variable with classification

$$A = \left\{ A_1, \dots, A_I : A_i \cap A_j = \emptyset, \quad i \neq j = 1, \dots, I, \quad \bigcup A_i = \Xi \right\}$$

where Ξ is the set of possible values for X . Similarly, let Y be a categorical variable with classification

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Psi \right\}$$

where Ψ is the set of possible values for Y .

Suppose $(X_1, Y_1), \dots, (X_n, Y_n)$ is a random sample of the random vector (X, Y) . Let N_{ij} denote the random variable which counts the number of observations that fall into the cross-category $A_i \times B_j$. That is,

$$N_{ij} = \sum_{k=1}^n I[X_k \in A_i, Y_k \in B_j], \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

where $I[\cdot]$ is the indicator function, defined as

$$I[A] = \begin{cases} 1 & \text{if the event } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Table 2.1 represents an $I \times J$ contingency table for X and Y where

Table 2.1: An $I \times J$ contingency table

	B_1	\cdots	B_j	\cdots	B_J	Total
A_1	N_{11}	\cdots	N_{1j}	\cdots	N_{1J}	$N_{1\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_i	N_{i1}	\cdots	N_{ij}	\cdots	N_{iJ}	$N_{i\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_I	N_{I1}	\cdots	N_{Ij}	\cdots	N_{IJ}	$N_{I\cdot}$
Total	$N_{\cdot 1}$	\cdots	$N_{\cdot j}$	\cdots	$N_{\cdot J}$	n

- N_{ij} represents the counts for the $A_i \times B_j$ cross-category
- $N_{i\cdot}$, called the i -th row total, represents the counts for the category A_i

$$N_{i\cdot} = \sum_{k=1}^n 1[X_k \in A_i] = N_{i1} + \cdots + N_{iJ} \quad i = 1, \dots, I;$$

- $N_{.j}$, called the j -th column total, represents the counts for the category B_j

$$N_{.j} = \sum_{k=1}^n 1[Y_k \in B_j] = N_{1j} + \cdots + N_{Ij} \quad j = 1, \dots, J;$$

- n represents the total number of observations, which is always fixed and known.

Parameters of interest in an $I \times J$ contingency table are the cell and marginal probabilities:

- π_{ij} denotes the probability that (X, Y) falls into the (i, j) -th cross-category

$$\pi_{ij} = P(X \in A_i, Y \in B_j), \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

- $\pi_{i.}$ denotes the probability that X is in category A_i : $\pi_{i.} = P(X \in A_i)$, $i = 1, \dots, I$. Since we have

$$P(X \in A_i) = P(X \in A_i, Y \in \Psi) = \sum_{j=1}^J P(X \in A_i, Y \in B_j)$$

where the B_j are disjoint, we have that

$$\pi_{i.} = \pi_{i1} + \cdots + \pi_{iJ}, \quad i = 1, \dots, I.$$

- $\pi_{.j}$ denotes the probability that Y is in category B_j : $\pi_{.j} = P(Y \in B_j)$, $j = 1, \dots, J$. Since we have

$$P(Y \in B_j) = P(X \in \Xi, Y \in B_j) = \sum_{i=1}^I P(X \in A_i, Y \in B_j)$$

where the A_i are disjoint, we have that

$$\pi_{.j} = \pi_{1j} + \cdots + \pi_{Ij}, \quad j = 1, \dots, J.$$

Here we have presented two-way contingency tables but this concept can be extended to *multi-way contingency tables* which cross-classify three or more categorical variables.

2.3 The Three Models Associated to Contingency Tables

As Lancaster[27] indicates, there are three probability models that allow for analysis on contingency tables. In the models, while n is always fixed, the sets of marginal totals may or may not be fixed.

2.3.1 The unrestricted bivariate sampling model

Suppose X and Y are categorical variables and the random vector (X, Y) has a joint distribution

$$P(X \in A_i, Y \in B_j) = \pi_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

such that $\pi_{11} + \cdots + \pi_{IJ} = 1$, where A_i , and B_j , are the categories for X and Y , respectively, $i = 1, \dots, I$, $j = 1, \dots, J$.

Let us assume n observations of (X, Y) are chosen by a random process with assigned weight π_{ij} . Let N_{ij} denote the number of observations in the (i, j) -th cross-category. The data can be arranged as in Table 2.1.

The probability for the cell counts N_{ij} of the $I \times J$ contingency table for X and Y is given by the multinomial distribution

$$P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, n) = \binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}} \quad (2.1)$$

where $\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{IJ})'$. The log-likelihood function for $\boldsymbol{\pi}$ is

$$l(\boldsymbol{\pi} | \{n_{ij}\}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij}.$$

Under the constraint $\pi_{11} + \dots + \pi_{IJ} = 1$, the Lagrange function is

$$L(\{n_{ij}\}, \lambda | \boldsymbol{\pi}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij} - \lambda \left(\sum_{i=1}^I \sum_{j=1}^J \pi_{ij} - 1 \right).$$

To find the ML estimator for π_{ij} , we maximise the Lagrange function:

$$\frac{\partial}{\partial \pi_{ij}} L(\{n_{ij}\}, \lambda | \boldsymbol{\pi}, n) = 0 \Leftrightarrow \frac{n_{ij}}{\pi_{ij}} - \lambda = 0 \Leftrightarrow \pi_{ij} = \frac{n_{ij}}{\lambda}.$$

Substituting into the constraint yields $(n_{11} + \dots + n_{IJ}) / \lambda = 1$, so $\lambda = n$ and the unrestricted ML estimators for the probabilities π_{ij} are

$$\hat{\pi}_{ij} = \frac{N_{ij}}{n} \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Given the cell counts have a multinomial distribution, the unrestricted ML estimators for the means are

$$\hat{\mu}_{ij} = n \hat{\pi}_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

and the unrestricted ML estimators for the covariances are

$$\begin{aligned} \text{cov}[N_{ij}, N_{ab}] &= n \begin{cases} \hat{\pi}_{ij}(1 - \hat{\pi}_{ij}), & i = a, \quad j = b, \\ -\hat{\pi}_{ij}\hat{\pi}_{ab}, & \text{otherwise,} \end{cases} \\ &= n \hat{\mathbf{R}} \end{aligned}$$

where \mathbf{R} is the $IJ \times IJ$ matrix given by

$$\mathbf{R} = [\text{diag}(\boldsymbol{\pi}) - (\boldsymbol{\pi}\boldsymbol{\pi}')].$$

In this model, the only fixed quantity is the total number of observations.

Example 2.1. Suppose we observe $n = 250$ individuals. We cross-classify each individual with respect to their sex (male or female) and their handedness (right-, left-handed or ambidextrous). Then, the 2×3 contingency table for the variables sex and handedness has an unrestricted bivariate sampling model.

2.3.2 The product-multinomial model

Lancaster[27] presents this as the *comparative trial model* however it is now more commonly known as the *product-multinomial model*.

Let X_1, \dots, X_I be random variables defined on the same outcome space Ξ which has the partition

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Xi \right\}$$

Suppose X_{i1}, \dots, X_{in_i} is a random sample for $X_i, i = 1, \dots, I$. Let N_{ij} denote the number of observations of the i -th sample which are in the j -th category. That is

$$N_{ij} = \sum_{k=1}^{n_i} 1 [X_{ik} \in B_j], \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Then the $I \times J$ contingency table for the variables X_1, \dots, X_I against the classification B is as Table 2.2.

Table 2.2: An $I \times J$ contingency table under the product-multinomial model

	B_1	\dots	B_j	\dots	B_J	Total
X_1	N_{11}	\dots	N_{1j}	\dots	N_{1J}	$n_{1\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_i	N_{i1}	\dots	N_{ij}	\dots	N_{iJ}	$n_{i\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_I	N_{I1}	\dots	N_{Ij}	\dots	N_{IJ}	$n_{I\cdot}$
Total	$N_{\cdot 1}$	\dots	$N_{\cdot j}$	\dots	$N_{\cdot J}$	n

In this model, the rows are independent. We denote the row totals as $n_{i\cdot}$, to emphasize that they are known and fixed

$$n_{i\cdot} = N_{i1} + \dots + N_{iJ} = n_{i\cdot}, \quad i = 1, \dots, I$$

and the product-multinomial model is called the *row-multinomial model*.

Let π_{ij} denote the probability that an observation of the i -th variable falls into the j -th category. That is,

$$P(X_i \in B_j) = \pi_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Since B partitions Ξ , the probabilities for each row sum to one

$$\pi_{i1} + \dots + \pi_{iJ} = 1, \quad i = 1, \dots, I.$$

Then, assuming that each observation is independent of every other observation, the i -th row \mathbf{N}_i has a *multinomial*($n_{i\cdot}, \boldsymbol{\pi}_i$) distribution, where $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{iJ})'$, $i = 1, \dots, I$, and the probability for the cell counts is given by the product of I multinomial distributions

$$P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{n_{i\cdot}\}, n) = \prod_{i=1}^I \left[\binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} \prod_{j=1}^J \pi_{ij}^{n_{ij}} \right] \quad (2.2)$$

where $\boldsymbol{\pi}' = (\boldsymbol{\pi}'_1, \dots, \boldsymbol{\pi}'_I)$. The log-likelihood function for $\boldsymbol{\pi}$ is

$$l(\boldsymbol{\pi} | \{n_{ij}\}, \{n_{i\cdot}\}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij}.$$

Under the constraints $\pi_{i1} + \dots + \pi_{iJ} = 1$, $i = 1, \dots, I$, the Lagrange function is

$$L(\{n_{ij}\}, \lambda | \boldsymbol{\pi}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij} - \sum_{i=1}^I \lambda_i \left(\sum_{j=1}^J \pi_{ij} - 1 \right).$$

To find the ML estimator for π_{ij} , we maximise the Lagrange function:

$$\frac{\partial}{\partial \pi_{ij}} L(\{n_{ij}\}, \lambda | \boldsymbol{\pi}, n) = 0 \Leftrightarrow \pi_{ij} = \frac{n_{ij}}{\lambda_i}.$$

Substituting into the i -th constraint yields $(n_{i1} + \dots + n_{iJ})/\lambda_i = 1$, so $\lambda_i = n_{i\cdot}$, $i = 1, \dots, I$, and the unrestricted ML estimator for the probabilities π_{ij} are

$$\hat{\pi}_{ij} = \frac{N_{ij}}{n_{i\cdot}} \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Given that the rows have independent multinomial distributions, the unrestricted ML estimators for the means are

$$\hat{\mu}_{ij} = n_{i\cdot} \hat{\pi}_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

and the unrestricted ML estimators for the covariances are

$$\begin{aligned} \hat{c}ov[N_{ij}, N_{ab}] &= n_{i\cdot} \begin{cases} \hat{\pi}_{ij}(1 - \hat{\pi}_{ij}), & i = a, \quad j = b, \\ -\hat{\pi}_{ij}\hat{\pi}_{ib}, & i = a, \quad j \neq b, \\ 0, & i \neq a, \end{cases} \\ &= n_{i\cdot} \begin{cases} \hat{\mathbf{R}}_i & i = a, \\ 0, & i \neq a. \end{cases} \end{aligned}$$

where \mathbf{R}_i is the $J \times J$ matrix given by

$$\mathbf{R}_i = [diag(\boldsymbol{\pi}_i) - (\boldsymbol{\pi}_i \boldsymbol{\pi}'_i)] \quad i = 1, \dots, I. \quad (2.3)$$

The same way the total number of observations in the rows were fixed, the total number of observations in the columns could be fixed instead, which would be denoted by $n_{\cdot j}$, $j = 1, \dots, J$. Then, the j -th column $\mathbf{N}_{\cdot j}$ has a *multinomial*($n_{\cdot j}, \boldsymbol{\pi}_{\cdot j}$) distribution, the probability for the cell counts would be given by

$$P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{n_{\cdot j}\}, n) = \prod_{j=1}^J \left[\binom{n_{\cdot j}}{n_{1j}, \dots, n_{Ij}} \prod_{i=1}^I \pi_{ij}^{n_{ij}} \right]$$

and the product-multinomial model is now called a *column-multinomial model*.

Example 2.2. Suppose we sample $n_i = 50$ individuals from $I = 5$ countries. For each country, we classify each individual with respect to their socio-economic class (upper, middle or lower class). Then, the 5×3 contingency table for the variables country and socio-economic class has a product-multinomial model.

2.3.3 The permutation model

Suppose we have n independent observations which can be categorised according to one of two possible classifications

$$A = \left\{ A_1, \dots, A_I : A_i \cap A_j = \emptyset, \quad i \neq j = 1, \dots, I, \quad \bigcup A_i = \Xi \right\}$$

and

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Xi \right\}$$

where Ξ is the set of possible values for the observations.

Let n_i , $i = 1, \dots, I$, and n_j , $j = 1, \dots, J$, be known and fixed positive non-zero integers such that $\sum_{i=1}^I n_i = \sum_{j=1}^J n_j = n$. From the n observations, select n_1 without replacement and assign them to A_1 , then select n_2 without replacement and assign them to A_2 , and so on. Now, we restart the process: from these same n observations select $n_{.1}$ without replacement and assign them to B_1 then select $n_{.2}$ without replacement and assign them to B_2 , and so on. This way, each object now has two labels. The $I \times J$ contingency table for this cross-classification is as Table 2.3.

Table 2.3: An $I \times J$ contingency table under the permutation model

	B_1	\cdots	B_j	\cdots	B_J	Total
A_1	N_{11}	\cdots	N_{1j}	\cdots	N_{1J}	$n_{1\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_i	N_{i1}	\cdots	N_{ij}	\cdots	N_{iJ}	$n_{i\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_I	N_{I1}	\cdots	N_{Ij}	\cdots	N_{IJ}	$n_{I\cdot}$
Total	$n_{\cdot 1}$	\cdots	$n_{\cdot j}$	\cdots	$n_{\cdot J}$	n

The row and column totals as are denoted by $n_{i\cdot}$ and $n_{\cdot j}$, respectively, to emphasize that they are fixed:

$$n_{i\cdot} = N_{i1} + \cdots + N_{iJ}, \quad n_{\cdot j} = N_{1j} + \cdots + N_{Ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

In this model, the fixed quantities are the total number of observations and both sets of marginal totals. The number of ways of obtaining cell counts for the i -th row is

$$\binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} = \frac{n_{i\cdot}!}{\prod_{j=1}^J n_{ij}!}$$

and so the number of ways obtaining an $I \times J$ contingency table with entries $\{n_{ij}\}$ is

$$\prod_{i=1}^I \binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} = \frac{\prod_{i=1}^I n_{i\cdot}!}{\prod_{i=1}^I \prod_{j=1}^J n_{ij}!}.$$

The number of ways of obtaining the column totals is

$$\binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}} = \frac{n!}{\prod_{j=1}^J n_{\cdot j}!}.$$

So, the probability for the cell counts is given by

$$\begin{aligned}
 P(\{N_{ij} = n_{ij}\} | \{n_{i\cdot}\}, \{n_{\cdot j}\}, n) &= \frac{1}{\binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}}} \prod_{i=1}^I \binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} \\
 &= \frac{\left(\prod_{i=1}^I n_{i\cdot}!\right) \left(\prod_{j=1}^J n_{\cdot j}!\right)}{n! \prod_{i=1}^I \prod_{j=1}^J n_{ij}!}.
 \end{aligned} \tag{2.4}$$

As per Lemma 5.1 of Lancaster's *The Chi-Squared Distribution*, the means for the cell counts are

$$\mu_{ij} = n\pi_{i\cdot}\pi_{\cdot j}$$

where, for the permutation model, the row and column probabilities are fixed at $\pi_{i\cdot} = n_{i\cdot}/n$ and $\pi_{\cdot j} = n_{\cdot j}/n$ respectively, $i = 1, \dots, I$, $j = 1, \dots, J$. Continuing from Lemma 5.1, the covariances are given by

$$\begin{aligned}
 cov[N_{ij}, N_{ab}] &= \frac{n^2}{n-1} \begin{cases} \pi_{i\cdot}(1-\pi_{i\cdot})\pi_{\cdot j}(1-\pi_{\cdot j}), & i = a, \quad j = b, \\ -\pi_{i\cdot}(1-\pi_{i\cdot})\pi_{\cdot j}\pi_{\cdot b}, & i = a, \quad j \neq b, \\ -\pi_{i\cdot}\pi_{a\cdot}\pi_{\cdot j}(1-\pi_{\cdot j}), & i \neq a, \quad j = b, \\ \pi_{i\cdot}\pi_{a\cdot}\pi_{\cdot j}\pi_{\cdot b}, & i \neq a, \quad j \neq b, \end{cases} \\
 &= \frac{n^2}{n-1} \begin{cases} \pi_{i\cdot}(1-\pi_{i\cdot})\mathbf{T}, & i = a, \\ -\pi_{i\cdot}\pi_{a\cdot}\mathbf{T}, & i \neq a, \end{cases} \\
 &= \frac{n^2}{n-1} \mathbf{S} \otimes \mathbf{T}
 \end{aligned}$$

where \mathbf{S} and \mathbf{T} are respectively the $I \times I$ and $J \times J$ matrices given by

$$\mathbf{S} = [diag(\pi_{i\cdot}) - (\pi_{i\cdot}\pi_{a\cdot})] \tag{2.5}$$

$$\mathbf{T} = [diag(\pi_{\cdot j}) - (\pi_{\cdot j}\pi_{\cdot b})]. \tag{2.6}$$

Example 2.3. Suppose we have two variables, sex and socio-economic class. We choose 150 males and 100 females such that we have 50 individuals classified as upper class, 75 middle class and 125 lower class. Then, the 2×3 contingency table for the variables sex and socio-economic class has a permutation model.

2.3.4 The link between the three models

While the cell counts have different distributions under the three models, the following theorem links the three models.

Theorem 2.5. Under the hypothesis of independence of the marginal variables, the distribution of the entries in a two-way contingency table does not involve the marginal parameters, $\{\pi_{i\cdot}\}$ and $\{\pi_{\cdot j}\}$ in the unrestricted bivariate sampling model.

Under the hypothesis of homogeneity, $\pi_{ij} = \pi_{\cdot j}$, the distribution of the entries in a two-way contingency table does not involve the marginal parameters $\{\pi_{\cdot j}\}$ in the product-multinomial model.

Proof. For the unrestricted bivariate sampling model, let us condition on both sets of marginal totals.

$$\begin{aligned}
P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{N_{i\cdot} = n_{i\cdot}\}, \{N_{\cdot j} = n_{\cdot j}\}, n) &= \frac{P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, n)}{P(\{N_{i\cdot} = n_{i\cdot}\} | \boldsymbol{\pi}, n) P(\{N_{\cdot j} = n_{\cdot j}\} | \boldsymbol{\pi}, n)} \\
&= \frac{\binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}}}{\left[\binom{n}{n_{1\cdot}, \dots, n_{I\cdot}} \prod_{i=1}^I \pi_{i\cdot}^{n_{i\cdot}} \right] \left[\binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}} \prod_{j=1}^J \pi_{\cdot j}^{n_{\cdot j}} \right]} \\
&= \frac{\binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}}}{\binom{n}{n_{1\cdot}, \dots, n_{I\cdot}} \binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}} \left[\prod_{i=1}^I \pi_{i\cdot}^{n_{i\cdot}} \right] \left[\prod_{j=1}^J \pi_{\cdot j}^{n_{\cdot j}} \right]} \\
P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{N_{i\cdot} = n_{i\cdot}\}, \{N_{\cdot j} = n_{\cdot j}\}, n) &= \frac{\left[\prod_{i=1}^I n_{i\cdot}! \right] \left[\prod_{j=1}^J n_{\cdot j}! \right]}{n! \prod_{i=1}^I \prod_{j=1}^J n_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \left[\frac{\pi_{ij}}{\pi_{i\cdot} \pi_{\cdot j}} \right]^{n_{ij}}
\end{aligned}$$

Under the hypothesis of independence, $\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$, and so the probability is now equal to (2.4), which does not involve the marginal parameters $\{\pi_{i\cdot}\}$ and $\{\pi_{\cdot j}\}$.

For the row-multinomial model, let us condition on the column totals.

$$\begin{aligned}
P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{n_{i\cdot}\}, \{N_{\cdot j} = n_{\cdot j}\}, n) &= \frac{P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{n_{i\cdot}\}, n)}{P(\{N_{\cdot j} = n_{\cdot j}\} | \boldsymbol{\pi}, n)} \\
&= \frac{\prod_{i=1}^I \left[\binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} \prod_{j=1}^J \pi_{ij}^{n_{ij}} \right]}{\binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}} \prod_{j=1}^J \pi_{\cdot j}^{n_{\cdot j}}} \\
&= \frac{\prod_{i=1}^I \binom{n_{i\cdot}}{n_{i1}, \dots, n_{iJ}} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}}}{\binom{n}{n_{\cdot 1}, \dots, n_{\cdot J}} \left[\prod_{j=1}^J \pi_{\cdot j}^{n_{\cdot j}} \right]} \\
P(\{N_{ij} = n_{ij}\} | \boldsymbol{\pi}, \{n_{i\cdot}\}, \{N_{\cdot j} = n_{\cdot j}\}, n) &= \frac{\left[\prod_{i=1}^I n_{i\cdot}! \right] \left[\prod_{j=1}^J n_{\cdot j}! \right]}{n! \prod_{i=1}^I \prod_{j=1}^J n_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \left[\frac{\pi_{ij}}{\pi_{\cdot j}} \right]^{n_{ij}}
\end{aligned}$$

Under the hypothesis of homogeneity, $\pi_{ij} = \pi_{\cdot j}$, and so the probability is now equal to (2.4) which does not involve the column parameters $\{\pi_{\cdot j}\}$. \square

This theorem, presented by Lancaster[27], tell us that under the appropriate hypothesis and given both sets of marginal totals, the cell counts under the three models have the same conditional probability,

$$P(\{N_{ij} = n_{ij}\} | \{N_{i\cdot} = n_{i\cdot}\}, \{N_{\cdot j} = n_{\cdot j}\}, n) = \frac{\left(\prod_{i=1}^I n_{i\cdot}! \right) \left(\prod_{j=1}^J n_{\cdot j}! \right)}{n! \prod_{i=1}^I \prod_{j=1}^J n_{ij}!}. \quad (2.7)$$

In fact, if we condition on both sets of marginal probabilities, we have the same result.

2.4 Chi-Squared Hypothesis Tests

We consider test statistics which asymptotically have a χ^2 distribution under the null hypothesis.

2.4.1 The chi-squared goodness-of-fit test

A goodness-of-fit test is used to decide if a sample comes from a specified distribution. This is done by verifying if the empirical distribution and the specified distribution give the same probabilities for a fixed number of sets.

The most widely used chi-squared test for goodness-of-fit, was introduced by Karl Pearson[32] in 1900.

Let X be a random variable, defined on the outcome space Ξ , with unknown cumulative distribution function F . Let X_1, \dots, X_n be a random sample of X . The null hypothesis for a chi-squared test of goodness-of-fit is

$$H_0 : F(x) = F_0(x) \quad \text{for all } x \in \mathbb{R}, \quad (2.8)$$

where F_0 is a completely specified distribution.

The data is grouped in the following way: let

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Xi \right\}$$

be a pre-specified partition of Ξ . Let N_j denote the random variable that counts the number of observations falling in the j -th category B_j :

$$N_j = \sum_{k=1}^n I[X_k \in B_j], \quad j = 1, \dots, J.$$

Then we have the following $1 \times J$ contingency table

	B_1	\dots	B_j	\dots	B_J	Total
X	N_1	\dots	N_j	\dots	N_J	n
Total	N_1	\dots	N_j	\dots	N_J	n

This contingency table falls under the row-multinomial model with $I = 1$, cell counts as $N_{1j} = N_j$, cell probabilities $\pi_{1j} = \pi_j = P(X \in B_j)$ which satisfy $\pi_1 + \dots + \pi_J = 1$, and a fixed (row) total of n .

Now, the null hypothesis for a chi-squared test of goodness-of-fit is equivalent to

$$H'_0 : \pi_j = \pi_{j0}, \quad j = 1, \dots, J. \quad (2.9)$$

where π_{j0} is the probability of the j -th cell under the distribution F_0 .

Since the hypothesis H'_0 is more restrictive than the hypothesis H_0 , if the hypothesis H'_0 is rejected then it is natural to reject the hypothesis H_0 . If the hypothesis H'_0 is not rejected, then we can say that the grouped data do not contradict the hypothesis H_0 .

The Pearson chi-squared statistic

Given that the above $1 \times J$ contingency table falls under the row-multinomial model, the random vector of the cell counts $\mathbf{N} = (N_1, \dots, N_J)'$ has a *multinomial*($n, \boldsymbol{\pi}$) distribution where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)'$.

Pearson suggested the following, the *Pearson chi-squared statistic*, as a statistic to test against the alternative hypothesis $H'_1 : \pi_j \neq \pi_{j0}$

$$X_P^2 = \sum_{j=1}^J \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}},$$

where $\hat{\mu}_{j0}$ is the ML estimator for $\mu_j = n\pi_j$ under $H'_0 : \pi_j = \pi_{j0}, j = 1, \dots, J$. This statistic looks at the difference between the observed counts and their expected value under H'_0 .

The expected value of N_j under H'_0 is $\hat{\mu}_{j0} = n\pi_{j0}$. Thus, the Pearson statistic for goodness-of-fit is

$$\begin{aligned} X_P^2 &= \sum_{j=1}^J \frac{(N_j - n\pi_{j0})^2}{n\pi_{j0}} \\ &= \sum_{j=1}^J \frac{N_j^2 - 2n\pi_{j0}N_j + (n\pi_{j0})^2}{n\pi_{j0}} \\ &= \sum_{j=1}^J \frac{N_j^2}{n\pi_{j0}} - 2 \sum_{j=1}^J N_j + n \sum_{j=1}^J \pi_{j0} \\ X_P^2 &= \sum_{j=1}^J \frac{N_j^2}{n\pi_{j0}} - n \end{aligned}$$

Theorem 2.6. *Under the null hypothesis*

$$H'_0 : \pi_j = \pi_{j0}, \quad j = 1, \dots, J,$$

the Pearson statistic X_P^2 asymptotically has a $\chi^2(J-1)$ distribution as $n \rightarrow \infty$.

Proof. The proof is presented in Kendall & Stuart[26]. □

The likelihood ratio test statistic

An alternative statistic for testing the above hypotheses has been derived by looking at this problem as a two-sided likelihood ratio test (LRT).

From subsection 2.3.2, the likelihood function for $\boldsymbol{\pi}$ is

$$L(\boldsymbol{\pi} | \{n_j\}, n) = \binom{n}{n_1, \dots, n_J} \prod_{j=1}^J \pi_j^{n_j}.$$

and the unrestricted ML estimators for the probabilities π_j are

$$\hat{\pi}_j = \frac{N_j}{n} \quad j = 1, \dots, J.$$

The ML estimators for the probabilities π_j under H'_0 are $\hat{\pi}_{j0} = \pi_{j0}, j = 1, \dots, J$. Therefore, the two-sided LRT statistic is

$$\Lambda(\mathbf{N}, n) = \frac{L(\hat{\pi}_{j0} | \mathbf{N}, n)}{L(\hat{\pi}_j | \mathbf{N}, n)} = \frac{\binom{n}{N_1, \dots, N_J} \prod_{j=1}^J \pi_{j0}^{N_j}}{\binom{n}{N_1, \dots, N_J} \prod_{j=1}^J \left(\frac{N_j}{n}\right)^{N_j}} = \prod_{j=1}^J \left(\frac{n\pi_{j0}}{N_j}\right)^{N_j}.$$

Wilks[49] showed that if the null hypothesis is true, $-2 \ln \Lambda$ has a chi-squared distribution as $n \rightarrow \infty$. The degrees of freedom are $\nu = \dim \Theta - \dim \Theta_0$, where Θ is the parameter space and Θ_0 is the parameter space under the null hypothesis.

The alternative statistic, called the *chi-squared likelihood ratio test statistic*, is

$$G^2 = -2 \ln \Lambda.$$

So, the chi-squared LRT statistic for goodness-of-fit is

$$G^2 = -2 \ln \Lambda = 2 \sum_{j=1}^J N_j \ln \left(\frac{N_j}{n\pi_{j0}} \right).$$

The parameter space is

$$\Theta = \{(\pi_1, \dots, \pi_J)' \in [0, 1]^J : \pi_1 + \dots + \pi_J = 1\}.$$

So $\dim \Theta = J - 1$ because of the relation $\sum \pi_j = 1$. Under $H_0' : \pi_j = \pi_{j0}$, the probabilities are fixed so $\dim \Theta_0 = 0$ and the chi-squared LRT statistic G^2 asymptotically has a $\chi^2(J - 1)$ as $n \rightarrow \infty$.

Comparison of the likelihood ratio test statistic and Pearson chi-squared statistic

When the null hypothesis of goodness-of-fit is true, both the Pearson statistic and the chi-squared LRT statistic asymptotically have a $\chi^2(J - 1)$.

Theorem 2.7. *Under the same null hypothesis, the statistics X_P^2 and G^2 are asymptotically equivalent.*

Proof. Let $\hat{\mu}_{j0}$ be the ML estimator of $\mu_j = E(N_j)$ under the null hypothesis. Then, the chi-squared LRT statistic is

$$\begin{aligned} G^2 &= 2 \sum_{j=1}^J N_j \ln \left(\frac{N_j}{\hat{\mu}_{j0}} \right) \\ &= 2 \sum_{j=1}^J (\hat{\mu}_{j0} + N_j - \hat{\mu}_{j0}) \ln \left(\frac{\hat{\mu}_{j0} + N_j - \hat{\mu}_{j0}}{\hat{\mu}_{j0}} \right) \\ &= 2 \sum_{j=1}^J (\hat{\mu}_{j0} + N_j - \hat{\mu}_{j0}) \ln \left(1 + \frac{N_j - \hat{\mu}_{j0}}{\hat{\mu}_{j0}} \right) \end{aligned}$$

Let $g(x) = \ln(1 + x)$. Then, the k -th derivative for g is

$$g^{(k)}(x) = (-1)^{k-1} \frac{(k-1)!}{(1+x)^k} \quad k = 1, 2, \dots$$

The Taylor expansion for g around a constant c is

$$T_g(x, c) = \sum_{i=1}^{\infty} g^{(i)}(c) \frac{(x-c)^i}{i!} = \sum_{i=1}^k \frac{(-1)^{i-1}}{i} \left(\frac{x-c}{1+c} \right)^i.$$

For $c = 0$,

$$T_g(x, 0) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} x^i = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Then the Taylor expansion around $c = 0$ for the j -th term of G^2 is

$$\begin{aligned}
(\hat{\mu}_{j0} + N_j - \hat{\mu}_{j0}) T_g \left(\frac{(N_j - \hat{\mu}_{j0})}{\hat{\mu}_{j0}}, 0 \right) &= (\hat{\mu}_{j0} + N_j - \hat{\mu}_{j0}) \left[\frac{(N_j - \hat{\mu}_{j0})}{\hat{\mu}_{j0}} - \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}^2} + \frac{1}{3} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^3} - \dots \right] \\
&= \hat{\mu}_{j0} \left[\frac{(N_j - \hat{\mu}_{j0})}{\hat{\mu}_{j0}} - \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}^2} + \frac{1}{3} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^3} - \dots \right] \\
&+ (N_j - \hat{\mu}_{j0}) \left[\frac{(N_j - \hat{\mu}_{j0})}{\hat{\mu}_{j0}} - \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}^2} + \frac{1}{3} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^3} - \dots \right] \\
&= (N_j - \hat{\mu}_{j0}) - \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}} + \frac{1}{3} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^2} - \dots \\
&+ \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}} - \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^2} + \frac{1}{3} \frac{(N_j - \hat{\mu}_{j0})^4}{\hat{\mu}_{j0}^3} - \dots \\
&= (N_j - \hat{\mu}_{j0}) + \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}} - \frac{1}{6} \frac{(N_j - \hat{\mu}_{j0})^3}{\hat{\mu}_{j0}^2} \dots \\
&= (N_j - \hat{\mu}_{j0}) + \frac{1}{2} \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}} + O \left[(N_j - \hat{\mu}_{j0})^3 \right]
\end{aligned}$$

Since under the null hypothesis $\sum_{j=1}^J (N_j - \hat{\mu}_{j0}) = 0$, as $n \rightarrow \infty$,

$$G^2 \approx \sum_{j=1}^J \frac{(N_j - \hat{\mu}_{j0})^2}{\hat{\mu}_{j0}} = X_P^2$$

□

Since the proof is shown without the use of explicit form of ML estimator $\hat{\mu}_{j0}$, X_P^2 and G^2 are always asymptotically equivalent under the same null hypothesis.

2.4.2 The chi-squared test for homogeneity

The objective of the chi-squared test for homogeneity is to decide if two or more samples come from the same population by determining if they give the same probabilities for a fixed number of sets.

Let X_1, \dots, X_I be random variables, defined on the same outcome space Ξ which has the partition

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Xi \right\}$$

with unknown cdf F_1, \dots, F_I , respectively. Let X_{i1}, \dots, X_{in_i} be a random sample of F_i , $i = 1, \dots, I$.

We set

$$\pi_{ij} = P_i(X_i \in B_j) \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

where P_i is the probability function associated with F_i , $i = 1, \dots, I$.

The data is grouped in the following way: let N_{ij} denote the number of observations in the i -th sample fall in the j -th category

$$N_{ij} = \sum_{k=1}^J 1[X_{ik} \in B_j] \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Then, we have an $I \times J$ contingency table as per Table 2.2.

The null hypothesis for a chi-squared test of homogeneity of the random variables is

$$H_0 : F_1(x) = \cdots = F_I(x) \quad \text{for all } x \in \mathbb{R}, \quad (2.10)$$

which is equivalent to

$$H'_0 : \pi_{1j} = \cdots = \pi_{Ij} = \pi_{.j}, \quad j = 1, \dots, J. \quad (2.11)$$

Since the hypothesis H'_0 is more restrictive than the hypothesis H_0 , if the hypothesis H'_0 is rejected then it is natural to reject the hypothesis H_0 . If the hypothesis H'_0 is not rejected, then we can say that the grouped data do not contradict the hypothesis H_0 .

The Pearson chi-squared statistic

Given that the row totals are fixed, the contingency table falls under the row-multinomial model and the i -th row \mathbf{N}_i has a *multinomial*($n_i, \boldsymbol{\pi}_i$) distribution, $i = 1, \dots, I$.

The Pearson statistic to test H'_0 against hypothesis the $H'_1 : \pi_{ij} \neq \pi_{.j}$ for some j , is

$$X_P^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - \hat{\mu}_{ij}^{(0)})^2}{\hat{\mu}_{ij}^{(0)}}$$

where $\hat{\mu}_{ij}^{(0)}$ is the ML estimator for $\mu_{ij} = n_i \pi_{ij}$ under H'_0 , $i = 1, \dots, I$, $j = 1, \dots, J$. We need to find the ML estimators $\hat{\pi}_{ij}^{(0)}$ under H'_0 .

