A MULTIVARIATE INDEX OF DISPERSION

A MULTIVARIATE INDEX OF DISPERSION

BY
DEAN HANSEN, B.Sc

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS AND STATISTICS
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Master of Science (2024)
(Mathematics & Statistics)

McMaster University
Hamilton, Ontario, Canada

TITLE:	A Multivariate Index of Dispersion
AUTHOR:	Dean Hansen Honours B.Sc. (Mathematics & Statistics), McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Narayanaswamy Balakrishnan
NUMBER OF PAGES:	x, 56

Abstract

The Fisher dispersion index is a widely used measure of dispersion in count data. In the univariate case, the dispersion index is equal to the ratio of the variance to the mean. In the multivariate case, such a widely used and applied dispersion index does not exist. In this thesis, we review some previously proposed multivariate extensions and introduce a new multivariate dispersion index. We illustrate its properties on some common multivariate discrete distributions and demonstrate its usefulness through a simulation study and real-world examples.

Keywords: Count Data · Dispersion · Hadamard Product

To Eden, Alan, Drew, Katie and Abby

Acknowledgements

I would like to express sincere thanks to my supervisor Dr. Balakrishnan for his support and encouragement during my graduate studies. When I needed inspiration or, more often than not, a new thesis topic, he was always there to steer me in the right direction. He is a truly great statistician, and more importantly, person. Finally, I would like to thank my family, my parents Eden and Alan Hansen, and my brother, Drew Hansen. I love you all very much, and I would be remiss if I did not tell you once in a while.

Table of Contents

List of Figures

List of Tables

Notation, and Abbreviations

Notation

$A \otimes B$	Hadamard Product of A and B
$\text{cov}(X_1, X_2)$	Covariance between X_1 and X_2.
$\text{corr}(X_1, X_2)$	Correlation between X_1 and X_2.

Abbreviations

DI	Fisher Dispersion Index
MLE	Maximum Likelihood Estimation
PGF	Probability Generating Function
PMF	Probability Mass Function
Pois(λ)	Poisson Distribution
NB(r,p)	Negative Binomial Distribution
BivBern($p_{00}, p_{10}, p_{01}, p_{11}$)	Bivariate Bernoulli Distribution
BivPois($\lambda_1, \lambda_2, \lambda_3$)	Bivariate Poisson Distribution
BivNB($\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2$)	Bivariate Negative Binomial Distribution
BivPA($\lambda_1, \lambda_2, \lambda_3, \rho$)	Bivariate Pólya-Aeppli Distribution
MVPois($\lambda_1, \ldots, \lambda_p$)	Multivariate Poisson Distribution

Declaration of Academic Achievement

This is a declaration that the contents of this thesis are my own.

Chapter 1

Introduction

Introduced by Ronald Fisher (1934), the dispersion index is defined as the ratio of the variance to the mean

$$\mathrm{DI} = \frac{\mathrm{Var}(Y)}{\mathrm{E}(Y)}. \tag{1.0.1}$$

When $\mathrm{DI} = 1$, we say that X is equi-dispersed. When $\mathrm{DI} < 1$ we say that X is under-dispersed and over-dispersed when $\mathrm{DI} > 1$. The dispersion index is often used to assess whether a set of univariate count data can be modelled using the Poisson distribution, which is always equi-dispersed. This assumption of the Poisson distribution is highly restrictive in practice as many datasets display over/under-dispersion (Hilbe (2014)).

A distribution is said to be over-dispersed relative to the Poisson distribution if it has $\mathrm{DI} > 1$. Similarly, a distribution is said to be under-dispersed relative to the Poisson distribution if $\mathrm{DI} < 1$. For example, the Bernoulli distribution is under-dispersed relative to the Poisson distribution since $\mathrm{Var}(X)/\mathrm{E}(X) = p(1-p)/p = (1-p) < 1$. Further, the negative binomial distribution is over-dispersed relative to the Poisson distribution since $\mathrm{Var}(X)/\mathrm{E}(X) = 1/p > 1$ when $p \in (0, 1)$.

In the multivariate setting, measuring dispersion is not so straightforward. One key challenge is that the variance and expected value of a distribution become matrix and vector valued, so care must be taken when defining an appropriate measure of dispersion. Further, a multivariate dispersion index should account for dependency between random variables, as this will influence the degree of dispersion present in data.

Considering these challenges, recent developments have been made in extending the dispersion index to a multivariate setting. Kokonendji and Puig (2018) have proposed the multivariate generalized and marginal dispersion indices (GDI and MDI) and Minkova and Balakrishnan (2014a) introduced FI_2

as a bivariate extension of the dispersion index. In this thesis, we will introduce a multivariate extension of FI_2 that provides a natural generalization of equi-dispersion from the bivariate to p-variate case. Key properties of GDI, MDI and FI_p will be explored in later sections, and their comparison will form the basis of our simulation study and examples.

1.1 Thesis Outline

The thesis is structured as follows. In Chapter 2, we introduce some bivariate discrete distributions and discuss their key properties. In Chapter 3, the methodology used for the simulation study is presented alongside the maximum likelihood estimates for some bivariate discrete distributions from Chapter 2. In Chapter 4, we introduce FI_p, our proposed multivariate dispersion index, and illustrate some of its key properties. In Chapter 5, we present a simulation study and applications of GDI, MDI and FI_p to some datasets of interest. Finally, Chapter 6 offers an overview of the thesis and suggestions for future research.

Chapter 2

Background

2.1 Bivariate Discrete Distributions

Here we introduce some bivariate and p-variate discrete distributions that will be of use in later sections. For a more detailed discussion of univariate and multivariate discrete distributions please refer to Johnson et al. (1997), Johnson et al. (2005) and Kocherlakota and Kocherlakota (2017).

2.1.1 Bivariate Bernoulli

Recall that Y follows a Bernoulli distribution if it has the following probability mass function (PMF) and probability generating function (PGF)

$$P(Y = y) = p^y(1-p)^{1-y} \tag{2.1.1}$$

$$\psi(s) = (1-p) + ps \tag{2.1.2}$$

where $y \in \{0, 1\}$, $p \in [0, 1]$ and $s \in \mathbb{R}$. The expected value and variance of the Bernoulli distribution are $\mathrm{E}(Y) = p$ and $\mathrm{Var}(Y) = p(1-p)$ respectively.

Consider the random vector $\mathbf{Y} = (Y_1, Y_2)$, where Y_i follows a Bernoulli distribution. The four possible outcomes for $\mathbf{Y}$ are $(0, 0), (0, 1), (1, 0), (1, 1)$ with probabilities p_{00}, p_{10}, p_{01} and p_{11} respectively. We say $\mathbf{Y} = (Y_1, Y_2)$ follows a bivariate Bernoulli distribution if it has the following joint probability mass function and joint probability generating function

$$P(y_1, y_2) = p_{11}^{y_1 y_2}\, p_{10}^{y_1(1-y_2)}\, p_{01}^{(1-y_1)y_2}\, p_{00}^{(1-y_1)(1-y_2)} \tag{2.1.3}$$

$$\psi(s_1, s_2) = \mathrm{E}(s_1^{y_1} s_2^{y_2}) = (s_1 s_2 p_{11} + s_1 p_{10} + s_2 p_{01} + p_{00}) \tag{2.1.4}$$

where $y_i \in \{0, 1\}$, $p_{ij} \in [0, 1]$ and $(s_1, s_2) \in \mathbb{R}^2$. The marginal distributions of $\mathbf{Y}$, $P(Y_1 = y_1)$ and $P(Y_2 = y_2)$, are readily seen to follow a Bernoulli distribution with parameters $p_{00}+p_{10}$ and $p_{01}+p_{11}$ respectively. The expected value and variance of Y_i are $\mathrm{E}(Y_i) = p_{0i}+p_{1i}$ and $\mathrm{Var}(Y_i) = (p_{0i}+p_{1i})(1-p_{0i}-p_{1i})$ for $i = 1, 2$. We have $\mathrm{E}(Y_1Y_2) = \Pr(Y_1 = 1, Y_2 = 1) = (p_{00}+p_{10})(p_{01}+p_{11})$. Thus, the covariance and correlation of Y_1 and Y_2 are given by

$$\mathrm{Cov}(Y_1, Y_2) = p_{00}p_{11} - p_{10}p_{01} \tag{2.1.5}$$

$$\mathrm{Corr}(Y_1, Y_2) = \frac{p_{00}p_{11} - p_{10}p_{01}}{\sqrt{(p_{00} + p_{10})(1 - p_{00} - p_{10})(p_{01} + p_{11})(1 - p_{01} - p_{11})}}. \tag{2.1.6}$$

2.1.2 Bivariate Poisson

Recall that Y follows a Poisson distribution if it has the following PMF and PGF

$$P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!} \tag{2.1.7}$$

$$\psi(s) = e^{\lambda(s-1)} \tag{2.1.8}$$

where $y \in \mathbb{N}$, $\lambda > 0$ and $s \in \mathbb{R}$. The expected value and variance of the Poisson distribution are $\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$.

The bivariate Poisson distribution arises in several ways though, here we follow the trivariate reduction method used originally by Holgate (1964). Let W_1, W_2 and W_3 be independent Poisson random variables with parameters λ_1, λ_2 and λ_3. Define the random vector $\mathbf{Y} = (Y_1, Y_2)$ where $Y_1 = W_1 + W_3$ and $Y_2 = W_2 + W_3$. We say $\mathbf{Y} = (Y_1, Y_2)$ follows a bivariate Poisson distribution if it has the following joint probability mass function and joint probability generating function

$$P(y_1, y_2) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{k=0}^{\min(y_1,y_2)} \frac{\lambda_1^{(y_1-k)}\lambda_2^{(y_2-k)}\lambda_3^k}{(y_1 - k)!\,(y_2 - k)!\,k!} \tag{2.1.9}$$

$$\psi(s_1, s_2) = e^{(\lambda_1+\lambda_3)(s_1-1)+(\lambda_2+\lambda_3)(s_2-1)+\lambda_3(s_1-1)(s_2-1)} \tag{2.1.10}$$

where $\lambda_i > 0$ and $(s_1, s_2) \in \mathbb{R}^2$. The marginal distributions of $\mathbf{Y}$, $P(Y_1 = y_1)$ and $P(Y_2 = y_2)$, are readily seen to follow Poisson distributions with parameters $\lambda_1 + \lambda_2$ and $\lambda_2 + \lambda_3$ respectively. The expected value and variance of Y_i are $\mathrm{E}(Y_i) = \mathrm{Var}(Y_i) = \lambda_i + \lambda_3$ for $i = 1, 2$. We have $\mathrm{E}(Y_1Y_2) = \mathrm{E}((W_1 + W_3)(W_2+W_3)) = \lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3+\lambda_3^2$. Thus, the covariance and correlation of Y_1 and Y_2 are given by

$$\mathrm{Cov}(Y_1, Y_2) = \lambda_3 \tag{2.1.11}$$

$$\text{Corr}(Y_1, Y_2) = \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}. \tag{2.1.12}$$

Below we visualize the PMF with $\lambda_1 = 2$, $\lambda_2 = 3$ and $\lambda_3 \in \{1, 2, 3, 4\}$. The lightly coloured cells indicate large values of the PMF, whereas the darker cells indicate areas with smaller values of the PMF.

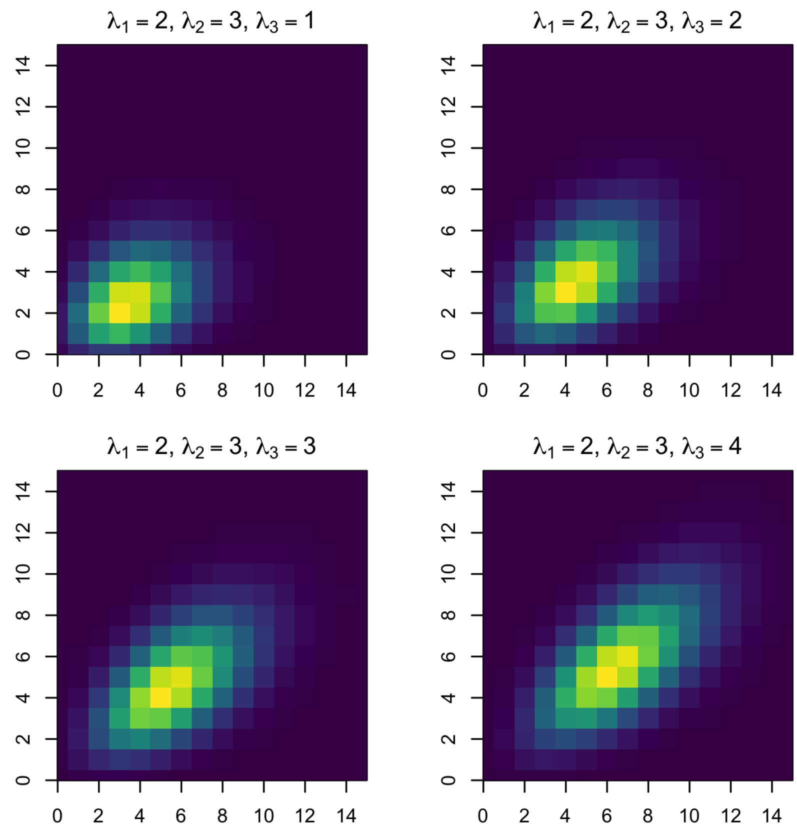

Figure 2.1: Joint PMF of Bivariate Poisson distribution with parameters $(\lambda_1, \lambda_2, \lambda_3)$ evaluated on $\{0, 15\}^2$

2.1.3 Bivariate Negative Binomial

Recall that Y follows a Negative Binomial distribution if it has the following PMF and PGF

$$P(Y = y) = \frac{(y + r - 1)!}{y!\,(r - 1)!}(1 - p)^y\, p^r \tag{2.1.13}$$

$$\psi(s) = \left\{ \frac{p}{1-(1-p)s} \right\}^r \tag{2.1.14}$$

where $y \in \{0, 1, 2, \ldots\}$, $p \in [0, 1]$ and $\mid s \mid < \frac{1}{p}$. The expected value and variance of this distribution are $\mathrm{E}(Y) = \frac{r(1-p)}{p}$ and $\mathrm{Var}(Y) = \frac{r(1-p)}{p^2}$ respectively.

The univariate Negative Binomial distribution arises in the analysis of wait times. If we consider an infinite number of Bernoulli trials with probability of success p, and let Y denote the number of failures before the r^{th} success, then Y follows a Negative Binomial distribution. An alternative way to derive the Negative Binomial distribution arises in the analysis of accident proneness.

Originally derived by Greenwood and Yule (1920), assume that the number of accidents experienced by α individuals in a period of time follows a Poisson distribution with rate parameter $\lambda > 0$. Let Y denote this random variable with probability mass function

$$P(Y = y \mid \lambda) = e^{-\alpha\lambda} \frac{(\alpha\lambda)^y}{y!} \tag{2.1.15}$$

where $y \in \mathbb{N}$. Let $\lambda \sim \mathrm{Gamma}(\nu, \beta)$ with probability density function

$$P(\lambda) = \frac{\beta^\nu}{\Gamma(\nu)} \lambda^{\nu-1} e^{-\beta\lambda} \tag{2.1.16}$$

where $\lambda > 0$, $\nu > 0$ and $\beta > 0$. Then the unconditional distribution of Y is given by

$$P(Y = y) = \frac{\Gamma(y+\nu)}{y! \cdot \Gamma(\nu)} \left(\frac{\alpha}{\alpha+\beta} \right)^y \left(\frac{\beta}{\alpha+\beta} \right)^v \tag{2.1.17}$$

$$= \frac{\Gamma(y+\nu)}{y! \cdot \Gamma(\nu)} \left(1 - \frac{\beta}{\alpha+\beta} \right)^y \left(\frac{\beta}{\alpha+\beta} \right)^v \tag{2.1.18}$$

which is Negative Binomial with parameters ν and $p = \frac{\beta}{\alpha+\beta}$ (from 2.1.13).

Similar to the bivariate Poisson distribution, the bivariate Negative Binomial distribution arises in several ways. Here we follow the compounding method originally introduced by Edwards and Gurland (1961) and later independently discovered by Subrahmaniam (1966). We note that the original derivation of the Negative Binomial distribution was given by Arbous and Kerrich (1951) and Bates and Neyman (1952).

