Statistical Analysis in Multivariate Sampling

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von M.Sc. Nuri Eltabit Mohamed

geb. am 15.07.1966 in Gurdha Shati

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. rer. nat. habil. Rainer Schwabe

Prof. Dr. rer. nat. habil. Waltraud Kahle

eingereicht am: 03.05.2010

Verteidgung am: 05.04.2011

Zusammenfassung

Im Rahmen der Dissertation mit dem Titel "Statistical Analysis in Multivariate Sampling" wird die Analyse von Zähldaten betrachtet. Hierbei werden drei Fälle unterschieden. Der univariate Fall, bei dem die m Beobachtungen durch Zufallsvariablen der Form (Y_i, X_i) beschrieben werden, sowie die bivariate Analyse und die multivariate Analyse, bei der die Daten durch Zufallvektoren $(Y_{ij}, X_{ij}), i = 1, ..., m,\ j = 1, ..., k,\ k = 2$ bzw $k > 2$ modelliert werden.
Ein grundlegendes Ziel dieser Arbeit ist es, basierend auf geeigneten Modellannahmen gute Schätzungen für die Häufigheit eines Merkmals zu erhalten (zum Beispiel: Schätzung der Anzahl an defekten oder schadhaften Teilen Y in einem bestimmten Werk, das eine bekannte Anzahl an Teilen produziert oder die Schätzung des Anteils in Bezug auf die Gesamtzahl). Ebenfalls von interesse ist, Die Konstruktion zuverlässiger Konfidenzintervalle für Anteile oder Linearkombinationen $\boldsymbol{\alpha}^T\mathbf{p}$ dieser Anteile, was Gewinne oder Verluste beschreiben kann. Hierbei ist $\boldsymbol{\alpha} \in \mathbf{R}^k,\ \boldsymbol{\alpha} \geq \mathbf{0}, k \geq 1$, und $\mathbf{p} \in [\mathbf{0}, \mathbf{1}]^k$, wobei k die Anzahl der Komponenten der Zähldaten darstellt, und $\mathbf{p}$ der Vektor der Anteile. Die Konstruktion der Konfidenzintervalle für die Anteile p_j folgt ebenfalls als Linearkombination, in dem man $\alpha_j = 1$ und $\boldsymbol{\alpha} = (0, ..., 0, \alpha_j, 0, ..., 0)^T$, $j = 1, ..., k$, wählt.

Abstract

Within the framework of this dissertation entitled 'Statistical Analysis in Multivariate Sampling', the analysis of univariate count data involves pairs of random variables (Y_i, X_i) of m observations, while in the bivariate and multivariate, analysis data of k pairs of random variables $(Y_{ij}, X_{ij}), i = 1, ..., m,\ j = 1, ..., k, k \geq 2$ are involved.

The fundamental goal of the work is, based on the appropriate model assumptions to obtain good estimates for the attribute totals such as: estimating the defective or damage totals, i.e, estimating the defective totals Y in a specific factory containing a known total amount of the productions, or estimating the proportions of those defective or damage totals with respect to the total amount of the items, as well as, constructing reliable confidence intervals for the proportion, or constructing confidence intervals for any linear combination of these proportions (which may describe some monetary gain or loss) $\boldsymbol{\alpha}^T\mathbf{p}$, where
$\boldsymbol{\alpha} \in \mathbf{R}^k,\ \boldsymbol{\alpha} \geq \mathbf{0}$, is a vector of constants, and $\mathbf{p} \in [\mathbf{0}, \mathbf{1}]^k, k \geq 1,\ k$ is the number of components of the count data, and $\mathbf{p}$ is the vector of the underlying proportions. Constructing confidence intervals for any proportion p_j can be obtained as a linear combination of the proportions by assigning the value $\alpha_j = 1$ in the vector
$\boldsymbol{\alpha} = (0, ..., 0, \alpha_j, 0, ..., 0)^T,\ j = 1, ..., k.$

Table of Contents

Introduction

Statistical analysis plays an important role in economics, biology, medicine, physics, and social sciences with a broad range in the other different fields.

In some circumstances as for example: In economic crises, it is important to obtain a reliable estimate or a function of estimates for the damage proportion(s) to determine which are compatible with the value of the damage proportion or a function of the damage proportions for various productions. In medicine, one can also produce good estimates of the proportions for the patients based on the collected sample data, to see how these estimates can be compatible with the corresponding proportions taking in account any dependencies between the data components, and so on.

The fundamental aims of the statistical analysis usually are:

- Estimating the coefficients of the considered model.
- Evaluate fitting the model to the data.
- Discovering or predicting further data.
- Making a statistical inference (i.e, confidence interval, test of hypothesis regarding the unknown coefficients) about the model coefficients.

The aims of the statistical analysis in this thesis are:

- Producing estimates for the components proportions based on the assumed model.
- Obtaining the asymptotic distribution of the estimators by involving asymptotic theory (asymptotic normality), as well as, constructing the approximate confidence intervals for the model coefficients, i.e, for the proportions or a linear combination of the proportions. Thus, all the procedure results will be approximate results.

This dissertation is devoted to studies and applies the statistical analysis methods to one or more dimensional data under the assumed model that describes the relationship among the variables.

On the one direction, the most famous and simple models have been used throughout this work are the linear models, and the bivariate, multivariate SUR models, which are the fundamental analyzing of the univariate, and the multivariate sampling data (non-linear models are not deal with here, nor included in this work).

On the other direction and according to the structure of data (count), the next involved models are the ' Univariate', ' Bivariate ', and the ' Multivariate' Poisson models (these models were discussed in: [6], and [9], [10], [11], and [13]).

Chapter 1

Univariate data analysis

1.1 Introduction

The univariate analysis deals with analysis of a single random variable, however in fact being analysis of pair random variables. In this chapter, we will analyze the collected sample count data (one dimension), sampling from a certain finite population.

Suppose, we are sampling from an infinite population, namely the *i.i.d* pairs $(Y_1, X_1), ..., (Y_m, X_m)$ is a random sample of size m drawn from an infinite population such that for each index i associated with the pair r.v's (Y_i, X_i) restricted by $0 \leq Y_i \leq X_i$, $\forall i$.

The count variables (Y_i, X_i) have the attributes, for instances:

$X_i \equiv$ No.of children in the family i, or No.of non defects of the product i for a specific factory.

$Y_i \equiv$ No.of male children in the same family i, or No.of defects for the same product i for a specific factory.

We will consider the common approach, $\hat{p} = \frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} X_i}$, as a sample proportion used to estimate the unknown population proportion p, where

$$p = \frac{E(Y_i)}{E(X_i)}, E(\hat{p}) = p.$$

We will start analyzing the sample data with the fundamental method of the analyzing. It will be assumed in the following section that the relationship between the random variable Y_i and the corresponding variable X_i is linear relationship, and the linear regression technique will be involved to analyze the data under the assumed linear model. Let us first introduce to the basic knowledge of the general linear model.

1.2 General linear model and linear ratio model

The general form of the univariate multiple linear regression model is written as

$$Y_i = \mathbf{f}(X_i)^T\boldsymbol{\beta} + \epsilon_i = \sum_{j=1}^{k} f_j(X_i)\beta_j + \epsilon_i,$$

where, $\boldsymbol{\beta} = (\beta_1, ..., \beta_k) \in \mathbf{R}^k$ is a vector of the unknown model coefficients, the real valued functions $\mathbf{f}(X_i) = (f_1(X_i), ..., f_k(X_i))^T$ are the regression functions linear in the $\beta_j' s$, and the random variable ϵ_i is the errors term of the model satisfying $E(Y_i) = \mathbf{f}(X_i)^T\boldsymbol{\beta}, \forall i = 1, ..., m$. We assume that there is no explicit intercept included in this model, it depends on the considered problem. In particular if, $k = 1$, this leads to $\boldsymbol{\beta} = \beta_1 = p$, $\mathbf{f}(X_i) = f_1(X_i) = f(X_i)$, so the model is reduced to the model

$$Y_i = f(X_i)p + \epsilon_i,$$

with the assumptions

$$E(\epsilon_i \mid f(X_i)) = 0, \text{ and, } Var(\epsilon_i \mid f(X_i)) = \sigma^2 f(X_i),$$

on the model errors $\epsilon_i, i = 1, ..., m$.

And further, assume the function $f(X_i) = X_i$, then the model becomes

$$Y_i = X_i p + \epsilon_i,$$

where, $E(\epsilon_i \mid X_i) = 0$, and $Var(\epsilon_i \mid X_i) = \sigma^2 X_i$, (either X_i fixed or random), $i = 1, ..., m$.

The unconditional variance, or marginal variance (by the law of total variance), and the unconditional expectation of the model errors are given by

$$\begin{aligned} Var(\epsilon_i) &= E\left(Var(\epsilon_i \mid X_i)\right) + Var\left(E(\epsilon_i \mid X_i)\right) = \sigma^2 E(X_i), \\ E(\epsilon_i) &= E\left(E(\epsilon_i \mid X_i)\right) = 0, \end{aligned} \tag{1.1}$$

as $E(\epsilon_i \mid X_i) = 0$.

1.2.1 Linear model (LM)

It will be assumed first the following univariate linear model

$$Y_i = x_i p + \epsilon_i, i = 1, ..., m, \tag{1.2}$$

with the assumptions:

$E(Y_i) = x_i p,\ E(\epsilon_i) = 0$, and with variance proportional to x_i (x_i is fixed variable), i.e, $Var(Y_i) = Var(\epsilon_i) = \sigma^2 x_i$.

Or, it would be convenient to compress the model in vector form

$$\mathbf{Y} = \mathbf{X}p + \boldsymbol{\epsilon}, \tag{1.3}$$

where, the $m \times 1$ dimension response vector $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$, and the $m \times 1$ design vector $\mathbf{X} = (x_1, \cdots, x_m)^T$, and the heteroscedastic errors $(\epsilon_1, \cdots, \epsilon_m)^T = \boldsymbol{\epsilon}$, with the assumptions, $E(\boldsymbol{\epsilon}) = \mathbf{0}_m$, and $Var(\boldsymbol{\epsilon}) = \sigma^2 W$, where $\mathbf{0}_m = (0, ..., 0)^T$, and $W = diag(x_i)$. In other words, the errors ϵ_i are uncorrelated and have variance proportional to the x_i.

1.2.2 Normal linear model (NLM)

Next, It will be assumed the normality of the errors of the linear model 1.3, given the fixed design vector $\mathbf{X} = (x_1, ..., x_m)^T$, i.e, $x_1, ..., x_m$ are fixed or non random variables, also called covariates or predictors. I.e, given fixed x_i, the errors ϵ_i are independently normally distributed with mean 0 and with variance proportional to x_i, i,e, given x_i, the response variable Y_i has $N(x_i p, \sigma^2 x_i)$, or, the errors ϵ_i are $i.i.d$ normally distributed with mean 0 and with variance $\sigma^2 E(x_i),\ i = 1, ..., m$.

The relationship between Y_i and the predictors x_i is postulate as the linear model

$$Y_i = x_i p + \epsilon_i,\ i = 1, ..., m, \tag{1.4}$$

with the assumptions

$\epsilon_i \sim N(0, \sigma^2 x_i)$, as well as $Y_i \sim N(x_i p, \sigma^2 x_i)$, it follows that $Var(Y_i) = Var(\epsilon_i) = \sigma^2 x_i$ (constants) $i = 1, ..., m$, the model is called ***Normal linear model*** (In fact, this model is called, an approximate NLM (ANLM), due to $P(Y_i < 0) = 0, \forall i = 1, ..., m$, which is not satisfied for normality of the model 1.4, but to use the Normal distribution tools, the model will be assumed as NLM).

The model (1.4) will compress in vector form

$$\mathbf{Y} = \mathbf{X}p + \boldsymbol{\epsilon},$$

where, $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$, and the design vector $\mathbf{X} = (x_1, \cdots, x_m)^T$, and the correlated normal errors $(\epsilon_1, \cdots, \epsilon_m)^T = \boldsymbol{\epsilon}$, with the assumptions

$E(\boldsymbol{\epsilon}) = \mathbf{0}_m$, and $Var(\boldsymbol{\epsilon}) = \sigma^2 W$, where $\mathbf{0}_m = (0, ..., 0)^T$, and $W = diag(x_i)$.

1.2.3 Standardizing the LM and NLM

It would be more convenient to work with models having constant variances or covariances rather than with variable ones.

In this paragraph, the univariate linear regression model will be standardized, where all covariates are weighted by the square root of the inverse of the function $f(X_i)$ in the conditional variance, where $f(X_i) = x_i$. Throughout we will only consider $x_i > 0$. Thus, the linear model 1.2 will be transformed by the transformation
$A_i Y_i = A_i x_i p + A_i \epsilon_i$, to obtain the ***weighted LM***

$$\widetilde{Y}_i = \widetilde{x}_i p + \widetilde{\epsilon}_i, i = 1, ..., m \tag{1.5}$$

where, $A_i = \frac{1}{\sqrt{x_i}}$, given that $x_i > 0, \widetilde{Y}_i = A_i Y_i = \frac{Y_i}{\sqrt{x_i}}$,$\widetilde{x}_i = A_i x_i = \sqrt{x_i}$, and $\widetilde{\epsilon}_i = A_i \epsilon_i = \frac{\epsilon_i}{\sqrt{x_i}}$, it follows that $E(\widetilde{\epsilon}_i) = 0$, and
$Var(\widetilde{\epsilon}_i) = Var(A_i \epsilon_i) = \sigma^2 \; \forall i = 1, ..., m$ (*homoscedastic errors*).

Similarly, the NLM 1.4 will be standardized to obtain the weighted normal linear model

$$\widetilde{Y}_i = \widetilde{x}_i p + \widetilde{\epsilon}_i, \tag{1.6}$$

with the assumptions, $\widetilde{\epsilon}_i \sim N(0, \sigma^2)$, $\widetilde{Y}_i \sim N(\widetilde{x}_i p, \sigma^2)$, $\widetilde{x}_i = \sqrt{x_i}$, $x_i > 0, \forall i = 1, ..., m$, i.e, the weighted errors are the *i.i.d* Normal random variables with mean 0 and finite variance σ^2, this model is called the ***weighted NLM***.

It would be convenient to rewrite the transformed linear models 1.5, and 1.6 in vector notation

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}} p + \widetilde{\boldsymbol{\epsilon}}, \tag{1.7}$$

where, $E(\widetilde{\boldsymbol{\epsilon}}) = \mathbf{0}_m$, and $E(\widetilde{\mathbf{Y}}) = \widetilde{\mathbf{X}} p$, as well as $Cov(\widetilde{\boldsymbol{\epsilon}}) = \sigma^2 I_m = Cov(\widetilde{\mathbf{Y}})$, where I_m is an identity matrix of dimension $m \times m$, and the weighted response vector
$\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \cdots, \widetilde{Y}_m)^T$, and the weighted design vector $\widetilde{\mathbf{X}} = (\widetilde{x}_1, \cdots, \widetilde{x}_m)^T$, as well as the weighted error vector $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\epsilon}_1, \cdots, \widetilde{\epsilon}_m)^T$.

It follows from the model 1.6 that, the error vector $\widetilde{\boldsymbol{\epsilon}}$ has the multivariate Normal distribution (also, as the AMVN (the approximate multivariate Normal distribution), but it will be much better to obtain exact results, therefore the model will be considered as the MVN) with $\mathbf{0}_m$ mean vector, and nonsingular covariance matrix $\sigma^2 I_m$, i.e $\widetilde{\boldsymbol{\epsilon}} \sim N_m(\mathbf{0}_m, \sigma^2 I_m)$.

1.2.4 Estimation in linear models

Estimation of the coefficient of the univariate linear ratio model deals with two cases, according whether the model errors are homoscedastic or heteroscedastic errors.

Homoscedasticity case

Under normality of the errors of the model 1.7, the weighted error vector $\widetilde{\boldsymbol{\epsilon}}$ has the multivariate Normal distribution with the $\mathbf{0}$ mean vector and the nonsingular covariance matrix $\sigma^2 I_m$, i.e, has variance proportional to the identity matrix I_m. In other words, the weighted errors are *iid* normal random variables with mean 0 and with constant variance σ^2 (*homoscedastic errors*).
On the other side, it is well-known that the WLSE (weighted least squares estimator) is the BLUE (the best linear unbiased estimator or the optimal estimator), and since the weighted errors are homoscedastic then, the WLSE applied to the weighted model 1.7 results in the OLSE (Ordinary least squares estimator), and hence is also the BLUE (the best linear unbiased estimator or the optimal estimator, according to Gauss-Markov's theorem, see [7], pp. 588-591 or, [20], pp. 35-42), i.e., $\hat{p}_{WLS} = \hat{p}_{OLS}$, in case of homoscedastic errors. Mathematically, one can investigate this as following:

Since, $Cov(\widetilde{\boldsymbol{\epsilon}}) = \sigma^2 I_m$, then

$$\begin{aligned}
\hat{p}_{WLS} &= \left(\widetilde{\mathbf{X}}^T \left(\sigma^2 I_m\right)^{-1} \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \left(\sigma^2 I_m\right)^{-1} \widetilde{\mathbf{Y}} \\
&= \left(\widetilde{\mathbf{X}}^T I_m \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T I_m \widetilde{\mathbf{Y}} = \left(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \widetilde{\mathbf{Y}} = \hat{p}_{OLS} \\
&= \left(\sum_{i=1}^{m} (\sqrt{x_i})^2\right)^{-1} \sum_{i=1}^{m} \sqrt{x_i} \widetilde{Y}_i = \frac{\sum_{i=1}^{m} \sqrt{x_i} \widetilde{Y}_i}{\sum_{i=1}^{m} x_i} = \frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} x_i} = \hat{p},
\end{aligned}$$

which results in the *ratio estimator*, where $\sqrt{x_i}\widetilde{Y}_i = Y_i$.

Heteroscedasticity case

Since, the error vector of the model 1.3 has covariance that is the variance not proportional to the identity matrix, i.e, proportional to the known invertible diagonal matrix W, so, it will be used the GLSE (Generalized least squares estimator), which is also the BLUE. Specifically, if $A = \Sigma^{-1/2}$ is a non singular symmetric positive definite matrix, then $A^T \Sigma A = I_m$.

The covariance structure of the model 1.3, is given by

$$Cov(\boldsymbol{\epsilon}) = \begin{pmatrix} \sigma^2 x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 x_m \end{pmatrix} = \sigma^2 \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_m \end{pmatrix} = \sigma^2 W,$$

i.e, proportional to an invertible matrix W, where $W = diag(x_i)$, x_i, are fixed, $i = 1, ..., m$, and $W^{-1}W = I_m$, as well as $\mathbf{X}^T W^{-1} = 1_m^T$, where I_m is the identity matrix, and $1_m = (1, ..., 1)^T$.
So, we have

$$\hat{p}_{GLS} = \left(\mathbf{X}^T \left(\sigma^2 W\right)^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \left(\sigma^2 W\right)^{-1} \mathbf{Y} = \left(\mathbf{X}^T W^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T W^{-1} \mathbf{Y}$$
$$= \left(1_m^T \mathbf{X}\right)^{-1} 1_m^T \mathbf{Y} = \left(\sum_{i=1}^m x_i\right)^{-1} \sum_{i=1}^m Y_i = \frac{\sum_{i=1}^m Y_i}{\sum_{i=1}^m x_i} = \hat{p},$$

which results in the *ratio estimator*.
Note that:

- The regression equation used to estimate the true mean value $E(\widetilde{\mathbf{Y}})$ in eq 1.7 can be written as $\widehat{\widetilde{\mathbf{Y}}} = \widetilde{\mathbf{X}}\hat{p}$.
- The variance of the estimator $\hat{p}_{WLS}$ in case of *homoscedasticity* is $\sigma^2 \left(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}\right)^{-1}$, whereas in the case of *heteroscedasticity* is $\sigma^2 \left(\mathbf{X}^T W^{-1} \mathbf{X}\right)^{-1}$, which are equal (for more details see [19] pp. 148-149).

1.2.5 Properties of the ratio estimator $\hat{p}$

Unbiasedness

The expected value of the *ratio estimator* $\hat{p}$ is the proportion p, i.e, the *ratio estimator* $\hat{p}$ is an unbiased

$$E(\hat{p}) = E\left(\frac{\sum_{i=1}^m \widetilde{Y}_i \sqrt{x_i}}{\sum_{i=1}^m x_i}\right) = \frac{\sum_{i=1}^m \sqrt{x_i} E(\widetilde{Y}_i)}{\sum_{i=1}^m x_i} = \frac{\sum_{i=1}^m \sqrt{x_i} p \sqrt{x_i}}{\sum_{i=1}^m x_i} = \frac{p \sum_{i=1}^m x_i}{\sum_{i=1}^m x_i} = p.$$

Variation

The variability of the *ratio estimator* from the proportion p is given by

$$Var(\hat{p}) = Var\left(\frac{\sum_{i=1}^m \widetilde{Y}_i \sqrt{x_i}}{\sum_{i=1}^m x_i}\right)$$
$$= \frac{1}{\left(\sum_{i=1}^m x_i\right)^2} \sum_{i=1}^m x_i Var(\widetilde{Y}_i) = \frac{\sigma^2 \sum_{i=1}^m x_i}{\left(\sum_{i=1}^m x_i\right)^2} = \frac{\sigma^2}{\sum_{i=1}^m x_i} = \sigma^2 \left(\widetilde{\mathbf{x}}^T \widetilde{\mathbf{x}}\right)^{-1},$$

derivation of the properties of the estimator $\hat{p}_{OLS}$ in the univariate multiple regression model with intercept is given in [19] pp. 129.

Given the fixed design vector $\mathbf{X}$ ($Y_i's$ are Normal), we have

$$\hat{p} \sim N\left(p, \frac{\sigma^2}{\sum_{i=1}^{m} x_i}\right), \tag{1.8}$$

where, σ^2 is unknown finite variance.

An unbiased consistent estimator $\hat{\sigma}^2$ for the unknown finite variance σ^2 based on the m residuals is given by

$$\hat{\sigma}^2 = s_e^2 = \frac{1}{m-1}\sum_{i=1}^{m}\frac{(Y_i - \hat{p}x_i)^2}{x_i} = \frac{1}{m-1}\sum_{i=1}^{m}(\widetilde{Y}_i - \hat{p}\widetilde{x}_i)^2 = s_{\widetilde{e}}^2,$$

where, $\widetilde{x}_i = \sqrt{x_i},\ x_i > 0,\ \forall i,\ \widetilde{Y}_i = \frac{Y_i}{\widetilde{x}_i}$.

For investigation: since $\frac{\sum_{i=1}^{m}\widetilde{e}_i^2}{\sigma^2} \sim \chi^2_{m-1}$, or $\frac{(m-1)s_{\widetilde{e}}^2}{\sigma^2} \sim \chi^2_{m-1}$, hence $E\left(\frac{(m-1)s_{\widetilde{e}}^2}{\sigma^2}\right) = m-1 \Rightarrow \frac{(m-1)}{\sigma^2}E(s_{\widetilde{e}}^2) = m-1 \Rightarrow E(s_{\widetilde{e}}^2) = \sigma^2$, and thus $s_{\widetilde{e}}^2$ is an unbiased estimator of σ^2 (which in turn, is the BUE according to the Lehmann-Scheffé theorem (see [8], pp. 426-430), the complete proving is given in the chapter 2 subsection 2.2.6).

In case of the ANLM ($Y_i's$ are not normal) one may obtain, asymptotically:

$$\sqrt{m}(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \frac{\sigma^2}{\mu}\right),$$

or approximately

$$\hat{p}_m \approx N\left(p, \frac{\sigma^2}{m\mu}\right),$$

provided that, $\frac{1}{m}\sum_{i=1}^{m} x_i \longrightarrow^{\mathcal{P}} \mu$, where, μ is known constant, and $\longrightarrow^{\mathcal{P}}$, $\longrightarrow^{\mathcal{D}}$ denote respectively, convergence in probability and in distribution.

1.2.6 Asymptotic normality of the estimator $\hat{p}_{OLS}$ (ratio estimator $\hat{p}_m$)

So far, it has been assumed that, given fixed x_i the errors are normally distributed in which one could obtain the exact distribution of the ratio estimator $\hat{p}$, while our interesting is to consider the asymptotic distribution of $\hat{p}_m$ in case of the non-normal errors, but under the stochasticity of X_i.

Thus, one would consider as well as identify the asymptotic Normal distribution of the ratio estimator $\hat{p}_m$ provided that the sample size is enough large.

Addition conditions on the pair of observations X_i, Y_i are required, namely (X_i, Y_i) are $i.i.d$ pairs of r.v's, and $E(X_i)$ exists $\Rightarrow (\widetilde{X}_i, \widetilde{Y}_i)$ are also $i.i.d$ random variables, and $E(\widetilde{X}_i^2)$ exists, $i = 1, ..., m$, $\widetilde{X}_i = \sqrt{X_i}, \widetilde{Y}_i = \frac{Y_i}{\sqrt{X_i}}$. Here we assume throughout that $X_i > 0$ almost surely.

To derive the asymptotic distribution, one rewrite first the estimator $\hat{p}_{OLS}$ as

$$\begin{aligned}\hat{p}_{OLS} &= \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}} = \left(\sum_{i=1}^{m}\widetilde{X}_i\widetilde{X}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{X}_i\widetilde{Y}_i \\ &= \left(\sum_{i=1}^{m}\widetilde{X}_i^2\right)^{-1}\sum_{i=1}^{m}\widetilde{X}_i\left(\widetilde{X}_i p + \widetilde{\epsilon}_i\right), \qquad (1.9)\end{aligned}$$

so

$$\begin{aligned}\sqrt{m}(\hat{p}_{OLS} - p) &= \left(\frac{1}{m}\sum_{i=1}^{m}\widetilde{X}_i^2\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{X}_i\widetilde{\epsilon}_i\right) \\ &= \left(\frac{1}{m}\sum_{i=1}^{m}\left(\sqrt{X_i}\right)^2\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sqrt{X_i}\widetilde{\epsilon}_i\right) \\ &= \left(\frac{1}{m}\sum_{i=1}^{m}X_i\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sqrt{X_i}\widetilde{\epsilon}_i\right). \qquad (1.10)\end{aligned}$$

Derivation of the asymptotic normality of the equation 1.10, needs to verify, the denominator in 1.10 is consistent, and the numerator obeys the Central limit theorem. It straightforward to see (by the LLN)

$$\left(\frac{1}{m}\sum_{i=1}^{m}X_i\right)^{-1} \longrightarrow^{\mathcal{P}} (E(X_i))^{-1},$$

provided that, $E(X_i) > 0$, also

$$\left(\frac{1}{m}\sum_{i=1}^{m}x_i\right)^{-1} \longrightarrow \mu^{-1}, \mu \text{ is constant.}$$

As well as the numerator

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sqrt{X_i}\widetilde{\epsilon}_i \longrightarrow^{\mathcal{D}} N\left(0, \sigma^2 E(X_i)\right),$$

where, the marginal or asymptotic variance

$$\begin{aligned}Cov\left(\sqrt{X_i}\widetilde{\epsilon}_i, \sqrt{X_i}\widetilde{\epsilon}_i\right) &= Var\left(\sqrt{X_i}\widetilde{\epsilon}_i\right) \\ &= E\left(X_i Var(\widetilde{\epsilon}_i \mid \widetilde{X}_i)\right) + Var(E\sqrt{X_i}\widetilde{\epsilon}_i \mid \widetilde{X}_i)) = \sigma^2 E(X_i).\end{aligned}$$

Therefore, with help of the so-called Slutsky's lemma (see [8], p. 342, or [14], pp. 119-120), one can obtain immediately the asymptotic Normal distribution of the ratio estimator $\hat{p}_m$, thus the equation 1.10 can be rewritten (since, $\hat{p}_{OLS} \equiv \hat{p}_m$) as

$$\sqrt{m}(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \sigma^2 E(X_i)(E(X_i))^{-2}\right) \equiv N\left(0, \sigma^2 (E(X_i))^{-1}\right). \quad (1.11)$$

1.2.7 Confidence intervals for the proportion p

We will give the outlines of the behave of the distribution of the ratio estimator with the proposed confidence intervals of the proportion based on the previous approaches, and on the observation X_i.

proposition 1

x_i are fixed variables:

Standardizing the expression 1.8 (Normal $Y_i's$) will gives

$$\sqrt{m}\,(\hat{p} - p) \sim N\left(0, \frac{\sigma^2}{\overline{x}_m}\right), \quad (1.12)$$

it follows that, the exact $(1-\alpha)\%$ confidence intervals for the proportion p are given by

$$\left[\hat{p} \pm z_{1-\frac{\alpha}{2}}\ S.E(\hat{p})\right], \text{as } \frac{\hat{p}-p}{\frac{\sigma}{\sqrt{\sum_{i=1}^m x_i}}} \sim N(0,1), \text{ when } \sigma^2 \text{ is known}$$

$$\left[\hat{p} \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p})\right], \text{as } \frac{\hat{p}-p}{\frac{s_e}{\sqrt{\sum_{i=1}^m x_i}}} \sim t_{m-1}, \quad \text{when } \sigma^2 \text{ is unknown,}$$

where, the Standard Error of $\hat{p}$, $S.E(\hat{p}) = \frac{\sigma}{\sqrt{\sum_{i=1}^m x_i}}$, and $s.e(\hat{p}) = \frac{s_{\tilde{e}}}{\sqrt{\sum_{i=1}^m x_i}}$.

For the non-Normal $Y_i's$ (since we have non negative count data), and since the sequence $\overline{x}_m = \frac{1}{m}\sum_{i=1}^m x_i \to \mu$, μ is constant, it follows that the sequence of variances $\frac{\sigma^2}{\overline{x}_m}$ converges to the asymptotic variance $\sigma^2(\mu) = \frac{\sigma^2}{\mu}$, as m tends to infinity. Asymptotically, 1.12 can be rewritten as

$$\sqrt{m}(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \frac{\sigma^2}{\mu}\right),$$

where, the asymptotic variance
$Var(\sqrt{m}\hat{p}_m) = E(Var\,(\sqrt{m}\hat{p}_m)) + Var(E(\sqrt{m}\hat{p}_m)) = \frac{\sigma^2}{\mu}$,
as, $Var(E(\sqrt{m}\hat{p}_m)) = 0$. It follows that, an approximate $(1-\alpha)\%$ of the asymptotic confidence interval for the proportion p (unknown σ^2) is given by

$$\left[\hat{p}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{p}_m)\right],$$

as

$$\frac{\hat{p}_m - p}{\frac{s_e}{\sqrt{\sum_{i=1}^{m} x_i}}} \longrightarrow^{\mathcal{D}} N(0,1),$$

where, $s.e(\hat{p}_m) = \frac{s_e}{\sqrt{\sum_{i=1}^{m} x_i}}$. As well as $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ percentile of the Standard Normal distribution, where the Standard Normal random variable Z has the Cumulative Standard Normal distribution function $\Phi : \Phi(z) = P(Z \leq z)$, where $P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha$. Further, one may rewrite (non-Normal $Y_i's$)

$$\frac{\hat{p}_m - p}{\sqrt{\widehat{Var(\hat{p}_m)}}} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1),$$

'$\simeq$' denotes as approximately equal to, where the estimator $\widehat{Var(\hat{p}_m)} = \frac{s_{\tilde{e}}^2}{\sum_{i=1}^{m} x_i}$ is a consistent estimator of the corresponding variance $Var(\hat{p}_m)$, and hence the suggested conservative confidence interval (safety bounds) is given by

$$\left[\hat{p}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p}_m)\right],$$

where, $s.e(\hat{p}_m) = \frac{s_e}{\sqrt{\sum_{i=1}^{m} x_i}}$, and $t_{(m-1,1-\frac{\alpha}{2})}$ is the $(1-\frac{\alpha}{2})$ percentile of the $t-$distribution with $(m-1)$ degrees of freedom.

proposition 2

(Y_i, X_i) are *iid* pairs random variables (not necessarily normal):

Since, $\overline{X}_m \to^{\mathcal{P}} E(X_i)$ (by the LLN, since X_i are $i.i.d$, and $\overline{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$), and hence the sequence of variances $\frac{\sigma^2}{\overline{X}_m}$ converges to the asymptotic variance $\frac{\sigma^2}{E(X_i)}$, as m tends to infinity, so by the Central limit theorem it follows that

$$\sqrt{m}(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \frac{\sigma^2}{E(X_i)}\right),$$

where, the asymptotic variance,
$Var(\sqrt{m}\hat{p}_m) = E(Var(\sqrt{m}\hat{p}_m \mid \mathbf{X}^T)) + Var(E(\sqrt{m}\hat{p}_m \mid \mathbf{X}^T)) = \frac{\sigma^2}{E(X_i)}$, as,
$Var(E(\sqrt{m}\hat{p}_m \mid \mathbf{X}^T)) = 0$, $\mathbf{X} = (X_1, ..., X_m)^T$.
And thus, the approximate $(1-\alpha)\%$ asymptotic confidence interval for the proportion p is given by

$$\left[\hat{p}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{p}_m)\right],$$

where, the standard error of $\hat{p}_m$, $s.e(\hat{p}_m) = \frac{s_m}{\sqrt{\sum_{i=1}^m X_i}}$, $s_m^2 = \frac{1}{m-1}\sum_{i=1}^m \frac{(Y_i - \hat{p}X_i)^2}{X_i}$, and the normal quantile $z_{1-\frac{\alpha}{2}}$ is defined as previous.

Further, and since $\overline{X}_m$ and s_m^2 are consistent estimators for $E(X_i)$ and σ^2 respectively, it follows that a consistent estimator of the $Var(\hat{p}_m)$ is $\frac{s_m^2}{\sum_{i=1}^m X_i}$, also, as

$$\frac{\hat{p}_m - p}{\sqrt{\frac{s_m^2}{\sum_{i=1}^m X_i}}} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1).$$

Hence, the interval whose safety bounds given by

$$\left[\hat{p}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p}_m)\right],$$

is the suggested more conservative confidence interval for the proportion p, where $s.e(\hat{p}_m) = \frac{s_m}{\sqrt{\sum_{i=1}^m X_i}}$.

For the Normal $Y_i's$, and from 1.12, we have

$$P^{\sqrt{m}(\hat{p}-p)|\mathbf{X}} \sim N\left(0, \frac{\sigma^2}{\overline{X}_m}\right),$$

so, the exact $(1-\alpha)\%$ confidence intervals for the proportion p, are

$$\left[\hat{p} \pm z_{1-\frac{\alpha}{2}}\ S.E(\hat{p})\right], \text{as } \frac{\hat{p}-p}{\frac{\sigma}{\sqrt{\sum_{i=1}^m X_i}}} \sim N(0,1), \text{ when } \sigma^2 \text{ is known}$$

$$\left[\hat{p} \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p})\right], \text{as } \frac{\hat{p}-p}{\frac{s_{\tilde{e}}}{\sqrt{\sum_{i=1}^m X_i}}} \sim t_{m-1}, \quad \text{when } \sigma^2 \text{ is unknown,}$$

where, $S.E(\hat{p}) = \frac{\sigma}{\sqrt{\sum_{i=1}^m X_i}}$, and $s.e(\hat{p}) = \frac{s_{\tilde{e}}}{\sqrt{\sum_{i=1}^m X_i}}$, $s_{\tilde{e}}^2 = \frac{1}{m-1}\sum_{i=1}^m (\widetilde{Y}_i - \hat{p}\widetilde{X}_i)^2$, $\widetilde{Y}_i = \frac{Y_i}{\sqrt{X_i}}$, $\widetilde{X}_i = \sqrt{X_i}$, as well as $\overline{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$.