Under the row-multinomial model, the log-likelihood function for $\boldsymbol{\pi}$ is

$$l(\boldsymbol{\pi} | \{n_{ij}\}, \{n_i\}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij}.$$

Under H'_0 , the log-likelihood is

$$l_0(\boldsymbol{\pi} | \{n_{ij}\}, \{n_i\}, n) \propto \sum_{j=1}^J n_{.j} \ln \pi_{.j}.$$

Under the constraint $\pi_{.1} + \cdots + \pi_{.J} = 1$, the Lagrange function under H'_0 is

$$L_0(\{n_{ij}\}, \lambda | \{\pi_{.j}\}, n) \propto \sum_{j=1}^J n_{.j} \ln \pi_{.j} - \lambda \left(\sum_{j=1}^J \pi_{.j} - 1 \right).$$

To find the ML estimator for $\pi_{.j}$, we maximise the Lagrange function:

$$\frac{\partial}{\partial \pi_{.j}} L_0(\{n_{ij}\}, \lambda | \{\pi_{.j}\}, n) = 0 \Leftrightarrow \pi_{.j} = \frac{n_{.j}}{\lambda}.$$

Substituting into the constraint yields $(n_{.1} + \cdots + n_{.J}) / \lambda = 1$, so $\lambda = n$ and the ML estimators for the marginal column probabilities $\pi_{.j}$ under H'_0 are

$$\hat{\pi}_{.j}^{(0)} = \frac{N_{.j}}{n} \quad j = 1, \dots, J.$$

Then, the ML estimators for the cell probabilities π_{ij} under $H'_0 : \pi_{1j} = \dots = \pi_{Ij} = \pi_{.j}$ are

$$\hat{\pi}_{ij}^{(0)} = \hat{\pi}_{.j}^{(0)} = \frac{N_{.j}}{n} \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

and the ML estimators for the cell means μ_{ij} under H'_0 are

$$\hat{\mu}_{ij}^{(0)} = \frac{n_{i.}N_{.j}}{n}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Thus, the Pearson statistic for homogeneity is

$$\begin{aligned} X_P^2 &= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J \frac{(nN_{ij} - n_{i.}N_{.j})^2}{n_{i.}N_{.j}} \\ &= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J \frac{n^2 N_{ij}^2 - 2nn_{i.}N_{.j}N_{ij} \cdot (n_{i.}N_{.j})^2}{n_{i.}N_{.j}} \\ &= n \sum_{i=1}^I \sum_{j=1}^J \frac{N_{ij}^2}{n_{i.}N_{.j}} - 2 \sum_{i=1}^I \sum_{j=1}^J N_{ij} \cdot \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J n_{i.}N_{.j} \\ X_P^2 &= n \left(\sum_{i=1}^I \sum_{j=1}^J \frac{N_{ij}^2}{n_{i.}N_{.j}} - 1 \right) \end{aligned}$$

Theorem 2.8. *Under the hypothesis,*

$$H'_0 : \pi_{1j} = \dots = \pi_{Ij} = \pi_{.j}, \quad j = 1, \dots, J.$$

the Pearson statistic X_P^2 has an approximately chi-squared distribution with $(I-1)(J-1)$ degrees of freedom as $n_{i.} \rightarrow \infty$, $i = 1, \dots, I$, and $n \rightarrow \infty$.

Proof. Under H'_0 , by the same approach as for Theorem 2.6, we obtain that for each row,

$$X_i^2 = \sum_{j=1}^J \frac{(N_{ij} - n_{i.}N_{.j})^2}{n_{i.}N_{.j}} \quad i = 1, \dots, I$$

approximately has a $\chi^2(J-1)$ distribution as $n_{i.} \rightarrow \infty$, $i = 1, \dots, I$. Since only $I-1$ of the rows are independent under H'_0 , $X_P^2 = \sum_{i=1}^I X_i^2$ approximately has a $\chi^2(I-1)(J-1)$ distribution as $n_{i.} \rightarrow \infty$, $i = 1, \dots, I$ and $n \rightarrow \infty$. \square

The likelihood ratio test statistic

Under the row-multinomial model, the likelihood function for $\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{IJ})'$ is

$$L(\boldsymbol{\pi} | \{n_{ij}\}, \{n_{i.}\}, n) = \prod_{i=1}^I \left[\binom{n_{i.}}{n_{i1}, \dots, n_{iJ}} \prod_{j=1}^J \pi_{ij}^{n_{ij}} \right],$$

and the unrestricted ML estimators for the probabilities π_{ij} are

$$\hat{\pi}_{ij} = \frac{N_{ij}}{n} \quad i = 1, \dots, I \quad j = 1, \dots, J.$$

As we saw, $\hat{\pi}_{ij}^{(0)} = \hat{\pi}_{.j}^{(0)} = N_{.j}/n$, $i = 1, \dots, I$ and $j = 1, \dots, J$. Therefore, the two-sided LRT statistic is

$$\Lambda(\mathbf{N}, n) = \frac{L(\hat{\pi}_{ij}^{(0)} | \mathbf{N}, n)}{L(\hat{\pi}_{ij} | \mathbf{N}, n)} = \frac{\prod_{i=1}^I \left[\binom{N_{i.}}{N_{i1}, \dots, N_{iJ}} \prod_{j=1}^J \left(\frac{N_{.j}}{n} \right)^{N_{ij}} \right]}{\prod_{i=1}^I \left[\binom{N_{i.}}{N_{i1}, \dots, N_{iJ}} \prod_{j=1}^J \left(\frac{N_{ij}}{N_{i.}} \right)^{N_{ij}} \right]} = \prod_{i=1}^I \prod_{j=1}^J \left(\frac{N_{.j}}{N_{ij}} \right)^{N_{ij}}$$

and the chi-squared LRT statistic for homogeneity is

$$G^2 = -2 \ln \Lambda = 2 \sum_{i=1}^I \sum_{j=1}^J N_{ij} \ln \left(\frac{N_{ij}}{N_{.j}} \right).$$

The parameter space is

$$\Theta = \left\{ (\pi_{11}, \dots, \pi_{IJ})' \in [0, 1]^{IJ} : \sum_{j=1}^J \pi_{ij} = 1 \quad i = 1, \dots, I \right\}$$

and $\dim \Theta = I(J - 1)$ because of the relation $\pi_{i.} = 1 - \sum_{j=1}^J \pi_{ij}$ for the I rows. The parameter space under H'_0 is

$$\Theta_0 = \left\{ (\pi_{.1}, \dots, \pi_{.J}) \in [0, 1]^J : \sum_{j=1}^J \pi_{.j} = 1 \right\}$$

and $\dim \Theta_0 = J - 1$ because of the relation $\pi_{.J} = 1 - \sum_{j=1}^{J-1} \pi_{.j}$. So under H'_0 , the chi-squared LRT statistic G^2 has a chi-squared distribution with $I(J - 1) - (J - 1) = (I - 1)(J - 1)$ degrees of freedom as $n_i \rightarrow \infty$, $i = 1, \dots, I$, and $n \rightarrow \infty$.

2.4.3 The chi-squared test for independence

The objective of the chi-squared test of independence is to decide if two samples are independent by comparing the observations to the expected values under the hypothesis of independence.

Let X be a random variable, defined on the outcome space Ξ which has partition

$$A = \left\{ A_1, \dots, A_I : A_i \cap A_j = \emptyset, \quad i \neq j = 1, \dots, I, \quad \bigcup A_i = \Xi \right\}.$$

Similarly, let Y random variable, defined on the outcome space Ψ which has partition

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Psi \right\}.$$

The null hypothesis of independence between X and Y is

$$H_0 : F(x, y) = F_X(x) F_Y(y) \quad \text{for all } (x, y) \in \mathbb{R}^2 \quad (2.12)$$

where F is the joint cdf, F_X and F_Y are the cdf for X and Y , respectively.

Suppose $(X_1, Y_1), \dots, (X_n, Y_n)$ is a random sample of the random vector (X, Y) . Let N_{ij} denote the random variable that counts the number of observations that fall in the (i, j) -th cross-category:

$$N_{ij} = \sum_{k=1}^n 1[X_k \in A_i, Y_k \in B_j] \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Then we have an $I \times J$ contingency table as per Table 2.1.

We set

$$\pi_{ij} = P(X \in A_i, Y \in B_j) \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

where P is the probability function associated with the joint cdf F . If the hypothesis H_0 is true, then we would have $\pi_{ij} = \pi_i \cdot \pi_j$ for all i and j . Thus a corresponding null hypothesis is

$$H'_0 : \pi_{ij} = \pi_i \cdot \pi_j \quad i = 1, \dots, I \quad j = 1, \dots, J. \quad (2.13)$$

Since the hypothesis H'_0 is narrower than the hypothesis H_0 , if the hypothesis H'_0 is rejected then it is natural to reject the hypothesis H_0 . If the hypothesis H'_0 is not rejected, then we can say that the grouped data do not contradict the hypothesis H_0 .

The Pearson chi-squared statistic

Given that only the total number of observations is fixed, the contingency table falls under the unrestricted bivariate sampling model and so the random vector of the cell counts $\mathbf{N} = (N_{11}, \dots, N_{IJ})'$ has a *multinomial*($n, \boldsymbol{\pi}$) such that $\pi_{11} + \dots + \pi_{IJ} = 1$.

The Pearson statistic to test H'_0 against $H'_1 : \pi_{ij} \neq \pi_i \cdot \pi_j$ for some (i, j) , is

$$X_P^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - \hat{\mu}_{ij}^{(0)})^2}{\hat{\mu}_{ij}^{(0)}}$$

where $\hat{\mu}_{ij}^{(0)}$ is the ML estimator for $\mu_{ij} = n\pi_{ij}$ under H'_0 , $i = 1, \dots, I$, $j = 1, \dots, J$. We need to find the ML estimators $\hat{\pi}_{ij}^{(0)}$ under $H'_0 : \pi_{ij} = \pi_i \cdot \pi_j$.

Under the unrestricted bivariate sampling model, the log-likelihood function for $\boldsymbol{\pi}$ is

$$l(\boldsymbol{\pi} | \{n_{ij}\}, n) \propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln \pi_{ij}.$$

Under H'_0 , the log-likelihood is

$$l_0(\boldsymbol{\pi} | \{n_{ij}\}, n) \propto \sum_{i=1}^I n_{i.} \ln \pi_i + \sum_{j=1}^J n_{.j} \ln \pi_j$$

Under the constraints $\pi_1 + \dots + \pi_I = 1$ and $\pi_{.1} + \dots + \pi_{.J} = 1$, the Lagrange function under H'_0 is

$$L_0(\{n_{ij}\}, \boldsymbol{\lambda} | \{\pi_i\}, \{\pi_{.j}\}, n) \propto \sum_{i=1}^I n_{i.} \ln \pi_i + \sum_{j=1}^J n_{.j} \ln \pi_{.j} - \lambda_1 \left(\sum_{i=1}^I \pi_i - 1 \right) - \lambda_2 \left(\sum_{j=1}^J \pi_{.j} - 1 \right).$$

To find the ML estimator for π_i , we maximise the Lagrange function:

$$\frac{\partial}{\partial \pi_i} L_0(\{n_{ij}\}, \boldsymbol{\lambda} | \{\pi_i\}, \{\pi_{.j}\}, n) = 0 \Leftrightarrow \pi_i = \frac{n_{i.}}{\lambda_1}.$$

Substituting into the first constraint yields $(n_{1.} + \dots + n_{I.}) / \lambda_1 = 1$, so $\lambda_1 = n$ and the ML estimators for the marginal row probabilities π_i under H'_0 are

$$\hat{\pi}_i^{(0)} = \frac{N_{i.}}{n} \quad i = 1, \dots, I.$$

Similarly, $\lambda_2 = n$ and the ML estimators for the marginal column probabilities $\pi_{.j}$ under H'_0 are

$$\hat{\pi}_{.j}^{(0)} = \frac{N_{.j}}{n} \quad j = 1, \dots, J.$$

Then, the ML estimators for the cell probabilities π_{ij} under H'_0 are

$$\hat{\pi}_{ij}^{(0)} = \hat{\pi}_i^{(0)} \hat{\pi}_{.j}^{(0)} = \frac{N_{i.} N_{.j}}{n} \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

and the ML estimators for the cell means μ_{ij} under H'_0 are

$$\hat{\mu}_{ij}^{(0)} = n \hat{\pi}_{ij}^{(0)} = \frac{N_{i.} N_{.j}}{n} \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Thus, the Pearson statistic for independence is

$$\begin{aligned} X_P^2 &= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J \frac{(nN_{ij} - N_{i.}N_{.j})^2}{N_{i.}N_{.j}} \\ &= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J \frac{(nN_{ij})^2 - 2nN_{ij}N_{i.}N_{.j} + t(N_{i.}N_{.j})^2}{N_{i.}N_{.j}} \\ &= n \sum_{i=1}^I \sum_{j=1}^J \frac{N_{ij}^2}{N_{i.}N_{.j}} - 2 \sum_{i=1}^I \sum_{j=1}^J N_{ij} + \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J N_{i.}N_{.j} \\ X_P^2 &= n \left(\sum_{i=1}^I \sum_{j=1}^J \frac{N_{ij}^2}{N_{i.}N_{.j}} - 1 \right) \end{aligned}$$

Theorem 2.9. *Under the hypothesis*

$$H'_0 : \pi_{ij} = \pi_i \pi_{.j} \quad i = 1, \dots, I \quad j = 1, \dots, J,$$

the Pearson statistic X_P^2 asymptotically has as a $\chi^2(I-1)(J-1)$ as $n_{i.} \rightarrow \infty$, $i = 1, \dots, I$, and $n \rightarrow \infty$.

Proof. Similar to the Pearson statistic for the test of homogeneity except with a vector of length IJ . \square

The likelihood ratio test statistic

Under the unrestricted bivariate sampling model, likelihood function for $\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{IJ})'$ is

$$L(\boldsymbol{\pi} | \{n_{ij}\}, n) = \binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}},$$

and the unrestricted ML estimators for the cell probabilities π_{ij} are

$$\hat{\pi}_{ij} = \frac{N_{ij}}{n} \quad i = 1, \dots, I \quad j = 1, \dots, J.$$

As we saw, $\hat{\pi}_{ij}^{(0)} = \hat{\pi}_i^{(0)} \hat{\pi}_{.j}^{(0)} = N_{i.}N_{.j}/n^2$. for all $i = 1, \dots, I$ and $j = 1, \dots, J$. Therefore, the two-sided LRT statistic is

$$\Lambda(\mathbf{N}, n) = \frac{L(\hat{\pi}_{ij}^{(0)} | \mathbf{N}, n)}{L(\hat{\pi}_{ij} | \mathbf{N}, n)} = \frac{\binom{n}{N_{11}, \dots, N_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \left(\frac{N_{i.}N_{.j}}{n^2}\right)^{N_{ij}}}{\binom{n}{N_{11}, \dots, N_{IJ}} \prod_{i=1}^I \prod_{j=1}^J \left(\frac{N_{ij}}{n}\right)^{N_{ij}}} = \prod_{i=1}^I \prod_{j=1}^J \left(\frac{N_{i.}N_{.j}}{nN_{ij}}\right)^{N_{ij}}$$

and the chi-squared LRT statistic for independence is

$$G^2 = -2 \ln \Lambda = 2 \sum_{i=1}^I \sum_{j=1}^J N_{ij} \ln \left(\frac{n N_{ij}}{N_{i.} N_{.j}} \right).$$

The parameter space is

$$\Theta = \left\{ (\pi_{11}, \dots, \pi_{IJ})' \in [0, 1]^{IJ} : \sum_{i=1}^I \sum_{j=1}^J \pi_{ij} = 1 \right\}$$

and $\dim \Theta = IJ - 1$ because of the relation $\pi_{IJ} = 1 - \sum \pi_{ij}$. The parameter space under H'_0 is

$$\Theta_0 = \left\{ (\pi_{1.}, \pi_{.1}, \dots, \pi_{I.}, \pi_{.J})' \in [0, 1]^{IJ} : \sum_{i=1}^I \pi_{i.} = 1 \quad \sum_{j=1}^J \pi_{.j} = 1 \right\}$$

and $\dim \Theta_0 = (I - 1) + (J - 1)$ because of the relations $\pi_{I.} = 1 - \sum \pi_{i.}$ and $\pi_{.J} = 1 - \sum \pi_{.j}$. So under H'_0 , the chi-squared LRT statistic G^2 has a chi-squared distribution with $IJ - 1 - (I - 1) - (J - 1) = (I - 1)(J - 1)$ degrees of freedom as $n \rightarrow \infty$.

Chapter 3

Tools for the Decomposition of the Pearson Statistic

In this chapter, we first discuss why sometimes the Pearson and likelihood ratio tests are not appropriate to use. We follow to give the three major concepts that we will rely on for the decomposition of the Pearson statistic: ranked data, smooth models and orthonormal polynomials with respect to a distribution.

3.1 Ordinal Categories and Testing for Trends

Chi-squared tests of independence merely indicate the degree of evidence of association between the row and column variables/classifications. One major fault of the Pearson and LRT chi-squared tests of independence is that the expected cell counts only depend on the marginal totals. The order of the categories of the rows and the columns is not taken into account. Thus Pearson and LRT statistics treat both the row and column variables/classifications as being nominal. If the chi-squared tests are applied when at least one variable/classification is ordinal, this information is ignored and may lead to a false decision.

Example 3.1. (From Agresti[1]) In the following table the variables are income and job satisfaction, measured for the black males in a national (U.S.A.) sample. Both classifications are ordinal, with the categories very dissatisfied (VD), little dissatisfied (LD), moderately satisfied (MS) and very satisfied (VS)

Job Satisfaction					
Income (USD)	VD	LD	MS	VS	Total
< 15,000	6	14	1	12	33
15,000 – 25,000	3	10	2	7	22
25,000 – 40,000	1	9	0	11	21
> 40,000	3	10	1	6	20
Total	13	43	4	36	96

The Pearson and LRT statistics for testing independence are $X_P^2 = 6.0$ and $G^2 = 6.8$ with 9 degrees of freedom and p -values are 0.74 and 0.66, respectively. The statistics show little evidence of association. However, we can permute the columns and rows to obtain the following table.

Income (USD)	Job Satisfaction				Total
	VD	LD	MS	VS	
< 15,000	1	3	10	6	20
15,000 – 25,000	2	3	10	7	22
25,000 – 40,000	1	6	14	12	33
> 40,000	0	1	9	11	21
Total	4	13	43	36	96

Since columns and rows of the second table are simply a permutation of the columns and rows from the first table, the cell counts are still being compared to the same estimated cell means and the Pearson and LRT statistics have the same value thus not rejecting the hypothesis of independence between income and job satisfaction, yet we can see there is an increasing trend.

In this thesis, we also look at chi-squared tests of homogeneity against monotonic trend. Let X_1, \dots, X_I be random variables, defined on the same outcome space Ξ which has the partition

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Xi \right\}$$

with unknown cdf F_1, \dots, F_I , respectively. We set

$$\pi_{ij} = P_i(X_i \in B_j) \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

where P_i is the probability function associated with F_i , $i = 1, \dots, I$. The null hypothesis for a chi-squared test of homogeneity of the random variables is

$$H_0 : F_1(x) = \dots = F_I(x) \text{ for all } x \in \mathbb{R}$$

which is equivalent to

$$H'_0 : \tau_{i1} = \dots = \tau_{Ij}, j = 1, \dots, J$$

where

$$\tau_{ij} = \begin{cases} \pi_{i1} + \dots + \pi_{ij} & j = 1, \dots, J-1, \\ 1 & j = J. \end{cases}$$

are the cumulative cell probabilities. The chi-squared test of homogeneity against monotonic trend tests H_0 against

$$H_1 : F_1(x) \leq \dots \leq F_I(x) \text{ for all } x \in \mathbb{R}$$

$$H_2 : F_1(x) \geq \dots \geq F_I(x) \text{ for all } x \in \mathbb{R}$$

$$\text{or } H_3 : F_1(x) \leq \dots \leq F_I(x) \text{ or } F_1(x) \geq \dots \geq F_I(x) \text{ for all } x \in \mathbb{R}$$

with at least one strict inequality in the alternatives. This is equivalent to testing H'_0 against

$$H'_1 : \tau_{i1} \leq \dots \leq \tau_{Ij} \quad j = 1, \dots, J, \tag{3.1}$$

$$H'_2 : \tau_{i1} \geq \dots \geq \tau_{Ij} \quad j = 1, \dots, J, \tag{3.2}$$

$$\text{or } H'_3 : \tau_{i1} \leq \dots \leq \tau_{Ij} \quad \text{or} \quad \tau_{i1} \geq \dots \geq \tau_{Ij} \quad j = 1, \dots, J \tag{3.3}$$

with at least one strict inequality in the alternative hypotheses. The alternative (3.1) is equivalent to a monotone increasing trend in the row distributions of a row-multinomial model, the alternative (3.2) is equivalent to a monotone decreasing trend, and the alternative (3.3) is equivalent to a monotonic trend. Once again, the Pearson and LRT chi-squared tests are invariant to the possible order of the columns when testing against the above directional alternatives and may result into a false decision.

Example 3.2. (From Bagdonavičius et al.[4]) Investigating the granular composition of quartz in Lithuanian and Saharan sand samples, the data on the lengths (in cm) of the maximal axis of the quartz grains were used. Using the granular composition of the samples, conclusions on the geological sand formation conditions were drawn. The data are grouped in intervals.

Interval midpoints	11	17	21	25	29	33	37	45	Total
Lithuanian Sand	16	35	61	52	23	7	4	3	201
Saharan Sand	6	10	12	13	12	15	12	22	102
Total	22	45	73	65	35	22	16	25	303

The Pearson and LRT statistics to verify the hypothesis that the length of the maximum axis has the same distribution in Lithuanian and Saharan sand are $X_P^2 = 75.2$ and $G^2 = 74.5$ with 7 degrees of freedom and p -values $< 10^{-12}$. The statistics show there is evidence of association. Looking at the table of cumulative cell probabilities,

Interval midpoints	11	17	21	25	29	33	37	45
Lithuanian Sand	0.0796	0.2537	0.5572	0.8159	0.9303	0.9652	0.9851	1.0000
Saharan Sand	0.0588	0.1569	0.2745	0.4020	0.5196	0.6667	0.7843	1.0000

an appropriate follow-up test would compare the hypothesis of homogeneity against decreasing trend in the row distributions.

For both these situations, ordinal variables or trend alternatives, alterations to the usual chi-squared tests have been made to increase the power of the test. Without loss of generality, we are interested in detecting monotone increasing trends.

3.2 Ranked Data and Contingency Tables

We consider nonparametric tests based on statistics which depend only on the location of observations in the ordered sample and not directly on their values.

3.2.1 Ranked data and their properties

Let X_1, \dots, X_n be a random sample of a random variable X and let $X_{(1)} \leq \dots \leq X_{(n)}$ be the order statistics.

If the distribution of X is absolutely continuous, i.e. the probability of having tied ranks is zero, and $X_{(1)} < \dots < X_{(n)}$, the rank R_i of X_i is the order number of X_i in the ordered sample $X_{(1)}, \dots, X_{(n)}$:

$$R_i = \sum_{j=1}^n jI [X_i = X_{(j)}].$$

Since ranks take the values $1, \dots, n$, the sum and the sum of squares of the ranks are constant

$$\sum_{i=1}^n R_i = \sum_{i=1}^n i = \frac{n(n+1)}{2}, \quad \sum_{i=1}^n R_i^2 = \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Then, if the distribution of the random variable X is absolutely continuous, then the ranks have probability $P(R = i) = 1/n$ and

$$E[R_i] = \frac{n+1}{2}, \quad \text{cov}[R_i, R_j] = \frac{n+1}{12} [n\mathbb{I}_n - \bar{\mathbf{1}}_n \bar{\mathbf{1}}_n'] \quad i, j = 1, \dots, n.$$

where \mathbb{I}_n is the $n \times n$ identity matrix and $\bar{\mathbf{1}}_n$ is the vector of n 1's.

The notion of rank can be generalized to the case where the distribution of the random variable X is not necessarily absolutely continuous, thus allowing for tied ranks. If there is a group of s coinciding at the j -th order statistic in the ordered sample $X_{(1)} \leq \dots \leq X_{(n)}$ and $X_i = X_{(j)}$ then R_i is defined as the arithmetic mean of the positions of coinciding observations:

$$R_i = \frac{j + (j+1) + \dots + (j+s-1)}{s} = \sum_{a=j}^{j+s-1} \frac{a}{s} = j + \sum_{a=1}^{s-1} \frac{a}{s} = j + \frac{s-1}{2}.$$

These adjusted ranks, called *mid-ranks*, have no effect on the mean of the ranks. Suppose $X_{i_1} = \dots = X_{i_s} = X_{(j)}$, then the sum of their ranks is

$$\sum_{b=1}^s R_{i_b} = \sum_{b=1}^s \left(j + \frac{s-1}{2} \right) = \sum_{b=1}^s \sum_{a=j}^{j+s-1} \frac{a}{s} = \sum_{a=j}^{j+s-1} a.$$

and the sum of all ranks remains the same.

Set

$$S = \sum_{l=1}^k s_l (s_l^2 - 1)$$

where k denotes the number of random groups with coinciding members in the ordered sample and s_l denotes the number of the members in the l -th group.

Theorem 3.1. *Conditional on the observed ties, the means and the covariances of the ranks are*

$$E[R_i] = \frac{n+1}{2} \quad \text{cov}[R_i, R_j] = \frac{n(n^2-1) - S}{12n(n-1)} [n\mathbb{I}_n - \mathbf{1}_n \mathbf{1}_n'] \quad i, j = 1, \dots, n.$$

Proof. The proof is provided by Bagdonavičius et al.[4] □

3.2.2 Advantages and disadvantages of working with ranked data

Conover[17] gives two reasons why one would prefer to work with ranks over the actual data:

- “If the numbers assigned to the observations have no meaning by themselves but rather attain meaning only in an ordinal comparison with other observations, then the number contains no more information than the ranks contain.”

- “Even if the numbers have meaning but the distribution function is not normal, the probability theory is beyond our reach when the statistic is based on the actual data. The probability theory based on ranks is relatively simple and does not depend on the distribution in many cases.”

Conover[17] defines a *nonparametric method* as follows:

Definition 3.1. *A statistical method is nonparametric if it satisfies at least one of the following criteria:*

- *The method may be used on data with a nominal scale of measurement.*
- *The method may be used on data with an ordinal scale of measurement.*
- *The method may be used on data with an interval or ratio scale of measurement, where the distribution function of the random variable producing the data is either unspecified or specified except for an infinite number of unknown parameters.*

Thus satisfying the second criteria, a statistical method using ranked data is considered nonparametric and has the following advantages:

- As given by the definition, nonparametric methods based on ranked data can be used on data with ordinal, interval or ratio scale of measurement.
- As they were originally developed before the wide use of computers, nonparametric methods are intuitive and are simple to carry out by hand, for small samples at least.
- There are no or very limited assumptions on the format of the data. A nonparametric method may be preferable when the assumptions required for a parametric method are not valid. For example,
 - the scale of the measurement: the Pearson correlation assumes data is at least interval while its nonparametric equivalent, the Spearman correlation, assumes the data is at least ordinal.
 - the underlying distribution: the one sample t-test requires observations to be drawn from a normally distributed population while its nonparametric equivalent, the Wilcoxon signed rank test, simply assumes the observations are drawn from the same population.

However, we need to acknowledge there are also some disadvantages to the use of nonparametric methods:

- If the sample size is too large, nonparametric methods may be difficult to compute by hand and unlike parametric methods, appropriate computer software for nonparametric methods are usually computationally intensive and can be limited, although the situation is improving.
- If the assumptions of the corresponding parametric method hold or the sample size is not large enough, nonparametric methods may lack power compared to their parametric equivalent.
- Tied values can be problematic when these are common, and adjustments to the test statistic may be necessary.

When the data follows a normal distribution, for example, the mean and standard deviation are all that is required to understand the distribution and make inference. With nonparametric methods, because there are only probabilities involved, reducing the data to a few numbers (e.g. the median) does not give an accurate picture. In this thesis, though there may be a large number of categories, with an ordinal variable and/or a trend, we reduce the amount of information necessary to better interpret the data.

3.3 Smooth Goodness-of-Fit Models

Neyman considered the problem of what constitutes an appropriate alternative when the null hypothesis is rejected in a goodness-of-fit test.

3.3.1 Neyman-type smooth models

Suppose we have n observations of a random variable U . We would like to test the null hypothesis that U has an underlying $U(0, 1)$ distribution against some alternative distribution which is close to uniform, i.e. *smooth*. The k -order alternative density function is given as

$$f(u|\boldsymbol{\theta}) = c(\boldsymbol{\theta}) \exp \left\{ \sum_{r=1}^k \theta_r p_r(u) \right\}, \quad 0 \leq u \leq 1$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$ is a vector of real parameters, the $\{p_r(u)\}$ are orthonormal polynomials with respect to the $U(0, 1)$ distribution, thus satisfying

$$p_0(y) = 1, \quad \int_0^1 p_r(y) p_s(u) du = \delta_{rs}, \quad r, s = 0, 1, \dots, k, \quad (3.4)$$

and $c(\boldsymbol{\theta})$ is a normalizing constant to make f a proper density. Let U_1, \dots, U_n be a random sample of the $U(0, 1)$ distribution. Testing the null hypothesis that the underlying distribution is uniform is equivalent to testing $H_0'' : \boldsymbol{\theta} = \vec{0}_k$. Neyman[31] derived the test statistic

$$\psi_k^2 = \sum_{r=1}^k V_r^2 \quad \text{where} \quad V_r = \frac{1}{\sqrt{n}} \sum_{j=1}^n p_r(U_j), \quad r = 1, \dots, k.$$

The statistics V_r are such that

$$E[V_r] = 0, \quad E[V_r^2] = 1, \quad r = 1, \dots, k.$$

Thomas & Pierce[47] show that as $n \rightarrow \infty$, ψ_k^2 asymptotically has a $\chi^2(k)$ distribution. The idea consists of comparing the sample means of the first k powers, u_1, u_1^2, \dots, u_1^k to their approximate expectations $\frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{k+1}$ under the null distribution. The comparisons are combined quadratically into a statistic[18].

Suppose now the random variable X has density $f_X(x)$ and cdf $F_X(x)$. We may write the alternative density f in terms of f_X as follows. Define the orthonormal polynomials g_r as

$$g_r(x) = p_r(F_X(x)), \quad r = 0, 1, \dots, k.$$

Then by (3.4)

$$g_0 = 1, \quad \int_{-\infty}^{\infty} g_r(x) g_s(x) f_X(x) dx = \delta_{rs}, \quad r, s = 0, 1, \dots, k$$

and the density f becomes a function of x :

$$f(x|\boldsymbol{\theta}) = f(F_X(x)|\boldsymbol{\theta}) \frac{du}{dx} = c(\boldsymbol{\theta}) \exp \left\{ \sum_{r=1}^k \theta_r g_r(x) \right\} f_0(x).$$

where $f_0(x)$ is the density of X under the null hypothesis. If $f_0(x)$ depends on a nuisance parameter $\boldsymbol{\beta}$, we may write

$$f(x|\boldsymbol{\theta}, \boldsymbol{\beta}) = c(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp \left\{ \sum_{r=1}^k \theta_r g_r(x|\boldsymbol{\beta}) \right\} f_0(x|\boldsymbol{\beta}). \quad (3.5)$$

Hence in general, suppose now we have a random sample X_1, \dots, X_n from $f_X(x|\boldsymbol{\beta})$. We wish to test the hypothesis

$$H_0 : f_X(x|\boldsymbol{\beta}) = f_0(x|\boldsymbol{\beta})$$

where f_0 is some specified density function. We write the k -order alternative density as per (3.5). Then testing H_0 is equivalent to testing $H_0'' : \boldsymbol{\theta} = \vec{0}_k$ against $H_1'' : \theta_r \neq 0$ for some r .

The log-likelihood for $(\boldsymbol{\theta}, \boldsymbol{\beta})$ is given by

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\beta} | \{x_i\}) &= \ln \left[\prod_{i=1}^n f(x_i | \boldsymbol{\theta}, \boldsymbol{\beta}) \right] \\ &= \sum_{i=1}^n \ln f(x_i | \boldsymbol{\theta}, \boldsymbol{\beta}) \\ &= n \ln c(\boldsymbol{\theta}, \boldsymbol{\beta}) + \sum_{i=1}^n \sum_{r=1}^k \theta_r h_r(x_i | \boldsymbol{\beta}) + \sum_{i=1}^n \ln f_0(x_i | \boldsymbol{\beta}). \end{aligned}$$

Under the constraint that $\boldsymbol{\theta} = \vec{0}_k$, Bouyn[14] maximizes the Lagrangian

$$L(\{x_i\}, \boldsymbol{\lambda} | \boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{1}{n} l(\boldsymbol{\theta}, \boldsymbol{\beta} | \{x_i\}) - \boldsymbol{\lambda}' \boldsymbol{\theta}$$

where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers, thus leading to an estimated score vector

$$\tilde{d}_{\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} L(\{x_i\}, \boldsymbol{\lambda} | \tilde{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n g(x_i | \tilde{\boldsymbol{\beta}})$$

where $\tilde{\boldsymbol{\beta}}$ is the restricted ML estimator of $\boldsymbol{\beta}$ when $\boldsymbol{\theta} = \vec{0}_k$. Setting $l_{11} = \text{var}[\tilde{d}_{\boldsymbol{\theta}}]$, the test statistic is then

$$\tilde{d}_{\boldsymbol{\theta}} \tilde{l}_{11}^{-1} \tilde{d}_{\boldsymbol{\theta}} = \frac{1}{n} \left(\sum_{i=1}^n \tilde{\mathbf{g}}_i \right)' \tilde{l}_{11}^{-1} \left(\sum_{i=1}^n \tilde{\mathbf{g}}_i \right)$$

where $\tilde{\mathbf{g}}_i = (g_1(x_i), \dots, g_k(x_i))'$, $i = 1, \dots, n$. Bouyn[14] shows that as $n \rightarrow \infty$, $\tilde{d}_{\boldsymbol{\theta}} \tilde{l}_{11}^{-1} \tilde{d}_{\boldsymbol{\theta}}$ asymptotically has a $\chi^2(k)$ distribution.

The smooth model with density as (3.5) is a *Neyman-type model*.

3.3.2 Barton-type smooth models

Alternatively, a *Barton-type model*, related to Barton[5], defines the k -order alternative density function as

$$f(u|\boldsymbol{\theta}) = \left\{ 1 + \sum_{r=1}^k \theta_r p_r(u) \right\}, \quad 0 \leq u \leq 1,$$

for the uniform variable U and

$$f(x|\boldsymbol{\theta}, \boldsymbol{\beta}) = \left\{ 1 + \sum_{r=1}^k \theta_r g_r(x|\boldsymbol{\beta}) \right\} f_0(x|\boldsymbol{\beta}) \quad (3.6)$$

for the random variable X . Since for $c \in \mathbb{R}$ around 0, $e^c \approx 1 + c$, the densities for the Neyman-type and Barton-type smooth models (and the score tests) derived are asymptotically equivalent.

3.3.3 Advantages and disadvantages of smooth models

Rayner et al.[39] discuss the advantages and disadvantages of the two types of the smooth model. The main disadvantage of the Barton-type model is that for possible values of θ the k -order alternative given by (3.6) can give negative probabilities. The Barton-type model is still used as a smooth model because some adjustments have been obtained to accommodate for the possible negative probabilities (see Rayner et al.[39]) and it is asymptotically equivalent to the Newton-type model which is a proper density¹. The Barton-type model is also used over the Newton-type model for the following reasons:

- The Barton-type model has no normalising constant $c(\theta, \beta)$ like the Neyman-type model, which is not always given and thus needs to be calculated.
- The Barton-type model provides simple ML estimators of θ : As an orthogonal series estimation of the density f_X (see Efromovich[21]), the k -order alternative f for the Barton-type model can be written as (3.6) where

$$\theta_r = \int g_r(x|\beta) f(x|\theta, \beta) dx \quad r = 0, 1, \dots, k.$$

Then for the r -th orthonormal function g_r ,

$$E[g_r(X)] = \int g_r(x) f(x|\theta, \beta) dx = \theta_r \quad r = 0, 1, \dots, k$$

and for a random sample X_1, \dots, X_n of the density of f_X , the ML estimator for the θ_r is simply

$$\hat{\theta}_r = \frac{1}{n} \sum_{i=1}^n g_r(X_i) \quad r = 1, \dots, k. \quad (3.7)$$

- For testing for trend, we have the choice of the orthonormal functions $\{g_r\}$ which in turn specify the alternative hypothesis. That is, if the $\{g_r\}$ are polynomial (trigonometric, logarithmic, etc.), we will then test against polynomial (trigonometric, logarithmic, etc.) departures from the null distribution.