Let $\mathbf{Y} = (Y_1, Y_2)$ represent the number of accidents incurred by the same individual for two types of accidents. Assume $\mathbf{Y}$ follows a bivariate Poisson distribution with parameter $\lambda \sim \text{Gamma}(\nu, \beta)$ and joint conditional probability generating function

$$\psi(s_1, s_2 \mid \lambda) = e^{\lambda[\alpha_0(s_1-1)+\alpha_1(s_2-1)+\alpha_2(s_1 s_2-1)]}. \tag{2.1.19}$$

We say $\mathbf{Y} = (Y_1, Y_2)$ follows a bivariate Negative Binomial distribution if it has the following joint probability mass function and the unconditional joint probability generating function

$$P(y_1, y_2) = \sum_{k=0}^{y_1} \sum_{m=0}^{y_2} \binom{\alpha_0 + y_1 + y_2 - k - m - 1}{\alpha_0 + y_2 - m - 1} \tag{2.1.20}$$

$$\times \binom{\alpha_0 + y_2 - m - 1}{\alpha_0 - 1} \binom{\alpha_1 + k - 1}{\alpha_1 - 1} \binom{\alpha_2 + m - 1}{\alpha_2 - 1} \tag{2.1.21}$$

$$\times \frac{\beta_1^{y_1} \beta_2^{y_2} (\beta_1 + \beta_2 + 1)^{k+m-y_1-y_2-\alpha_0}}{(\beta_1 + 1)^{k+\alpha_1} (\beta_2 + 1)^{m+\alpha_2}} \times \mathbf{1}_{(y_1, y_2) \in \mathbb{N}^2} \tag{2.1.22}$$

$$\psi(s_1, s_2) = \left[1 + \frac{\alpha_0}{\beta} + \frac{\alpha_1}{\beta} + \frac{\alpha_2}{\beta} - \frac{\alpha_0}{\beta} s_1 - \frac{\alpha_1}{\beta} s_2 - \frac{\alpha_2}{\beta} s_1 s_2\right]^{-\nu} \tag{2.1.23}$$

where $\beta_2 = \delta * \beta_1$, $\alpha_i > 0$, $\beta_j > 0$ for $i = 0, 1, 2$ and $j = 1, 2$ and $(s_1, s_2) \in \mathbb{R}^2$. The marginal distributions of $\mathbf{Y}$, $P(Y_1 = y_1)$ and $P(Y_2 = y_2)$, are Negative Binomial distributions with $Y_i \sim \text{NB}(\alpha_0 + \alpha_i, \frac{1}{\beta_i + 1})$ for $i = 1, 2$. The expected value and variance of Y_i are $\text{E}(Y_i) = (\alpha_0 + \alpha_i)\beta_i$ and $\text{Var}(Y_i) = (\alpha_0 + \alpha_i)\beta_i(\beta_i + 1)$. The covariance and correlation of Y_1 and Y_2 are given by

$$\text{Cov}(Y_1, Y_2) = \alpha_0 \beta_1 \beta_2 \tag{2.1.24}$$

$$\text{Corr}(Y_1, Y_2) = \frac{\alpha_0}{\sqrt{(\alpha_0 + \alpha_1)(\alpha_0 + \alpha_2)}} \sqrt{\frac{\beta_1 \beta_2}{(\beta_1 + 1)(\beta_2 + 1)}}. \tag{2.1.25}$$

Below we visualize the PMF with parameters $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$.

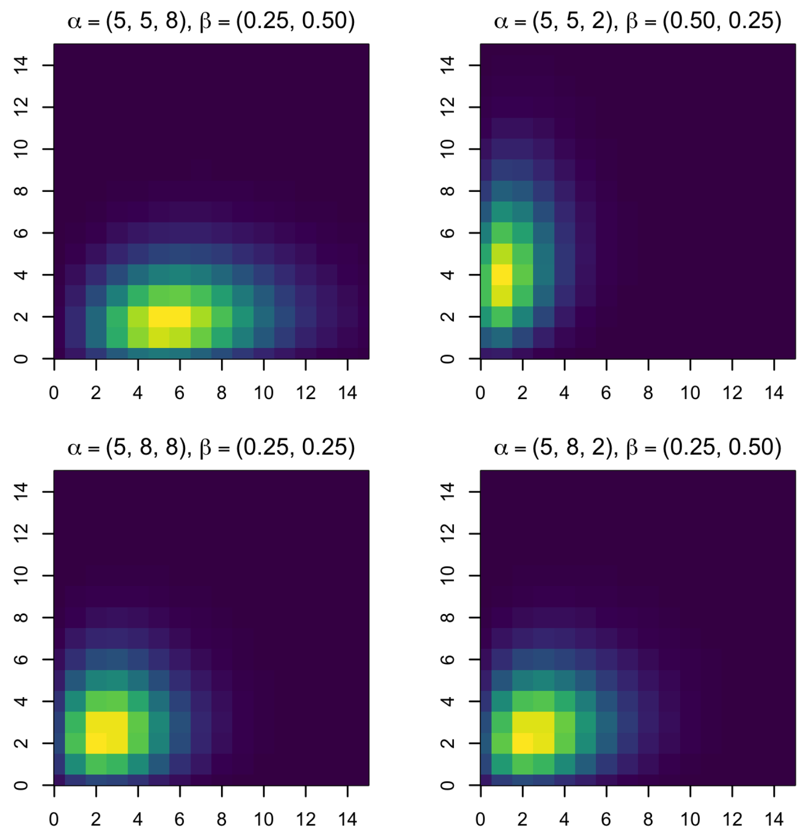

Figure 2.2: Joint PMF of Bivariate Negative Binomial distribution with parameters $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ evaluated on $\{0, 15\}^2$

2.1.4 Bivariate Type I Pólya-Aeppli

Recall that Y follows a Pólya-Aeppli distribution if it has the following PMF and PGF

$$P(Y = m) = \begin{cases} e^{-\lambda} & \text{if } m = 0 \\ e^{-\lambda} \sum_{k=1}^{m} \binom{m-1}{k-1} \frac{[\lambda(1-\rho)]^k}{k!} \rho^{m-k} & \text{if } m \in \{1, 2, \ldots\} \end{cases} \tag{2.1.26}$$

$$\psi(s) = e^{-\lambda(1 - \frac{(1-\rho)s}{1-\rho s})} \tag{2.1.27}$$

where $\lambda > 0$, $\rho \in (0, 1)$, and $s \in \mathbb{R}$. The expected value and variance of the Pólya-Aeppli distribution are $\mathrm{E}(Y) = \frac{\lambda}{1-\rho}$ and $\mathrm{Var}(Y) = \frac{\lambda(1+\rho)}{(1-\rho)^2}$. We note that the Pólya-Aeppli distribution belongs to a larger class of compound distributions. In this case, the distribution is found by summing Y independent and identically distributed Geometric random variables, where the number of random variables to sum has a Poisson distribution.

The bivariate Pólya–Aeppli distribution arises by considering a compound bivariate Poisson distribution with geometric compounding distribution as shown in Minkova and Balakrishnan (2014b). Let $\mathbf{Y} = (Y_1, Y_2)$ follow a bivariate Poisson distribution with parameters λ_1, λ_2 and λ_3. We say $\mathbf{Y} = (Y_1, Y_2)$ follows a bivariate Pólya–Aeppli distribution if it has the following joint probability generating function

$$\psi(s_1, s_2) = e^{-(\lambda_1+\lambda_2+\lambda_3)+(\lambda_1+\psi_1(s_1)+\lambda_2\psi_1(s_2)+\lambda_3\psi_1(s_1)\psi_1(s_2))} \tag{2.1.28}$$

where $\lambda_i > 0$, $(s_1, s_2) \in \mathbb{R}^2$ and $\psi_1(s) = \frac{(1-\rho)s}{1-\rho s}$ is the PGF of a Geometric random variable. The marginal distributions of $\mathbf{Y}$, $P(Y_1 = y_1)$ and $P(Y_2 = y_2)$, follow Pólya-Aeppli distributions with parameters $\{\lambda_1+\lambda_3, \rho\}$ and $\{\lambda_2+\lambda_3, \rho\}$, respectively. The expected value and variance of Y_i are $\mathrm{E}(Y_i) = \frac{(\lambda_i+\lambda_3)}{1-\rho}$ and $\mathrm{Var}(Y_i) = \frac{(\lambda_i+\lambda_3)(1+\rho)}{(1-\rho)^2}$ for $i = 1, 2$. We have $\mathrm{E}(Y_1Y_2) = \frac{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)+\lambda_3}{(1-\rho)^2}$. Thus, the covariance and correlation of Y_1 and Y_2 are given by

$$\mathrm{Cov}(Y_1, Y_2) = \frac{\lambda_3}{(1-\rho)^2} \tag{2.1.29}$$

$$\mathrm{Corr}(Y_1, Y_2) = \frac{\lambda_3}{(1+\rho)\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}. \tag{2.1.30}$$

Below we visualize the PMF with parameters $\lambda_1 = 1$, $\lambda_2 \in \{1, 3\}$, $\lambda_3 \in \{1, 3\}$ and $\rho \in \{0.25, 0.50\}$.

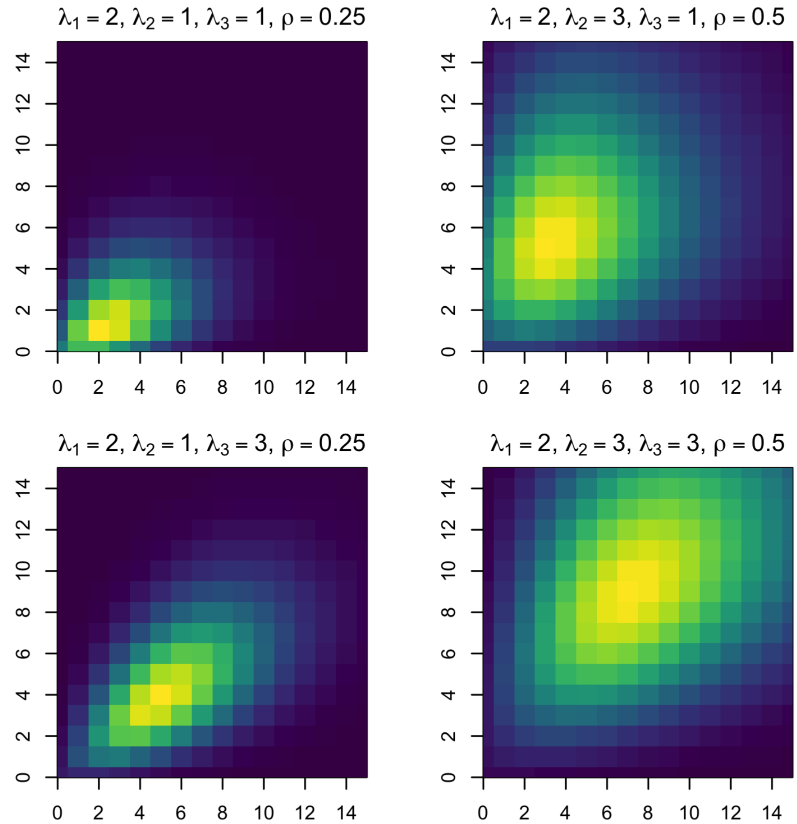

Figure 2.3: Joint PMF of Type I Bivariate Pólya-Aeppli distribution with parameters $(\lambda_1, \lambda_2, \lambda_3, \rho)$ evaluated on $\{0, 15\}^2$

2.2 Multivariate Discrete Distributions

2.2.1 Multivariate Poisson Distribution

Let $Y_1, \ldots, Y_p \sim \text{Pois}(\lambda_i)$. Then we may construct the following multivariate extensions of the Poisson distribution. Let $\text{cov}(\mathbf{Y})$ be the p-dimensional covariance matrix of $\mathbf{Y} = (Y_1, ..., Y_p)$. If $\text{cov}(\mathbf{Y}) = \mathbf{I}_p$, then the joint PMF is given by

$$P(y_1, \ldots, y_p) = \prod_{i=1}^{p} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}. \tag{2.2.1}$$

The second multivariate extension follows from the trivariate reduction method, which was used to construct the bivariate Poisson distribution. Let $Z_i = Y_i + Y$ where $Y \sim \text{Pois}(a)$. Then $\mathbf{Z} = (Z_1, \ldots, Z_p)$ follows the classical p-variate

Poisson distribution, denoted $\text{MVPois}(\lambda_1, \ldots, \lambda_p)$, with mean vector $\mathrm{E}(\mathbf{Z}) = (\lambda_1 + a, \ldots, \lambda_p + a)$ and covariance matrix

$$(\mathrm{cov}(\mathbf{Z}))_{ij} = \begin{cases} \lambda_i + a & \text{if } i = j \\ a & \text{if } i \neq j. \end{cases} \tag{2.2.2}$$

Clearly, each Z_i marginally follows a Poisson distribution with mean and variance equal to $\lambda_i + a$.

2.3 Extensions of the Fisher Dispersion Index

In this section, we review some proposed multivariate extensions of the Fisher dispersion index, then apply them to the bivariate discrete distributions introduced in Section 2.1.

2.3.1 FI_2

Introduced by Minkova and Balakrishnan (2014b), they proposed the following bivariate dispersion index

$$\mathrm{FI}_2(\mathbf{Y}) = \left[\mathrm{DI}_1 + \mathrm{DI}_2 - 2R\frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathrm{E}(Y_1)}\sqrt{\mathrm{E}(Y_2)}}\right]\frac{1}{(1 - R^2)} \tag{2.3.1}$$

where $R = \mathrm{Corr}(Y_1, Y_2)$ and $\mathrm{DI}_i = \mathrm{Var}(Y_i)/\,\mathrm{E}(Y_i)$. The definition of this measure is motivated by the observation that the univariate Poisson distribution is equi-dispersed, and thus the bivariate Poisson distribution should possess a similar property (shown in section 4.3).

Using this index, a bivariate count distribution is over-/equi-/under-dispersed if $\mathrm{FI}_2 > 2$, $\mathrm{FI}_2 = 2$ and $\mathrm{FI}_2 < 2$, respectively. The quantity $\mathrm{FI}_2(\mathbf{Y})/2$ is a measure of dispersion relative to the uncorrelated bivariate Poisson distribution. This connection is made clearer in later sections, and is extended to the p-variate case.

2.3.2 GDI and MDI

Introduced by Kokonendji and Puig (2018), they proposed the following generalized dispersion index and multiple marginal dispersion index defined by

$$\mathrm{GDI}(\mathbf{Y}) = \frac{\sqrt{\mathrm{E}(\mathbf{Y})}^T \operatorname{cov}(\mathbf{Y}) \sqrt{\mathrm{E}(\mathbf{Y})}}{\mathrm{E}(\mathbf{Y})^T \,\mathrm{E}(\mathbf{Y})} \tag{2.3.2}$$

$$\mathrm{MDI}(\mathbf{Y}) = \frac{\sqrt{\mathrm{E}(\mathbf{Y})}^T \operatorname{diag}(\operatorname{cov}(\mathbf{Y})) \sqrt{\mathrm{E}(\mathbf{Y})}}{\mathrm{E}(\mathbf{Y})^T \,\mathrm{E}(\mathbf{Y})} \tag{2.3.3}$$

where $\mathbf{Y} \in \mathbb{N}^p, p \geq 1$. Using GDI, a p-variate count distribution is over-/equi-/under-dispersed if $\mathrm{GDI}(\mathbf{Y}) > 1$, $\mathrm{GDI}(\mathbf{Y}) = 1$ and $\mathrm{GDI}(\mathbf{Y}) < 1$, respectively (same interpretation for $\mathrm{MDI}(\mathbf{Y})$).

Compared to FI_2, GDI and MDI do not require the inverse covariance matrix, which can be singular in some cases. Further, under some assumptions on the moments, Kokonendji and Puig (2018) provide explicit formulas for the asymptotic covariance matrices via the multivariate delta method. To verify the formulas, the authors conducted a simulation study which empirically verified the asymptotic results for large n.

In the next chapter, we introduce maximum likelihood estimation (MLE), the delta method and the bootstrap. Further, MLEs for the bivariate discrete distributions used in the simulation study will be discussed.