1.3 Poisson model

This section, deals with the next assumed model in which the data analysis based on, with a new assumption on the variable X_i, namely, the variable X_i assumes as a Poisson random variable, $X_i \geq 0$, (Note that, we are not interest in a relationship between the observations Y_i and X_i, only in the conditional distribution of Y_i given X_i).

Let us describe the so-called 'Univariate Poisson model', and then we will obtain the conditional distribution of the r.v Y_i given the Poisson random variable X_i, as well as, we will construct the approximate confidence intervals for a function of the Poisson model parameters.

1.3.1 The Univariate Poisson model

Assume that the count r.v X_i has a Poisson distribution, which decomposes additively into two independent Poisson random variables Y_i, Z_i with means, λ_1, λ_2 respectively. I.e, $X_i = Y_i + Z_i$, it follows that $X_i \sim Poiss(\lambda_1 + \lambda_2)$, and hence, $E(Y_i) = Var(Y_i) = \lambda_1$, and $E(Z_i) = Var(Z_i) = \lambda_2$, consequently, $E(X_i) = Var(X_i) = \lambda_1 + \lambda_2$.
Further, one may obtain the conditional distribution of a sub count random variable Y_i given the count r.v X_i, which may be summarizing as:

$P(Y_i \mid X_i) \sim Bin(X_i, \frac{\lambda_1}{\lambda_1+\lambda_2})$, with the Binomial proportion $p = g(\lambda_1, \lambda_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$, $X_i > 0$ (where, $Bin(n,p)$ denotes the Binomial distribution with sample size n and success probability p, in case, $X_i \equiv 0 \Rightarrow Y_i \equiv 0$).
The Maximum likelihood estimators $\hat{\lambda}_{1,ML}, \hat{\lambda}_{2,ML}$ of the means λ_1, λ_2 are $\overline{Y}, \overline{Z}$ respectively. The Maximum likelihood estimator
$\widehat{\lambda_1 + \lambda_2} = \hat{\lambda}_{1,ML} + \hat{\lambda}_{2,ML} = \overline{Y} + \overline{Z} = \overline{X}$, consequently

$$\hat{p}_{ML} = g(\widehat{\lambda_1, \lambda_2}) = g(\hat{\lambda}_{1,ML}, \hat{\lambda}_{2,ML}) = \frac{\hat{\lambda}_{1,ML}}{\hat{\lambda}_{1,ML} + \hat{\lambda}_{2,ML}} = \frac{\overline{Y}}{\overline{Y} + \overline{Z}} = \frac{\overline{Y}}{\overline{X}},$$

where, $\hat{p}_{ML}$ denotes the Maximum likelihood estimator of the conditional Binomial proportion p.

1.3.2 Approximate confidence intervals for the conditional Binomial proportion p

We will try to identify the asymptotic distribution for the distribution of the estimator $\hat{p}_m \mid X_1, ..., X_m$, where $X_1, ..., X_m$ are $i.i.d$ Poisson r.v's.

In the univariate Poisson model, obtaining the conditional distributions is usually possible and is fairly straightforward, however in a higher dimension can not be specify explicitly, due to the dependence of the components, i.e, correlations between them. We summarize the following steps:
Since, we have

$$P^{Y_i|X_i} \sim Bin\left(X_i, p\right), \ X_i > 0,$$

subsequently

$$P^{\sum_{i=1}^{m} Y_i|X_1,...,X_m} \sim Bin\left(\sum_{i=1}^{m} X_i, p\right),$$

(given that (Y_i, X_i) are independent), it implies that

$$P^{\hat{p}_m|X_1,\ldots,X_m} \sim \frac{Bin\left(N_m, p\right)}{N_m},$$

where, $N_m = \sum_{i=1}^{m} X_i$ (random sample size), which diverges as sample size 'm' becomes large.
To demonstrate this formally, since (by the law of large numbers)

$$P\left(\left|\frac{\sum_{i=1}^{m} X_i}{mE(X_i)} - 1\right| > \epsilon\right) \longrightarrow^{P} 0, \text{ for each, } \epsilon > 0,$$

or one might rewrite it as, $P\left(\frac{\sum_{i=1}^{m} X_i}{mE(X_i)} < 1 - \epsilon\right) \longrightarrow^{P} 0$, it follows that

$$P\left(\sum_{i=1}^{m} X_i < (1-\epsilon).mE(X_i)\right) = 0,$$

or

$$P\left(\sum_{i=1}^{m} X_i < c.m, \text{ infinitely many times}\right) = 0,$$

where the constant, $c = (1-\epsilon)E(X_i)$, it follows that, $P\left(\sum_{i=1}^{m} X_i \longrightarrow \infty\right) = 1$, i.e, $N_m \longrightarrow^{a.s} \infty$.

On the one hand, and by applying the Central limit theorem to the conditional Binomial, we will obtain the following:

$$P^{\frac{\sqrt{N_m}(\hat{p}_m - p)}{\sqrt{\hat{p}_m(1-\hat{p}_m)}}|N_m} = P^{\frac{\sqrt{N_m}\left(\frac{W_{N_m}}{N_m} - p\right)}{\sqrt{\hat{p}_m(1-\hat{p}_m)}}|N_m} = P^{\frac{W_{N_m} - N_m p}{\sqrt{N_m \hat{p}_m(1-\hat{p}_m)}}|N_m} \longrightarrow^{\mathcal{D}} N(0,1), \tag{1.13}$$

[where, $W_{N_m} = \sum_{i=1}^{m} Y_i = \sum_{i=1}^{m} \widetilde{Y}_{X_i} = \sum_{i=1}^{m} \sum_{j=1}^{X_i} \widetilde{\widetilde{Y}}_{ij} = \sum_{i=1}^{N_m} \widetilde{\widetilde{\widetilde{Y}}}_{ij}$, so, $\widetilde{Y}_{X_i} = \sum_{j=1}^{X_i} \widetilde{\widetilde{Y}}_{ij}$, $P^{Y_i|X_i} \equiv P^{Y_{X_i}|X_i} \sim Bin(X_i, p)$, $P^{\widetilde{\widetilde{Y}}_{ij}|X_{ij}} \sim Bern(1,p)$, as well as $\widetilde{\widetilde{\widetilde{Y}}}_{ij}$ are *iid* $Bern(1,p)$, (where, $Bern(1,p)$ denotes the Bernoulli distribution with probability of success p, i.e Binomial distribution with one sample observation or outcome (success or failure) with probability p of success), $i = 1, ..., m,\ j = 1, ..., X_i$].
I.e, the asymptotic distribution of the conditional distribution

$$P^{\frac{\sqrt{N_m}(\hat{p}_m - p)}{\sqrt{\hat{p}_m(1-\hat{p}_m)}}|N_m} = P^{\frac{W_{N_m} - N_m p}{\sqrt{N_m \hat{p}_m(1-\hat{p}_m)}}|N_m} \longrightarrow^{\mathcal{D}} N(0,1),$$

in words, the asymptotic normality of the conditional Binomial distribution, $P^{\hat{p}_m|X_1,\ldots,X_m} \sim \frac{Bin(N_m,p)}{N_m}$, conditionally holds.

On the other hand, deriving the asymptotic normality of the unconditional distribution of random variables, by the CLT with the random summation index have been proved by Landers and Rogge (1976), see [15], 269-271.
According to this paper one might restate the following:

$$P\left(\frac{W_{N_m} - N_m p}{\sqrt{N_m \hat{p}(1-\hat{p})}} \le x\right) \longrightarrow \Phi(x)$$
$$\Leftarrow P\left(\left|\frac{N_m p(1-p)}{mE(X_i)p(1-p)} - 1\right| > \epsilon\right) \longrightarrow^{\mathcal{P}} 0,$$

where

$$P\left(\left|\frac{N_m p(1-p)}{mE(X_i)p(1-p)} - 1\right| > \epsilon\right) \longrightarrow^{\mathcal{P}} 0, \text{ for each } \epsilon > 0. \tag{1.14}$$

Expression 1.14 will proves the consequence $\frac{N_m p(1-p)}{m} \longrightarrow^{\mathcal{P}} E(X_i)p(1-p)$, which is the LLN, and itself can be proved by Chebyshev's or Markov's inequality (see[14], pp. 123-125), as

$$P\left(\left|\frac{N_m}{m} - E(X_i)\right| > \epsilon\right) \le \frac{m(\lambda_1+\lambda_2)}{m^2\epsilon^2} \longrightarrow 0, \quad as,\ m \longrightarrow \infty.$$

The conditional expectation and the conditional variance respectively of the conditional distribution given by expression 1.13 are obtained by

$$E\left(\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right) \mid N_m\right) = E\left(\sqrt{N_m}\left(\frac{W_{N_m}}{N_m} - p\right) \mid N_m\right) = \sqrt{N_m}\left(\frac{N_m p}{N_m} - p\right)$$
$$= 0,$$
$$Var\left(\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right) \mid N_m\right) = Var\left(\sqrt{N_m}\left(\frac{W_{N_m}}{N_m} - p\right) \mid N_m\right) = \frac{N_m N_m p(1-p)}{N_m^2}$$
$$= p(1-p),$$

which in contrast, the obtained conditional variance is equal to the unconditional variance (the asymptotic variance)

$$Var\left(\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right)\right) = E\left(Var\left(\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right) \mid N_m\right)\right) +$$
$$Var\left(E\left(\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right) \mid N_m\right)\right) = E\left(p(1-p)\right) + 0$$
$$= p(1-p).$$

So, the asymptotic Normal distribution of the conditional Binomial distribution of the r.v, $\frac{W_{N_m}}{N_m} \mid N_m$ is given by

$$P^{\sqrt{N_m}\left(\frac{\overline{Y}_m}{\overline{X}_m}-p\right)|N_m} = P^{\sqrt{N_m}\left(\frac{W_{N_m}}{N_m}-p\right)|N_m} \longrightarrow^{\mathcal{D}} N\left(0, p(1-p)\right) \tag{1.15}$$

provided that, the random sample size $N_m = \sum_{i=1}^{m} X_i \longrightarrow \infty$, as $m \longrightarrow \infty$, as well as

$$P^{\frac{W_{N_m}}{N_m}|N_m} \sim \frac{Bin\left(N_m, p\right)}{N_m}.$$

It follows from (1.15), that the conditional distribution of $\sqrt{N_m}\left(\frac{W_{N_m}}{N_m} - p\right) \mid N_m$ has an asymptotic $N\left(0, p(1-p)\right)$, or in other words, one can say:

In large sample size N_m, the conditional distribution of the r.v $\frac{W_{N_m}}{N_m} \mid N_m$ is approximately Normal distribution with mean is the proportion p and with variance $\frac{p(1-p)}{N_m}$, i.e $\approx N\left(p, \frac{p(1-p)}{N_m}\right)$.

Therefore, the estimated asymptotic confidence interval for the proportion $p = g(\lambda_1, \lambda_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$ in this situation will be obtained by

$$\left[\hat{p}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{p}_m)\right],$$

where, the standard error $s.e(\hat{p}_m) = \sqrt{\frac{\hat{p}_m(1-\hat{p}_m)}{N_m}}$.

Further and according to 1.20, the suggested confidence interval (conservative) for the proportion p can be obtained by

$$\left[\hat{p}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p}_m)\right].$$

1.3.3 Approximate confidence intervals for the proportion p by the Delta method

At the end of this chapter and looking from other angel, one attempt to obtain approximate confidence intervals for the proportion p, when the exact distributions of either X_i nor Y_i are not necessary known.

The (Delta or δ) method help to obtain the asymptotic distribution for any non-linear transformation of the pair random variables (Y_i, X_i) regardless the exact distribution of the r.v's.

Let the pairs $\mathbf{H}_i = (Y_i, X_i)^T, i = 1, ..., m$ are $i.i.d$ pairs of random variables,

and, $\mathbf{C} = as\ Cov(\mathbf{H}_i) = \begin{pmatrix} \lambda_1 & \lambda_1 \\ \lambda_1 & \lambda_1 + \lambda_2 \end{pmatrix}$.
According to the multivariate central limit theorem we have

$$\sqrt{m}\left(\overline{\mathbf{H}} - E(\mathbf{H})\right) \longrightarrow^{\mathcal{D}} N_2\left(\mathbf{0}, \mathbf{C}\right),$$

or, decomposes to

$$\sqrt{m}\left(\begin{pmatrix} \overline{Y}_m \\ \overline{X}_m \end{pmatrix} - \begin{pmatrix} \lambda_1 \\ \lambda_1 + \lambda_2 \end{pmatrix}\right) \longrightarrow^{\mathcal{D}} N_2\left(\mathbf{0}, \mathbf{C}\right),$$

The Delta method

Suppose that, $\{\hat{\boldsymbol{\theta}}_n\}$ is a sequence of random vectors with $\hat{\boldsymbol{\theta}}_n \longrightarrow^{\mathcal{P}} \boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a vector of parameters. Let $\{a_n\}$ be a sequence of constants, $a_n \uparrow \infty$.
If, $a_n\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \longrightarrow^{\mathcal{D}} \mathbf{Z} \sim N_p\left(\mathbf{0}, \mathbf{C}\right)$, where $\mathbf{C}$ is a $p \times p$ covariance matrix, and let $g(\boldsymbol{\theta}) : R^p \to R^k$ be a real valued function that is continuously differentiable at vector $\boldsymbol{\theta} \in R^p$. The matrix of partial derivatives of the function g with respect to the vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^T$ is given by

$$\nabla_g(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial g_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{bmatrix},$$

or, as $\boldsymbol{\theta}^T = (\theta_1, ..., \theta_p)$, we have

$$\nabla_g^T(\boldsymbol{\theta}) = \left(\frac{\partial g_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)_{i=1,\ldots,k} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_k(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_k(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix},$$

i.e,

$$\nabla_g^T(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix},$$

as, $g_1 = g$ (univariate delta method).

In general, for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k)^T$, where $\boldsymbol{\theta}_j$ is $1 \times p$ vector of constants, $g(\boldsymbol{\theta}) : R^{pk} \to R^k$, we have

$$\nabla_g(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial g_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{bmatrix},$$

or

$$\nabla_g^T(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \frac{\partial g_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} & \cdots & \frac{\partial g_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \end{bmatrix}_{k \times pk}, \tag{1.16}$$

where, $\frac{\partial g_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}$ is the vector of partial derivatives of g_i with respect to the elements of $\boldsymbol{\theta}_j, i, j = 1, ..., k$.

Then

$$a_n \left(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}) \right) \longrightarrow^{\mathcal{D}} \nabla_g(\boldsymbol{\theta})\mathbf{Z} \sim N_k \left(\mathbf{0}, \nabla_g^T(\boldsymbol{\theta})\mathbf{C}\nabla_g(\boldsymbol{\theta}) \right). \tag{1.17}$$

Proof:

see [14], pp. 120-121, and pp. 148-149.

The Univariate Delta method

Define the following notations

$$\boldsymbol{\theta} = E(\mathbf{H}_i) = (\theta_1, \theta_2)^T, \quad \theta_1 = E(Y_i),\ \theta_2 = E(X_i)$$
$$g(\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2}, \quad g(\boldsymbol{\theta}) : \mathbf{R}^2 \longrightarrow \mathbf{R}, \text{ is continuously differentiable at } \boldsymbol{\theta},\ \theta_2 > 0.$$

The vector of partial derivatives of the continuous differentiable function g with respect to the components of $\boldsymbol{\theta}$, i.e

$$\nabla_g^T(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial g}{\partial \theta_1} & \frac{\partial g}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\theta_2} & -\frac{\theta_1}{\theta_2{}^2} \end{pmatrix}.$$

Also by plugging in the elements of $\boldsymbol{\theta}$ to the matrix $\mathbf{C}$ we get $\mathbf{C} = \mathbf{C}_\theta = \begin{pmatrix} \theta_1 & \theta_1 \\ \theta_1 & \theta_2 \end{pmatrix}$,

For, $p = \frac{E(Y_i)}{E(X_i)}$, and according to the delta method with its notation in univariate case, we have

$$g(\boldsymbol{\theta}) = p, \quad \text{and } \hat{p}_m = g(\hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}} = \overline{\mathbf{H}}_m = (\overline{Y}_m, \overline{X}_m)^T$ is a consistent estimator of $\boldsymbol{\theta}$, as well as $\hat{p}_m = \frac{\overline{Y}_m}{\overline{X}_m}$. Hence, it follows that

$$\sqrt{m}(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \nabla_g^T(\boldsymbol{\theta})\mathbf{C}_\theta \nabla_g(\boldsymbol{\theta})\right), \tag{1.18}$$

where, the variance in the equation 1.18 simplifies to

$$\begin{aligned}\nabla_g^T(\boldsymbol{\theta})\mathbf{C}_\theta \nabla_g(\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2^2} - \frac{2\theta_1^2}{\theta_2^3} + \frac{\theta_1^2}{\theta_2^3} = \frac{\theta_1}{\theta_2^2} - \frac{\theta_1^2}{\theta_2^3} = \frac{\theta_1}{\theta_2^2}\left(1 - \frac{\theta_1}{\theta_2}\right) &= \frac{1}{E(X_i)}p(1-p) \\ &= \frac{p(1-p)}{\lambda_1 + \lambda_2},\end{aligned}$$

and, consequently

$$\sqrt{m}\,(\hat{p}_m - p) \longrightarrow^{\mathcal{D}} N\left(0, \frac{p(1-p)}{\lambda_1 + \lambda_2}\right),$$

which results in the asymptotic Normal distribution of the ratio estimator.

On the same context, it is also possible to derive this consequence directly with helps of the Slutsky's Theorem.

Since

$$\begin{aligned}\sqrt{m}\,(\hat{p}_m - p) &= \sqrt{m}\left(\frac{\overline{Y}_m}{\overline{X}_m} - p\right) = \sqrt{m}\frac{1}{\overline{X}_m}\left(\overline{Y}_m - p\overline{X}_m\right) \\ &= \sqrt{m}\frac{1}{\overline{X}_m}\frac{1}{m}\sum_{i=1}^{m}(Y_i - pX_i) = \sqrt{m}\frac{1}{\overline{X}_m}\frac{1}{m}\sum_{i=1}^{m}\left((1-p)Y_i - pZ_i\right), \end{aligned} \tag{1.19}$$

let, $G_i = (1-p)Y_i - pZ_i$, where, G_i are *iid* r.v's, with mean

$$\begin{aligned}E(G_i) &= (1-p)\lambda_1 - p\lambda_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2}\lambda_1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\lambda_2 = 0, \text{ and variance} \\ Var(G_i) &= \frac{\lambda_2^2}{(\lambda_1 + \lambda_2)^2}\lambda_1 + \frac{\lambda_1^2}{(\lambda_1 + \lambda_2)^2}\lambda_2 = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} = p(1-p)(\lambda_1 + \lambda_2).\end{aligned}$$

From equation 1.19 we have

$$\sqrt{m}\frac{1}{m}\sum_{i=1}^{m}\left((1-p)Y_i - pZ_i\right) \longrightarrow^{\mathcal{D}} N\left(0, p(1-p)(\lambda_1 + \lambda_2)\right),$$

it follows (by Slutsky's lemma, since $\overline{X}_m \longrightarrow^{\mathcal{P}} \lambda_1 + \lambda_2$) that

$$\sqrt{m}\frac{1}{\overline{X}_m}\frac{1}{m}\sum_{i=1}^{m}\left((1-p)Y_i - pZ_i\right) \longrightarrow^{\mathcal{D}} N\left(0, \frac{p(1-p)}{\lambda_1 + \lambda_2}\right).$$

And hence, $\sqrt{m}(\hat{p}_m - p)$ has an asymptotic Normal distribution with mean zero and asymptotic variance $asVar(\hat{p}_m) = \frac{p(1-p)}{E(X_i)}$, whence, $E(X_i) = \lambda_1 + \lambda_2$. Or in other words (in terms of $\hat{p}_m$), one can say that:

In large sample size m, the approximated Normal distribution for the estimator $\hat{p}_m$ is, $N\left(p, \frac{p(1-p)}{mE(X_i)}\right)$. The consistent variance estimator is obtained by plugging in the estimate for the corresponding individual parameters in the approximated variance, i.e $\widehat{Var(\hat{p}_m)} = \frac{\hat{p}_m(1-\hat{p}_m)}{m\overline{X}_m} = \frac{\hat{p}_m(1-\hat{p}_m)}{\sum_{i=1}^m X_i}$, and hence, the standard error $s.e(\hat{p}_m) = \sqrt{\frac{\hat{p}_m(1-\hat{p}_m)}{\sum_{i=1}^m X_i}}$. Therefore and on one hand, the asymptotic confidence interval for the proportion p is given by

$$\left[\hat{p}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{p}_m)\right],$$

which is the previous confidence interval for the proportion p of the conditional Binomial distribution. As well as, it follows that, $\hat{p}_m = \frac{\overline{Y}_m}{\overline{X}_m} = \hat{p}_{ML} = \hat{p}$ (*the ratio estimator*).

On the other hand, it may be possible to write the following expression

$$\frac{\sqrt{m}\,(\hat{p}_m - p)}{\sqrt{\frac{\hat{p}_m(1-\hat{p}_m)}{\sum_{i=1}^m X_i}}} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1), \text{ as}, m \longrightarrow \infty, \tag{1.20}$$

which suggest the conservative confidence interval (safety bounds) for the proportion p which is given by

$$\left[\hat{p}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{p}_m)\right],$$

where, $t_{(m-1,1-\frac{\alpha}{2})}$ is the $(1-\frac{\alpha}{2})$ quantile of a $t-$distribution with $(m-1)$ degrees of freedom.

Chapter 2

Bivariate data analysis

2.1 Introduction

In this chapter, it will be extended the univariate count data to the bivariate setting, where the bivariate analysis involves data in two dimensional setup.

Consider, there are two relevant components of count data sampling from an infinite population, taking in account any vertical dependencies between the components, namely the $i.i.d$ two dimensional pairs $((Y_{11}, X_{11}), (Y_{12}, X_{12})), ..., ((Y_{m1}, X_{m1}), (Y_{m2}, X_{m2}))$ is a random sample of size m drawn from the same population, such that each observational unit indexed by the subscript i associated with those random variables restricted by $0 \leq Y_{ij} \leq X_{ij}$, $\forall i, j$.

For illustration, the i^{th} individual represents count variables for example:

$X_{i1}, X_{i2} \equiv$ No.of children and No.of dogs respectively in the family i.

$Y_{i1}, Y_{i2} \equiv$ No.of male children and white dogs respectively in the same family.

Also denote that, $\hat{p}_1 = \frac{\sum_{i=1}^{m} Y_{i1}}{\sum_{i=1}^{m} X_{i1}}$, $\hat{p}_2 = \frac{\sum_{i=1}^{m} Y_{i2}}{\sum_{i=1}^{m} X_{i2}}$ as the sample proportions corresponding to the population proportions p_1, p_2, where $p_j = \frac{E(Y_{ij})}{E(X_{ij})}$, $E(\hat{p}_j) = p_j$, $0 \leq p_j \leq 1$, $j = 1, 2$.

In the next section, it will be used the fundamental approach (the SUR model) for analyzing the collected two dimensional count data. Hence, it will be assumed throughout the first section that the relationship between the pairs of random variables Y_{i1}, Y_{i2}, and the corresponding pairs X_{i1}, X_{i2} (the dependence between the pairs (Y_{i1}, X_{i1}) and (Y_{i2}, X_{i2}) is crucial) are linearly modeled and then we will use the linear regression technique to analyze the collected sample points bases on the SUR model.

2.2 Bivariate SUR (Seemingly Unrelated Regression) Model

In the two dimensional case, the bivariate SUR model (this model was introduced by Zellner (1962), see also [21], or [25] for more details) based on m observations can be modeled in the next steps:

2.2.1 The bivariate linear model

The pairs of the univariate linear models will be considered first, as

$$Y_{i1} = x_{i1}p_1 + \epsilon_{i1}$$
$$Y_{i2} = x_{i2}p_2 + \epsilon_{i2}$$

with the following assumptions:
$E(\epsilon_{i1}) = 0,\ E(\epsilon_{i2}) = 0$ and with the variances proportional to x_{i1}, x_{i2} respectively (x_{i1}, x_{i2} are fixed variables), i.e
$Var(\epsilon_{i1}) = Var(Y_{i1}) = \sigma_1^2 x_{i1},\ Var(\epsilon_{i2}) = Var(Y_{i2}) = \sigma_2^2 x_{i2},\ Cov(\epsilon_{i1}, \epsilon_{i2}) = \sigma_{12}\sqrt{x_{i1}x_{i2}}$, and $Cov(\epsilon_{ij}, \epsilon_{i'j}) = 0,\ \forall i \neq i',\ i, i' = 1, ..., m, j = 1, 2.$
We merge these equations into a single bivariate model (for the i^{th}observation)

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{p} + \boldsymbol{\epsilon}_i,\ (i = 1, ..., m), \tag{2.1}$$

where, the response variable $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$, the design matrix $\mathbf{X}_i = \begin{pmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{pmatrix}$, and the model coefficient $\mathbf{p} = (p_1, p_2)^T$, as well as, the error component $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2})^T$, $i = 1, ..., m$.
The error component $\boldsymbol{\epsilon}_i$ has the variance-covariance matrix given by

$$\Sigma_i = Cov(\boldsymbol{\epsilon}_i) = \begin{pmatrix} \sigma_1^2 x_{i1} & \sigma_{12}\sqrt{x_{i1}x_{i2}} \\ \sigma_{12}\sqrt{x_{i1}x_{i2}} & \sigma_2^2 x_{i2} \end{pmatrix},$$

$Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_{i'}) = \mathbf{0}_{2\times 2}$, when $i \neq i' = 1, ..., m.$

2.2.2 The weighted bivariate linear model

In the same manner as in the univariate linear model, the model 2.1 (we assume that $x_{ij} > 0\, \forall i, j$) will be standardized by the linear transformation

$$\mathbf{A}_i\mathbf{Y}_i = \mathbf{A}_i\mathbf{X}_i\mathbf{p} + \mathbf{A}_i\boldsymbol{\epsilon}_i,$$

where, the diagonal transformation matrix $\mathbf{A}_i = \mathbf{X}_i^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{x_{i1}}} & 0 \\ 0 & \frac{1}{\sqrt{x_{i2}}} \end{pmatrix}$, such that $\mathbf{A}_i\mathbf{X}_i\mathbf{A}_i^T = I_2$. Hence, the weighted linear model becomes

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{X}}_i\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}_i, \tag{2.2}$$

where, the weighed response variable $\widetilde{\mathbf{Y}}_i = \mathbf{A}_i\mathbf{Y}_i = \left(\widetilde{Y}_{i1}, \widetilde{Y}_{i2}\right)^T$, the weighted error component $\widetilde{\boldsymbol{\epsilon}}_i = \mathbf{A}_i\boldsymbol{\epsilon}_i = \left(\ \widetilde{\epsilon}_{i1}, \widetilde{\epsilon}_{i2}\ \right)^T$, and the weighted design matrix $\widetilde{\mathbf{X}}_i = \mathbf{A}_i\mathbf{X}_i = \begin{pmatrix} \sqrt{x_{i1}} & 0 \\ 0 & \sqrt{x_{i2}} \end{pmatrix} = \begin{pmatrix} \widetilde{x}_{i1} & 0 \\ 0 & \widetilde{x}_{i2} \end{pmatrix}$, $\widetilde{x}_{ij} = \sqrt{x_{ij}}$, and $\widetilde{Y}_{ij} = \frac{Y_{ij}}{\sqrt{x_{ij}}}$, $\widetilde{\epsilon}_{ij} = \frac{\epsilon_{ij}}{\sqrt{x_{ij}}}$, given that $x_{ij} > 0\ \forall i, j,\ i = 1, ..., m, j = 1, 2$.
The covariance of the weighted error vector is given by

$$\widetilde{\Sigma} = Cov(\widetilde{\boldsymbol{\epsilon}}_i) = Cov(\widetilde{\mathbf{Y}}_i) = \mathbf{A}_iCov(\boldsymbol{\epsilon}_i)\mathbf{A}_i^T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\ (homoscedastic\ error\ vectors),$$

and, $Cov(\widetilde{\boldsymbol{\epsilon}}_i, \widetilde{\boldsymbol{\epsilon}}_{i'}) = \mathbf{0}_{2\times 2},\ i \neq i' = 1, ..., m$, where $\sigma_j^2 = E\left(\widetilde{Y}_{ij} - \widetilde{x}_{ij}p_j\right)^2$, $\sigma_{12} = E\left((\widetilde{Y}_{i1} - \widetilde{x}_{i1}p_1)(\widetilde{Y}_{i2} - \widetilde{x}_{i2}p_2)\right), j = 1, 2$.

2.2.3 The SUR Model

The bivariate SUR model can be established (via stacking in column wise of the bivariate equations 2.1 in to a single model) as:

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} x_{11} & 0 \\ 0 & x_{12} \end{pmatrix} \\ \begin{pmatrix} x_{21} & 0 \\ 0 & x_{22} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} x_{m1} & 0 \\ 0 & x_{m2} \end{pmatrix} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{pmatrix},$$

or, it would be convenient if we rewrite the model compactly in the vector form

$$\mathbf{Y} = \mathbf{X}\mathbf{p} + \boldsymbol{\epsilon}, \tag{2.3}$$

where, the $2m \times 1$ dimension response vector $\mathbf{Y} = \left(\mathbf{Y}_1^T, \mathbf{Y}_2^T, \cdots, \mathbf{Y}_m^T\right)^T$, the $2m \times 2$ dimension design matrix $\mathbf{X} = \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \cdots, \mathbf{X}_m^T\right)^T$, so, 2×1 dimension SUR model parameter $\mathbf{p} = (p_1, p_2)^T$, and the, $2m \times 1$ dimension SUR model error vector $\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}_1^T, \boldsymbol{\epsilon}_2^T, \cdots, \boldsymbol{\epsilon}_m^T\right)^T$, as well as, the $2m \times 2m$ dimension covariance structure of the SUR model error vector is given by

$$\boldsymbol{\Sigma} = Cov(\boldsymbol{\epsilon}) = \begin{pmatrix} \Sigma_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma_m \end{pmatrix}.$$

Similarly, the weighted bivariate SUR model (or simply the bivariate SUR model)

$$\begin{pmatrix} \widetilde{\mathbf{Y}}_1 \\ \widetilde{\mathbf{Y}}_2 \\ \vdots \\ \widetilde{\mathbf{Y}}_m \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \sqrt{x_{11}} & 0 \\ 0 & \sqrt{x_{12}} \end{pmatrix} \\ \begin{pmatrix} \sqrt{x_{21}} & 0 \\ 0 & \sqrt{x_{22}} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \sqrt{x_{m1}} & 0 \\ 0 & \sqrt{x_{m2}} \end{pmatrix} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} \widetilde{\boldsymbol{\epsilon}}_1 \\ \widetilde{\boldsymbol{\epsilon}}_2 \\ \vdots \\ \widetilde{\boldsymbol{\epsilon}}_m \end{pmatrix},$$

the model compresses in the compact form

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}, \tag{2.4}$$

where, the $2m \times 1$ dimension weighted response vector $\widetilde{\mathbf{Y}} = \left(\widetilde{\mathbf{Y}}_1^T, \widetilde{\mathbf{Y}}_2^T, \cdots, \widetilde{\mathbf{Y}}_m^T\right)^T$, the $2m \times 2$ dimension weighted design matrix, $\widetilde{\mathbf{X}} = \left(\widetilde{\mathbf{X}}_1^T, \widetilde{\mathbf{X}}_2^T,, \cdots, \widetilde{\mathbf{X}}_m^T\right)^T$, the 2×1 dimension SUR model parameter vector $\mathbf{p} = (p_1, p_2)^T$, and the $2m \times 1$ dimension SUR model error vector $\widetilde{\boldsymbol{\epsilon}} = \left(\widetilde{\boldsymbol{\epsilon}}_1^T, \widetilde{\boldsymbol{\epsilon}}_2^T, \cdots, \widetilde{\boldsymbol{\epsilon}}_m^T\right)^T$, with the $2m \times 2m$ covariance matrix of the weighted error vector

$$\widetilde{\boldsymbol{\Sigma}} = Cov(\widetilde{\boldsymbol{\epsilon}}) = \begin{pmatrix} \widetilde{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \widetilde{\Sigma} \end{pmatrix} = I_m \otimes \widetilde{\Sigma},$$

i.e, $Var(\widetilde{\boldsymbol{\epsilon}}_i) = \widetilde{\Sigma}_{2\times2}$, and $Cov(\widetilde{\boldsymbol{\epsilon}}_i, \widetilde{\boldsymbol{\epsilon}}_i{}') = \mathbf{0}_{2\times2}$ $(\forall i \neq i',\ i, i' = 1, ..., m)$.

2.2.4 Estimation in the SUR models

Under the assumption of normality of the linear model errors, the least squares estimators of the parameters in $\mathbf{p}$ are in fact the corresponding MLE's, i.e, the least squares estimators coincide with maximum likelihood estimators.

Further, it well-known that the WLSE (weighted least squires estimator) is the BLUE (optimal) of the parameter $\mathbf{p}$, however, the WLSE of the SUR model parameter vector results in not the sample ratio estimator vector in question (which is the OLSE, and the equality WLSE $\equiv$ OLSE holds if $\sigma_{12} = 0$, i.e, the error covariance matrices are diagonal).

The OLSE will be used, although it is not the optimal estimator nevertheless produces the ratio estimators.

So, the required estimator of the model parameter (proportion $\mathbf{p}$) will now obtained from the SUR model 2.4

$$\hat{\mathbf{p}}_{OLS} = (\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}} = \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i = \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^2\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i$$

$$= \begin{pmatrix} \sum_{i=1}^m x_{i1} & 0 \\ 0 & \sum_{i=1}^m x_{i2} \end{pmatrix}^{-1} \sum_{i=1}^{m} \begin{pmatrix} \sqrt{x_{i1}}\widetilde{Y}_{i1} \\ \sqrt{x_{i2}}\widetilde{Y}_{i2} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^m \sqrt{x_{i1}}\widetilde{Y}_{i1}}{\sum_{i=1}^m x_{i1}} \\ \frac{\sum_{i=1}^m \sqrt{x_{i2}}\widetilde{Y}_{i2}}{\sum_{i=1}^m x_{i2}} \end{pmatrix} = \hat{\mathbf{p}},$$

which results in the ratio estimator vector, where $\sqrt{x_{i1}}\widetilde{Y}_{ij} = Y_{ij},\ j = 1, 2.$

2.2.5 Properties of the estimator $\hat{\mathbf{p}}_{OLS}$ (ratio estimator vector $\hat{\mathbf{p}}$)

Unbiasedness

As in the univariate case, the expectation of the ratio estimator $\hat{p}_j$ is given by

$$E\left(\hat{p}_j\right) = E\left(\frac{\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}}\right) = \frac{\sum_{i=1}^m \sqrt{x_{ij}}E(\widetilde{Y}_{ij})}{\sum_{i=1}^m x_{ij}} = \frac{\sum_{i=1}^m \sqrt{x_{ij}}p_j\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}} = \frac{p_j\sum_{i=1}^m x_{ij}}{\sum_{i=1}^m x_{ij}}$$

$$= p_j,$$

$j = 1, 2.$

Dispersion

The variance of the sample ratio $\hat{p}_j$ can be obtained by

$$Var\left(\hat{p}_j\right) = Var\left(\frac{\sum_{i=1}^{m}\widetilde{Y}_{ij}\sqrt{x_{ij}}}{\sum_{i=1}^{m}x_{ij}}\right)$$
$$= \frac{1}{\left(\sum_{i=1}^{m}x_{ij}\right)^2}\sum_{i=1}^{m}x_{ij}Var(\widetilde{Y}_{ij}) = \frac{\sigma_j^2\sum_{i=1}^{m}x_{ij}}{\left(\sum_{i=1}^{m}x_{ij}\right)^2} = \frac{\sigma_j^2}{\sum_{i=1}^{m}x_{ij}},$$

see also [19] pp. 129.