3.4 Orthonormal Polynomials Defined From Distributional Moments

We consider the polynomials $\{g_r(x)\}$ given by

$$g_r(x) = a_{r0} + a_{r1}x + \dots + a_{rr}x^r \quad r = 0, 1, \dots, I-1 \quad (3.8)$$

where the a_{rj} are constants, $j = 0, 1, \dots, r$. As a polynomial, $g_r(x)$ is asymptotically equivalent to $(x-c)^r$ for any constant c . Also, for all $r < s$, there is an x_0 such that $g_r(x) < g_s(x)$ for all $x \geq x_0$. Let X be a categorical variable with scores x_1, \dots, x_I and probabilities π_1, \dots, π_I , $\pi_i > 0$. Then, by these two properties of polynomials, we have the following:

- $E[g_r(X)]$ is asymptotically equivalent to $\mu_r = E[(X-\mu)^r]$, the r -th central moment of X ;
- A statistic based on $g_r(X)$ will in fact detect deviations for the r first central moments, $r = 1, \dots, I-1$.

¹The Newton-type model is a proper density in the sense that it is a non-negative function that integrates to 1.

Let the polynomials $\{h_s(y)\}$ be given by

$$h_s(y) = b_{s0} + b_{s1}y + \cdots + b_{ss}y^s \quad s = 0, 1, \dots, J-1$$

and let Y be a categorical variable with scores y_1, \dots, y_J and probabilities π_1, \dots, π_J , $\pi_j > 0$. Then, similarly as above,

- $E[g_r(X)h_s(Y)]$ is asymptotically equivalent to $\mu_{rs} = E[(X - \mu_X)^r(Y - \mu_Y)^s]$, the (r, s) -th bivariate moment for the random vector (X, Y) ;
- A statistic based on $g_r(X)h_s(Y)$ will detect deviations up to the (r, s) -th bivariate moment, $r = 1, \dots, I-1$, $s = 1, \dots, J-1$.

For the above categorical variable X , a recursion developed by Emerson[22] can be used to find the constants a_{rj} , $j = 0, \dots, r$ for the polynomials given by (3.8), such that the set $\{g_r(x)\}$ is orthonormal on the probabilities π_1, \dots, π_I .

The r -th polynomial g_r is defined recursively in terms of the previous g_{r-1} and g_{r-2} by

$$g_r(x) = (A_r x + B_r)g_{r-1}(x) - C_r g_{r-2}(x), \quad r = 1, 2, \dots, I-1 \quad (3.9)$$

where constants A_r , B_r and C_r are defined by

$$A_r = \left\{ \sum_{i=1}^I \pi_i x_i^2 g_{r-1}^2(x_i) - \left[\sum_{i=1}^I \pi_i x_i g_{r-1}^2(x_i) \right]^2 - \left[\sum_{i=1}^I \pi_i x_i g_{r-1}(x_i) g_{r-2}(x_i) \right]^2 \right\}^{-1/2} \quad (3.10)$$

$$B_r = -A_r \sum_{i=1}^I \pi_i x_i g_{r-1}^2(x_i) \quad (3.11)$$

$$C_r = A_r \sum_{i=1}^I \pi_i x_i g_{r-1}(x_i) g_{r-2}(x_i) \quad (3.12)$$

with $g_{-1}(x) = 0$ and $g_0(x) = 1$.

By (3.9),

$$g_1(x) = \frac{x - \mu}{\sqrt{\mu_2}} \quad \text{and} \quad g_2(x) = a \left[(x - \mu)^2 - \frac{\mu_3}{\mu_2} (x - \mu) - \mu_2 \right]$$

where

$$a = \sqrt{\frac{\mu_2}{\mu_2 \mu_4 - \mu_3^2 - \mu_2^3}}$$

Proof. See Appendix A. □

Example 3.3. Let X have a discrete uniform distribution on $\{1, \dots, I\}$. Then, the mean and second through fourth central moments are

$$\mu_X = \frac{I+1}{2}, \quad \mu_2 = \frac{I^2-1}{12}, \quad \mu_3 = 0, \quad \text{and} \quad \mu_4 = \frac{(I^2-1)(3I^2-7)}{240}.$$

The first three orthonormal polynomials obtained by Emerson's recursion are

$$g_0(i) = 1, \quad g_1(i) = \left(\frac{12}{I^2-1} \right)^{1/2} \left(i - \frac{I+1}{2} \right) \quad \text{and} \quad g_2(i) = a \left[\left(i - \frac{I+1}{2} \right)^2 - \frac{12}{I^2-1} \right]$$

where

$$a = \sqrt{\frac{12}{(I^2 - 1)(I^2 - 4)}}.$$

Example 3.4. Let $X \sim N(\mu, \sigma^2)$. Then, the mean and second through fourth central moments are

$$\mu_X = \mu, \quad \mu_2 = \sigma^2, \quad \mu_3 = 0, \quad \text{and} \quad \mu_4 = 3\sigma^4.$$

The first three orthonormal polynomials obtained by Emerson's recursion are

$$g_0(x) = 1, \quad g_1(x) = \frac{x - \mu}{\sigma} \quad \text{and} \quad g_2(x) = \frac{1}{\sqrt{2}} \left[\left(\frac{x - \mu}{\sigma} \right)^2 - 1 \right].$$

Chapter 4

Decompositions of Some Chi-Squared Statistics

The widely used chi-squared tests are generally multilateral hypothesis tests and thus are omnibus. That is, given that the alternative hypothesis is composite, we have difficulty restraining the hypothesis test to test against an alternative which better reflects the reality of the problem.

The ANOVA test, when it is significant, indicates the existence of a difference between treatments but fails to indicate which treatments are different. For that reason, the ANOVA test is followed by a series of multiple comparisons between treatment means. A similar process exists for chi-squared tests, the *decomposition of the chi-squared statistic*, also known as the *partition* of the chi-squared statistic.

In number theory, a partition of a positive integer n is a way of writing n as a sum of positive integers. In set theory, a partition of a set is its division into non-overlapping and non-empty subsets. In the context of chi-squared tests, a partition is defined as follows:

Definition 4.1. *A chi-squared statistic W is said to be partitioned, or decomposed, if there exists independent (at least asymptotically) chi-squared statistics W_1, \dots, W_k , $k \geq 2$, such that*

$$W = W_1 + \dots + W_k$$

where the W_1, \dots, W_k are called the components of W .

The concept of the decomposition of the chi-squared statistic is possible by Theorem 2.1 and Remark 2.2 by which W can be decomposed up to $k = \nu$ components where $W \sim \chi^2(\nu)$ distribution. In this case, the components would be iid with a $\chi^2(1)$ distribution.

In this chapter, we present a few decompositions for some chi-squared statistics, including the Pearson statistic.

4.1 The Decomposition Into Subtables

Suppose we have an $I \times J$ contingency table under the permutation model. We would like to test the dependence relationship between the row and column variables.

Iversen[25] describes how the Pearson and LRT statistics can be decomposed in a way such that each component is associated to a subtable of the $I \times J$ contingency table and thus allowing to determine which cross-categories are significant in describing the dependence relationship between the row and column variables.

In order to have independent components, subtables must be carefully chosen to be independent. That is, the product of their probabilities is equal to the probability of the original $I \times J$ contingency table.

The decomposition of a contingency table into independent subtables can be done recursively. Suppose we have a $2 \times J$ contingency table. By subsection 2.3.4, given the marginal totals, the random vector $\tilde{\mathbf{N}}_i$, which is the vector obtained by removing the last element of \mathbf{N}_i , has a multivariate hypergeometric distribution. We choose 2×2 subtables in the following way: the first subtable compares the first two columns, the second subtable combines the first two columns and compares them to the third column, \dots , and the $(J - 1)$ -th component combines the $J - 1$ first columns and compares them to the last. The s -th subtable is given by Table 4.1

Table 4.1: The s -th 2×2 subtable

	$Y_{(1:s)}$	Y_{s+1}	Total
X_1	M_{s1}	$N_{1,s+1}$	$M_{(s+1)1}$
X_2	M_{s2}	$N_{2,s+1}$	$M_{(s+1)2}$
Total	$M_{s\cdot}$	$n_{\cdot,s+1}$	$M_{(s+1)\cdot}$

where for the i -th row, M_{si} denotes the cumulative count up to the s -th column:

$$M_{si} = \begin{cases} N_{i1} + \dots + N_{is} & s = 1, \dots, J - 1, \\ N_{i\cdot} & s = J \end{cases}$$

and $M_{s\cdot}$ denotes the cumulative column total for the s -th column:

$$M_{s\cdot} = \begin{cases} N_{\cdot 1} + \dots + N_{\cdot s} & s = 1, \dots, J - 1, \\ n & s = J. \end{cases}$$

This decomposition satisfies the equality

$$P_1 P_2 \dots P_{(J-1)} = P\left(\tilde{\mathbf{N}}_{1\cdot} = \tilde{\mathbf{n}}_{1\cdot}, \{n_{i\cdot}\}, \{n_{\cdot j}\}, n\right),$$

where P_s is the probability of the s -th subtable, $s = 1, \dots, J - 1$.

For an $I \times J$ contingency table, the independent subtables result in the comparison of first two columns, combining them and comparing them to the third column, so on and so forth. Each of the $J - 1$ associated test statistics now have $I - 1$ degrees of freedom. A more detailed decomposition as mentioned by Iversen[25] decomposes these $I \times 2$ subtables each into $I - 1$ independent 2×2 subtables each with a test statistics with a $\chi^2(1)$ distribution. This is now a decomposition of the original contingency table into $(I - 1)(J - 1)$ subtables which corresponds to the decomposition of the chi-square statistic of the original table into $(I - 1)(J - 1)$ components. As Iversen[25] demonstrates, it is not always necessary to do such a detailed decomposition if certain row and/or column categories are grouped.

The rules for decomposing the contingency table into independent subtables are

- Each cell count in the original table must be a cell count in one and only one subtable.
- Each marginal total in the original table must be a marginal total in one and only one subtable.
- The sum of the degrees of freedom for each subtables must equal to the degrees of freedom for the original table.

The first two rules guarantee that the product of the probabilities for the component tables is equal to the probability of the original table, i.e. the subtables are independent. The last rule guarantees that

the sum of the test statistics for the subtables is equal to the test statistic for the original table, at least approximately.

For the LRT statistic, the decomposition is exact. On the other hand, the sum of the Pearson statistics for the subtables is approximately equal to the Pearson statistic for the original table. It is still valid to use the Pearson statistic since by section 2.4.1 it is asymptotically equivalent to the LRT statistic when the null hypothesis is true.

As per subsection 2.3.4, once we condition on both sets of marginal totals, the conditional probabilities of the cell counts in the three models are equal. In that case, this decomposition can be used to both test the hypothesis of independence and the hypothesis of homogeneity, with the use of the appropriate ML estimators under the null hypothesis.

An example of the decomposition

Example 4.1. (From Agresti[1].) We classify a sample of psychiatrists by their school of psychiatric thought and by their opinion on the origin of schizophrenia.

School of Thought	Origin of Schizophrenia			Total
	Biogenetic	Environmental	Combination	
Eclectic	90	12	78	180
Medical	13	1	6	20
Psychoanalytic	19	13	50	82
Total	122	26	134	282

$G_4^2 = 23.04$

The LRT statistic for testing independence is $G^2 = 23.04$ with four degrees of freedom and a p -value 1.25×10^{-4} . The statistic shows there is evidence of an association. The table suggests a division by number of sources (single origin: biogenetic, environmental versus combination).

The single sources give a 3×2 table with $G_2^2 = 13.25$, where the subscript indicates the degrees of freedom. This table is decomposed into two subtables.

School	Biogenetic	Environmental	Total
Eclectic	90	12	102
Medical	13	1	14
Psychoanalytic	19	13	32
Total	122	26	148

$G_2^2 = 13.25$

School	Biogenetic	Environmental	Total	School	Biogenetic	Environmental	Total
Eclectic	90	12	102	Ecl + Med	103	13	116
Medical	13	1	14	Psychoanalytic	19	13	32
Total	103	13	116	Total	122	26	148

$G_1^2 = 0.29$ $G_1^2 = 12.95$

From these subtables, we see that the significant chi-square on two degrees of freedom for the single sources is mainly due to the difference between the psychoanalytical school of thought and the other schools.

Of those who chose either biogenetic or environmental origins, member of the psychoanalytical school are more likely to choose the environmental origin of schizophrenia.

We now have $23.04 - 13.25 = 9.79$ on $4 - 2 = 2$ degrees of freedom to compare the single origins versus a combination of biogenetics and environment. A decomposition of this chi-square is given by the following tables.

School	Bio + Env	Combination	Total
Eclectic	102	78	180
Medical	14	6	20
Psychoanalytic	32	50	82
Total	148	134	282

$$G_1^2 = 9.79$$

School	Bio + Env	Combination	Total
Eclectic	102	78	180
Medical	14	6	20
Total	116	84	200

$$G_1^2 = 1.36$$

School	Bio + Env	Combination	Total
Ecl + Med	116	84	200
Psychoanalytic	32	50	82
Total	148	134	282

$$G_1^2 = 8.43$$

From these subtables, we see the significant chi-square of two degrees of freedom is once again mainly due to the difference between the psychoanalytical school of thought and the other schools. Members of the psychoanalytical school are more likely to attributing the origin of schizophrenia as a combination rather than only biogenetics or environment.

The various chi-squares are summarised in the following table.

Source	df	Value	P-value
Biogenetic vs. Environment			
Eclectic vs. Medical	1	0.29	0.5875
Ecl + Med vs. Psychoanalytical	1	12.95	0.0003
Bio + Env vs. Combination			
Eclectic vs. Medical	1	1.36	0.2437
Ecl + Med vs. Psychoanalytical	1	8.43	0.0037
Total	4	23.04	0.0001

From this table, we see that adding the least significant chi-squares, we get a total of 3.66 on two degrees of freedom which is the chi-square that indicates there is no significant difference between the eclectic and medical school of thoughts. The remaining two degrees of freedom produce a significant chi-square of 19.24 which indicates that the distribution for the psychoanalytical school of thought on the origin of schizophrenia is different from the other two schools.

4.2 The Decomposition of the Pearson Statistic

4.2.1 Decomposition of the Pearson statistic in Two-Way Tables

Suppose we have an $I \times J$ contingency table. Let x_1, \dots, x_I be the scores for the row variable X . Similarly, let y_1, \dots, y_J be the scores for the column variable Y . We have the following property, given by Lancaster[27],

Property 4.1. *For any set of k orthonormal functions $\{p_k(x, y)\}$, $k \leq (I - 1)(J - 1)$, we can write*

$$X_P^2 = \sum_{r=1}^k V_r^2 \quad \text{where} \quad V_r = \sum_{i=1}^I \sum_{j=1}^J p_r(x_i, y_j) \quad r = 1, \dots, k$$

are components of the Pearson statistic.

In other words, the decomposition of the Pearson statistic by orthonormal functions is invariant to the choice of the set of orthonormal functions.

To test the dependence relationship between X and Y , we set a Barton-type smooth model of the form (3.6) however defined for a bivariate density function given by $f_{XY}(x_i, y_j)$ which is equal to $f_0(x_i, y_j)$ under the null hypothesis, $i = 1, \dots, I, j = 1, \dots, J$. The (k_1, k_2) -order alternative density is

$$f(x_i, y_j | \boldsymbol{\theta}) = \left\{ 1 + \sum_{s=1}^{k_2} \sum_{r=1}^{k_1} \theta_{sr} h_s(y_j) g_r(x_i) \right\} f_0(x_i, y_j) \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (4.1)$$

where

- $k_1 \leq I - 1$ and $k_2 \leq J - 1$ are chosen to be at most 4 (usually 2),
- $\{\theta_{sr} : s = 1, \dots, k_2, r = 1, \dots, k_1\}$ are real valued parameters,
- h_1, \dots, h_{k_2} are functions orthonormal with respect to the column probabilities $\pi_{.1}, \dots, \pi_{.J}$, and
- g_1, \dots, g_{k_1} are functions orthonormal with respect to the row probabilities $\pi_{1.}, \dots, \pi_{I.}$.

The null hypothesis is now equivalent to testing

$$H_0'' : \boldsymbol{\theta} = \vec{0}_{k_1 k_2} \quad \text{against} \quad H_1'' : \theta_{sr} \neq 0 \text{ for some } (s, r)$$

where $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1k_1}, \dots, \theta_{k_2 1}, \dots, \theta_{k_2 k_1})'$ is the vector of the θ_{sr} in hierarchical order and the marginal probabilities $\{\pi_{i.}\}$ and $\{\pi_{.j}\}$ are nuisance parameters. In fact, we can remove $\pi_{i.}$ and $\pi_{.j}$ as nuisance parameters since under any model from section 2.3, the marginal probabilities satisfy

$$\pi_{1.} + \dots + \pi_{I.} = 1 \quad \text{and} \quad \pi_{.1} + \dots + \pi_{.J} = 1.$$

Let \mathbf{h}_s be the column vector with elements $h_s(y_j)$:

$$\mathbf{h}_s = (h_s(y_1), \dots, h_s(y_J))' \quad s = 1, \dots, k_2. \quad (4.2)$$

Similarly, let \mathbf{g}_r be the column vector with elements $g_r(x_i)$:

$$\mathbf{g}_r = (g_r(x_1), \dots, g_r(x_I))' \quad r = 1, \dots, k_1.$$

We denote by $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}$ the column vectors of the elements $\hat{h}_s(y_j)$ and $\hat{g}_r(x_i)$ where \hat{h} and \hat{g} are the functions orthonormal with respect to the ML estimators $\{\hat{\pi}_{\cdot j}\}$ and $\{\hat{\pi}_{i\cdot}\}$, respectively. Let \mathbf{H}_s be the $IJ \times IJ$ matrix defined by

$$\mathbf{H}_s = \text{diag}(\mathbf{h}_s) = \begin{bmatrix} \mathbf{h}_s & 0 & \cdots & 0 \\ 0 & \mathbf{h}_s & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{h}_s \end{bmatrix} \quad s = 1, \dots, k_2. \quad (4.3)$$

Then, for $k_1 = I - 1$ and $k_2 = J - 1$, the Pearson statistic is the score statistic to test the model

$$f(x_i, y_j | \boldsymbol{\theta}) = \left\{ 1 + \sum_{s=1}^{k_2} \sum_{r=1}^{k_1} \theta_{sr} h_s(y_j) g_r(x_i) \right\} f_0(x_i, y_j) \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1,$$

and by Property 4.1 it has components

$$\hat{V}_{sr} = \frac{1}{\sqrt{n}} \sum_{i=1}^I \sum_{j=1}^J N_{ij} \hat{h}_s(y_j) \hat{g}_r(x_i) = \frac{1}{\sqrt{n}} \hat{\mathbf{g}}_r' \hat{\mathbf{H}}_s \mathbf{N} \quad s = 1, \dots, J - 1, \quad r = 1, \dots, I - 1 \quad (4.4)$$

which, under the null hypothesis, are asymptotically iid with a $N(0, 1)$ distribution, such that

$$X_P^2 = \sum_{s=1}^{J-1} \sum_{r=1}^{I-1} \hat{V}_{sr}^2.$$

Proof. The proof is done by Rayner & Best[36]. □

As a consequence, the decomposition of the Pearson statistic is useful in tests against monotonic trend since by the asymptotic independence of the components, we can construct statistics which have good power against directional alternatives on the the elements of $\boldsymbol{\theta}$.

- \hat{V}_{sr} detects deviations of θ_{sr} because in a similar way as in (3.7), \hat{V}_{sr} is an ML estimator for θ_{sr} .
- $\hat{\mathbf{V}}_{s\cdot} = (\hat{V}_{s1}, \dots, \hat{V}_{s(I-1)})'$ will detect significant deviations of $\boldsymbol{\theta}_{s\cdot} = (\theta_{s1}, \dots, \theta_{s(I-1)})'$, $s = 1, \dots, J - 1$. Since the \hat{V}_{sr}^2 are asymptotically iid with a $\chi^2(1)$, the $\hat{\mathbf{V}}_{s\cdot}' \hat{\mathbf{V}}_{s\cdot}$, are also iid with a $\chi^2(I - 1)$, $s = 1, \dots, J - 1$. Then, the $\hat{\mathbf{V}}_{s\cdot}$ are also components of the Pearson statistic since

$$\sum_{s=1}^{J-1} \hat{\mathbf{V}}_{s\cdot}' \hat{\mathbf{V}}_{s\cdot} = \sum_{s=1}^{J-1} \sum_{r=1}^{I-1} \hat{V}_{sr}^2 = X_P^2.$$

- $\hat{\mathbf{V}}_{(1:s)\cdot} = (\hat{\mathbf{V}}_{1\cdot}', \dots, \hat{\mathbf{V}}_{s\cdot}')'$ will detect significant deviations of $\boldsymbol{\theta}_{(1:s)\cdot} = (\boldsymbol{\theta}'_{1\cdot}, \dots, \boldsymbol{\theta}'_{s\cdot})'$, $s = 1, \dots, J - 1$. In a similar way as for $\hat{\mathbf{V}}_{s\cdot}$, $\hat{\mathbf{V}}_{(1:s)\cdot}$ and $\hat{\mathbf{V}}_{(s+1:J-1)\cdot}$ are asymptotically independent with $\chi^2((I - 1)s)$ and $\chi^2((I - 1)(J - 1 - s))$ distributions, respectively, and are components of the Pearson statistic. In particular, $\hat{V}'_{(1:J-1)\cdot} = X_P^2$ which we know detects extreme deviations in $\boldsymbol{\theta}_{(1:J-1)\cdot} = \boldsymbol{\theta}$.

While the degrees of freedom increase from \hat{V}_{sr} to $\hat{\mathbf{V}}_{(1:s)\cdot}$, the sensitivity of the components decreases with respect to alternatives on parameter spaces with dimension smaller than their degrees of freedom (Agresti[1]):

- A test based on \hat{V}_{sr} has more power than those based on $\hat{\mathbf{V}}_{s\cdot}$, $\hat{\mathbf{V}}_{(1:s)\cdot}$ or the Pearson statistic against the specified alternative $H_1'' : \theta_{sr} \neq 0$ since \hat{V}_{sr} is more sensitive to deviations of θ_{sr} .

¹Similarly, $\hat{\mathbf{V}}_{\cdot r} = (\hat{V}_{1r}, \dots, \hat{V}_{(J-1)r})'$ will detect significant deviations of $\boldsymbol{\theta}_{\cdot r} = (\theta_{1r}, \dots, \theta_{(J-1)r})'$, $r = 1, \dots, I - 1$.

- Similarly, a test based on $\hat{\mathbf{V}}_s$ has more power than those based on $\hat{\mathbf{V}}_{(1:s)}$ or the Pearson statistic against the specified alternative $H_1'' : \theta_{sr} \neq 0$, for some r .

And so, while the decomposition of the Pearson statistic was derived to better understand why the null hypothesis is rejected, we should still do the decomposition since the components will be more powerful against directional alternatives for their corresponding parameter while the Pearson statistic, as an omnibus test, has moderate power against all alternative hypothesis.

Decomposition of the Pearson statistic by orthonormal polynomials

The decompositions of the Pearson statistic derived from the bivariate Barton-type smooth model in subsection 4.2.1 that we will present mostly use the orthonormal polynomials from section 3.4. Since these polynomials are defined from distributional moments, the (s, r) -th component \hat{V}_{sr} will detect deviations in the (r, s) -th central moment.

Continuing from subsection 4.2.1, the decomposition of the Pearson by the orthonormal polynomials from section 3.4 is useful in tests against directional alternatives in the bivariate central moment as

$$\theta_{sr} = \sum_{i=1}^I \sum_{j=1}^J h_s(y_j) g_r(x_i) f(x_i, y_j) \quad r = 1, \dots, I-1, \quad s = 1, \dots, J-1.$$

is asymptotically equivalent to the (r, s) -th bivariate moment, $r = 1, \dots, I-1$, $s = 1, \dots, J-1$. Even though we use the same set of orthonormal functions, the change of the contingency table model and the null hypothesis, and thus the smooth model, will result in the first component corresponding to different known statistics and the subsequent components being extensions of these statistics. That is, if the first component detects linear (quadratic or logarithmic, etc.) deviations for the first moment, the subsequent components will detect deviations of the same type for their corresponding moment.

4.2.2 Under the unrestricted bivariate sampling model

Rayner & Best[36] present a partition of the Pearson statistic, based on a bivariate Barton-type smooth model as in subsection 4.2.1 to better understand why the hypothesis of independence would be rejected.

Under the alternative $H_1' : \pi_{ij} \neq \pi_{i \cdot} \pi_{\cdot j}$, we set the (k_1, k_2) -order alternative as (4.1) with density $f_{XY}(x_i, y_j) = \pi_{ij}$ and null density $f_0(x_i, y_j) = \pi_{i \cdot} \pi_{\cdot j}$, $i = 1, \dots, I$, $j = 1, \dots, J$. Then, for $k_1 = I-1$ and $k_2 = J-1$, the Pearson statistic is the score statistic to test the model

$$\pi_{ij} = \left\{ 1 + \sum_{s=1}^{J-1} \sum_{r=1}^{I-1} \theta_{sr} h_s(y_j) g_r(x_i) \right\} \pi_{i \cdot} \pi_{\cdot j} \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1$$

and its components are given by (4.4) which, under the hypothesis of independence, are asymptotically iid with a $N(0, 1)$ distribution.

Theorem 4.2. *Each \hat{V}_{sr} can be derived as the score statistic to test the hypothesis $H_{0(sr)}'' : \theta_{sr} = 0$ where the $(1, 1)$ -order alternative is given by*

$$\pi_{ij} = \{1 + \theta_{sr} h_s(y_j) g_r(x_i)\} \pi_{i \cdot} \pi_{\cdot j}, \quad i = 1, \dots, I, \quad j = 1, \dots, J. \quad (4.5)$$

Proof. Under the unrestricted bivariate sampling model, the cell counts $\mathbf{N} = (N_{11}, \dots, N_{IJ})'$ have a *multinomial* $(n, \boldsymbol{\pi})$ distribution. Then, the likelihood function for θ_{sr} is

$$L(\theta_{sr} | \{n_{ij}\}, n) = \prod_{i=1}^I \prod_{j=1}^J \frac{\pi_{ij}^{n_{ij}}}{n_{ij}!}.$$

With the cell probabilities given by (4.5), the log-likelihood function for θ_{sr} is

$$\begin{aligned} l(\theta_{sr} | \{n_{ij}\}, n) &\propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln(\pi_{ij}) \\ &\propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln(\{1 + \theta_{sr} h_s(y_j) g_r(x_i)\} \pi_{i\cdot} \pi_{\cdot j}) \\ &\propto \sum_{i=1}^I \sum_{j=1}^J n_{ij} \ln(1 + \theta_{sr} h_s(y_j) g_r(x_i)). \end{aligned}$$

Then, the score function and information matrix for θ_{sr} are

$$\begin{aligned} U(\theta_{sr} | \{n_{ij}\}, n) &= \sum_{i=1}^I \sum_{j=1}^J \frac{n_{ij} h_s(y_j) g_r(x_i)}{1 + \theta_{sr} h_s(y_j) g_r(x_i)} \\ I(\theta_{sr} | \{n_{ij}\}, n) &= n \sum_{i=1}^I \sum_{j=1}^J \frac{\pi_{ij} h_s^2(y_j) g_r^2(x_i)}{(1 + \theta_{sr} h_s(y_j) g_r(x_i))^2}. \end{aligned}$$

The ML estimator of the score function under $H''_{0(sr)} : \theta_{sr} = 0$ is

$$\hat{U}(0 | \{n_{ij}\}, n) = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \hat{h}_s(y_j) \hat{g}_r(x_i).$$

Since the ML estimators of π_{ij} under (4.5) are

$$\begin{aligned} \hat{\pi}_{ij} &= \left\{ 1 + \theta_{sr} \hat{h}_s(y_j) \hat{g}_r(x_i) \right\} \hat{\pi}_{i\cdot} \hat{\pi}_{\cdot j} \\ &= \left\{ 1 + \theta_{sr} \hat{h}_s(y_j) \hat{g}_r(x_i) \right\} \frac{N_{i\cdot} N_{\cdot j}}{n}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \end{aligned}$$

the ML estimator of the information matrix under $H''_{0(sr)} : \theta_{sr} = 0$ is

$$\begin{aligned} \hat{I}(0 | \{n_{ij}\}, n) &= n \sum_{i=1}^I \sum_{j=1}^J \hat{\pi}_{ij} \hat{h}_s^2(y_j) \hat{g}_r^2(x_i) \\ &= n \sum_{j=1}^J \frac{N_{\cdot j}}{n} \hat{h}_s^2(y_j) \sum_{i=1}^I \frac{N_{i\cdot}}{n} \hat{g}_r^2(x_i) \\ &= n \end{aligned}$$

where we use the fact that \hat{h}_s and \hat{g}_r are orthonormal on the column and row probabilities, respectively. Then, the score statistic to test $H''_{0(sr)} : \theta_{sr} = 0$ is the (s, r) -th component

$$\hat{S}^*(0) = \frac{\hat{U}(0)}{\hat{I}^{1/2}(0)} = \frac{1}{\sqrt{n}} \sum_{i=1}^I \sum_{j=1}^J N_{ij} \hat{h}_s(y_j) \hat{g}_r(x_i) = \hat{V}_{sr}$$

Applying the Central Limit Theorem to the N_{ij} , which have a binomial distribution under $H''_{0(sr)} : \theta_{sr} = 0$, we can say that \hat{V}_{sr} asymptotically has a normal distribution. \square

Consequently,

- we reject $H''_{0(sr)} : \theta_{sr} = 0$ in favor of $H''_{1(sr)} : \theta_{sr} < 0$, when \hat{V}_{sr} is significantly small;
- we reject $H''_{0(sr)} : \theta_{sr} = 0$ in favor of $H''_{1(sr)} : \theta_{sr} > 0$, when \hat{V}_{sr} is significantly large.

This is why we can say that each of the component \hat{V}_{sr} of X_P^2 forms the basis for a strong directional test for the (r, s) -th bivariate moment (see Rayner & Best[36]), $r = 1, \dots, I - 1$, $s = 1, \dots, J - 1$.

Remark 4.3. *When the marginal totals are fixed, we can still use the above decomposition of the Pearson statistic where the components are now*

$$V_{sr} = \frac{1}{\sqrt{n}} \sum_{i=1}^I \sum_{j=1}^J N_{ij} h_s(y_j) g_r(x_i), \quad s = 1, \dots, J - 1, \quad r = 1, \dots, I - 1,$$

and the orthonormal polynomials $h_s(y)$ and $g_r(x)$ are based on the fixed marginal probabilities, $\{\pi_{.j} = n_{.j}/n\}$ and $\{\pi_{i.} = n_{i.}/n\}$, respectively (see Best & Rayner[8]).

An extension of the Pearson product-moment correlation coefficient

Let X be a categorical variable with classification

$$A = \left\{ A_1, \dots, A_I : A_i \cap A_j = \emptyset, \quad i \neq j = 1, \dots, I, \quad \bigcup A_i = \Xi \right\}$$

where Ξ is the set of possible values for X . Similarly, let Y be a categorical variable with classification

$$B = \left\{ B_1, \dots, B_J : B_i \cap B_j = \emptyset, \quad i \neq j = 1, \dots, J, \quad \bigcup B_j = \Psi \right\}$$

where Ψ is the set of possible values for Y .

Suppose $(X_1, Y_1), \dots, (X_n, Y_n)$ is a random sample of the random vector (X, Y) . Let N_{ij} be the counts for the cross-category $A_i \times B_j$. The counts are arranged as in Table 2.1 which is under the unrestricted bivariate sampling model. We would like to test the hypothesis of independence against increasing trend. Rayner & Best's[36] above partition of the Pearson statistic can be used here.

Let the mean and variance for X be defined as

$$\mu_X = \sum_{i=1}^I x_i \pi_{i.}, \quad \sigma_X^2 = \sum_{i=1}^I x_i^2 \pi_{i.} - \mu_X^2$$

where x_1, \dots, x_I are scores for X with probabilities $\pi_{i.}, \dots, \pi_{I.}$, $\pi_{i.} > 0$. Similarly, we have

$$\mu_Y = \sum_{j=1}^J y_j \pi_{.j}, \quad \sigma_Y^2 = \sum_{j=1}^J y_j^2 \pi_{.j} - \mu_Y^2.$$

where y_1, \dots, y_J are scores for Y with probabilities $\pi_{.1}, \dots, \pi_{.J}$, $\pi_{.j} > 0$. Let $\{g_r(x)\}$ and $\{h_s(y)\}$ be the set orthonormal polynomials defined as in section 3.4. That is,

$$g_0(x) = 1, \quad \text{and} \quad g_1(x) = \frac{x - \mu_X}{\sigma_X}$$

and

$$h_0(y) = 1, \quad \text{and} \quad h_1(y) = \frac{y - \mu_Y}{\sigma_Y}.$$

Then, we can show that the first component \hat{V}_{11} is a multiple of the Pearson product-moment correlation coefficient for grouped data.

Proof.

$$\begin{aligned}
r_{xy} &= \frac{\sum_{i=1}^I \sum_{j=1}^J N_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{i=1}^I N_{i.} (x_i - \bar{x})^2 \sum_{j=1}^J N_{.j} (y_j - \bar{y})^2}} \\
&= \frac{1}{n} \frac{\sum_{i=1}^I \sum_{j=1}^J N_{ij} (x_i - \hat{\mu}_X)(y_j - \hat{\mu}_Y)}{\sqrt{\sum_{i=1}^I \hat{\pi}_{i.} (x_i - \hat{\mu}_X)^2 \sum_{j=1}^J \hat{\pi}_{.j} (y_j - \hat{\mu}_Y)^2}} \\
&= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J N_{ij} \frac{(x_i - \hat{\mu}_X)}{\hat{\sigma}_X} \frac{(y_j - \hat{\mu}_Y)}{\hat{\sigma}_Y} \\
r_{xy} &= \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J N_{ij} \hat{g}_1(x_i) \hat{h}_1(y_j) = \frac{1}{\sqrt{n}} \hat{V}_{11}
\end{aligned}$$

where $\hat{\mu}_X$, $\hat{\mu}_Y$, $\hat{\sigma}_X$ and $\hat{\sigma}_Y$ are the ML estimators under the null hypothesis. \square

Since the Pearson correlation detects linear shifts in the means, so does \hat{V}_{11} and the subsequent components \hat{V}_{sr} detect linear shifts in the (r, s) -th bivariate moment, $r = 1, \dots, I - 1$, $s = 1, \dots, J - 1$.

An extension of the Spearman correlation coefficient

Suppose now that the categories of the classifications A and B are respectively defined as

$$\begin{aligned}
A_i &= \{R_a = i\} \quad i = 1, \dots, I = n \\
B_j &= \{S_a = j\} \quad j = 1, \dots, J = n
\end{aligned}$$

where R_a denotes the rank of X_a with respect to X_1, \dots, X_n and S_a denotes the rank of Y_a with respect to Y_1, \dots, Y_n , not allowing for tied ranks, $a = 1, \dots, n$. The counts are now given by

$$N_{ij} = \sum_{a=1}^n I[R_a = i, S_a = j], \quad i, j = 1, \dots, n.$$

The resulting $n \times n$ contingency table is a doubly-ordered².

Since there are no tied ranks,

- for the i -th row, i.e. we fix $R_a = i$, N_{ij} is equal to 1 for only one column:

$$N_{ij} = \sum_{a=1}^n I[R_a = i, S_a = j] = \sum_{a=1}^n I[S_a = j] = \begin{cases} 1 & S_a = j, \\ 0 & \text{otherwise;} \end{cases}$$

- Similarly, for the j -th column, i.e. we fix $S_a = j$, N_{ij} is equal to 1 for only one row:

$$N_{ij} = \sum_{a=1}^n I[R_a = i, S_a = j] = \sum_{a=1}^n I[R_a = i] = \begin{cases} 1 & R_a = i, \\ 0 & \text{otherwise.} \end{cases}$$

²A doubly-ordered contingency table is defined as a contingency table in which both row and column classifications are ordinal.