Chapter 3

Methodology

In this section we introduce maximum likelihood estimation (MLE) as a method for fitting model parameters from data. MLE has a long and storied history which can be traced to several famous scientists such as Lagrange, Bernoulli, Laplace and Gauss. The central role MLE plays in modern statistical inference is largely due to Ronald Fisher, who popularized it during the 1920s (Pfanzagl (1994)). Important properties of MLEs will be discussed, in particular the invariance property and asymptotic normality of MLEs. Using these properties, we will introduce the delta method as a way to construct approximate confidence intervals for FI_p, GDI and MDI. Lastly, the bootstrap will be introduced as an alternative to the delta method for constructing confidence intervals for the indices of dispersion mentioned previously.

3.1 Maximum Likelihood Estimation

Assume $y_1, \ldots, y_n$ are independent and identically distributed random variables (i.i.d) with parameters $\boldsymbol{\theta}$ and probability density function $f(\mathbf{y}|\boldsymbol{\theta})$. The maximum likelihood estimate of θ is defined to be the value $\widehat{\boldsymbol{\theta}}$ which maximizes the likelihood function $L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$. Intuitively, the MLE is the set of parameters that make the observed data most likely to have occurred under the assumed probability model. Further, since the product of density functions is always positive, the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \log(f(y_i|\boldsymbol{\theta}))$ is often used instead of the likelihood function as it is monotonic and easier to optimize.

There are several asymptotic properties of MLEs that are useful for constructing confidence intervals and performing hypothesis tests. MLEs are asymptotically normally distributed and converge to the true population parameter as the sample sizes goes to infinity. Often we are interested in a function of the MLEs. In this case, the invariance property of MLEs states that if $\widehat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}$ and $f(\boldsymbol{\theta})$ is any function of $\boldsymbol{\theta}$, then $f(\widehat{\boldsymbol{\theta}})$ is the MLE for $f(\boldsymbol{\theta})$ (Casella and Berger (2002)). Since the MLE is asymptotically normal, for a continuously differentiable function f of the MLEs $\widehat{\boldsymbol{\theta}}$, a first-order Taylor series expansion can be taken to find an estimate of the asymptotic variance of f about the MLE. This procedure is referred to as the delta method in statistics. Here is a formal definition from Casella and Berger (2002).

Theorem 3.1.1. *(Multivariate Delta Method) Let* $\mathbf{Y}_1, .., \mathbf{Y}_n$ *be a random sample with* $\mathrm{E}(Y_{ij}) = \mu_{ij}$ *and* $\mathrm{Cov}(Y_{ik}, Y_{jk}) = \sigma_{ij}$*. For a given function* f *with continuous partial derivatives and a specific value* $\widehat{\boldsymbol{\theta}} = (\theta_1, \ldots, \theta_p)$ *for which* $\tau^2 = \sum\sum \sigma_{ij} \frac{\partial f(\widehat{\boldsymbol{\theta}})}{\partial \theta_i} \cdot \frac{\partial f(\widehat{\boldsymbol{\theta}})}{\partial \theta_j} > 0$,

$$\sqrt{n}\left[f(\hat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta})\right] \overset{d}{\sim} N(0, \tau^2) \tag{3.1.1}$$

Using the delta method, we can construct confidence intervals based on the normal approximation for FI_2, GDI and MDI. Further, we can use the observed Fisher information matrix defined by

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\theta}}) \approx -\left(\frac{\partial^2 \log\left[L(\boldsymbol{\theta}|\mathbf{y})\right]}{\partial \boldsymbol{\theta}^2}\right)^{-1} \tag{3.1.2}$$

to calculate the asymptotic covariance matrix of the MLEs, evaluated at $\widehat{\boldsymbol{\theta}}$. The estimated covariance matrix is used in the delta method to obtain the value of τ^2 for a given dataset.

3.1.1 Bootstrap Method

The bootstrap method is a general purpose tool based on resampling. Given observed data, the bootstrap method proceeds by producing many bootstrap samples of the original data with replacement. Using these samples, various quantities of interest such as bias, variance, confidence intervals can be constructed. Let B denote the number of bootstrap samples, and $\boldsymbol{\theta}$ for the parameters of interest. Suppose we are given a sample $y_1, ..., y_n$ from an absolutely continuous distribution with probability density function $f(y; \boldsymbol{\theta})$. In the context of this thesis, we first find the MLEs for the original dataset. Then, we simulate B samples of size m from $f(y; \widehat{\boldsymbol{\theta}})$ where $\widehat{\boldsymbol{\theta}}$ is the MLE for the original data. For each bootstrap sample, we calculate the MLE for the sample, and thus end up with a sequence of MLEs $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_B$ for which various quantities can be calculated such as the bias and variance. In this thesis, we are interested in functions of the MLEs, so we the sequence will be $f(\widehat{\boldsymbol{\theta}})_1, ..., f(\widehat{\boldsymbol{\theta}})_B$. In practice, we assume the number of bootstrap samples B are sufficiently large for the bootstrap to yield accurate results.

There are various ways to form confidence intervals using bootstrap estimates. The simplest, which we consider here, is the percentile based confidence interval. To form a percentile confidence interval, we sort the bootstrap estimates in increasing order $f(\widehat{\boldsymbol{\theta}})_{1'}, ..., f(\widehat{\boldsymbol{\theta}})_{B'}$. Then, using the $l = (B \times \alpha)^{th}$ and $u = (B \times (1-\alpha))^{th}$ values, a $100(1-\alpha)\%$ percentile confidence interval is given by

$$\left[f(\widehat{\boldsymbol{\theta}})_l,\, f(\widehat{\boldsymbol{\theta}})_u\right].$$

For example, when $B = 10,000$ and $\alpha = 0.05$, then the $l = 250^{th}$ and $u = 9,750^{th}$ sorted values are the upper and lower bounds for the percentile confidence interval (Davison and Hinkley (1997)).

3.1.2 Bivariate Bernoulli

For the purposes of this thesis, the bivariate Bernoulli maximum likelihood estimates are omitted. Our simulation study only covers distributions with support in $\mathbb{N}^2$, whereas the bivariate Bernoulli distribution has support in the lattice $\{0,1\}^2$.

3.1.3 Bivariate Poisson

Recall from subsection 2.1.2 the probability mass function given by

$$P(y_1, y_2) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{k=0}^{\min(y_1,y_2)} \frac{\lambda_1^{(y_1-k)} \lambda_2^{(y_2-k)} \lambda_3^k}{(y_1-k)!\,(y_2-k)!\,k!},\; (y_1, y_2) \in \mathbb{N}^2. \quad (3.1.3)$$

Holgate (1964) first obtained the maximum likelihood estimates of $P(y_1, y_2)$ using the slightly modified version

$$P(y_1, y_2) = e^{-(\lambda_1^*+\lambda_2^*-\lambda_3)} \sum_{k=0}^{\min(y_1,y_2)} \frac{(\lambda_1^*-\lambda_3)^{(y_1-k)} (\lambda_2^*-\lambda_3)^{(y_2-k)} \lambda_3^k}{(y_1-k)!\,(y_2-k)!\,k!} \quad (3.1.4)$$

where $\lambda_i^* = \lambda_i + \lambda_3$ for $i \in \{1,2\}$. Using this form of the probability mass function, the recurrence relations from Teicher (1954) become

$$y_1 P(y_1, y_2) = (\lambda_1^* - \lambda_3) P(y_1 - 1, y_2) + \lambda_3 P(y_1 - 1, y_2 - 1) \quad (3.1.5)$$

$$y_2 P(y_1, y_2) = (\lambda_2^* - \lambda_3) P(y_1, y_2 - 1) + \lambda_3 P(y_1 - 1, y_2 - 1). \quad (3.1.6)$$

Upon differentiating (3.1.4) with respect to each parameter λ_i, we have

$$\frac{\partial P(y_1, y_2)}{\partial \lambda_1^*} = P(y_1 - 1, y_2) - P(y_1, y_2) \quad (3.1.7)$$

$$\frac{\partial P(y_1, y_2)}{\partial \lambda_2^*} = P(y_1, y_2 - 1) - P(y_1, y_2) \tag{3.1.8}$$

$$\begin{aligned}\frac{\partial P(y_1, y_2)}{\partial \lambda_3} &= P(y_1, y_2) - P(y_1, y_2 - 1) \\ &- P(y_1 - 1, y_2) + P(y_1 - 1, y_2 - 1)\end{aligned} \tag{3.1.9}$$

Using the recurrence relations and partial derivatives, the following likelihood equations were obtained by Holgate (1964) by differentiating the log-likelihood function of (3.1.4)

$$\frac{\bar{y}_1}{\widehat{\lambda_1^*} - \widehat{\lambda_3}} - \frac{\widehat{\lambda_3}\bar{R}}{\widehat{\lambda_1^*} - \widehat{\lambda_3}} - 1 = 0 \tag{3.1.10}$$

$$\frac{\bar{y}_2}{\widehat{\lambda_2^*} - \widehat{\lambda_3}} - \frac{\widehat{\lambda_3}\bar{R}}{\widehat{\lambda_2^*} - \widehat{\lambda_3}} - 1 = 0. \tag{3.1.11}$$

Adding the equations together, we obtain

$$\frac{\bar{y}_1}{\widehat{\lambda_1^*} - \widehat{\lambda_3}} + \frac{\bar{y}_2}{\widehat{\lambda_2^*} - \widehat{\lambda_3}} - \left[1 + \frac{\widehat{\lambda_3}}{\widehat{\lambda_1^*} - \widehat{\lambda_3}} + \frac{\widehat{\lambda_3}}{\widehat{\lambda_2^*} - \widehat{\lambda_3}}\right]\bar{R} - 1 = 0, \tag{3.1.12}$$

where

$$R(y_1, y_2) = \frac{P(y_1 - 1, y_2 - 1)}{P(y_1, y_2)} \tag{3.1.13}$$

$$\bar{R} = \frac{1}{n}\sum_{y_1, y_2} R(y_1, y_2), \tag{3.1.14}$$

Solving (3.1.10) and (3.1.11), we obtain

$$\bar{y}_1 = \widehat{\lambda_3}(\bar{R} - 1) + \widehat{\lambda_1^*} \tag{3.1.15}$$

$$\bar{y}_2 = \widehat{\lambda_3}(\bar{R} - 1) + \widehat{\lambda_2^*} \tag{3.1.16}$$

where the maximum likelihood estimates are $\widehat{\lambda_1^*} = \bar{y}_1$ and $\widehat{\lambda_2^*} = \bar{y}_2$ given that $\bar{R} = 1$. The maximum likelihood estimate for $\widehat{\lambda_3}$ can be found by solving $\bar{R} = 1$ using an iterative method (since $\bar{R}$ depends on λ_1 and λ_2 through $P(y_1, y_2)$).

3.1.4 Bivariate Negative Binomial

Following the approach of Cho et al. (2023), we obtain the maximum likehood estimates via the expectation-maximization (EM) algorithm using the `bzinb` package available in `R` (Cho et al. (2022)). The EM algorithm is a general purpose approach to iteratively computing MLEs, especially in cases where there is missing data (McLachlan and Krishnan (2008)). There are two key steps in the EM algorithm, the expectation step and the maximization step, which are alternated repeatedly until a desired error tolerance is reached.

Following Cho et al. (2023), the complete data density is given by

$$f(Y_1, Y_2, Y_1, Y_2, R_0, R_1, R_2, E_1, E_2, E_3, E_4) = \frac{(R_0+R_1)^{Y_1}(R_0+R_2)^{Y_2} R_0^{\alpha_0-1} R_1^{\alpha_1-1} R_2^{\alpha_2-1} \beta_2^{Y_2} \prod_{k=1}^{4} \pi_k^{E_k}}{Y_1!\, Y_2! \Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2) \exp\{R_0 \frac{(1+\beta_1+\beta_2)}{\beta_1} + R_1 \frac{(1+\beta_1)}{\beta_1} + R_2 \frac{(1+\beta_2)}{\beta_1}\} \beta_1^{Y_2+\alpha_0+\alpha_1+\alpha_2}} \times 1_{\sum_{k=1}^{4} E_k = 1}\,. \quad (3.1.17)$$

The full individual log-likelihood for the i^{th} entry (i.e. observation) is given by

$$\begin{aligned} \ell_i^{\text{Full}} &= Y_{1,i} \log(R_{0,i} + R_{1,i}) + Y_{2,i} \log(R_{0,i} + R_{2,i}) + (\alpha_0 - 1)\log(R_{0,i}) \\ &+ (\alpha_1 - 1)\log(R_{1,i}) + (\alpha_2 - 1)\log(R_{2,i}) + Y_{2,i}\log(\beta_2) \\ &- (Y_{2,i} + \alpha_0 + \alpha_1 + \alpha_2)\log(\beta_1) + \sum_{k=1}^{4} E_{k,i}\log(\pi_k) - \log(Y_{1,i}!) - \log(Y_{2,i}!) \\ &- \log(\Gamma(\alpha_0)) - \log(\Gamma(\alpha_1)) - \log(\Gamma(\alpha_2)) \\ &- R_{0,i}\frac{(1+\beta_1+\beta_2)}{\beta_1} - R_{1,i}\frac{(1+\beta_1)}{\beta_1} - R_{2,i}\frac{(1+\beta_2)}{\beta_1} \\ &+ \log(1_{(Y_{1,i}=Y_{1,i}(E_{1,i}+E_{2,i}))}) + \log(1_{(Y_{2,i}=Y_{2,i}(E_{1,i}+E_{3,i}))}) \\ &\quad + \log(1_{(Y_{1,i}=Y_{1,i}(E_{1,i}+E_{2,i}))}) + \log(1_{\sum_{k=1}^{4} E_k=1}). \end{aligned} \quad (3.1.18)$$

The conditional expected log-likelihood is linear given $(X_{1,i}, X_{2,i}\,; \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi})$ contains the current MLEs. The authors propose solving the following score equations iteratively to find the MLEs

$$\partial_{\alpha_j} E\left[\ell_i^{\text{Full}} \middle| \cdot\right] = E\left[\log(R_{j,i})|\cdot\right] - \log(\beta_1) - \psi(\alpha_j) \quad (3.1.19)$$

$$\partial_{\beta_1} E\left[\ell_i^{\text{Full}} \middle| \cdot\right] = \frac{E\left[R_{0,i} + R_{2,i}|\cdot\right](1+\beta_2) + E\left[R_{1,i}|\cdot\right]}{\beta_1^2}$$

$$- \frac{\alpha_0 + \alpha_1 + \alpha_2 + E\left[Y_{2,i}|\cdot\right]}{\beta_1} \tag{3.1.20}$$

$$\partial_{\beta_2} E\left[\ell_i^{\text{Full}}\middle|\cdot\right] = \frac{E\left[Y_{2,i}|\cdot\right]}{\beta_2} - \frac{E\left[R_{0,i} + R_{2,i}|\cdot\right]}{\beta_1} \tag{3.1.21}$$

$$\partial_{\pi_j} E\left[\ell_i^{\text{Full}}\middle|\cdot\right] = \frac{E\left[E_{j,i}|\cdot\right]}{\pi_j} - \frac{1 - E\left[E_{j,i}|\cdot\right]}{1-\pi_j} \tag{3.1.22}$$

for $\alpha_j = \{\alpha_0, \alpha_1, \alpha_2\}$ and $\pi_j = \{\pi_1, \pi_2, \pi_3\}$, where $(\cdot)$ represents conditioning on $(\mathbf{X}_1, \mathbf{X}_2\,; \boldsymbol{\theta})$. The final MLE are determined based on a given maximum number of iterations and error tolerance.

3.1.5 Bivariate Type I Polyá-Aeppli

The maximum likelihood estimates for $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3, \rho)$ can be found using an iterative procedure as proposed by Balakrishnan et al. (2017). Due to the recursive nature of the probability mass function (PMF) and likelihood equations, the Newton-Raphson (N-R) algorithm was proposed to solve the following non-linear system of equations

$$\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}^{(n-1)} - \mathbf{J}^{-1}(\boldsymbol{\theta}^{(n-1)}) \cdot \mathbf{F}(\boldsymbol{\theta}^{(n-1)}), \tag{3.1.23}$$

where $\mathbf{J}^{-1}$ is the Jacobian matrix and $\mathbf{F}$ is the gradient of the likelihood function. The algorithm stops once a desired error-threshold has been reached between the nth and $(n+1)$st step or a maximum number of iterations has been reached. In the latter case, convergence of the MLEs is said to have failed and a grid search can be used instead as proposed by the authors. Initial values, $\boldsymbol{\theta}^{(0)}$, for the Newton-Raphson algorithm can be found in some cases using the following method of moments (MoM) estimates

$$\widehat{\rho} = \frac{s_1^2 + s_2^2 - (\bar{n}_1 + \bar{n}_2)}{s_1^2 + s_2^2 + \bar{n}_1 + \bar{n}_2} \tag{3.1.24}$$

$$\widehat{\lambda_1} = (1-\widehat{\rho})\widehat{n}_1 - (1-\widehat{\rho})^2 s_{12} \tag{3.1.25}$$

$$\widehat{\lambda_2} = (1-\widehat{\rho})\widehat{n}_2 - (1-\widehat{\rho})^2 s_{12} \tag{3.1.26}$$

$$\widehat{\lambda_3} = (1-\widehat{\rho})^2 s_{12}. \tag{3.1.27}$$

When the MoM estimates produce an inadmissible initial value, the authors suggest using a parameter grid search to find a suitable starting value for the NR algorithm (refer to Balakrishnan et al. (2017) for more details).

