

Parameters Estimation of a Bivariate Generalized Poisson Distribution with Applications to Metabolic Syndrome Data

Mohamed M. Shoukri

Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada

Email: mmshoukr@uwo.ca, Shoukri.mohamed@gmail.com

How to cite this paper: Shoukri, M.M. (2024) Parameters Estimation of a Bivariate Generalized Poisson Distribution with Applications to Metabolic Syndrome Data. *Open Journal of Statistics*, 14, 467-480.
<https://doi.org/10.4236/ojs.2024.145020>

Received: September 17, 2024

Accepted: October 25, 2024

Published: October 28, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: Bivariate count data are commonly encountered in medicine, biology, engineering, epidemiology and many other applications. The Poisson distribution has been the model of choice to analyze such data. In most cases mutual independence among the variables is assumed, however this fails to take into accounts the correlation between the outcomes of interests. A special bivariate form of the multivariate Lagrange family of distribution, names Generalized Bivariate Poisson Distribution, is considered in this paper. **Objectives:** We estimate the model parameters using the method of maximum likelihood and show that the model fits the count variables representing components of metabolic syndrome in spousal pairs. We use the likelihood local score to test the significance of the correlation between the counts. We also construct confidence interval on the ratio of the two correlated Poisson means. **Methods:** Based on a random sample of pairs of count data, we show that the score test of independence is locally most powerful. We also provide a formula for sample size estimation for given level of significance and given power. The confidence intervals on the ratio of correlated Poisson means are constructed using the delta method, the Fieller's theorem, and the nonparametric bootstrap. We illustrate the methodologies on metabolic syndrome data collected from 4000 spousal pairs. **Results:** The bivariate Poisson model fitted the metabolic syndrome data quite satisfactorily. Moreover, the three methods of confidence interval estimation were almost identical, meaning that they have the same interval width.

Keywords

Lagrange Distributions, Double Poisson, Maximum Likelihood Estimation, Score Test of Independence, Higher Order Moments, Non-Parametric

1. Introduction

For many years there has been an increase in the statistical research to develop multivariate distributions for count data. There are numerous ways by which we can generalize the distributions of univariate counts to multivariate forms. The applications of these multivariate distributions are found in many fields such as engineering, bioinformatics, genetic epidemiology and biomedicine. An interested reader is referred to the recent editions of the book by Winkelmann [1] and the book by Cameron and Trivedi [2] for thorough reviews of regression models for multivariate count data and the very recent work by Tzougas *et al.* [3]. Needless to say, fitting these multivariate distribution to data is quite complicated and requires a great deal of computing resources. Therefore there is a need to develop more flexible bivariate models and demonstrate their applicability to real life data.

The main purpose of this article is to investigate the statistical characteristics of a class of bivariate generalized Poisson models with a correlation parameter that offers sufficient flexibility for accommodating overdispersion and accounting for the positive correlation between the variables of interest. The distribution is named “Bivariate Generalized Poisson Distribution”, BGPLD, which is a special member of a larger family of distributions developed by Shenton and Consul [4]. Some of the interesting properties of this class were studied further by Jain and Sing [5], and by Shoukri [6] who investigated some of the properties of the distribution as a member of the class of bivariate modified power series distributions.

This paper has three-fold objectives. Firstly, we derive the locally powerful score test on the hypothesis that the correlation parameter is significantly different from zero, and then assess the power of this test. Secondly, construct confidence interval on the ratio of the means of the two correlated Poisson variables. Thirdly, we fit the model to metabolic syndrome data available from a random sample of spousal pairs.

2. The Bivariate Generalized Poisson Distribution (BGPLD)

A bivariate random variable (X, Y) is said to have BGPLD when the joint probability distribution is given by (1):

$$P(X = x, Y = y) = \frac{(1 + m_1x + m_2y)^{x+y-1}}{x!y!} \theta_1^x \theta_2^y \exp[-(\theta_1 + \theta_2) \cdot (1 + m_1x + m_2y)] \quad (1)$$

$x, y \geq 0$, $0 < \theta_i < \infty$, $m_i > 0 (i = 1, 2)$, $0 < \theta_1 m_1 + \theta_2 m_2 < 1$, and zero otherwise. The BGPLD can be obtained as follows:

By expanding the bivariate probability generating function (BPGF)

$\phi(t_1, t_2) = \exp[\theta_1(t_1 - 1) + \theta_2(t_2 - 1)]$, using the bivariate Lagrange expansion of an implicit function, as was shown in [4].