And so we can redefine the cell counts as,

$$N_{ij} = \begin{cases} 1 & (R_a, S_a) = (i, j), \\ 0 & \text{otherwise,} \end{cases} \quad (4.6)$$

and all the marginal totals are equal to 1. The $n \times n$ contingency table formed from the counts N_{ij} is as Table 4.2 which is under the unrestricted bivariate sampling model with fixed marginal totals by construction. By Remark 4.3, Rayner & Best's[36] above partition of the Pearson statistic can be used here.

Table 4.2: A rank-by-rank table (without ties)

	1	...	j	...	n	Total
1	N_{11}	...	N_{1j}	...	N_{1n}	1
\vdots	\vdots		\vdots		\vdots	\vdots
i	N_{i1}	...	N_{ij}	...	N_{in}	1
\vdots	\vdots		\vdots		\vdots	\vdots
n	N_{n1}	...	N_{nj}	...	N_{nn}	1
Total	1	...	1	...	1	n

With the same set of orthonormal polynomials as for the Pearson correlation coefficient, we can show that \hat{V}_{11}/\sqrt{n} is now the Spearman rank correlation coefficient.

Proof. The proof is done by Rayner & Best[36]. □

Since the Spearman correlation detects monotonic shifts in the means, so does \hat{V}_{11} and the subsequent components \hat{V}_{sr} detect monotonic shifts in the (r, s) -th bivariate moment, $r, s = 1, \dots, n - 1$.

An example of the decomposition

Example 4.2. (From Bagdonavičius et al.[4].) The quantity of starch was determined in 16 potatoes using two methods. The results are given in the following table

i	1	2	3	4	5	6	7	8
X_i	21.8	18.7	18.3	17.5	18.5	15.5	17.0	16.6
Y_i	21.5	18.7	18.2	17.4	18.4	15.4	16.7	16.9
i	9	10	11	12	13	14	15	16
X_i	14.0	17.2	21.7	18.6	17.9	17.7	18.4	15.6
Y_i	13.9	17.0	21.4	18.6	18.0	17.6	18.5	15.5

The 16×16 rank-by-rank contingency table is

Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
2	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
3	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1
4	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	1
5	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	1
6	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	1
7	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	1
8	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	1
9	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	1
10	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	1
11	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	1
12	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	1
13	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	1
14	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	1
15	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	1
16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	1
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16

The Pearson statistic for testing independence between the methods of measurements is valued at 240 with 225 degrees of freedom. With a p-value of 0.2348, the Pearson test says there is little evidence of an association. Looking at the contingency table, this is very doubtful and so we calculate the following components of the Pearson statistic and their associated p-values:

Component	df	Value	P-value
\hat{V}_{11}^2	1	15.81	< 0.0001
\hat{V}_{12}^2	1	< 0.0001	0.9988
\hat{V}_{21}^2	1	< 0.0001	0.9988
\hat{V}_{22}^2	1	0.35	0.5500
Remainder	221	223.83	0.4342
X_P^2	225	240	0.2348

The highly significant (1,1)-th component suggests there is a monotonic association between the means of the two methods. Given the Spearman correlation coefficient is given by $r_S = \hat{V}_{11}/4 = 0.9941$, there is a strong increasing monotonic trend between the two methods, which is in accordance with the rank-by-rank table.

4.2.3 Under the multinomial model

Suppose we have an $I \times J$ contingency table under the product-multinomial model (here, we assume we have a row-multinomial model). We would like to test the hypothesis of homogeneity for distributions of the row variables $X_i, i = 1, \dots, I$. The general hypotheses for a chi-squared test of homogeneity for X_1, \dots, X_I are

$$H_0 : F_1(x) = \dots = F_I(x), \quad \text{for all } x \in \mathbb{R},$$

$$H_1 : \text{at least one } F_i \text{ is different, for at least one } x \in \mathbb{R},$$

which are equivalent to testing

$$H'_0 : \pi_{1j} = \dots = \pi_{Ij} \quad j = 1, \dots, J, \tag{4.7}$$

$$H'_1 : \text{at least one } \pi_{ij} \text{ is different, for at least one } (i, j). \tag{4.8}$$

However, we are interested in testing the hypothesis of homogeneity against the hypothesis of increasing trend in the row distributions. This is equivalent to testing the hypotheses

$$\begin{aligned} H'_0 &: \tau_{1i} = \cdots = \tau_{sI} \quad s = 1, \dots, J, \\ H'_1 &: \tau_{1i} \leq \cdots \leq \tau_{sI} \quad s = 1, \dots, J, \end{aligned}$$

with at least one inequality in the alternative, where

$$\tau_{si} = \begin{cases} \pi_{i1} + \cdots + \pi_{is} & s = 1, \dots, J-1, \\ 1 & s = J. \end{cases}$$

τ_{si} is the s -th cumulative cell probability for the i -th row, $i = 1, \dots, I$, $s = 1, \dots, J$.

An extension of the Yates test

Under the row-multinomial model, Rayner & Best[38] present a partition of the Pearson statistic, based on a bivariate Barton-type smooth model, to test the hypothesis of homogeneity of the row distributions against increasing trend.

Under the alternative hypothesis (4.8), we set the alternative density as (4.1) with density $f_{XY}(x_i, y_j) = \pi_{ij}$ and null density $f_0(x_i, y_j) = \pi_{.j}$, $i = 1, \dots, I$, $j = 1, \dots, J$, and $g_r \equiv 1$ since we want the cell probabilities to be a function of only the column probabilities. The k -order alternative is given by

$$\pi_{ij} = \left\{ 1 + \sum_{s=1}^k \theta_{si} h_s(y_j) \right\} \pi_{.j}, \quad i = 1, \dots, I, \quad j = 1, \dots, J \quad (4.9)$$

where

- $k \leq J-1$ is chosen to be at most 4,
- $\{\theta_{si} : i = 1, \dots, I, s = 1, \dots, k\}$ are real valued parameters,
- $y_1 < \cdots < y_J$ are the ordered scores for the columns,
- h_1, \dots, h_k are orthonormal functions on the column probabilities $\pi_{.1}, \dots, \pi_{.J}$.

Testing the hypothesis of homogeneity is now equivalent to

$$H''_0 : \boldsymbol{\theta} = \vec{0}_k \quad \text{against} \quad H''_1 : \theta_{si} \neq 0 \text{ for some } (s, i),$$

where $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1I}, \dots, \theta_{k1}, \dots, \theta_{kI})'$ and the column probabilities $\pi_{.1}, \dots, \pi_{.(J-1)}$ are nuisance parameters.

For $k = J-1$, Rayner & Best[35] show that the Pearson statistic for the i -th row,

$$X_i^2 = \sum_{j=1}^J \frac{\left(N_{ij} - \mu_{ij}^{(0)} \right)^2}{\mu_{ij}^{(0)}}$$

is the score statistic to test the hypothesis of homogeneity for the i -th row,

$$H'_{0i} : \pi_{ij} = \pi_{.j}, \quad j = 1, \dots, J$$

against the corresponding $(J-1)$ -order alternative density

$$\pi_{ij} = \left\{ 1 + \sum_{s=1}^{J-1} \theta_{si} h_s(y_j) \right\} \pi_{.j}, \quad j = 1, \dots, J$$

and it has components

$$\hat{V}_{si} = \frac{1}{\sqrt{n_i}} \sum_{j=1}^J N_{ij} \hat{h}_s(y_j), \quad s = 1, \dots, J-1$$

which, under H'_{0i} , are asymptotically iid with a $N(0, 1)$ distribution such that

$$X_i^2 = \hat{V}_{1i}^2 + \dots + \hat{V}_{(J-1)i}^2 \quad i = 1, \dots, I.$$

Then, the Pearson statistic to test the hypothesis of homogeneity for the $I \times J$ row-multinomial contingency table is

$$X_P^2 = \sum_{i=1}^I X_i^2 = \sum_{s=1}^{J-1} \sum_{i=1}^I \hat{V}_{si}^2.$$

Let \mathbf{h}_s be the column vector defined as (4.2), $s = 1, \dots, J-1$. Then, the \hat{V}_{si} can be written as

$$\hat{V}_{si} = \frac{1}{\sqrt{n_i}} \hat{\mathbf{h}}'_s \mathbf{N}_i, \quad i = 1, \dots, I, \quad s = 1, \dots, J-1.$$

Since the cell counts are correlated, even asymptotically, as linear combinations of the cell counts, the \hat{V}_{si} are also correlated. Thus, while they are components for the X_i^2 , $i = 1, \dots, I$ and their sum of their squares is equal to X_P^2 , they are not components for the Pearson statistic.

Rayner & Best[37] show that for $k = J-1$, the score statistic for the bivariate Barton-type smooth model on the $I-1$ first rows

$$\pi_{ij} = \left\{ 1 + \sum_{s=1}^k \theta_{si} h_s(y_j) \right\} \pi_{.j}, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1$$

is the Pearson statistic which can be partitioned by the random vectors $\hat{\mathbf{V}}_s = (\hat{V}_{s1}, \dots, \hat{V}_{sI})'$, $s = 1, \dots, J-1$, such that

$$\hat{\mathbf{V}}'_1 \hat{\mathbf{V}}_1 + \dots + \hat{\mathbf{V}}'_{(J-1)} \hat{\mathbf{V}}_{(J-1)} = \sum_{s=1}^{J-1} \sum_{i=1}^I \hat{V}_{si}^2 = X_P^2.$$

The $\hat{\mathbf{V}}_s$ are asymptotically iid as I -variate normal distribution with mean $\vec{0}_I$ and covariance matrix $\mathbb{I}_{I-1} + \mathbf{f}\mathbf{f}'$ where

$$\mathbf{f} = \frac{1}{\sqrt{n_I}} (\sqrt{n_1}, \dots, \sqrt{n_{(I-1)}})'$$

Consequently, the $\hat{\mathbf{V}}'_s \hat{\mathbf{V}}_s = \hat{V}_{s1}^2 + \dots + \hat{V}_{sI}^2$ are asymptotically iid with a $\chi^2(I-1)$ distribution, $s = 1, \dots, J-1$.

In a similar way as in Theorem 4.2, the \hat{V}_{sa} can be derived as score statistic to test homogeneity under the Barton-type smooth model

$$\pi_{aj} = \begin{cases} \{1 + \theta_{sa} h_s(y_j)\} \pi_{.j} & a = i, \quad j = 1, \dots, J-1, \\ \pi_{.j} & a \neq i, \quad j = 1, \dots, J-1 \end{cases}$$

for the $I-1$ first rows. Therefore, the \hat{V}_{si} form the basis for a strong directional test which is sensitive to deviations of θ_{si} , which is asymptotically equivalent to the s -th moment for the i -th row distribution, $s = 1, \dots, J-1$, $i = 1, \dots, I-1$.

Analogously, $\hat{\mathbf{V}}_s$ can be derived as a score statistic to test homogeneity for all the first $I - 1$ rows simultaneously under the Barton-type smooth model

$$\pi_{ij} = \{1 + \theta_{si} h_s(y_j)\} \pi_{.j}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1$$

and the $\hat{\mathbf{V}}_s$ form the basis for a test detecting significant deviations of the s -th moment for the first $I - 1$ rows, $s = 1, \dots, J - 1$.

The extension Let the column scores be $y_j = j$, $j = 1, \dots, J$. We choose $\{h_s(j)\}$ to be the set of polynomials orthonormal on the discrete uniform distribution on $\{1, \dots, J\}$. Then, we can show that the first component is Yates' Q statistic.

An extension of the Kruskal-Wallis test

Under a row-multinomial model with both sets of marginal totals fixed, Rayner & Best[37] present a partition of the Pearson statistic, based on a Barton-type smooth model, to test homogeneity of the row distributions against increasing trend.

Under the alternative hypothesis (4.8), we set the k -order alternative density as (4.19) and testing the hypothesis of homogeneity is once again equivalent to

$$H_0'' : \boldsymbol{\theta} = \vec{0}_k \quad \text{against} \quad H_1'' : \theta_{si} \neq 0 \text{ for some } (s, i),$$

where $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1I}, \dots, \theta_{k1}, \dots, \theta_{kI})'$ and the column totals $\{n_{.j}\}$ are nuisance parameters.

Let $\mathbf{Z} = (Z_{11}, \dots, Z_{IJ})'$ be the vector of standardised cell counts, that is

$$Z_{ij} = \frac{N_{ij} - \mu_{ij}}{\sqrt{\mu_{ij}}}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Then,

$$E[\mathbf{Z}] = \frac{1}{\sqrt{\mu_{ij}}} E[N_{ij} - \mu_{ij}] = \vec{0}_{IJ}$$

and, from subsection 2.3.3,

$$\begin{aligned} \text{var}[\mathbf{Z}] &= \frac{1}{n} \frac{1}{\sqrt{\pi_{i.} \pi_{.j}}} \frac{1}{\sqrt{\pi_{a.} \pi_{.b}}} \text{cov}[N_{ij}, N_{ab}] \\ &= \frac{n}{n-1} \begin{cases} (1 - \pi_{i.})(1 - \pi_{.j}), & i = a, \quad j = b, \\ -(1 - \pi_{i.}) \sqrt{\pi_{.j} \pi_{.b}}, & i = a, \quad j \neq b, \\ -\sqrt{\pi_{i.} \pi_{a.}} (1 - \pi_{.j}), & i \neq a, \quad j = b, \\ \sqrt{\pi_{i.} \pi_{a.}} \sqrt{\pi_{.j} \pi_{.b}}, & i \neq a, \quad j \neq b, \end{cases} \\ &= \frac{n}{n-1} [\mathbb{I}_I - \sqrt{\pi_{i.} \pi_{a.}}] \otimes [\mathbb{I}_J - \sqrt{\pi_{.j} \pi_{.b}}]. \end{aligned}$$

Rayner & Best[37] explain that under the null hypothesis, \mathbf{Z} asymptotically has multinormal distribution with mean $\vec{0}_{IJ}$ and variance matrix asymptotically equivalent to $\mathbb{I}_{(I-1)(J-1)} \oplus \mathbf{0}_{I+J-1}$ where $\mathbf{0}$ is the matrix of 0's. Then, by construction, under the null hypothesis, $\mathbf{Z}'\mathbf{Z}$ asymptotically has a $\chi^2(I-1)(J-1)$ distribution such that

$$\mathbf{Z}'\mathbf{Z} = X_P^2.$$

Let \mathbf{h}_s and \mathbf{H}_s be as (4.2) and (4.3). Let \mathbf{H} be the $IJ \times IJ$ matrix given by

$$\mathbf{H} = [\mathbf{H}_1 \quad \cdots \quad \mathbf{H}_J].$$

where \mathbf{H}_J is defined on $\mathbf{h}_J = \vec{1}_J$. Then, \mathbf{H} is an orthogonal matrix³.

Proof. Let $s, b = 1, \dots, J$. Then by the orthonormality of the polynomials $\{h_s(y)\}$,

$$\begin{aligned} \mathbf{H}'_s \mathbf{H}_b &= \text{diag}(\mathbf{h}'_s) \text{diag}(\mathbf{h}_b) \\ &= \text{diag}(\mathbf{h}'_s \mathbf{h}_b) \\ &= \begin{cases} \text{diag}(\sum h_s^2(y_j)), & s = b, \\ \text{diag}(\sum h_s(y_j) h_b(y_j)), & s \neq b = 1, \dots, J-1, \\ \text{diag}(\sum h_s(y_j)), & s \neq b = J, \end{cases} \\ &= \begin{cases} \mathbb{I}_I, & s = b, \\ \mathbf{0}_I, & s \neq b = 1, \dots, J, \end{cases} \end{aligned}$$

And so,

$$\mathbf{H}'\mathbf{H} = [\mathbf{H}'_s \mathbf{H}_b] = \mathbb{I}_{IJ}.$$

□

Let $\mathbf{V} = \mathbf{H}'\mathbf{Z}$. Then,

$$E[\mathbf{V}] = \mathbf{H}'E[\mathbf{Z}] = \vec{0}_{IJ} \quad \text{and} \quad \text{var}[\mathbf{V}] = \mathbf{H}' \text{var}[\mathbf{Z}] \mathbf{H}.$$

Then, under the null hypothesis, \mathbf{V} asymptotically has a multinormal distribution with mean $\vec{0}_{IJ}$ and variance matrix asymptotically equivalent to $\mathbb{I}_{(I-1)(J-1)} \oplus \mathbf{0}_{I+J-1}$ since by the orthogonality of \mathbf{H}

$$\text{var}[\mathbf{V}] = \mathbf{H}' \text{var}[\mathbf{Z}] \mathbf{H} \sim \mathbf{H}' (\mathbb{I}_{(I-1)(J-1)} \oplus \mathbf{0}_{I+J-1}) \mathbf{H} = \mathbb{I}_{(I-1)(J-1)} \oplus \mathbf{0}_{I+J-1}.$$

By construction, under the null hypothesis, $\mathbf{V}'\mathbf{V}$ asymptotically has a $\chi^2(I-1)(J-1)$ distribution such that

$$\mathbf{V}'\mathbf{V} = (\mathbf{Z}'\mathbf{H})(\mathbf{H}'\mathbf{Z}) = \mathbf{Z}'\mathbf{Z} = X_P^2.$$

Then, the Pearson statistic can be partitioned by the random vectors $\mathbf{V}_s = \mathbf{H}'_s \mathbf{Z}$, $s = 1, \dots, J$ such that $\mathbf{V}' = (\mathbf{V}'_1, \dots, \mathbf{V}'_J)$ and

$$X_P^2 = \mathbf{V}'\mathbf{V} = \mathbf{V}'_1 \mathbf{V}_1 + \cdots + \mathbf{V}'_{(J-1)} \mathbf{V}_{(J-1)}.$$

where $\mathbf{V}_J = \vec{0}_I$. The components \mathbf{V}_s are asymptotically iid with an $N(\vec{0}_I, \mathbb{I}_{I-1} \oplus 0)$ distribution and so the $\mathbf{V}'_s \mathbf{V}_s$ are asymptotically iid with a $\chi^2(I-1)$ distribution, $s = 1, \dots, J-1$.

Explicitly, the components are given by

$$\mathbf{V}_s = \mathbf{H}'_s \mathbf{Z} = (\mathbf{h}'_s \mathbf{Z}_1, \dots, \mathbf{h}'_s \mathbf{Z}_I)', \quad s = 1, \dots, J-1.$$

Similarly as Theorem 4.2, the i -th element of the s -th component, $V_{si} = \mathbf{h}'_s \mathbf{Z}_i$ can be derived as a score statistic to test the hypothesis $H''_{0(si)} : \theta_{si} = 0$ and can then detect deviations of the s -th moment for the i -th row, $s = 1, \dots, J-1$, $i = 1, \dots, I$.

Analogously, the s -th component \mathbf{V}_s can be derived as a score statistic to test the hypothesis $H''_{0s} : \boldsymbol{\theta}_s = \vec{0}_I$, where $\boldsymbol{\theta}_s = (\theta_{s1}, \dots, \theta_{sI})'$ and can then detect significant deviations of the s -th moment for the I rows simultaneously, $s = 1, \dots, J-1$.

³A matrix is orthogonal if its inverse is its transpose.

The extension of the Kruskal-Wallis statistic when there are no tied ranks Let X_{i1}, \dots, X_{in_i} be a random sample of the random variable X_i , $i = 1, \dots, I$, where the X_i are defined on the same outcome space. We would like to test the hypothesis of homogeneity for the distributions of the X_i , $i = 1, \dots, I$.

The observations can be arranged into a singly-ordered contingency table in the following way. Let R_{ij} denote the rank assigned to X_{ij} with respect to the unified sample $(X_{11}, \dots, X_{In_i})$, not allowing for ties. Let N_{ij} be the number of observations in the i -th sample which have the j -th rank in the unified sample:

$$N_{ij} = \sum_{a=1}^{n_i} I[R_{ia} = j] \quad i = 1, \dots, I, \quad j = 1, \dots, n. \quad (4.10)$$

Then the resulting $I \times n$ contingency table is of the same form of Table 4.3 where the cell counts are restricted to zero and one, and by design, both sets of marginal totals are fixed.

Table 4.3: A treatment-by-rank table (without ties)

	1	...	j	...	n	Total
X_1	N_{11}	...	N_{1j}	...	N_{1n}	n_1
\vdots	\vdots		\vdots		\vdots	\vdots
X_i	N_{i1}	...	N_{ij}	...	N_{in}	n_i
\vdots	\vdots		\vdots		\vdots	\vdots
X_I	N_{I1}	...	N_{Ij}	...	N_{In}	n_I
Total	1	...	1	...	1	n

For Table 4.3, no matter the values of the cell counts given by (4.10), the Pearson statistic is constant.

Proof. For Table 4.3, the ML estimators for the expected cell counts under the null hypothesis are

$$\mu_{ij}^{(0)} = \frac{n_i \cdot n_{\cdot j}}{n} = \frac{n_i}{n} \quad i = 1, \dots, I, \quad j = 1, \dots, n$$

since $n_{\cdot j} = 1$, $j = 1, \dots, n$. Also, we note that since N_{ij} are restricted to 0 and 1, $N_{ij}^2 = N_{ij}$, $i = 1, \dots, I$, $j = 1, \dots, J$.

$$\begin{aligned} X_P^2 &= \sum_{i=1}^I \sum_{j=1}^n \frac{(N_{ij} - \hat{\mu}_{ij}^{(0)})^2}{\hat{\mu}_{ij}^{(0)}} \\ &= \sum_{i=1}^I \sum_{j=1}^n \frac{(N_{ij} - n_i/n)^2}{n_i/n} \\ &= \sum_{i=1}^I \sum_{j=1}^n \frac{N_{ij}^2 - 2N_{ij}n_i/n + n_i^2/n^2}{n_i/n} \\ &= n \sum_{i=1}^I \frac{1}{n_i} \sum_{j=1}^n N_{ij} - 2 \sum_{i=1}^I \sum_{j=1}^n N_{ij} + \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^n n_i \\ &= In - 2n + n \\ X_P^2 &= (I - 1)n \end{aligned}$$

□

While the Pearson statistic is never appropriate to use when doing a test of homogeneity in Table 4.3, it can still be partitioned by the above decomposition now that both sets of marginal totals are fixed.

We choose $\{h_s(y)\}$ to be the set of polynomials, defined as in section 3.4, orthonormal on the discrete uniform distribution on $\{1, \dots, n\}$ since the ordered column scores are the ranks $1, \dots, n$, $n_{.j} = 1$ and thus $\pi_{.j} = 1/n$, $j = 1, \dots, n$. Then, it can be shown that given there are no ties, the first component for Table 4.3 is the Kruskal-Wallis statistic.

Proof. The proof is done by Rayner & Best[37]. □

As the Kruskal-Wallis statistic detects significant deviations in the shifts in the medians, by section 4.2.1, the subsequent components will detect shifts in the dispersion, skewness, etc.

The extension of the Kruskal-Wallis statistic when there are tied ranks Let X_{i1}, \dots, X_{in_i} be a random sample of the random variable X_i , $i = 1, \dots, I$, where the X_i are defined on the same outcome space. We would like to test the hypothesis of homogeneity for the distributions of the X_i , $i = 1, \dots, I$.

The observations can be arranged into a singly-ordered contingency table in the following way. Let R_{ia} denote the rank assigned to X_{ia} with respect to the unified sample $(X_{11}, \dots, X_{In_I})$. We denote the mid-ranks by $y_1 < \dots < y_J$. Let N_{ij} be the number of observations in the i -th sample which have the j -th rank in the unified sample:

$$N_{ij} = \sum_{a=1}^{n_i} I[R_{ia} = y_j] \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Then the resulting $I \times J$ contingency table is of the same form of Table 2.2. For this contingency table, the above decomposition of the Pearson statistic, which will no longer be constant, is still valid, however the set of polynomials $\{h_s(y)\}$, defined on the scores/midranks $y_1 < \dots < y_J$, will be orthonormal on the column probabilities $\{\pi_{.j} = n_{.j}/n\}$, and the first component is a multiple of the Kruskal-Wallis statistic, now adjusted for ties.

Proof. Let $R_{i.}$ denote the sum of the ranks for the i -th sample, $i = 1, \dots, I$. Since the mid-ranks y_j are defined as the arithmetic mean of the positions of coinciding observations⁴

$$R_{i.} = \sum_{a=1}^{n_i} R_{ia} = \sum_{j=1}^J N_{ij} y_j \quad i = 1, \dots, I.$$

Let $\bar{R}_{i.} = R_{i.}/n_i$ denote the average of the ranks associated to the i -th sample, $i = 1, \dots, I$, and let $\bar{R} = (n+1)/2$ denote the overall average of the ranks in the united sample.

The mean and variance for the column variable Y are defined as

$$\mu_Y = \sum_{j=1}^J y_j \pi_{.j} = \frac{n+1}{2} \quad \text{and} \quad \sigma_Y^2 = \sum_{j=1}^J (y_j - \mu_Y)^2 \pi_{.j}$$

where y_1, \dots, y_J have probabilities $\pi_{.1}, \dots, \pi_{.J}$, $\pi_{.j} > 0$. Let $\{h_s\}$ be the set of polynomials defined as in section 3.4. That is

$$h_0(y) = 1, \quad \text{and} \quad h_1(y) = \frac{y - \mu_Y}{\sigma_Y}.$$

⁴see subsection 3.2.1.

The numerator of the Kruskal-Wallis statistic is

$$\begin{aligned}
 (n-1) \sum_{i=1}^I n_i (\bar{R}_i - \bar{R})^2 &= (n-1) \sum_{i=1}^I n_i \left(\frac{R_{i\cdot}}{n_i} - \frac{n+1}{2} \right)^2 \\
 &= (n-1) \sum_{i=1}^I \frac{1}{n_i} \left(R_{i\cdot} - n_i \frac{n+1}{2} \right)^2 \\
 &= (n-1) \sum_{i=1}^I \frac{1}{n_i} \left(\sum_{j=1}^J N_{ij} (y_j - \hat{\mu}_Y) \right)^2.
 \end{aligned} \tag{4.11}$$

The denominator of the Kruskal-Wallis statistic is

$$\begin{aligned}
 \sum_{i=1}^I \sum_{a=1}^{n_i} (R_{ia} - \bar{R})^2 &= \sum_{i=1}^I \sum_{a=1}^{n_i} \left(R_{ia} - \frac{n+1}{2} \right)^2 \\
 &= \sum_{i=1}^I \sum_{j=1}^J N_{ij} \left(y_j - \frac{n+1}{2} \right)^2 \\
 &= \sum_{j=1}^J N_{\cdot j} \left(y_j - \frac{n+1}{2} \right)^2 \\
 &= n \hat{\sigma}_Y^2
 \end{aligned} \tag{4.12}$$

Then, with (4.11) and (4.12) the Kruskal-Wallis statistic, adjusted for ties, can be written as

$$\begin{aligned}
 S_{KW} &= (n-1) \frac{\sum_{i=1}^I n_i (\bar{R}_i - \bar{R})^2}{\sum_{i=1}^I \sum_{a=1}^{n_i} (R_{ia} - \bar{R})^2} \\
 &= \frac{n-1}{n} \sum_{i=1}^I \frac{1}{n_i} \left(\sum_{j=1}^J N_{ij} \left(\frac{y_j - \hat{\mu}_Y}{\hat{\sigma}_Y} \right) \right)^2 \\
 &= \frac{n-1}{n} \sum_{i=1}^I \hat{V}_{1i}^2 \\
 &= \frac{n-1}{n} \hat{\mathbf{V}}_1' \hat{\mathbf{V}}_1.
 \end{aligned}$$

which is a multiple of the first component of the decomposition. \square

An example of the decomposition

Example 4.3. (Example 3.2 cont'd.) Investigating the granular composition of quartz in Lithuanian and Saharan sand samples, the data on the lengths of the maximal axis of the quartz grains were used. In section 3.1 the Pearson test rejected the hypothesis that the length of the maximum axis has the same distribution in Lithuanian and Saharan sand. From the table of cumulative cell probabilities,

Interval midpoints	11	17	21	25	29	33	37	45
Lithuanian Sand	0.0796	0.2537	0.5572	0.8159	0.9303	0.9652	0.9851	1.0000
Saharan Sand	0.0588	0.1569	0.2745	0.4020	0.5196	0.6667	0.7843	1.0000

we suspected that an appropriate follow-up test would compare the hypothesis of homogeneity against decreasing trend in the row distributions. In particular, we test the hypotheses

$$H'_0 : \tau_{s1} = \tau_{s2} \quad \text{vs.} \quad H'_1 : \tau_{s1} \geq \tau_{s2} \quad s = 1, \dots, 8.$$

In this situation, we can use the decomposition of the Pearson statistic from section 4.2.3 which extends the Kruskal-Wallis statistic.

For the decomposition of the Pearson we change the column scores to be the mid-ranks for the unified sample, i.e. the 22 grains that are in the interval with midpoint length of 11 have mid-rank $1 + (22 - 1)/2 = 11.5$, the following 45 have mid-rank $23 + (45 - 1)/2 = 45$, and so on.

Mid-ranks	11.5	45	99	173	223	251.5	269.5	291	Total
Lithuanian Sand	16	35	61	52	23	7	4	3	201
Saharan Sand	6	10	12	13	12	15	12	22	102
Total	22	45	73	65	35	22	16	25	303

We calculate the first three components of the statistics and their associated p -values.

The Pearson statistic			
Component	df	Value	P-value
$\hat{\mathbf{V}}'_1 \cdot \hat{\mathbf{V}}_1$	1	42.78	$\ll 0.0001$
$\hat{\mathbf{V}}'_2 \cdot \hat{\mathbf{V}}_2$	1	30.75	$\ll 0.0001$
$\hat{\mathbf{V}}'_3 \cdot \hat{\mathbf{V}}_3$	1	0.45	0.5035
Remainder	4	1.24	0.8709
X^2_P	7	75.21	$\ll 0.0001$

The Pearson statistic for testing homogeneity of the row distributions is valued at 75.21 with 7 degrees of freedom. With a p -value significantly smaller than 0.0001, the Pearson test suggest there is an association. The highly significant first and second components suggest there are associations between the medians and dispersions. In fact, with the median maximal axis length of 21 and 29 cm for the Lithuanian and Saharan grains, respectively, we can say that the Lithuanian grains have significant smaller maximal axis length than the Saharan grains. Additionally, with sample standard deviations of 6.22 and 10.34 cm, the Lithuanian grains have significant less variability in the maximal axis lengths than the Sahara grains.

4.2.4 Under the binomial model

For a row-binomial model, the test of homogeneity against increasing trend is equivalent to testing

$$H'_0 : \pi_{1j} = \dots = \pi_{Ij}$$

$$H'_1 : \pi_{1j} \leq \dots \leq \pi_{Ij} \quad j = 1, \dots, J$$

with at least one inequality in the alternative H'_1 .

An extension of the median test

Let X_{i1}, \dots, X_{in_i} be a random sample of the random variable X_i , $i = 1, \dots, I$, where the X_i are defined on the same outcome space. We would like to test the hypothesis of homogeneity for the distributions of the X_i , $i = 1, \dots, I$.

The observations can be arranged into a singly-ordered contingency table in the following way. We choose a median M of the unified sample X_{11}, \dots, X_{In_I} such that $a = \lceil 0.50n \rceil$ of the observations are less or equal to M and $b = n - a$ are greater than M ⁵. Let N_{i1} denote the number of observations in the i -th sample less than M and let $N_{i2} = n_i - N_{i1}$ denote the number of observations in the i -th sample greater than M :

$$N_{i1} = \sum_{a=1}^{n_i} I[X_{ia} \leq M], \quad N_{i2} = \sum_{a=1}^{n_i} I[X_{ia} > M], \quad i = 1, \dots, I.$$

Then the resulting $I \times 2$ contingency table is as Table 4.4. While this contingency table has a row-binomial model where the column totals are also fixed by construction. Thus, the decomposition presented in section 4.2.3 can be used and it can be shown that the first component for Table 4.4 is the Median statistic as given by Conover[17]⁶.

Table 4.4: The $I \times 2$ table around the overall sample median

	1	2	Total
X_1	N_{11}	N_{12}	n_1
\vdots	\vdots	\vdots	\vdots
X_i	N_{i1}	N_{i2}	n_i
\vdots	\vdots	\vdots	\vdots
X_I	N_{I1}	N_{I2}	n_I
Total	a	b	n

Proof. The proof is done by Rayner & Best[37]. □

While the Median statistic detects shifts in the medians, the subsequent components will detect shifts in the dispersions, skewness, etc.

4.2.5 Under the randomised design model

Suppose we have a randomised block design model

$$X_{ib} = \mu + \alpha_i + \beta_b + \varepsilon_{ib} \quad i = 1, \dots, I, \quad b = 1, \dots, B$$

where

- μ is the overall mean;

⁵ $a = b = n/2$ if n is even; $a = (n + 1)/2$ and $b = (n - 1)/2$ if n is odd.

⁶“The median test is designed to examine whether several samples came from populations having the same median.” (Conover[17]).

- $\alpha_1, \dots, \alpha_I$ are the treatment effects;
- β_1, \dots, β_B are the block effects;
- $\varepsilon_{11}, \dots, \varepsilon_{IB}$ are the iid standard normal error terms.

Example 4.4. Suppose we have I independent varieties of a product (treatments) that we want ranked, not allowing for ties, by B consumers (blocks). Then $(R_{1b}, \dots, R_{Ib})'$ is the ranking of the I varieties by the b -th consumer, $b = 1, \dots, B$.

We arrange the data according to Schach's[41] method. For the b -th block, let $N_{ij}^{(b)}$ denote the contribution of the i -th treatment to the m -way tie at the j -th rank

$$N_{ij}^{(b)} = \begin{cases} 1/m & \text{if treatment } i \text{ is tied in a group of } m \text{ tied observations including rank } j \\ 0 & \text{otherwise} \end{cases}$$

where $m = 1$ if the i -th treatment is not tied with any other treatment. The $I \times I$ table of the $N_{ij}^{(b)}$ is the treatment-by-rank for the b -th block. The resulting table of the $N_{ij} = \sum_{k=1}^b N_{ij}^{(k)}$ is the singly-ordered $I \times I$ treatment-by-rank table of *fractional counts*.

We would like to test the hypothesis of homogeneity of the distributions of the treatments, X_i , $i = 1, \dots, I$. Using the Pearson statistic in this $I \times I$ contingency table is not appropriate. Under the permutation model, the usual model when both sets of marginal totals are fixed, the cell counts would be associated to the permutation of $n = BI$ objects. However, in this situation, the blocks are independent for whom there are $I!$ possible ways to rank the treatments. For this contingency table, Anderson[3] shows that

$$A = \frac{I-1}{I} X_P^2$$

is the appropriate chi-squared statistic, which has a $\chi^2(I-1)^2$ distribution, unlike the Pearson statistic.

The decomposition of the Anderson statistic when there are no tied ranks

Not allowing for tied ranks, the counts of the treatment-by-rank for the b -th block are given by

$$N_{ij}^{(b)} = \begin{cases} 1 & R_{ib} = j \\ 0 & \text{otherwise} \end{cases}$$

and the count in the treatment-by-rank table are

$$N_{ij} = \sum_{b=1}^B I [R_{ib} = j], \quad i, j = 1, \dots, I$$

which will always be integers. Since there are no ties,

- each treatment receives a distinct rank in each block: the row totals are all B ;
- each rank is given to only one treatment in each block: the column totals are all B .

All the marginal totals are fixed at $n_{i\cdot} = n_{\cdot j} = B$, $i, j = 1, \dots, I$, and the total number of observations is $n = BI$. The resulting $I \times I$ treatment-by-rank table formed from the counts N_{ij} is as Table 4.5 which is an $I \times I$ contingency table under a row-multinomial model with fixed column totals.