Chapter 4

A Multivariate Index of Dispersion

4.1 Preliminaries

Before introducing our proposed extension of the Fisher dispersion index, we first review some matrix theory that will be useful for our purposes. For further details, please refer to Styan (1973) and Horn and Johnson (2012).

Definition 4.1.1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $m \times p$ *real-valued matrices. The Hadamard product is defined as* $(\mathbf{A} \otimes \mathbf{B})_{ij} = a_{ij}b_{ij}$ *for* $1 \leq i \leq m$*,* $1 \leq j \leq p$*.*

Compared to the usual matrix product, the Hadamard product requires both the row and column dimensions be equal. Further, the Hadamard product is commutative, whereas the regular matrix product is not. This result follows directly from the commutativity of scalar multiplication.

$$(\mathbf{A} \otimes \mathbf{B})_{ij} = a_{ij}b_{ij} = b_{ij}a_{ij} = (\mathbf{B} \otimes \mathbf{A})_{ij} \tag{4.1.1}$$

The role of the identity matrix for Hadamard products is played by the matrix $\mathbf{J}$. This matrix has elements $(\mathbf{J})_{ij} = 1$ for all $1 \leq i \leq m$, $1 \leq j \leq p$.

$$(\mathbf{A} \otimes \mathbf{J})_{ij} = a_{ij} \times 1 = a_{ij} = (\mathbf{A})_{ij} \tag{4.1.2}$$

We will assume in the proceeding discussion that matrices $\mathbf{A}$ and $\mathbf{B}$ are square symmetric matrices of dimension p.

4.1.1 Important Results

The Hadamard product has many applications in statistics due to the following theorem proved in 1911 by the mathematician Issai Schur (Schur (1911)).

Theorem 4.1.1. *(Shur Product Theorem) When* $\mathbf{A}$ *and* $\mathbf{B}$ *be are positive semidefinite, then so is* $\mathbf{A} \otimes \mathbf{B}$*. When* $\mathbf{A}$ *and* $\mathbf{B}$ *are positive definite, then so is* $\mathbf{A} \otimes \mathbf{B}$*.*

Proof. (from Styan (1973)) Suppose $\mathbf{A}$ and $\mathbf{B}$ be are positive semidefinite matrices. Let $\mathbf{x} \neq \mathbf{0}$ be $p \times 1$ and consider the quadratic form

$$\mathbf{x}^T (\mathbf{A} \otimes \mathbf{B})\, \mathbf{x}. \tag{4.1.3}$$

Since $\mathbf{B}$ is positive semidefinite, there exists a $p \times p$ matrix $\mathbf{V}$ such that $\mathbf{B} = \mathbf{V}\mathbf{V}^T$, where $\mathbf{V}\mathbf{V}^T$ is symmetric. Plugging this into (4.1.3), we find

$$\mathbf{x}^T (\mathbf{A} \otimes (\mathbf{V}\mathbf{V}^T))\, \mathbf{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} x_i a_{ij} \left(\sum_{k=1}^{p} v_{ik} v_{jk} \right) x_j \tag{4.1.4}$$

$$= \sum_{k=1}^{p} \left(\sum_{i=1}^{p}\sum_{j=1}^{p} x_i a_{ij} v_{ik} v_{jk} x_j \right) \tag{4.1.5}$$

$$= \sum_{k=1}^{p} \left(\sum_{i=1}^{p}\sum_{j=1}^{p} x_i v_{ik} a_{ij} x_j v_{jk} \right) \tag{4.1.6}$$

$$= \sum_{k=1}^{p} (\mathbf{x} \otimes \mathbf{v}_k)^T \mathbf{A}\, (\mathbf{x} \otimes \mathbf{v}_k) \geq 0\,, \tag{4.1.7}$$

where $\mathbf{v}_k$ denotes the k^{th} column of $\mathbf{V}$. Since $\mathbf{A}$ is positive semidefinite, (4.1.7) is the sum of k non-negative quadratic forms, and is always greater than or equal to zero. When $\mathbf{A}$ and $\mathbf{B}$ are non-singular, $\mathbf{V}\mathbf{V}^T$ is non-singular and (4.1.7) becomes strictly positive. □

Positive semidefinite matrices arise naturally in statistics through the covariance matrix. Since the eigenvalues of the covariance matrix are always greater than or equal to zero, we know that non-singular covariance matrices and their inverses are always positive definite. This leads us to the following useful corollary.

Corollary 4.1.2. *Suppose* $\text{cov}(\mathbf{Y})$ *is a non-singular covariance matrix of size* p *with inverse matrix given by* $\text{cov}(\mathbf{Y})^{-1}$*. Then* $(\text{cov}(\mathbf{Y}) \otimes \text{cov}(\mathbf{Y})^{-1})$ *is a positive definite matrix of size* p*.*

Proof. Follows immediately from Theorem 4.1.1 □

Proved by Oppenheim (1930), we have an interesting lower bound on the determinant of $\mathbf{A} \otimes \mathbf{B}$.

Theorem 4.1.3. *When* $\mathbf{A}$ *and* $\mathbf{B}$ *are positive semidefinite matrices*

$$\det(\mathbf{A} \otimes \mathbf{B}) \geq \det(\mathbf{A}) \det(\mathbf{B}) \tag{4.1.8}$$

Proof. See Theorem 3.7 in Styan (1973) □

A nice application of Theorem 4.1.3 is that $\det(\text{cov}(\mathbf{Y}) \otimes \text{cov}(\mathbf{Y})^{-1}) \geq 1$ for every non-singular covariance matrix $\mathbf{Y}$.

With these results in hand, we introduce our proposed multivariate extension of the Fisher dispersion index.

4.2 FI_p - A New Multivariate Dispersion Index

We start this section by introducing a new multivariate index of dispersion denoted FI_p.

Definition 4.2.1. *Let* $\mathbf{Y}$ *be a p-variate random vector defined on* $\mathbb{N}^p$, $p \geq 1$. *Let* $\text{cov}(\mathbf{Y})$ *be the covariance matrix of* $\mathbf{Y}$ *and* $\sqrt{\mathbf{d}} = (\sqrt{\text{DI}_1}, \ldots, \sqrt{\text{DI}_p})$ *be the element-wise square-root vector of marginal dispersion indices. We define the following quadratic form*

$$\text{FI}_p(\mathbf{Y}) = \sqrt{\mathbf{d}}^T \left(\text{cov}(\mathbf{Y}) \otimes \text{cov}(\mathbf{Y})^{-1}\right) \sqrt{\mathbf{d}} \tag{4.2.1}$$

as a multivariate dispersion index.

As mentioned in subsection 2.3.1, the bivariate dispersion index FI_2 introduced by Minkova and Balakrishnan was motivated by the equi-dispersion property of the Poisson distribution. They found that when $\mathbf{Y}$ follows a bivariate Poisson distribution, $\text{FI}_2(\mathbf{Y}) = 2$. It turns out that this result can be generalized to the multivariate Poisson distribution as follows.

Lemma 4.2.1. *Let* $\mathbf{Y} \sim MVPois(\lambda_1, \ldots, \lambda_p)$ *with* $\text{E}(Y_i) = \text{Var}(Y_i) = \lambda_i + a$ *and* $\text{cov}(Y_i, Y_j) = a$. *Then* $\text{FI}_p(\mathbf{Y}) = p$.

Proof. See Appendix B.

This important result extends the equi-dispersion property of the bivariate Poisson distribution to the multivariate Poisson distribution using FI_p. The importance of this result is that it justifies normalizing FI_p by p to measure dispersion relative to the p-variate Poisson distribution. Next we show that FI_2 is a special case of FI_p when $p = 2$.

Lemma 4.2.2. *When* $p = 2$, FI_p *is* FI_2.

Proof. Let $\mathbf{Y} = (Y_1, Y_2)$ be a random vector in $\mathbb{N}^2$ and $\mathrm{cov}(\mathbf{Y})$ be the associated 2×2 covariance matrix.

$$\mathrm{cov}(\mathbf{Y}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

The determinant of $\mathrm{cov}(\mathbf{Y})$ is $\det(\mathrm{cov}(\mathbf{Y})) = \sigma_{11}\sigma_{22} - \sigma_{12}^2$ and the inverse is given by

$$\mathrm{cov}(\mathbf{Y})^{-1} = \frac{1}{\det(\mathrm{cov}(\mathbf{Y}))} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

Taking the Hadamard product of the two matrices we have

$$(\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}) = \frac{1}{\det(\mathrm{cov}(\mathbf{Y}))} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \tag{4.2.2}$$

$$= \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{11}\sigma_{22} & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}\sigma_{22} \end{bmatrix} \tag{4.2.3}$$

Letting $\mathrm{R} = \sigma_{12} / (\sqrt{\sigma_{11}}\sqrt{\sigma_{22}})$ denote the correlation coefficient of $\mathbf{Y}$, the above becomes

$$= \begin{bmatrix} 1 & -R^2 \\ -R^2 & 1 \end{bmatrix} \frac{1}{(1 - R^2)} \tag{4.2.4}$$

We now compute FI_p below where $p = 2$.

$$\mathrm{FI}_p(\mathbf{Y}) = \sqrt{\mathbf{d}}^T \, (\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}) \sqrt{\mathbf{d}} \tag{4.2.5}$$

$$= \begin{bmatrix} \sqrt{\mathrm{DI}_1} & \sqrt{\mathrm{DI}_2} \end{bmatrix} \cdot \begin{bmatrix} 1 & -R^2 \\ -R^2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\mathrm{DI}_1} \\ \sqrt{\mathrm{DI}_2} \end{bmatrix} \frac{1}{(1 - R^2)} \tag{4.2.6}$$

$$= \begin{bmatrix} \sqrt{\mathrm{DI}_1} & \sqrt{\mathrm{DI}_2} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\mathrm{DI}_1} - R^2\sqrt{\mathrm{DI}_2} \\ \sqrt{\mathrm{DI}_2} - R^2\sqrt{\mathrm{DI}_1} \end{bmatrix} \frac{1}{(1 - R^2)} \tag{4.2.7}$$

$$= \left[\mathrm{DI}_1 + \mathrm{DI}_2 - 2R^2\sqrt{\mathrm{DI}_1}\sqrt{\mathrm{DI}_2}\right] \frac{1}{(1 - R^2)} \tag{4.2.8}$$

$$= \left[\mathrm{DI}_1 + \mathrm{DI}_2 - 2R^2\sqrt{\frac{\sigma_{11}}{\mu_1}}\sqrt{\frac{\sigma_{22}}{\mu_2}}\right] \frac{1}{(1 - R^2)} \tag{4.2.9}$$

$$= \left[\mathrm{DI}_1 + \mathrm{DI}_2 - 2R\frac{\sigma_{12}}{\sqrt{\mu_1}\sqrt{\mu_2}}\right] \frac{1}{(1 - R^2)} \tag{4.2.10}$$

$$= \left[\mathrm{DI}_1 + \mathrm{DI}_2 - 2R \frac{\mathrm{cov}(Y_1, Y_2)}{\sqrt{\mathrm{E}(Y_1)}\sqrt{\mathrm{E}(Y_2)}} \right] \frac{1}{(1 - R^2)} \tag{4.2.11}$$

which is the same expression as FI_2 □

For certain classes of matrices, FI_p admits a simple form. When the covariance matrix is diagonal, we have the following result.

Corollary 4.2.3. *Let* $\mathrm{cov}(\mathbf{Y})$ *be a diagonal covariance matrix of size* p*. Then for* $\mathbf{Y} \in \mathbb{N}^p$*, the multivariate dispersion index equals*

$$\mathrm{FI}_p(\mathbf{Y}) = \frac{\sqrt{\mathbf{d}}^T \cdot \sqrt{\mathbf{d}}}{\det(\mathrm{cov}(\mathbf{Y}))}. \tag{4.2.12}$$

Proof. Let $\mathrm{cov}(\mathbf{Y})$ be the non-singular diagonal covariance matrix of $\mathbf{Y}$. Since $\mathrm{cov}(\mathbf{Y})$ is a diagonal matrix, we have that $(\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}) = \mathbf{I}_p$, which implies that the multivariate dispersion index is given by

$$\mathrm{FI}_p(\mathbf{Y}) = \frac{\sqrt{\mathbf{d}}^T (\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}) \sqrt{\mathbf{d}}}{\det(\mathrm{cov}(\mathbf{Y}))} \tag{4.2.13}$$

$$= \frac{\sqrt{\mathbf{d}}^T \cdot \sqrt{\mathbf{d}}}{\det(\mathrm{cov}(\mathbf{Y}))} \square \tag{4.2.14}$$

From the previous corollary, we have the following result which we state without proof.

Corollary 4.2.4. *Suppose* $\mathrm{cov}(\mathbf{Y}) = \mathbf{I}_p$ *is the identity matrix of size* p*. Then* $(\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}) = \mathbf{I}_p$ *and* $\mathrm{FI}_p(\mathbf{Y}) = \sum_{k=1}^{p} \mathrm{DI}_k$.

For joint distributions constructed via products of independent and identically distributed random variables, Corollary 4.2.4 says that FI_p will always equal the sum of the marginal dispersion indices. The importance of this result is that for uncorrelated random variables, FI_p will always equal the sum of the marginal dispersion indices. If each marginal dispersion index is over/equi/under-dispersed, then FI_p will be over/equi/under-dispersed.

Next we apply GDI, MDI and FI_p to some common bivariate and multivariate discrete distributions.

4.3 Illustrative Examples

In this section, we derive expressions for FI_p, GDI and MDI for the bivariate discrete distributions introduced in Section 2.1 and the p-variate Poisson distribution introduced in Section 2.2.