Furthermore, the covariance between the estimators $\hat{p}_1, \hat{p}_2$ is obtained by

$$Cov\left(\hat{p}_1, \hat{p}_2\right) = Cov\left(\frac{\sum_{i=1}^{m}\widetilde{Y}_{i1}\sqrt{x_{i1}}}{\sum_{i=1}^{m}x_{i1}}, \frac{\sum_{i=1}^{m}\widetilde{Y}_{i2}\sqrt{x_{i2}}}{\sum_{i=1}^{m}x_{i2}}\right)$$
$$= \frac{1}{\sum_{i=1}^{m}x_{i1}\sum_{i=1}^{m}x_{i2}}Cov\left(\sum_{i=1}^{m}\sqrt{x_{i1}}\widetilde{Y}_{i1}, \sum_{i=1}^{m}\sqrt{x_{i2}}\widetilde{Y}_{i2}\right)$$
$$= \frac{1}{\sum_{i=1}^{m}x_{i1}\sum_{i=1}^{m}x_{i2}}\sum_{i=1}^{m}Cov\left(\widetilde{Y}_{i1}, \widetilde{Y}_{i2}\right)\sqrt{x_{i1}x_{i2}} = \frac{\sigma_{12}\sum_{i=1}^{m}\sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^{m}x_{i1}\sum_{i=1}^{m}x_{i2}}.$$

The covariance matrix of the ratio vector $\hat{\mathbf{p}}$ is obtained as

$$\Sigma_{\mathbf{x}} = Cov\left(\hat{\mathbf{p}}\right) = Cov\left(\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}}\right) = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{X}}\right)\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}$$
$$= \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i\left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}$$
$$= \begin{pmatrix}\sum_{i=1}^{m}x_{i1} & 0\\ 0 & \sum_{i=1}^{m}x_{i2}\end{pmatrix}^{-1}\begin{pmatrix}\sigma_1^2\sum_{i=1}^{m}x_{i1} & \sigma_{12}\sum_{i=1}^{m}\sqrt{x_{i1}x_{i2}}\\ \cdots & \sigma_2^2\sum_{i=1}^{m}x_{i2}\end{pmatrix}$$
$$\begin{pmatrix}\sum_{i=1}^{m}x_{i1} & 0\\ 0 & \sum_{i=1}^{m}x_{i2}\end{pmatrix}^{-1}$$
$$= \begin{pmatrix}\frac{\sigma_1^2}{\sum_{i=1}^{m}x_{i1}} & \frac{\sigma_{12}\sum_{i=1}^{m}\sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^{m}x_{i1}\sum_{i=1}^{m}x_{i2}}\\ \cdots & \frac{\sigma_2^2}{\sum_{i=1}^{m}x_{i2}}\end{pmatrix}. \tag{2.5}$$

Note that, the last covariance matrix is larger than the covariance matrix $\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{X}}\right)^{-1}$, when we use the weighted least squares estimator $\hat{\mathbf{p}}_{WLS} = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{Y}}$ rather than 'OLSE' based on the SUR model, where $\widetilde{\mathbf{\Sigma}} = I_m \otimes \widetilde{\Sigma}$, (more explanation with a simple example in the univariate linear model with intercept is given in [19] pp. 151-153.

Exact and Asymptotic distributions

If one assumed that $\boldsymbol{\epsilon}_i \sim N_2(\mathbf{0}, \Sigma_i)$, so $\epsilon_{i1} \sim N(0, \sigma_1^2 x_{i1})$, and $\epsilon_{i2} \sim N(0, \sigma_2^2 x_{i2})$, where $\mathbf{0} = (0,0)^T$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2})^T$, as well as, the design vectors, $\mathbf{x}_i = (x_{i1}, x_{i2})^T$, $\mathbf{x}_i > 0$ are fixed, $i = 1, ..., m$, then the model is called the BNLM, (the Biv-Normal linear model). Further, given fixed weighted design vector $\widetilde{\mathbf{x}}_i^T = (\widetilde{x}_{i1}, \widetilde{x}_{i2})$, $\widetilde{x}_{i1} = \sqrt{x_{i1}}, \widetilde{x}_{i2} = \sqrt{x_{i2}}$ the weighted error vectors $\widetilde{\boldsymbol{\epsilon}}_i$ are $i.i.d$ Biv-Normal random vectors, namely $\widetilde{\boldsymbol{\epsilon}}_i \sim N_2(\mathbf{0}, \widetilde{\Sigma})$, $\forall i = 1, ..., m$, where $\mathbf{0} = (0,0)^T$.
Also given fixed $\widetilde{\mathbf{x}}_i^T$, the $i.d$ weighted random component $\widetilde{\mathbf{Y}}_i$ has a biv-Normal distribution i.e, $\widetilde{\mathbf{Y}}_i \sim N_2\left(\begin{pmatrix} \widetilde{x}_{i1} p_1 \\ \widetilde{x}_{i2} p_2 \end{pmatrix}, \widetilde{\Sigma}\right)$.
So, $\hat{\mathbf{p}} \sim N_2(\mathbf{p}, \Sigma_{\mathbf{x}})$, where, $\mathbf{p} = (p_1, p_2)^T$, and $\Sigma_{\mathbf{x}}$ is given by the matrix 2.5.
It follows that

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N_2\left(\mathbf{0}, \begin{pmatrix} \frac{\sigma_1^2}{\mu_1} & \frac{\sigma_{12}\mu_{12}}{\mu_1\mu_2} \\ \frac{\sigma_{12}\mu_{12}}{\mu_1\mu_2} & \frac{\sigma_2^2}{\mu_2} \end{pmatrix}\right),$$

provided that, $\frac{1}{m}\sum_{i=1}^m x_{i1} \longrightarrow \mu_1$, $\frac{1}{m}\sum_{i=1}^m x_{i2} \longrightarrow \mu_2$, $\frac{1}{m}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}} \longrightarrow \mu_{12}$, where μ_1, μ_2, μ_{12} are constants.

For the necessity, one might ask the following question:
Are the conditional estimators based on the residuals the best unbiased estimators for the corresponding parameters $\sigma_1^2, \sigma_2^2, \sigma_{12}$ that are need for statistical inference about the model coefficients?
The answer is not explicit and it needs to investigated, whether these estimators are the BUE's or not, i.e $\widehat{\widetilde{\Sigma}} = \begin{pmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{pmatrix}$ is the best estimator of $\widetilde{\Sigma}$.
These estimators incidentally are defined by the formulae

$$\hat{\sigma}_1^2 = s_1^2 = \frac{1}{m-1}\sum_{i=1}^m \frac{(Y_{i1} - \hat{p}_1 x_{i1})^2}{x_{i1}} = \frac{1}{m-1}\sum_{i=1}^m \left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{x_{i1}}\right)^2$$

$$\hat{\sigma}_2^2 = s_2^2 = \frac{1}{m-1}\sum_{i=1}^m \frac{(Y_{i2} - \hat{p}_2 x_{i2})^2}{x_{i2}} = \frac{1}{m-1}\sum_{i=1}^m \left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{x_{i2}}\right)^2$$

$$\hat{\sigma}_{12} = s_{12} = \frac{1}{m-1}\sum_{i=1}^m \frac{(Y_{i1} - \hat{p}_1 x_{i1})(Y_{i2} - \hat{p}_2 x_{i2})}{\sqrt{x_{i1}x_{i2}}}$$

$$= \frac{1}{m-1}\sum_{i=1}^m \left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{x_{i1}}\right)\left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{x_{i2}}\right),$$

where, $\widetilde{Y}_{ij} = \frac{Y_{ij}}{\sqrt{x_{ij}}}, j = 1, 2,\ x_{i1}, x_{i2} > 0\ \forall i$.
It is necessary to demonstrate the unbiasedness and consistence properties of the MSE's estimators.
Recall that, the least squares regression model (sample linear regression model or fitted model) corresponding to the observed or the true model (2.2) is written as $\widehat{\widetilde{\mathbf{Y}}}_i = \widetilde{X}_i\hat{\mathbf{p}}$, with the corresponding residuals $\widetilde{\mathbf{e}}_i = \widetilde{\mathbf{Y}}_i - \widehat{\widetilde{\mathbf{Y}}}_i$, where $\widetilde{\mathbf{e}}_i = (\widetilde{e}_{i1}, \widetilde{e}_{i2})^T$. For simplicity we will defined the vectors $\widetilde{e}_j = (\widetilde{Y}_j - \widetilde{X}_j\hat{p}_j)$, where $\widetilde{Y}_j = (\widetilde{Y}_{1j}, ..., \widetilde{Y}_{mj})^T$, as well as $\widetilde{X}_j = (\widetilde{x}_{1j}, ..., \widetilde{x}_{mj})^T, i = 1, ..., m, j = 1, 2$.

2.2.6 Properties of the estimator of the covariance structure of the Bivariate SUR model

Theorem

The conditional estimators s_1^2, s_2^2, namely the diagonal elements of the estimator matrix $\widehat{\widetilde{\Sigma}}$ are consistent unbiased estimators of the corresponding parameters (disjointly are also BUE's according to the Lehmann-Scheffe theorem), while s_{12}, s_{21}, i.e., the off diagonal elements are consistent but not necessarily unbiased i.e., asymptotically unbiased.

Proof

Starting with the diagonal entries, we have

$$(m-1)s_1^2 = \widetilde{e}_1^T\widetilde{e}_1 = (\widetilde{Y}_1 - \widetilde{X}_1\hat{p}_1)^T(\widetilde{Y}_1 - \widetilde{X}_1\hat{p}_1) = \sum_{i=1}^m \widetilde{e}_{i1}^2 = \sum_{i=1}^m \left(\widetilde{Y}_{i1} - \sqrt{x_{i1}}\hat{p}_1\right)^2$$
$$= \sum_{i=1}^m \left(\widetilde{Y}_{i1}^2 - 2\widetilde{Y}_{i1}\sqrt{x_{i1}}\hat{p}_1 + x_{i1}\hat{p}_1^2\right) = \sum_{i=1}^m \widetilde{Y}_{i1}^2 - 2\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{i1}}\hat{p}_1 + \sum_{i=1}^m x_{i1}\hat{p}_1^2$$
$$= \sum_{i=1}^m \widetilde{Y}_{i1}^2 - \sum_{i=1}^m x_{i1}\hat{p}_1^2,$$

since

$$\frac{m}{m-1}\frac{1}{m}\sum_{i=1}^m \widetilde{Y}_{i1}^2 \longrightarrow^{\mathcal{P}} E(\widetilde{Y}_{i1}^2) = \sigma_1^2 + p_1^2\mu_1,$$

as well as

$$\frac{m}{m-1}\frac{1}{m}\sum_{i=1}^m x_{i1}\hat{p}_1^2 \longrightarrow^{\mathcal{P}} \mu_1 p_1^2,$$

then, $s_1^2 \longrightarrow^{\mathcal{P}} \sigma_1^2 + p_1^2\mu_1 - p_1^2\mu_1 = \sigma_1^2$, and thus s_1^2 is a consistent estimator of σ_1^2, likewise s_2^2. Furthermore,

$$E\left((m-1)s_1^2\right) = (m-1)E(s_1^2) = \sum_{i=1}^{m} E(\widetilde{Y}_{i1}^2) - \sum_{i=1}^{m} x_{i1}E(\hat{p}_1^2),$$

since

$$E(\widetilde{Y}_{i1}^2) = \sigma_1^2 + \left(E(\widetilde{Y}_{i1})\right)^2 = \sigma_1^2 + p_1^2 x_{i1}, \text{ as well as, } E(\hat{p}_1^2) = \frac{\sigma_1^2}{\sum_{i=1}^{m} x_{i1}} + p_1^2,$$

hence

$$\begin{aligned}(m-1)E(s_1^2) &= \sum_{i=1}^{m}\left(\sigma_1^2 + p_1^2 x_{i1}\right) - \sum_{i=1}^{m} x_{i1}\left(\frac{\sigma_1^2}{\sum_{i=1}^{m} x_{i1}} + p_1^2\right)\\ &= m\sigma_1^2 + p_1^2\sum_{i=1}^{m} x_{i1} - \sigma_1^2 - \sum_{i=1}^{m} x_{i1}p_1^2 = (m-1)\sigma_1^2,\end{aligned}$$

consequently, s_1^2 is an unbiased estimator of σ_1^2, likewise s_2^2, or in general s_j^2 is an unbiased estimator of σ_j^2, $j = 1, 2$.

One may verifying these results via matrix notations

$$\begin{aligned}SSE = \widetilde{e}_1^T\widetilde{e}_1 &= (\widetilde{Y}_1 - \hat{\widetilde{Y}}_1)^T(\widetilde{Y}_1 - \hat{\widetilde{Y}}_1) = (\widetilde{Y}_1 - \widetilde{X}_1\hat{p}_1)^T(\widetilde{Y}_1 - \widetilde{X}_1\hat{p}_1)\\ &= \left(\widetilde{Y}_1 - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{Y}_1\right)^T\left(\widetilde{Y}_1 - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{Y}_1\right)\\ &= \widetilde{Y}_1^T\left(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\right)\widetilde{Y}_1,\end{aligned}$$

the square symmetric nonnegative definite matrix $H_1 = \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T$ is called a projection (hat) matrix, so

$$E(SSE) = E\left(\widetilde{Y}_1^T(I_m - H_1)\widetilde{Y}_1\right) = trace[(I_m - H_1)Var(\widetilde{Y}_1)] + (E(\widetilde{Y}_1))^T(I_m - H_1)E(\widetilde{Y}_1),$$

$$\text{where, } Var(\widetilde{Y}_1) = \sigma_1^2 I_m,\ E(\widetilde{Y}_1) = p_1\widetilde{X}_1^T,$$

hence

$$\begin{aligned}E(SSE) = \sigma_1^2(m-1), \text{as, } & trace(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T) =\\ & trace(I_m) - trace(\widetilde{X}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}) = m-1,\end{aligned}$$

as well as

$$(E(\widetilde{Y}_1))^T(I_m - H_1)E(\widetilde{Y}_1) = p_1\left[\widetilde{X}_1^T\left(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\right)\widetilde{X}_1\right] = 0$$

as

$$\widetilde{X}_1^T \left(I - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\right)\widetilde{X}_1 = \left(\widetilde{X}_1^T\widetilde{X}_1 - \widetilde{X}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_1\right) = 0,$$

therefore s_1^2 is an unbiased estimator of σ_1^2, likewise s_2^2.
Moreover, since

$$SSE = \widetilde{\epsilon}_1^T \left(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\right)\widetilde{\epsilon}_1 = \widetilde{\epsilon}_1^T\widetilde{\epsilon}_1 - \widetilde{\epsilon}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{\epsilon}_1,$$

it follows that
$\frac{m}{m-1}\frac{1}{m}\widetilde{\epsilon}_1^T\widetilde{\epsilon}_1 \longrightarrow^{\mathcal{P}} \sigma_1^2$, and if $\widetilde{Y}_1 = \widetilde{X}_1 p_1 + \widetilde{\epsilon}_1$, then

$$(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{Y}_1 \longrightarrow^{\mathcal{P}} p_1 \Rightarrow (\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{\epsilon}_1 \longrightarrow^{\mathcal{P}} 0, \text{ if } p_1 = 0,$$

and thus s_1^2 is a consistent estimator of σ_1^2, likewise s_2^2.

Unbiasedness property for the estimator s_{12} is not satisfies unless asymptotically. One can investigate this as following (see also [22]):

For the off-diagonal $s_{12} = s_{21}$, we have

$$\widetilde{e}_1^T\widetilde{e}_2 = (\widetilde{Y}_1 - \widetilde{X}_1\hat{p}_1)^T(\widetilde{Y}_2 - \widetilde{X}_2\hat{p}_2) = \widetilde{Y}_1^T(I_m - H_1)(I_m - H_2)\widetilde{Y}_2,$$

so

$$E\left(\widetilde{e}_1^T\widetilde{e}_2\right) = trace\left((I_m - H_1)(I_m - H_2)Cov(\widetilde{Y}_1, \widetilde{Y}_2)\right) + (E(\widetilde{Y}_1))^T(I_m - H_1)(I_m - H_2)E(\widetilde{Y}_2),$$

where

$$Cov(\widetilde{Y}_1, \widetilde{Y}_2) = Cov(\widetilde{Y}_{i1}, \widetilde{Y}_{i2})I_m = \sigma_{12}I_m,\ H_j = \widetilde{X}_j(\widetilde{X}_j^T\widetilde{X}_j)^{-1}\widetilde{X}_j^T,$$

since, $(I_m - H_1)(I_m - H_2) =$

$$I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T - \widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T + \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T,$$

we have

$$\begin{aligned} trace((I_m - H_1)(I_m - H_2)I_m) &= trace((I_m - H_1)(I_m - H_2)) \\ &= m - 2 + \frac{\left(\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}\right)^2}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}}, \end{aligned}$$

it follows that

$$E\left(\widetilde{e}_1^T\widetilde{e}_2\right) = \sigma_{12}\left[m - 2 + \frac{\left(\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}\right)^2}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}}\right],$$

as

$$
\begin{aligned}
(E(\widetilde{Y}_1))^T(I_m - H_1)(I_m - H_2)E(\widetilde{Y}_2) =& p_1\widetilde{X}_1^T[I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T - \widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T \\
&+\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T]\widetilde{X}_2p_2 \\
=& p_1p_2[\widetilde{X}_1^T\widetilde{X}_2 - \widetilde{X}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2 \\
&-\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{X}_2 \\
&+\widetilde{X}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{X}_2] \\
=& p_1p_2[\widetilde{X}_1^T\widetilde{X}_2 - \widetilde{X}_1^T\widetilde{X}_2 - \widetilde{X}_1^T\widetilde{X}_2 + \widetilde{X}_1^T\widetilde{X}_2] = 0,
\end{aligned}
$$

or, as

$$
\begin{aligned}
\widetilde{X}_1^T[I_m - H_1] = \widetilde{X}_1^T[I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T] = [\widetilde{X}_1^T - \widetilde{X}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T] =& \widetilde{X}_1^T - \widetilde{X}_1^T \\
=& \mathbf{0}^T,
\end{aligned}
$$

i.e, $\widetilde{X}_1$ is orthogonal to each column of the projection matrix $[I_m - H_1]$, likewise $\widetilde{X}_2$ is orthogonal to each column of the projection matrix $[I_m - H_2]$.

Corollary

The residuals are orthogonal (perpendicular) to the fitted values of $\widetilde{Y}_1$, also to the design matrix, in other words: $\hat{\widetilde{Y}}_1$ is independent of the error $\widetilde{e}_1$.

Proof

Its enough to show that $\hat{\widetilde{Y}}_1^T\widetilde{e}_1 = 0$.
Since

$$
\hat{\widetilde{Y}}_1^T\widetilde{e}_1 = \widetilde{Y}_1^T H_1(I_m - H_1)\widetilde{Y}_1 = \widetilde{Y}_1^T H_1\widetilde{Y}_1 - \widetilde{Y}_1^T H_1\widetilde{Y}_1 = 0, \text{as, } H_1^2 = H_1,
$$

as well as

$$
\widetilde{X}_1^T\widetilde{e}_1 = \widetilde{X}_1^T(I - H_1)\widetilde{Y}_1 = (\widetilde{X}_1^T - \widetilde{X}_1^T)\widetilde{Y}_1 = \mathbf{0}^T\widetilde{Y}_1 = 0, \text{as, } \widetilde{X}_1^T H_1 = \widetilde{X}_1^T.
$$

Therefore

$$
E(\widetilde{e}_1^T\widetilde{e}_2) = \sigma_{12}\left[m - 1 + \frac{\left(\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}\right)^2}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}} - 1\right] = \sigma_{12}\left[m - 1 + c_m - 1\right],
$$

and hence $s_{12} = \frac{1}{m-1}\widetilde{e}_1^T\widetilde{e}_2$ is a biased estimator of σ_{12} with the biased correction $c_m - 1$, where $c_m = \frac{\left(\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}\right)^2}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}}$.

Furthermore, $(m-1)\widetilde{e}_1^T\widetilde{e}_2=$

$$=\widetilde{Y}_1^T(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T - \widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T + \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T)\widetilde{Y}_2$$
$$=\widetilde{\epsilon}_1^T(I_m - \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T - \widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T + \widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T)\widetilde{\epsilon}_2$$
$$=(\widetilde{\epsilon}_1^T\widetilde{\epsilon}_2 - \widetilde{\epsilon}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{\epsilon}_2 - \widetilde{\epsilon}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{\epsilon}_2 +$$
$$\widetilde{\epsilon}_1^T\widetilde{X}_1(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{X}_2(\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{\epsilon}_2),$$

and since

$$\frac{m}{m-1}\frac{1}{m}\widetilde{\epsilon}_1^T\widetilde{\epsilon}_2 \longrightarrow^{\mathcal{P}} \sigma_{12},$$

and, if

$$\widetilde{Y}_2 = \widetilde{X}_2 p_2 + \widetilde{\epsilon}_2, \text{ then, } (\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{Y}_2 \longrightarrow^{\mathcal{P}} p_2 \Rightarrow (\widetilde{X}_2^T\widetilde{X}_2)^{-1}\widetilde{X}_2^T\widetilde{\epsilon}_2 \longrightarrow^{\mathcal{P}} 0, \text{if, } p_2 = 0,$$
$$\text{and, } \frac{1}{m}\widetilde{X}_1^T\widetilde{X}_2 \longrightarrow^{\mathcal{P}} E(\widetilde{X}_1^T\widetilde{X}_2),$$
$$\text{as well as, } \frac{1}{m}(\widetilde{X}_1^T\widetilde{X}_1)(\widetilde{X}_1^T\widetilde{X}_1)^{-1}\widetilde{X}_1^T\widetilde{\epsilon}_2 \longrightarrow^{\mathcal{P}} 0, \text{as, } \frac{1}{m}(\widetilde{X}_1^T\widetilde{X}_1) \longrightarrow^{\mathcal{P}} E(\widetilde{X}_1^T\widetilde{X}_1),$$

and therefore s_{12} is a consistent estimator of σ_{12}.
From the previous theorem, one may establish the following consequences:

result I

s_{12} is asymptotically unbiased, as $E(s_{12}) = \frac{m-1-(1-c_m)}{m-1}\sigma_{12} \longrightarrow \sigma_{12}$ for any sequence c_m, $0 \leq c_m \leq 1$.

result II

One can observe that, if the relation between x_1, x_2, say linear relationship (in fact r^2 measures the strength of the linear association between $\sqrt{x_{i1}}, \sqrt{x_{i2}}$), i.e, $x_{i2} = x_{i1}$, or $x_{i2} = cx_{i1}, \forall i = 1, ..., m$ for any $c > 0$, this implies that $r^2 = 1$, then the bias term is vanishes and the estimator s_{12} is unbiased.

result III

The coefficient of determination (say the square of the correlation coefficient between the observed values and the fitted values $\hat{\widetilde{Y}}_i$) measures the goodness-of-fit of $\widetilde{Y}_i$, $\hat{\widetilde{Y}}_i$ or measures the strength of the relationship between $\widetilde{Y}_i$, and $\widetilde{X}_i$, which equals

$\frac{SSR}{SST} = 1 - \frac{SSE}{SST}$, hence

$$R^2 = \left(Corr(\widetilde{Y}_{i1}, \hat{\widetilde{Y}}_{i1})\right)^2 = 1 - \frac{\widetilde{Y}_1^T\,(I_m - H_1)\,\widetilde{Y}_1}{\widetilde{Y}_1^T\widetilde{Y}_1} = \frac{(\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{i1}})^2}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m \widetilde{Y}_{i1}^2} = \frac{\sum_{i=1}^m(\hat{\widetilde{Y}}_{i1})^2}{\sum_{i=1}^m \widetilde{Y}_{i1}^2} = \frac{SSR}{SST}.$$

similarly

$$R_c^2 = \left(Corr(\widetilde{Y}_{i1}, \widetilde{Y}_{i2})\right)^2 = 1 - \frac{\widetilde{Y}_1^T(I_m - H_1)(I_m - H_2)\widetilde{Y}_2}{\widetilde{Y}_1^T\widetilde{Y}_2}$$

$$= \frac{\sum_{i=1}^m x_{i2}(\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{i1}}\sum_{i=1}^m \widetilde{Y}_{i2}\sqrt{x_{i2}}) + \sum_{i=1}^m x_{i1}(\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{i2}}\sum_{i=1}^m \widetilde{Y}_{i2}\sqrt{x_{i2}})}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}\left(\sum_{i=1}^m \widetilde{Y}_{i1}\widetilde{Y}_{i2}\right)}$$
$$\frac{-\sum_{i=1}^m \widetilde{Y}_{i1}\sqrt{x_{i1}}\sum_{i=1}^m \widetilde{Y}_{i2}\sqrt{x_{i2}}(\sum_{i=1}^m \sqrt{x_{i1}x_{i2}})}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}\left(\sum_{i=1}^m \widetilde{Y}_{i1}\widetilde{Y}_{i2}\right)},$$

$0 \le R_c^2 \le 1$.

2.2.7 Asymptotic normality of the ratio estimator vector $\hat{\mathbf{p}}_m$

As in the univariate case, it will be assumed that the random error vectors $\widetilde{\boldsymbol{\epsilon}}_i$ are not normally distributed but $i.i.d$ random vectors, $i = 1, ..., m$, i.e $E\left(\widetilde{\boldsymbol{\epsilon}}_i\right) = \mathbf{0}$, and $Cov(\widetilde{\boldsymbol{\epsilon}}_i) = \widetilde{\Sigma}$. Moreover, under certain conditions on the design matrix $\mathbf{X}_i$ one can show that in large sample size, $\hat{\mathbf{p}}_m$ has the asymptotic Normal distribution. These conditions namely, the pairs
$(\mathbf{X}_i, \mathbf{Y}_i)$ are $i.i.d \Rightarrow (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Y}}_i)$ are also $i.i.d$, where $\widetilde{\mathbf{X}}_i = \begin{pmatrix} \widetilde{X}_{i1} & 0 \\ 0 & \widetilde{X}_{i2} \end{pmatrix}$, $\widetilde{X}_{ij} = \sqrt{X_{ij}}, X_{ij} > 0, i = 1, ..., m,\ j = 1, 2$, as well as $E(\widetilde{X}_{ij}\widetilde{X}_{ij'})$ exists $\forall j, j' = 1, 2$.

$\hat{\mathbf{p}}_{OLS}$ can be rewritten as

$$\begin{aligned} \hat{\mathbf{p}}_{OLS} &= \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}} = \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^n \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i \\ &= \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\left(\widetilde{\mathbf{X}}_i\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}_i\right) \\ &= \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\mathbf{p} + \sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right), \end{aligned}$$

and this can be reexpression as

$$\sqrt{m}\left(\hat{\mathbf{p}}_{OLS}-\mathbf{p}\right)=\left(\frac{1}{m}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right)$$

$$=\left(\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} X_{i1} & 0 \\ 0 & X_{i2}\end{pmatrix}\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\begin{pmatrix}\widetilde{X}_{i1}\widetilde{\epsilon}_{i1} \\ \widetilde{X}_{i2}\widetilde{\epsilon}_{i2}\end{pmatrix}\right). \quad (2.6)$$

To derive the required asymptotic distribution, it needs to investigate, first the denominator matrix in eq. 2.6 is consistent, and second the numerator obeys the Central limit theorem.

Thus, in large m, its straightforward to see that the denominator of eq. 2.6 is consistent. By the LLN, we have

$$\left(\frac{1}{m}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\longrightarrow^{\mathcal{P}}\left(E(\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i)\right)^{-1},$$

provided that $X_{ij}>0$ almost surely, thus yields

$$\left(\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} X_{i1} & 0 \\ 0 & X_{i2}\end{pmatrix}\right)^{-1}\longrightarrow^{\mathcal{P}}\begin{pmatrix} E(X_1) & 0 \\ 0 & E(X_2)\end{pmatrix}^{-1}.$$

It may need to mention that, in case of the $\widetilde{\mathbf{X}}_i=\begin{pmatrix}\widetilde{x}_{i1} & 0 \\ 0 & \widetilde{x}_{i2}\end{pmatrix}$, $\widetilde{x}_{ij}=\sqrt{x_{ij}}, x_{ij}>0$ are fixed variables, $i=1,...,m,\ j=1,2$, then also satisfies

$$\left(\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} x_{i1} & 0 \\ 0 & x_{i2}\end{pmatrix}\right)^{-1}\longrightarrow\begin{pmatrix}\mu_1 & 0 \\ 0 & \mu_2\end{pmatrix}^{-1}.$$

provided that, $\frac{1}{m}\sum_{i=1}^{m}x_{i1}\longrightarrow\mu_1$, $\frac{1}{m}\sum_{i=1}^{m}x_{i2}\longrightarrow\mu_2$, where μ_1,μ_2 are constants.

The numerator obeys the CLT

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\longrightarrow^{\mathcal{D}}N\left(\mathbf{0},E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i\right)\right),$$

where, the marginal or the asymptotic covariance

$$Cov\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right)=E\left(Cov\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\mid\widetilde{X}_{i1},\widetilde{X}_{i2}\right)\right)+Cov\left(E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\mid\widetilde{X}_{i1},\widetilde{X}_{i2}\right)\right)$$

$$=E\left(\widetilde{\mathbf{X}}_i^T Cov\left(\widetilde{\boldsymbol{\epsilon}}_i\mid\widetilde{X}_{i1},\widetilde{X}_{i2}\right)\widetilde{\mathbf{X}}_i\right)=E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i\right),$$

as

$$E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1}, \widetilde{X}_{i2}\right) = \widetilde{X}_i^T E\left(\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1}, \widetilde{X}_{i2}\right) = \mathbf{0}_2.$$

Thus, the numerator of 2.6 yields

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\begin{pmatrix}\sqrt{X_{i1}}\widetilde{\epsilon}_{i1}\\ \sqrt{X_{i2}}\widetilde{\epsilon}_{i2}\end{pmatrix} \longrightarrow^{\mathcal{D}} N\left(\mathbf{0}, \begin{pmatrix}\sigma_1^2 E(X_{i1}) & \sigma_{12}E\left(\sqrt{X_{i1}X_{i2}}\right)\\ \sigma_{12}E\left(\sqrt{X_{i1}X_{i2}}\right) & \sigma_2^2 E(X_{i2})\end{pmatrix}\right)$$
$$\equiv N\left(\mathbf{0}, \begin{pmatrix}\sigma_1^2 E(X_1) & \sigma_{12}E\left(\sqrt{X_1X_2}\right)\\ \sigma_{12}E\left(\sqrt{X_1X_2}\right) & \sigma_2^2 E(X_2)\end{pmatrix}\right).$$

Therefore, with help of the known Slutsky's lemma, equation (2.6) can be rewritten (since, $\hat{\mathbf{p}}_{OLS}$ is the ratio estimator $\hat{\mathbf{p}}_m$) as

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N(\mathbf{0}, \begin{pmatrix}E(X_1) & 0\\ 0 & E(X_2)\end{pmatrix}^{-1}\begin{pmatrix}\sigma_1^2 E(X_1) & \sigma_{12}E(\sqrt{X_1X_2})\\ \sigma_{12}E(\sqrt{X_1X_2}) & \sigma_2^2 E(X_2)\end{pmatrix}$$
$$\begin{pmatrix}E(X_1) & 0\\ 0 & E(X_2)\end{pmatrix}^{-1})$$
$$\equiv N\left(\mathbf{0}, \begin{pmatrix}\frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)}\\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)}\end{pmatrix}\right) \equiv N\left(\mathbf{0}, \Sigma_{\mathbf{p}}\right),$$

which results in the asymptotic Normal with the asymptotic covariance matrix

$$\Sigma_{\mathbf{p}} = Cov(\sqrt{m}\hat{\mathbf{p}}_m) = \begin{pmatrix}\frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)}\\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)}\end{pmatrix}.$$

2.2.8 Approximate confidence intervals for a linear combination of the proportions

From the last result, we have

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N(\mathbf{0}, \Sigma_{\mathbf{p}}). \tag{2.7}$$

By applying the extremely useful result called, the Cramer-Wold device (see [14] pp. 147), it essentially reduces multivariate CLTs to a special case of univariate CLTs. Hence, this result shows that the expr.2.7 holds iff, $\forall \boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T \in \mathbf{R}^2$, such that

$\| \boldsymbol{\alpha} \| > 0$, so we have

$$\begin{aligned}
\sqrt{m}\left(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m - \boldsymbol{\alpha}^T\mathbf{p}\right) &= \sqrt{m}(\hat{\vartheta}_m - \vartheta) \longrightarrow^{\mathcal{D}} N\left(0, \boldsymbol{\alpha}^T\Sigma_{\mathbf{p}}\boldsymbol{\alpha}\right) \\
&\equiv N\left(0, (\alpha_1 \quad \alpha_2)\begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} \\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}\right) \\
&\equiv N\left(0, \alpha_1^2\frac{\sigma_1^2}{E(X_1)} + \alpha_2^2\frac{\sigma_2^2}{E(X_2)} + 2\alpha_1\alpha_2\frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)}\right), \quad (2.8)
\end{aligned}$$

where, $\hat{\vartheta}_m = \boldsymbol{\alpha}^T\hat{\mathbf{p}}_m$, and $\vartheta = \boldsymbol{\alpha}^T\mathbf{p}$. It follows that, the asymptotic variance of $\hat{\vartheta}_m$ is given by

$$Var(\sqrt{m}\hat{\vartheta}_m) = \sigma_{\vartheta}^2 = \alpha_1^2\frac{\sigma_1^2}{E(X_1)} + \alpha_2^2\frac{\sigma_2^2}{E(X_2)} + 2\alpha_1\alpha_2\frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)}.$$

Constructing the approximate confidence intervals for the linear combination ϑ, needs to estimate the asymptotic covariance matrix $\Sigma_{\mathbf{p}}$, by plugging in estimates for the individual parameters $\sigma_1^2, \sigma_2^2, \sigma_{12}, E(X_1), E(X_2)$, and $E(\sqrt{X_1X_2})$.
Explicitly, these consistent estimators are s_1^2, s_2^2, s_{12}, $\overline{X}_{.1}, \overline{X}_{.2}$, and $\overline{X}_{.12}$, where

$$\begin{aligned}
s_1^2 &= \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{X_{i1}}\right)^2, s_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{X_{i2}}\right)^2, \\
s_{12} &= \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{X_{i1}}\right)\left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{X_{i2}}\right), \ \widetilde{Y}_{ij} = \frac{Y_{ij}}{\sqrt{X_{ij}}}, \ X_{ij} > 0, j = 1, 2,
\end{aligned}$$

$\overline{X}_{.1} = \frac{1}{m}\sum_{i=1}^{m} X_{i1}$, $\overline{X}_{.2} = \frac{1}{m}\sum_{i=1}^{m} X_{i2}$, $\overline{X}_{.12} = \frac{1}{m}\sum_{i=1}^{m}\sqrt{X_{i1}X_{i2}}$. The consistent estimator matrix for the covariance matrix $\Sigma_{\mathbf{p}}$ is given by

$$\widehat{\Sigma}_{\mathbf{p}} = \begin{pmatrix} \frac{s_1^2}{\overline{X}_{.1}} & \frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}} \\ \frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}} & \frac{s_2^2}{\overline{X}_{.2}} \end{pmatrix}, \quad (2.9)$$

also it follows that, the standard error of the linear combination of the sample ratios is given by

$$\widehat{\sigma}_{\vartheta} = s.e(\hat{\vartheta}_m) = \sqrt{\frac{1}{m}\left(\alpha_1^2\frac{s_1^2}{\overline{X}_{.1}} + \alpha_2^2\frac{s_2^2}{\overline{X}_{.2}} + 2\alpha_1\alpha_2\frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}}\right)},$$

provided that, $\overline{X}_{.1}, \overline{X}_{.2}, \overline{X}_{.12}, s_1^2, s_2^2$, and s_{12} are consistent estimators of the corresponding parameters $E(X_1), E(X_2), E(\sqrt{X_1X_2}), \sigma_1^2, \sigma_2^2$, and σ_{12}, and hence the standard error is also a consistent estimator for the corresponding asymptotic standard deviation.