Best[7] presents a partition of the Anderson statistic derived from a Barton-type smooth model to test the hypothesis of homogeneity of the treatments. Under the alternative hypothesis (4.8), we set the

Table 4.5: A treatment-by-rank table (without ties)

	1	...	j	...	I	Total
Treatment 1	N_{11}	...	N_{1j}	...	N_{1I}	B
\vdots	\vdots		\vdots		\vdots	\vdots
Treatment i	N_{i1}	...	N_{ij}	...	N_{iI}	B
\vdots	\vdots		\vdots		\vdots	\vdots
Treatment I	N_{I1}	...	N_{Ij}	...	N_{II}	B
Total	B	...	B	...	B	BI

alternative we set the k -order alternative density as (4.19) and testing the hypothesis of homogeneity is once again equivalent to

$$H_0'' : \boldsymbol{\theta} = \vec{0}_k \quad \text{against} \quad H_1'' : \theta_{si} \neq 0 \text{ for some } (s, i),$$

where $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1I}, \dots, \theta_{k1}, \dots, \theta_{kI})'$ and the column totals $\{n_{.j}\}$ are nuisance parameters.

By the same approach as in section 4.2.3, the components for the Anderson statistic are

$$\mathbf{V}_s = \sqrt{\frac{I-1}{I}} \mathbf{H}'_s \mathbf{Z}_i \quad s = 1, \dots, I-1, \quad i = 1, \dots, I$$

where \mathbf{H}_s is defined on the functions $\{h_s(j)\}$ orthonormal on the fixed and equal column probabilities $\pi_{.j} = 1/I$, $j = 1, \dots, I$. The $\mathbf{V}'_s \mathbf{V}_s$ are asymptotically iid with a $\chi^2(I-1)$ distribution such that

$$\mathbf{V}'_1 \mathbf{V}_1 + \dots + \mathbf{V}'_{(I-1)} \mathbf{V}_{(I-1)} = A$$

The extension We choose $\{h_s(j)\}$ to be the set of polynomials orthonormal on the discrete uniform distribution on $\{1, \dots, I\}$, as in Example 3.3. Then, it can be shown that given there are no ties, the first component for Table 4.5 is the Friedman statistic.

Proof. Let $R_{i.}$ denote the sum of the ranks for the i -th treatment, $i = 1, \dots, I$

$$R_{i.} = \sum_{b=1}^B R_{ib} = \sum_{j=1}^I j N_{ij} \quad i = 1, \dots, I.$$

Let $\bar{R}_{i.} = R_{i.}/B$ denote the average of the ranks associated to the i -th treatment, $i = 1, \dots, I$, and let $\bar{R} = (I+1)/2$ denote the overall average of the ranks.

The mean and variance for the column variable Y are defined as

$$\mu_Y = \sum_{j=1}^I \frac{1}{I} j = \frac{I+1}{2} \quad \text{and} \quad \sigma_Y^2 = \sum_{j=1}^I \frac{1}{I} (j - \mu_Y)^2.$$

When there are no tied ranks, the numerator of the Friedman statistic is

$$\begin{aligned}
B \sum_{i=1}^I (\bar{R}_i - \bar{R})^2 &= B \sum_{i=1}^I \left(\frac{R_{i.}}{B} - \frac{I+1}{2} \right)^2 \\
&= \sum_{i=1}^I \frac{1}{B} \left(R_{i.} - B \frac{I+1}{2} \right)^2 \\
&= \sum_{i=1}^I \frac{1}{B} \left(\sum_{j=1}^I N_{ij} (j - \hat{\mu}_Y) \right)^2
\end{aligned} \tag{4.13}$$

and the denominator of the Friedman statistic is

$$\begin{aligned}
\frac{1}{B(I-1)} \sum_{i=1}^I \sum_{b=1}^B (R_{ib} - \bar{R})^2 &= \frac{1}{B(I-1)} \sum_{i=1}^I \sum_{b=1}^B \left(R_{ib} - \frac{I+1}{2} \right)^2 \\
&= \frac{1}{B(I-1)} \sum_{i=1}^I \sum_{j=1}^I N_{ij} \left(y - \frac{I+1}{2} \right)^2 \\
&= \frac{1}{B(I-1)} \sum_{i=1}^I n_{.j} \left(j - \frac{I+1}{2} \right)^2 \\
&= \frac{I}{I-1} \hat{\sigma}_Y^2.
\end{aligned} \tag{4.14}$$

Then, with (4.13) and (4.14) the Friedman statistic, when there are tied ranks, can be written as

$$\begin{aligned}
S_F &= \frac{B \sum_{i=1}^I (\bar{R}_i - \bar{R})^2}{\frac{1}{B(I-1)} \sum_{i=1}^I \sum_{b=1}^B (R_{ib} - \bar{R})^2} \\
&= \frac{I-1}{I} \sum_{i=1}^I \frac{1}{B} \left(\sum_{j=1}^J N_{ij} \frac{j - \hat{\mu}_Y}{\hat{\sigma}_Y} \right)^2 \\
&= \hat{\mathbf{V}}_1' \hat{\mathbf{V}}_1.
\end{aligned}$$

which is the first component of the decomposition. \square

As the Friedman statistic detects significant deviations in the means, by section 4.2.1, the s -th component will detect significant deviations of the s -th moment, $s = 1, \dots, J-1$.

The decomposition of the Anderson statistic when there are tied ranks

Suppose we allow for ties. Then in the treatment-by-rank table for the b -th block, we have

- for the i -th row, $N_{ij}^{(b)} = 1/m$ for the m consecutive ranks which are tied with the j -th rank and zero otherwise.
- for the j -th column, $N_{ij}^{(b)} = 1/m$ for the m treatments included in the m -way tie including the j -th rank and zero otherwise.

Thus, for the treatment-by-rank table for the b -th block, all the row and column totals, denoted by $n_{i.}^{(b)}$ and $n_{.j}^{(b)}$ respectively, are equal 1 and the total number of observations is I . The row totals for the

treatment-by-rank fractional counts are then given by

$$n_{i\cdot} = \sum_{j=1}^I N_{ij} = \sum_{j=1}^I \sum_{b=1}^B N_{ij}^{(b)} = \sum_{b=1}^B n_{i\cdot}^{(b)} = B, \quad i = 1, \dots, I.$$

Similarly, the column totals are given by

$$n_{\cdot j} = \sum_{i=1}^I N_{ij} = \sum_{i=1}^I \sum_{b=1}^B N_{ij}^{(b)} = \sum_{b=1}^B n_{\cdot j}^{(b)} = B, \quad j = 1, \dots, I.$$

Although the entries in the $I \times I$ table of treatment-by-rank fractional counts are non-integer valued, the marginal totals are still equal to B and the total number of observations is

$$n = \sum_{i=1}^I \sum_{j=1}^I N_{ij} = \sum_{i=1}^I n_{i\cdot} = BI$$

like in the case of ranked data without ties.

For the b -th block, let $U_{ij}^{(b)}$ denote the contribution of the i -th rank to the m -way tie at the j -th rank

$$U_{ij}^{(b)} = \begin{cases} 1/m & \text{if rank } i \text{ is tied in a group of } m \text{ tied observations including rank } j \\ 0 & \text{otherwise} \end{cases}$$

As it is, $U_{ij}^{(b)}$ is a function of the fractional counts through the expression $U_{ij}^{(b)} = \sum_{a=1}^I N_{ai}^{(b)} N_{aj}^{(b)}$. The $I \times I$ matrix of the $U_{ij}^{(b)}$ is the *tie structure* for the b -th block. The resulting $I \times I$ matrix of the $U_{ij} = \sum_{b=1}^B U_{ij}^{(b)}$ is the tie structure matrix \mathbf{U} for the $I \times I$ table of fractional counts. The i -th row total for the tie structure for b -th block, $i = 1, \dots, I$, $b = 1, \dots, B$, is

$$U_{i\cdot}^{(b)} = \sum_{j=1}^I U_{ij}^{(b)} = \sum_{j=1}^I \sum_{a=1}^I N_{ai}^{(b)} N_{aj}^{(b)} = \sum_{a=1}^I N_{ai}^{(b)} \sum_{j=1}^I N_{aj}^{(b)} = n_{i\cdot}^{(b)} n_{a\cdot}^{(b)} = 1$$

and the row totals for the tie structure \mathbf{U} are fixed at B , $i = 1, \dots, I$. Similarly, the column totals, $U_{\cdot j}^{(b)} = 1$ such that the column totals for the tie structure \mathbf{U} are also fixed at B , $j = 1, \dots, I$.

Assuming there are no tied ranks, Brockhoff et al[16] generalise the Anderson statistic and present a partition of the generalised Anderson statistic based on smooth models to test the hypothesis of homogeneity of the treatments, conditional on the tie structure matrix \mathbf{U} .

Let \mathbf{R} be the $I \times I$ matrix given by

$$\mathbf{R} = \frac{1}{B} \mathbf{U} - \frac{1}{I} \vec{1}_I \vec{1}'_I$$

With ties in the data, the matrix \mathbf{U} may have less than full rank I . However, if for all the pairs of consecutive ranks, $(1, 2), (2, 3), \dots, (I-1, I)$, there is a block where the pair is not tied together, then $\text{rank}(\mathbf{U}) = I$.

Theorem 4.4. *Assume that all consecutive pairs of ranks are untied once. Let the generalised Anderson statistic be defined as*

$$A_G = \frac{I-1}{I} \sum_{i=1}^I \tilde{\mathbf{Z}}'_i \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{Z}}_i.$$

where $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iI})'$ is the vector of standardized fractional counts for the i -th treatment:

$$Z_{ij} = \frac{N_{ij} - B/I}{\sqrt{B/I}}, \quad i, j = 1, \dots, I,$$

$\tilde{\mathbf{Z}}_i$ is the vector obtained by removing the last element of \mathbf{Z}_i and $\tilde{\mathbf{R}}$ is \mathbf{R} with the last row and column removed.

- The asymptotic distribution of A_G as $n \rightarrow \infty$, conditional on the tie structure, is a chi-squared distribution with $(I - 1)^2$ degrees of freedom.
- When there are no ties in the data, the generalized statistic A_G simplifies to the usual Anderson statistic.

The extension Let $\{h_s(y) : s = 1, \dots, q - 1\}$, where $\text{rank}(\mathbf{U}) = q \leq I$, be the set of polynomials orthogonal with respect to the probabilities given by $\pi_{ij} = U_{ij}/I$, $i, j = 1, \dots, I$, i.e.

$$\sum_{j=1}^I h_s(j) = 0 \quad \text{and} \quad \mathbf{h}_s \mathbf{U} \mathbf{h}_b = I^2 \delta_{sb} \quad s, b = 1, \dots, q - 1$$

and let \mathbf{H}_s be defined as (4.3).

Theorem 4.5. The components $\mathbf{V}'_1, \mathbf{V}_1, \dots, \mathbf{V}'_{(q-1)}, \mathbf{V}_{(q-1)}$ where \mathbf{V}_s is defined as

$$\mathbf{V}_s = \sqrt{\frac{I-1}{I}} \mathbf{H}'_s \mathbf{Z}, \quad s = 1, \dots, q - 1,$$

are asymptotically mutually independent $\chi^2(q - 1)$ variables and constitute a decomposition of the generalized Anderson statistic via

$$A_G = \sum_{s=1}^{q-1} \mathbf{V}'_s \mathbf{V}_s.$$

The first component $\mathbf{V}'_1, \mathbf{V}_1$ is the Friedman rank statistic corrected for ties

$$\mathbf{V}'_1 \mathbf{V}_1 = \frac{12/BI(I-1)}{1 - \sum_{b=1}^B \sum_{j=1}^I \left(\left(U_{jj}^{(b)} \right)^2 - 1 \right) / BI(I^2 - 1)} \sum_{i=1}^I R_i^2 - 3B(I+1)$$

where R_i is the mid-rank sum for the i -th treatment, $i = 1, \dots, t$.

The proofs for Theorem 4.4 and Theorem 4.5 are done by Brockhoff et al[16].

If we use this decomposition when there are no ties, the treatment-by-rank table is simply Table 4.5, the tie structure matrix is $\mathbf{U} = \mathbb{I}_I$, $A_G = A$, this becomes the decomposition from Best[7] and the first component is the again the Friedman statistic, adjusted for no ties.

4.2.6 Decomposition of the Pearson Statistic in Three-Way Tables

Three-way contingency tables

Let X, Y and Z be categorical variables with I, J and K categories, respectively. Suppose $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$ is a sample of the random vector (X, Y, Z) . Then the cell counts for the $I \times J \times K$ contingency table are given by

$$N_{ijk} = \sum_{l=1}^n 1[X_l \in A_i, Y_l \in B_j, Z_l \in C_k] \quad \text{for all } i, j, k$$

and the associated cell probabilities are

$$\pi_{ijk} = P(X \in A_i, Y \in B_j, Z \in C_k) \quad \text{for all } i, j, k.$$

The three models for the two-way contingency tables from section 2.3 can be generalised to three-way contingency tables (Lancaster [27]):

1. *Unrestricted Sampling.* n independent observations are made on the parent population.
2. *Product-Multinomial.*
 - (a) $n_{i.}$ observations are chosen arbitrarily from the i -th row but no selection is made with respect to columns or layers. The row totals satisfy $n_{1.} + \dots + n_{I.} = n$.
 - (b) $n_{.j}$ observations are chosen arbitrarily from the intersection of the i -th row and the j -th column but no selection is made with respect to layers. The $\{n_{ij.}\}$ satisfy $n_{11.} + \dots + n_{IJ.} = n$.
3. *Permutation.* The one dimension frequencies are given, namely $\{n_{i.}\}$, $\{n_{.j}\}$ and $\{n_{..k}\}$, such that $n_{1.} + \dots + n_{I.} = n_{.1} + \dots + n_{.J} = n_{..1} + \dots + n_{..K} = n$.

The Pearson and chi-squared LRT statistics for testing homogeneity and independence in an $I \times J \times K$ contingency table are given by

$$X_P^2 = \sum_{i,j,k} \frac{(N_{ijk} - \hat{\mu}_{ijk}^{(0)})^2}{\hat{\mu}_{ijk}^{(0)}} \quad \text{and} \quad G^2 = 2 \sum_{i,j,k} N_{ijk} \ln \left(\frac{N_{ijk}}{\hat{\mu}_{ijk}^{(0)}} \right)$$

where $\hat{\mu}_{ijk}^{(0)}$ are the ML estimates of the expected cell counts μ_{ijk} under the null hypothesis. Once again, the Pearson and chi-squared LRT statistics are asymptotically equivalent under the same null hypothesis now with an asymptotic $\chi^2[(I-1)(J-1)(K-1)]$ distribution as the sample sizes increase.

From a three-way contingency table, we can obtain many two-way contingency tables in which we can study the relationship of two of the variables given the third (Agresti [1]). For example, to study the relationship of X and Y given Z :

- We have K *partial tables* which cross-classify X and Y at the separate levels of Z . In which case, the cell counts for the k -th partial table are given by

$$N_{ij(k)} = \sum_{l=1}^n 1[X_l \in A_i, Y_l \in B_j, Z_l \in C_k] \quad i = 1, \dots, I \quad j = 1, \dots, J,$$

which have associated cell probabilities given by

$$\pi_{ij(k)} = P(X \in A_i, Y \in B_j | Z \in C_k) \quad i = 1, \dots, I \quad j = 1, \dots, J.$$

- We have the *marginal table*, obtained by combining the partial tables, which has cell counts given by

$$N_{ij.} = \sum_{l=1}^n 1[X_l \in A_i, Y_l \in B_j] \quad i = 1, \dots, I \quad j = 1, \dots, J,$$

which have associated cell probabilities given by

$$\pi_{ij.} = P(X \in A_i, Y \in B_j) \quad i = 1, \dots, I \quad j = 1, \dots, J.$$

The associations in partial tables are called *conditional associations* because they refer to the effect of X on Y conditional on fixing Z at some level while the associations in the marginal table, called *marginal associations*, refer to the effect of X on Y while ignoring Z .

To test conditional and marginal associations with respect to Z , we can use the Pearson and chi-squared LRT statistic from section 2.4. Then, the partition from subsection 4.2.1 can be used for studying the conditional and marginal associations. However, we cannot describe the relationship of the three variables solely on the two-variable relationships.

Extension of the decomposition of the Pearson statistic

Suppose we want to test the relationship in an $I \times J \times K$ contingency table. Let $\{x_i\}$ be the scores for the row variable X , $\{y_j\}$ the scores for the column variable Y and $\{z_k\}$ the scores for the layer variable Z . We set the Barton-type model for the trivariate density function $f_{XYZ}(x, y, z)$ and null density $f_0(x, y, z)$. The (k_1, k_2, k_3) -order alternative density function is

$$f_{XYZ}(x_i, y_j, z_k) = \left\{ 1 + \sum_{t=1}^{k_3} \sum_{s=1}^{k_2} \sum_{r=1}^{k_1} \theta_{tsr} c_t(z_k) b_s(y_j) a_r(x_i) \right\} f_0(x_i, y_j, z_k) \quad \text{for all } i, j, k$$

where

- $k_1 \leq I - 1$, $k_2 \leq J - 1$ and $k_3 \leq K - 1$;
- $\{\theta_{tsr}\}$ are real valued parameters,
- $\{c_t\}$ are functions orthonormal with respect to the layer probabilities $\{\pi_{..k}\}$,
- $\{b_s\}$ are functions orthonormal with respect to the column probabilities $\{\pi_{.j}\}$, and
- $\{a_r\}$ are functions orthonormal with respect to the row probabilities $\{\pi_{i..}\}$.

The hypothesis $H_0 : f_{XYZ}(x, y, z) = f_0(x, y, z)$ is now equivalent to $H_0'' : \boldsymbol{\theta} = \vec{0}_{k_1 k_2 k_3}$ where $\boldsymbol{\theta}$ is the $k_1 k_2 k_3$ -dimensional vector of the θ_{tsr} in hierarchical order and the marginal probabilities $\{\pi_{i..}\}$, $\{\pi_{.j}\}$, and $\{\pi_{..k}\}$ are nuisance parameters. Then, for $k_1 = I - 1$, $k_2 = J - 1$ and $k_3 = K - 1$, similarly as subsection 4.2.1, the Pearson statistic is the score statistic to test the model

$$f_{XYZ}(x_i, y_j, z_k) = \left\{ 1 + \sum_{t=1}^{K-1} \sum_{s=1}^{I-1} \sum_{r=1}^{I-1} \theta_{tsr} c_t(z_k) b_s(y_j) a_r(x_i) \right\} f_0(x_i, y_j, z_k)$$

for $i = 1, \dots, I - 1$, $j = 1, \dots, J - 1$ and $k = 1, \dots, K - 1$, and, by extension of Property 4.1 for three-way contingency tables, it has components

$$\hat{V}_{tsr} = \frac{1}{\sqrt{n}} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K N_{ijk} \hat{c}_t(z_k) \hat{b}_s(y_j) \hat{a}_r(x_i) \quad t = 1, \dots, K - 1, s = 1, \dots, J - 1, r = 1, \dots, I - 1$$

which, under the null hypothesis, are asymptotically iid with a $N(0, 1)$ distribution, such that

$$X_P^2 = \sum_{t=1}^{K-1} \sum_{s=1}^{I-1} \sum_{r=1}^{I-1} \hat{V}_{tsr}^2.$$

Proof. The proof is provided by Rayner & Best [33]. □

By the asymptotic independence of the components, we can construct statistics that have good power against directional alternatives on the elements of θ since the \hat{V}_{tsr} are score statistics for testing $\theta_{tsr} = E[c_t(Z)b_s(Y)a_r(X)]$ under the appropriate model, for all t, s and r .

Similarly as section 4.2.1, if $\{c_t\}$, $\{b_s\}$ and $\{a_r\}$ are the orthonormal polynomials obtained from section 3.4, the components \hat{V}_{tsr} will detect deviations for trivariate central moments under the unrestricted sampling model, and bivariate or univariate central moments under the product-multinomial models.

This partition of the Pearson statistic for three-way contingency tables is a direct extension of the partition of the Pearson statistic by smooth models for two-way contingency tables from subsection 4.2.1. From the approach for three-way contingency tables, we can see this can be extended to m -way contingency tables, $m = 4, 5, \dots$

Adjustments to this partition of the Pearson statistic for testing the relationship in a three-way contingency tables in the presence of at least one ordinal variable are discussed by Beh & Davy [10, 11], Rayner & Bey [33] and Rayner & Best [34].

4.3 The Decomposition of the Cumulative Chi-Squared Statistic

Suppose we have an $I \times J$ contingency table under the product-multinomial model with an ordinal variable. We are interested in the test of homogeneity against monotonic trend. As stated in section 3.1, the Pearson statistic does not take into account the order of the categories of the ordinal variable. Without loss of generality, assuming the column variable is ordinal with increasing scores, Taguchi[45, 46] developed a statistic that does account for the order by considering the cumulative sum of the cell counts across the columns.

For the i -th row, let M_{si} denotes the cumulative count up to the s -th column:

$$M_{si} = \begin{cases} N_{i1} + \dots + N_{is} & s = 1, \dots, J-1, \\ N_{i.} & s = J, \end{cases}$$

let $M_s.$ denotes the cumulative column total for the s -th column:

$$M_s. = \begin{cases} N_{.1} + \dots + N_{.s} & s = 1, \dots, J-1, \\ n & s = J, \end{cases}$$

and let $\tau_s. = M_s./n$ denote the cumulative column proportion, $s = 1, \dots, J$. Then, the statistic proposed by Taguchi[45, 46] is given by

$$T_E = \sum_{s=1}^{J-1} \sum_{i=1}^I \frac{N_{i.}(M_{si}/N_{i.} - \tau_s.)^2}{\tau_s.(1 - \tau_s.)}. \quad (4.15)$$

There is a link between the Taguchi and Pearson statistics. Conditional on both sets of marginal totals, we can write

$$T_E = \sum_{s=1}^{J-1} X_s^2$$

where X_s^2 is the Pearson statistic for the $I \times 2$ contingency table as Table 4.6 that compares the sum of the s first columns to the sum of the $J - s$ last columns.

Proof. Conditional on both sets of marginal totals, from subsection 2.3.3, the expected cell counts for Table 4.6 are

$$\begin{aligned} E[M_{si}] &= N_{i.}\tau_s. \\ E[N_{i.} - M_{si}] &= N_{i.}(1 - \tau_s.) \quad i = 1, \dots, I, \quad s = 1, \dots, J-1. \end{aligned}$$

Table 4.6: The s -th collapsed $I \times 2$ contingency table

	$Y_{(1:s)}$	$Y_{(s+1:-J)}$	Total
A_1	M_{1s}	$N_{1\cdot} - M_{1s}$	$N_{1\cdot}$
\vdots	\vdots	\vdots	\vdots
A_i	M_{si}	$N_{i\cdot} - M_{si}$	$N_{i\cdot}$
\vdots	\vdots	\vdots	\vdots
A_I	M_{sI}	$N_{I\cdot} - M_{sI}$	$N_{I\cdot}$
Total	$M_{s\cdot}$	$n - M_{s\cdot}$	n

Then,

$$\begin{aligned}
X_s^2 &= \sum_{i=1}^I \left[\frac{(M_{si} - N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}\tau_{s\cdot}} + \frac{(N_{i\cdot} - M_{si} - N_{i\cdot}(1 - \tau_{s\cdot}))^2}{N_{i\cdot}(1 - \tau_{s\cdot})} \right] \\
&= \sum_{i=1}^I \left[\frac{(M_{si} - N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}\tau_{s\cdot}} + \frac{(N_{i\cdot} - M_{si} - N_{i\cdot} \cdot N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}(1 - \tau_{s\cdot})} \right] \\
&= \sum_{i=1}^I \left[\frac{(M_{si} - N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}\tau_{s\cdot}} + \frac{(M_{si} - N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}(1 - \tau_{s\cdot})} \right] \\
&= \sum_{i=1}^I \frac{(M_{si} - N_{i\cdot}\tau_{s\cdot})^2}{N_{i\cdot}\tau_{s\cdot}(1 - \tau_{s\cdot})} = \sum_{i=1}^I \frac{N_{i\cdot}(M_{si}/N_{i\cdot} - \tau_{s\cdot})^2}{\tau_{s\cdot}(1 - \tau_{s\cdot})}
\end{aligned}$$

□

This is why the Taguchi statistic is called the *cumulative chi-squared (CCS) statistic*.

The Taguchi statistic (4.15) can be generalised to give the class of CCS-type tests given by

$$T = \sum_{s=1}^{J-1} \sum_{i=1}^I w_s N_{i\cdot} (M_{si}/N_{i\cdot} - \tau_{s\cdot})^2 = \sum_{s=1}^{J-1} \sum_{i=1}^I w_s (M_{si} - N_{i\cdot}\tau_{s\cdot})^2 / N_{i\cdot} \quad (4.16)$$

for some weights w_1, \dots, w_{J-1} , $w_s > 0$. One such CCS-type test is the simple alternative to the Taguchi CCS statistic proposed by Nair[30]

$$T_C = \frac{1}{J} \sum_{s=1}^{J-1} \sum_{i=1}^I (M_{si}/N_{i\cdot} - \tau_{s\cdot})^2$$

which has constant weights $w_s = 1/J$, $s = 1, \dots, J - 1$.

4.3.1 General form of the decomposition of the cumulative chi-squared statistic

To study the properties of the CCS test, Nair[30] decomposes the CCS statistic (4.16) into orthogonal components by first providing it a matrix form. Let \mathbf{W} be the $(J - 1) \times (J - 1)$ diagonal matrix of the

weights:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{J-1} \end{bmatrix}.$$

Let \mathbf{A} be the $(J-1) \times J$ matrix of the cumulative column proportions given by

$$\mathbf{A} = \begin{bmatrix} 1 - \tau_1. & -\tau_1. & -\tau_1. & \cdots & -\tau_1. & -\tau_1. \\ 1 - \tau_2. & 1 - \tau_2. & -\tau_2. & \cdots & -\tau_2. & -\tau_2. \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 - \tau_{(J-1)}. & 1 - \tau_{(J-1)}. & 1 - \tau_{(J-1)}. & \cdots & 1 - \tau_{(J-1)}. & -\tau_{(J-1)}. \end{bmatrix}.$$

Then, the CCS statistic (4.16) can be written as

$$T = \sum_{i=1}^I \mathbf{N}'_i \mathbf{A}' \mathbf{W} \mathbf{A} \mathbf{N}_i / N_i.$$

Proof. $\mathbf{A} \mathbf{N}_i$ is the $(J-1) \times 1$ matrix whose elements are $\mathbf{A}_s \mathbf{N}_i$, where \mathbf{A}_s is the s -th row of \mathbf{A} , $s = 1, \dots, J-1$.

$$\begin{aligned} \mathbf{A}_s \mathbf{N}_i &= [1 - \tau_s. \quad \cdots \quad 1 - \tau_s. \quad -\tau_s. \quad \cdots \quad -\tau_s.] \begin{bmatrix} N_{i1} \\ \vdots \\ N_{is} \\ N_{i(s+1)} \\ \vdots \\ N_{iJ} \end{bmatrix} \\ &= (1 - \tau_s.) \sum_{j=1}^s N_{ij} - \tau_s. \sum_{j=s+1}^J N_{ij} \\ &= M_{si} - \tau_s. N_i. \end{aligned}$$

And so,

$$\mathbf{A} \mathbf{N}_i = \begin{bmatrix} M_{1i} - \tau_1. N_i. \\ \vdots \\ M_{(J-1)i} - \tau_{(J-1)}. N_i. \end{bmatrix}$$

and consequently,

$$\begin{aligned} \mathbf{N}'_i \mathbf{A}' \mathbf{W} &= [w_1 (M_{1i} - \tau_1. N_i.) \quad \cdots \quad w_{J-1} (M_{i(J-1)} - \tau_{(J-1)}. N_i.)] \\ \mathbf{N}'_i \mathbf{A}' \mathbf{W} \mathbf{A} \mathbf{N}_i / N_i &= \sum_{s=1}^{J-1} w_s (M_{si} - N_i. \tau_s.)^2 / N_i. \\ \sum_{i=1}^I \mathbf{N}'_i \mathbf{A}' \mathbf{W} \mathbf{A} \mathbf{N}_i / N_i &= \sum_{i=1}^I \sum_{s=1}^{J-1} w_s (M_{si} - N_i. \tau_s.)^2 / N_i. = T \end{aligned}$$

□

Now that we have the CCS statistic in matrix form, it can be decomposed through the matrix decomposition of the $J \times J$ symmetric matrix $\mathbf{A}'\mathbf{W}\mathbf{A}$. Let $\mathbf{\Lambda}$ be the $J \times J$ diagonal matrix of the column probabilities:

$$\mathbf{\Lambda} = \begin{bmatrix} \hat{\pi}_{.1} & 0 & \cdots & 0 \\ 0 & \hat{\pi}_{.2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_{.J} \end{bmatrix} \quad \text{where } \hat{\pi}_{.j} = \frac{N_{.j}}{n}, \quad j = 1, \dots, J.$$

Let $\mathbf{\Gamma}$ be the $(J-1) \times (J-1)$ diagonal matrix of the nonzero eigenvalues of $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{\Lambda}^7$ and $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_{J-1}]$ the $J \times (J-1)$ matrix of the associated eigenvectors such that

$$\bar{\mathbf{Q}}'\mathbf{\Lambda}\bar{\mathbf{Q}} = \mathbb{I}_J \quad (4.17)$$

where $\bar{\mathbf{Q}} = [\mathbf{1}_J | \mathbf{Q}]$, i.e. the eigenvectors $\mathbf{q}_1, \dots, \mathbf{q}_{J-1}$ are orthogonal with respect to the column probabilities. Then, $\mathbf{A}'\mathbf{W}\mathbf{A} = \mathbf{Q}\mathbf{\Gamma}\mathbf{Q}'$ and

$$T = \sum_{i=1}^I \mathbf{N}'_i \mathbf{Q}\mathbf{\Gamma}\mathbf{Q}' \mathbf{N}_i / N_i = \sum_{s=1}^{J-1} \gamma_s \mathbf{V}'_s \mathbf{V}_s.$$

where $\gamma_s = \mathbf{\Gamma}_{ss}$ is the s -th nonzero eigenvalue of $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{\Lambda}$, $s = 1, \dots, J-1$, and the i -th element of \mathbf{V}_s is

$$V_{si} = \frac{1}{\sqrt{N_i}} \mathbf{q}'_s \mathbf{N}_i, \quad s = 1, \dots, J-1, \quad i = 1, \dots, I.$$

4.3.2 The distribution of the cumulative chi-squared statistic

The random vector of the summands associated to the i -th row \mathbf{N}_i is

$$\mathbf{V}_{.i} = \frac{1}{\sqrt{N_i}} \mathbf{Q} \mathbf{N}_i, \quad i = 1, \dots, I.$$

Given the row and column probabilities, \mathbf{N}_i can be approximated by a multinomial distribution which has a limiting multinormal distribution as $n \rightarrow \infty$. Then, since $\mathbf{\Gamma}$ and \mathbf{Q} only depend on the weights and the column probabilities, \mathbf{V}_i also has a limiting multinormal distribution as $n \rightarrow \infty$. From subsection 2.3.3, the mean for V_{si} is

$$E[V_{si}] = \frac{1}{\sqrt{N_i}} \mathbf{q}'_s E[\mathbf{N}_i] = \sqrt{N_i} \mathbf{q}'_s (\hat{\pi}_{.1}, \dots, \hat{\pi}_{.J})' = \vec{0} \quad i = 1, \dots, I, \quad s = 1, \dots, J-1$$

since the columns of \mathbf{Q} are orthogonal with respect to the column probabilities by (4.17). Let $\hat{\mathbf{T}}$ be the ML estimator of the $J \times J$ matrix \mathbf{T} given by (2.6). Since

$$\mathbf{q}'_s \hat{\mathbf{T}} \mathbf{q}_b = \mathbf{q}'_s \mathbf{\Lambda} \mathbf{q}_b - \mathbf{q}'_s [\hat{\pi}_{.j} \hat{\pi}_{.t}] \mathbf{q}_b = \delta_{sb},$$

⁷There are exactly $J-1$ eigenvalues for $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{\Lambda}$ since \mathbf{W} and $\mathbf{\Lambda}$ are $(J-1) \times (J-1)$ diagonal matrices and $\text{rank}(\mathbf{A}) = J-1$ by the additive nature of the τ_s .

by the orthogonality of the columns of \mathbf{Q} with respect to the column probabilities, the covariance for V_{si} and V_{ba} is

$$\begin{aligned} \text{cov}[V_{si}, V_{ba}] &= \frac{1}{\sqrt{N_{i\cdot}N_{\cdot a}}} \mathbf{q}'_s \text{cov}[\mathbf{N}_{i\cdot}, \mathbf{N}_{\cdot a}] \mathbf{q}_b \\ &= \frac{1}{n-1} \begin{cases} (n - N_{i\cdot}) \mathbf{q}'_s \hat{\mathbf{T}} \mathbf{q}_b, & i = a, \\ -\sqrt{N_{i\cdot}N_{\cdot a}} \mathbf{q}'_s \hat{\mathbf{T}} \mathbf{q}_b, & i \neq a, \end{cases} \\ &= \frac{n}{n-1} \left[\mathbb{I}_I - \sqrt{\hat{\pi}_i \hat{\pi}_a} \right] \otimes \mathbb{I}_{J-1}. \end{aligned}$$

The V_{si} are asymptotically iid with a $N(0, 1)$ distribution as $n \rightarrow \infty$, $s = 1, \dots, J-1$, $i = 1, \dots, I-1$. Then, under the hypothesis of homogeneity and given the row and column probabilities, the components $\mathbf{V}'_s \mathbf{V}_s = \sum_{i=1}^I V_{si}^2$ are asymptotically iid with a $\chi^2(I-1)$ distribution.

Consequently, under the null hypothesis, the limiting distribution of the CCS statistic is the linear combination of chi-squared distributions:

$$T = \sum_{s=1}^{J-1} \gamma_s \left(\sum_{i=1}^I \mathbf{v}'_s \mathbf{v}_s \right) \xrightarrow[H_0]{d} \sum_{s=1}^{J-1} \gamma_s \chi_s^2(I-1)$$

where $\chi_s^2(I-1)$ is the chi-squared distribution for the s -th component, $s = 1, \dots, J-1$. By the Satterthwaite approximation[40], the asymptotic distribution of T can be approximated by $d\chi^2(v)$ where

$$d = \frac{1}{I-1} \frac{\sum_{s=1}^{J-1} \gamma_s^2}{\sum_{s=1}^{J-1} \gamma_s} \quad \text{and} \quad v = \frac{1}{d} \sum_{s=1}^{J-1} \gamma_s. \quad (4.18)$$

4.3.3 Link to smooth models

Suppose we have an $I \times J$ contingency table under the row-multinomial model with an ordinal column variable. We are interested in testing the hypothesis of homogeneity of the row distributions against increasing trend. Under the alternate hypothesis

$$H'_1 : \text{at least one } \pi_{ij} \text{ is different, for at least one } (i, j)$$

we set the alternative density as (4.1) with density $f_{XY}(x_i, y_j) = \pi_{ij}$ and null density $f_0(x_i, y_j) = \pi_{\cdot j}$, $i = 1, \dots, I$, $j = 1, \dots, J$, and $g_r \equiv 1$ since we want the cell probabilities to be a function of only the column probabilities. The k -order alternative is given by

$$\pi_{ij} = \left\{ 1 + \sum_{s=1}^k \theta_{si} h_s(y_j) \right\} \pi_{\cdot j}, \quad i = 1, \dots, I, \quad j = 1, \dots, J \quad (4.19)$$

where

- $k \leq J-1$ is chosen to be at most 4,
- $\{\theta_{si} : s = 1, \dots, k, i = 1, \dots, I\}$ are real valued parameters,
- $y_1 < \dots < y_J$ are the ordered scores for the columns, and
- h_1, \dots, h_k are orthonormal functions on the column probabilities $\pi_{\cdot 1}, \dots, \pi_{\cdot J}$.