Example 4.3.1 (Bivariate Bernoulli)**.** Let $\mathbf{Y} \sim \text{BivBern}(p_{00}, p_{10}, p_{01}, p_{11})$. Then $\mu_i = \text{E}(Y_i) = p_{0i} + p_{1i}$, $\sigma_i = \text{Var}(Y_i) = \mu_i(1-\mu_i)$ and $a = \text{cov}(Y_1, Y_2) = p_{00}p_{11} - p_{01}p_{10}$ for $i = 1, 2$. First, we derive an expression for FI_2.

$$\text{FI}_2(\mathbf{Y}) = \sqrt{\mathbf{d}}^T \left(\text{cov}(\mathbf{Y}) \otimes \text{cov}(\mathbf{Y})^{-1}\right) \sqrt{\mathbf{d}} \tag{4.3.1}$$

$$= \frac{1}{(\mu_1\mu_2 - a^2)} \begin{bmatrix} \sqrt{1-\mu_1} & \sqrt{1-\mu_2} \end{bmatrix} \cdot \begin{bmatrix} \mu_1\mu_2 & -a^2 \\ -a^2 & \mu_1\mu_2 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{1-\mu_1} \\ \sqrt{1-\mu_2} \end{bmatrix} \tag{4.3.2}$$

$$= \frac{\mu_1\mu_2(1-\mu_1) + \mu_1\mu_2(1-\mu_2) - 2a^2\sqrt{1-\mu_1}\sqrt{1-\mu_2}}{\mu_1\mu_2 - a^2} \tag{4.3.3}$$

$$= \frac{\mu_1\mu_2(2-\mu_1-\mu_2) - 2a^2\sqrt{\mu_2}\sqrt{\mu_1}}{\mu_1\mu_2 - a^2} \tag{4.3.4}$$

$$= \frac{(p_{00}+p_{10})(p_{01}+p_{11})}{(p_{00}+p_{10})(p_{01}+p_{11}) - (p_{00}p_{11} - p_{01}p_{10})^2} \tag{4.3.5}$$

$$- \frac{2(p_{00}p_{11} - p_{01}p_{10})^2\sqrt{(p_{00}+p_{10})(p_{01}+p_{11})}}{(p_{00}+p_{10})(p_{01}+p_{11}) - (p_{00}p_{11} - p_{01}p_{10})^2}. \tag{4.3.6}$$

The bivariate Bernoulli is always under-dispersed relative to the bivariate Poisson using FI_2. For example, when $(p_{00}, p_{10}, p_{01}, p_{11}) = (0.25, 0.25, 0.25, 0.25)$, $\text{FI}_2 = 1 < 2$. Next, we derive an expression for GDI.

$$\text{GDI}(\mathbf{Y}) = \frac{\sqrt{\text{E}(\mathbf{Y})}^T \text{cov}(\mathbf{Y}) \sqrt{\text{E}(\mathbf{Y})}}{\text{E}(\mathbf{Y})^T \, \text{E}(\mathbf{Y})} \tag{4.3.7}$$

$$= \frac{\begin{bmatrix} \sqrt{\mu_1} & \sqrt{\mu_2} \end{bmatrix} \cdot \begin{bmatrix} \mu_1(1-\mu_1) & a \\ a & \mu_2(1-\mu_2) \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\mu_1} \\ \sqrt{\mu_2} \end{bmatrix}}{\begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \cdot \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}} \tag{4.3.8}$$

$$= \frac{\mu_1^2(1-\mu_1) + \mu_2^2(1-\mu_2) + 2a\sqrt{\mu_1\mu_2}}{\mu_1^2 + \mu_2^2} \tag{4.3.9}$$

$$= 1 + \frac{2a\sqrt{\mu_1\mu_2} - (\mu_1^3 + \mu_2^3)}{\mu_1^2 + \mu_2^2} \tag{4.3.10}$$

$$= 1 + \frac{2(p_{00}p_{11} - p_{01}p_{10})\sqrt{(p_{00}+p_{10})(p_{01}+p_{11})}}{(p_{00}+p_{10})^2 + (p_{01}+p_{11})^2} \tag{4.3.11}$$

$$- \frac{(p_{00}+p_{10})^3+(p_{01}+p_{11})^3}{(p_{00}+p_{10})^2+(p_{01}+p_{11})^2}. \tag{4.3.12}$$

Similar to FI_2, the bivariate Bernoulli is always under-dispersed relative to the bivariate Poisson using GDI. For example, when $(p_{00}, p_{10}, p_{01}, p_{11}) = (0.25, 0.25, 0.25, 0.25)$, GDI $= 1/2 < 1$. Lastly, we derive an expression for MDI, which is the same as GDI with $a = 0$.

$$\text{MDI}(\mathbf{Y}) = \frac{\sqrt{\text{E}(\mathbf{Y})}^T \operatorname{diag}(\text{cov}(\mathbf{Y}))\sqrt{\text{E}(\mathbf{Y})}}{\text{E}(\mathbf{Y})^T \, \text{E}(\mathbf{Y})} \tag{4.3.13}$$

$$= \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1(1-\mu_1) & 0 \\ 0 & \mu_2(1-\mu_2)\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.14}$$

$$= \frac{\mu_1^2(1-\mu_1)+\mu_2^2(1-\mu_2)}{\mu_1^2+\mu_2^2} \tag{4.3.15}$$

$$= 1 - \frac{\mu_1^3+\mu_2^3}{\mu_1^2+\mu_2^2} \tag{4.3.16}$$

$$= 1 - \frac{(p_{00}+p_{10})^3+(p_{01}+p_{11})^3}{(p_{00}+p_{10})^2+(p_{01}+p_{11})^2}. \tag{4.3.17}$$

Since $(\mu_1^3+\mu_2^3)/(\mu_1^2+\mu_2^2) > 0$, MDI has the same interpretation as GDI for the bivariate Bernoulli (i.e., it is always under-dispersed).

Example 4.3.2 (Bivariate Poisson)**.** Let $\mathbf{Y} \sim \text{BivPois}(\lambda_1, \lambda_2, \lambda_3)$. Then $\mu_i = \text{E}(Y_i) = \text{Var}(Y_i) = \lambda_i + \lambda_3$ and $a = \text{cov}(Y_1, Y_2) = \lambda_3$ for $i = 1, 2$. As in the previous section, we derive expressions for FI_2, GDI and MDI.

$$\text{FI}_2(\mathbf{Y}) = \frac{1}{(\mu_1\mu_2 - \lambda_3^2)} \begin{bmatrix}1 & 1\end{bmatrix} \cdot \begin{bmatrix}\mu_1\mu_2 & -a^2 \\ -a^2 & \mu_1\mu_2\end{bmatrix} \cdot \begin{bmatrix}1 \\ 1\end{bmatrix} \tag{4.3.18}$$

$$= \frac{2\mu_1\mu_2 - 2a^2}{(\mu_1\mu_2 - a^2)} \tag{4.3.19}$$

$$= \frac{2(\mu_1\mu_2 - a^2)}{(\mu_1\mu_2 - a^2)} \tag{4.3.20}$$

$$= 2. \tag{4.3.21}$$

The bivariate Poisson distribution is always equi-dispersed using FI_2 (by definition). This result is extended to the p-variate uncorrelated Poisson distribution by Corollary 4.2.3. Next we derive an expression for GDI.

$$\mathrm{GDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1 & a \\ a & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.22}$$

$$= \frac{\mu_1^2 + 2\mu_1\mu_2 a + \mu_2^2}{\mu_1^2 + \mu_2^2} \tag{4.3.23}$$

$$= 1 + \frac{2\lambda_3\mu_1\mu_2}{\mu_1^2 + \mu_2^2} \tag{4.3.24}$$

$$= 1 + \frac{2\lambda_3(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}{(\lambda_1 + \lambda_3)^2 + (\lambda_2 + \lambda_3)^2}. \tag{4.3.25}$$

Depending on λ_3, the bivariate Poisson can exhibit over/equi-dispersion to the uncorrelated bivariate Poisson distribution using GDI. For example, when $(\lambda_1, \lambda_2, \lambda_3) = (1, 1, 3)$ we have GDI $= 4 > 1$ and for $(\lambda_1, \lambda_2, \lambda_3) = (1, 1, 0)$ we have GDI $= 1$. Next we derive an expression for MDI, which is the same as GDI with $\lambda_3 = 0$.

$$\mathrm{MDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1 & 0 \\ 0 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.26}$$

$$= \frac{\mu_1^2 + \mu_2^2}{\mu_1^2 + \mu_2^2} \tag{4.3.27}$$

$$= 1. \tag{4.3.28}$$

Similar to FI_2, the bivariate Poisson distribution is equi-dispersed relative to the uncorrelated Poisson distribution using MDI.

Example 4.3.3 (Bivariate Negative Binomial)**.** Let $\mathbf{Y} \sim \mathrm{BivNB}(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$. Then $\mu_i = \mathrm{E}(Y_i) = (\alpha_0 + \alpha_i)\beta_i$, $\sigma_i = \mathrm{Var}(Y_i) = (\alpha_0 + \alpha_i)\beta_i(\beta_i + 1)$ and $a = \mathrm{cov}(Y_1, Y_2) = \alpha_0\beta_1\beta_2$ for $i = 1, 2$. Further, set $w = (\alpha_0 + \alpha_1)(\alpha_0 +$

$\alpha_2)(\beta_1+1)(\beta_2+1)$. First we derive an expression for FI_2.

$$\mathrm{FI}_2(\mathbf{Y}) = \frac{1}{\beta_1\beta_2 w-\alpha_0^2\beta_1^2\beta_2^2}\begin{bmatrix}\sqrt{\beta_1+1} & \sqrt{\beta_2+1}\end{bmatrix}\cdot\begin{bmatrix}\sigma_1\sigma_2 & -a^2\\ -a^2 & \sigma_1\sigma_2\end{bmatrix}\cdot\begin{bmatrix}\sqrt{\beta_1+1}\\ \sqrt{\beta_2+1}\end{bmatrix} \tag{4.3.29}$$

$$= \frac{\sigma_1\sigma_2(\beta_1+1)+\sigma_1\sigma_2(\beta_2+1)-2a^2\sqrt{\beta_1+1}\sqrt{\beta_2+1}}{\beta_1\beta_2 w-\alpha_0^2\beta_1^2\beta_2^2} \tag{4.3.30}$$

$$= \frac{w(\beta_1+1)+w(\beta_2+1)-2\alpha_0^2\beta_1\beta_2\sqrt{\beta_1+1}\sqrt{\beta_2+1}}{w-\alpha_0^2\beta_1\beta_2} \tag{4.3.31}$$

$$= \frac{w(2+\beta_1+\beta_2)-2\alpha_0^2\beta_1\beta_2\sqrt{\beta_1+1}\sqrt{\beta_2+1}}{w-\alpha_0^2\beta_1\beta_2} \tag{4.3.32}$$

$$= \frac{(\alpha_0+\alpha_1)(\alpha_0+\alpha_2)(\beta_1+1)(\beta_2+1)(2+\beta_1+\beta_2)}{(\alpha_0+\alpha_1)(\alpha_0+\alpha_2)(\beta_1+1)(\beta_2+1)-\alpha_0^2\beta_1\beta_2} \tag{4.3.33}$$

$$- \frac{2\alpha_0^2\beta_1\beta_2\sqrt{\beta_1+1}\sqrt{\beta_2+1}}{(\alpha_0+\alpha_1)(\alpha_0+\alpha_2)(\beta_1+1)(\beta_2+1)-\alpha_0^2\beta_1\beta_2}. \tag{4.3.34}$$

The bivariate Negative Binomial distribution is over-dispersed relative to the bivariate Poisson distribution using FI_2. For example, when $(\alpha_0,\alpha_1,\alpha_2,\beta_1,\beta_2) = (2,2,2,0.25,0.25)$ we have $\mathrm{FI}_2 = 2.5 > 2$. Next we derive an expression for GDI.

$$\mathrm{GDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix}\cdot\begin{bmatrix}\mu_1(\beta_1+1) & a\\ a & \mu_2(\beta_2+1)\end{bmatrix}\cdot\begin{bmatrix}\sqrt{\mu_1}\\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix}\cdot\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}} \tag{4.3.35}$$

$$= \frac{\mu_1^2(\beta_1+1)+\mu_2^2(\beta_2+1)+2a\sqrt{\mu_1\mu_2}}{\mu_1^2+\mu_2^2} \tag{4.3.36}$$

$$= 1+\frac{\mu_1^2\beta_1+\mu_2^2\beta_2+2\alpha_0\beta_1\beta_2\sqrt{\mu_1\mu_2}}{\mu_1^2+\mu_2^2} \tag{4.3.37}$$

$$= 1+\frac{(\alpha_0+\alpha_1)^2\beta_1^3+(\alpha_0+\alpha_2)^2\beta_2^3}{(\alpha_0+\alpha_1)^2\beta_1^2+(\alpha_0+\alpha_2)^2\beta_2^2} \tag{4.3.38}$$

$$+ \frac{2\alpha_0\beta_1^{\frac{3}{2}}\beta_2^{\frac{3}{2}}\sqrt{(\alpha_0+\alpha_1)(\alpha_0+\alpha_2)}}{(\alpha_0+\alpha_1)^2\beta_1^2+(\alpha_0+\alpha_2)^2\beta_2^2}. \tag{4.3.39}$$

The bivariate Negative Binomial distribution is over-dispersed relative to the bivariate Poisson distribution using GDI (this follows since the second term is always positive). For example, when $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2) = (2, 2, 2, 0.25, 0.25)$ we have GDI $= 1.28125 > 1$. Next we derive an expression for MDI which is the same as GDI with $a = 0$.

$$\mathrm{MDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1(\beta_1+1) & 0 \\ 0 & \mu_2(\beta_2+1)\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.40}$$

$$= \frac{\mu_1^2(\beta_1+1) + \mu_2^2(\beta_2+1)}{\mu_1^2 + \mu_2^2} \tag{4.3.41}$$

$$= 1 + \frac{\mu_1^2\beta_1 + \mu_2^2\beta_2}{\mu_1^2 + \mu_2^2} \tag{4.3.42}$$

$$= 1 + \frac{(\alpha_0+\alpha_1)^2\beta_1^3 + (\alpha_0+\alpha_2)^2\beta_2^3}{(\alpha_0+\alpha_1)^2\beta_1^2 + (\alpha_0+\alpha_2)^2\beta_2^2}. \tag{4.3.43}$$

Similar to GDI, the bivariate Negative Binomial distribution is always over-dispersed relative to the bivariate Poisson distribution using MDI. For example, when $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2) = (2, 2, 2, 0.25, 0.25)$ we have MDI $= 1.25 > 1$.

Example 4.3.4 (Bivariate Type I Pólya-Aeppli)**.** Let $\mathbf{Y} \sim \mathrm{BivPA}(\lambda_1, \lambda_2, \lambda_3, \rho)$. Then $\mu_i = \mathrm{E}(Y_i) = \frac{(\lambda_i+\lambda_3)}{1-\rho}$, $\sigma_i = \mathrm{Var}(Y_i) = \frac{(\lambda_i+\lambda_3)(1+\rho)}{(1-\rho)^2}$ and $a = \mathrm{cov}(Y_1, Y_2) = \frac{\lambda_3}{(1-\rho)^2}$ for $i = 1, 2$. First we derive an expression for FI_2.

$$\mathrm{FI}_2(\mathbf{Y}) = \frac{1}{(\sigma_1\sigma_2 - \sigma_{12}^2)} \begin{bmatrix}\sqrt{\frac{1+\rho}{1-\rho}} & \sqrt{\frac{1+\rho}{1-\rho}}\end{bmatrix} \cdot \begin{bmatrix}\sigma_1\sigma_2 & -a^2 \\ -a^2 & \sigma_1\sigma_2\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\frac{1+\rho}{1-\rho}} \\ \sqrt{\frac{1+\rho}{1-\rho}}\end{bmatrix} \tag{4.3.44}$$

$$= \frac{2\,\sigma_1\sigma_2(\frac{1+\rho}{1-\rho}) - 2a^2(\frac{1+\rho}{1-\rho})}{(\sigma_1\sigma_2 - a^2)} \tag{4.3.45}$$

$$= 2\left(\frac{1+\rho}{1-\rho}\right). \tag{4.3.46}$$

The Type I bivariate Pólya-Aeppli distribution is always over-dispersed relative to the bivariate Poisson distribution (since $\rho \in (0, 1)$). For example, when

$\rho = 0.5$ we have $\mathrm{FI}_2 = 6 > 2$. Next we derive an expression for GDI.

$$\mathrm{GDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \frac{(1+\rho)}{(1-\rho)} & a \\ a & \mu_2 \frac{(1+\rho)}{(1-\rho)}\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.47}$$

$$= \frac{\mu_1^2 \frac{(1+\rho)}{(1-\rho)} + \mu_2^2 \frac{(1+\rho)}{(1-\rho)} + 2a\sqrt{\mu_1}\sqrt{\mu_2}}{\mu_1^2 + \mu_2^2} \tag{4.3.48}$$

$$= \frac{\mu_1^2 \frac{(1-\rho+2\rho)}{(1-\rho)} + \mu_2^2 \frac{(1-\rho+2\rho)}{(1-\rho)} + 2a\sqrt{\frac{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{(1-\rho)^2}}}{\mu_1^2 + \mu_2^2} \tag{4.3.49}$$

$$= \frac{\mu_1^2 + \mu_2^2 + \frac{2\rho\mu_1^2 + 2\rho\mu_2^2}{(1-\rho)} + 2a\sqrt{\frac{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{(1-\rho)^2}}}{\mu_1^2 + \mu_2^2} \tag{4.3.50}$$

$$= 1 + 2\frac{(\mu_1^2 + \mu_2^2)\rho + a\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}{(1-\rho)(\mu_1^2 + \mu_2^2)} \tag{4.3.51}$$

$$\begin{aligned}= 1 &+ \frac{2\rho\left[(\lambda_1+\lambda_3)^2 + (\lambda_2+\lambda_3)^2\right]}{(1-\rho)(\lambda_1+\lambda_3)^2 + (1-\rho)(\lambda_2+\lambda_3)^2} \\ &+ \frac{2\lambda_3\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}{(1-\rho)(\lambda_1+\lambda_3)^2 + (1-\rho)(\lambda_2+\lambda_3)^2}\end{aligned} \tag{4.3.52}$$

$$= 1 + \frac{2\rho}{(1-\rho)} + \frac{2\lambda_3\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}{(1-\rho)(\lambda_1+\lambda_3)^2 + (1-\rho)(\lambda_2+\lambda_3)^2}. \tag{4.3.53}$$

The Type I bivariate Pólya-Aeppli distribution is always over-dispersed relative to the bivariate Poisson distribution using GDI. For example, when $(\lambda_1, \lambda_2, \lambda_3, \rho) = (1, 1, 1, 0.5)$ we have $\mathrm{GDI} = 4 > 1$. Lastly, we derive an expression for MDI which is the same as GDI with $a = 0$.

$$\mathrm{MDI}(\mathbf{Y}) = \frac{\begin{bmatrix}\sqrt{\mu_1} & \sqrt{\mu_2}\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \frac{(1+\rho)}{(1-\rho)} & 0 \\ 0 & \mu_2 \frac{(1+\rho)}{(1-\rho)}\end{bmatrix} \cdot \begin{bmatrix}\sqrt{\mu_1} \\ \sqrt{\mu_2}\end{bmatrix}}{\begin{bmatrix}\mu_1 & \mu_2\end{bmatrix} \cdot \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}} \tag{4.3.54}$$

$$= \frac{\mu_1^2 \frac{(1-\rho+2\rho)}{(1-\rho)} + \mu_2^2 \frac{(1-\rho+2\rho)}{(1-\rho)}}{\mu_1^2 + \mu_2^2} \tag{4.3.55}$$

$$= 1 + \frac{2\rho(\mu_1^2 + \mu_2^2)}{(1-\rho)(\mu_1^2 + \mu_2^2)} \tag{4.3.56}$$

$$= 1 + \frac{2\rho}{(1-\rho)}. \tag{4.3.57}$$

Similar to GDI, the Type I bivariate Pólya-Aeppli distribution is always over-dispersed relative to the bivariate Poisson distribution using MDI. For example, when $(\lambda_1, \lambda_2, \lambda_3, \rho) = (1, 1, 1, 0.5)$ we have GDI $= 3 > 1$.