Further, one may say that the consistent matrix (2.9) is to be positive definite or at least positive semi definite (see A.1), to ensure $\frac{\boldsymbol{\alpha}^T \widehat{\Sigma_{\mathbf{p}}} \boldsymbol{\alpha}}{m} \geq 0,\ \boldsymbol{\alpha} \geq \mathbf{0}$, otherwise, we have to exclude the negative variances this will be explicitly clarifying in the chapter 4.

Finally, and based on these estimators, an approximate asymptotic normal confidence interval for the linear combination $\vartheta = \alpha_1 p_1 + \alpha_2 p_2$ can be established by

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{\vartheta}_m)\right],$$

as well as, the suggested conservative confidence interval (safety bounds) by the t-quantile is given by

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

where, $t_{(m-1,1-\frac{\alpha}{2})}$ is the $\left(1-\frac{\alpha}{2}\right)$ quantile of the $t-$distribution with $(m-1)$ $d.f.$

2.2.9 Derivation of confidence intervals for the linear combination of the proportions

One like to summarize the derivation of the confidence intervals for the linear combination $\vartheta = \boldsymbol{\alpha}^T \mathbf{p} = \alpha_1 p_1 + \alpha_2 p_2$ in the cases:

case I

On one hand and on one side, we will obtain the distribution of $\hat{\vartheta} = \boldsymbol{\alpha}^T \hat{\mathbf{p}}$ (unbiased estimator of ϑ) given the fixed pair of design vectors $(\mathbf{x}_1^T, \mathbf{x}_2^T)$, $\mathbf{x}_j = (x_{1j}, ..., x_{mj})^T, j = 1, 2$, and $\mathbf{Y}_i$ has a biv-Normal distribution.
Since

$$\hat{\vartheta} \sim N\left(\alpha_1 p_1 + \alpha_2 p_2, \alpha_1^2 \frac{\sigma_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{\sigma_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12} \sum_{i=1}^m \sqrt{x_{i1} x_{i2}}}{\sum_{i=1}^m x_{i1} \sum_{i=1}^m x_{i2}}\right) \quad (2.10)$$

then 2.10 will rewrite after its standardization, as

$$\begin{aligned}\sqrt{m}\left(\hat{\vartheta} - \vartheta\right) &\sim N\left(0, \alpha_1^2 \frac{m\sigma_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{m\sigma_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{m\sigma_{12} \sum_{i=1}^m \sqrt{x_{i1} x_{i2}}}{\sum_{i=1}^m x_{i1} \sum_{i=1}^m x_{i2}}\right) \\ &\equiv N\left(0, \sigma_{\mathbf{x}}^2\right),\end{aligned}$$

where, the distribution variance

$$\sigma_{\mathbf{x}}^2 = \alpha_1^2 \frac{m\sigma_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{m\sigma_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{m\sigma_{12} \sum_{i=1}^m \sqrt{x_{i1} x_{i2}}}{\sum_{i=1}^m x_{i1} \sum_{i=1}^m x_{i2}}.$$

It follows that, the exact $(1-\alpha)\%$ confidence intervals for the linear combination ϑ are given by

$$\left[\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}}\; S.E(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta}-\vartheta}{\sigma_{\mathbf{x}}} \sim N(0,1), \text{ when } \sigma_1^2, \sigma_2^2, \sigma_{12} \text{ are known}$$
$$\left[\hat{\vartheta} \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta}-\vartheta}{\widehat{\sigma}_{\mathbf{x}}} \sim t_{m-1}, \quad \text{when } \sigma_1^2, \sigma_2^2, \sigma_{12} \text{ are unknown},$$

where, the standard errors of $\hat{\vartheta}$ are obtained by

$$S.E(\hat{\vartheta}) = \sqrt{\alpha_1^2 \frac{\sigma_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{\sigma_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}}},$$

as well as

$$s.e(\hat{\vartheta}) = \sqrt{\alpha_1^2 \frac{s_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{s_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{s_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}}}.$$

On the other side, and for non-bivariate Normal of $\mathbf{Y}_i's$, and if the sequences, say $\frac{1}{m}\sum_{i=1}^m x_{i1} \longrightarrow \mu_1$, $\frac{1}{m}\sum_{i=1}^m x_{i2} \longrightarrow \mu_2$, $\frac{1}{m}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}} \longrightarrow \mu_{12}$, where μ_1, μ_2, μ_{12} are some constants, then the sequence of the approximate variances $\sigma_{\mathbf{x}}^2$ converges to the corresponding constant variance

$$\sigma_{\vartheta}^2 = \alpha_1^2 \frac{\sigma_1^2}{\mu_1} + \alpha_2^2 \frac{\sigma_2^2}{\mu_2} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\mu_{12}}{\mu_1\mu_2}, \tag{2.11}$$

as m tends to infinity. So, the asymptotic distribution for the distribution of the estimator $\hat{\vartheta}_m$ can be obtained as

$$\sqrt{m}\left(\hat{\vartheta}_m - \vartheta\right) \longrightarrow^{\mathcal{D}} N(0, \sigma_{\vartheta}^2),$$

as 2.11 is its asymptotic variance.

Therefore, the approximate $(1-\alpha)\%$ confidence intervals for ϑ (with the unknown $\sigma_1^2, \sigma_2^2, \sigma_{12}$) are given by

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{\vartheta}_m)\right],$$

or

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

provided that the estimator $\widehat{Var(\hat{\vartheta}_m)} = \frac{\widehat{Var(\sqrt{m}\hat{\vartheta}_m)}}{m}$ is a consistent estimator of σ_μ^2, and

$$\frac{\hat{\vartheta}_m - \vartheta}{\sqrt{\widehat{Var(\hat{\vartheta}_m)}}} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1),$$

as well as, the standard error of $\hat{\vartheta}_m$,

$$s.e(\hat{\vartheta}_m) = \sqrt{\alpha_1^2 \frac{s_1^2}{\sum_{i=1}^m x_{i1}} + \alpha_2^2 \frac{s_2^2}{\sum_{i=1}^m x_{i2}} + 2\alpha_1\alpha_2 \frac{s_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^m x_{i1} \sum_{i=1}^m x_{i2}}}.$$

case II

On the second hand, the conditional variance of $\hat{\vartheta}$ given the pair random vectors $(\mathbf{X}_1^T, \mathbf{X}_2^T)$, $\mathbf{X}_j = (X_{1j}, ..., X_{mj})^T, j = 1,2$, when there vectors $\mathbf{X}_i = (X_{i1}, X_{i2})^T, i = 1, ..., m$ are $i.i.d$ random vectors with non-biv normal of $\mathbf{Y}_i's$ is given by

$$\alpha_1^2 \frac{\sigma_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{\sigma_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1} \sum_{i=1}^m X_{i2}},$$

asymptotically

$$Var(\sqrt{m}\hat{\vartheta}) = mVar(\hat{\vartheta}) = \alpha_1^2 \frac{\sigma_1^2}{\overline{X}_{.1}} + \alpha_2^2 \frac{\sigma_2^2}{\overline{X}_{.2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}}. \tag{2.12}$$

And from the expression 2.8, one can rewrite the asymptotic distribution of $\hat{\vartheta}_m$ as

$$\sqrt{m}\left(\hat{\vartheta}_m - \vartheta\right) \longrightarrow^{\mathcal{D}} N\left(0, \sigma_\vartheta^2\right),$$

where, the asymptotic variance σ_ϑ^2 is given by

$$\sigma_\vartheta^2 = \alpha_1^2 \frac{\sigma_1^2}{E(X_1)} + \alpha_2^2 \frac{\sigma_2^2}{E(X_2)} + 2\alpha_1\alpha_2 \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)}, \tag{2.13}$$

however, since, $\overline{X}_{.1} = \frac{1}{m}\sum_{i=1}^m X_{i1} \longrightarrow^{\mathcal{P}} E(X_1)$, $\overline{X}_{.2} = \frac{1}{m}\sum_{i=1}^m X_{i2} \longrightarrow^{\mathcal{P}} E(X_2)$, as well as $\overline{X}_{.12} = \frac{1}{m}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}} \longrightarrow^{\mathcal{P}} E(\sqrt{X_1X_2})$, (LLN), then the variance 2.12 converges in probability to the corresponding asymptotic variance 2.13.
And thus, a consistent variance estimator for σ_ϑ^2 is given by

$$\widehat{\sigma_\vartheta^2} = \alpha_1^2 \frac{s_1^2}{\overline{X}_{.1}} + \alpha_2^2 \frac{s_2^2}{\overline{X}_{.2}} + 2\alpha_1\alpha_2 \frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}},$$

provided that, the estimators

$$s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{X_{i1}}\right)^2, s_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{X_{i2}}\right)^2, \text{ and}$$

$$s_{12} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{i1} - \hat{p}_1\sqrt{X_{i1}}\right)\left(\widetilde{Y}_{i2} - \hat{p}_2\sqrt{X_{i2}}\right), \; X_{i1}, X_{i2} > 0,$$

as well as, $\overline{X}_{.1}, \overline{X}_{.2}, \overline{X}_{.12}$, are all consistent estimators for the corresponding parameters σ_1^2, σ_2^2, $\sigma_{12}, E(X_1), E(X_2), E(\sqrt{X_1X_2})$, subsequently it follows that the standard error of $\hat{\vartheta}$, $s.e(\hat{\vartheta}) = \sqrt{\frac{1}{m}\widehat{\sigma_{\vartheta}^2}}$.
The approximate confidence intervals for the linear combination $\vartheta = \alpha_1 p_1 + \alpha_2 p_2$ can be constructed based on the following cases:

- If $\sigma_1^2, \sigma_2^2, \sigma_{12}$ are known, then the approximate confidence bounds for ϑ is given by

$$\left[\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}} S.E(\hat{\vartheta})\right],$$

  where, the Standard Error

$$S.E(\hat{\vartheta}) = \sqrt{\frac{1}{m}\left(\alpha_1^2\frac{\sigma_1^2}{\overline{X}_{.1}} + \alpha_2^2\frac{\sigma_2^2}{\overline{X}_{.2}} + 2\alpha_1\alpha_2\frac{\sigma_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}}\right)},$$

  and the quantile $z_{1-\frac{\alpha}{2}}$ is defined as previous.

- In case of the unknown parameters $\sigma_1^2, \sigma_2^2, \sigma_{12}$, again the asymptotic theory given here can be used to obtain the approximate confidence interval for ϑ, which is given by

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{\vartheta}_m)\right],$$

  where

$$s.e(\hat{\vartheta}_m) = \sqrt{\frac{1}{m}\left(\alpha_1^2\frac{s_1^2}{\overline{X}_{.1}} + \alpha_2^2\frac{s_2^2}{\overline{X}_{.2}} + 2\alpha_1\alpha_2\frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}}\right)}.$$

- For small sample sizes, and with all these available consistent estimators, given

$$\frac{\hat{\vartheta}_m - \vartheta}{s.e(\hat{\vartheta}_m)} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1),$$

a statistic $t-$distribution can be used to obtain the conservative confidence interval

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

where, $t_{(m-1,1-\frac{\alpha}{2})}$ is the $\left(1-\frac{\alpha}{2}\right)$ quantile of the $t-$ distribution with $(m-1)$ degrees of freedom.

Finally, when $\mathbf{Y}_i's$ has a biv-Normal distribution, then we have

$$\hat{\vartheta} \sim N\left(\alpha_1 p_1 + \alpha_2 p_2, \alpha_1^2 \frac{\sigma_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{\sigma_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1}\sum_{i=1}^m X_{i2}}\right),$$

it follows that

$$\begin{aligned}\sqrt{m}\left(\hat{\vartheta}-\vartheta\right) &\sim N\left(0, \alpha_1^2 \frac{m\sigma_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{m\sigma_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{m\sigma_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1}\sum_{i=1}^m X_{i2}}\right) \\ &\equiv N\left(0, \sigma_{\mathbf{X}}^2\right),\end{aligned}$$

where, the distribution variance

$$\sigma_{\mathbf{X}}^2 = \alpha_1^2 \frac{m\sigma_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{m\sigma_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{m\sigma_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1}\sum_{i=1}^m X_{i2}}.$$

Thus, the exact $(1-\alpha)\%$ confidence intervals for ϑ, are given by

$$\begin{aligned}&\left[\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}}\ S.E(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta}-\vartheta}{\sigma_{\mathbf{X}}} \sim N(0,1), \text{ when } \sigma_1^2, \sigma_2^2, \sigma_{12} \text{ are known} \\ &\left[\hat{\vartheta} \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta}-\vartheta}{\widehat{\sigma}_{\mathbf{X}}} \sim t_{m-1}, \quad \text{when } \sigma_1^2, \sigma_2^2, \sigma_{12} \text{ are unknown},\end{aligned}$$

where, the standard errors

$$S.E(\hat{\vartheta}) = \sqrt{\alpha_1^2 \frac{\sigma_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{\sigma_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{\sigma_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1}\sum_{i=1}^m X_{i2}}},$$

as well as

$$s.e(\hat{\vartheta}) = \sqrt{\alpha_1^2 \frac{s_1^2}{\sum_{i=1}^m X_{i1}} + \alpha_2^2 \frac{s_2^2}{\sum_{i=1}^m X_{i2}} + 2\alpha_1\alpha_2 \frac{s_{12}\sum_{i=1}^m \sqrt{X_{i1}X_{i2}}}{\sum_{i=1}^m X_{i1}\sum_{i=1}^m X_{i2}}}.$$

2.3 Bivariate Poisson model

A class of bivariate Poisson distributions was introduced and investigated by Aitken (1936), Campbell (1938), Consael (1952), and Holgate (1964). Recently introduced and discussed by N. Johnson, S. Kotz, and N. Balakrishnan (1997). Poisson models were discussed by Karlis and Ntzoufras (2000), they have many researches and articles in this field. There are many applications of bivariate Poisson models in which Bivariate count data arise for example; in Medicine: paired count data in medical research; Epidemiology: joint concurrence of two different diseases; Marketing: joint purchases of two products; sports especially soccer, football, handball, etc. See also [11].
Karlis and Ntzoufras have been considered independent variables that are Poisson distributed. They also considered discrete bivariate and multivariate count data.

In this section, we will extend the univariate Poisson model to the bivariate setting, deals with two dimensional count data that are Poisson distributed.
In shortcut, this is another method for the analysis based on other model.

2.3.1 Description of the model

The model is considered by extending the univariate Poisson model to the model with two components of marked count data where each decomposes additively into two disjoint groups of data, and thus each individual in the group is independently distributed Poisson random variable, so each two groups of the components can be respectively include for instance the events:

- No.of success and failure.
- No.of defect and non-defect.
- Count with property and count with out property, etc.

Let us consider the following observable random variables:
$X_{i1}, X_{i2} \equiv$ total amount of counts of $1^{st}, 2^{nd}$ component respectively for individual i,
$Y_{i1}, Y_{i2} \equiv$ No.of successes of the $1^{st}, 2^{nd}$comp. respectively for individual i,
$Z_{i1}, Z_{i2} \equiv$No.of failures of the $1^{st}, 2^{nd}$comp. respectively for individual i, with the latent variables:

$W_i \equiv$No.of successes simultaneously in both components for individual i,
$V_{i1} \equiv$ No.of success in the 1^{st}, and failure in 2^{nd} component for individual i.
$V_{i2} \equiv$ No.of failure in the 1^{st}, and success in 2^{nd} component for individual i,
$i = 1, ..., m$, where m is the sample size, i.e, W_i, V_{i1}, V_{i2} are unobservable independent Poisson random variables with parameters $\lambda_0, \lambda_1, \lambda_2$, as well as Z_{i1}, Z_{i2} are observable independent Poisson r.v's with parameters μ_1, μ_2 respectively. Define the random variables
$Y_{ij} = W_i + V_{ij}$, additionally $X_{ij} = Y_{ij} + Z_{ij}$, where, $0 \leq Y_{ij} \leq X_{ij}, j = 1, 2$. All, $W_i, V_{i1}, V_{i2}, Z_{i1}, Z_{i2}$ are independent Poisson r.v's with parameters $\lambda_0, \lambda_1, \lambda_2, \mu_1, \mu_2$ respectively.

The random variables Y_{i1}, Y_{i2} have jointly a bivariate Poisson distribution denoted as $BPoiss(\lambda_0, \lambda_1, \lambda_2)$, if they have joint probability function (see [6], or [10]):

$$\begin{aligned} P\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}\right) &= P\left(W_i + V_{i1} = y_{i1}, W_i + V_{i2} = y_{i2}\right) \\ &= \sum_{w_i} P\left(W_i = w_i, V_{i1} = y_{i1} - w_i, V_{i2} = y_{i2} - w_i\right) \\ &= e^{-(\lambda_0+\lambda_1+\lambda_2)} \sum_{w_i=0}^{\min(y_{i1},y_{i2})} \frac{\lambda_0^{w_i}}{w_i!} \frac{\lambda_1^{y_{i1}-w_i}}{(y_{i1}-w_i)!} \frac{\lambda_2^{y_{i2}-w_i}}{(y_{i2}-w_i)!}. \end{aligned} \tag{2.14}$$

The function (2.14) is computational demanding, and very complicated for estimation purposes. Moreover, $(X_{i1}, X_{i2}) \sim BPoiss(\lambda_0, \lambda_1 + \mu_1, \lambda_2 + \mu_2)$, where $\lambda_0, \lambda_1 + \mu_1, \lambda_2 + \mu_2$ are the parameters of the corresponding independent Poisson variables $W_i, V_{i1} + Z_{i1}, V_{i2} + Z_{i2}$.

2.3.2 Properties of the model

With the properties given in [10], one can list some interesting properties of the model

- The marginal distributions are Poisson, namely $Y_{ij} \sim Poiss(\lambda_0 + \lambda_j)$, as well as $X_{ij} \sim Poiss(\lambda_0 + \lambda_j + \mu_j)$. This implies that $E(Y_{ij}) = Var(Y_{ij}) = \lambda_0 + \lambda_j$, and $E(X_{ij}) = Var(X_{ij}) = \lambda_0 + \lambda_j + \mu_j$.

- $Cov(X_{i1}, X_{i2}) = Cov(Y_{i1}, Y_{i2}) = \lambda_0$, (see [6] pp. 126).

- The marginal conditional distributions of Y_{ij} given X_{ij} are given by: $P^{(Y_{ij}|X_{ij})} \sim Bin\left(X_{ij}, \frac{\lambda_0+\lambda_j}{\lambda_0+\lambda_j+\mu_j}\right) \equiv Bin(X_{ij}, p_j)$, $p_j = \frac{\lambda_0+\lambda_j}{\lambda_0+\lambda_j+\mu_j}$ is the Binomial proportion of the j^{th} component. It follows that $E(Y_{ij} \mid X_{ij}) = p_j X_{ij}$, as well as

$Var(Y_{ij} \mid X_{ij}) = p_j(1-p_j)X_{ij}, j = 1, 2$. However the joint conditional distribution $P^{(Y_{i1},Y_{i2}|X_{i1},X_{i2})}$ as well as the conditional covariance $Cov(Y_{i1}, Y_{i2} \mid X_{i1}, X_{i2})$ can not be explicitly calculated.

- Maximum likelihood estimation.
  It is too complicated to derive the MLE's of the parameters $\lambda_0, \lambda_1, \lambda_2$ from the probability function (2.14), due containing the latent variable W_i, whereas, $\hat{\mu}_{j,ML} = \overline{Z}_{.j}$ (from the independence). Karlis and Ntzoufras (2003) had been described ML estimation for bivariate Poisson model via an EM algorithm, which does not need calculation of the function (2.14), (for more details on the EM algorithm see [17])

- Unconditional consistent estimators of the combinations:
  $\lambda_0, \lambda_0 + \lambda_j, \lambda_0 + \lambda_j + \mu_j$, are: $s_{Y_1,Y_2}, \overline{Y}_{.j}, \overline{X}_{.j}$, respectively ([6], pp. 129) i.e: $s_{Y_1,Y_2} \longrightarrow^{\mathcal{P}} \lambda_0, \overline{Y}_{.j} \longrightarrow^{\mathcal{P}} \lambda_0 + \lambda_j, \overline{Y}_{.j} - s_{Y_1,Y_2} \longrightarrow^{\mathcal{P}} \lambda_j$, as well as $\overline{X}_{.j} \longrightarrow^{\mathcal{P}} \lambda_0 + \lambda_j + \mu_j$. Furthermore, $\overline{X}_{.j} - s_{Y_1,Y_2} \longrightarrow^{\mathcal{P}} \lambda_j + \mu_j$, $\overline{Z}_{.j} \longrightarrow^{\mathcal{P}} \mu_j$, where $\overline{Y}_{.j} = \frac{1}{m}\sum_{i=1}^m Y_{ij}$, $\overline{X}_{.j} = \frac{1}{m}\sum_{i=1}^m X_{ij}$, and $\overline{Z}_{.j} = \frac{1}{m}\sum_{i=1}^m Z_{ij}$, $j = 1, 2$, as well as the unconditional sample covariance $s_{Y_1,Y_2} = \frac{1}{m-1}\sum_{i=1}^m (Y_{i1} - \overline{Y}_{.1})(Y_{i2} - \overline{Y}_{.2})$, and also $\hat{p}_j = \frac{\widehat{\lambda_0+\lambda_j}}{\widehat{\lambda_0+\lambda_j+\mu_j}} = \frac{\overline{Y}_{.j}}{\overline{X}_{.j}}, j = 1, 2$ are the commonly used ratio estimators for the proportions p_j.

2.3.3 The Bivariate Poisson distribution

One may describe the model through an 2×5 matrix $\mathbf{A}_y$ the elements of $\mathbf{A}_y$ are zero and ones, no duplicate rows exist, and the vector $\mathbf{T}_i = (W_i, V_{i1}, V_{i2}, Z_{i1}, Z_{i2})^T$, $i = 1, ..., m$. So

$$\mathbf{A}_y = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix},$$

and thus, the linear equations

$$\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T = \mathbf{A}_y \mathbf{T}_i$$

follow a bivariate Poisson distribution with parameters $\lambda_0, \lambda_1, \lambda_2$. Furthermore, define the matrix

$$\mathbf{A}_x = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

also, the linear equations

$$\mathbf{X}_i = (X_{i1}, X_{i2})^T = \mathbf{A}_x \mathbf{T}_i,$$

follows also a bivariate Poisson distribution with parameters $\lambda_0, \lambda_1 + \mu_1, \lambda_2 + \mu_2$. Further, define the vector

$$\mathbf{H}_i = (Y_{i1}, X_{i1}, Y_{i2}, X_{i2})^T, \text{ and } \mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Similarly, $\mathbf{H}_i = \mathbf{A}\mathbf{T}_i$.

In order to obtain the approximate confidence intervals for a linear combination of the proportions, One shall first consider the asymptotic distribution of the proportions estimators by the delta method (since the joint conditional probability distributions can not be obtained).

2.3.4 Asymptotic normality of the proportion estimator

For the linear equations $\mathbf{H}_i = \mathbf{A}\mathbf{T}_i$, the covariance structure is obtained by

$$\begin{aligned}
\Sigma^* &= Cov(\mathbf{H}_i) = \mathbf{A}\ Cov(\mathbf{T}_i)\ \mathbf{A}^T \\
&= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & \mu_2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \lambda_0 + \lambda_1 & \lambda_0 + \lambda_1 & \lambda_0 & \lambda_0 \\ \lambda_0 + \lambda_1 & \lambda_0 + \lambda_1 + \mu_1 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_0 + \lambda_2 & \lambda_0 + \lambda_2 \\ \lambda_0 & \lambda_0 & \lambda_0 + \lambda_2 & \lambda_0 + \lambda_2 + \mu_2 \end{pmatrix} \\
&= \begin{pmatrix} E(Y_1) & E(Y_1) & \lambda_0 & \lambda_0 \\ E(Y_1) & E(X_1) & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & E(Y_2) & E(Y_2) \\ \lambda_0 & \lambda_0 & E(Y_2) & E(X_2) \end{pmatrix}.
\end{aligned}$$

By the Multivariate central limit theorem, we have

$$\sqrt{m}\left(\overline{\mathbf{H}} - E(\mathbf{H})\right) \longrightarrow^{\mathcal{D}} N_4\left(\mathbf{0}, \Sigma^*\right), \text{ or}$$

$$\sqrt{m}\left(\begin{pmatrix}\overline{Y}_{.1}\\ \overline{X}_{.1}\\ \overline{Y}_{.2}\\ \overline{X}_{.2}\end{pmatrix} - \begin{pmatrix}E(Y_1)\\ E(X_1)\\ E(Y_2)\\ E(X_2)\end{pmatrix}\right) \longrightarrow^{\mathcal{D}} N_4\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \Sigma^*\right),$$

where, Σ^* is defined above.

The Multivariate Delta method

The delta method (see the subsection 1.3.3), is a method for deriving the asymptotic Normal distribution for any statistical estimator and gives knowledge about the asymptotic variance. For obtaining the asymptotic normality of the estimator $\hat{\mathbf{p}}_m$ via a non-linear transformation, we will introduce to the multivariate delta method $k = 2$, where k is the number of columns of the data matrix. Here, is its notations:

$$\boldsymbol{\theta} = E(\mathbf{H}_i) = (\theta_1, \theta_2, \theta_3, \theta_4)^T, \quad \theta_1 = E(Y_1),\ \theta_2 = E(X_1)$$
$$, \quad \theta_3 = E(Y_2),\ \theta_4 = E(X_2)$$
$$g(\boldsymbol{\theta}) = \left(\frac{\theta_1}{\theta_2}, \frac{\theta_3}{\theta_4}\right),\ g(\boldsymbol{\theta}) : \mathbf{R}^4 \longrightarrow \mathbf{R}^2,$$

is a two dimension vector real-valued function that is continuously differentiable at $\boldsymbol{\theta}$, $\theta_2, \theta_4 > 0$. The matrix of partial derivatives of the function g with respect to the components of $\boldsymbol{\theta}$ is given by

$$\nabla_g^T(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_1}{\partial \theta_3} & \frac{\partial g_1}{\partial \theta_4} \\ \frac{\partial g_2}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_3} & \frac{\partial g_2}{\partial \theta_4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\theta_2} & -\frac{\theta_1}{\theta_2{}^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\theta_4} & -\frac{\theta_3}{\theta_4{}^2} \end{bmatrix},$$

also by plugging in the elements of $\boldsymbol{\theta}$ into the matrix Σ^*, we get the covariance matrix

$$\Sigma^* = \Sigma_{\boldsymbol{\theta}, \boldsymbol{\lambda_0}} = \begin{pmatrix} \theta_1 & \theta_1 & \lambda_0 & \lambda_0 \\ \theta_1 & \theta_2 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \theta_3 & \theta_3 \\ \lambda_0 & \lambda_0 & \theta_3 & \theta_4 \end{pmatrix}.$$

For, $\mathbf{p} = (p_1, p_2)^T$, and according to the delta method with its notation, we have

$$g(\boldsymbol{\theta}) = \mathbf{p}, \quad \text{and } \hat{\mathbf{p}}_m = (\hat{p}_{m1}, \hat{p}_{m2})^T = g(\hat{\boldsymbol{\theta}}_m), \hat{\boldsymbol{\theta}}_m = \overline{\mathbf{H}},$$

with the corresponding estimators $\hat{p}_{mj} = \frac{\overline{Y}_{.j}}{\overline{X}_{.j}}$, $j = 1, 2$. Hence, it follows that

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N_2\left(\mathbf{0}, \nabla_g^T(\boldsymbol{\theta})\Sigma_{\boldsymbol{\theta},\boldsymbol{\lambda_0}}\nabla_g(\boldsymbol{\theta})\right), \tag{2.15}$$

where, the asymptotic covariance matrix of expression (2.15), equals

$$\nabla_g^T(\boldsymbol{\theta})\,\Sigma_{\boldsymbol{\theta},\boldsymbol{\lambda_0}}\,\nabla_g(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\theta_2} & -\frac{\theta_1}{{\theta_2}^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\theta_4} & -\frac{\theta_3}{{\theta_4}^2} \end{bmatrix} \times \begin{pmatrix} \theta_1 & \theta_1 & \lambda_0 & \lambda_0 \\ \theta_1 & \theta_2 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \theta_3 & \theta_3 \\ \lambda_0 & \lambda_0 & \theta_3 & \theta_4 \end{pmatrix}$$

$$\times \begin{bmatrix} \frac{1}{\theta_2} & 0 \\ -\frac{\theta_1}{{\theta_2}^2} & 0 \\ 0 & \frac{1}{\theta_4} \\ 0 & -\frac{\theta_3}{{\theta_4}^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\theta_1}{\theta_2^2} - \frac{2\theta_1^2}{\theta_2^3} + \frac{\theta_1^2}{\theta_2^3} & \frac{\lambda_0}{\theta_2\theta_4} - \frac{\lambda_0\theta_1}{\theta_2^2\theta_4} - \frac{\lambda_0\theta_3}{\theta_2\theta_4^2} + \frac{\lambda_0\theta_1\theta_3}{\theta_2^2\theta_4^2} \\ \frac{\lambda_0}{\theta_2\theta_4} - \frac{\lambda_0\theta_3}{\theta_2\theta_4^2} - \frac{\lambda_0\theta_1}{\theta_2^2\theta_4} + \frac{\lambda_0\theta_1\theta_3}{\theta_2^2\theta_4^2} & \frac{\theta_3}{\theta_4^2} - \frac{2\theta_3^2}{\theta_4^3} + \frac{\theta_3^2}{\theta_4^3} \end{bmatrix},$$

There first diagonal element simplifies to

$$\frac{\theta_1}{\theta_2^2} - \frac{2\theta_1^2}{\theta_2^3} + \frac{\theta_1^2}{\theta_2^3} = \frac{\theta_1}{\theta_2^2} - \frac{\theta_1^2}{\theta_2^3} = \frac{\theta_1}{\theta_2^2}\left(1 - \frac{\theta_1}{\theta_2}\right) = \frac{1}{E(X_1)}p_1(1 - p_1),$$

and similarly, the second diagonal element

$$\frac{\theta_3}{\theta_4^2} - \frac{2\theta_3^2}{\theta_4^3} + \frac{\theta_3^2}{\theta_4^3} = \frac{\theta_3}{\theta_4^2} - \frac{\theta_3^2}{\theta_4^3} = \frac{\theta_3}{\theta_4^2}\left(1 - \frac{\theta_3}{\theta_4}\right) = \frac{1}{E(X_2)}p_2(1 - p_2),$$

as well as, the off-diagonal elements are symmetric, so we have

$$\frac{\lambda_0}{\theta_2\theta_4} - \frac{\lambda_0\theta_1}{\theta_2^2\theta_4} - \frac{\lambda_0\theta_3}{\theta_2\theta_4^2} + \frac{\lambda_0\theta_1\theta_3}{\theta_2^2\theta_4^2} = \frac{\lambda_0}{\theta_2\theta_4}\left[1 - \frac{\theta_1}{\theta_2} - \frac{\theta_3}{\theta_4} + \frac{\theta_1\theta_3}{\theta_2\theta_4}\right] =$$

$$\frac{\lambda_0}{\theta_2\theta_4}\left[\left(1 - \frac{\theta_1}{\theta_2}\right) - \frac{\theta_3}{\theta_4}\left(1 - \frac{\theta_1}{\theta_2}\right)\right] = \frac{\lambda_0}{\theta_2\theta_4}\left(1 - \frac{\theta_1}{\theta_2}\right)\left(1 - \frac{\theta_3}{\theta_4}\right)$$

$$= \frac{\lambda_0}{E(X_1)E(X_2)}\left(1 - p_1\right)\left(1 - p_2\right).$$

Thus, 2.15 can be rewritten as

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N\left(\mathbf{0}, \begin{pmatrix} \frac{p_1(1-p_1)}{E(X_1)} & \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} \\ \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \frac{p_2(1-p_2)}{E(X_2)} \end{pmatrix}\right), \tag{2.16}$$

which is the asymptotic Normal distribution of the estimator $\hat{\mathbf{p}}_m$, where $\lambda_0 = Cov(Y_1, Y_2)$. For statistical inference, it needs to estimate the asymptotic covariance matrix in 2.16, i.e, $asCov(\hat{\mathbf{p}}_m) = \begin{pmatrix} \frac{p_1(1-p_1)}{E(X_1)} & \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} \\ \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \frac{p_2(1-p_2)}{E(X_2)} \end{pmatrix}$, which can be obtained by plugging in the estimator for each individual parameter

$$\widehat{asCov(\hat{\mathbf{p}}_m)} = \begin{pmatrix} \frac{\hat{p}_1(1-\hat{p}_1)}{\overline{X}_{.1}} & \frac{s_{Y_1,Y_2}(1-\hat{p}_1)(1-\hat{p}_2)}{\overline{X}_{.1}\overline{X}_{.2}} \\ \frac{s_{Y_1,Y_2}(1-\hat{p}_1)(1-\hat{p}_2)}{\overline{X}_{.1}\overline{X}_{.2}} & \frac{\hat{p}_2(1-\hat{p}_2)}{\overline{X}_{.2}} \end{pmatrix}, \tag{2.17}$$

where,

$$\overline{X}_{.1} \longrightarrow^{\mathcal{P}} E(X_1), \overline{X}_{.2} \longrightarrow^{\mathcal{P}} E(X_2), \hat{p}_1 \longrightarrow^{\mathcal{P}} p_1, \hat{p}_2 \longrightarrow^{\mathcal{P}} p_2, \text{and } s_{Y_1,Y_2} \longrightarrow^{\mathcal{P}} \lambda_0,$$

consequently the estimator matrix 2.17 is consistent.