Testing the hypothesis of homogeneity is now equivalent to

$$H''_0 : \boldsymbol{\theta} = \vec{0}_k \quad \text{against} \quad H''_1 : \theta_{si} \neq 0 \text{ for some } (s, i),$$

where $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1I}, \dots, \theta_{k1}, \dots, \theta_{kI})'$ and the column probabilities $\pi_{\cdot 1}, \dots, \pi_{\cdot (J-1)}$ are nuisance parameters. Then, for $k = J - 1$, and by Property 4.1, the $\hat{\mathbf{V}}_s$ are components for the Pearson and CCS statistics by

$$X_P^2 = \sum_{s=1}^{J-1} \hat{\mathbf{V}}_s' \hat{\mathbf{V}}_s \quad \text{and} \quad T = \sum_{s=1}^{J-1} \gamma_s \hat{\mathbf{V}}_s' \hat{\mathbf{V}}_s.$$

by taking $\mathbf{q}_s = \hat{\mathbf{h}}_s$ as the columns of \mathbf{Q} , $s = 1, \dots, J - 1$. By extension, the CCS statistic and its components have the same properties as the Pearson statistic and its components. In particular, in a similar way as for the Pearson statistic, it is still recommended to decompose a CCS statistic when the null hypothesis of homogeneous distributions is not rejected since the components have equal or better power than the CCS statistic against specified alternatives on the parameter space corresponding to the component.

4.3.4 Explicit solutions when the column are equiprobables

To compute the CCS statistic, we need to find the nonzero eigenvalues of $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{A}$ which form the diagonal of $\boldsymbol{\Gamma}$ and the corresponding eigenvectors which form the columns \mathbf{Q} that satisfy

$$\mathbf{A}'\mathbf{W}\mathbf{A} = \mathbf{Q}\boldsymbol{\Gamma}\mathbf{Q}' \quad \text{and} \quad \bar{\mathbf{Q}}'\boldsymbol{\Lambda}\bar{\mathbf{Q}} = \mathbb{I}_J. \quad (4.20)$$

Let γ be an eigenvalue in $\boldsymbol{\Gamma}$ and \mathbf{q} the corresponding eigenvector in \mathbf{Q} . Then the eigenproblem given by (4.20) is equivalent to

$$\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{q} = \gamma\mathbf{q}. \quad (4.21)$$

We note that

$$\begin{aligned} \mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{q} = \gamma\mathbf{q} &\Leftrightarrow \mathbf{q} = \gamma(\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{q} \\ &\Leftrightarrow (\mathbf{A}'\mathbf{W})^{-1}\mathbf{q} = \gamma(\mathbf{A}'\mathbf{W})^{-1}(\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{q} \\ &\Leftrightarrow (\mathbf{A}'\mathbf{W})^{-1}\mathbf{q} = \gamma\mathbf{W}^{-1}(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{A}'\mathbf{W})^{-1}\mathbf{q} \end{aligned}$$

We define $\mathbf{v} = (\mathbf{A}'\mathbf{W})^{-1}\mathbf{q}$. Then, the eigenproblem (4.21) is now equivalent to solving

$$\mathbf{v} = \gamma\mathbf{W}^{-1}(\mathbf{A}\mathbf{A}')^{-1}\mathbf{v}. \quad (4.22)$$

In general, this eigenproblem can be solved only numerically. However, when the column proportions are equal, the cumulative column proportions are $\tau_s = s/J$, $s = 1, \dots, J - 1$ and the matrix $(\mathbf{A}\mathbf{A}')^{-1}$ is always symmetric such that

$$(\mathbf{A}\mathbf{A}')^{-1} = \begin{cases} 2 & \text{on the diagonal,} \\ -1 & \text{off the diagonal,} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. See Appendix B. □

With $(\mathbf{A}\mathbf{A}')^{-1}$ symmetric, (4.22) can be solved algebraically and can be written as

$$\gamma \begin{bmatrix} \frac{2}{w_1} & -\frac{1}{w_1} & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{w_2} & \frac{2}{w_2} & -\frac{1}{w_2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{w_{J-2}} & \frac{2}{w_{J-2}} & -\frac{1}{w_{J-2}} \\ 0 & 0 & 0 & \cdots & 0 & -\frac{1}{w_{J-1}} & \frac{2}{w_{J-1}} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{J-2} \\ v_{J-1} \end{bmatrix}$$

where the s -th line is equivalent to

$$\begin{aligned} -\frac{\gamma}{w_s}(v_{s-1} - 2v_s + v_{s+1}) = v_s &\Leftrightarrow \gamma \nabla^2 v_{s+1} = -w_s v_s \\ &\Leftrightarrow \gamma \nabla^2 v_{s+1} + w_s v_s = 0 \end{aligned} \quad (4.23)$$

where ∇ is the backward difference operator defined by $\nabla v_s = v_s - v_{s-1}$ with $v_0 = v_J = 0$. Given the values of the weights $\{w_s\}$, we then solve the system of $J - 1$ equations given by (4.23) to obtain the columns of the matrix \mathbf{Q} by $\mathbf{q} = \mathbf{A}'\mathbf{W}\mathbf{v}$.

Under a product-multinomial model with equal column proportions, Nair[30] presents a partition of the CCS statistic to test homogeneity against increasing trend. He also gives explicit solutions for the Taguchi and Nair CCS statistics under the special case of equiprobable columns.

Under the row-multinomial model

Explicit solutions for the Taguchi CCS statistic Under the assumption of equiprobable columns, the weights of the Taguchi statistic are

$$w_s = \frac{1}{\tau_s \cdot (1 - \tau_s)} = \frac{J}{s(J - s)} \quad s = 1, \dots, J - 1$$

and the corresponding eigenvalues and eigenvectors for (4.22) have been obtained by Whittaker & Watson[48]:

$$\begin{aligned} \gamma_{E,s} &= \frac{J}{s(s+1)} \quad s = 1, \dots, J - 1 \\ \mathbf{v}_{E,s}(j) &\propto v(J - s) \nabla \mathbf{e}_s(j) \quad s = 1, \dots, J - 1 \quad j = 1, \dots, J \end{aligned}$$

where $\mathbf{e}_s = (e_s(1), \dots, e_s(J))'$ is the s -th degree Chebychev polynomial on the integers $\{1, \dots, J\}$. We obtain the column of \mathbf{Q} by $\mathbf{q} = \mathbf{A}'\mathbf{W}\mathbf{v}$ which, for the Taguchi CCS statistic under the row-multinomial model with equiprobable columns, are given by

$$\mathbf{q}_{E,s} = \frac{1}{\sqrt{J}} \frac{\mathbf{e}_s}{\|\mathbf{e}_s\|} \quad s = 1, \dots, J - 1.$$

Then, the Taguchi CCS statistic is

$$T_E = \sum_{s=1}^{J-1} \frac{J}{s(s+1)} \mathbf{V}'_{E,s} \cdot \mathbf{V}_{E,s}$$

where the i -th element of $\mathbf{V}_{E,s}$ is

$$V_{E,si} = \frac{1}{\sqrt{n_i}} \mathbf{q}'_{E,s} \mathbf{N}_i = \frac{1}{\sqrt{n_i}} \frac{1}{\sqrt{J}} \frac{1}{\|\mathbf{e}_s\|} \sum_{j=1}^J e_s(j) N_{ij} \quad s = 1, \dots, J - 1, \quad i = 1, \dots, I.$$

Since $\mathbf{q}_{E,1}$ is linear, $V_{E,1i}$ is equivalent to the Wilcoxon statistic $W_i = \sum_{j=1}^J j N_{ij}$ for the $2 \times J$ table given by Table 4.7 with equiprobable columns, $i = 1, \dots, I$.

Under the row-multinomial model, the Wilcoxon statistic is known to detect linear shifts in the means. Similarly, since $\mathbf{q}_{E,2}$ is quadratic, $V_{E,2i}$ is equivalent to the Mood statistic $M_i = \sum_{j=1}^J (j - \frac{J+1}{2})^2 N_{ij}$ for Table 4.7 which, assuming the means are equal, detects linear shifts in the variances, $i = 1, \dots, I$. Under the row-multinomial model with equiprobable columns, the Taguchi CCS statistic is decomposed into orthogonal components such that the s -th component detects linear shifts in the s -th moment, $s = 1, \dots, J - 1$.

Table 4.7: The $2 \times J$ collapsed table around the i -th row

	Y_1	\cdots	Y_j	\cdots	Y_J	Total
X_i	N_{i1}	\cdots	N_{ij}	\cdots	N_{iJ}	n_i
X'_i	$n_{\cdot 1} - N_{i1}$	\cdots	$n_{\cdot 1} - N_{ij}$	\cdots	$n_{\cdot 1} - N_{iJ}$	$n - n_i$
Total	$n_{\cdot 1}$	\cdots	$n_{\cdot j}$	\cdots	$n_{\cdot J}$	n

As a telescopic series, the sum of the nonzero eigenvalues is

$$\sum_{s=1}^{J-1} \gamma_{E,s} = \sum_{s=1}^{J-1} \frac{J}{s(s+1)} = J - 1$$

and the weight of the s -th nonzero eigenvalue in the sum is given by

$$\frac{\gamma_{E,s}}{\sum_{s=1}^{J-1} \gamma_{E,s}} = \frac{1}{s(s+1)} \frac{J}{J-1} \quad s = 1, \dots, J-1$$

which is a decreasing sequence. Since the Taguchi CCS statistic puts most of its weight on its first component, $\gamma_{E,1}/\sum \gamma_{E,s} = J/[2(J-1)]$, it behave like $\sum_{i=1}^I V_{E,i1}^2$ and is then powerful for tests for shifts in the means. On the other hand, as explained in subsection 4.3.3, the Pearson statistic, which can obtained by

$$X_P^2 = \sum_{s=1}^{J-1} \sum_{i=1}^I V_{E,is}^2$$

only has eigenvalue $\gamma_P = 1$ with multiplicity $J-1$ and thus puts an equal weight of $1/(J-1)$ to each component. Since

$$\frac{\gamma_{E,1}}{\sum_{s=1}^{J-1} \gamma_{E,s}} = \frac{1}{2} \frac{J}{J-1} \geq \frac{1}{J-1} = \frac{\gamma_{P,1}}{\sum_{s=1}^{J-1} \gamma_{P,s}}$$

the Taguchi CCS statistic is equally or more powerful than the Pearson statistic for detecting shifts in the means. However, since

$$\frac{\gamma_{E,2}}{\sum_{s=1}^{J-1} \gamma_{E,v}} = \frac{1}{6} \frac{J}{J-1} > \frac{1}{J-1} = \frac{\gamma_{P,2}}{\sum_{s=1}^{J-1} \gamma_{P,s}} \quad \text{only if } J \geq 7$$

the Taguchi CCS statistic may sometimes be less powerful than the Pearson statistic for detecting shifts in the variances.

Explicit solutions for the Nair CCS statistic Under the assumption of equiprobable columns, the weights for the Nair statistic are still constant at $w_s = 1/J$, $s = 1, \dots, J-1$. The corresponding eigenvalues and eigenvectors are

$$\gamma_{C,s} = \frac{1}{4J \sin^2(s\pi/2J)} \quad s = 1, \dots, J-1$$

$$\mathbf{v}_{C,s}(j) \propto \sin\left(\frac{s\pi j}{J}\right) \quad s = 1, \dots, J-1 \quad j = 1, \dots, J$$

We obtain the columns of \mathbf{Q} by $\mathbf{q} = \mathbf{A}'\mathbf{W}\mathbf{v}$ which, for the Nair CCS statistic under the row-multinomial model with equiprobable columns, are given by

$$q_{C,s}(j) = \sqrt{2J} \cos\left(\frac{s\pi(j-1)}{2J}\right).$$

Then, the Nair CCS statistic is

$$T_C = \sum_{s=1}^{J-1} \frac{1}{4J \sin^2(s\pi/2J)} \mathbf{V}'_{C,s} \mathbf{V}_{C,s}.$$

where the i -th element of $\mathbf{V}_{C,s}$ is

$$V_{C,si} = \frac{1}{\sqrt{n_i}} \mathbf{q}'_{C,s} \mathbf{N}_{i\cdot} = \sqrt{\frac{2J}{n_i}} \sum_{j=1}^J \cos\left(\frac{s\pi(j-1)}{2J}\right) N_{ij}$$

Under the row-multinomial model with equiprobable columns, the Nair CCS statistic is decomposed into components such that the s -th component can detect cosinusoidal deviances in the s -th moment.

Since $s\pi/2J \rightarrow 0$, $\gamma_{C,s} \rightarrow J/(s\pi)^2$ as $J \rightarrow \infty$ and the sum of the nonzero eigenvalues is

$$\sum_{s=1}^{J-1} \gamma_{C,s} \rightarrow \sum_{s=1}^{J-1} \frac{J}{(s\pi)^2} \approx \frac{J}{6} \quad J \rightarrow \infty$$

and the weight of the s -th nonzero eigenvalue in the sum is given by

$$\frac{\gamma_{C,s}}{\sum_{s=1}^{J-1} \gamma_{C,s}} \rightarrow \frac{1}{s^2} \frac{6}{\pi^2} \quad s = 1, \dots, J-1$$

which is a decreasing sequence. Since the Nair CCS statistic puts most of its weight on its first component, $\gamma_{C,1}/\sum \gamma_{C,s} = 6/\pi^2$, it behave like $\sum_{i=1}^I V_{E,1i}^2$ and is then powerful for tests for shifts in the means.

Since

$$\frac{\gamma_{C,1}}{\sum_{s=1}^{J-1} \gamma_{C,s}} \rightarrow \frac{6}{\pi^2} > \frac{1}{J-1} = \frac{\gamma_{P,1}}{\sum_{s=1}^{J-1} \gamma_{P,s}},$$

the Nair CCS statistic is more powerful than the Pearson statistic for detecting shifts in the means. However, since

$$\frac{\gamma_{C,2}}{\sum_{s=1}^{J-1} \gamma_{C,s}} \rightarrow \frac{1}{2} \frac{3}{\pi^2} > \frac{1}{J-1} = \frac{\gamma_{P,2}}{\sum_{s=1}^{J-1} \gamma_{P,s}} \quad \text{only if } J \geq 8$$

the Nair CCS statistic may sometimes be less powerful than the Pearson statistic for detecting shifts in the variances.

Under the column-binomial model

When we have only two rows, $I = 2$, the CCS statistic gets reduces to considering only the cumulative sum of the first row.

Proof. First, we note that for $I = 2$, the cumulative total $M_{s\cdot}$ can be written as

$$M_{s\cdot} = \tau_{s\cdot} N_1 + \tau_{s\cdot} N_2.$$

such that $M_{s1} - \tau_s \cdot N_{1\cdot} = -(M_{s2} - \tau_s \cdot N_{2\cdot})$. Then,

$$\begin{aligned} T &= \sum_{s=1}^{J-1} \sum_{i=1}^2 w_s \frac{(M_{si} - \tau_s \cdot N_{i\cdot})^2}{N_{i\cdot}} \\ &= \sum_{s=1}^{J-1} w_s \left\{ \frac{(M_{s1} - \tau_s \cdot N_{1\cdot})^2}{N_{1\cdot}} + \frac{(M_{s2} - \tau_s \cdot N_{2\cdot})^2}{N_{2\cdot}} \right\} \\ &= \left(\frac{1}{N_{1\cdot}} + \frac{1}{N_{2\cdot}} \right) \sum_{s=1}^{J-1} w_s (M_{s1} - \tau_s \cdot N_{1\cdot})^2 \\ &= \frac{n}{N_{1\cdot} N_{2\cdot}} \sum_{s=1}^{J-1} w_s (M_{s1} - \tau_s \cdot N_{1\cdot})^2 \end{aligned}$$

□

Then, for $I = 2$, the CCS statistic can be written as

$$T = \sum_{s=1}^{J-1} \gamma_s V_s^2$$

where

$$V_s = \frac{1}{\sqrt{N_{1\cdot}}} \mathbf{q}'_s \mathbf{N}_{1\cdot} \quad s = 1, \dots, J-1$$

which, under the null hypothesis and given the row and column proportions, are asymptotically iid with a $N(0, 1)$ distribution. Then, the components V_s^2 are asymptotically iid with a $\chi^2(1)$ distribution and

$$T = \sum_{s=1}^{J-1} \gamma_s V_s^2 \xrightarrow[H_0]{d} \sum_{s=1}^{J-1} \gamma_s \chi_s^2(1)$$

where $\chi_s^2(1)$ is the distribution for the s -th component. By the Satterthwaite approximation[40], the asymptotic distribution of T can be approximated by $d\chi^2(v)$ where

$$d = \frac{\sum_{s=1}^{J-1} \gamma_s^2}{\sum_{s=1}^{J-1} \gamma_s} \quad \text{and} \quad v = \frac{1}{d} \sum_{s=1}^{J-1} \gamma_s.$$

The computation of the CCS statistic is the same problem as presented in subsection 4.3.3. Again, when the column proportions are equal, explicit solutions have been found. In the column-binomial model, the equal column proportions imply that the J column totals are equal. In fact, the eigenvalues and eigenvectors obtained in section 4.3.4 are the same found under the column-binomial model with equal column totals. However, the interpretation of the components is different.

We consider the logistic model

$$\text{logit}(\pi_{1j}) = \theta_0 + \sum_{s=1}^k \theta_s h_s(y_j) \quad j = 1, \dots, J$$

which can be thought of as a k -order smooth alternative for $g(x|\boldsymbol{\theta}) = \text{logit}(\pi_{1j})$ instead of $g(x|\boldsymbol{\theta}) = \pi_{1j}$. Then as per usual, the Pearson statistic is the score statistic for

$$H_0'' : \boldsymbol{\theta} = \vec{0}_{J-1} \quad \text{against} \quad H_1'' : \theta_s \neq 0 \text{ for some } s$$

with components

$$V_s = \frac{1}{\sqrt{N_1}} \sum_{j=1}^J h_s(y_j) N_{1j} \quad s = 1, \dots, J-1$$

which are again powerful for testing on θ_s . If we choose \mathbf{h}_s to be columns of \mathbf{Q} and the column proportions are equal, then the components of the Pearson statistic are also the components of the CCS statistics as per subsection 4.3.3. In this case the components of T_E will be powerful for polynomial deviations of the s -th coefficient θ_s of the logit model and the components of T_C will be powerful for testing for cosinusoidal deviations of the s -th coefficient θ_s , $s = 1, \dots, J-1$.

Under the one-way ANOVA model

Suppose we have K observations per level of a one-factor experiment with I levels

$$X_{ik} = \mu + \alpha_i + \varepsilon_{ik} \quad i = 1, \dots, I \quad k = 1, \dots, K \quad (4.24)$$

where

- μ is the overall mean;
- $\alpha_1, \dots, \alpha_I$ are the treatment effects of a factor A ;
- $\varepsilon_{11}, \dots, \varepsilon_{IK}$ are the iid error terms with a $N(0, \sigma^2)$ distribution.

As it so happens, the Taguchi CCS statistic was originally developed for the ANOVA for industrial experiments modelled as (4.24) to test the hypothesis of homogeneity against monotonicity in the treatment effects. Without loss of generality, we consider the test of homogeneity against increasing trend:

$$H_0^I : \alpha_1 = \dots = \alpha_I \quad \text{against} \quad H_1^I : \alpha_1 \leq \dots \leq \alpha_I$$

with at least one inequality in the alternative. Suppose the outcome sample for the observations can be partitioned by $B = \{B_1, \dots, B_J\}$. Let N_{ij} denote the number of observations for the i -th level that fall into the j -th category B_j :

$$N_{ij} = \sum_{k=1}^K I[X_{ik} \in B_j] \quad i = 1, \dots, I \quad j = 1, \dots, J.$$

We obtain an $I \times J$ contingency table with row-multinomial model with equal row totals: $n_{i.} = K$, $i = 1, \dots, I$.

For the ANOVA model, Nair[29] gives that the sum of squares for the factor A is

$$SSA = n \sum_{s=1}^{J-1} \sum_{i=1}^I \frac{(M_{si} - \bar{M}_{s.})^2}{\bar{M}_{s.} (K - \bar{M}_{s.})}$$

where $n = IK$ is the total number of observations and $\bar{M}_{s.} = \sum_{i=1}^I M_{si}/I$ is the average cumulative frequency for the s -th column, $s = 1, \dots, J-1$.

$$\begin{aligned} SSA &= n \sum_{s=1}^{J-1} \sum_{i=1}^I \frac{(M_{si} - \bar{M}_{s.})^2}{\bar{M}_{s.} (K - \bar{M}_{s.})} \\ &= n \sum_{s=1}^{J-1} \sum_{i=1}^I \frac{(M_{si} - K\tau_{s.})^2}{K\tau_{s.} (K - K\tau_{s.})} \\ &= n \sum_{s=1}^{J-1} \sum_{i=1}^I \frac{(M_{si}/K - \tau_{s.})^2}{\tau_{s.} (1 - \tau_{s.})} \end{aligned}$$

which is the Taguchi CCS statistic with fixed and equal row totals. Thus, the decomposition of the Taguchi CCS statistic presented by Nair[30] is valid here.

For the model (4.24), Hirotsu[24] decomposes the statistic

$$T_H = K \|\mathbf{P}'\bar{\mathbf{X}}\|^2$$

where, for $\sigma^2 = 1$, $\mathbf{P}' = \mathbf{W}^{1/2}\mathbf{A}$ with \mathbf{W} and \mathbf{A} defined for the Taguchi weights and equally likely columns as given by Nair[30] (see subsection 4.3.1). Hirotsu obtains components defined on the Chebychev polynomials as done in section 4.3.4, with the polynomials given by a recursion formula.

Assuming $\sigma^2 = 1$ and $I = 1$, let N_j denote the number of observations that fall into the j -th category B_j :

$$N_j = \sum_{k=1}^K I[X_{1k} \in B_j] \quad j = 1, \dots, J.$$

For the $1 \times J$ contingency table of the N_j , Nair[28] presents a cumulative chi-squared statistic for the ANOVA model which is simply the Nair CCS statistic as provided earlier.

We can obtain the equivalent of the Nair CCS statistic for the model (4.24) by applying the same approach by Nair[28] to the model on the treatment means given by

$$\bar{X}_{i.} = \mu + \alpha_i + \bar{\varepsilon}_i. \quad i = 1, \dots, I$$

where $\bar{\varepsilon}_i. = \sum_{k=1}^K \varepsilon_{ik}/K$ are iid with a $N(0, \sigma^2)$ distribution.

An example of the decomposition

Example 4.5. (Example 3.2 and 4.3 cont'd.) Investigating the granular composition of quartz in Lithuanian and Saharan sand samples, the data on the lengths of the maximal axis of the quartz grains were used. In section 3.1 the Pearson test rejected the hypothesis that the length of the maximum axis has the same distribution in Lithuanian and Saharan sand. From the table of cumulative cell probabilities, we suspected that an appropriate follow-up test would compare the hypothesis of homogeneity against decreasing trend in the row distributions. In particular, we test the hypotheses

$$H'_0 : \tau_{s1} = \tau_{s2} \quad \text{vs.} \quad H'_1 : \tau_{s1} \geq \tau_{s2} \quad s = 1, \dots, 8.$$

In this situation, where the columns are not equiprobable, we use the general form of the decomposition for the Taguchi and Nair CCS statistics, as subsection 4.3.1. We denote the adjusted Taguchi CCS statistic by $T_E^* = T_E/d_E$ which has a $\chi^2(v_E)$, where d_E and v_E are given by (4.18) for the Satterthwaite approximation[40] for the distribution of T_E . We have,

$$d_E = 28.8466 \quad \text{and} \quad v_E = 2.7159.$$

Similarly, the adjusted Nair CCS statistic $T_C^* = T_C/0.6051$ and has an $\chi^2(2.3101)$ distribution.

The remainders are given by

$$\text{Remainder} = \frac{1}{d_R} \sum_{s=4}^7 \gamma_s \mathbf{V}'_s \cdot \mathbf{V}_s.$$

The Taguchi CCS statistic				The Nair CCS statistic			
Component	df	Value	P-value	Component	df	Value	P-value
\hat{V}'_1, \hat{V}_1	1	4.94	0.0262	\hat{V}'_1, \hat{V}_1	1	5.02	0.0251
\hat{V}'_2, \hat{V}_2	1	1.03	0.3109	\hat{V}'_2, \hat{V}_2	1	1.21	0.2714
\hat{V}'_3, \hat{V}_3	1	0.25	0.6168	\hat{V}'_3, \hat{V}_3	1	0.21	0.6477
Remainder	3.3465	0.60	0.9275	Remainder	3.7459	0.39	0.9771
T_E^*	2.7160	8.17	0.0337	T_C^*	2.3101	7.82	0.0275

where similarly as (4.18),

$$d_R = \frac{\sum_{s=4}^7 \gamma_s^2}{\sum_{s=1}^{J-1} \gamma_s} \quad \text{and} \quad v_R = \frac{1}{d_R} \sum_{s=4}^7 \gamma_s. \quad (4.25)$$

such that the distribution of the remainder is approximated by $\chi^2(v_R)$.

The adjusted Taguchi and Nair CCS statistics reject the hypothesis of homogeneity of the row distributions are valued at 8.17 and 7.82 with p -values of 0.0337 and 0.0275, respectively. The highly significant first component suggests there is an association between the means. With the mean maximal axis length of 22.55 and 30.29 cm for the Lithuanian and Saharan grains, respectively, we can say that the Lithuanian grains have significant smaller maximal axis length than the Saharan grains. We notice that the second component for the CCS component do not reject the hypothesis of homogeneity in the dispersions while the one for the Pearson statistic does. This is because the Taguchi and Nair CCS statistics' second components determine if the scale parameters differ from the null due to polynomial and cosinusoidal deviations, respectively, while the Pearson statistic's second component determine if the scale parameters differ from the null due to any type of deviation. From this, we conclude there is a difference in the scale parameters however it not due to polynomial or cosinusoidal deviation.

4.4 The Decomposition of the Chi-Squared Likelihood-Ratio Test Statistic

4.4.1 Log-linear models

Log-linear models for two-way contingency tables

Consider an $I \times J$ contingency table with n observations. Let X and Y be the row and column variables in the table. Then the expected cell counts μ_{ij} have the multiplicative form given by

$$\mu_{ij} = n\alpha_i^X \alpha_j^Y \beta_{ij}^{XY} \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

where

- $\{\alpha_i^X\}$ are the row effects such that $\sum_i \alpha_i^X = 1$;
- $\{\alpha_j^Y\}$ are the column effects such that $\sum_j \alpha_j^Y = 1$; and
- $\{\beta_{ij}^{XY}\}$ are the interaction effects such that $\sum_i \beta_{ij}^{XY} = \sum_j \beta_{ij}^{XY} = 1$.

Then, the $\ln \mu_{ij}$ have the following additive form

$$\ln \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad i = 1, \dots, I, \quad j = 1, \dots, J \quad (4.26)$$

where $\lambda = \ln n$, $\lambda_i^X = \ln \alpha_i^X$, $\lambda_j^Y = \ln \alpha_j^Y$ and $\lambda_{ij}^{XY} = \ln \beta_{ij}^{XY}$. (4.26) is the *saturated model* which perfectly describes any $\{\mu_{ij} > 0\}$ in terms of IJ parameters. In practice, unsaturated models are preferable, since their fit smooths the sample data and has simpler interpretations.

Log-linear model for three-way contingency tables

Consider an $I \times J \times K$ contingency table with n observations. The expected cell counts μ_{ijk} have the multiplicative form given by

$$\mu_{ijk} = n \alpha_i^X \alpha_j^Y \alpha_k^Z \beta_{ij}^{XY} \beta_{ik}^{XZ} \beta_{jk}^{YZ} \gamma_{ijk}^{XYZ} \quad \text{for all } i, j, k$$

where

- $\{\alpha_i^X\}$, $\{\alpha_j^Y\}$ and $\{\alpha_k^Z\}$ are the row, column and layer effects;
- $\{\beta_{ij}^{XY}\}$, $\{\beta_{ik}^{XZ}\}$ and $\{\beta_{jk}^{YZ}\}$ are the two-factor interaction effects; and
- $\{\gamma_{ijk}^{XYZ}\}$ are the three-way interaction effects.

Then the saturated model for $\ln \mu_{ijk}$ is

$$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \quad \text{for all } i, j, k$$

which has the constraints

$$\begin{aligned} \sum_{i=1}^I \lambda_i^X &= \sum_{j=1}^J \lambda_j^Y = \sum_{k=1}^K \lambda_k^Z = 0, \\ \sum_{i=1}^I \lambda_{ij}^{XY} &= \sum_{j=1}^J \lambda_{ij}^{XY} = 0, \quad \sum_{i=1}^I \lambda_{ik}^{XZ} = \sum_{k=1}^K \lambda_{ik}^{XZ} = 0, \quad \sum_{j=1}^J \lambda_{jk}^{YZ} = \sum_{k=1}^K \lambda_{jk}^{YZ} = 0, \\ \sum_{i=1}^I \lambda_{ijk}^{XYZ} &= \sum_{j=1}^J \lambda_{ijk}^{XYZ} = \sum_{k=1}^K \lambda_{ijk}^{XYZ} = 0. \end{aligned}$$

Once again, the number of parameters for the saturated model is given by the number of cells in the table, here equal to IJK .

4.4.2 Decomposition of the chi-squared likelihood-ratio test statistic

Suppose we want test the goodness-of-fit of a log-linear model, say M , in an $I \times J$ contingency table. This is equivalent to testing the hypotheses

$$H_0 : \mu_{ij} \text{ satisfy the model(M)} \quad \text{against} \quad H_1 : \mu_{ij} \text{ satisfy the saturated model}$$

To test the goodness-of-fit of the model(M), we can use the Pearson statistic

$$X_P^2(M) = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(N_{ij} - \hat{\mu}_{ij}^{(M)} \right)^2}{\hat{\mu}_{ij}^{(M)}}$$

or the chi-squared LRT statistic

$$G^2(M) = -2[L(\hat{\mu}_{ij}^{(M)}) - L(N_{ij})] = -2 \sum_{i=1}^I \sum_{j=1}^J N_{ij} \ln \left(\frac{\hat{\mu}_{ij}^{(M)}}{N_{ij}} \right) \quad (4.27)$$

where $\hat{\mu}_{ij}^{(M)}$ are the ML estimates of the expected cell counts under the log-linear model M^8 . Under the null hypothesis, $X_P^2(M)$ and $G^2(M)$ are asymptotically equivalent with a $\chi^2(v_S - v_M)$ distribution where $v_S = IJ$ is the number of independent parameters for the saturated model and v_M is the number of independent parameters for the model(M).

For nested⁹ log-linear models, the chi-squared LRT statistic can be partitioned conditionally as follows. Suppose we want to test

$$H_0 : \mu_{ij} \text{ satisfy the model(2)} \quad \text{against} \quad H_1 : \mu_{ij} \text{ satisfy the saturated model}$$

If model(2) is nested in a model(1), the chi-squared LRT statistic can be expressed as

$$\begin{aligned} G^2(2) &= -2[L(\hat{\mu}_{ij}^{(2)}) - L(N_{ij})] \\ &= -2[L(\hat{\mu}_{ij}^{(2)}) - L(\hat{\mu}_{ij}^{(1)})] - 2[L(\hat{\mu}_{ij}^{(1)}) - L(N_{ij})] \\ &= G^2(2|1) + G^2(1) \end{aligned}$$

where $G^2(2|1)$ is the chi-squared LRT statistic for testing

$$H_0 : \mu_{ij} \text{ satisfy the model(2)} \quad \text{against} \quad H_1 : \mu_{ij} \text{ satisfy the model(1)}$$

and $G^2(1)$ is the chi-squared LRT statistic for testing

$$H_0 : \mu_{ij} \text{ satisfy the model(1)} \quad \text{against} \quad H_1 : \mu_{ij} \text{ satisfy the saturated model.}$$

This is a partition of $G^2(2)$ into two parts:

1. $G^2(2|1)$, a measure of distance of the estimates from model(2) from those of model(1); and
2. $G^2(1)$, a measure of distance of the estimates from model(1) from the observations.

The conditional measure of model(2) given model(1), is given by

$$G^2(2|1) = -2[L(\hat{\mu}_{ij}^{(2)}) - L(\hat{\mu}_{ij}^{(1)})] = -2 \sum_{i=1}^I \sum_{j=1}^J \hat{\mu}_{ij}^{(1)} \ln \left(\frac{\hat{\mu}_{ij}^{(2)}}{\hat{\mu}_{ij}^{(1)}} \right)$$

which, under $H_0 : \mu_{ij}$ satisfy the model(1), asymptotically has a $\chi^2(v_1 - v_2)$ distribution where v_i is the number of independent parameters for model(i), $i = 1, 2$.

An advantage of this partition of the chi-squared LRT statistic is that allows us to determine if the model under the null hypothesis, here model(2), is being rejected since another model, model(1), is a better fit for the data. Another advantage of the partition is that if model(1) is itself nested in another model, $G^2(1)$ can be similarly partitioned into two parts.

In particular, we consider only hierarchical models. That is, if a model includes an interaction term, it includes all the low-term included in the interaction. For example, if a model contains λ_{ij}^{XY} , it also contains λ_i^X and λ_j^Y . "A reason for including lower-order terms is that, otherwise, the statistical significance and the interpretation of a higher-order term depends on how variables are coded (Agresti [1])".

For an $I \times J$ contingency table, we have the following four types of hierarchical log-linear models:

⁸The ML estimates of the expected values under the saturated model are the cell counts N_{ij} .

⁹For two statistical models, a model(2) is *nested* in model(1) if model(2) contains only a subset of the terms in model(1).

Table 4.8: Hierarchical log-linear models for an $I \times J$ contingency table

Log-linear Model	Interpretation	Symbol
$\ln \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$	The saturated model	(XY)
$\ln \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$	Independence of X and Y	(X, Y)
$\ln \mu_{ij} = \lambda + \lambda_i^X$	Homogeneity of the row distributions	(X)
$\ln \mu_{ij} = \lambda$	Equiprobability of the cells	(*)

4.4.3 Extension of the decomposition

The hierarchical log-linear models for an $I \times J \times K$ contingency table are given by Table 4.9. Once again, the Pearson and chi-squared LRT statistics can be used to test the goodness-of-fit of any model on the expected cell counts, including the three-factor hierarchical log-linear models Table 4.9. The partition of the chi-squared LRT statistic in subsection 4.4.2 can once again be used here. The usual use of this partition (presented by Bishop et al[13]) is for comparing nested hierarchical log-linear models that differ by a single λ -term. However, it is a special case of the partition for hierarchical log-linear models presented by Goodman[23] where a hypothesis H on an m -way contingency table can be partitioned into two parts:

- a corresponding hypothesis H' in a marginal table; and
- a hypothesis H'' about independence, conditional independence or conditional equiprobability in the m -way table.

In which case,

$$G^2(H) = G^2(H') + G^2(H''),$$

where H'' can itself be partitioned using the same method described by Goodman. This more general form of the partition has the advantage that while H is a special case of H'' , the difference between the corresponding hierarchical log-linear models may be in terms of more than one λ -term.