Example 4.3.5 (Multivariate Poisson)**.** Let $\mathbf{Y} \sim \text{MVPois}(\lambda_1, \ldots, \lambda_p)$. Then $\mu_i = \text{E}(Z_i) = \text{Var}(Y_i) = \lambda_i + a$ and $\text{cov}(Y_i, Y_j) = a$. Again, we first derive an expression for FI_p.

$$\text{FI}_2(\mathbf{Y}) = \mathbf{1}^T \left[(D + a \cdot \mathbf{1}\mathbf{1}^T) \otimes \left(D^{-1} - \frac{D^{-1}(a \cdot \mathbf{1}\mathbf{1}^T)D^{-1}}{1 + a \cdot \mathbf{1}^T D^{-1} \mathbf{1}} \right) \right] \mathbf{1} \tag{4.3.58}$$

$$= p. \tag{4.3.59}$$

As mentioned, a proof of this result is given in Appendix B. The key ingredient to the proof is recognizing that the covariance matrix can be written as the sum of a diagonal and a rank-one matrix. Next we derive an expression for GDI.

$$\text{GDI}(\mathbf{Y}) = \frac{\sum_{k=1}^{p}(\lambda_k + a)^2 + a \cdot \left(\sum_{k=1}^{p} \sqrt{\lambda_k + a}\right) \left(\sum_{j \neq k} \sqrt{\lambda_j + a}\right)}{(\lambda_1 + a)^2 + \ldots + (\lambda_k + a)^2} \tag{4.3.60}$$

$$= 1 + a \frac{\left(\sum_{k=1}^{p} \sqrt{\lambda_k + a}\right) \left(\sum_{j \neq k} \sqrt{\lambda_j + a}\right)}{(\lambda_1 + a)^2 + \ldots + (\lambda_k + a)^2}. \tag{4.3.61}$$

When $a = 0$, the p-variate Poisson distribution is equi-dispersed. When $a > 0$, then it is over-dispersed relative to the uncorrelated p-variate Poisson distribution using GDI. Lastly, we derive an expression for MDI, which is the same as GDI with $a = 0$.

$$\text{MDI}(\mathbf{Y}) = \frac{(\lambda_1 + a)^2 + \ldots + (\lambda_k + a)^2}{(\lambda_1 + a)^2 + \ldots + (\lambda_k + a)^2} \tag{4.3.62}$$

$$= 1. \tag{4.3.63}$$

Similar to FI_p, the p-variate classical Poisson is always equi-dispersed using MDI. The authors Kokonendji and Puig (2018) commented on the similar behaviour of FI_2 and MDI during simulations. That both measures treat the p-variate classical Poisson distribution as always equi-dispersed reveals yet another interesting connection between FI_p and MDI.

Chapter 5

Results

In this chapter, we simulate datasets from known over/equi-dispersed models and measure how well FI_2 can discriminate between proposed models compared to GDI and MDI. Further, we compare the dispersion indices on a stock trading volume dataset, football dataset and insurance claims dataset. The following analysis was conducted using `R` (R Core Team (2023)).

5.1 Simulation Study

In this simulation study, the goal is to investigate the performance of FI_2 in detecting over/equi-dispersion given the true data distribution is known to be over/equi-dispersed. The original methodology included a bootstrap component of the simulation study. However, due to technical difficulties with the numerical optimization, bootstrap estimation was not used. Instead, we present three datasets simulated from known over/equi-dispersed models and fit the Poisson, Pólya-Aeppli and Negative Binomial models to the entire dataset (which did not have issues). Using the fitted models, we compute FI_2, GDI and MDI for each model, then make some comments on their interpretation.

For the first case, we generated $N = 500$ observations from a bivariate Poisson distribution with parameters $(\lambda_1, \lambda_2, \lambda_3) = (2, 4, 1)$ using the `rbvpois` function in the `bivpois` package (Tsagris (2023)). The observed range of values for Y_1 and Y_2 are 0 to 8 and 0 to 15, respectively. The mean and variance for Y_1 are 2.94 and 2.58, while the mean and variance for Y_2 are 4.98 and 5.27. The marginal dispersion indices for Y_1 and Y_2 are 0.88 and 1.06, respectively. Table 5.1 displays the MLEs for the Bivariate Poisson, Pólya-Aeppli and Negative Binomial distributions.

MLEs	Poisson	Pólya-Aeppli	Neg. Bin.
$\widehat{\lambda_1}$	1.9843	2.0200	
$\widehat{\lambda_2}$	4.0283	4.0742	
$\widehat{\lambda_3}$	0.9517	0.8940	
$\widehat{\rho}$		0.0005	
$\widehat{\alpha_0}$			36.1570
$\widehat{\alpha_1}$			1.0358
$\widehat{\alpha_2}$			0.8099
$\widehat{\beta_1}$			0.0789
$\widehat{\beta_2}$			0.1347
$-\log L$	-2047.87	-2033.05	-2042.96

Table 5.1: MLEs for the three bivariate models.

Table 5.2 displays the indices evaluated at the MLEs found in Table 5.1. The true index values for this dataset are $\mathrm{FI}_2 = 1$, GDI = 1.88 and MDI = 1. We see FI_2 and GDI are close to unity for each model, indicating the data are equi-dispersed. GDI has a large range of values indicating varying levels of over-dispersion under each model. All three indices are close to their true values, but the different interpretations of FI_2, GDI and MDI under the Poisson model make direct comparison difficult.

Index	Poisson	Pólya-Aeppli	Neg. Bin.
FI_2	1.0000	1.0011	1.1068
GDI	1.8327	1.2063	1.1170
MDI	1.0000	1.0011	1.1067

Table 5.2: Indices evaluated at the MLEs for each model presented in Table 5.1.

For the second case, we generated $N = 500$ observations from the bivariate negative binomial distribution with parameters $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2) = (4, 4, 6, 0.25, 0.25)$ using the `rbnb` function in the `bzinb` package (Cho et al. (2022)). The observed range of values for Y_1 and Y_2 are 0 to 9 and 0 to 9, respectively. The mean and variance for Y_1 are 2.12 and 2.81, while the mean and variance for Y_2 are 2.58 and 3.25. The marginal dispersion indices for Y_1

and Y_2 are 1.33 and 1.26, respectively. Table 5.3 displays the MLEs for the Bivariate Poisson, Pólya-Aeppli and Negative Binomial distributions.

MLEs	Poisson	Pólya-Aeppli	Neg. Bin.
$\widehat{\lambda_1}$	1.9460	0.6437	
$\widehat{\lambda_2}$	2.4020	1.1291	
$\widehat{\lambda_3}$	0.1760	0.9475	
$\widehat{\rho}$		0.1947	
$\widehat{\alpha_0}$			3.1287
$\widehat{\alpha_1}$			3.1292
$\widehat{\alpha_2}$			6.7100
$\widehat{\beta_1}$			0.3391
$\widehat{\beta_2}$			0.2620
$-\log L$	-1902.46	-1928.70	-1882.69

Table 5.3: MLEs for the three bivariate models.

Table 5.4 displays the indices evaluated at the MLEs found in Table 5.3. The true index values for this dataset are $\text{FI}_2 = 1.25$, GDI $= 1.28$ and MDI $=$ 1.25. Under the Poisson model, FI_2 and MDI do not provide any information about the underlying dispersion since they are constant (as a function of the MLEs). Further, we see that all three indices are close to their trues values for the Negative Binomial model, indicating the data is over-dispersed.

Index	Poisson	Pólya-Aeppli	Neg. Bin.
FI_2	1.0000	1.4834	1.3006
GDI	1.1727	2.1083	1.3163
MDI	1.0000	1.4834	1.2842

Table 5.4: Indices evaluated at the MLEs for each model presented in Table 5.3.

For the third case, we generated $N = 500$ observations from the bivariate Pólya-Aeppli distribution with parameters $(\lambda_1, \lambda_2, \lambda_3, \rho) = (0.6, 0.6, 0.9, 0.25)$ using the method described in Balakrishnan et al. (2017). The observed range of values for Y_1 and Y_2 are 0 to 14 and 0 to 11, respectively. The mean and variance for Y_1 are 2.02 and 3.53, while the mean and variance for Y_2 are 1.81

and 2.88. The marginal dispersion indices for Y_1 and Y_2 are 1.75 and 1.59, respectively. Table 5.5 displays the MLEs for the Bivariate Poisson, Pólya-Aeppli and Negative Binomial distributions.

MLEs	Poisson	Pólya-Aeppli	Neg. Bin.
$\widehat{\lambda_1}$	1.2804	0.6671	
$\widehat{\lambda_2}$	1.0724	0.5225	
$\widehat{\lambda_3}$	0.7356	0.8383	
$\widehat{\rho}$		0.2498	
$\widehat{\alpha_0}$			2.2907
$\widehat{\alpha_1}$			0.1016
$\widehat{\alpha_2}$			0.0824
$\widehat{\beta_1}$			0.8422
$\widehat{\beta_2}$			0.7618
$-\log L$	-1888.34	-1745.30	-1758.01

Table 5.5: MLEs for the three bivariate models.

Table 5.6 displays the indices evaluated at the MLEs found in Table 5.5. The true index values for this dataset are $\mathrm{FI}_2 = 1.67$, $\mathrm{GDI} = 2.47$ and $\mathrm{MDI} = 1.67$. Under the Pólya-Aeppli model, all three indices underestimate the true index value whereas under the Negative Binomial model, all three indices overestimate the true index value. In either case, the indices correctly classify the underlying data as over-dispersed. Further, we see that FI_2 and GDI under the Poisson model imply the data are equi-dispersed, whereas GDI correctly identifies over-dispersion.

Index	Poisson	Pólya-Aeppli	Neg. Bin.
FI_2	1.0000	1.6661	1.8021
GDI	1.7313	2.4428	2.4192
MDI	1.0000	1.6661	1.8023

Table 5.6: Indices evaluated at the MLEs for each model presented in Table 5.5.

5.2 Simulated Quintivariate Count Dataset

Kokonendji and Puig (2018) used the `NORTARA` package to simulate a quintivariate dataset of size $n = 300$ (Su (2014)). The marginal distributions of the quintivariate dataset are Poisson, binomial and negative binomial (three of the marginal distributions are negative binomial). Table 5.7 displays a summary of the dataset and results of the dispersion indices using sample estimates of the mean vector and covariance matrix.

X_j	Mean	Variance	Dispersion Index
1	1.2633	1.6930	
2	4.9467	5.3483	$\widehat{\text{GDI}} = 1.8995$
3	5.0300	6.3703	$\widehat{\text{MDI}} = 1.3464$
4	0.4767	0.2503	$\widehat{\text{FI}_5} = 3.4779$
5	2.7300	6.8667	

Table 5.7: Simulated quintivariate data. Expected value and variance of marginal distributions presented with multivariate dispersion indices.

We see from Table 5.7 that $\text{GDI} > 1$, $\text{MDI} > 1$ and $\text{FI}_5 > 1$, all indicating over-dispersion. The value of FI_5 is almost double the values of GDI and MDI, indicating that GDI and MDI under-estimate the amount of dispersion present in the data.

5.3 Intra-day Trading Volume Dataset

From `www.finam.ru`, Kokonendji and Puig (2018) downloaded the intra-day common stock trade volumes from May 6, 2015 to June 6, 2015 for JPMorgan Chase & Co (NYSE: JPM), Bank of America (NYSE: BAC) and Wells Fargo (NYSE: WFC). Table 5.8 displays a summary of the dataset and results of the dispersion indices using sample estimates of the mean vector and covariance matrix.

Bank	Mean	Variance	Correlation	Dispersion Index
JPM	11,965.52	96,707,437	JPM & BAC: 0.381	$\widehat{\text{GDI}} = 59975.32$
BAC	44,455.59	2,550,265,224	JPM & WFC 0.565	$\widehat{\text{MDI}} = 51350.61$
WFC	11,654.23	109,833,650	BAC & WFC: 0.388	$\widehat{\text{FI}_3} = 26618.91$

Table 5.8: Summary statistics from the intra-day trading volume dataset.

We see from Table 5.8 that all three indices indicate significant over-dispersion. The value of FI_3 is less than half that of GDI and MDI, indicating less over-dispersion is present in the data.

5.4 Football Reference Dataset

European football has an abundance of count data. Many statistics of interest are naturally represented as counts such as the number of goals scored in a game and the number of successful passes by each team. In the last decade, the rivalry between Real Madrid and Barcelona in Spain was marked by the presence of Cristiano Ronaldo and Lionel Messi, two of the greatest men's goalscorers of all time (IFFHS (2023)). The three seasons between 2014-2015 and 2017-2018 saw the height of the rivalry between Real Madrid and Barcelona as they both had formidable attacking trios up front. Real Madrid with Cristiano Ronaldo, Karim Benzema and Gareth Bale, Barcelona with Lionel Messi, Luis Súarez and Neymar.

Using match logs from FBref (2024), we created a dataset containing all of the goals scored per game in La Liga by the Real Madrid and Barcelona attacking trios between 2014-2015 and 2017-2018. Games where all three players did not appear were removed, and only games played in La Liga were included. The resulting dataset contains 115 matches for Barcelona and 67 matches for Real Madrid. Our goal is to assess the amount of dispersion present in the data using bootstrap estimates of FI_3, GDI and MDI for Barcelona and Real Madrid respectively.

Summary statistics for the Real Madrid (RM) and Barcelona (BAR) attacking trio datasets are presented in Table 5.9. We see that Ronaldo and Messi have the highest average goals on their team. Further, Bale and Súarez have the largest dispersion indices on their respective teams.

#	Team	Player	$\bar{x}$	s^2	Dispersion Index
1	RM	Gareth Bale	0.4627	0.6463	1.3969
2	RM	Karim Benzema	0.5672	0.5219	0.9203
3	RM	Cristiano Ronaldo	1.0896	1.3252	1.2163
4	BAR	Lionel Messi	0.9130	0.8871	0.9716
5	BAR	Luis Súarez	0.7565	0.9402	1.2428
6	BAR	Neymar	0.5304	0.4267	0.8044

Table 5.9: Summary statistics by team and player. Sample mean, sample variance and (marginal) dispersion index shown.