2.3.5 Approximate confidence intervals for a linear combination of the proportions

A necessary condition for constructing confidence intervals for a linear combination $\vartheta = \boldsymbol{\alpha}^T\mathbf{p}$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T \geq \mathbf{0}$, is the matrix (2.17) be positive definite or at least positive semi definite (see A.1, or [16]) to ensure $\frac{asVar(\widehat{\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m})}{m} \geq 0$, otherwise we have to truncate the corresponding intervals by taking only the positive variances. Hence, to be on the safe side from the undefined s.e's ($s.e = \sqrt{\frac{as\widehat{Var}}{m}}$) during the confidence intervals evaluation, one should take only the positive variances, this will be explained in chapter 4. Obviously

$$\begin{aligned} asVar(\hat{\vartheta}_m) &= asVar(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m) = \boldsymbol{\alpha}^T asVar(\hat{\mathbf{p}}_m)\boldsymbol{\alpha} \\ &= (\alpha_1 \quad \alpha_2) \begin{pmatrix} \frac{p_1(1-p_1)}{E(X_1)} & \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} \\ \frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \frac{p_2(1-p_2)}{E(X_2)} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \\ &= \alpha_1^2\frac{p_1(1-p_1)}{E(X_1)} + \alpha_2^2\frac{p_2(1-p_2)}{E(X_2)} + 2\alpha_1\alpha_2\frac{\lambda_0(1-p_1)(1-p_2)}{E(X_1)E(X_2)}, \end{aligned}$$

it follows that, the standard error of the linear combination $\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m$ (if defined) is given by

$$\begin{aligned} s.e(\hat{\vartheta}_m) &= \sqrt{\frac{1}{m}\widehat{asVar(\hat{\vartheta}_m)}} = \sqrt{\frac{1}{m}\widehat{asVar(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m)}} \\ &= \sqrt{\frac{1}{m}\left(\alpha_1^2\frac{\hat{p}_1(1-\hat{p}_1)}{\overline{X}_{.1}} + \alpha_2^2\frac{\hat{p}_2(1-\hat{p}_2)}{\overline{X}_{.2}} + 2\alpha_1\alpha_2\frac{s_{Y_1,Y_2}(1-\hat{p}_1)(1-\hat{p}_2)}{\overline{X}_{.1}\overline{X}_{.2}}\right)}. \end{aligned}$$

Finally, the approximate $(1-\alpha)\,\%$ asymptotic normal confidence interval by the normal quantile $z_{1-\frac{\alpha}{2}}$ for the linear combination ϑ is given by

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}}\, s.e(\hat{\vartheta}_m)\right],$$

or, the proposed conservative confidence interval by the t-quantile

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

where, $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ percentile of a standard normal distribution, as well as $t_{(m-1,1-\frac{\alpha}{2})}$ is the $(1-\frac{\alpha}{2})$ percentile of a $t-$distribution with $(m-1)$ degrees of freedom.

Chapter 3

Multivariate data analysis

3.1 Introduction

In the multivariate analysis being extending the analysis of a bivariate count data to the analysis of a multivariate count data. It will be assumed that there are k components of marked count data. Clearly, we have $2k$ dimension random sample of size m drawn from an infinite population, i.e, $(Y_{i1}, X_{i1}), ..., (Y_{ik}, X_{ik}), i = 1, ..., m$ where $\mathbf{H}_i = ((Y_{i1}, X_{i1}), ..., (Y_{ik}, X_{ik}))$ are $i.i.d$ sets of k pairs of random variables. The data are displayed in a matrix of dimension $m \times 2k$ of marked count data, where the m rows represent the individuals, and $2k$ columns represent the k dimension of the pairs count data, with the restrictions, $0 \leq Y_{ij} \leq X_{ij}$, $\forall i, j,\ j = 1, ..., k$.
The estimator $\hat{p}_j = \frac{\sum_{i=1}^{m} Y_{ij}}{\sum_{i=1}^{m} X_{ij}}$ is the commonly sample proportion corresponding to the proportion p_j, where $p_j = \frac{E(Y_{ij})}{E(X_{ij})}$, $E(\hat{p}_j) = p_j$,
$0 \leq p_j \leq 1,\ j = 1, 2, ..., k$.

In the following section, we will analyze the marked count data matrix using the multivariate SUR model, assuming that the relation between each of the response random variables $Y_{i1}, ..., Y_{ik}$ and the corresponding variables $X_{i1}, ..., X_{ik}$ for the i^{th} individual is linearly modeled. We assume that $X_{ij} > 0$ almost surely. The stacked SUR equations will be considered to obtain optimal estimators if exists for the (coefficients) proportions $p_j, j = 1, ..., k$ of the SUR model, namely, the LSE's including there asymptotic properties.

3.2 The SUR (Seemingly Unrelated Regression) Model

The SUR model (k correlated regression equations) based on the m observations can be modeled next:

3.2.1 The multiple linear model

Consider the k linear equations

$$\begin{aligned} Y_{i1} =& x_{i1}p_{1} + \epsilon_{i1} \\ \vdots & \quad \vdots \\ Y_{ik} =& x_{ik}p_{k} + \epsilon_{ik} \end{aligned}$$

, with the assumptions: $E(\epsilon_{ij}) = 0$, and with variances proportional to x_{ij}, where x_{ij}, are the fixed variables, i.e

$$\begin{aligned} &Var(\epsilon_{ij}) = \sigma_j^2 x_{ij},\ Cov(\epsilon_{ij}, \epsilon_{ij'}) = \sigma_{jj'}\sqrt{x_{ij}x_{ij'}},\ \forall j \neq j^{'}, \\ &Cov(\epsilon_{ij}, \epsilon_{i^{'}j}) = 0,\ \forall i \neq i^{'},\ i, i^{'} = 1, ..., m, j, j^{'} = 1, ..., k. \end{aligned}$$

We merge these linear models compactly into a single multivariate linear model (for the i^{th} observation)

$$\mathbf{Y}_i = \mathbf{X}_i \mathbf{p} + \boldsymbol{\epsilon}_i,\ i = 1, ..., m, \tag{3.1}$$

where, the response variable, $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{ik})^T$, the observed design matrix

$$\mathbf{X}_i = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_{ik} \end{pmatrix},$$

and the model coefficients (proportions), $\mathbf{p} = (p_1, \cdots, p_k)^T$, as well as the error component, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{ik})^T,\ \ i = 1, ..., m$. The error vector $\boldsymbol{\epsilon}_i$ has the variance-covariance matrix given by

$$\Sigma_i = Var(\boldsymbol{\epsilon}_i) = \begin{pmatrix} \sigma_1^2 x_{i1} & \sigma_{12}\sqrt{x_{i1}x_{i2}} & \cdots & \sigma_{1k}\sqrt{x_{i1}x_{ik}} \\ \sigma_{12}\sqrt{x_{i1}x_{i2}} & \sigma_2^2 x_{i2} & \cdots & \sigma_{2k}\sqrt{x_{i2}x_{ik}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k}\sqrt{x_{i1}x_{ik}} & \sigma_{2k}\sqrt{x_{i2}x_{ik}} & \cdots & \sigma_k^2 x_{ik} \end{pmatrix},$$

where, $Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_{i'}) = \mathbf{0}_{k\times k}$, when, $i \neq i^{'} = 1, ..., m$.

3.2.2 The weighted multiple linear model

In the same manner as in the bivariate linear model, the model 3.1 will be standardized by the linear transformation $\mathbf{A}_i\mathbf{Y}_i = \mathbf{A}_i\mathbf{X}_i\mathbf{p} + \mathbf{A}_i\boldsymbol{\epsilon}_i$, where the transformation matrix

$$\mathbf{A}_i = \mathbf{X}_i^{-\frac{1}{2}} = \text{diag}\left(x_{ij}^{-\frac{1}{2}}\right)_{j=1,\dots,k} = \begin{pmatrix} \frac{1}{\sqrt{x_{i1}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{x_{i2}}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{x_{ik}}} \end{pmatrix},$$

(given that $x_{ij} > 0 \; \forall i, j, \; i = 1, ..., m, j = 1, ..., k$) to obtain the weighted multivariate linear model

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{X}}_i\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}_i, \; i = 1, ..., m, \tag{3.2}$$

one can also observe that $\mathbf{A}_i\mathbf{X}_i\mathbf{A}_i^T = I_k$, where I_k is the k dimension identity matrix, the weighed response variable

$$\widetilde{\mathbf{Y}}_i = \mathbf{A}_i\mathbf{Y}_i = \left(\frac{Y_{i1}}{\sqrt{x_{i1}}}, \cdots, \frac{Y_{ik}}{\sqrt{x_{ik}}}\right)^T = (\widetilde{Y}_{i1}, ..., \widetilde{Y}_{ik})^T,$$

and the weighted error component

$$\widetilde{\boldsymbol{\epsilon}}_i = \mathbf{A}_i\boldsymbol{\epsilon}_i = \left(\frac{\epsilon_{i1}}{\sqrt{x_{i1}}}, \cdots, \frac{\epsilon_{ik}}{\sqrt{x_{ik}}}\right)^T = (\widetilde{\epsilon}_{i1}, \cdots, \widetilde{\epsilon}_{ik})^T,$$

as well as, the weighted design matrix

$$\widetilde{\mathbf{X}}_i = \mathbf{A}_i\mathbf{X}_i = \begin{pmatrix} \widetilde{x}_{i1} & 0 & \cdots & 0 \\ 0 & \widetilde{x}_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \widetilde{x}_{ik} \end{pmatrix} = diag\left(\widetilde{x}_{ij}\right)_{j=1,\dots,k},$$

where, $\widetilde{x}_{ij} = \sqrt{x_{ij}}$, provided that, $x_{ij} > 0, \; \forall i, j, \; i = 1, ..., m, \; j = 1, ..., k$.
The covariance of the weighted error vector $\widetilde{\boldsymbol{\epsilon}}_i$ is given by

$$\begin{aligned} \widetilde{\Sigma}_{k\times k} &= Cov(\widetilde{\boldsymbol{\epsilon}}_i) = \mathbf{A}_i Cov(\boldsymbol{\epsilon}_i)\mathbf{A}_i^T \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_k^2 \end{pmatrix} \; (homoscedastic\ error\ vectors), \end{aligned}$$

and, $Cov(\widetilde{\boldsymbol{\epsilon}}_i, \widetilde{\boldsymbol{\epsilon}}_{i'}) = \mathbf{0}_{k\times k}, \; i \neq i'$.

3.2.3 The SUR Model

And thus, one can stacking the multivariate linear equations 3.2 in the form

$$\begin{pmatrix} \widetilde{\mathbf{Y}}_1 \\ \widetilde{\mathbf{Y}}_2 \\ \vdots \\ \widetilde{\mathbf{Y}}_m \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \sqrt{x_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{x_{12}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{x_{1k}} \end{pmatrix} \\ \begin{pmatrix} \sqrt{x_{21}} & 0 & \cdots & 0 \\ 0 & \sqrt{x_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{x_{2k}} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \sqrt{x_{m1}} & 0 & \cdots & 0 \\ 0 & \sqrt{x_{m2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{x_{mk}} \end{pmatrix} \end{pmatrix} \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} + \begin{pmatrix} \widetilde{\boldsymbol{\epsilon}}_1 \\ \widetilde{\boldsymbol{\epsilon}}_2 \\ \vdots \\ \widetilde{\boldsymbol{\epsilon}}_m \end{pmatrix},$$

the model can be compressed in to a single model as (see also [21] , for more details)

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}, \tag{3.3}$$

where, the $mk \times 1$ dimension SUR model response vector $\widetilde{\mathbf{Y}} = \left(\widetilde{\mathbf{Y}}_1^T, \widetilde{\mathbf{Y}}_2^T, \cdots, \widetilde{\mathbf{Y}}_m^T\right)^T$, the $mk \times k$ dimension design matrix $\widetilde{\mathbf{X}} = \left(\widetilde{\mathbf{X}}_1^T, \widetilde{\mathbf{X}}_2^T, \cdots, \widetilde{\mathbf{X}}_m^T\right)^T$, so $k \times 1$ dimension SUR model parameter vector $\mathbf{p} = (p_1, ..., p_k)^T$, and the $mk \times 1$ dimension SUR model error vector $\widetilde{\boldsymbol{\epsilon}} = \left(\widetilde{\boldsymbol{\epsilon}}_1^T, \widetilde{\boldsymbol{\epsilon}}_2^T, \cdots, \widetilde{\boldsymbol{\epsilon}}_m^T\right)^T$ of the $i.i.d$ error components $\widetilde{\boldsymbol{\epsilon}}_i$, as well as the $mk \times mk$ dimension covariance structure of the SUR model error vector is given by

$$\widetilde{\boldsymbol{\Sigma}} = Cov(\widetilde{\boldsymbol{\epsilon}}) = \begin{pmatrix} \widetilde{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \widetilde{\Sigma} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \widetilde{\Sigma} \end{pmatrix} = I_m \otimes \widetilde{\Sigma},$$

i.e, $Var(\widetilde{\boldsymbol{\epsilon}}_i) = \widetilde{\Sigma}_{k\times k}$, and $Cov(\widetilde{\boldsymbol{\epsilon}}_i, \widetilde{\boldsymbol{\epsilon}}_{i'}) = \mathbf{0}_{k\times k}(\forall i \neq i',\ i, i' = 1, ..., m)$, where $\mathbf{0}_{k\times k}$ is the square matrix of dimension $k \times k$ of zero's.

3.2.4 Estimation of the parameter vector p in the SUR model

The ordinary least squares estimator of the model parameter (proportion $\mathbf{p}$) derived from the SUR model 3.3 is given by

$$\hat{\mathbf{p}}_{OLS} = (\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}} = \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i = \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^2\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i$$

$$= \begin{pmatrix} \sum_{i=1}^m x_{i1} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^m x_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{i=1}^m x_{ik} \end{pmatrix}^{-1} \sum_{i=1}^{m}\begin{pmatrix} \sqrt{x_{i1}}\widetilde{Y}_{i1} \\ \vdots \\ \sqrt{x_{ik}}\widetilde{Y}_{ik} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sum_{i=1}^m \sqrt{x_{i1}}\widetilde{Y}_{i1}}{\sum_{i=1}^m x_{i1}} \\ \vdots \\ \frac{\sum_{i=1}^m \sqrt{x_{ik}}\widetilde{Y}_{ik}}{\sum_{i=1}^m x_{ik}} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^m Y_{i1}}{\sum_{i=1}^m x_{i1}} \\ \vdots \\ \frac{\sum_{i=1}^m Y_{ik}}{\sum_{i=1}^m x_{ik}} \end{pmatrix} = \hat{\mathbf{p}},$$

which results in the ratio estimator vector, where $Y_{ij} = \sqrt{x_{i1}}\widetilde{Y}_{ij},\ j = 1, ..., k.$

3.2.5 Properties of the estimator $\hat{\mathbf{p}}_{OLS}$ (ratio vector $\hat{\mathbf{p}}$)

Unbiasedness

The expectation of the ratio estimator $\hat{p}_j$ is given by

$$E\left(\hat{p}_j\right) = E\left(\frac{\sum_{i=1}^m \widetilde{Y}_{ij}\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}}\right) = \frac{\sum_{i=1}^m \sqrt{x_{ij}}E\left(\widetilde{Y}_{ij}\right)}{\sum_{i=1}^m x_{ij}} = \frac{\sum_{i=1}^m \sqrt{x_{ij}}p_j\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}} = \frac{p_j\sum_{i=1}^m x_{ij}}{\sum_{i=1}^m x_{ij}}$$
$$= p_j,$$

$j = 1, ..., k.$

Dispersion

The variance of the ratio estimator $\hat{p}_j$ is obtained by

$$Var\left(\hat{p}_j\right) = Var\left(\frac{\sum_{i=1}^m \widetilde{Y}_{ij}\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}}\right) = \frac{1}{\left(\sum_{i=1}^m x_{ij}\right)^2}\sum_{i=1}^{m} x_{ij}Var(\widetilde{Y}_{ij}) = \frac{\sigma_j^2\sum_{i=1}^m x_{ij}}{\left(\sum_{i=1}^m x_{ij}\right)^2}$$
$$= \frac{\sigma_j^2}{\sum_{i=1}^m x_{ij}}.$$

Furthermore, the covariance of the ratio estimators $\hat{p}_j, \hat{p}_{j'}, j \neq j'$, and $j, j' = 1, ..., k$, will be obtained as

$$\begin{aligned} Cov\left(\hat{p}_j, \hat{p}_{j'}\right) &= Cov\left(\frac{\sum_{i=1}^m \widetilde{Y}_{ij}\sqrt{x_{ij}}}{\sum_{i=1}^m x_{ij}}, \frac{\sum_{i=1}^m \widetilde{Y}_{ij'}\sqrt{x_{ij'}}}{\sum_{i=1}^m x_{ij'}}\right) \\ &= \frac{1}{\sum_{i=1}^m x_{ij}\sum_{i=1}^m x_{ij'}} Cov\left(\sum_{i=1}^m \sqrt{x_{ij}}\widetilde{Y}_{ij}, \sum_{i=1}^m \sqrt{x_{ij'}}\widetilde{Y}_{ij'}\right) \\ &= \frac{1}{\sum_{i=1}^m X_{ij}\sum_{i=1}^m x_{ij'}} \sum_{i=1}^m Cov\left(\widetilde{Y}_{ij}, \widetilde{Y}_{ij'}\right)\sqrt{x_{ij}x_{ij'}} \\ &= \frac{\sigma_{jj'}\sum_{i=1}^m \sqrt{x_{ij}x_{ij'}}}{\sum_{i=1}^m x_{ij}\sum_{i=1}^m x_{ij'}}. \end{aligned}$$

Consequently, the covariance matrix of the ratio vector $\hat{\mathbf{p}}$ can be established as

$$\Sigma_{\mathbf{x}} = Cov\left(\hat{\mathbf{p}}\right) = Cov\left(\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}}\right) = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{X}}\right)\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}$$

$$= \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1} =$$

$$\begin{pmatrix} \sum_{i=1}^m x_{i1} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^m x_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{i=1}^m x_{ik} \end{pmatrix}^{-1} \times$$

$$\begin{pmatrix} \sigma_1^2\sum_{i=1}^m x_{i1} & \sigma_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}} & \cdots & \sigma_{1k}\sum_{i=1}^m \sqrt{x_{i1}x_{ik}} \\ \sigma_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}} & \sigma_2^2\sum_{i=1}^m x_{i2} & \cdots & \sigma_{2k}\sum_{i=1}^m \sqrt{x_{i2}x_{ik}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k}\sum_{i=1}^m \sqrt{x_{i1}x_{ik}} & \sigma_{2k}\sum_{i=1}^m \sqrt{x_{i2}x_{ik}} & \cdots & \sigma_k^2\sum_{i=1}^m x_{ik} \end{pmatrix} \times$$

$$\begin{pmatrix} \sum_{i=1}^m x_{i1} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^m x_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{i=1}^m x_{ik} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{\sigma_1^2}{\sum_{i=1}^m x_{i1}} & \frac{\sigma_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}} & \cdots & \frac{\sigma_{1k}\sum_{i=1}^m \sqrt{x_{i1}x_{ik}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{ik}} \\ \frac{\sigma_{12}\sum_{i=1}^m \sqrt{x_{i1}x_{i2}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{i2}} & \frac{\sigma_2^2}{\sum_{i=1}^m x_{i2}} & \cdots & \frac{\sigma_{2k}\sum_{i=1}^m \sqrt{x_{i2}x_{ik}}}{\sum_{i=1}^m x_{i2}\sum_{i=1}^m x_{ik}} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\sigma_{1k}\sum_{i=1}^m \sqrt{x_{i1}x_{ik}}}{\sum_{i=1}^m x_{i1}\sum_{i=1}^m x_{ik}} & \frac{\sigma_{2k}\sum_{i=1}^m \sqrt{x_{i2}x_{ik}}}{\sum_{i=1}^m x_{i2}\sum_{i=1}^m x_{ik}} & \cdots & \frac{\sigma_k^2}{\sum_{i=1}^m x_{ik}} \end{pmatrix}. \tag{3.4}$$

As mentioned in chapter 2, the covariance matrix 3.4 is larger than the covariance matrix $\left(\widetilde{\mathbf{X}}^T\, \widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{X}}\right)^{-1}$, when we use the weighted least squares estimator $\hat{\mathbf{p}}_{WLS}$.

Recall also that, the conditional consistent estimators based on m residuals for the corresponding diagonal and off-diagonal entries, namely σ_j^2 and $\sigma_{jj'}$ respectively for the covariance matrix of the ratio vector $\hat{\mathbf{p}}$ are given by:

$$\begin{aligned}
\hat{\sigma}_j^2 = s_j^2 &= \frac{1}{m-1}\sum_{i=1}^{m}\frac{\left(Y_{ij}-\hat{p}_j x_{ij}\right)^2}{x_{ij}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij}-\hat{p}_j\sqrt{x_{ij}}\right)^2,\\
\hat{\sigma}_{jj'} = s_{jj'} &= \frac{1}{m-1}\sum_{i=1}^{m}\frac{\left(Y_{ij}-\hat{p}_j x_{ij}\right)\left(Y_{ij'}-\hat{p}_{j'} x_{ij'}\right)}{\sqrt{x_{ij}}\sqrt{x_{ij'}}}\\
&= \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij}-\hat{p}_j\sqrt{x_{ij}}\right)\left(\widetilde{Y}_{ij'}-\hat{p}_{j'}\sqrt{x_{ij'}}\right),
\end{aligned}$$

provided that, $x_{ij}, x_{ij'} > 0\ \forall j \neq j',\ j, j^{'} = 1, ..., k, i = 1, ..., m,\ \widetilde{Y}_{ij} = \frac{Y_{ij}}{\sqrt{x_{ij}}},\ \widetilde{Y}_{ij'} = \frac{Y_{ij'}}{\sqrt{x_{ij'}}}$. They were well demonstrated as seen in chapter 2 that, s_j^2 is consistent unbiased estimator of the corresponding parameter σ_j^2, while $s_{jj'}$ is consistent but only asymptotically unbiased estimator of the corresponding $\sigma_{jj'}$, $j \neq j'$.

Exact and Asymptotic distributions

If one assumed that, $\boldsymbol{\epsilon}_i \sim N_k\left(\mathbf{0}, \Sigma_i\right)$, so $\epsilon_{ij} \sim N(0, \sigma_j^2 x_{ij}),\ j = 1, ..., k, \forall i = 1, ..., m$, where, $N_k\left(\mathbf{0}, \Sigma_i\right)$ denotes, the K-variate Normal distribution with mean vector $\mathbf{0} = (0, ..., 0)^T$, and with the symmetric covariance matrix Σ_i, this is called the MNLM (Multivariate Normal linear model).

Further, given the design vectors $\widetilde{\mathbf{x}}_i^T = (\widetilde{x}_{i1}, ..., \widetilde{x}_{ik})$, are fixed, $i = 1, ..., m$, the weighted error vectors $\widetilde{\boldsymbol{\epsilon}}_i$ are $i.i.d$, k-variate Normal random vectors, i.e
$\widetilde{\boldsymbol{\epsilon}}_i \sim N_k(\mathbf{0}_k, \widetilde{\Sigma})\ \forall i = 1, ..., m$, where $\mathbf{0}_k = (0, ..., 0)^T$. Also, one may say that, given $\widetilde{\mathbf{x}}_i^T$ the $i.d$ weighted random component $\widetilde{\mathbf{Y}}_i$ has the multivariate Normal distribution i.e, $\widetilde{\mathbf{Y}}_i \sim N_k\left(\begin{pmatrix}\widetilde{x}_{i1}p_1\\ \vdots \\ \widetilde{x}_{ik}p_k\end{pmatrix}, \widetilde{\Sigma}\right)$, $\widetilde{x}_{ij} = \sqrt{x_{ij}}, x_{ij} > 0$. The distribution of $\hat{\mathbf{p}} \sim N_k\left(\mathbf{p}, \Sigma_{\mathbf{x}}\right)$, where $\Sigma_{\mathbf{x}}$ is given by the matrix 3.4 , as well as $\mathbf{p} = (p_1, \cdots, p_k)^T$. Asymptotically

one can obtain

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N_k\left(\mathbf{0}, \begin{pmatrix} \frac{\sigma_1^2}{\mu_1} & \frac{\sigma_{12}\mu_{12}}{\mu_1\mu_2} & \dots & \frac{\sigma_{1k}\mu_{1k}}{\mu_1\mu_k} \\ \frac{\sigma_{12}\mu_{12}}{\mu_1\mu_2} & \frac{\sigma_2^2}{\mu_2} & \dots & \frac{\sigma_{2k}\mu_{2k}}{\mu_2\mu_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1k}\mu_{1k}}{\mu_1\mu_k} & \frac{\sigma_{2k}\mu_{2k}}{\mu_2\mu_k} & \dots & \frac{\sigma_k^2}{\mu_k} \end{pmatrix}\right),$$

provided that, $\frac{1}{m}\sum_{i=1}^{m} x_{ij} \longrightarrow \mu_j$, $\frac{1}{m}\sum_{i=1}^{m} \sqrt{x_{ij}x_{ij'}} \longrightarrow \mu_{jj'}$, where $\mu_j, \mu_{jj'}$, $j < j' = 1, ..., k$, are constants.

3.2.6 Multivariate asymptotic normality of the ratio vector $\hat{\mathbf{p}}_m$

In a similar way as in the bivariate case, it will be assumed that the random error components $\widetilde{\boldsymbol{\epsilon}}_i$ are not normally distributed, but are *i.i.d* random vectors, i.e, $E\left(\widetilde{\boldsymbol{\epsilon}}_i\right) = \mathbf{0}_k$, and $Cov(\widetilde{\boldsymbol{\epsilon}}_i) = \widetilde{\Sigma}_{k\times k}$, $i = 1, ..., m$.
Moreover, under conditions on the weighted design matrices $\widetilde{\mathbf{X}}_i$, we will show that in large sample size m, $\hat{\mathbf{p}}_m$ has the multivariate Normal asymptotic distribution. These conditions, namely the pairs
$(\mathbf{X}_i, \mathbf{Y}_i)$ are *i.i.d* $\Rightarrow$ the pairs $(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Y}}_i)$ are also *i.i.d*, where, $\widetilde{\mathbf{X}}_i = diag\left(\widetilde{X}_{ij}\right)_{j=1,...,k}$, $\widetilde{X}_{ij} = \sqrt{X_{ij}}$, $X_{ij} > 0$, $\forall i, j$, $i = 1, ..., m, j = 1, ..., k$. As well as, $E(\widetilde{X}_{ij}\widetilde{X}_{ij'})$ i.e, $E\left(\sqrt{X_{ij}X_{ij'}}\right)$ exists, $\forall j, j' = 1, ..., k$.
So again, one can rewrite $\hat{\mathbf{p}}_{OLS}$ as

$$\begin{aligned} \hat{\mathbf{p}}_{OLS} = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{Y}} &= \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{Y}}_i \\ &= \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\left(\widetilde{\mathbf{X}}_i\mathbf{p} + \widetilde{\boldsymbol{\epsilon}}_i\right) \\ &= \left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\mathbf{p} + \sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right), \end{aligned}$$

and this can be rewritten as

$$\sqrt{m}\,(\hat{\mathbf{p}}_{OLS} - \mathbf{p}) = \left(\frac{1}{m}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right)$$
$$= \left(\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} X_{i1} & 0 & \cdots & 0 \\ 0 & X_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X_{ik}\end{pmatrix}\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\begin{pmatrix}\widetilde{X}_{i1}\widetilde{\epsilon}_{i1} \\ \vdots \\ \widetilde{X}_{ik}\widetilde{\epsilon}_{ik}\end{pmatrix}\right). \tag{3.5}$$

To derive the multivariate asymptotic distribution it needs to investigate first, the denominator matrix in expression 3.5 is consistent, and second the numerator obeys the multivariate central limit theorem.
Thus in large m, and by following the LLN, the denominator of 3.5 is consistent. Since

$$\left(\frac{1}{m}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1} \longrightarrow^{\mathcal{P}} \left(E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)\right)^{-1} \tag{3.6}$$

provided that, $\left(E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\right) > 0$, subsequently, eq. 3.6 can be written as

$$\left(\frac{1}{m}\sum_{i=1}^{m}(diag(X_{ij}))_{j=1,\ldots,k}\right)^{-1} \longrightarrow^{\mathcal{P}} ((diag(E(X_{ij})))_{j=1,\ldots,k})^{-1} > 0,$$

or

$$\left(\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} X_{i1} & 0 & \cdots & 0 \\ 0 & X_{i2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X_{ik}\end{pmatrix}\right)^{-1} \longrightarrow^{\mathcal{P}} \begin{pmatrix} E(X_{i1}) & 0 & \cdots & 0 \\ 0 & E(X_{i2}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & E(X_{ik})\end{pmatrix}^{-1}.$$

The numerator obeys the Multivariate Central limit theorem

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i \longrightarrow^{\mathcal{D}} N\left(\mathbf{0}_k, E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i\right)\right)$$
$$\equiv N\left(\mathbf{0}_k, \left(\sigma_{jj'}E\left(\sqrt{X_{ij}X_{ij'}}\right)\right)_{j,j'=1,\ldots,k}\right), \tag{3.7}$$

where, the asymptotic covariance

$$\begin{aligned} Cov\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i\right) &= E\left(Cov\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1},...,\widetilde{X}_{ik}\right)\right) + Cov\left(E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1},...,\widetilde{X}_{ik}\right)\right) \\ &= E\left(\widetilde{\mathbf{X}}_i^T Cov\left(\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1},...,\widetilde{X}_{ik}\right)\widetilde{\mathbf{X}}_i\right) = E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\Sigma}\widetilde{\mathbf{X}}_i\right), \end{aligned}$$

as

$$E\left(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1},...,\widetilde{X}_{ik}\right) = \widetilde{\mathbf{X}}_i^T E\left(\widetilde{\boldsymbol{\epsilon}}_i \mid \widetilde{X}_{i1},...,\widetilde{X}_{ik}\right) = \mathbf{0}_k.$$

The asymptotic covariance in the consequence 3.7 can be derived as following: From the MVSUR model 3.3, and for $i = 1, ..., m$, we have

$$\begin{aligned} Cov(\widetilde{\mathbf{X}}_i^T\widetilde{\boldsymbol{\epsilon}}_i) &= Cov\begin{pmatrix} \sqrt{X_{i1}}\widetilde{\epsilon}_{i1} \\ \vdots \\ \sqrt{X_{ik}}\widetilde{\epsilon}_{ik} \end{pmatrix} = \left(Cov\left(\sqrt{X_{ij}}\widetilde{\epsilon}_{ij}, \sqrt{X_{ij'}}\widetilde{\epsilon}_{ij'}\right)\right)_{j,j'=1,...,k} \\ &= \left(E\left(\sqrt{X_{ij}X_{ij'}}\widetilde{\epsilon}_{ij}\widetilde{\epsilon}_{ij'}\right)\right)_{j,j'=1,...,k} - 0 = \left(E(\sqrt{X_{ij}X_{ij'}})E\left(\widetilde{\epsilon}_{ij}\widetilde{\epsilon}_{ij'}\right)\right)_{j,j'=1,...,k} \\ &= \left(\sigma_{jj'}E(\sqrt{X_{ij}X_{ij'}})\right)_{j,j'=1,...,k}, \text{ as,} \quad E(\widetilde{\epsilon}_{ij}) = 0, E(\widetilde{\epsilon}_{ij'}) = 0. \end{aligned}$$

It follows that

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\begin{pmatrix} \sqrt{X_{i1}}\widetilde{\epsilon}_{i1} \\ \vdots \\ \sqrt{X_{ik}}\widetilde{\epsilon}_{ik} \end{pmatrix} \longrightarrow^{\mathcal{D}}$$

$$N\left(\mathbf{0}_k, \begin{pmatrix} \sigma_1^2E(X_{i1}) & \sigma_{12}E(\sqrt{X_{i1}X_{i2}}) & \cdots & \sigma_{1k}E(\sqrt{X_{i1}X_{ik}}) \\ \sigma_{12}E(\sqrt{X_{i1}X_{i2}}) & \sigma_2^2E(X_{i2}) & \cdots & \sigma_{2k}E(\sqrt{X_{i2}X_{ik}}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k}E(\sqrt{X_{i1}X_{ik}}) & \sigma_{2k}E(\sqrt{X_{i2}X_{ik}}) & \cdots & \sigma_k^2E(X_{ik}) \end{pmatrix}\right).$$

Therefore (with help of the known Slutsky's lemma), one can obtain the asymptotic Normal distribution of the sample ratio estimator $\hat{\mathbf{p}}_m$, hence the expression 3.5 can be rewritten as

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N(\mathbf{0}, \begin{pmatrix} E(X_1) & 0 & \cdots & 0 \\ 0 & E(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & E(X_k) \end{pmatrix}^{-1} \times$$

$$\begin{pmatrix} \sigma_1^2 E(X_1) & \sigma_{12}E(\sqrt{X_1X_2}) & \cdots & \sigma_{1k}E(\sqrt{X_1X_k}) \\ \sigma_{12}E(\sqrt{X_1X_2}) & \sigma_2^2 E(X_2) & \cdots & \sigma_{2k}E(\sqrt{X_2X_k}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k}E(\sqrt{X_1X_k}) & \sigma_{2k}E(\sqrt{X_2X_k}) & \cdots & \sigma_k^2 E(X_k) \end{pmatrix} \times$$

$$\begin{pmatrix} E(X_1) & 0 & \cdots & 0 \\ 0 & E(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & E(X_k) \end{pmatrix}^{-1})$$

$$\equiv N(\mathbf{0}, \begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \cdots & \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} \\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} & \cdots & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} & \cdots & \frac{\sigma_k^2}{E(X_k)} \end{pmatrix}) \tag{3.8}$$

is the multivariate asymptotic Normal distribution of the ratio estimator vector $\hat{\mathbf{p}}_m$ with the asymptotic covariance matrix

$$\Sigma_{\mathbf{p}} = \begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \cdots & \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} \\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} & \cdots & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} & \cdots & \frac{\sigma_k^2}{E(X_k)} \end{pmatrix}.$$

3.2.7 Approximate confidence intervals for the linear combination $\vartheta = \alpha^T\mathbf{p}$

By recalling the last consequence

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N(\mathbf{0}, \Sigma_{\mathbf{p}}), \tag{3.9}$$

and again by applying the Cramer-Wold device, so Cramer-Wold device shows that the exp 3.9 holds iff, $\forall \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^T \in \mathbf{R}^k$ such that $\| \boldsymbol{\alpha} \| > 0$, so we have

$$\begin{aligned}
\sqrt{m}(\hat{\vartheta}_m - \vartheta) &= \sqrt{m}\left(\boldsymbol{\alpha}^T \hat{\mathbf{p}}_m - \boldsymbol{\alpha}^T \mathbf{p}\right) \xrightarrow{\mathcal{D}} N\left(0, \boldsymbol{\alpha}^T \Sigma_{\mathbf{p}} \boldsymbol{\alpha}\right) \\
&\equiv N\left(0, (\alpha_1 \cdots \alpha_k)\begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \cdots & \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} \\ \frac{\sigma_{12}E(\sqrt{X_1X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} & \cdots & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1k}E(\sqrt{X_1X_k})}{E(X_1)E(X_k)} & \frac{\sigma_{2k}E(\sqrt{X_2X_k})}{E(X_2)E(X_k)} & \cdots & \frac{\sigma_k^2}{E(X_k)} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}\right) \\
&\equiv N\left(0, \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{E(X_j)} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}E(\sqrt{X_j X_{j'}})}{E(X_j)E(X_{j'})}\right), \qquad (3.10)
\end{aligned}$$

and hence, the asymptotic variance of $\hat{\vartheta}_m = \boldsymbol{\alpha}^T \hat{\mathbf{p}}_m$ is given by

$$\sigma_\vartheta^2 = \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{E(X_j)} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}E(\sqrt{X_j X_{j'}})}{E(X_j)E(X_{j'})}, \qquad (3.11)$$

which is, the asymptotic variance of $\hat{\vartheta}_m$, when the error components are Normally distributed.