Table 4.9: Hierarchical log-linear models for an $I \times J \times K$ contingency table

Log-linear Model	Interpretation	Symbol
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$	The saturated model	(XYZ)
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$	The no-three factor interaction model	(XY,XY,XZ)
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$	Conditional independence of X and Y given Z	(XZ, YZ)
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{YZ}$	X independent of Y and Z	(X, YZ)
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	Mutual independence of X , Y and Z	(X, Y, Z)
$\ln \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y$	Homogeneity of the $I \times J$ partial tables	(X, Y)
$\ln \mu_{ijk} = \lambda + \lambda_i^X$	Homogeneity of the row-distributions	(X)
$\ln \mu_{ijk} = \lambda$	Equiprobability of cells	(*)

Chapter 5

Comparisons of the CCS Statistics and a Correlation Based Statistic

Suppose we have an $I \times J$ contingency table. We are interested in testing the dependence relationship between the row and column variables. The advantages of the decomposition of the Pearson and CCS statistics, as presented in Chapter 4, is that through the components, we are able to identify the sources of deviations when testing against a directional alternative on the distributions. However, if we are interested in studying the dependence relationship through only one parameter, the Pearson and CCS statistics and their decompositions are not the best options:

- As omnibus tests, the Pearson and CCS statistics are not sensitive to deviations on only one parameter, unless the deviation is extreme. “The statistics are designed to detect any pattern on $[(I - 1)(J - 1)]$ parameters. In achieving this generality, they sacrifice sensitivity for detecting particular patterns.” Agresti[1].
- For a test on one parameter, the appropriate component does have better power than the Pearson and CCS statistics¹. However, the process of obtaining the components may not be computationally convenient.
 - There are no issues for the computation of a component which is equivalent to a known statistic (e.g. the first component of the decompositions of the Pearson statistic, shown in Chapter 4). Otherwise, appropriate statistical software is not necessarily implemented to compute the other components.
 - By computing the Pearson/CCS statistics then the appropriate component, we are performing multiple tests. “The rejection threshold [i.e. the significance level] (α) is valid for one test; it corresponds to the error rate [...] of each test. If two tests are carried out at the threshold 0.05, the error rate for each test remains at 0.05 but the family-wise error rate (FWER), the probability of wrongly rejecting the null hypothesis in at least one test, is much higher as it equals to $1 - (1 - 0.05)^2$ (or 0.0975).” Bestgen[9].

For studying the dependence relationship in an $I \times J$ contingency table through only one parameter, we suggest that alternate and simpler chi-squared statistics perform better than the Pearson and CCS statistics and their corresponding components.

- A chi-squared test based on a statistic designed to detect patterns on one parameter has one degree of

¹In the appropriate articles, simulations were done to show this to be true.

freedom and has power advantage on the Pearson and CCS tests. “Since df [degrees of freedom] equal the mean of the chi-squared distribution, a relatively large value with $df = 1$ falls farther out in its right-hand tail than a comparable value of [the Pearson and CCS statistics] with $df = (I - 1)(J - 1)$ [...]” Agresti[1].

- A statistic which has simple form is easier to program if the statistical software is not readily available.
- A chi-squared statistic with only one degree of freedom cannot be decomposed and the FWER is always equal to α .

In this chapter, we show that for studying the dependence relationship in an $I \times J$ contingency table through a hypothesis on one parameter rather than $(I - 1)(J - 1)$, a simpler chi-squared statistic will have better power than the Pearson and CCS statistics. In particular, for a $2 \times J$ contingency table under the column-binomial model, we compare the significance levels and powers of the Pearson statistic, the Taguchi and Nair CCS statistics to the statistic developed by Alvo & Berthelot[2, 6] to test the hypothesis of homogeneity in the success probabilities.

5.1 The Alvo & Berthelot Spearman Test

Suppose we have $2 \times J$ contingency table under the column-binomial model,

	y_1	y_2	\dots	y_J	Total
Successes	N_{11}	N_{12}	\dots	N_{1J}	$N_{1\cdot}$
Failures	N_{21}	N_{22}	\dots	N_{2J}	$N_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	\dots	$n_{\cdot J}$	n

where the column variable is ordinal with increasing scores: $y_1 < \dots < y_J$. For the j -th column, the cell counts are obtained by

$$N_{1j} = \sum_{k=1}^{n_{\cdot j}} Y_{jk} \quad N_{2j} = n_{\cdot j} - N_{1j} \quad j = 1, \dots, J$$

where the $\{Y_{jk}\}$ are iid with a *Bernoulli*(π_{1j}) distribution, $j = 1, \dots, J$. Then, $N_{1j} \sim \text{binomial}(n_{\cdot j}, \pi_{1j})$, $j = 1, \dots, J$. We want to test the hypothesis of equal success probabilities against increasing trend:

$$H_0 : \pi_{11} = \dots = \pi_{1J} \quad \text{against} \quad H_1 : \pi_{11} \leq \dots \leq \pi_{1J}$$

with at least one inequality in the alternative.

Example 5.1. Suppose we have to following 2×2 contingency table under a column-binomial model

	y_1	y_2	Total
Successes	2	1	3
Failures	1	1	2
Total	3	2	5

Let ϖ and ν be the ranking of the Bernoulli observations $\{Y_{jk}\}$ with respect to the columns and rows, respectively, allowing for ties. The compatibility class corresponding to ϖ contains the 12 permutations obtained by permuting ranks 1,2,3 among themselves and ranks 4,5 among themselves

$$C(\varpi) = \{(123|45), (132|45), (213|45), \dots, (312|54), (321|54)\}.$$

On the other hand, there are a total of 6 patterns with (2 successes, 1 failure) for the first column and (1 success, 1 failure) for the second column

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

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

This concept of compatibility to tied rankings is generalised by Alvo & Berthelot[2] where they develop the following statistic, based on the Spearman similarity,

$$S = \sum_{j=1}^J \left(c_j - \frac{n+1}{2} \right) N_{1j} = \sum_{j=1}^J c_j^* N_{1j}$$

where c_j is the average of the compatible ranks for the j -th column:

$$c_j = \sum_{b=1}^{j-1} n_{.b} + \frac{n_{.j} + 1}{2} \quad j = 1, \dots, J.$$

and c_j^* is the centralised average of the compatible ranks for the j -th column:

$$c_j^* = c_j - \frac{n+1}{2}, \quad j = 1, \dots, J.$$

As $N_{1.}, N_{2.} \rightarrow \infty$, S asymptotically has a normal distribution with mean and variance given by

$$\mu_S = \sum_{j=1}^J n_{.j} c_j \pi_{1j} \quad \text{and} \quad \sigma_S^2 = \sum_{j=1}^J n_{.j} c_j^{*2} \pi_{1j} (1 - \pi_{1j}).$$

Under the null hypothesis, $\mu_S = 0$ and hence

$$S_* = \frac{S}{\sigma_S} = \frac{\sum_{j=1}^J c_j^* N_{1j}}{\sqrt{\sum_{j=1}^J n_{.j} c_j^{*2} \pi_{1j} (1 - \pi_{1j})}} \xrightarrow{H_0} N(0, 1)$$

where we reject the null hypothesis for large values of S_* .

5.2 The Comparisons

For the comparisons, we consider 2×5 contingency tables under the column-binomial model with equal column totals, $n_{.j} = n/5$, $j = 1, \dots, 5$. To obtain the cell counts, we generate random observations from the *binomial*($\pi_{1j}, n/5$) distributions, $j = 1, \dots, 5$.

When the column totals are too small and/or the success probabilities are too small, it is possible to obtain zeros in the first row of the randomly generated contingency table. In which case, $\hat{\sigma}_S^2 = 0$ and S_* cannot be computed. Thus, we consider the adjustment

$$\tilde{\pi}_{1j} = \frac{N_{1j} + 1}{n_{.j} + 2} \quad j = 1, \dots, 5$$

when $n_{.j} \hat{\pi}_{1j} < 5$ for at least one j and compute \tilde{S}_* based on $\tilde{\sigma}_S^2 = \sum_{j=1}^5 n_{.j} c_j^{*2} \tilde{\pi}_{1j} (1 - \tilde{\pi}_{1j})$.

5.2.1 The distributions under the null hypothesis

The distribution of the Spearman statistic has been confirmed to be normal under the null hypothesis by Berthelot[6].

For equal column totals in an $I \times J$ contingency table, the $J \times J$ matrix $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{A}$ depends only on the number of columns and the choice of the weights defining \mathbf{W} . The matrix \mathbf{A} is proportional to the identity matrix

$$\mathbf{A} = \begin{bmatrix} \hat{\pi}_{.1} & 0 & \cdots & 0 \\ 0 & \hat{\pi}_{.2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_{.J} \end{bmatrix} = \frac{1}{J}\mathbb{I}_J$$

since $\hat{\pi}_{.j} = (n/J)/n = 1/J$, $j = 1, \dots, J$. Consequently, the cumulative column probabilities are $\tau_{.s} = s/J$ and the matrix \mathbf{A} is given by

$$\mathbf{A} = \frac{1}{J} \begin{bmatrix} J-1 & -1 & -1 & \cdots & -1 & -1 \\ J-2 & J-2 & -2 & \cdots & -2 & -2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & -(J-1) \end{bmatrix}.$$

Then, for equal column totals, the distribution of the CCS statistic now only depends on I , J and the matrix \mathbf{W} . Therefore, for a 2×5 contingency table with equal column totals,

$$d_E = \frac{\sum_{s=1}^4 \gamma_{E,s}^2}{\sum_{s=1}^4 \gamma_{E,s}} = 1.7951 \quad \text{and} \quad v_E = \frac{1}{d_E} \sum_{s=1}^4 \gamma_{E,s} = 2.2282,$$

where d_E and v_E are given by (4.18) in subsection 4.3.2, and

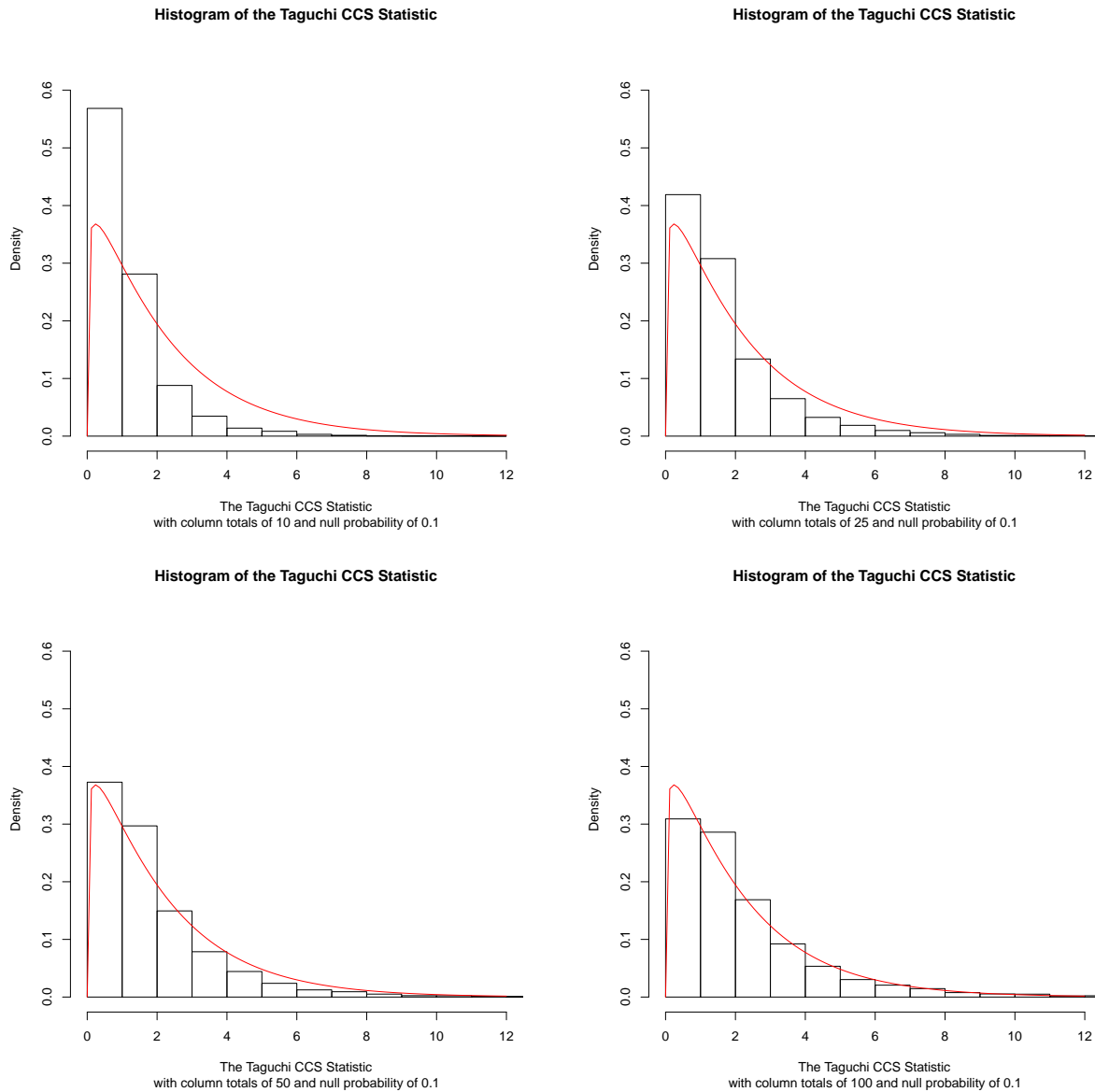
$$d_C = \frac{\sum_{s=1}^4 \gamma_{C,s}^2}{\sum_{s=1}^4 \gamma_{C,s}} = 0.0760 \quad \text{and} \quad v_C = \frac{1}{d_C} \sum_{s=1}^4 \gamma_{C,s} = 2.1052.$$

Thus, the distributions for the Taguchi and Nair CCS statistics for a 2×5 contingency table with equal column totals are

$$T_E \sim 1.7951\chi^2(2.2282) \quad \text{and} \quad T_C \sim 0.0760\chi^2(2.1052)$$

under the null hypothesis of homogeneity.

The simulations are done with R. After calculating the Taguchi CCS statistic for different values of $n_{.j}$ (10, 25, 50 and 100) and π_0 (0.1, 0.2, 0.3, 0.4 and 0.5), we can see from histograms that the values of $T_E/1.7951$, under the null hypothesis $H_0 : \pi_{11} = \cdots = \pi_{15} = 0.1$, seem to be well approximated by the $\chi^2(2.2282)$ distribution. Similar results were obtained for the other null probabilities.

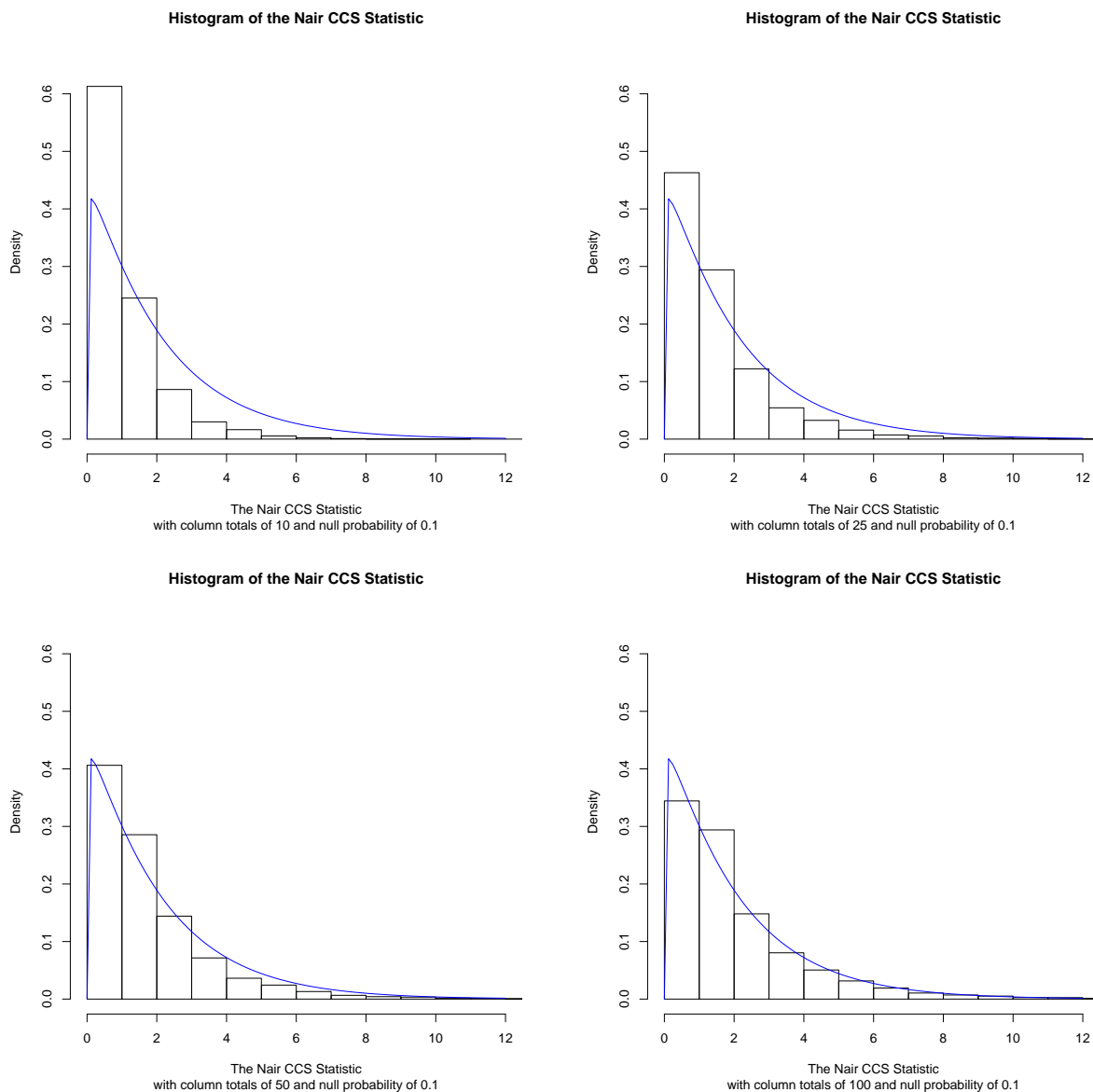


Similarly, after calculating the Nair CCS statistic for different values of $n_{.j}$ and π_0 , we can see from the histograms that the values of $T_C/0.0760$, under the null hypothesis $H_0 : \pi_{11} = \dots = \pi_{15} = 0.1$ seem to be well approximated by the $\chi^2(2.1052)$ distribution. Again, similar results were obtained for the other null probabilities.

5.2.2 Comparison of the significance levels

Definition 5.1. *The significance level of a test is the probability of rejecting the null hypothesis when it is true.*

We check that the Taguchi CCS, Nair CCS and Spearman statistics reach the nominal significance level $\alpha = 0.05$. In order to estimate significance levels, we compute the proportion of K simulations that reject



the null hypothesis. A lower bound for K is z_α^2/E where z_α is the upper $100(1 - \alpha)$ -th percentage point for the $N(0,1)$ distribution and E is the margin of error. The Table 5.1 gives the minimum number of simulations to compute the proportion for three significance levels and three margins of error. We run 10,000 simulations to compare the significance levels at $\alpha = 0.05$ with a margin of error of at most 0.01.

The Spearman statistic has been confirmed to reach the nominal level of $\alpha = 0.05$ by Berthelot[6]. The Taguchi and Nair CCS statistics have been confirmed to reach the nominal level of $\alpha = 0.10$ for a logistic location model by Nair[30]. However there are no details to the comparison and so, we do our own simulation and comparisons.

The simulated significance levels for the Taguchi and Nair CCS statistics are given in Table 5.2. We note that for both the Taguchi and Nair CCS statistics, the significance levels are very small for small n_j and small π_0 but do increase as the n_j and π_0 increase, though, not always actually reaching the nominal

level. However, since the 95% confidence interval for $\alpha = 0.05$ at 10,000 simulations, is (0.0487, 0.0514), we are confident the significance levels are not significantly different from 0.05.

Table 5.3 allows us to compare the significance levels of the Pearson statistic, the Taguchi and Nair CCS statistics and the Spearman statistic. Once again, all the statistics have small significance levels for small $n_{.j}$ and small π_0 and increase as $n_{.j}$ and π_0 increase and although the significance levels do not all reach 0.05, we are confident they are not significantly different from 0.05 since they are included in the above 95% confidence interval. However, their rates of increase are significantly different. In particular:

- for small $n_{.j}$ and small π_0 , the significance levels for the Pearson statistic are significantly smaller than those for the Taguchi CCS statistic;
- for all $n_{.j}$ and all π_0 , the significance levels for the Taguchi and Nair CCS statistics are not significantly different;
- for small $n_{.j}$ and small π_0 , the significance levels for the Taguchi CCS statistic (and the Nair CCS statistic) are significantly smaller than those for the Spearman statistic.

Table 5.1: Minimum number of simulations

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$E = 0.10$	42	68	136
$E = 0.05$	165	271	542
$E = 0.01$	4106	6764	13530

Table 5.2: Significance levels for the Taguchi and Nair CCS statistics

The Taguchi CCS statistic					The Nair CCS statistic				
π_0	$n_{.j} = 10$	$n_{.j} = 25$	$n_{.j} = 50$	$n_{.j} = 100$	π_0	$n_{.j} = 10$	$n_{.j} = 25$	$n_{.j} = 50$	$n_{.j} = 100$
0.10	0.0044	0.0160	0.0261	0.0514	0.10	0.0049	0.0151	0.0290	0.0525
0.20	0.0142	0.0274	0.0442	0.0509	0.20	0.0147	0.0283	0.0436	0.0514
0.30	0.0224	0.0493	0.0516	0.0477	0.30	0.0239	0.0490	0.0528	0.0470
0.40	0.0238	0.0487	0.0481	0.0483	0.40	0.0246	0.0456	0.0482	0.0494
0.50	0.0280	0.0499	0.0459	0.0476	0.50	0.0287	0.0494	0.0472	0.0471

Table 5.3: Significance levels for the Pearson and Spearman statistics

The Pearson statistic					The Spearman statistic				
π_0	$n_{.j} = 10$	$n_{.j} = 25$	$n_{.j} = 50$	$n_{.j} = 100$	π_0	$n_{.j} = 10$	$n_{.j} = 25$	$n_{.j} = 50$	$n_{.j} = 100$
0.10	0.0017	0.0088	0.0191	0.0491	0.10	0.0346	0.0363	0.0420	0.0527
0.20	0.0091	0.0234	0.0472	0.0531	0.20	0.0284	0.0461	0.0524	0.0531
0.30	0.0139	0.0538	0.0491	0.0490	0.30	0.0313	0.0593	0.0516	0.0508
0.40	0.0187	0.0468	0.0502	0.0496	0.40	0.0388	0.0522	0.0464	0.0502
0.50	0.0201	0.0495	0.0472	0.0514	0.50	0.0284	0.0423	0.0501	0.0469

5.2.3 Comparison of the powers

Definition 5.2. *The power of a test is the probability of rejecting the null hypothesis when it is not true.*

For the comparison of the powers, we consider three types of alternative hypotheses:

- the success probabilities are strictly increasing,
- the success probabilities are non-decreasing with consecutive repeated values, and
- the success probabilities are non-decreasing and non-increasing.

For each case, we consider 2×5 contingency tables with $n_{.j} = 10, 25$ and 100 , $j = 1, \dots, 5$.

Strictly increasing success probabilities The simulated powers for the Pearson statistic, the Taguchi and Nair CCS statistics and the Spearman statistic are given in Table 5.4. We note that there are significant differences between the powers. First, the Pearson statistic always has the lowest power since the Taguchi and Nair CCS statistics put 37.5% and 60.79% of the weight on their first component which detects deviations on the means, which here is equivalent to detecting deviations of the success probabilities since the column totals are equal, while the Pearson statistic only puts 25% of its weight on its first component². Secondly, for all column totals, the Taguchi and Nair CCS statistics are not significantly different. To finish, the Spearman statistic has the highest power because with $df = 1$ the large values of the statistic are more likely to fall further out in its right tail.

Non-decreasing success probabilities with consecutive repeated values The simulated powers for the Pearson statistic, the Taguchi and Nair CCS statistics and the Spearman statistic are given in Table 5.5. Similarly as to when we had strictly increasing success probabilities, for small column totals, the Pearson statistic has the smallest power and the Spearman statistic has the highest power. However, the results are mixed for the Taguchi and Nair CCS statistics. In some cases, they are not significantly different, in others the Taguchi statistic has slightly better power.

Non-increasing and non-decreasing success probabilities The simulated powers for the Pearson statistic, the Taguchi and Nair CCS statistics and the Spearman statistic are given in Table 5.6. We note that the order of the powers has reversed. The Spearman statistic now has the lowest power. The Taguchi and Nair CCS statistics have the advantage on the Spearman statistic since they are also detecting deviations for higher moments. The Nair CCS statistic does better for the alternative hypothesis where the success probabilities were obtained from a cosinusoidal and a sinusoidal function (such as lines 10 and 16) while the Taguchi CCS statistic does better for the alternative hypotheses where the success probabilities were obtained from polynomial functions (such as lines 6 and 23). The Pearson statistic has the highest power. This is due to the fact that like the Taguchi and Nair CCS statistics, the Pearson statistic detects deviations for the higher moments, but also by section 4.3.4, since $J = 5 < 8$, the Pearson statistic has more weight on its second, third and fourth components than the Taguchi and Nair statistic do and thus is better at detecting deviations for the corresponding moments.

From the power simulations, we confirm that for testing the hypothesis of homogeneity of the column distributions with respect to only one parameter, here the success probabilities, a simple statistic, such as the Spearman statistic developed by Alvo & Berthelot[2, 6], has better power than the Pearson and CCS statistics against alternatives of monotonic trend. However, as omnibus tests, the Pearson and CCS

²See section 4.3.4.

statistics have the advantage when the alternatives are non-monotonic since they also detect deviations on other parameters/moments.

The simulations also confirm that, as we stated in section 4.3.4, as the Taguchi and Nair CCS statistics lean heavily onto their first component, they are better detectors of deviation in location than the Pearson statistic, as we saw in the comparisons with strictly increasing and, non-decreasing success probabilities with repeated values. However, here the Taguchi and Nair CCS statistics do not have as much power the Pearson statistic which equally detects deviations on dispersion, skewness, etc. as it does for location.

Table 5.4: Comparison of the powers – strictly increasing success probabilities

					Pearson			Taguchi			Nair			Spearman		
π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	10	25	100	10	25	100	10	25	100	10	25	100
0.05	0.10	0.15	0.20	0.25	0.0265	0.2208	0.9634	0.1027	0.4775	0.9945	0.1047	0.4839	0.9946	0.3255	0.7022	0.9987
0.05	0.20	0.30	0.40	0.50	0.2543	0.9058	1	0.5669	0.9814	1	0.5647	0.9794	1	0.8122	0.9973	1
0.10	0.20	0.30	0.40	0.50	0.1946	0.7696	1	0.4624	0.9299	1	0.4641	0.9296	1	0.7275	0.9777	1
0.10	0.20	0.40	0.50	0.60	0.4007	0.9613	1	0.7135	0.9934	1	0.7227	0.9937	1	0.9023	0.9989	1
0.10	0.40	0.65	0.85	0.95	0.9803	1	1	0.9986	1	1	0.9986	1	1	1	1	1
0.15	0.20	0.25	0.30	0.35	0.0352	0.1972	0.5098	0.1088	0.3773	0.7177	0.1113	0.3773	0.7313	0.2789	0.5675	0.8312
0.15	0.35	0.45	0.55	0.75	0.4745	0.9759	1	0.7616	0.9966	1	0.7527	0.9959	1	0.9323	0.9997	1
0.20	0.25	0.30	0.35	0.40	0.0404	0.1835	0.8065	0.1075	0.3503	0.9295	0.1111	0.3548	0.9285	0.2168	0.5089	0.9698
0.20	0.35	0.50	0.65	0.80	0.5491	0.9860	1	0.8238	0.9986	1	0.8273	0.9983	1	0.9602	1	1
0.20	0.50	0.60	0.70	0.80	0.5239	0.9819	1	0.7769	0.9968	1	0.7682	0.9956	1	0.9342	0.9997	1
0.30	0.45	0.55	0.65	0.70	0.1973	0.7364	0.9749	0.4211	0.8997	0.9966	0.4240	0.8984	0.9964	0.6272	0.9570	0.9990
0.30	0.45	0.60	0.75	0.90	0.5670	0.9956	1	0.8373	0.9996	1	0.8377	0.9995	1	0.9613	1	1
0.35	0.45	0.60	0.75	0.95	0.5842	0.9975	1	0.8619	0.9999	1	0.8568	0.9999	1	0.9670	1	1
0.35	0.55	0.70	0.80	0.85	0.3916	0.9958	1	0.6639	0.9917	1	0.6701	0.9915	1	0.8724	0.9978	1
0.35	0.70	0.75	0.80	0.90	0.4460	0.9752	1	0.6727	0.9931	1	0.6485	0.9907	1	0.8539	0.9986	1
0.40	0.55	0.60	0.65	0.70	0.0917	0.4247	0.9784	0.2102	0.6109	0.9947	0.2107	0.5999	0.9938	0.3851	0.7476	0.9985
0.40	0.55	0.65	0.75	0.80	0.2091	0.7786	1	0.4397	0.9106	1	0.4450	0.9095	1	0.6328	0.9683	1
0.45	0.50	0.60	0.75	0.95	0.3871	0.9841	1	0.7011	0.9981	1	0.6971	0.9977	1	0.8895	0.9993	1
0.45	0.60	0.70	0.75	0.80	0.1449	0.6447	0.9991	0.3223	0.8239	0.9999	0.3227	0.8189	0.9999	0.5034	0.9144	1
0.50	0.60	0.70	0.80	0.90	0.2033	0.8259	1	0.4583	0.9466	1	0.4589	0.9446	1	0.7279	0.9842	1
0.50	0.70	0.75	0.85	0.90	0.2010	0.8339	1	0.4433	0.9379	1	0.4360	0.9352	1	0.6882	0.9776	1
0.55	0.60	0.70	0.80	0.90	0.2830	0.7410	1	0.5031	0.8995	1	0.5088	0.9006	1	0.7318	0.9644	1
0.55	0.70	0.75	0.85	0.90	0.2836	0.7207	0.9999	0.4999	0.8775	1	0.4962	0.8764	1	0.6843	0.9475	1
0.60	0.65	0.70	0.75	0.80	0.0995	0.2366	0.8063	0.1833	0.3964	0.9307	0.1812	0.3977	0.9321	0.2927	0.5596	0.9680
0.60	0.70	0.75	0.85	0.90	0.2163	0.5913	0.9987	0.4054	0.8033	1	0.4065	0.8045	1	0.6007	0.9007	1

Table 5.5: Comparison of the powers – increasing success probabilities with repeated values

					Pearson			Taguchi			Nair			Spearman		
π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	10	25	100	10	25	100	10	25	100	10	25	100
0.05	0.10	0.10	0.20	0.30	0.0540	0.4230	0.9972	0.1838	0.6807	0.9997	0.1816	0.6797	0.9997	0.4332	0.8394	0.9999
0.05	0.55	0.85	0.95	0.95	0.9998	1	1	1	1	1	1	1	1	1	1	1
0.10	0.10	0.10	0.10	0.80	0.9283	1	1	0.9456	1	1	0.9253	1	1	0.9754	1	1
0.10	0.20	0.40	0.40	0.60	0.3519	0.9312	1	0.6300	0.9859	1	0.6285	0.9847	1	0.8551	0.9974	1
0.10	0.30	0.30	0.40	0.50	0.1519	0.7034	0.9744	0.3613	0.8695	0.9949	0.3508	0.8527	0.9927	0.6395	0.9503	0.9992
0.10	0.30	0.30	0.50	0.50	0.2180	0.8214	1	0.4430	0.9230	1	0.4447	0.9204	1	0.7230	0.9757	1
0.15	0.20	0.20	0.30	0.40	0.0652	0.3332	0.9715	0.1736	0.5275	0.9939	0.1733	0.5219	0.9934	0.3704	0.6997	0.9971
0.20	0.20	0.30	0.40	0.50	0.1232	0.5234	0.9981	0.2978	0.7425	0.9996	0.3061	0.7477	0.9997	0.4648	0.8528	0.9998
0.20	0.30	0.30	0.40	0.50	0.0919	0.3967	0.9882	0.2163	0.6127	0.9971	0.2153	0.6031	0.9965	0.3826	0.7670	0.9994
0.20	0.30	0.30	0.50	0.60	0.2195	0.7506	1	0.4350	0.8959	1	0.4447	0.8962	1	0.6323	0.9567	1
0.30	0.40	0.40	0.40	0.40	0.0269	0.0899	0.2771	0.0371	0.1009	0.2996	0.0370	0.0959	0.2725	0.0834	0.1583	0.3890
0.30	0.40	0.50	0.50	0.70	0.1611	0.6622	0.9998	0.3477	0.8213	0.9999	0.3413	0.8091	0.9999	0.5595	0.9095	1
0.30	0.50	0.50	0.50	0.90	0.4658	0.9871	1	0.6435	0.9939	1	0.6009	0.9890	1	0.8641	0.9978	1
0.35	0.50	0.60	0.60	0.80	0.2100	0.7915	0.9999	0.4241	0.9101	1	0.4170	0.9006	1	0.6369	0.9685	1
0.35	0.60	0.60	0.60	0.70	0.1278	0.5359	0.9953	0.2196	0.6455	0.9985	0.2033	0.6072	0.9971	0.3880	0.7470	0.9993
0.40	0.50	0.60	0.60	0.80	0.1613	0.6854	0.9999	0.3390	0.8473	1	0.3342	0.8349	0.9999	0.5489	0.9365	1
0.40	0.50	0.60	0.80	0.80	0.2659	0.8698	0.9965	0.5240	0.9560	0.9991	0.5427	0.9614	0.9995	0.7171	0.9852	0.9997
0.45	0.50	0.50	0.70	0.75	0.1290	0.5738	0.9979	0.2654	0.7331	0.9994	0.2770	0.7434	0.9995	0.4594	0.8366	0.9997
0.50	0.50	0.85	0.85	0.85	0.3493	0.9465	1	0.5415	0.9678	1	0.5684	0.9730	1	0.7518	0.9796	1
0.50	0.60	0.70	0.80	0.80	0.1249	0.5935	0.9988	0.2948	0.7795	1	0.3047	0.7876	1	0.4627	0.8817	1
0.55	0.60	0.60	0.70	0.90	0.2429	0.6983	0.9999	0.3991	0.8269	1	0.3878	0.8061	1	0.6374	0.8900	1
0.60	0.60	0.70	0.80	0.90	0.2334	0.6416	0.9493	0.4220	0.8285	0.9887	0.4281	0.8315	0.9885	0.6458	0.9247	0.9957
0.60	0.70	0.70	0.80	0.90	0.1870	0.5235	0.9972	0.3406	0.7313	0.9997	0.3351	0.7209	0.9997	0.5419	0.8618	1
0.60	0.70	0.80	0.80	0.90	0.2001	0.5469	0.9968	0.3592	0.7446	0.9999	0.3508	0.7378	0.9999	0.5487	0.8613	1
0.70	0.70	0.85	0.85	0.85	0.1223	0.3174	0.9192	0.1975	0.4219	0.9497	0.2034	0.4389	0.9585	0.3129	0.5451	0.9648

Table 5.6: Comparison of the powers – non-increasing and non-decreasing success probabilities

					Pearson			Taguchi			Nair			Spearman		
π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	10	25	100	10	25	100	10	25	100	10	25	100
0.05	0.40	0.20	0.50	0.90	0.9609	1	1	0.9905	1	1	0.9849	1	1	0.9999	1	1
0.05	0.30	0.20	0.10	0.45	0.2200	0.8535	1	0.1896	0.7061	1	0.1505	0.6039	1	0.4490	0.7652	0.9994
0.10	0.90	0.20	0.60	0.50	0.9948	1	1	0.2087	0.9897	1	0.1616	0.9896	1	0.2660	0.5693	0.9869
0.10	0.80	0.50	0.80	0.10	0.9700	1	1	0.6036	0.9998	1	0.3252	0.9984	1	0.0427	0.0442	0.0542
0.20	0.50	0.30	0.80	0.40	0.5177	0.9796	1	0.2582	0.8327	1	0.2656	0.8220	1	0.3625	0.7598	0.9996
0.20	0.90	0.50	0.90	0.20	0.9716	1	1	0.5042	0.9994	1	0.2630	0.9958	1	0.0466	0.0412	0.0426
0.30	0.90	0.70	0.50	0.60	0.4745	0.9883	1	0.1143	0.7559	1	0.0697	0.6195	1	0.0753	0.1588	0.3816
0.30	0.70	0.90	0.40	0.10	0.9345	1	1	0.7672	0.9999	1	0.7234	0.9998	1	0.0001	0.0000	0.0000
0.35	0.90	0.50	0.80	0.70	0.4944	0.9872	1	0.2331	0.7782	1	0.1893	0.7064	1	0.2972	0.6928	0.9949
0.35	0.65	0.35	0.05	0.35	0.4346	0.9815	1	0.2101	0.7590	0.9967	0.2689	0.8158	0.9976	0.0010	0.0001	0.0000
0.40	0.20	0.20	0.10	0.20	0.0789	0.4190	0.9934	0.1497	0.4919	0.9923	0.1440	0.4696	0.9894	0.0021	0.0004	0.0000
0.40	0.60	0.30	0.70	0.50	0.1940	0.7234	0.9998	0.0597	0.2109	0.8828	0.0640	0.2220	0.9069	0.1275	0.2550	0.6097
0.50	0.20	0.50	0.50	0.20	0.2167	0.7548	1	0.0616	0.2875	0.9946	0.0450	0.1969	0.9867	0.0052	0.0040	0.0001
0.50	0.40	0.10	0.80	0.30	0.6921	0.9985	1	0.0492	0.4176	1	0.0536	0.4512	1	0.0352	0.0478	0.0495
0.55	0.90	0.60	0.80	0.70	0.2811	0.7425	0.9999	0.1062	0.2253	0.9262	0.0891	0.1740	0.8742	0.1131	0.1823	0.4021
0.60	0.20	0.40	0.80	0.50	0.4872	0.9738	1	0.1316	0.6198	1	0.1634	0.6624	1	0.1731	0.3700	0.8479
0.60	0.60	0.95	0.60	0.60	0.2948	0.8763	1	0.1023	0.2294	0.9963	0.1008	0.2362	0.9988	0.0423	0.0525	0.0545
0.60	0.45	0.40	0.45	0.60	0.0619	0.2766	0.8718	0.0368	0.1433	0.6848	0.0334	0.1147	0.5793	0.0401	0.0532	0.0446
0.60	0.90	0.55	0.70	0.90	0.4035	0.8915	1	0.1571	0.4867	0.9996	0.1281	0.4084	0.9988	0.2744	0.4599	0.9398
0.70	0.50	0.40	0.80	0.60	0.1957	0.7592	1	0.0423	0.2161	0.9655	0.0481	0.2299	0.9705	0.0608	0.0949	0.1658
0.70	0.50	0.50	0.70	0.50	0.0879	0.3878	0.9669	0.0409	0.1405	0.6475	0.0366	0.1078	0.5347	0.0172	0.0118	0.0010
0.80	0.30	0.30	0.50	0.30	0.4520	0.9723	1	0.3718	0.9328	1	0.2987	0.8874	1	0.0007	0.0000	0.0000
0.90	0.30	0.10	0.30	0.90	0.9903	1	1	0.9130	1	1	0.8215	1	1	0.0300	0.0395	0.0528
0.90	0.50	0.50	0.10	0.70	0.8814	1	1	0.6432	0.9991	1	0.5804	0.9977	1	0.0000	0.0000	0.0000
0.95	0.20	0.40	0.65	0.80	0.8805	1	1	0.3786	0.9980	1	0.2647	0.9895	1	0.0744	0.1540	0.3777

Chapter 6

Conclusion

This thesis presents different chi-squared statistics and their decompositions for hypothesis tests in $I \times J$ contingency tables. Since it is omnibus, the Pearson test has moderate power for all alternatives on $(I - 1)(J - 1)$ parameters and thus does not indicate the strength or type of relationship in an $I \times J$ contingency table. This is why we explore the decomposition of the Pearson statistic from which were derived chi-squared statistics that have better power for directional alternatives on smaller sets of parameters than the Pearson statistic.