Below in Figure 5.1 we plot the frequency of goals scored per game per player. We see Ronaldo and Messi have similar numbers of zero and one goal games, whereas Bale has many zero goal games and some game where he scores one or more goals.

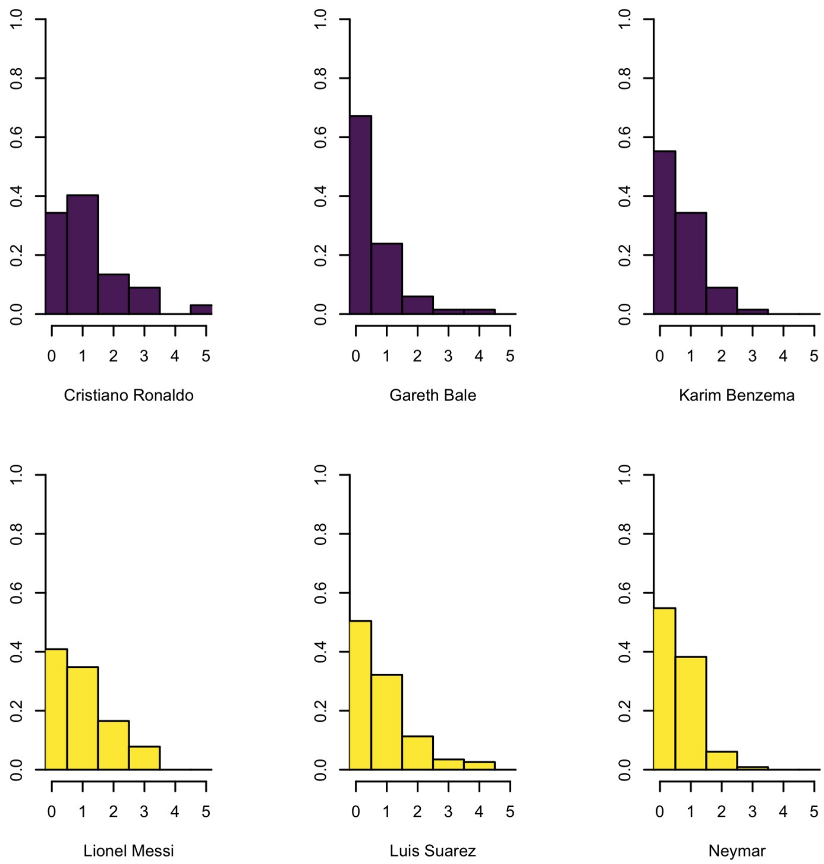

Figure 5.1: Goal frequency for each Real Madrid (Purple) and Barcelona (Yellow) player.

Below in Figure 5.2 we plot the bootstrap distributions of FI_3, GDI and MDI. The plots indicate that goal counts are approximately equi-dispersed for Barcelona, and slightly over-dispersed for Real Madrid. This indicates that a trivariate Poisson model may be suitable for Barcelona, whereas a trivariate distribution that can account for over-dispersion would be suitable for Real Madrid.

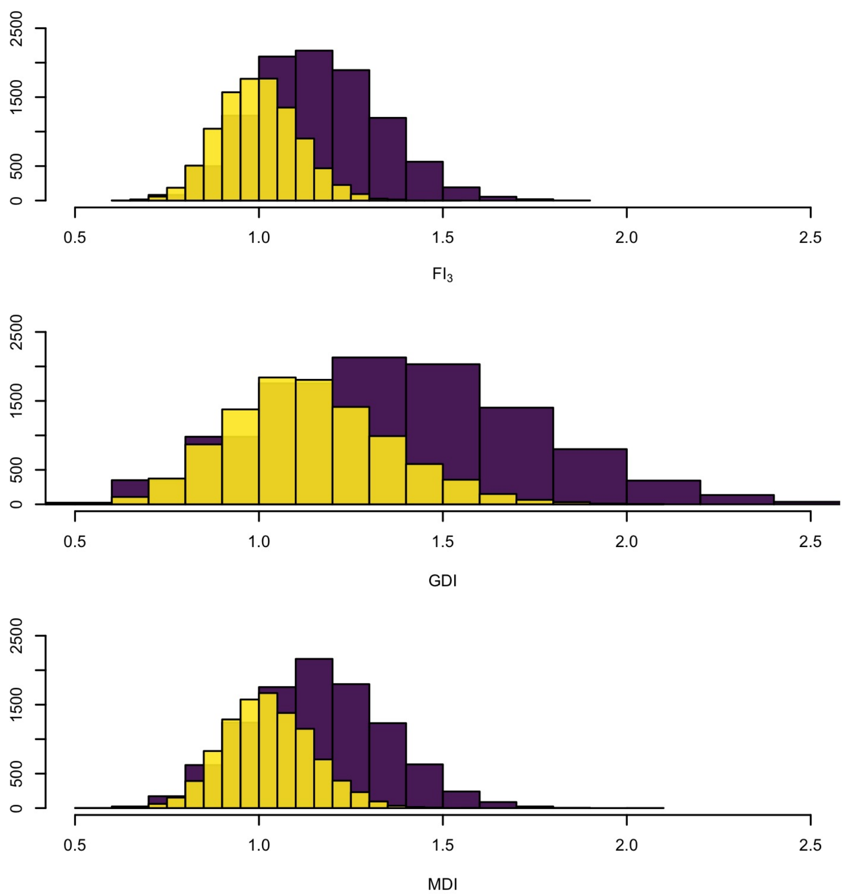

Figure 5.2: Histogram of bootstrap FI_3, GDI and MDI values for Real Madrid (Purple) and Barcelona (Yellow)

We provide a summary of results for the B = 10,000 bootstrap samples in Table 5.10 below. Based on the confidence intervals, we cannot reject the null hypothesis that both datasets are equi-dispersed. Further, we find that confidence interval for FI_3 is shorter than GDI and MDI across both datasets.

#	Team	Index	$\bar{x}$	$\hat{\sigma}_{se}$	95% CI	CI Length
1	RM	FI_3	1.1534	0.1677	(0.8465, 1.4924)	0.6459
2	RM	GDI	1.3868	0.3498	(0.7548, 2.1032)	1.3484
3	RM	MDI	1.1556	0.1862	(0.8027, 1.5345)	0.7318
4	BAR	FI_3	0.9988	0.1082	(0.7981, 1.2181)	0.4200
5	BAR	GDI	1.1370	0.2205	(0.7491, 1.6056)	0.8565
6	BAR	MDI	1.0257	0.1205	(0.8052, 1.2734)	0.4682

Table 5.10: Bootstrap summary statistics. $100(1-\alpha)\%$ confidence interval provided for index value using percentile method.

Below we display the bias and MSE for the bootstrap estimates. We found the bias and MSE were smallest for FI_3 and MDI across both datasets. Further, the bias and MSE are small and positive for each index, indicating overestimation of the dispersion in the data.

#	Team	Index	Bias	MSE
1	RM	FI_3	0.0278	0.0289
2	RM	GDI	0.0434	0.1242
3	RM	MDI	0.0278	0.0354
4	BAR	FI_3	0.0076	0.0118
5	BAR	GDI	0.0128	0.0246
6	BAR	MDI	0.0100	0.0112

Table 5.11: Bootstrap summary statistics. $100(1-\alpha)\%$ confidence interval provided for index value using percentile method.

5.5 Posterior Ratemaking Dataset

The posterior ratemaking dataset contains $80,994$ automobile insurance claims made to a Spanish insurance company in 1995. Each observation is represented as a bivariate count vector, denoted $\mathbf{N} = (N_1, N_2)$, where N_1 represents claims made under third-party liability policies and N_2 represents claims made under guarantees such as emergency roadside assistance and medical cost coverage (Bermúdez i Morata (2009)). The cross-tabulation of claims from the two types of policies is shown in Table A.1. Table 5.12 presents the MLEs for the Bivariate Poisson, Pólya–Aeppli and Negative Binomial models.

MLEs	Poisson	Pólya-Aeppli	Neg. Bin.
$\widehat{\lambda_1}$	0.0670	0.0516	
$\widehat{\lambda_2}$	0.0884	0.0658	
$\widehat{\lambda_3}$	0.0140	0.0132	
$\widehat{\rho}$		0.2167	
$\widehat{\alpha_0}$			0.0822
$\widehat{\alpha_1}$			0.0684
$\widehat{\alpha_2}$			0.0710
$\widehat{\beta_1}$			0.5372
$\widehat{\beta_2}$			0.6695
$-\log L$	-53271.05	-48087.48	-48050.41

Table 5.12: MLEs for the three bivariate models.

For each fitted model, we present the indices evaluated using the MLEs in Table 5.12. We see that each index indicates over-dispersion in the data, except for Poisson where FI_2 and MDI are always equi-dispersed (using the MLEs).

Index	Poisson	Pólya-Aeppli	Neg. Bin.
FI_2	1.0000	1.5533	1.6034
GDI	1.0136	1.7849	1.7991
MDI	1.0000	1.5533	1.6044

Table 5.13: Indices evaluated at the MLEs for each model presented in Table 5.12.

Chapter 6

Conclusion

In this thesis, we have proposed a new multivariate extension of the Fisher dispersion index. Some interesting properties of FI_p have been discussed, with applications to some common bivariate and multivariate discrete distributions and real-world datasets. One potential issue in practice is the need to compute the inverse of the covariance matrix. This can present a challenge when using numerical linear algebra routines since they may be unable to compute a matrix inverse, though it is known to exist. In practice, many matrices are invertible, and thus this issue of minor importance. Further, when suitable, the generalized inverse can be used in place of the usual inverse covariance matrix to compute the index.

To conclude, FI_p has several interesting mathematical properties, and presents a natural extension of the univariate dispersion index to the multivariate setting. Previously proposed indices such as GDI and MDI do not take into account crucial dependency information present in the covariance matrix, making it simpler to use computationally but potentially misleading. This dependency information is naturally included in FI_p through the covariance matrix and marginal dispersion indices. Future research is required to develop hypothesis testing and goodness-of-fit tests. Further, exploration of the mathematical properties of FI_p may also provide a fruitful research direction.

Appendix A

Supplementary Data

A.1 Posterior Ratemaking Dataset

	N_2							
N_1	0	1	2	3	4	5	6	7
0	71,087	3,722	807	219	51	14	4	0
1	3,022	686	184	71	26	10	3	1
2	574	138	55	15	8	4	1	1
3	149	42	21	6	6	1	0	1
4	29	15	3	2	1	1	0	0
5	4	1	0	0	0	0	2	0
6	2	1	0	1	0	0	0	0
7	1	0	0	1	0	0	0	0
8	0	0	1	0	0	0	0	0

Table A.1: Automobile insurance database cross-tabulation.
N_1: number of claims for third-party liability.
N_2: number of claims for the rest of guarantees.

Appendix B

Proofs

B.1 Delta Method Std. Errors

B.1.1 Bivariate Poisson

Let $\mu_i = \lambda_i + \lambda_3$. The gradient vector of $\mathrm{FI}_2(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_1} = 0 \tag{B.1.1}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_2} = 0 \tag{B.1.2}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_3} = 0. \tag{B.1.3}$$

Let $D = \mu_1^2 + \mu_2^2$. The gradient vector of $\mathrm{GDI}(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \lambda_1} = \frac{2\mu_2\lambda_3(\mu_2^2 - \mu_1^2)}{D^2} \tag{B.1.4}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \lambda_2} = \frac{2\mu_1\lambda_3(\mu_1^2 - \mu_2^2)}{D^2} \tag{B.1.5}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \lambda_3} = \frac{2\mu_1\mu_2 \cdot D + 2\lambda_3(\mu_1 + \mu_2)^2(\mu_1 - \mu_2)}{D^2}. \tag{B.1.6}$$

The gradient vector of $\mathrm{MDI}(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{MDI}(\boldsymbol{\theta})}{\partial \lambda_1} = 0 \tag{B.1.7}$$

$$\frac{\partial \mathrm{MDI}(\boldsymbol{\theta})}{\partial \lambda_2} = 0 \tag{B.1.8}$$

$$\frac{\partial \mathrm{MDI}(\boldsymbol{\theta})}{\partial \lambda_3} = 0. \tag{B.1.9}$$

B.1.2 Bivariate Negative Binomial

Let $\mu_i = (\alpha_0 + \alpha_i)$, $\bar{\beta}_i = (\beta_i + 1)$, $N = (\mu_1\mu_2\bar{\beta}_1\bar{\beta}_2(\bar{\beta}_1 + \bar{\beta}_1) - 2\alpha_0^2\beta_1\beta_2\sqrt{\bar{\beta}_1\bar{\beta}_2})$ and $D = (\mu_1\mu_2\bar{\beta}_1\bar{\beta}_2 - \alpha_0^2\beta_1\beta_2)$. The gradient vector of $\mathrm{FI}_2(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \alpha_0} = \frac{2\mu_1\mu_2\beta_1^2\beta_2^2(\mu_1(\beta_2 - \beta_1) + \mu_2(\beta_1 - \beta_2))}{D^2} \tag{B.1.10}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \alpha_1} = \frac{\mu_2\bar{\beta}_1\bar{\beta}_2\alpha_0^2\beta_1\beta_2(2\sqrt{\bar{\beta}_1\bar{\beta}_2} - (\bar{\beta}_1 + \bar{\beta}_2))}{D^2} \tag{B.1.11}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \alpha_2} = \frac{\mu_1\bar{\beta}_1\bar{\beta}_2\alpha_0^2\beta_1\beta_2(2\sqrt{\bar{\beta}_1\bar{\beta}_2} - (\bar{\beta}_1 + \bar{\beta}_2))}{D^2} \tag{B.1.12}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \beta_1} = \frac{\mu_1\mu_2\bar{\beta}_2(2\bar{\beta}_1 + \bar{\beta}_2) - \alpha_0^2\beta_2\sqrt{\bar{\beta}_1\bar{\beta}_2}\left(1 - \beta_1\bar{\beta}_1^{-1}\right)}{D} \tag{B.1.13}$$

$$- \frac{N \cdot (\mu_1\mu_2\bar{\beta}_2 - \alpha_0^2\beta_2)}{D^2} \tag{B.1.14}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \beta_2} = \frac{\mu_1\mu_2\bar{\beta}_1(2\bar{\beta}_2 + \bar{\beta}_1) - \alpha_0^2\beta_1\sqrt{\bar{\beta}_1\bar{\beta}_2}\left(1 - \beta_2\bar{\beta}_2^{-1}\right)}{D} \tag{B.1.15}$$

$$- \frac{N \cdot (\mu_1\mu_2\bar{\beta}_1 - \alpha_0^2\beta_1)}{D^2}. \tag{B.1.16}$$

Let $N = (\mu_1^2\beta_1^3 + \mu_2^2\beta_2^3 + 2\alpha_0\beta_1^{\frac{3}{2}}\beta_2^{\frac{3}{2}}\sqrt{\mu_1\mu_2})$ and $D = (\mu_1^2\beta_1^2 + \mu_2^2\beta_2^2)$. The gradient vector of $\mathrm{GDI}(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \alpha_0} = \frac{(2\mu_1^{\frac{3}{2}}\sqrt{\mu_2}\beta_1^3 + 2\sqrt{\mu_1}\mu_2^{\frac{3}{2}}\beta_2^3 + \beta_1^{\frac{3}{2}}\beta_2^{\frac{3}{2}}(2\mu_1\mu_2 + \alpha_0(\mu_1 + \mu_2))}{\sqrt{\mu_1\mu_2} \cdot D} \tag{B.1.17}$$

$$- \frac{2\sqrt{\mu_1\mu_2}(\mu_1\beta_1^2 + \mu_2\beta_2^2) \cdot N}{D^2} \tag{B.1.18}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \alpha_1} = \frac{2\mu_1^{\frac{3}{2}}\beta_1^3 + \alpha_0\beta_1^{\frac{3}{2}}\beta_2^{\frac{3}{2}}\sqrt{\mu_2}}{\sqrt{\mu_1}\cdot D} - \frac{2\mu_1\beta_1^2\cdot N}{D^2} \tag{B.1.19}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \alpha_2} = \frac{2\mu_2^{\frac{3}{2}}\beta_2^3 + \alpha_0\beta_1^{\frac{3}{2}}\beta_2^{\frac{3}{2}}\sqrt{\mu_1}}{\sqrt{\mu_2}\cdot D} - \frac{2\mu_2\beta_2^2\cdot N}{D^2} \tag{B.1.20}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \beta_1} = \frac{3(\mu_1^2\beta_1^2 + \alpha_0\sqrt{\beta_1}\beta_2^{\frac{3}{2}}\sqrt{\mu_1\mu_2})}{D} - \frac{2\mu_1^2\beta_1\cdot N}{D^2} \tag{B.1.21}$$

$$\frac{\partial \mathrm{GDI}(\boldsymbol{\theta})}{\partial \beta_2} = \frac{3(\mu_2^2\beta_2^2 + \alpha_0\beta_1^{\frac{3}{2}}\sqrt{\beta_2}\sqrt{\mu_1\mu_2})}{D} - \frac{2\mu_2^2\beta_2\cdot N}{D^2}. \tag{B.1.22}$$