To obtain the approximate confidence intervals for the linear combination of the proportions, it needs to estimate the asymptotic covariance matrix $\Sigma_{\mathbf{p}}$ by plugging in estimates for each corresponding individual parameter, which are

$$\overline{X}_{.j} = \frac{1}{m}\sum_{i=1}^{m} X_{ij},\ \overline{X}_{.jj'} = \frac{1}{m}\sum_{i=1}^{m}\sqrt{X_{ij}X_{ij'}},\ j \neq j', j, j' = 1, ..., k,$$

as well as

$$\begin{aligned}
s_j^2 &= \frac{1}{m-1}\sum_{i=1}^{m}\frac{\left(Y_{ij} - \hat{p}_j X_{ij}\right)^2}{X_{ij}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij} - \hat{p}_j\sqrt{X_{ij}}\right)^2 \\
s_{jj'} &= \frac{1}{m-1}\sum_{i=1}^{m}\frac{\left(Y_{ij} - \hat{p}_j X_{ij}\right)\left(Y_{ij'} - \hat{p}_{j'} X_{ij'}\right)}{\sqrt{X_{ij}}\sqrt{X_{ij'}}} \\
&= \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij} - \hat{p}_j\sqrt{X_{ij}}\right)\left(\widetilde{Y}_{ij'} - \hat{p}_{j'}\sqrt{X_{ij'}}\right),
\end{aligned}$$

provided that,
$X_{ij}, X_{ij'} > 0 \; \forall j \neq j', \; j, j^{'} = 1, ..., k, i = 1, ..., m, \; \widetilde{Y}_{ij} = \frac{Y_{ij}}{\sqrt{X_{ij}}}, \; \widetilde{Y}_{ij'} = \frac{Y_{ij'}}{\sqrt{X_{ij'}}}$, and hence

$$\widehat{\Sigma}_{\mathbf{p}} = \begin{pmatrix} \frac{s_1^2}{\overline{X}_{.1}} & \frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}} & \cdots & \frac{s_{1k}\overline{X}_{.1k}}{\overline{X}_{.1}\overline{X}_{.k}} \\ \frac{s_{12}\overline{X}_{.12}}{\overline{X}_{.1}\overline{X}_{.2}} & \frac{s_2^2}{\overline{X}_{.2}} & \cdots & \frac{s_{2k}\overline{X}_{.2k}}{\overline{X}_{.2}\overline{X}_{.k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{1k}\overline{X}_{.1k}}{\overline{X}_{.1}\overline{X}_{.k}} & \frac{s_{2k}\overline{X}_{.2k}}{\overline{X}_{.2}\overline{X}_{.k}} & \cdots & \frac{s_k^2}{\overline{X}_{.k}} \end{pmatrix}.$$

Consequently, from 3.11, one can obtain the standard error of $\hat{\vartheta}_m$ as

$$\widehat{\sigma}_{\vartheta} = s.e(\hat{\vartheta}_m) = \sqrt{\frac{1}{m}\left(\sum_{j=1}^{k} \alpha_j^2 \frac{s_j^2}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{s_{jj'}\overline{X}_{.jj'}}{\overline{X}_{.j}\overline{X}_{.j'}}\right)},$$

where, the estimators namely $\overline{X}_{.j}, \overline{X}_{.jj'}, s_j^2$, and $s_{jj'}$ are the consistent estimators of the corresponding parameters $E(X_j), E(\sqrt{X_j X_{j'}}), \sigma_j^2$, and $\sigma_{jj'}$.
One also should mention here that, during the intervals evaluation, the estimator covariance matrix have to be positive or at least positive semi definite, to ensure that $\frac{\boldsymbol{\alpha}^T \widehat{\Sigma_{\mathbf{p}}} \boldsymbol{\alpha}}{m} \geq 0$, otherwise, we have to exclude the negative variances, this will be clarified in chapter 4.

Finally, The approximate confidence interval for the linear combination $\vartheta = \boldsymbol{\alpha}^T \mathbf{p}$, can be obtained by the normal quantile $z_{1-\frac{\alpha}{2}}$ as

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{\vartheta}_m)\right],$$

or the suggested conservative confidence interval by the t-quantile which is given by

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

as, $\frac{\hat{\vartheta}_m - \vartheta}{s.e(\hat{\vartheta}_m)} \simeq t_{m-1}$, where, $t_{(m-1,1-\frac{\alpha}{2})}$ is a $\left(1 - \frac{\alpha}{2}\right)$ quantile of the $t-$distribution with $(m-1)$ degrees of freedom.

3.2.8 Derivation of confidence intervals for ϑ

In continuous context, one may gives the outline for deriving the confidence intervals for the linear combination ϑ of the proportions, where $\vartheta = \boldsymbol{\alpha}^T \mathbf{p}$, $\boldsymbol{\alpha} \geq \mathbf{0}$ at the following cases:

case I

Firstly and on one hand, the distribution of $\hat{\vartheta} = \boldsymbol{\alpha}^T \hat{\mathbf{p}}$ ($\hat{\vartheta}$ is an unbiased estimator of ϑ), given the design vectors $(\mathbf{x}_1^T, ..., \mathbf{x}_k^T)$, $\mathbf{x}_j = (x_{1j}, ..., x_{mj})^T, j = 1, ..., k$, where $\hat{\mathbf{p}} \sim N_k(\mathbf{p}, \Sigma_{\mathbf{x}})$, $\Sigma_{\mathbf{x}}$ is the covariance matrix given by 3.4, will obtained as

$$\hat{\vartheta} \sim N\left(\sum_{j=1}^{k} \alpha_j p_j, \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\sum_{i=1}^{m} x_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} \sum_{i=1}^{m} \sqrt{x_{ij} x_{ij'}}}{\sum_{i=1}^{m} x_{ij} \sum_{i=1}^{m} x_{ij'}}\right), \tag{3.12}$$

the expression 3.12 rewritten after it is standardization, as

$$\begin{aligned}\sqrt{m}\left(\hat{\vartheta} - \vartheta\right) &\sim N\left(0, \sum_{j=1}^{k} \alpha_j^2 \frac{m\sigma_j^2}{\sum_{i=1}^{m} x_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} m \sum_{i=1}^{m} \sqrt{x_{ij} x_{ij'}}}{\sum_{i=1}^{m} x_{ij} \sum_{i=1}^{m} x_{ij'}}\right) \\ &\equiv N\left(0, \sigma_{\mathbf{x}}^2\right),\end{aligned}$$

where, the variance

$$\sigma_{\mathbf{x}}^2 = \sum_{j=1}^{k} \alpha_j^2 \frac{m\sigma_j^2}{\sum_{i=1}^{m} x_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} m \sum_{i=1}^{m} \sqrt{x_{ij} x_{ij'}}}{\sum_{i=1}^{m} x_{ij} \sum_{i=1}^{m} x_{ij'}}.$$

It follows that, the exact $(1-\alpha)\%$ confidence intervals for ϑ are obtained by

$$\left[\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}}\ S.E(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta} - \vartheta}{\sigma_{\mathbf{x}}} \sim N(0,1),$$

when $\sigma_j^2, \sigma_{jj'}, j \neq j', j, j' = 1, ..., k$ are known,

$$\left[\hat{\vartheta} \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta})\right], \text{as}, \quad \frac{\hat{\vartheta} - \vartheta}{\widehat{\sigma}_{\mathbf{x}}} \sim t_{m-1},$$

when $\sigma_j^2, \sigma_{jj'}, j \neq j', j, j' = 1, ..., k$ are unknown,

where, the standard errors of $\hat{\vartheta}$ are given by

$$S.E(\hat{\vartheta}) = \sqrt{\sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\sum_{i=1}^{m} x_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} \sum_{i=1}^{m} \sqrt{x_{ij} x_{ij'}}}{\sum_{i=1}^{m} x_{ij} \sum_{i=1}^{m} x_{ij'}}},$$

and

$$s.e(\hat{\vartheta}) = \sqrt{\sum_{j=1}^{k} \alpha_j^2 \frac{s_j^2}{\sum_{i=1}^{m} x_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{s_{jj'} \sum_{i=1}^{m} \sqrt{x_{ij} x_{ij'}}}{\sum_{i=1}^{m} x_{ij} \sum_{i=1}^{m} x_{ij'}}}.$$

For the non multivariate Normal distribution of $\mathbf{Y}_i$ or non multivariate Normal distribution of $\hat{\mathbf{p}}$, and if these sequences say converge to the corresponding constants, namely:

$$\frac{1}{m}\sum_{i=1}^{m} x_{ij} \longrightarrow \mu_j, \quad \frac{1}{m}\sum_{i=1}^{m} \sqrt{x_{ij}x_{ij'}} \longrightarrow \mu_{jj'},$$

where, $\mu_j, \mu_{jj'}$ are constants, $j < j^{'} = 1, ..., k$, then the sequence of variances $\sigma^2_{\mathbf{x}}$ converges to the corresponding variance σ^2_{μ} as m tends to infinity, where

$$\sigma^2_{\mu} = \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\mu_j} + 2\sum_{j^{'}=2}^{k}\sum_{j=1}^{j^{'}-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}\mu_{jj'}}{\mu_j \mu_{j'}}. \tag{3.13}$$

And hence, the asymptotic Normal distribution of the estimator $\hat{\vartheta}_m$ can be obtained by

$$\sqrt{m}\left(\hat{\vartheta}_m - \vartheta\right) \longrightarrow^{\mathcal{D}} N(0, \sigma^2_{\mu}),$$

where the asymptotic variance σ^2_{μ} is given by 3.13.
The obtained confidence interval is an approximate confidence interval, since the asymptotic Normal of $\hat{\vartheta}_m$ is involved.

case II

On the second hand, the conditional variance of $\hat{\vartheta}$ given the random vectors $(\mathbf{X}_1^T, ..., \mathbf{X}_k^T)$, $\mathbf{X}_j = (X_{1j}, ..., X_{mj})^T, j = 1, ..., k$, when the vectors $\mathbf{X}_i = (X_{i1}, ..., X_{ik})^T, i = 1, ..., m$ are $i.i.d$ random vectors (non multivariate Normal distribution of $\mathbf{Y}_i$) would be obtained by

$$\sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\sum_{i=1}^{m} X_{ij}} + 2\sum_{j^{'}=2}^{k}\sum_{j=1}^{j^{'}-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}\sum_{i=1}^{m}\sqrt{X_{ij}X_{ij'}}}{\sum_{i=1}^{m} X_{ij}\sum_{i=1}^{m} X_{ij'}},$$

subsequently, it follows that

$$Var(\sqrt{m}\hat{\vartheta}) = mVar(\hat{\vartheta}) = \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\overline{X}_{.j}} + 2\sum_{j^{'}=2}^{k}\sum_{j=1}^{j^{'}-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}\overline{X}_{.jj'}}{\overline{X}_{.j}\overline{X}_{.j'}}, \tag{3.14}$$

And from the expression 3.10, one can rewrite the asymptotic distribution of the distribution of $\hat{\vartheta}_m$ as

$$\sqrt{m}\left(\hat{\vartheta}_m - \vartheta\right) \longrightarrow^{\mathcal{D}} N\left(0, \sigma^2_{\vartheta}\right), \tag{3.15}$$

where, the asymptotic variance σ_ϑ^2 is given by

$$Var(\sqrt{m}\hat{\vartheta}_m) = \sigma_\vartheta^2 = \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{E(X_j)} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} E(\sqrt{X_j X_{j'}})}{E(X_j)E(X_{j'})}. \quad (3.16)$$

However, since

$$\overline{X}_{.j} \longrightarrow^{\mathcal{P}} E(X_j), \overline{X}_{.j'} \longrightarrow^{\mathcal{P}} E(X_{j'}), \text{ as well as}$$

$$\overline{X}_{.jj'} = \frac{1}{m}\sum_{i=1}^{m} \sqrt{X_{ij}X_{ij'}} \longrightarrow^{\mathcal{P}} E\left(\sqrt{X_j X_{j'}}\right), \text{ and } j \neq j' \text{ (with the LLN)},$$

then the variance in 3.14 converges also in probability to the corresponding asymptotic variance 3.16.

And thus, the corresponding variance estimator for σ_ϑ^2 is obtained by plugging in the estimator for the individual parameter in 3.16 as

$$\widehat{\sigma_\vartheta^2} = \sum_{j=1}^{k} \alpha_j^2 \frac{s_j^2}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{s_{jj'}\overline{X}_{.jj'}}{\overline{X}_{.j}\overline{X}_{.j'}},$$

provided that, the estimators

$$s_j^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij} - \hat{p}_j\sqrt{X_{ij}}\right)^2,$$

$$s_{jj'} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\widetilde{Y}_{ij} - \hat{p}_j\sqrt{X_{ij}}\right)\left(\widetilde{Y}_{ij'} - \hat{p}_{j'}\sqrt{X_{ij'}}\right), X_{ij}, X_{ij'} > 0$$

as well as, $\overline{X}_{.j}, \overline{X}_{.j'}, \overline{X}_{.jj'}$, are all consistent estimators for the corresponding parameters σ_j^2 and $\sigma_{jj'}, E(X_j), E(X_{j'}), E\left(\sqrt{X_j X_{j'}}\right)$. It follows that, the estimator $\widehat{\sigma_\vartheta^2}$ is also a consistent estimator of σ_ϑ^2, and hence the standard error of $\hat{\vartheta}$, $s.e(\hat{\vartheta}) = \sqrt{\frac{1}{m}\widehat{\sigma_\vartheta^2}}$.

One can now construct the approximate confidence intervals for the linear combination ϑ as following:

When, $\sigma_j^2, \sigma_{jj'}, \forall j, j' = 1, ..., k, j \neq j'$ are known then, the approximate confidence bounds for ϑ is given by

$$\left[\hat{\vartheta} \pm z_{1-\frac{\alpha}{2}} S.E(\hat{\vartheta})\right],$$

where, the Standard Error

$$S.E(\hat{\vartheta}) = \sqrt{\frac{1}{m}\left(\sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'}\overline{X}_{.jj'}}{\overline{X}_{.j}\overline{X}_{.j'}}\right)},$$

and the quantile, $z_{1-\frac{\alpha}{2}}$ is defined as previous.

In case of unknown parameters $\sigma_j^2, \sigma_{jj'}$, again the asymptotic theory given here will be involved to obtain the approximate confidence interval for ϑ, which is given by

$$\left[\hat{\vartheta}_m \pm z_{1-\frac{\alpha}{2}} s.e(\hat{\vartheta}_m)\right],$$

where,

$$s.e(\hat{\vartheta}_m) = \sqrt{\frac{1}{m}\left(\sum_{j=1}^{k} \alpha_j^2 \frac{s_j}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{s_{jj'}\overline{X}_{.jj'}}{\overline{X}_{.j}\overline{X}_{.j'}}\right)}.$$

For a small sample size with all these available consistent estimators, the $t-$distribution can be involved to obtain a conservative confidence interval for the linear combination $\vartheta = \sum_{j=1}^{k} \alpha_j p_j$, which is given by

$$\left[\hat{\vartheta}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\hat{\vartheta}_m)\right],$$

as

$$\frac{\hat{\vartheta}_m - \vartheta}{s.e(\hat{\vartheta}_m)} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1).$$

Finally, when the $i.d$ weighted random component $\widetilde{\mathbf{Y}}_i$ has the multivariate Normal distribution ($\hat{\mathbf{p}}$ has the multivariate Normal distribution), then given $(\mathbf{X}_1^T, ..., \mathbf{X}_k^T)$ we will have

$$\hat{\vartheta} \sim N\left(\sum_{j=1}^{k} \alpha_j p_j, \sum_{j=1}^{k} \alpha_j^2 \frac{\sigma_j^2}{\sum_{i=1}^{m} X_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} \sum_{i=1}^{m}\sqrt{X_{ij}X_{ij'}}}{\sum_{i=1}^{m} X_{ij} \sum_{i=1}^{m} X_{ij'}}\right)$$

this expression rewritten after its standardization, as

$$\begin{aligned}\sqrt{m}\left(\hat{\vartheta} - \vartheta\right) &\sim N\left(0, \sum_{j=1}^{k} \alpha_j^2 \frac{m\sigma_j^2}{\sum_{i=1}^{m} X_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} m \sum_{i=1}^{m}\sqrt{X_{ij}X_{ij'}}}{\sum_{i=1}^{m} X_{ij} \sum_{i=1}^{m} X_{ij'}}\right) \\ &\equiv N\left(0, \sigma_{\mathbf{X}}^2\right),\end{aligned}$$

where, the variance

$$\sigma_{\mathbf{X}}^2 = \sum_{j=1}^{k} \alpha_j^2 \frac{m\sigma_j^2}{\sum_{i=1}^{m} X_{ij}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1} \alpha_j \alpha_{j'} \frac{\sigma_{jj'} m \sum_{i=1}^{m}\sqrt{X_{ij}X_{ij'}}}{\sum_{i=1}^{m} X_{ij} \sum_{i=1}^{m} X_{ij'}}.$$

Similar to the case I, the constructing confidence intervals in this case are exact confidence intervals, since it was assumed that the random components $\mathbf{Y}_i$ has the multivariate Normal distribution.

3.3 Multivariate Poisson model

There are many fields in which the multivariate Poisson data arises, i.e: Epidemiology: Incidences of different diseases across in time; Crime data: rapes, arson, manslaughter, smuggling; marketing: Purchases of different products; economics: Different types of faults in production system; sports: Football data, etc.

For a comprehensive discussion of the bivariate Poisson model and its multivariate extensions see [6].

3.3.1 The Multivariate Poisson model

In order to extend the bivariate Poisson model to the multivariate extension, one shall consider the following random variables:

$$\begin{aligned}
Y_1 &= W + V_1, & X_1 &= Y_1 + Z_1 \\
Y_2 &= W + V_2, & X_2 &= Y_2 + Z_2 \\
&\vdots & &\vdots \\
Y_k &= W + V_k, & X_k &= Y_k + Z_k
\end{aligned}$$

where, W, V_j, Z_j are independent Poisson distributed random variables with the parameters $\lambda_0, \lambda_j, \mu_j$ respectively, W, V_j are latent variables, while Y_j, Z_j are observable, $j = 1, ..., k$.

The random variables $Y_1, ..., Y_k$ follow jointly a Multivariate Poisson distribution with the joint probability function is given by

$$\begin{aligned}
P(\mathbf{Y} = \mathbf{y}) = P(Y_1 = y_1, Y_2 = y_2, ..., Y_k = y_k) &= \sum_{w=0}^{s} P(W = w, V_j = y_j - w) \\
&= e^{-(\lambda_0 + \sum_{j=1}^{k} \lambda_j)} \prod_{j=1}^{k} \frac{\lambda_j^{y_j}}{y_j!} \sum_{w=0}^{s} \left[\prod_{l=1}^{k} \binom{y_l}{w} w^{k-1}! \left(\frac{\lambda_0}{\prod_{j=1}^{k} \lambda_j} \right)^{w} \right] \qquad (3.17)
\end{aligned}$$

where, $s = \min(y_1, y_2, ..., y_k)$.

The probability function (3.17) is quite complicated for calculation and for obtaining

the Likelihood function for maximization (containing the latent variable W), nevertheless one can say that $Y_1, ..., Y_k \sim MulPoiss(\lambda_0, \lambda_1, ..., \lambda_k)$, where, $MulPoiss$ denotes the multivariate Poisson distribution, and λ_0, λ_j are the parameters of the independent Poisson random variables W, V_j respectively.
Moreover, $X_1, ..., X_k \sim MulPoiss(\lambda_0, \lambda_1 + \mu_1, ..., \lambda_k + \mu_k)$, where $\lambda_0, \lambda_j + \mu_j$ are the parameters of the corresponding independent Poisson variables $W, V_j + Z_j, j = 1, ..., k$.

3.3.2 Properties of the model

- Marginally, $Y_j \sim Poiss(\lambda_0 + \lambda_j)$, as well as $X_j \sim Poiss(\lambda_0 + \lambda_j + \mu_j)$. This implies that $E(Y_j) = Var(Y_j) = \lambda_0 + \lambda_j$, and
$E(X_j) = Var(X_j) = \lambda_0 + \lambda_j + \mu_j, j = 1, ..., k$.

- $Cov(X_j, X_{j'}) = Cov(Y_j, Y_{j'}) = \lambda_0, \forall j \neq j', j, j' = 1, ..., k$,
i.e, the parameter λ_0 is the covariance between all the pairs of the random variable Y_j, and all the pairs of the random variable X_j. For different covariance structure for each pair of the variables, see [13].

- The Marginal Conditional distributions are given by:
$P^{(Y_j|X_j)} \sim Bin\left(X_j, \frac{\lambda_0+\lambda_j}{\lambda_0+\lambda_j+\mu_j}\right) \equiv Bin(X_j, p_j)$, where the Binomial proportion $p_j = \frac{\lambda_0+\lambda_j}{\lambda_0+\lambda_j+\mu_j}$, further, it follows that $E(Y_j \mid X_j) = p_j X_j$, as well as
$Var(Y_j \mid X_j) = p_j(1 - p_j)X_j$, however the joint conditional distribution
$P^{\left(Y_j, Y_{j'} | X_j, X_{j'}\right)}$ and the pair conditional covariance $Cov(Y_j, Y_{j'} \mid X_j, X_{j'})$, $j \neq j' = 1, ..., k$ can not be explicitly calculated.

- Unconditional consistent estimators of the combinations:
$\lambda_0, \lambda_0 + \lambda_j, \lambda_0 + \lambda_j + \mu_j$, are: $s_{Y_j, Y_{j'}}, \overline{Y}_{.j}, \overline{X}_{.j}$, respectively i.e:
$s_{Y_j, Y_{j'}} \longrightarrow^{\mathcal{P}} \lambda_0, \overline{Y}_{.j} \longrightarrow^{\mathcal{P}} \lambda_0 + \lambda_j, \overline{Y}_{.j} - s_{Y_j, Y_{j'}} \longrightarrow^{\mathcal{P}} \lambda_j$, as well as
$\overline{X}_{.j} \longrightarrow^{\mathcal{P}} \lambda_0 + \lambda_j + \mu_j$. Furthermore $\overline{X}_{.j} - s_{Y_j, Y_{j'}} \longrightarrow^{\mathcal{P}} \lambda_j + \mu_j$, $\overline{Z}_{.j} \longrightarrow^{\mathcal{P}} \mu_j$, where $\overline{Y}_{.j} = \frac{1}{m}\sum_{i=1}^{m} Y_{ij} = \overline{W} + \overline{V}_{.j}$, $\overline{X}_{.j} = \frac{1}{m}\sum_{i=1}^{m} X_{ij} = \overline{W} + \overline{V}_{.j} + \overline{Z}_{.j}$, and
$\overline{Z}_{.j} = \frac{1}{m}\sum_{i=1}^{m} Z_{ij}$, as well as the unconditional sample covariance
$s_{Y_j, Y_{j'}} = \frac{1}{m-1}\sum_{i=1}^{m}(Y_{ij} - \overline{Y}_{.j})(Y_{ij'} - \overline{Y}_{.j'})$, $j < j', j, j' = 1, ..., k$, and also
$\hat{p}_j = \frac{\widehat{\lambda_0+\lambda_j}}{\lambda_0+\widehat{\lambda_j}+\mu_j} = \frac{\overline{Y}_{.j}}{\overline{X}_{.j}}$ are the commonly used ratio estimators for the proportions $p_j, j = 1, ..., k$.

3.3.3 Description of the model

One may describe the model through the vector, $\mathbf{T} = (W, V_1, ..., V_k, Z_1, ...Z_k)^T$, and an $k \times (2k+1)$ matrix $\mathbf{A}$ (i.e, the elements of $\mathbf{A}$ are zero and ones no duplicate rows exist, (see [12], for more details).

Define the matrices $\mathbf{A}_y, \mathbf{A}_x$, having the forms $\mathbf{A}_y = [1_k \; I_k \; \mathbf{0}_{k\times k}]$, $\mathbf{A}_x = [1_k \; I_k \; I_k]$, where $1_k = (1, ..., 1)^T$, I_k is the identity matrix of size $k \times k$, as well as $\mathbf{0}_{k\times k}$ is a matrix of $k \times k$ zero's.

The vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_k)^T = \mathbf{A}_y\mathbf{T}$ follows a multivariate Poisson distribution with parameters $\lambda_0, \lambda_1, ..., \lambda_k$.

Furthermore, $\mathbf{X} = (X_1, X_2, ..., X_k)^T = \mathbf{A}_x\mathbf{T}$ follows also a multivariate Poisson distribution with parameters $\lambda_0, \lambda_1 + \mu_1, ..., \lambda_k + \mu_k$. The number of the model parameters is exactly $2k + 1, k \geq 2$.

Further, one may describe the hole model by the vector $(\mathbf{Y}, \mathbf{X})^T$, as

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_y \\ \mathbf{A}_x \end{pmatrix} \begin{pmatrix} W \\ \mathbf{V} \\ \mathbf{Z} \end{pmatrix},$$

where, $\mathbf{Y} = (Y_1, ..., Y_k)^T, \mathbf{X} = (X_1, ..., X_k)^T, \mathbf{V} = (V_1, ..., V_k)^T$, $\mathbf{Z} = (Z_1, ..., Z_k)^T$, and $\mathbf{A}_y, \mathbf{A}_x$ are defined above.

3.3.4 Multivariate asymptotic normality of the estimator vector $\hat{\mathbf{p}}_m$

We will derive the asymptotic distribution of the estimator vector $\hat{\mathbf{p}}_m$ of the proportion vector $\mathbf{p} = (p_1, ..., p_k)^T$, where, $\hat{\mathbf{p}}_m = (\hat{p}_{m1}, ..., \hat{p}_{mk})^T$, as well as, $\hat{p}_{mj}$ is the ratio estimator of the corresponding proportion $p_j, j = 1, ..., k$.

For the independent r.v's W, V_j and Z_j of the vector $\mathbf{T} = (W, V_1, ..., V_k, Z_1, ...Z_k)^T$, we have the asymptotic covariance matrix

$$\Lambda = Cov(\mathbf{T}) =$$

$$\begin{bmatrix} \lambda_0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_1 & 0 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \lambda_k & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \mu_1 & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & \cdots & 0 & \mu_k \end{bmatrix}_{(2k+1)\times(2k+1)} = diag\left(\lambda_0, \lambda_1, ..., \lambda_k, ..., \mu_k\right),$$

and further one can apply the multivariate central limit theorem to the vector $\overline{\mathbf{T}}_m = \frac{1}{m}\sum_{i=1}^{m}\mathbf{T}_i$, where the $i.i.d$ random vectors $\mathbf{T}_i = (W_i, V_{i1}, ..., V_{ik}, Z_{i1}, ...Z_{ik})^T$:

$$\sqrt{m}\left(\overline{\mathbf{T}}_m - E(\mathbf{T})\right) \longrightarrow^{\mathcal{D}} N_{2k+1}\left(\mathbf{0}_{2k+1}, \Lambda\right),$$

where, the covariance matrix Λ is defined above.

Define the vector $\mathbf{H} = (Y_1, X_1, ..., Y_k, X_k)^T$, with the covariance matrix

$$\mathbf{\Sigma}_\lambda = Cov(\mathbf{H}) = Cov\begin{pmatrix} Y_1 \\ X_1 \\ \cdots \\ \vdots \\ \cdots \\ Y_k \\ X_k \end{pmatrix}$$

$$= \begin{bmatrix} \mathbf{C}_1 & \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} & \cdots & \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} \\ \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} & \mathbf{C}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} \\ \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} & \cdots & \begin{bmatrix} \lambda_o & \lambda_o \\ \lambda_o & \lambda_o \end{bmatrix} & \mathbf{C}_k \end{bmatrix}_{(2k\times 2k)},$$

where, the diagonal elements of the covariance matrix are given by

$$\mathbf{C}_j = \begin{bmatrix} \lambda_o + \lambda_j & \lambda_o + \lambda_j \\ \lambda_o + \lambda_j & \lambda_o + \lambda_j + \mu_j \end{bmatrix} = \begin{bmatrix} E(Y_j) & E(Y_j) \\ E(Y_j) & E(X_j) \end{bmatrix}, \ j = 1, ..., k.$$

Thus, by plugging in these diagonal entries to the covariance matrix $\mathbf{\Sigma}_{\lambda}$, we obtain

$$\mathbf{\Sigma}_{\lambda} = diag\left(\lambda_j 1_2 1_2^T + \begin{bmatrix} 0 & 0 \\ 0 & \mu_j \end{bmatrix}\right)_{j=1,...,k} + \lambda_0 1_{2k} 1_{2k}^T. \tag{3.18}$$

And in same manner, by applying the Multivariate Central limit theorem to the $i.i.d$ random vectors $\mathbf{H}_i = (Y_{i1}, X_{i1}, ..., Y_{ik}, X_{ik})^T$, $\overline{\mathbf{H}}_m = \frac{1}{m}\sum_{i=1}^{m} \mathbf{H}_i$, yields

$$\sqrt{m}\left(\overline{\mathbf{H}}_m - E(\mathbf{H})\right) \longrightarrow^{\mathcal{D}} N_{2k}\left(\mathbf{0}, \mathbf{\Sigma}_{\lambda}\right),$$

or,

$$\sqrt{m}\left(\begin{bmatrix} \overline{Y}_{.1} \\ \overline{X}_{.1} \\ \cdots \\ \vdots \\ \cdots \\ \overline{Y}_{.k} \\ \overline{X}_{.k} \end{bmatrix}_{2k\times 1} - \begin{bmatrix} \lambda_0 + \lambda_1 \\ \lambda_0 + \lambda_1 + \mu_1 \\ \cdots \\ \vdots \\ \cdots \\ \lambda_0 + \lambda_k \\ \lambda_0 + \lambda_k + \mu_k \end{bmatrix}\right) \longrightarrow^{\mathcal{D}} N_{2k}\left(\begin{bmatrix} 0 \\ 0 \\ \cdots \\ \vdots \\ \cdots \\ 0 \\ 0 \end{bmatrix}_{2k\times 1}, \mathbf{\Sigma}_{\lambda}\right),$$

The Multivariate Delta method

So far, it was just applying the MVCLT to the pairs $(\overline{Y}_j, \overline{X}_j)$ to obtain the asymptotic normality of the vector estimator $\hat{\mathbf{p}}$, we will introduce to the extended δ-method called the multivariate δ-method (see the subsection 1.3.3 or [14]) which is applied to obtain the asymptotic distribution of a k dimensional non-linear mappings of the pairs random variables (Y_j, X_j), i.e, asymptotic distribution of the non-linear transformations $\frac{\overline{Y}_{.j}}{\overline{X}_{.j}}, j = 1, ..., k$.
We will define the following notations:

$$\boldsymbol{\theta} = E(\mathbf{H}) = (\theta_{11}, \theta_{21}, ..., \theta_{1k}, \theta_{2k})^T, \quad \theta_{1j} = E(Y_j),\ \theta_{2j} = E(X_j), \quad j = 1, ..., k,$$
$$\boldsymbol{\theta} \in \mathbf{R}^{2k},\ \ g(\boldsymbol{\theta}) = \left(\frac{\theta_{1j}}{\theta_{2j}}\right)_{j=1,...,k},\ g(\boldsymbol{\theta}) : \mathbf{R}^{2k} \longrightarrow \mathbf{R}^{k}, \tag{3.19}$$

is a vector-valued function that is continuously differentiable at $\boldsymbol{\theta}$ such that $\theta_{2j} > 0\ \forall j$. By plugging in the elements of $\boldsymbol{\theta}$ of notation 3.19 to the covariance matrix 3.18, we obtain

$$\boldsymbol{\Sigma}_{\lambda} = \boldsymbol{\Sigma}_{\theta,\lambda_0} = \begin{bmatrix} \begin{bmatrix} \theta_{11} & \theta_{11} \\ \theta_{11} & \theta_{21} \end{bmatrix} & \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} & \cdots & \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} \\ \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} & \begin{bmatrix} \theta_{12} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{bmatrix} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} \\ \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} & \cdots & \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix} & \begin{bmatrix} \theta_{1k} & \theta_{1k} \\ \theta_{1k} & \theta_{2k} \end{bmatrix} \end{bmatrix}_{2k \times 2k},$$

where,

$$\theta_{sj} = \begin{cases} E(Y_j) & ,\text{for } s = 1 \\ E(X_j) & ,\text{for } s = 2 \end{cases} \quad ,\forall\ j = 1, ..., k.$$

Let, $\boldsymbol{\theta}_j = \begin{bmatrix} \theta_{1j} & \theta_{1j} \\ \theta_{1j} & \theta_{2j} \end{bmatrix}$, and $\boldsymbol{\lambda}_0 = \begin{bmatrix} \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 \end{bmatrix}$, then the squared block matrix

$$\boldsymbol{\Sigma}_{\theta,\lambda_0} = \begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{\lambda}_0 & \cdots & \boldsymbol{\lambda}_0 \\ \boldsymbol{\lambda}_0 & \boldsymbol{\theta}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\lambda}_0 \\ \boldsymbol{\lambda}_0 & \cdots & \boldsymbol{\lambda}_0 & \boldsymbol{\theta}_k \end{bmatrix}_{2k \times 2k}. \tag{3.20}$$

The matrix $\nabla_g^T(\boldsymbol{\theta})$ of partial derivatives of the continuous differentiable function g with respect to $\boldsymbol{\theta}$ is obtained by

$$\begin{aligned} \nabla_g^T(\boldsymbol{\theta}) = \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} &= \begin{bmatrix} \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_{1j}} & \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_{2j}} \end{bmatrix}_{j=1,\ldots,k} \\ &= \begin{bmatrix} \frac{\partial g_1}{\partial \theta_{11}} & \frac{\partial g_1}{\partial \theta_{21}} & \frac{\partial g_1}{\partial \theta_{12}} & \frac{\partial g_1}{\partial \theta_{22}} & \cdots & \frac{\partial g_1}{\partial \theta_{1k}} & \frac{\partial g_1}{\partial \theta_{2k}} \\ \frac{\partial g_2}{\partial \theta_{11}} & \frac{\partial g_2}{\partial \theta_{21}} & \frac{\partial g_2}{\partial \theta_{12}} & \frac{\partial g_2}{\partial \theta_{22}} & \cdots & \frac{\partial g_2}{\partial \theta_{1k}} & \frac{\partial g_2}{\partial \theta_{2k}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial g_k}{\partial \theta_{11}} & \frac{\partial g_k}{\partial \theta_{21}} & \frac{\partial g_k}{\partial \theta_{12}} & \frac{\partial g_k}{\partial \theta_{22}} & \cdots & \frac{\partial g_k}{\partial \theta_{1k}} & \frac{\partial g_k}{\partial \theta_{2k}} \end{bmatrix}_{(k \times 2k)} \\ &= \begin{bmatrix} \frac{1}{\theta_{21}} & -\frac{\theta_{11}}{\theta_{21}^2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{\theta_{22}} & -\frac{\theta_{12}}{\theta_{22}^2} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \frac{1}{\theta_{2k}} & -\frac{\theta_{1k}}{\theta_{2k}^2} \end{bmatrix}_{(k \times 2k)}. \end{aligned}$$