This thesis has two main contributions. First, we review and gather different methods of decomposing some chi-squared statistics. Iversen[25] describes a partition of the Pearson and likelihood ratio test statistics by the decomposition of the $(I - 1)(J - 1)$ contingency table. Rayner & Best develop a general form of decomposition of the Pearson statistic by orthogonal polynomials, defined on distribution moments, derived from smooth models, which lead to the extension of the well-known statistics when the decomposition is applied to different models on the contingency table. For the chi-squared test of independence, it leads to extension of Pearson product-moment and Spearman correlations. For the chi-squared test of homogeneity, it leads to various other extensions (Yates, Kruskal-Wallis and Friedman statistics) depending on whether the column scores are ranks and/or the marginal total are fixed. We also explore of the class of CCS statistics, which unlike the Pearson statistic are used to test homogeneity against monotonic trend. Their decomposition is done by writing the CCS statistic as a matrix multiplication and find eigendecomposition of the central matrix. In particular, we look at the decomposition of two examples of CCS statistics when the columns are equiprobable.

The second contribution of this thesis is the comparison of the Pearson and CCS statistics to the Spearman statistics developed by Alvo & Berthelot[2]. The choice of orthogonal functions for the decomposition of PCS and the weights for CCS gives more flexibility for the choice of alternatives when we want to test distribution/population(s) in the table on a large number of parameters. However, if we know we are interested in testing the distribution/populations(s) on a single parameter, the decomposition and the omnibus tests have significantly less power than chi-squared statistic with one degree of freedom. The results show that the Spearman statistic has better power when the success probabilities are monotonically increasing. However, when the probabilities are not monotonic, the omnibus tests have better power since they are detecting deviations other than those in the means/success probabilities.

Some further research interest would start with developing other decompositions of the Pearson or the CCS statistics to obtain components that detect other types of deviations. In particular, we would look at different orthogonal functions for the Pearson statistic and different weights for the class of CCS statistics. We note that with natural scores, the decomposition by the orthogonal polynomials defined on distributional moments always leads to the Spearman metric (which compares the scores to the average score). It would be

interesting to investigate the Pearson statistic when the data in a contingency table are arranged by number of pairwise inversion between the row and column rankings and see if the decomposition by orthogonal polynomials possibly leads instead to the Kendall metric. On the more applied side, we could consider the software implementation for the orthogonal decomposition of chi-squared statistics. We would look into the R-package ‘orthopolynom’ which can construct sets of orthogonal polynomials and their recurrence relations and further explore *Smooth Tests of Goodness of Fit: Using R* from Rayner et al[39]. This would lead to easier application to real data sets.

Appendix A

Obtaining the first two orthogonal polynomials from the Emerson recursion

We consider the polynomials $\{g_r(x)\}$ given by

$$g_r(x) = a_{r0} + a_{r1}x + \cdots + a_{rr}x^r \quad r = 0, 1, \dots, I-1 \quad (\text{A.1})$$

where the a_{rj} are constants, $j = 0, 1, \dots, r$. Let X be a categorical variable with scores x_1, \dots, x_I and probabilities π_1, \dots, π_I , $\pi_i > 0$. A recursion developed by Emerson[22] can be used to find the constants a_{rj} for the polynomials given by (A.1), such that the set $\{g_r(x)\}$ is orthonormal on the probabilities π_1, \dots, π_I , that is

$$g_0(x) = 1, \quad \sum_{i=1}^I g_r(x_i) g_s(x_i) \pi_i = \delta_{rs}, \quad r, s = 0, 1, \dots, I-1. \quad (\text{A.2})$$

The r -th polynomial g_r is defined recursively in terms of the previous g_{r-1} and g_{r-2} by

$$g_r(x) = (A_r x + B_r) g_{r-1}(x) - C_r g_{r-2}(x), \quad r = 1, 2, \dots, I-1 \quad (\text{A.3})$$

where constants A_r , B_r and C_r are defined by

$$A_r = \left\{ \sum_{i=1}^I \pi_i x_i^2 g_{r-1}^2(x_i) - \left[\sum_{i=1}^I \pi_i x_i g_{r-1}^2(x_i) \right]^2 - \left[\sum_{i=1}^I \pi_i x_i g_{r-1}(x_i) g_{r-2}(x_i) \right]^2 \right\}^{-1/2} \quad (\text{A.4})$$

$$B_r = -A_r \sum_{i=1}^I \pi_i x_i g_{r-1}^2(x_i) \quad (\text{A.5})$$

$$C_r = A_r \sum_{i=1}^I \pi_i x_i g_{r-1}(x_i) g_{r-2}(x_i) \quad (\text{A.6})$$

with $g_{-1}(x) = 0$ and $g_0(x) = 1$.

By (A.3), the first polynomial $g_1(x)$ is defined by

$$g_1(x) = (A_1 x + B_1) g_0(x) - C_1 g_{-1}(x) = A_1 x + B_1$$

Let us find A_1 and B_1 . From (A.4),

$$\begin{aligned} A_1 &= \left\{ \sum \pi x^2 g_0^2(x) - \left[\sum \pi x g_0^2(x) \right]^2 - \left[\sum \pi x g_0(x) g_{-1}(x) \right]^2 \right\}^{-1/2} \\ &= \left\{ \sum \pi x^2 - \left[\sum \pi x \right]^2 \right\}^{-1/2} \\ A_1 &= \frac{1}{\sqrt{\mu_2}} \end{aligned}$$

Consequently, by (A.5),

$$B_1 = -A_1 \sum \pi x g_0^2(x) = -\frac{\mu}{\sqrt{\mu_2}}$$

and the first polynomial $g_1(x)$ is

$$g_1(x) = A_1 x + B_1 = \frac{x - \mu}{\sqrt{\mu_2}} \quad (\text{A.7})$$

By (3.9), the second polynomial $g_2(x)$ is defined by

$$\begin{aligned} g_2(x) &= (A_2 x + B_2) g_1(x) - C_2 g_0(x) \\ &= (A_2 x + B_2) (A_1 x + B_1) - C_2 \\ g_2(x) &= A_1 A_2 x^2 + (A_1 B_2 + A_2 B_1) x + B_1 B_2 - C_2 \end{aligned}$$

Let us find A_2 and B_2 . From (A.4),

$$\begin{aligned} A_2 &= \left\{ \sum \pi x^2 g_1^2(x) - \left[\sum \pi x g_1^2(x) \right]^2 - \left[\sum \pi x g_0(x) g_1(x) \right]^2 \right\}^{-1/2} \\ &= \mu_2 \left\{ \mu_2 \sum \pi x^2 (x - \mu)^2 - \left[\sum \pi x (x - \mu)^2 \right]^2 - \left[\sum \pi x (x - \mu) \right]^2 \right\}^{-1/2} \end{aligned}$$

The explicit formula for the second, third (and its square) and fourth central moments are

$$\mu_2 = \left(\sum \pi x^2 \right) - \mu^2,$$

$$\mu_3 = \left(\sum \pi x^3 \right) - 3\mu \left(\sum \pi x^2 \right) + 2\mu^3,$$

$$\begin{aligned} \mu_3^2 &= \left(\sum \pi x^3 \right)^2 + 9\mu^2 \left(\sum \pi x^2 \right)^2 + 4\mu^6 \\ &\quad - 6\mu \left(\sum \pi x^3 \right) \left(\sum \pi x^2 \right) + 4\mu^3 \left(\sum \pi x^3 \right) - 12\mu^4 \left(\sum \pi x^2 \right), \end{aligned}$$

$$\mu_4 = \left(\sum \pi x^4 \right) - 4\mu \left(\sum \pi x^3 \right) + 6\mu^2 \left(\sum \pi x^2 \right) - 3\mu^4.$$

Let

$$A_{21} = \sum \pi x^2 (x - \mu)^2, \quad A_{22} = \sum \pi x (x - \mu)^2, \quad A_{23} = \sum \pi x (x - \mu) \text{ and } A = \mu_2 A_{21} - A_{22}^2 - A_{23}^2$$

such that $A_2 = \mu_2/\sqrt{A}$. Then,

$$\begin{aligned} A_{21} &= \left(\sum \pi x^4\right) - 2\mu \left(\sum \pi x^3\right) + \mu^2 \left(\sum \pi x^2\right) \\ &= \mu_4 + 2\mu \left(\sum \pi x^3\right) - 5\mu^2 \left(\sum \pi x^2\right) + 3\mu^4 \end{aligned} \quad (\text{A.8})$$

$$\begin{aligned} A_{22}^2 &= \left[\sum \pi x(x - \mu)^2\right]^2 \\ &= \left[\left(\sum \pi x^3\right) - 2\mu \left(\sum \pi x^2\right) + \mu^3\right]^2 \\ &= \left(\sum \pi x^3\right)^2 + 2\mu^3 \left(\sum \pi x^3\right) + 4\mu^2 \left(\sum \pi x^2\right)^2 - 4\mu^4 \left(\sum \pi x^2\right) \\ &\quad - 4\mu \left(\sum \pi x^3\right) \left(\sum \pi x^2\right) + \mu^6 \end{aligned} \quad (\text{A.9})$$

$$A_{23}^2 = \left[\sum \pi x(x - \mu)\right]^2 = \left[\left(\sum \pi x^2\right) - \mu^2\right]^2 = \mu_2^2 \quad (\text{A.10})$$

and

$$\begin{aligned} A &= \mu_2 \left[\mu_4 + 2\mu \left(\sum \pi x^3\right) - 5\mu^2 \left(\sum \pi x^2\right) + 3\mu^4\right] \\ &\quad - \left(\sum \pi x^3\right)^2 - 2\mu^3 \left(\sum \pi x^3\right) - 4\mu^2 \left(\sum \pi x^2\right)^2 + 4\mu^4 \left(\sum \pi x^2\right) \\ &\quad + 4\mu \left(\sum \pi x^3\right) \left(\sum \pi x^2\right) - \mu^6 - \mu_2^3 \\ &= \mu_2 \mu_4 + 2\mu \left(\sum \pi x^3\right) \left(\sum \pi x^2\right) - 5\mu^2 \left(\sum \pi x^2\right)^2 + 3\mu^4 \left(\sum \pi x^2\right) \\ &\quad - 2\mu^3 \left(\sum \pi x^3\right) + 5\mu^4 \left(\sum \pi x^2\right) - 3\mu^6 - \left(\sum \pi x^3\right)^2 \\ &\quad - 2\mu^3 \left(\sum \pi x^3\right) - 4\mu^2 \left(\sum \pi x^2\right)^2 + 4\mu^4 \left(\sum \pi x^2\right) + 4\mu \left(\sum \pi x^3\right) \left(\sum \pi x^2\right) \\ &\quad - \mu^6 - \mu_2^3 \\ &= \mu_2 \mu_4 - \left(\sum \pi x^3\right)^2 - 4\mu^3 \left(\sum \pi x^3\right) + 6\mu \left(\sum \pi x^3\right) \left(\sum \pi x^2\right) \\ &\quad - 9\mu^2 \left(\sum \pi x^2\right)^2 + 12\mu^4 \left(\sum \pi x^2\right) - 4\mu^6 - \mu_2^3 \\ A &= \mu_2 \mu_4 - \mu_3^2 - \mu_2^3. \end{aligned} \quad (\text{A.11})$$

Then,

$$A_2 = \frac{\mu_2}{\sqrt{\mu_2 \mu_4 - \mu_3^2 - \mu_2^3}}. \quad (\text{A.12})$$

Consequently, by (A.5) and (A.6)

$$\begin{aligned}
B_2 &= -A_2 \sum \pi x g_1^2(x) \\
&= -\frac{a}{\sqrt{\mu_2}} \left[\left(\sum \pi x^3 \right) - 2\mu \left(\sum \pi x^2 \right) + \mu^3 \right] \\
&= -\frac{a}{\sqrt{\mu_2}} \left[\mu_3 + \mu \left(\sum \pi x^2 \right) - \mu^3 \right] \\
&= -\frac{a}{\sqrt{\mu_2}} (\mu\mu_2 + \mu_3),
\end{aligned} \tag{A.13}$$

$$\begin{aligned}
C_2 &= A_2 \sum \pi x g_1(x) g_0(x) \\
&= a \left[\left(\sum \pi x^2 \right) - \mu^2 \right] \\
&= a\mu_2
\end{aligned} \tag{A.14}$$

where

$$a = \frac{A_2}{\sqrt{\mu_2}} = \sqrt{\frac{\mu_2}{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}}, \tag{A.15}$$

and the second polynomial $g_2(x)$ is

$$\begin{aligned}
g_2(x) &= A_1 A_2 x^2 + (A_1 B_2 + A_2 B_1)x + B_1 B_2 - C_2 \\
&= \frac{1}{\sqrt{\mu_2}} \frac{\mu_2}{\sqrt{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}} x^2 - \left(\frac{1}{\sqrt{\mu_2}} \frac{\mu\mu_2 + \mu_3}{\sqrt{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}} + \frac{\mu_2}{\sqrt{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}} \frac{\mu}{\sqrt{\mu_2}} \right) x \\
&\quad + \frac{\mu}{\sqrt{\mu_2}} \frac{\mu\mu_2 + \mu_3}{\sqrt{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}} - \mu_2 \sqrt{\frac{\mu_2}{\mu_2\mu_4 - \mu_3^2 - \mu_2^3}} \\
&= ax^2 - a \frac{2\mu\mu_2 + \mu_3}{\mu_2} x + a \left(\frac{\mu^2\mu_2 + \mu\mu_3 - \mu_2^2}{\mu_2} \right) \\
&= a \left[(x^2 - 2\mu x + \mu^2) - \frac{\mu_3}{\mu_2} x + \frac{\mu\mu_3}{\mu_2} - \mu_2 \right] \\
g_2(x) &= a \left[(x - \mu)^2 - \frac{\mu_3}{\mu_2} (x - \mu) - \mu_2 \right].
\end{aligned} \tag{A.16}$$

Let us show that $\{g_0, g_1, g_2\}$ is an orthonormal set of polynomials.

Proof. The polynomials are orthogonal:

$$\sum \pi g_0(x) g_1(x) = \sum \pi (x - \mu) / \mu_2^{1/2} = 0$$

$$\begin{aligned}
\sum \pi g_0(x) g_2(x) &= \sum a\pi(x - \mu)^2 - \sum a\mu_3(x - \mu) / \mu_2 - a\mu_2 \\
&= a\mu_2 - a\mu_2 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\sum \pi g_1(x) g_2(x) &= \sum a\pi(x - \mu)^3 / \mu_2^{1/2} - \sum a\mu_3(x - \mu)^2 / \mu_2 - \sum a\pi(x - \mu) \mu_2 / \mu_2^{1/2} \\
&= a\mu_3 / \mu_2^{1/2} - a\mu_3 / \mu_2^{1/2} \\
&= 0
\end{aligned}$$

and they are normalised:

$$\sum \pi g_0^2(x) = \sum \pi = 1$$

$$\sum \pi g_1^2(x) = \mu_2^{-1} \sum \pi (x - \mu)^2 = \mu_2^{-1} \mu_2 = 1$$

$$\begin{aligned} \sum \pi g_2^2(x) &= \frac{\mu_2}{\mu_2 \mu_4 - \mu_3^2 - \mu_2^3} \sum \pi \left[(x - \mu)^2 - \mu_3(x - \mu)/\mu_2 - \mu_2 \right]^2 \\ &= \frac{1}{\mu_2 (\mu_2 \mu_4 - \mu_3^2 - \mu_2^3)} \sum \pi \left[\mu_2(x - \mu)^2 - \mu_3(x - \mu) - \mu_2^2 \right]^2 \\ &= \frac{1}{\mu_2 (\mu_2 \mu_4 - \mu_3^2 - \mu_2^3)} \left[\mu_2^2 \mu_4 \cdot \mu_2 \mu_3^2 \cdot \mu_2^4 - 2\mu_2 \mu_3^2 - 2\mu_2^4 \right] \\ &= \frac{1}{\mu_2 (\mu_2 \mu_4 - \mu_3^2 - \mu_2^3)} \mu_2 (\mu_2 \mu_4 - \mu_3^2 - \mu_2^3) \\ \sum \pi g_2^2(x) &= 1 \end{aligned}$$

□

Appendix B

Proof of symmetry of $(\mathbf{A}\mathbf{A}')^{-1}$ when the columns are equiprobable

Let $\tau_{s\cdot}$ denote the s -th cumulative column proportion:

$$\tau_{s\cdot} = \begin{cases} \pi_{\cdot 1} + \cdots + \pi_{\cdot j} & s = 1, \dots, J-1, \\ 1 & s = J \end{cases}$$

and let \mathbf{A} be the $(J-1) \times J$ matrix of the cumulative column proportions given by

$$\mathbf{A} = \begin{bmatrix} 1 - \tau_{1\cdot} & -\tau_{1\cdot} & -\tau_{1\cdot} & \cdots & -\tau_{1\cdot} & -\tau_{1\cdot} \\ 1 - \tau_{2\cdot} & 1 - \tau_{2\cdot} & -\tau_{2\cdot} & \cdots & -\tau_{2\cdot} & -\tau_{2\cdot} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 - \tau_{(J-1)\cdot} & 1 - \tau_{(J-1)\cdot} & 1 - \tau_{(J-1)\cdot} & \cdots & 1 - \tau_{(J-1)\cdot} & -\tau_{(J-1)\cdot} \end{bmatrix}.$$

We show that if $\tau_{s\cdot} = s/J$, $i = 1, \dots, J-1$, $(\mathbf{A}\mathbf{A}')^{-1}$ is always symmetric such that

$$(\mathbf{A}\mathbf{A}')^{-1} = \begin{cases} 2 & \text{on the diagonal} \\ -1 & \text{off the diagonal} \\ 0 & \text{otherwise} \end{cases}$$

For example, for $J = 3$,

$$(\mathbf{A}\mathbf{A}')^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Proof. We want to find the value of the (s, t) -th cell in $\mathbf{A}\mathbf{A}'$, denoted by $\mathbf{A}\mathbf{A}'_{st}$. Let $\mathbf{A}_{s\cdot}$ be the s -th row of \mathbf{A} and let $\mathbf{A}'_{\cdot t}$ be the t -th column of \mathbf{A}' . Then $(\mathbf{A}\mathbf{A}')_{st} = \mathbf{A}_{s\cdot} \mathbf{A}'_{\cdot t}$.

Suppose $s = t$. Then

$$\begin{aligned} (\mathbf{A}\mathbf{A}')_{ii} &= i(1 - \tau_{i\cdot})^2 + (J - i)\tau_{i\cdot}^2 \\ &= i \frac{(J - i)^2}{J^2} + (J - i) \frac{i^2}{J^2} \\ &= \frac{i(J - i)}{J^2} (J - i + i) \\ (\mathbf{A}\mathbf{A}')_{ii} &= \frac{i(J - i)}{J} \end{aligned}$$

Suppose $i < j$. Then

$$\begin{aligned} (\mathbf{AA}')_{ij} &= i(1 - \tau_i)(1 - \tau_j) - (j - i)\tau_i(1 - \tau_j) + (J - j)\tau_i\tau_j \\ &= i\frac{(J - i)(J - j)}{J^2} - (j - i)\frac{i(J - j)}{J^2} + (J - j)\frac{ij}{J^2} \\ &= \frac{i(J - j)}{J^2}(J - i - j + i + j) \\ (\mathbf{AA}')_{ij} &= \frac{i(J - j)}{J} \end{aligned}$$

Suppose $j < i$. Since $\mathbf{A}'_j = \mathbf{A}_j$ and $\mathbf{A}_i = \mathbf{A}'_i$,

$$(\mathbf{AA}')_{ij} = \mathbf{A}_i \mathbf{A}'_j = (\mathbf{A}'_i)' (\mathbf{A}_j)' = (\mathbf{A}_j \mathbf{A}'_i)' = \mathbf{A}_j \mathbf{A}'_i.$$

And so by the previous, if $j < i$,

$$(\mathbf{AA}')_{ij} = \frac{j(J - i)}{J}.$$

\mathbf{AA}' is the $(J - 1) \times (J - 1)$ matrix of the following form:

$$\frac{1}{J} \begin{bmatrix} (J-1) & (J-2) & (J-3) & (J-4) & \cdots & 4 & 3 & 2 & 1 \\ (J-2) & 2(J-2) & 2(J-3) & 2(J-4) & \cdots & 8 & 6 & 4 & 2 \\ (J-3) & 2(J-3) & 3(J-3) & 3(J-4) & \cdots & 12 & 9 & 6 & 3 \\ (J-4) & 2(J-4) & 3(J-4) & 4(J-4) & \cdots & 16 & 12 & 8 & 4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 4 & 8 & 12 & 16 & \cdots & 4(J-4) & 3(J-4) & 2(J-4) & (J-4) \\ 3 & 6 & 9 & 12 & \cdots & 3(J-4) & 3(J-3) & 2(J-3) & (J-3) \\ 2 & 4 & 6 & 8 & \cdots & 2(J-4) & 2(J-3) & 2(J-2) & (J-2) \\ 1 & 2 & 3 & 4 & \cdots & (J-4) & (J-3) & (J-2) & (J-1) \end{bmatrix}$$

To find the inverse of \mathbf{AA}' , we apply the Gauss-Jordan reduction on $\mathbf{B} = J[\mathbf{AA}' \mathbb{I}_{J-1}]$.

Step 1: for $i = 1, \dots, J - 2$, $\mathbf{B}_i - \mathbf{B}_{i1} \times \mathbf{B}_{(J-1)}$, and $-\mathbf{B}_i \rightarrow \mathbf{B}_i$, resulting in

$$\mathbf{B} \sim \left[\begin{array}{cccccccc|cccccccc} 0 & J & 2J & 3J & \cdots & (J-5)J & (J-4)J & (J-3)J & (J-2)J & -J & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & (J-1)J \\ 0 & 0 & J & 2J & \cdots & (J-6)J & (J-5)J & (J-4)J & (J-3)J & 0 & -J & 0 & 0 & \cdots & 0 & 0 & 0 & (J-2)J \\ 0 & 0 & 0 & J & \cdots & (J-7)J & (J-6)J & (J-5)J & (J-4)J & 0 & 0 & -J & 0 & \cdots & 0 & 0 & 0 & (J-3)J \\ 0 & 0 & 0 & 0 & \cdots & (J-8)J & (J-7)J & (J-6)J & (J-5)J & 0 & 0 & 0 & -J & \cdots & 0 & 0 & 0 & (J-4)J \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & J & 2J & 3J & 4J & 0 & 0 & 0 & 0 & \cdots & -J & 0 & 0 & 4J \\ 0 & 0 & 0 & 0 & \cdots & 0 & J & 2J & 3J & 0 & 0 & 0 & 0 & \cdots & 0 & -J & 0 & 3J \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & J & 2J & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -J & 2J \\ 1 & 2 & 3 & 4 & \cdots & (J-4) & (J-3) & (J-2) & (J-1) & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & J \end{array} \right]$$

Step 2: for $i = 1, \dots, J - 2$, $\mathbf{B}_i \rightarrow \mathbf{B}_{(i+1)}$, and $\mathbf{B}_{(J-1)} \rightarrow \mathbf{B}_1$, resulting in

$$\mathbf{B} \sim \left[\begin{array}{cccccccc|cccccccc} 1 & 2 & 3 & 4 & \cdots & (J-4) & (J-3) & (J-2) & (J-1) & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & J \\ 0 & J & 2J & 3J & \cdots & (J-5)J & (J-4)J & (J-3)J & (J-2)J & -J & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & (J-1)J \\ 0 & 0 & J & 2J & \cdots & (J-6)J & (J-5)J & (J-4)J & (J-3)J & 0 & -J & 0 & 0 & \cdots & 0 & 0 & 0 & (J-2)J \\ 0 & 0 & 0 & J & \cdots & (J-7)J & (J-6)J & (J-5)J & (J-4)J & 0 & 0 & -J & 0 & \cdots & 0 & 0 & 0 & (J-3)J \\ 0 & 0 & 0 & 0 & \cdots & (J-8)J & (J-7)J & (J-6)J & (J-5)J & 0 & 0 & 0 & -J & \cdots & 0 & 0 & 0 & (J-4)J \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & J & 2J & 3J & 0 & 0 & 0 & 0 & \cdots & -J & 0 & 0 & 4J \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & J & 2J & 0 & 0 & 0 & 0 & \cdots & 0 & -J & 0 & 3J \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & J & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -J & 2J \end{array} \right]$$

Bibliography

- [1] AGRESTI, A. *Categorical Data Analysis*, 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2002.
- [2] ALVO, M., AND BERTHELOT, M.-P. Nonparametric tests of trend for proportions. *International Journal of Statistics and Probability* 1, 1 (2012), 92–104.
- [3] ANDERSON, R. L. Use of contingency tables in the analysis of consumer preference studies. *Biometrics* 15, 4 (1959), 582–590.
- [4] BAGDONAVIČIUS, V., KRUPIS, J., AND NIKULIN, M. S. *Non-parametric Tests for Complete Data*, 1st ed. Wiley & Sons, Inc. and ISTE, Hoboken, USA and London, UK, 2011.
- [5] BARTON, D. E. A form of Neyman’s ψ_k^2 test of goodness of fit applicable to grouped and discrete data. *Scandinavian Actuarial Journal 1955*, 1-2 (1955), 1–16.
- [6] BERTHELOT, M.-P. Tests non-paramétriques de tendance sur les proportions. Master’s thesis, University of Ottawa, 2012.
- [7] BEST, D. J. Extended analysis for ranked data. *The Australian Journal of Statistics* 35, 3 (1993), 257–262.
- [8] BEST, D. J., AND RAYNER, J. C. W. Nonparametric analysis of doubly ordered two-way contingency tables. *Biometrics* 52, 3 (1996), 1153–1156.
- [9] BESTGEN, Y. Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary & Linguistic Computing* (2013).
- [10] BEY, E., AND DAVY, P. Partitioning Pearson’s chi-squared statistic for a completely ordered three-way contingency table. *Australian & New Zealand Journal of Statistics* 40, 4 (1998).
- [11] BEY, E., AND DAVY, P. Partitioning Pearson’s chi-squared statistic for a partially ordered three-way contingency table. *Australian & New Zealand Journal of Statistics* 41, 2 (1999).
- [12] BIENAYMÉ, J. Sur la probabilité des résultats moyens des observations; démonstration directe de la règle de laplace. *Mémoires Sav. Étrangers Acad. Sci.* 5 (1838), 513–558.
- [13] BISHOP, Y., FIENBERG, S. E., AND HOLLAND, P. W. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge, 1975.
- [14] BOUYN, S. Smooth goodness-of-fit specification tests under the Lagrange multiplier principle. *Communications in Statistics - Theory and Methods* 3, 37 (2008), 443–459.
- [15] BRAVAIS, A. Analyse mathématique sur les probabilités des erreurs de situation d’un point. *Mém. de l’Institut de France* 9 (1846), 255–332.

- [16] BROCKHOFF, P. B., BEST, D. J., AND RAYNER, J. C. W. Partitioning Anderson's statistic for tied data. *Journal of Statistical Planning and Inference* 121 (2004), 93–111.
- [17] CONOVER, W. J. *Practical Nonparametric Statistics*. Wiley & Sons, Inc., New York, USA, 1971.
- [18] COX, D. R., AND HINKLEY, D. V. *Theoretical Statistics*. Chapman and Hall, 1974, p. 315.
- [19] DE MOIVRE, A. *Miscellanea Analytica*. London, 1733, ch. Approximatio ad summam terminorum $(a + b)^n$ in seriem expansi.
- [20] EDGEWORTH, F. Y. On the mathematical representation of statistical data. *Philosophical Magazine* 5, 34 (1892), 429–438 and 518–526.
- [21] EFROMOVICH, S. Orthogonal series density estimation. *Wiley Interdisciplinary Reviews: Computational Statistics* 2, 4 (2010), 467–476.
- [22] EMERSON, P. L. Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics* 24, 3 (1968), 695–701.
- [23] GOODMAN, L. Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *Journal of the American Statistical Association* 66, 334 (1971).
- [24] HIROTSU, C. Cumulative chi-squared statistic as a tool for testing goodness of fit. *Biometrika* 73, 1 (1986), 165–173.
- [25] IVERSEN, G. R. Decomposing chi-square: A forgotten technique. *Sociological Methods Research* 8, 2 (1979), 143–157.
- [26] KENDALL, SIR. M. G., AND STUART, A. *The Advanced Theory of Statistics*, 4th ed., vol. 1-2. C Griffin, 1977.
- [27] LANCASTER, H. O. *The Chi-Squared Distribution*. John Wiley & Sons, Inc., New York, USA, 1969.
- [28] NAIR, V. On testing against ordered alternatives in analysis of variance models. *Biometrika* 73, 2 (1986), 493–499.
- [29] NAIR, V. Testing in industrial experiments with ordered categorical data. *Technometrics* 28, 4 (1986), 283–291.
- [30] NAIR, V. Chi-squared-type test for ordered alternatives in contingency tables. *Journal of the American Statistical Association* 82, 397 (1987), 293–291.
- [31] NEYMAN, J. Smooth test for goodness of fit. *Scandinavian Actuarial Journal* 1937, 3-4 (1937), 149–199.
- [32] PEARSON, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* 50, 2 (1900).
- [33] RAYNER, J., AND BEH, E. Components of Pearson's statistic for at least partially ordered m -way contingency tables. *Journal of Applied Mathematics and Decision Sciences* 2009, 980706 (2009).
- [34] RAYNER, J., AND BEST, D. *A Contingency Table Approach to Nonparametric Testing*. Chapman & Hall, Boca Raton, FL, USA, 2001.
- [35] RAYNER, J. C. W., AND BEST, D. J. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York, 1989.

- [36] RAYNER, J. C. W., AND BEST, D. J. Smooth extensions of Pearson's product moment correlation and Spearman's rho. *Statistics & Probability Letters* 30 (1996), 171–177.
- [37] RAYNER, J. C. W., AND BEST, D. J. Extensions to the Kruskal-Wallis test and a generalised median test with extensions. *Journal of Applied Mathematics & Decision Sciences* 1, 1 (1997), 13–25.
- [38] RAYNER, J. C. W., AND BEST, D. J. Analysis of singly ordered two-way contingency tables. *Journal of Applied Mathematics & Decision Sciences* 4, 1 (2000), 83–98.
- [39] RAYNER, J. C. W., THAS, O., AND BEST, D. J. *Smooth Tests of Goodness of Fit: Using R*, 2nd ed. John Wiley & Sons, 2009.
- [40] SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2 (1946), 110–114.
- [41] SCHACH, S. An alternative to the Friedman test with certain optimality properties. *The Annals of Statistics* 7, 3 (1979), 537–550.
- [42] SCHOLS, C. M. Over de theorie der fouten in de ruimte en in het platte vlak. *Verh. Kon. Akad. Wetens., Amsterdam* 15 (1875), 67.
- [43] SHEPPARD, W. F. On the application of the theory of error to cases of normal distribution and normal correlation. *Philosophical Transactions of the Royal Society of London* 192 (1898), 101–167.
- [44] SHEPPARD, W. F. On the geometrical treatment of the 'normal curve' of statistics with special reference to correlation and to the theory of errors. *Proceedings of the Royal Society of London* 62 (1898), 170–173.
- [45] TAGUCHI, G. A new statistical analysis for clinical data, the accumulating analysis, in contrast with the chi-square test. *Saishin Igaku (The Newest Medicine)*, 29 (1974), 806–813.
- [46] TAGUCHI, G. *Statistical Analysis (in Japanese)*. Maruzen, Tokyo, Japan, 2000.
- [47] THOMAS, D. R., AND PIERCE, D. A. Neyman's smooth goodness-of-fit test when the hypothesis is composite. *Journal of the American Statistical Association* 74, 366 (1979), 441–445.
- [48] WHITTAKER, E. T., AND WATSON, G. N. *A Course of Modern Analysis*. Cambridge University Press, 1927.
- [49] WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* 9 (1938), 60–62.