The gradient vector of $\mathrm{MDI}(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{MDI}(\theta)}{\partial \alpha_0} = \frac{2\mu_1\mu_2\beta_1^2\beta_2^2(\mu_1(\beta_2-\beta_1)+\mu_2(\beta_1-\beta_2))}{D^2} \tag{B.1.23}$$

$$\frac{\partial \mathrm{MDI}(\theta)}{\partial \alpha_1} = \frac{2\mu_1\mu_2^2\beta_1^2\beta_2^2(\beta_1-\beta_2)}{D^2} \tag{B.1.24}$$

$$\frac{\partial \mathrm{MDI}(\theta)}{\partial \alpha_2} = \frac{2\mu_1^2\mu_2\beta_1^2\beta_2^2(\beta_2-\beta_1)}{D^2} \tag{B.1.25}$$

$$\frac{\partial \mathrm{MDI}(\theta)}{\partial \beta_1} = \frac{\mu_1^2\beta_1(\mu_1^2\beta_1^3+\mu_2^2\beta_2^2(3\beta_1-2\beta_2))}{D^2} \tag{B.1.26}$$

$$\frac{\partial \mathrm{MDI}(\theta)}{\partial \beta_2} = \frac{\mu_2^2\beta_2(\mu_2^2\beta_2^3+\mu_1^2\beta_1^2(3\beta_2-2\beta_1))}{D^2}. \tag{B.1.27}$$

B.1.3 Bivariate Pólya-Aeppli

Let $\mu_i = \lambda_i + \lambda_3$. The gradient vector of $\mathrm{FI}_2(\boldsymbol{\theta})$ has components

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_1} = 0 \tag{B.1.28}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_2} = 0 \tag{B.1.29}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \lambda_3} = 0 \tag{B.1.30}$$

$$\frac{\partial \mathrm{FI}_2(\boldsymbol{\theta})}{\partial \rho} = \frac{4}{(1-\rho)^2}. \tag{B.1.31}$$

Let $D = (1-\rho)(\mu_1^2 + \mu_2^2)$. The gradient vector of GDI($\boldsymbol{\theta}$) has components

$$\frac{\partial \text{GDI}(\boldsymbol{\theta})}{\partial \lambda_1} = \sqrt{\frac{\mu_1}{\mu_2}} \cdot \frac{2(1-\rho)(\mu_1^2 - 3\mu_2^2)}{D^2} \tag{B.1.32}$$

$$\frac{\partial \text{GDI}(\boldsymbol{\theta})}{\partial \lambda_2} = \sqrt{\frac{\mu_2}{\mu_1}} \cdot \frac{2(1-\rho)(\mu_2^2 - 3\mu_1^2)}{D^2} \tag{B.1.33}$$

$$\frac{\partial \text{GDI}(\boldsymbol{\theta})}{\partial \lambda_3} = \frac{2\mu_1\mu_2 \cdot D + \lambda_3(1-\rho)(\mu_1 - \mu_2)^2(\mu_1 + \mu_2)}{\sqrt{\mu_1\mu_2} \cdot D^2} \tag{B.1.34}$$

$$\frac{\partial \text{GDI}(\boldsymbol{\theta})}{\partial \rho} = \frac{2(\mu_1^2 + \mu_2^2) - 2\lambda_3\sqrt{\mu_1\mu_2}}{(1-\rho) \cdot D} \tag{B.1.35}$$

The gradient vector of MDI($\boldsymbol{\theta}$) has components

$$\frac{\partial \text{MDI}(\boldsymbol{\theta})}{\partial \lambda_1} = 0 \tag{B.1.36}$$

$$\frac{\partial \text{MDI}(\boldsymbol{\theta})}{\partial \lambda_2} = 0 \tag{B.1.37}$$

$$\frac{\partial \text{MDI}(\boldsymbol{\theta})}{\partial \lambda_3} = 0 \tag{B.1.38}$$

$$\frac{\partial \text{MDI}(\boldsymbol{\theta})}{\partial \rho} = \frac{2}{(1-\rho)^2}. \tag{B.1.39}$$

B.2 FI_p for p-variate Poisson

First we demonstrate the results for $p = 3$. Using trivariate reduction, the covariance matrix is given by

$$\Sigma = \begin{bmatrix} \lambda_1 + a & a & a \\ a & \lambda_2 + a & a \\ a & a & \lambda_3 + a \end{bmatrix} \tag{B.2.1}$$

$$= \begin{bmatrix} \lambda_1 & 0 & -\lambda_3 \\ 0 & \lambda_2 & -\lambda_3 \\ a & a & \lambda_3 + a \end{bmatrix}. \tag{B.2.2}$$

The determinant of Σ is given by

$$|\Sigma| = \lambda_1\lambda_2\lambda_3 + a(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\,, \tag{B.2.3}$$

and the inverse matrix is given by

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} \lambda_2\lambda_3 + a\lambda_2 + a\lambda_3 & -a\lambda_3 & -a\lambda_2 \\ -a\lambda_3 & \lambda_1\lambda_3 + a\lambda_1 + a\lambda_3 & -a\lambda_1 \\ \lambda_2\lambda_3 & \lambda_1\lambda_3 & \lambda_1\lambda_2 \end{bmatrix}. \tag{B.2.4}$$

Plugging this into the formula for FI_3, we have that

$$\mathrm{FI}_p(\mathbf{Y}) = \sqrt{\mathbf{d}}^T\left(\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1}\right)\sqrt{\mathbf{d}} \tag{B.2.5}$$

$$= \frac{1}{|\Sigma|}\mathbf{1}^T \begin{bmatrix} \lambda_1\lambda_2\lambda_3+a\lambda_1\lambda_2+a\lambda_1\lambda_3 & 0 & a\lambda_2\lambda_3 \\ 0 & \lambda_1\lambda_2\lambda_3+a\lambda_1\lambda_2+a\lambda_2\lambda_3 & a\lambda_1\lambda_3 \\ a\lambda_2\lambda_3 & a\lambda_1\lambda_3 & \lambda_1\lambda_2\lambda_3+a\lambda_1\lambda_2 \end{bmatrix}\mathbf{1} \tag{B.2.6}$$

$$= \frac{1}{|\Sigma|}\cdot\left[3\lambda_1\lambda_2\lambda_3 + 3a(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\right] \tag{B.2.7}$$

$$= \frac{3}{|\Sigma|}\cdot\left[\lambda_1\lambda_2\lambda_3 + a(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\right] \tag{B.2.8}$$

$$= 3\cdot\frac{|\Sigma|}{|\Sigma|} \tag{B.2.9}$$

$$= 3. \tag{B.2.10}$$

The key observation here is that the vector of ones $\mathbf{1}^T = (1, \dots, 1)$ on both sides of the quadratic form changes the product into a sum over all of the entries of $(\mathrm{cov}(\mathbf{Y}) \otimes \mathrm{cov}(\mathbf{Y})^{-1})$.

For the p-variate case, we have the covariance matrix

$$\Sigma = \begin{bmatrix} \lambda_1 + a & a & a & \dots & a \\ a & \lambda_2 + a & a & \dots & a \\ a & a & \lambda_3 + a & \dots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \dots & \lambda_p + a \end{bmatrix}. \tag{B.2.11}$$

Observe that Σ is the sum of a diagonal and a rank-one matrix. Thus we can write Σ as the sum of two matrices

$$\Sigma = D + a \cdot \mathbf{1}\mathbf{1}^T \tag{B.2.12}$$

$$= \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_p \end{bmatrix} + \begin{bmatrix} a & a & a & a & a \\ a & a & a & \dots & a \\ a & a & a & \dots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \dots & a \end{bmatrix}. \tag{B.2.13}$$

By the Sherman-Morrison-Woodbury matrix inverse formula (Horn and Johnson (2012)), the inverse of this matrix is given by

$$\Sigma^{-1} = D^{-1} - \frac{D^{-1}(a \cdot \mathbf{1}\mathbf{1}^T)D^{-1}}{1 + a \cdot \mathbf{1}^T D^{-1}\mathbf{1}}. \tag{B.2.14}$$

To simplify the derivation, we note the following useful results

$$1 + a \cdot \mathbf{1}^T D^{-1}\mathbf{1} = 1 + a\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_p}\right) \tag{B.2.15}$$

$$= 1 + aq \tag{B.2.16}$$

$$q^2 = \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_p}\right)\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_p}\right) \tag{B.2.17}$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{1}{\lambda_i}\frac{1}{\lambda_j}. \tag{B.2.18}$$

Plugging these results into $\text{FI}_p(\mathbf{Y})$, we have for any classical p-variate Poisson distribution that

$$\text{FI}_p(\mathbf{Y}) = \sqrt{\mathbf{d}}^T \left(\text{cov}(\mathbf{Y}) \otimes \text{cov}(\mathbf{Y})^{-1}\right) \sqrt{\mathbf{d}} \tag{B.2.19}$$

$$= \mathbf{1}^T \left[(D + a \cdot \mathbf{1}\mathbf{1}^T) \otimes \left(D^{-1} - \frac{D^{-1}(a \cdot \mathbf{1}\mathbf{1}^T)D^{-1}}{1 + a \cdot \mathbf{1}^T D^{-1}\mathbf{1}} \right) \right] \mathbf{1} \tag{B.2.20}$$

$$= \mathbf{1}^T \left[(D + a \cdot \mathbf{1}\mathbf{1}^T) \otimes \left(D^{-1} - a \cdot \frac{D^{-1}\mathbf{1}\mathbf{1}^T D^{-1}}{1 + aq} \right) \right] \mathbf{1} \tag{B.2.21}$$

$$= \mathbf{1}^T(D \otimes D^{-1})\mathbf{1} - \frac{a}{1+aq}\mathbf{1}^T(D \otimes (D^{-1}\mathbf{1}\mathbf{1}^T D^{-1}))\mathbf{1} \tag{B.2.22}$$

$$a \cdot \mathbf{1}^T(\mathbf{1}\mathbf{1}^T \otimes D^{-1})\mathbf{1} - \frac{a^2}{1+aq}\mathbf{1}^T(\mathbf{1}\mathbf{1}^T \otimes (D^{-1}\mathbf{1}\mathbf{1}^T D^{-1}))\mathbf{1} \tag{B.2.23}$$

$$= p - \frac{aq}{1+aq} + aq - \frac{a^2}{1+aq}\mathbf{1}^T \begin{bmatrix} \frac{1}{\lambda_1^2} & \frac{1}{\lambda_1}\frac{1}{\lambda_2} & \cdots & \cdots & \frac{1}{\lambda_1}\frac{1}{\lambda_p} \\ \frac{1}{\lambda_1}\frac{1}{\lambda_2} & \frac{1}{\lambda_2^2} & \frac{1}{\lambda_2}\frac{1}{\lambda_3} & \cdots & \frac{1}{\lambda_2}\frac{1}{\lambda_p} \\ \frac{1}{\lambda_1}\frac{1}{\lambda_3} & \frac{1}{\lambda_2}\frac{1}{\lambda_3} & \frac{1}{\lambda_3^2} & \cdots & \frac{1}{\lambda_3}\frac{1}{\lambda_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\lambda_1}\frac{1}{\lambda_p} & \frac{1}{\lambda_2}\frac{1}{\lambda_p} & \frac{1}{\lambda_3}\frac{1}{\lambda_p} & \cdots & \frac{1}{\lambda_p^2} \end{bmatrix} \mathbf{1} \tag{B.2.24}$$

$$= p + \frac{-aq + aq + a^2q^2 - a^2q^2}{1+aq} \tag{B.2.25}$$

$$= p + \frac{0}{1+aq} \tag{B.2.26}$$

$$= p. \tag{B.2.27}$$

This result proves that the classical p-variate Poisson distribution is always equi-dispersed.

Appendix C

Code

Code to generate the plots and analysis is available at *github.com/deanhansen/MSc*.

Bibliography

Arbous, A. G. and Kerrich, J. E. (1951). Accident Statistics and the Concept of Accident-Proneness. *Biometrics*, 7(4):340.

Balakrishnan, N., Ye, Y., and AL-Jarallah, R. A. (2017). Likelihood inference for Type I bivariate Pólya–Aeppli distribution. *Communications in Statistics - Simulation and Computation*, 46(8):6436–6453.

Bates, G. and Neyman, J. (1952). *Contributions to the Theory of Accident Proneness*. University of California publications in statistics. University of Calif. Press, 1st edition.

Bermúdez i Morata, L. (2009). A priori ratemaking using bivariate Poisson regression models. *Insurance: Mathematics and Economics*, 44(1):135–141.

Casella, G. and Berger, R. L. (2002). *Statistical inference*. Thomson Learning, Australia; Pacific Grove, CA, 2nd edition.

Cho, H., Liu, C., Park, J., and Wu, D. (2022). bzinb: Bivariate Zero-Inflated Negative Binomial Model Estimator.

Cho, H., Liu, C., Preisser, J. S., and Wu, D. (2023). A bivariate zero-inflated negative binomial model and its applications to biomedical settings. *Statistical Methods in Medical Research*, 32(7):1300–1317.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, 1st edition.

Edwards, C. B. and Gurland, J. (1961). A Class of Distributions Applicable to Accidents. *Journal of the American Statistical Association*, 56(295):503–517.

FBref (2024). Football Statistics and History.

Fisher, R. A. (1934). THE EFFECT OF METHODS OF ASCERTAINMENT UPON THE ESTIMATION OF FREQUENCIES. *Annals of Eugenics*, 6(1):13–25.

Greenwood, M. and Yule, G. U. (1920). An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society*, 83(2):255.

Hilbe, J. M. (2014). *Modeling Count Data*. Cambridge University Press, 1st edition.

Holgate, P. (1964). Estimation for the Bivariate Poisson Distribution. *Biometrika*, 51(1/2):241.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press, 2nd edition.

IFFHS (2023). IFFHS All Time Ranking.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley, 3rd edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley Series in Probability and Statistics. Wiley, New York, 1st edition.

Kocherlakota, S. and Kocherlakota, K. (2017). *Bivariate Discrete Distributions*. CRC Press, 1st edition.

Kokonendji, C. C. and Puig, P. (2018). Fisher dispersion index for multivariate count distributions: A review and a new proposal. *Journal of Multivariate Analysis*, 165:180–193.

McLachlan, G. J. and Krishnan, T. (2008). *The EM algorithm and extensions*. Wiley series in probability and statistics. Wiley, Hoboken, 2nd edition.

Minkova, L. D. and Balakrishnan, N. (2014a). On a Bivariate Pólya-Aeppli Distribution. *Communications in Statistics - Theory and Methods*, 43(23):5026–5038.

Minkova, L. D. and Balakrishnan, N. (2014b). Type II bivariate Pólya–Aeppli distribution. *Statistics & Probability Letters*, 88:40–49.

Oppenheim, A. (1930). Inequalities Connected with Definite Hermitian Forms. *Journal of the London Mathematical Society*, s1-5(2):114–119.

Pfanzagl, J. (1994). *Parametric Statistical Theory*. DE GRUYTER, 1st edition.

R Core Team (2023). R: A Language and Environment for Statistical Computing.

Schur, J. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *crll*, 1911(140):1–28.

Styan, G. P. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, 6:217–240.

Su, P. (2014). NORTARA: Generation of Multivariate Data with Arbitrary Marginals.

Subrahmaniam, K. (1966). A Test for "Intrinsic Correlation" in the Theory of Accident Proneness. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):180–189.

Teicher, H. (1954). On the multivariate poisson distribution. *Scandinavian Actuarial Journal*, 1954(1):1–9.

Tsagris, M. (2023). bivpois: Bivariate Poisson Distribution.