Let, $\boldsymbol{\eta}_j = \begin{pmatrix} \frac{1}{\theta_{2j}} & -\frac{\theta_{1j}}{\theta_{2j}^2} \end{pmatrix}_{j=1,...,k}$, the block diagonal matrix

$$\nabla_g^T(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\eta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\eta}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\eta}_k \end{bmatrix}_{(k\times 2k)} = block.diag\left(\boldsymbol{\eta}_j\right)_{j=1,...,k},$$ whose the off-diagonal matrices $\mathbf{0} = (0 \quad 0)$, as well as the transpose of $\nabla_g^T(\boldsymbol{\theta})$ is given by

$$\nabla_g(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\eta}_1^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \boldsymbol{\eta}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \boldsymbol{\eta}_k^T \end{bmatrix}.$$

For, $\mathbf{p} = (p_1, ..., p_k)^T$, where $p_j = \frac{E(Y_j)}{E(X_j)} = \frac{\lambda_0+\lambda_j}{\lambda_0+\lambda_j+\mu_j}$, and according to the MVδ-method we have

$$g(\boldsymbol{\theta}) = \mathbf{p} \Rightarrow g(\hat{\boldsymbol{\theta}}_m) = \hat{\mathbf{p}}_m, \hat{\boldsymbol{\theta}}_m = \overline{\mathbf{H}}_m,$$

with the corresponding estimators of $\hat{\mathbf{p}}_m = (\hat{p}_{m1}, ..., \hat{p}_{mk})^T$, $\hat{p}_{mj} = \frac{\overline{Y}_{.j}}{\overline{X}_{.j}}$, $j = 1, ..., k$. Hence, it follows that

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}} N_k\left(\mathbf{0}, \nabla_g^T(\boldsymbol{\theta})\, \boldsymbol{\Sigma}_{\theta,\lambda_0} \nabla_g(\boldsymbol{\theta})\right), \tag{3.21}$$

where, $\boldsymbol{\Sigma}_{\theta,\lambda_0}$ is given by the matrix 3.20. So, the asymptotic covariance matrix of expression (3.21) equals

$$\nabla_g^T(\boldsymbol{\theta})\, \boldsymbol{\Sigma}_{\theta,\lambda_0} \nabla_g(\boldsymbol{\theta}) =$$

$$= \begin{bmatrix} \boldsymbol{\eta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\eta}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\eta}_k \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{\lambda}_0 & \cdots & \boldsymbol{\lambda}_0 \\ \boldsymbol{\lambda}_0 & \boldsymbol{\theta}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\lambda}_0 \\ \boldsymbol{\lambda}_0 & \cdots & \boldsymbol{\lambda}_0 & \boldsymbol{\theta}_k \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\eta}_1^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \boldsymbol{\eta}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \boldsymbol{\eta}_k^T \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\eta}_1\boldsymbol{\theta}_1\boldsymbol{\eta}_1^T & \boldsymbol{\eta}_1\boldsymbol{\lambda}_0\boldsymbol{\eta}_2^T & \cdots & \boldsymbol{\eta}_1\boldsymbol{\lambda}_0\boldsymbol{\eta}_k^T \\ \boldsymbol{\eta}_2\boldsymbol{\lambda}_0\boldsymbol{\eta}_1^T & \boldsymbol{\eta}_2\boldsymbol{\theta}_2\boldsymbol{\eta}_2^T & \cdots & \boldsymbol{\eta}_2\boldsymbol{\lambda}_0\boldsymbol{\eta}_k^T \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\eta}_k\boldsymbol{\lambda}_0\boldsymbol{\eta}_1^T & \boldsymbol{\eta}_k\boldsymbol{\lambda}_0\boldsymbol{\eta}_2^T & \cdots & \boldsymbol{\eta}_k\boldsymbol{\theta}_k\boldsymbol{\eta}_k^T \end{bmatrix}. \tag{3.22}$$

There j^{th} diagonal element simplifies to

$$\boldsymbol{\eta}_j \boldsymbol{\theta}_j \boldsymbol{\eta}_j^T = \frac{\theta_{1j}}{\theta_{2j}^2} - \frac{2\theta_{1j}^2}{\theta_{2j}^3} + \frac{\theta_{1j}^2}{\theta_{2j}^3} = \frac{\theta_{1j}}{\theta_{2j}^2} - \frac{\theta_{1j}^2}{\theta_{2j}^3} = \frac{\theta_{1j}}{\theta_{2j}^2}\left(1 - \frac{\theta_{1j}}{\theta_{2j}}\right)$$
$$= \frac{E(Y_j)}{(E(X_j))^2}\left(1 - \frac{E(Y_j)}{E(X_j)}\right) = \frac{1}{E(X_j)} p_j(1-p_j),\ j = 1, ..., k,$$

and whose off-diagonal elements simplify to

$$\boldsymbol{\eta}_j \boldsymbol{\lambda}_0 \boldsymbol{\eta}_{j'}^T = \boldsymbol{\eta}_{j'} \boldsymbol{\lambda}_0 \boldsymbol{\eta}_j^T = \frac{\lambda_0}{\theta_{2j}\theta_{2j'}} - \frac{\lambda_0\theta_{1j}}{\theta_{2j}^2\theta_{2j'}} - \frac{\lambda_0\theta_{1j'}}{\theta_{2j}\theta_{2j'}^2} + \frac{\lambda_0\theta_{1j}\theta_{1j'}}{\theta_{2j}^2\theta_{2j'}^2}$$
$$= \frac{\lambda_0}{\theta_{2j}\theta_{2j'}}\left[1 - \frac{\theta_{1j}}{\theta_{2j}} - \frac{\theta_{1j'}}{\theta_{2j'}} + \frac{\theta_{1j}\theta_{1j'}}{\theta_{2j}\theta_{2j'}}\right]$$
$$= \frac{\lambda_0}{\theta_{2j}\theta_{2j'}}\left[\left(1 - \frac{\theta_{1j}}{\theta_{2j}}\right) - \frac{\theta_{1j'}}{\theta_{2j'}}\left(1 - \frac{\theta_{1j}}{\theta_{2j}}\right)\right]$$
$$= \frac{\lambda_0}{\theta_{2j}\theta_{2j'}}\left(1 - \frac{\theta_{1j}}{\theta_{2j}}\right)\left(1 - \frac{\theta_{1j'}}{\theta_{2j'}}\right)$$
$$= \frac{\lambda_0}{E(X_j)E(X_{j'})}(1-p_j)(1-p_{j'}),\ \forall j \neq j', j, j' = 1, ..., k.$$

The multivariate asymptotic Normal distribution follows now by plugging in the above diagonal and off-diagonal entries to the matrix 3.22 of the expression 3.21

$$\sqrt{m}(\hat{\mathbf{p}}_m - \mathbf{p}) \longrightarrow^{\mathcal{D}}$$
$$N_k\left(\mathbf{0}, \begin{bmatrix} \frac{p_1(1-p_1)}{E(X_1)} & \frac{Cov(Y_1,Y_2)(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \cdots & \frac{Cov(Y_1,Y_k)(1-p_1)(1-p_k)}{E(X_1)E(X_k)} \\ \frac{Cov(Y_1,Y_2)(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \frac{p_2(1-p_2)}{E(X_2)} & \cdots & \frac{Cov(Y_2,Y_k)(1-p_2)(1-p_k)}{E(X_2)E(X_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{Cov(Y_1,Y_k)(1-p_1)(1-p_k)}{E(X_1)E(X_k)} & \frac{Cov(Y_2,Y_k)(1-p_2)(1-p_k)}{E(X_2)E(X_k)} & \cdots & \frac{p_k(1-p_k)}{E(X_k)} \end{bmatrix}\right),$$

where, $Cov(Y_j, Y_{j'}) = \lambda_0,\ \forall j < j', j, j' = 1, ..., k$. The asymptotic covariance matrix of the ratio estimator vector $\hat{\mathbf{p}}_m$ will be

$$asCov(\hat{\mathbf{p}}_m) = \begin{bmatrix} \frac{p_1(1-p_1)}{E(X_1)} & \frac{Cov(Y_1,Y_2)(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \cdots & \frac{Cov(Y_1,Y_k)(1-p_1)(1-p_k)}{E(X_1)E(X_k)} \\ \frac{Cov(Y_1,Y_2)(1-p_1)(1-p_2)}{E(X_1)E(X_2)} & \frac{p_2(1-p_2)}{E(X_2)} & \cdots & \frac{Cov(Y_2,Y_k)(1-p_2)(1-p_k)}{E(X_2)E(X_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{Cov(Y_1,Y_k)(1-p_1)(1-p_k)}{E(X_1)E(X_k)} & \frac{Cov(Y_2,Y_k)(1-p_2)(1-p_k)}{E(X_2)E(X_k)} & \cdots & \frac{p_k(1-p_k)}{E(X_k)} \end{bmatrix}.$$

Again the estimator for the asymptotic covariance matrix can be obtained by plugging in the estimates for each individual parameter in the matrix $asCov(\hat{\mathbf{p}}_m)$

$$as\widehat{Cov}(\hat{\mathbf{p}}_m) = \begin{bmatrix} \frac{\hat{p}_1(1-\hat{p}_1)}{\overline{X}_{.1}} & \frac{s_{Y_1,Y_2}(1-\hat{p}_1)(1-\hat{p}_2)}{\overline{X}_{.1}\overline{X}_{.2}} & \cdots & \frac{s_{Y_1,Y_k}(1-\hat{p}_1)(1-\hat{p}_k)}{\overline{X}_{.1}\overline{X}_{.k}} \\ \frac{s_{Y_1,Y_2}(1-\hat{p}_1)(1-\hat{p}_2)}{\overline{X}_{.1}\overline{X}_{.2}} & \frac{\hat{p}_2(1-\hat{p}_2)}{\overline{X}_{.2}} & \cdots & \frac{s_{Y_2,Y_k}(1-\hat{p}_2)(1-\hat{p}_k)}{\overline{X}_{.2}\overline{X}_{.k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{Y_1,Y_k}(1-\hat{p}_1)(1-\hat{p}_k)}{\overline{X}_{.1}\overline{X}_{.k}} & \frac{s_{Y_2,Y_k}(1-\hat{p}_2)(1-\hat{p}_k)}{\overline{X}_{.2}\overline{X}_{.k}} & \cdots & \frac{\hat{p}_k(1-\hat{p}_k)}{\overline{X}_{.k}} \end{bmatrix}, \quad (3.23)$$

where

$$\overline{X}_{.j} \longrightarrow^{\mathcal{P}} E(X_j),\ \hat{p}_j \longrightarrow^{\mathcal{P}} p_j,\ s_{Y_j,Y_{j'}} \longrightarrow^{\mathcal{P}} Cov(Y_j, Y_{j'}),$$

and the unconditional sample covariances

$$s_{Y_j,Y_{j'}} = \frac{1}{m-1}\sum_{i=1}^{m}(Y_{ij} - \overline{Y}_{.j})(Y_{ij'} - \overline{Y}_{.j'}),\ j < j', j, j' = 1, ..., k,$$

alternatively, one may consider a joint estimator for λ_0

$$s_{j,j'} = \frac{1}{\frac{k(k-1)}{2}}\sum_{j=1}^{k}\sum_{j'=j-1}^{k} s_{Y_j,Y_{j'}}.$$

I.e, the matrix 3.23 is also a consistent.

3.3.5 Approximate confidence intervals for a 'k' linear combination of the proportions

For constructing the confidence intervals for the linear combination of the proportions one should mention that, the matrix 3.23 have to be positive or at least positive semi definite, to ensure, $\frac{as\widehat{Var(\boldsymbol{\alpha}^T\hat{\mathbf{p}})}}{m} \geq 0$, otherwise we have to exclude the negative variances (see chapter 2, subsection 2.3.5, also it will be clarified in the chapter 4). The estimated variance of the linear combination of the ratio estimators of $\hat{\mathbf{p}}_m$, such that $\boldsymbol{\alpha} \geq \mathbf{0}$ is given by

$$as\widehat{Var(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m)} = \sum_{j=1}^{k}\alpha_j^2\frac{\hat{p}_j(1-\hat{p}_j)}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1}\alpha_j\alpha_{j'}\frac{s_{Y_j,Y_{j'}}(1-\hat{p}_j)(1-\hat{p}_{j'})}{\overline{X}_{.j}\overline{X}_{.j'}}, \quad (3.24)$$

and consequently, the standard error of the linear combination $\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m$ (if defined), is giiven by the square root of the expression 3.24 divided by the $\sqrt{m}$,

$$s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m) = \sqrt{\frac{1}{m}\left(\sum_{j=1}^{k}\alpha_j^2\frac{\hat{p}_j(1-\hat{p}_j)}{\overline{X}_{.j}} + 2\sum_{j'=2}^{k}\sum_{j=1}^{j'-1}\alpha_j\alpha_{j'}\frac{s_{Y_j,Y_{j'}}(1-\hat{p}_j)(1-\hat{p}_{j'})}{\overline{X}_{.j}\overline{X}_{.j'}}\right)}.$$

Finally, the approximate $(1-\alpha)\,\%$ confidence intervals for the linear combination of the proportions $\boldsymbol{\alpha}^T\mathbf{p}$ are given by

$$\left[\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m \pm z_{1-\frac{\alpha}{2}}\, s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m)\right],$$

Or, the suggested conservative confidence intervals (safety bounds) by the t-quantiles

$$\left[\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m \pm t_{(m-1,1-\frac{\alpha}{2})} s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m)\right],$$

as, $\frac{\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m - \boldsymbol{\alpha}^T\mathbf{p}}{s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}}_m)} \simeq t_{m-1} \longrightarrow^{\mathcal{D}} N(0,1)$, and the quantiles $z_{1-\frac{\alpha}{2}}$, $t_{(m-1,1-\frac{\alpha}{2})}$ are previously defined.

Chapter 4

Simulations for approximating the "true" coverage

4.1 Introduction

An extremely powerful application of modern computers is in the field of simulation. A simulation is a computer experiment which mirrors some aspect of real life data (which is complicated to manipulate in real life) that appears to be based on random processes.

Computer simulation tools can be used to compare the observed coverage ('coverage probability' or simply 'coverage', which is a number of the covering intervals for the parameter divided by the total replications or loops, or a percent of the covering intervals with respect to the total number of these intervals) of the confidence intervals with the corresponding nominal value (the true coverage).

The programs instructions are performed with the R language to run the R software packages. They made to compare a curve plotted by the coverage of the corresponding confidence interval with the nominal value of the true coverage (1-α), where α is the confidence level, with the fact: under repeated sampling, (1-α)% of these intervals will contain the proportion p or a linear function of the proportions.

Aim of the simulation

the main goal of the simulations is to validate approximate confidence intervals for a linear combination of the proportions based on Poisson models. The simulation study was performed and evaluated to obtain a more reasonable and appropriate coverage close to the true coverage. For approximating the true coverage for different sample sizes and model parameter values, I considered two confidence intervals, approximate (with the normal quantile) and conservative (with the t- quantile).

Thus, 'With large replications, samples of different sizes are taken from the Poisson distribution with different parameter values at different levels of the confidence intervals'.

To display the results graphically, the results plotted in figures with horizontal line to indicate the nominal values and by the corresponding tables for more illustration.

4.2 The Univariate case

Simulation in a univariate case does not need much work. Theoretically, the sample variance based on the univariate residuals as well as the variance estimator based on the univariate Poisson model are both BUE's.

The program is running with 10,000 replications at the nominal value 0.95, for small and large values of the Poisson parameter for different sample size m, taking in account only the valid intervals (runs) with valid standard errors, i.e., $\sum_{i=1}^{m} X_i > 0$, if $\sum_{i=1}^{m} X_i = 0$ in a sample, presumably nobody would like to calculate a confidence interval in that case, as no real observations are available). The t-quantile is adjusted on the reduced actual sample size m_1 which contains only the informative observations (a noninformative observation is the random variable associates with the event $A_i = \{X_i = 0\}$ with probability of occurrence $P(A_i) = e^{-\nu}$, and $P(\overline{A_i}) = 1 - e^{-\nu}$, where $\overline{A_i} = \{X_i > 0\}, i = 1, ..., m)$ (we only accept runs as valid, for which the reduced sample size m_1 of informative observations is at least 9). Also one can calculate the percentage of the excluded runs (non valid).

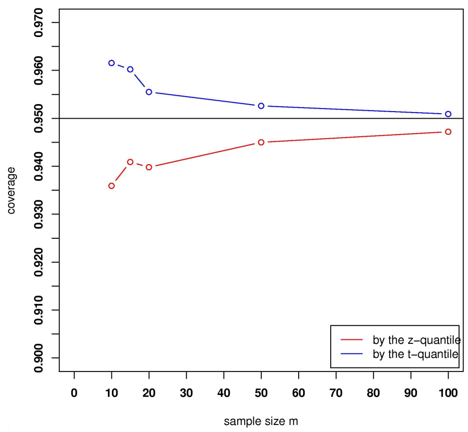

Figure 4.1: Coverage of the confidence interval for the proportion for the parameter values $\lambda = 1$, $m = 10, 15, 20, 50, 100$ with the variance estimate based on the Poisson model.

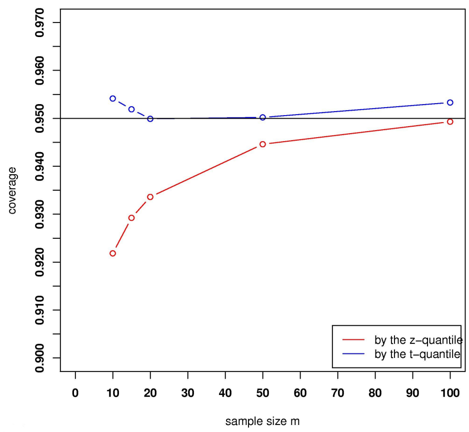

Figure 4.2: Coverage of the confidence interval for the proportion for the parameter values $\lambda = 1$, $m = 10, 15, 20, 50, 100$ with the sample variance.

For fixed $\lambda_1 = \lambda_2 = \lambda = 1$ for different sample sizes with the variance estimate based on the Poisson model figure 4.1, the coverage by the adjusted t-quantile method are larger than the nominal value, while by z-quantile are smaller. And with the sample variance (figure 4.2) is nearly similar to the figure 4.1. Further, there is no excluded runs, when the sample sizes become larger for any λ as seen in tables 4.1 and 4.2.

	sample size m				
Coverage	10	15	20	50	100
by the z-quantile	0.9359	0.9409	0.9398	0.9450	0.9472
by the adjusted t-quantile	0.9616	0.9602	0.9555	0.9526	0.9509
The percentage of the excluded runs	0.3914	0.0018	0	0	0

Table 4.1: Coverage of the confidence interval for the proportion and the percentage of the excluded runs for the parameter values $\lambda = 1$ for sample sizes $m = 10, 15, 20, 50, 100$ with the variance estimate based on the Poisson model.

	sample size m				
Coverage	10	15	20	50	100
by the z-quantile	0.9218	0.9292	0.9336	0.9446	0.9493
by the adjusted t-quantile	0.9541	0.9519	0.9499	0.9502	0.9533
The percentage of the excluded runs	0.4026	0.0023	0	0	0

Table 4.2: Coverage of the confidence interval for the proportion and the percentage of the excluded runs for the parameter values $\lambda = 1$ for sample sizes $m = 10, 15, 20, 50, 100$ with the sample variance.

Further, the coverage for more parameter values, i.e., $\lambda = 0.5, 2$, with the both variances estimates are given in the figures 4.3,4.4, 4.5, and 4.6.

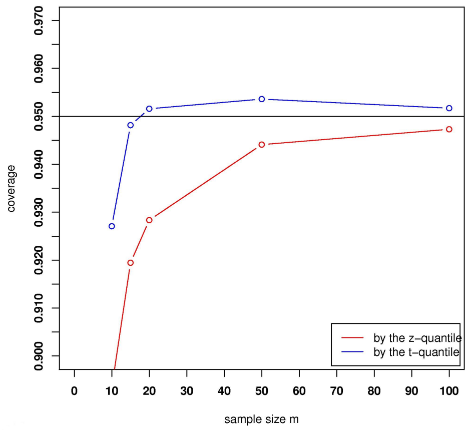

Figure 4.3: Coverage for $\lambda = 0.5$, $m = 10, 15, 20, 50, 100$ with the variance estimate based on the Poisson model.

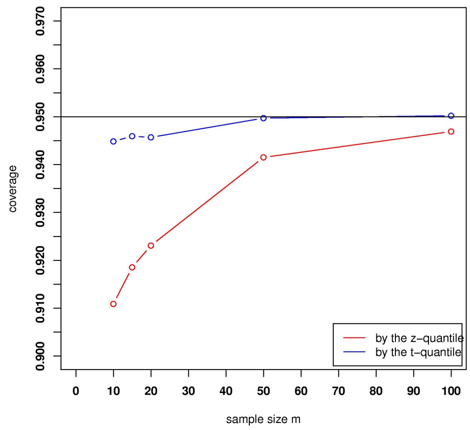

Figure 4.4: Coverage for $\lambda = 0.5$, $m = 10, 15, 20, 50, 100$ with the sample variance.

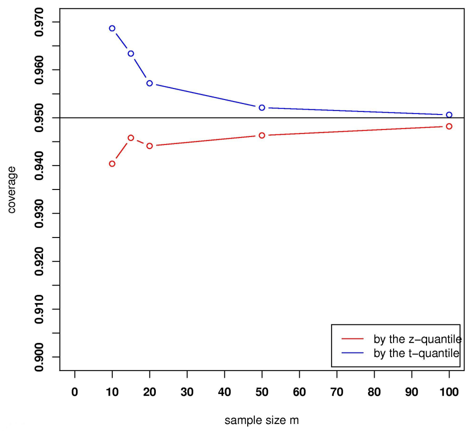

Figure 4.5: Coverage for $\lambda = 2$, $m = 10, 15, 20, 50, 100$ with the variance estimate based on the Poisson model.

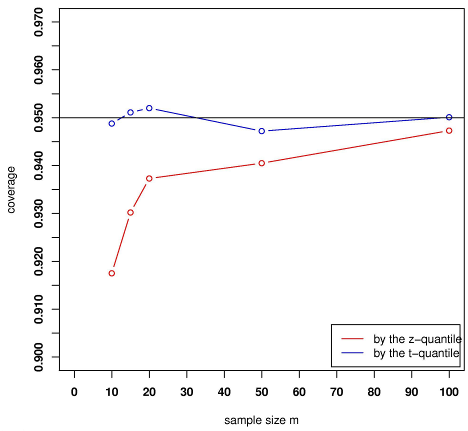

Figure 4.6: Coverage for $\lambda = 2$, $m = 10, 15, 20, 50, 100$ with the sample variance.

On the other hand, for the fixed sample sizes $10, 50$, and $\lambda = \{0.5, 0.8, 1, 2, 5\}$, we simulate the true coverage as following:

a) Using the sample variance.

From the figure 4.7, the coverage by the adjusted t-quantile method is conservative and gives more reasonable coverage especially for large λ, while by the z-quantile method the coverage is dramatically smaller than the nominal value, also there is a big difference between the both methods. For larger sample size (figure 4.8), the coverage by the t-quantile method are again more reasonable while by the z-quantile are slightly smaller with small difference between the coverage of both methods. Further, the tables 4.4 and 4.5 show that for all $\lambda = \{0.5, 0.8, 1, 2, 5\}$, there is no excluded runs when lambda or sample sizes are large.

Similar results hold at different confidence levels (because, the confidence intervals demonstrated the same pattern for all confidence levels, we focus on the level 0.95). See for example table 4.3 for more coverage at different confidence levels.

	Coverage by the z-quantile at the confidence level			
m	0.90	0.95	0.99	0.999
10	0.8697	0.9184	0.9728	0.9896
15	0.8795	0.9262	0.9740	0.9940
20	0.8863	0.9313	0.9820	0.9953
50	0.8909	0.9449	0.9872	0.9982
100	0.8967	0.9488	0.9879	0.9982
	Coverage by the adjusted t-quantile at the confidence level			
m	0.90	0.95	0.99	0.999
10	0.9041	0.9496	0.9904	0.9988
15	0.9053	0.9501	0.9888	0.9992
20	0.9045	0.9502	0.9904	0.9990
50	0.8997	0.9501	0.9902	0.9989
100	0.9006	0.9516	0.9899	0.9989

Table 4.3: Coverage for $\lambda = 1$, with the sample variance

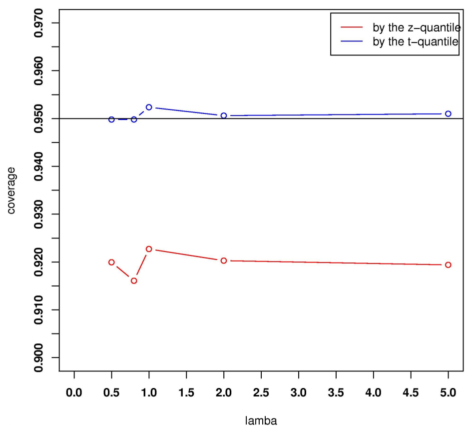

Figure 4.7: Coverage for the parameter values $\lambda = \{0.5, 0.8, 1, 2, 5\}$, $m = 10$ with the sample variance.

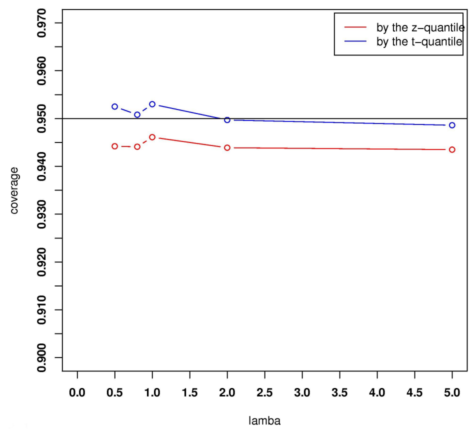

Figure 4.8: Coverage for the parameter values $\lambda = \{0.5, 0.8, 1, 2, 5\}$, $m = 50$ with the sample variance.

	λ				
Coverage	0.5	0.8	1	2	5
by the z-quantile	0.9199	0.9161	0.9227	0.9203	0.9194
by the adjusted t-quantile	0.9498	0.9498	0.9524	0.9506	0.9510
The percentage of the excluded runs	0.9263	0.6294	0.3996	0.0140	0

Table 4.4: Coverage and percentage of the excluded runs for parameter values $\lambda = \{0.5, 0.8, 1, 2, 5\}$, $m = 10$ with the sample variance.

b) Using the variance estimate based on the Poisson model.

Figure 4.9 shows that the t-quantile method is more conservative and gives larger coverage than that by the z-quantile which gives small coverage for all λ and smaller coverage when λ is small. When the sample size is larger (figure 4.10), the coverage are similar to that in the figure 4.8 but slightly larger (the estimated variance based on the Poisson model is larger than the sample variance). Also, from the tables 4.6 and 4.7, one can see that, when λ or sample sizes are larger then there is no excluded runs by the both methods.

Coverage	λ 0.5	0.8	1	2	5
by the z-quantile	0.9442	0.9441	0.9461	0.9439	0.9435
by adjusted the t-quantile	0.9525	0.9508	0.9530	0.9497	0.9486
The percentage of the excluded runs	0	0	0	0	0

Table 4.5: Coverage and percentage of the excluded runs for parameter values $\lambda = \{0.5, 0.8, 1, 2, 5\}$, $m = 50$ with the sample variance

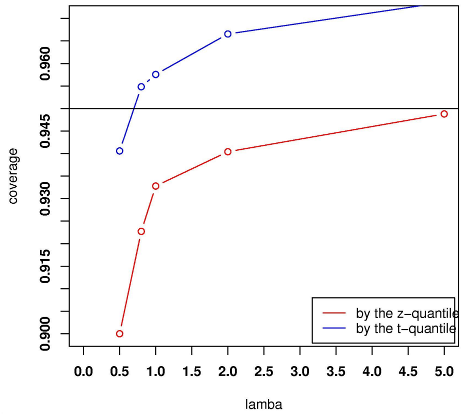

Figure 4.9: Coverage for λ, $m = 10$ with the variance estimate based on the Poisson model.

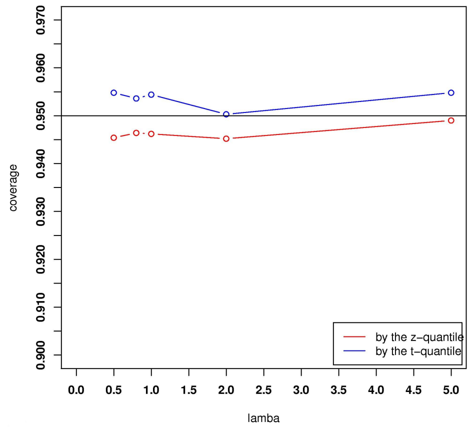

Figure 4.10: Coverage for λ, $m = 50$ with the variance estimate based on the Poisson model.

Coverage	λ 0.5	0.8	1	2	5
by the z-quantile	0.9000	0.9227	0.9328	0.9404	0.9488
by the adjusted t-quantile	0.9406	0.9548	0.9576	0.9666	0.9736
The percentage of the excluded runs	0.9310	0.6325	0.3944	0.0133	0

Table 4.6: Coverage and percentage of the excluded runs for λ for $m = 10$ with the variance estimate based on the Poisson model.

Coverage	λ 0.5	0.8	1	2	5
by the z-quantile	0.9454	0.9464	0.9462	0.9452	0.9490
by the adjusted t-quantile	0.9548	0.9536	0.9544	0.9503	0.9548
The percentage of the excluded runs	0	0	0	0	0

Table 4.7: Coverage and percentage of the excluded runs for λ for $m = 50$ with the variance estimate based on the Poisson model.

4.3 The Bivariate case

In this section, one would consider the coverage of the confidence intervals for the linear combination of the proportions $a_1 p_1 + a_2 p_2$, so we will simulate the sample data of sizes 10,15,20,50,100 of the data points (W_i, V_{i1}, V_{i2}) from Poisson distribution with 10,000 replications to consider closeness of the coverage to the nominal value. The purpose of this section is to compare the coverage of the runs using the Poisson estimated covariance or the SUR estimated covariance by the both quantiles with the corresponding nominal value.

Note that, small values of the model parameter will produce more noninformative observations with high probability (the noninformative observations is the set of events $A_i = \{X_{ij} = 0\}$, with probability of success $P(A_i) = e^{-\nu}$, and $P(\overline{A_i}) = 1 - e^{-\nu}$, where $\overline{A_i} = \{X_{ij} > 0\}, i = 1, ..., m, j = 1, 2$), or more unwanted negative variances produce invalid runs with invalid s.e's during the runs session in which will be excluded as well as a run is excluded, if $\sum_{i=1}^{m} X_{ij} = 0$ for any component j, by exclude these runs and will not be counted (also we mention here that we only accept runs as valid, for which the reduced sample size $m_j \, \forall j$ of informative observations is at least 9 to justify the use of the asymptotic approach), in addition one can also specify the percentage of the excluded runs. The procedure called 'A truncation of the invalid runs'.

Moreover, we need to take the following considerations:

- Taking only the informative observations will reduce the actual sample size m to the random sample size m_j. Theoretically, one can calculate the random number of the noninformative observations in each sample, which equals $m - m_j$, where $m_j \leq m$, as well as, the random percentage of the noninformative observations is $1 - \frac{m_j}{m}$.

- The valid runs are based on the positive variances (to ensure that, we will take only the positive estimated covariances) and the positive summations of the observations X_j. Further, the standard error is a consistent estimator for the corresponding positive asymptotic variance, so it converges to a positive number as sample size m tends to infinity, this will ensure the validity of the confidence interval at large sample sizes.

- We will adjust the degrees of freedom for the t-quantile on the reduced sample size m_1, where m_1 is the number of the informative observations of X_1. So, the conservative confidence interval (safety bounds) becomes: $\left[\boldsymbol{\alpha}^T\hat{\mathbf{p}} \pm t_{(m1-1,1-\frac{\alpha}{2})} s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}})\right]$, as well as, the approximate confidence interval is $\left[\boldsymbol{\alpha}^T\hat{\mathbf{p}} \pm z_{1-\frac{\alpha}{2}} s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}})\right]$, where $s.e(\boldsymbol{\alpha}^T\hat{\mathbf{p}}) = \sqrt{\frac{1}{m} as\widehat{Var(\boldsymbol{\alpha}^T\hat{\mathbf{p}})}}$.

a) For very small values of the Poisson parameter the procedure results in many non informative observations with high probability.

b) For small values (< 1), for example $\lambda_0 = 0.4, \lambda_1 = 0.5, \lambda_2 = 0.4, \mu_1 = 0.4, \mu_2 = 0.5$, the coverage by the z-quantiles method are smaller than the nominal level especially for $m \leq 20$, figure 4.11, because the simulation produces many noninformative observations or non positive covariances which may cause invalid s.e's, and hence the corresponding runs were excluded. From figure 4.12 (using the SUR estimated covariance), the both method gave coverage less than the nominal.

Tables 4.8, 4.9 show also the excluded runs decrease as sample sizes become larger by the both estimated covariances, as well as show larger exclusions for

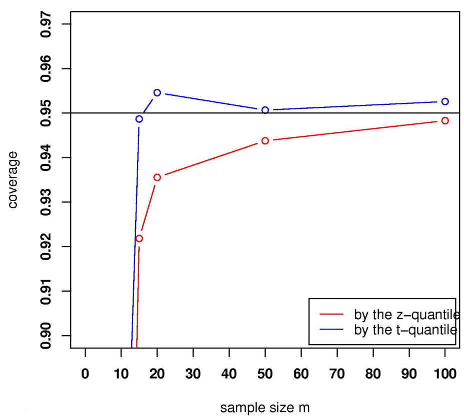

Figure 4.11: Coverage of the combination $a_1 p_1 + a_2 p_2$ using the Poisson estimated covariance at nominal value 0.95, for $\lambda_0 = 0.4, \lambda_1 = 0.5, \lambda_2 = 0.4, \mu_1 = 0.4, \mu_2 = 0.5, \alpha_1 = 1, \alpha_2 = 2$.

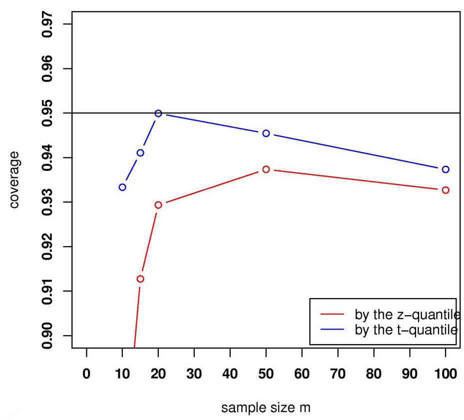

Figure 4.12: Coverage of the combination $a_1 p_1 + a_2 p_2$ using the SUR estimated covariance at nominal value 0.95, for $\lambda_0 = 0.4, \lambda_1 = 0.5, \lambda_2 = 0.4, \mu_1 = 0.4, \mu_2 = 0.5, \alpha_1 = 1, \alpha_2 = 2$.

c) For larger parameter values (≥ 1), for example $\lambda_0 = 4, \lambda_1 = 5, \lambda_2 = 4, \mu_1 = 4, \mu_2 = 5$ (figure 4.13), the confidence interval

Coverage	sample size m 10	15	20	50	100
by the z-quantile	0.7391	0.9218	0.9356	0.9438	0.9483
by the adjusted t-quantile	0.8261	0.9487	0.9546	0.9507	0.9526
The percentage of the excluded runs	0.9977	0.7058	0.2226	0.0005	0

Table 4.8: Coverage of the combination $a_1p_1 + a_2p_2$ using the Poisson estimated covariance at nominal value 0.95, for $\lambda_0 = 0.4, \lambda_1 = 0.5, \lambda_2 = 0.4, \mu_1 = 0.4, \mu_2 = 0.5, \alpha_1 = 1, \alpha_2 = 2$.

Coverage	sample size m 10	15	20	50	100
by the z-quantile	0.8667	0.9127	0.9294	0.9374	0.9327
by the adjusted t-quantile	0.9333	0.9411	0.9499	0.9455	0.9373
The percentage of the excluded runs	0.9970	0.7811	0.4168	0.1347	0.0551

Table 4.9: Coverage of the combination $a_1p_1 + a_2p_2$ using the the SUR estimated covariance at nominal value 0.95, for $\lambda_0 = 0.4, \lambda_1 = 0.5, \lambda_2 = 0.4, \mu_1 = 0.4, \mu_2 = 0.5, \alpha_1 = 1, \alpha_2 = 2$.

by t-quantile method is more conservative especially for small sample sizes and gives larger coverage, however by z-quantiles gives coverage slightly smaller but more reasonable. In figure 4.14, the SUR estimated covariance has been used, it is look like that the coverage in 4.13 are shifted down. Table 4.10 shows also the percentage of the truncated runs which tend to zero as sample sizes become larger, the interesting things from table 4.11 is that the percentage of the exclusions is larger due to the estimated conditional SUR covariance depends strictly on the observations of X_j which may causes many negative covariances that have been removed.

Coverage	sample size m 10	15	20	50	100
by the z-quantile	0.9435	0.9475	0.9473	0.9519	0.9485
by the adjusted t-quantile	0.9719	0.9663	0.9632	0.9579	0.9515
The percentage of the excluded runs	0.0731	0.0320	0.0159	0.0001	0

Table 4.10: Coverage using the Poisson estimated covariance, for $\lambda_0 = 4, \lambda_1 = 5, \lambda_2 = 4, \mu_1 = 4, \mu_2 = 5$.

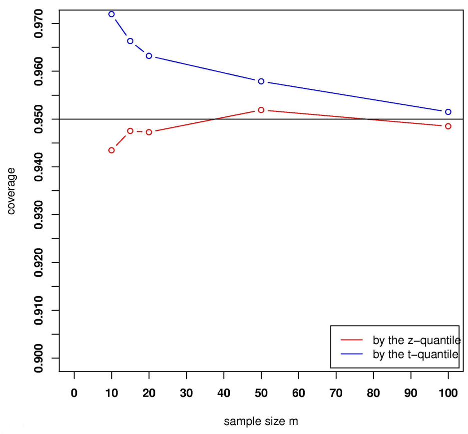

Figure 4.13: Coverage using the Poisson estimated covariance, for $\lambda_0 = 4, \lambda_1 = 5, \lambda_2 = 4, \mu_1 = 4, \mu_2 = 5$.

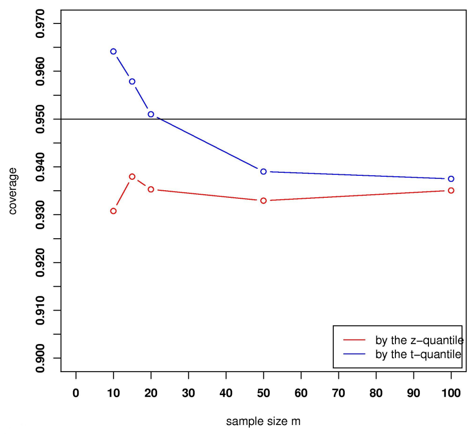

Figure 4.14: Coverage using the SUR estimated covariance, for $\lambda_0 = 4, \lambda_1 = 5, \lambda_2 = 4, \mu_1 = 4, \mu_2 = 5$.

	sample size m				
Coverage	10	15	20	50	100
by the z-quantile	0.9308	0.9380	0.9353	0.9329	0.9351
by the adjusted t-quantile	0.9641	0.9579	0.9510	0.9390	0.9375
The percentage of the excluded runs	0.3167	0.2810	0.2427	0.1292	0.0544

Table 4.11: Coverage using the SUR estimated covariance, for $\lambda_0 = 4, \lambda_1 = 5, \lambda_2 = 4, \mu_1 = 4, \mu_2 = 5$.

For $\lambda = 0.5, 0.8, 1, 2, 5$, sample sizes $m = 10, 50$:

- (with the Poisson estimated covariance):

  For the sample size 10 for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, figure 4.15, the coverage by the adjusted t-quantiles is more conservative and give larger coverage except for small parameter values, while by z-quantiles gives dramatically smaller coverage specially for small λ, there is also a big difference between the coverage by the two methods. However, for the sample size 50, figures 4.16 shows that both coverage by the both quantiles are more close to each other and to the nominal value.

  One can see from the tables 4.12, and 4.13, for small λ the percentage of excluded runs is more than that of large λ, as well as the exclusions decrease as λ or sample sizes become larger.

  This can be theocratically demonstrate as:

$P(X_{ij} = 0) = e^{(-\lambda_0+\lambda_j+\mu_j)} = e^{-3\lambda}$, $P(\sum_{i=1}^{m} X_{ij} = 0) = P(X_{ij} = 0 \text{ for all } i = 1, ..., m) = (e^{-3\lambda})^m$, and $P(\exists_j \sum_{i=1}^{m} X_{ij} = 0) \leq \sum_{j=1}^{2} P(\sum_{i=1}^{m} X_{ij} = 0) =$

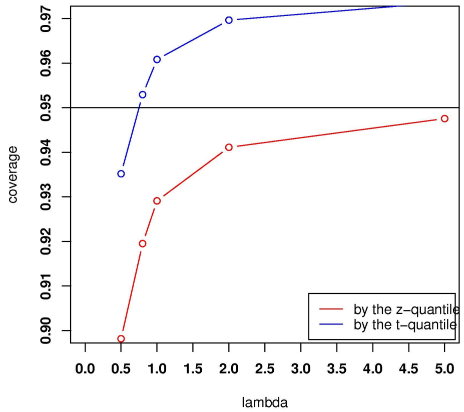

Figure 4.15: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=10 (with the Poisson estimated covariance).

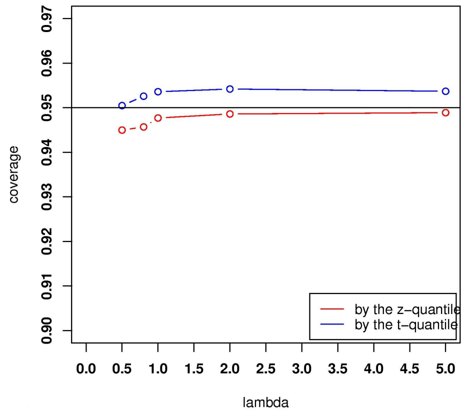

Figure 4.16: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=50 (with the Poisson estimated covariance).

	λ				
Coverage	0.5	0.8	1	2	5
by the z-quantile	0.8981	0.9195	0.9291	0.9411	0.9476
by the adjusted t-quantile	0.9352	0.9529	0.9608	0.9697	0.9737
The percentage of the excluded runs	0.9892	0.8385	0.6502	0.1067	0.0618

Table 4.12: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=10 (with the Poisson estimated covariance).

	λ				
Coverage	0.5	0.8	1	2	5
by the z-quantile	0.9450	0.9457	0.9477	0.9486	0.9489
by the adjusted t-quantile	0.9505	0.9526	0.9536	0.9542	0.9537
The percentage of the excluded runs	3e-04	3e-04	1e-04	0	1e-04

Table 4.13: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=50 (with the Poisson estimated covariance).

- (with the SUR estimated covariance):

  Figures 4.17 shows that for small sample size '10', the coverage by the z-quantiles are dramatically smaller than the nominal value, but by the t-quantile is conservative for large λ and gives larger coverage except for small λ, while for larger sample size '50' the both methods produce coverage smaller than the nominal

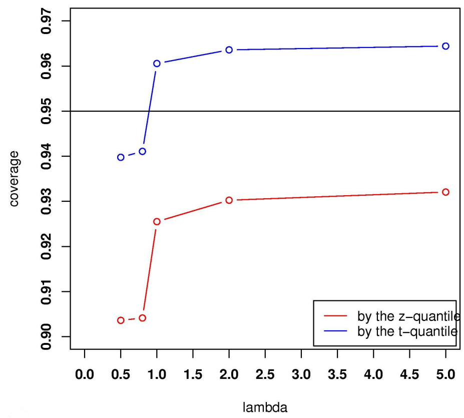

Figure 4.17: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=10 (wwith the SUR estimated covariance).

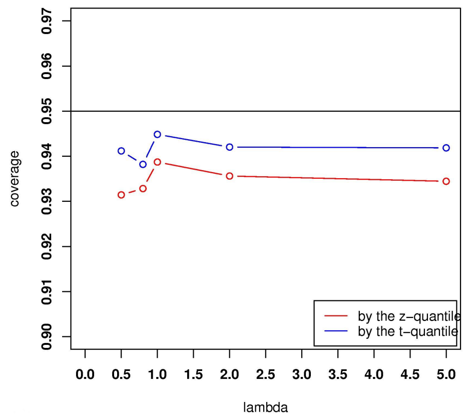

Figure 4.18: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=50 (with the SUR estimated covariance).

	λ				
Coverage	0.5	0.8	1	2	5
by the z-quantile	0.9036	0.9041	0.9255	0.9302	0.9321
by the adjusted t-quantile	0.9398	0.9411	0.9606	0.9636	0.9644
The percentage of the excluded runs	0.9917	0.8863	0.7490	0.3520	0.3171

Table 4.14: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=10 (with the SUR estimated covariance).

	λ				
Coverage	0.5	0.8	1	2	5
by the z-quantile	0.9314	0.9328	0.9387	0.9356	0.9344
by the adjusted t-quantile	0.9412	0.9382	0.9448	0.9420	0.9419
The percentage of the excluded runs	0.1279	0.1246	0.1352	0.1272	0.1229

Table 4.15: Coverage for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$, m=50 (with the SUR estimated covariance).

The same case for the percentage of excluded runs, the percentage in tables 4.14 and 4.15 are larger than that in the tables 4.12 and 4.13 respectively, because the conditional covariance estimators are used which are depend on the observations of X_j that based on the reduced sample size $m12$, where $m12$ is the number of the informative observations of X_1 and X_2 simultaneously, but also the exclusions are decrease as sample sizes and λ decrease.

Similar results hold for $\lambda = 0.5, 0.8, 1, 2, 5$, $m = 10, 15, 20, 50, 100$ for both estimated covariances. Figure 4.19 plotted for $\lambda = 2$ using the Poisson covariance, while using the SUR covariance shown in figure 4.20.
It seems that the coverage in figure 4.20 similar to 4.19 but shifted down.

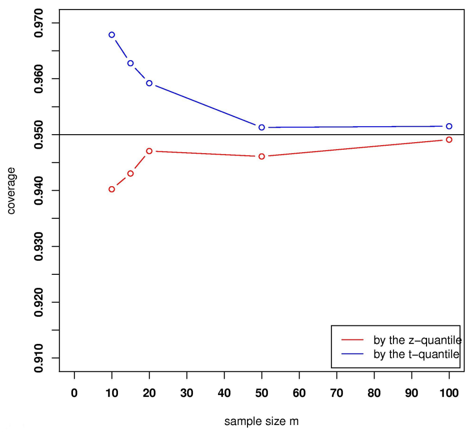

Figure 4.19: Coverage using the Poisson estimated covariance for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 2$.

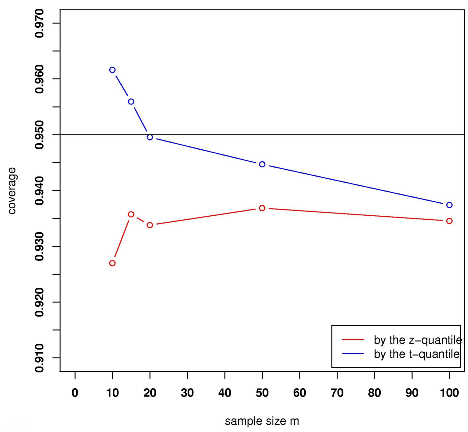

Figure 4.20: Coverage using the SUR covariance term for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 2$.

The corresponding coverage using the Poisson estimated covariance shown in table 4.16.

	Coverage by the z-quantile for λ				
m	0.5	0.8	1	2	5
10	0.9115	0.9305	0.9378	0.9402	0.9502
15	0.9247	0.9407	0.9401	0.9430	0.9501
20	0.9323	0.9412	0.9445	0.9471	0.9470
50	0.9448	0.9473	0.9450	0.9461	0.9478
100	0.9515	0.9463	0.9464	0.9491	0.9489
	Coverage by the adjusted t-quantile for λ				
m	0.5	0.8	1	2	5
10	0.9506	0.9617	0.9668	0.9679	0.9770
15	0.9532	0.9608	0.9616	0.9628	0.9655
20	0.9525	0.9556	0.9590	0.9592	0.9612
50	0.9526	0.9534	0.9507	0.9513	0.9531
100	0.9551	0.9485	0.9488	0.9515	0.9513

Table 4.16: Coverage using the Poisson estimated covariance at 0.95 for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$.

See tables 4.17 and 4.18 respectively, for more coverage on other confidence levels.

	Coverage by the z-quantile at confidence level			
m	0.90	0.95	0.99	0.999
10	0.8940	0.9398	0.9839	0.9980
15	0.8929	0.9468	0.9868	0.9984
20	0.8963	0.9442	0.9879	0.9988
50	0.8951	0.9491	0.9860	0.9984
100	0.8946	0.9480	0.9894	0.9989

	Coverage by the adjusted t-quantile at confidence level			
m	0.90	0.95	0.99	0.999
10	0.9258	0.9695	0.9967	1.0000
15	0.9151	0.9647	0.9959	1.0000
20	0.9147	0.9598	0.9931	0.9995
50	0.9036	0.9550	0.9894	0.9991
100	0.8978	0.9507	0.9906	0.9992

Table 4.17: Coverage for the parameter values $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 2$, using the Poisson estimated covariance.

	Coverage by the z-quantile at confidence level			
m	0.90	0.95	0.99	0.999
10	0.8728	0.9302	0.9789	0.9962
15	0.8752	0.9294	0.9806	0.9972
20	0.8709	0.9309	0.9811	0.9973
50	0.8773	0.9348	0.9857	0.9978
100	0.8742	0.9321	0.9832	0.9980

	Coverage by the adjusted t-quantile at confidence level			
m	0.90	0.95	0.99	0.999
10	0.9120	0.9642	0.9969	0.9999
15	0.9005	0.9492	0.9910	0.9997
20	0.8905	0.9472	0.9894	0.9995
50	0.8859	0.9425	0.9890	0.9990
100	0.8777	0.9360	0.9857	0.9983

Table 4.18: Coverage for the parameter values $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 2$, using the SUR estimated covariance.

4.4 Conclusions

Simulations are made to evaluate and validate the confidence intervals. The programs instructions made with flexible choices of the model parameter values and the sample sizes as well as the confidence levels, which enable achieving a more reasonable and appropriate coverage being close to the nominal value.

The confidence intervals generated by the adjusted t-quantiles method are more conservative (unless when the sample sizes or the parameter values are small), and give always larger coverage than the that by the z-quantiles method .

Small values of the model parameter produce results in many noninformative observations with high probability which can be theoretically calculated, or produce runs with invalid standard errors (negative variances) that will not be counted during the runs session, and will be excluded by taking the runs based only on the positive variances and positive summations of the observations of $X_j, j = 1, ..., k$.

As the sample size m or λ become larger, the percentage of the excluded runs become smaller and tends to zero. Further for large λ, the sample size $m \approx m_j, \forall j$.

The suggested confidence interval by the t-quantile method is too conservative for large λ (not recommended for a given parameters), also the corresponding coverage always larger than that by the z-quantile. While by the z-quantile is recommended when λ is large. To get better coverage, one may propose to take the average of the coverage of the both methods simultaneously.

Finally, the exclusions using the Poisson estimated covariance are less than the exclusion using the SUR covariance for all sample sizes. Both exclusions are decreasing as m, λ become larger.

Chapter 5

Contributions and Results

In each chapter of this dissertation two methods were used for the analysis of count data, one concerns the Linear or the SUR model, while the other concerns the Poisson model. Further, In this chapter we would mention that the assumption $X_{ij} > 0$ almost surely in our theoretical derivations, but in the Poisson model, which we employ in the simulation, this condition is violated, as $P(X_{ij} = 0) > 0$.

In chapter 1, the normality of the estimator $\hat{p}$ of the corresponding proportion p was studied based on the assumed linear model, assuming in the first part the normality of the errors of the linear model given that X_i are fixed variables, and the exact confidence intervals of the model coefficient (proportion) constructed. Further, the asymptotic normality of $\hat{p}_m$ under the non normal errors assumption given the $i.i.d$ of the observations X_i are obtained, and approximate confidence intervals of the proportion are constructed. In the second part, the distribution of the estimator $\hat{p}_m$ was discussed given that the observations X_i having Poisson distribution (univariate Poisson model), which results in the conditional Binomial distribution, and consequently the asymptotic normality of the conditional Binomial distribution of $\hat{p}_m$ was obtained, which was the asymptotic normality of a non-linear transformation of the pairs (Y_i, X_i) by using the Delta-method regardless the exact distribution of Y_i, and X_i. Consequently, the approximate confidence intervals of the proportions were identical.

In chapter 2 and 3, the bivariate and the multivariate normal distributions of the estimator vector $\hat{\mathbf{p}}$ of the proportion vector $\mathbf{p}$ have been assumed based on the SUR model, by assuming in the first part of each chapters, the normality of the error vectors of the SUR model given the fixed design vectors, where the constructed confidence intervals of the SUR model coefficient vector (proportion vector) were exact. The asymptotic bivariate and multivariate normality of $\hat{\mathbf{p}}_m$ under the non normal error vectors given that *iid* diagonal design matrices $\widetilde{X}_i$, are derived, and the corresponding approximate confidence intervals of a linear combination of the proportion vector have been constructed.

In the second part, the bivariate and the multivariate Poisson models in both chapters respectively including there definitions and properties were introduced and discussed. It was not explicitly possible to calculate the conditional distributions and the conditional covariance $Cov(Y_{i1}, Y_{i2} \mid X_{i1}, X_{i2})$ either in the multivariate case, and hence, the approximate confidence intervals for the linear combination of the proportions based on the models have been constructed by the asymptotic normality using the ***multivariate delta method***.

On the other side, in the both assumed models, the observations Y_{ij} conditionally depending on the observations X_{ij}, in other words, in the linear and SUR models, the observations X_{ij} considered as the constants or random variables, while in the Poisson models, the observations X_{ij} considered as Poisson random variables, $j = 1, ..., k$, and due to the correlations between the count data, the data were conditionally analyzed.

Furthermore, due to the dependence between the components, the BLUE's, BUE's which are, the best linear unbiased estimators, the best unbiased estimators of the proportions, and the variances respectively, satisfied only in the univariate case, while in a higher dimension case were not satisfied.

In the simulation chapter, it was taken the public available statistical software comprehensive R program to evaluate the approximate confidence intervals, and aid to see how the proportion or a linear combination of the proportions confidently fall in intervals having coverage closed to the nominal value.

Finally, this work may not considered as broader than that the wider contains many different techniques, however some were described and the required assumptions were given. Although some of the derivations were not included within the text but was refereed the reader to the reference where can he found the derivations, or the source of the used technique, further some knowledge of matrix algebra are covered in the appendix.

One may mention, that the open problems which can not be explicitly calculated:

- $P^{(Y_{ij} \mid X_{ij}, X_{ij'})} \sim ??$

- $P^{(Y_{ij},Y_{ij'}|X_{ij},X_{ij'})} \sim ??$
- $E(Y_{ij} \mid X_{ij}, X_{ij'}) = ??$
- $Var(Y_{ij} \mid X_{ij}, X_{ij'}) = ??$
- $Cov(Y_{ij}, Y_{ij'} \mid X_{ij}, X_{ij'}) = ??, j \neq j' = 1, ..., k.$

One may look for the BLUE estimators of p or the vector $\mathbf{p}$, namely

$\hat{\mathbf{p}}_{WLS} = (\widetilde{\mathbf{X}}^T \mathbf{\Sigma}^{-1} \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \mathbf{\Sigma}^{-1} \widetilde{\mathbf{Y}}$, or the estimator

$\widehat{\hat{\mathbf{p}}}_{WLS} = (\widetilde{\mathbf{X}}^T \widehat{\mathbf{\Sigma}}^{-1} \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \widehat{\mathbf{\Sigma}}^{-1} \widetilde{\mathbf{Y}}$, which are at least asymptotically efficient.

They are not the ratio estimator vectors, but give more appropriate confidence intervals for the linear combination of $\mathbf{p}$.

Appendix A

Supplementary Material

A.1 Background from the theory of Matrix Algebra

Let, $\mathbf{w}$ and $\mathbf{v}$ be two vectors having the same order, and let $\mathbf{A}$ and $\mathbf{B}$ are two squared symmetric matrices of the same dimension $n \times n$, then the following are available:

The inverse of A is denoted by $\mathbf{A}^{-1}$, the inverse exists and unique Iff $\mathbf{A}$ is non-singular, where $\mathbf{A}$ is a nonsingular Iff its determinate $\mid \mathbf{A} \mid \neq 0$, for which $\mid \mathbf{A}^{-1} \mid = \frac{1}{|\mathbf{A}|}$, and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = I_n$. Furthermore if $\mathbf{A}$, and $\mathbf{B}$ are invertible or nonsingular, then

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Idempotent matrix A is called ***symmetric idempotent*** if $\mathbf{A} = \mathbf{AA} = \mathbf{A}^T\mathbf{A}$, so it follows that:

- $I_n - \mathbf{A}$ is symmetric idempotent.
- $\mathbf{A}(I_n - \mathbf{A}) = \mathbf{0}$, and $(I_n - \mathbf{A})\mathbf{A} = \mathbf{0}$.
- $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is also symmetric idempotent.

Trace of A is denoted by $trace(\mathbf{A})$ or $tr(\mathbf{A})$, where $tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$, a_{ii} are the diagonal elements of A. Some properties of the trace are given by the following:

- $tr(\mathbf{A} \pm \mathbf{B}) = tr(\mathbf{B} \pm \mathbf{A}) = tr(\mathbf{A}) \pm tr(\mathbf{B})$.
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$[1].

1 it holds also for any matrices $\mathbf{A}, \mathbf{B}$ for dimensional $n \times p, p \times n$ respectively, where also for any $n < p, n > p, n = p$.

- $tr(\mathbf{A}^T) = tr(\mathbf{A})$.
- $tr(\mathbf{A}^T\mathbf{A}) = tr(\mathbf{A}\mathbf{A}^T) = \sum_{i=1}^{n} \mathbf{a}_i^T\mathbf{a}_i = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$, where $\mathbf{a}_i$ is the i^{th} row vector of $\mathbf{A}$.
- $tr(k\mathbf{A}) = k\ tr(\mathbf{A})$, where k is a real number.

Rank of an idempotent matrix A is its trace, where the ***rank of A*** is the number of linearly independent columns, or the no.of linearly independent rows.

orthogonal vectors w and v are ***orthogonal vectors*** if the vector product

$\mathbf{w}^T\mathbf{v} = \mathbf{v}^T\mathbf{w} = 0$.

Quadratic Form, the function $\mathbf{w}^T\mathbf{A}\mathbf{w} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}w_iw_j$ is called the ***quadratic form***, and $\mathbf{A}$ is called the ***matrix of quadratic form***.

Positive definite and positive semi definite matrices, matrix $\mathbf{A}$ is said to be ***positive definite*** if $\mathbf{w}^T\mathbf{A}\mathbf{w} > 0\ \forall\mathbf{w} \in R^n,\ \mathbf{w} \neq 0$, and said to be ***positive semi definite*** if $\mathbf{w}^T\mathbf{A}\mathbf{w} = 0$, for some $\mathbf{w} \neq 0$.

if A is positive definite matrix, then $\mid \mathbf{A} \mid > 0$, $\mid \mathbf{A}^{-1} \mid > 0$, and it follows that all its diagonal elements $a_{ii} > 0, \forall i = 1, ..., n$, similarly for ***positive semi definite matrix,*** we replace $>$ by $\geq$. Further if $\mathbf{A}$ is diagonal matrix then $\mid \mathbf{A} \mid = \prod_{i=1}^{n} a_{ii}$.

A.2 Background from the theory of the linear models and the MSUR model

- expectation of the quadratic forms
  let $\mathbf{Y} = (Y_1, ..., Y_n)^T$ be the univariate random vector of size n with mean vector $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, and variance-covariance matrix $Var(\mathbf{Y}) = \sigma^2 I_n$, where $\mathbf{X}$ is the design matrix of $n \times k$ covariates, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^T$, then

$$E(\mathbf{Y}^T\mathbf{A}\mathbf{Y}) = \sigma^2 tr(\mathbf{A}I_n) + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{A}\mathbf{X}\boldsymbol{\beta} = \sigma^2 tr(\mathbf{A}) + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{A}\mathbf{X}\boldsymbol{\beta} \qquad (A.1)$$

- if $\mathbf{Z} = \mathbf{AY}$, then

$$
\begin{aligned}
Cov(\mathbf{Z}) &= E((\mathbf{Z} - E(\mathbf{Z}))(\mathbf{Z} - E(\mathbf{Z}))^T) = E\left((\mathbf{AY} - \mathbf{AX}\boldsymbol{\beta})(\mathbf{AY} - \mathbf{AX}\boldsymbol{\beta})^T\right) \\
&= E(\mathbf{A}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{A}^T) \\
&= \mathbf{A}(E(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T)A^T = \mathbf{A}Cov(\mathbf{Y})\mathbf{A}^T = \sigma^2\mathbf{A}\mathbf{A}^T \qquad (A.2)
\end{aligned}
$$

The least squares estimator $\hat{\boldsymbol{\beta}}_{OLS}$ is obtained by minimizing the sum of the squired deviations of the observations from their expected values. Hence minimizing
$S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ which leads to the system of normal equations $(\mathbf{X}^T\mathbf{X})^{-1}\hat{\boldsymbol{\beta}}_{OLS} = \mathbf{X}^T\mathbf{Y}$, assuming that $(\mathbf{X}^T\mathbf{X})^{-1}$ is invertible, the OLSE $\hat{\boldsymbol{\beta}}_{OLS}$ can be written explicitly as
$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, which is a linear function of $\mathbf{Y}$, the vector of fitted values $\hat{\mathbf{Y}}$ corresponding to the observed $\mathbf{Y}$ is $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = H\mathbf{Y}$, where $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is known as the hat or projection matrix which plays a central role in linear model analysis, the vector of residuals is given by

$$
\begin{aligned}
\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - H\mathbf{Y} = (I_n - H)\mathbf{Y}, \text{ and more} \\
E(\hat{\mathbf{Y}}) = HE(\mathbf{Y}) = H\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}, \qquad (A.3)
\end{aligned}
$$

and thus, $\hat{\mathbf{Y}}$ is an unbiased estimator of the mean of $\mathbf{Y}$.

On the other side, if $Var(\mathbf{Y}) = \sigma^2\Omega$, where Ω is a positive definite matrix but not equal to I_n, then it may be possible to implement a generalized least squares (GLSE) estimator that is the BLUE (at least asymptotically), so the GLSE estimator $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{Y}$ is the BLUE, with the variance-covariance matrix $(\mathbf{X}^T\Omega^{-1}\mathbf{X})^{-1}$. Note that when $\Omega = I_n$, then the GLSE = OLSE with the covariance $= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$(i.e, OLS is a special case of the more general estimator).

- If $\mathbf{Y} = (\mathbf{Y}_1^T, ..., \mathbf{Y}_k^T)^T$ is the multivariate columns wise expansion of the random vector $\mathbf{Y}$ of dimension $nk \times 1$ with mean vector $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, and variance-covariance matrix $Var(\mathbf{Y}) = \Sigma \otimes I_n$, (assuming $\Sigma = \widetilde{\Sigma}$), where $\mathbf{X}$ is the design matrix of dimension $nk \times k$ covariates, and $\boldsymbol{\beta}$ is a vector of $k \times 1$ of the unknown parameters. Then, the least squares estimator $\hat{\boldsymbol{\beta}}$ is obtained by minimizing the sum of squared deviations of the observations from their expected values. Hence

minimizing
$S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ which leads to the system of normal equations $(\mathbf{X}^T\mathbf{X})^{-1}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$, assuming that $(\mathbf{X}^T\mathbf{X})^{-1}$ is invertible, the OLSE $\hat{\boldsymbol{\beta}}$ of the SUR model parameter vector can be written explicitly as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, which is the linear function of $\mathbf{Y}$, as well as, the WLSE is given by:
$\hat{\boldsymbol{\beta}}_{WLS} = \left(\mathbf{X}^T(\Sigma \otimes I_n)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T(\Sigma \otimes I_n)^{-1}\mathbf{Y}$.
There are two cases where the WLSE reduces to the OLSE:

- if $\sigma_{jj'} = 0$, $\forall j \neq j' = 1, ..., k$, i.e, Σ is diagonal.
- if $\mathbf{X}_1 = \mathbf{X}_2 = ... = \mathbf{X}_n = \mathbf{X}_0$.

These two cases are proved by Zellner (1962), see [25], [26].

Furthermore:

- If $\mathbf{Y} = (\mathbf{Y}_1^T, ..., \mathbf{Y}_n^T)^T$ is the multivariate rows wise random vector of dimension $nk \times 1$ with mean vector $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, and with variance-covariance matrix $Var(\mathbf{Y}) = I_n \otimes \Sigma$, (assuming $\Sigma = \widetilde{\Sigma}$), where $\mathbf{X}$ is the design matrix of $nk \times k$ covariates, and $\boldsymbol{\beta}$ is a vector of $nk \times 1$ of unknown parameters. Then, the WLSE of the SUR model parameter vector is given by

$$\hat{\boldsymbol{\beta}}_{WLS} = \left(\mathbf{X}^T(I_n \otimes \Sigma)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T(I_n \otimes \Sigma)^{-1}\mathbf{Y},$$

consequently

$$Cov(\hat{\boldsymbol{\beta}}_{WLS}) = \left(\mathbf{X}^T(I_n \otimes \Sigma)^{-1}\mathbf{X}\right)^{-1} = \left(\mathbf{X}^T(I_n \otimes \Sigma^{-1})\mathbf{X}\right)^{-1} = \left(\sum_{i=1}^{n}\mathbf{X}_i^T\Sigma^{-1}\mathbf{X}_i\right)^{-1}, \tag{A.4}$$

whereas

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}_{OLS}) &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T(I_n \otimes \Sigma)\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \\ &= \left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\Sigma\mathbf{X}_i\left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i\right)^{-1}, \end{aligned} \tag{A.5}$$

and are equivalent when Σ is diagonal matrix.

Bibliography

[1] Aitken, A.C. (1936). A further note on multivariate selection. *Proceedings of the Edinburgh Mathematical Society*, **5**, pp. 37-40.

[2] Campbell, J.T. (1938). The Poisson correlation function. *Proccedings of the Edinburgh Mathematical Society (series 2)*, **4**, pp. 18-26.

[3] Consael, R. (1952). Sur les processus composés de Poisson á deux variables aléatoires. Académic Royale de Belgique, Classe des Sciences, Mémoires 27, No 6, 4-43.

[4] Finkelstein M, Tucker H.G and Veeh J.A. (2000). Conservative confidence intervals for a single parameter.*Communications in Statistics - Theory and Methods, Volume 29, Issue 8,* pp. 1911 - 1928.

[5] Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika*, **51**, pp. 241-245.

[6] Johnson. N, Kotz. S and Balakrishnan. N. (1997). Discrete Multivariate Distributions. *Wiley*, New York.

[7] Johnson R.A and Wichern D.W. (1988). Applied Multivariate Statistical Analysis. Third Edition. *Prentice-Hall.*

[8] Kapadia A.S, Chan W and Moye L. (2005). Mathematical Statistics with Application. *Taylor & Francis Group*, Houston.

[9] Karlis, D and Ntzoufras, J. (2000). On modelling soccer data. *Student*, 3, 229-244.

[10] Karlis, D and Ntzoufras, J. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician*, **52**, part **3**, pp. 381-393.

[11] Karlis, D and Tsiamyrtzis, P. (2007). Exact Bayesian modeling for bivariate Poisson data and extensions. *Stat Comput 18:* pp. 27-40.

[12] Karlis, D and Meligkotsidou, L. (2003). Model based clustering for multivariate count data. *Proceedings of the 18th International Workshop on Statistical Modelling, Leuven, Belgium,* pp. 375-382.

[13] Karlis, D and Meligkotsidou, L. (2005). Multivariate Poisson Model with covariance structure. Statistics and Computing **15**, pp. 255-265.

[14] Knight, K.C. (2000). Mathematical Statistics. *Chapman & Hall/CRC.*

[15] Landers, D and Rogge, C. (1976). The exact approximation order in the Central Limit theorem for random summation. *Wahrscheinlichkeitstheorie verw.Gebiete,* **36**, pp. 269-283.

[16] Montgomery, D.C , Peck, E.A and Vining, G.G. (2001). Introduction to linear Regression analysis. Third Edition, *Wiley*, New York.

[17] McLachlan G and Krishnan T. (1997). The EM Algorithm and Extensions, Wiley, New York.

[18] Rawlings J.O, Pantula S.G and Dickey D.A. (1998). Applied Regression analysis. A research Tool. *Springer*, New York.

[19] Rencher, A.C, Schaalje, G.B. (2000). Linear models in Statistics. *Wiley*, New York.

[20] Sen, A.K. and Srivastava, M.S. (1990). Regression Analysis: Theory, Methods, and Application . *Springer, New York.*

[21] Srivastava, V.K and Upadhyaya, S. (1978). Large Sample approximation in Seemingly Unrelated Regression Equations. *Ann. Inst. Statist. Math.* Part A, 89-96.

[22] Srivastava, V.K and Dwivedi, T.D. (1979). Estimation of seemingly unrelated regression equations. *Journal of Econometrics 10*, pp. 15-32.

[23] Stock, J.H and Watson, M.W. (2003). Introduction to econometrics. *Addison-Wesley, Boston, Mass.*

[24] Timm, N.H. (2002). Applied Multivariate analysis. *Springer*, New York.

[25] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *J. Amer. Statist. Ass.* **57**, 348-368.

[26] Zellner, A. (1963). Estimators for seemingly unrelated regression equations: Some exact finite sample results, *J. Amer. Statist. Ass.* **58**, 977-992, Corrigenda **67**, 255.