3. Recurrence Relation among Moments

From (1) we can write

$$\exp(\theta_1 + \theta_2) = \sum_{x,y} (x!y!)^{-1} (1 + m_1x + m_2y)^{x+y-1} \cdot \theta_1^x \theta_2^y \exp[-(\theta_1 + \theta_2)(1 + m_1x + m_2y)] \quad (2)$$

Therefore, from (2) we can see that the distribution possesses a power series expansion in terms of the parameters $(\theta_1 \ \& \ \theta_2)$.

On differentiating the above equation {2} partially *w.r.t.* θ_1 and θ_2 respectively, dividing by $\exp(\theta_1 + \theta_2)$, and on summation, we get by solving for $\mu_{1,0}$ and $\mu_{0,1}$.

$$\mu_{1,0} = \theta_1 (1 - m_1\theta_1 - m_2\theta_2)^{-1} \quad (3)$$

$$\mu_{0,1} = \theta_2 (1 - m_1\theta_1 - m_2\theta_2)^{-1} \quad (4)$$

To obtain a recurrence relation among the higher non-central product moments, we write

$$\mu_{r,s} = \sum_{x,y} x^r y^s P(X = x, Y = y) \quad (5)$$

Define the cross product moments:

$$\mu_{r+1,s} = E[(x - \mu'_{10})^{r+1} (y - \mu'_{01})^s].$$

We can show that

$$\mu_{r+1,s} = \theta_1 (1 - m_1\theta_1 - m_2\theta_2)^{-1} \left[\{(1 - m_2\theta_2)\partial_1 + m_2\theta_2\partial_2\} \mu'_{r,s} \right] + \mu'_{r,s} \mu'_{1,0} \quad (6)$$

By symmetry we can write:

$$\mu'_{r,s+1} = \theta_2 (1 - m_1\theta_1 - m_2\theta_2)^{-1} \left[\{(1 - m_1\theta_1)\partial_2 + m_1\theta_1\partial_1\} \mu'_{r,s} \right] + \mu'_{r,s} \mu'_{0,1} \quad (7)$$

These relationships among the moments are obtained on following Kendall [7].

In particular, from (6) and (7):

$$\mu_{1,0} = \theta_1 (1 - m_1\theta_1 - m_2\theta_2)^{-1} \quad (8)$$

$$\mu_{2,0} = \theta_1 (1 - m_1\theta_1 - m_2\theta_2)^{-3} \left[(1 - m_2\theta_2)^2 + \theta_1\theta_2 m_2^2 \right] \quad (9)$$

are respectively the mean and variance of X in terms of the model parameters.

And the values of $\mu_{0,1}$ and $\mu_{0,2}$ can be written down by symmetry. The coefficient of correlation ρ , between X and Y is given by

$$\rho = \frac{\theta_1\theta_2 [m_1(1 - m_2\theta_2) + m_2(1 - m_1\theta_1)]}{\left[\theta_1\theta_2 (1 - 2m_1\theta_1 + m_1^2\theta_1^2 + \theta_1\theta_2 m_1^2) (1 - 2m_2\theta_2 + m_2^2\theta_2^2 + \theta_1\theta_2 m_2^2) \right]^{1/2}} \quad (10)$$

As can be seen, from the restrictions on the model parameters, the correlation (10) cannot be negative.

3.1. Maximum Likelihood Estimators

Assuming m_1 and m_2 to be known constants, let $(X_i, Y_i), i = 1, 2, \dots, N$ be a random sample of fixed size n , taken from the BGPD family given in (1). It can be easily shown that the maximum likelihood (ML) estimators for θ_1 and θ_2 are as shown in (11) and (12):

$$\hat{\theta}_1 = z_1 (n + m_1 z_1 + m_2 z_2)^{-1} \tag{11}$$

$$\hat{\theta}_2 = z_2 (n + m_1 z_1 + m_2 z_2)^{-1} \tag{12}$$

With some considerable algebra, converting the Fisher's information matrix we can show that the asymptotic variances and covariances of the ML estimators of θ_1 and θ_2 are given respectively in (13)-(15):

$$\text{Var}(\hat{\theta}_1) = \frac{n\theta_1}{(n + m_1)^2} - \frac{n^2 m_1 \theta_1^2}{(n + m_1)^2 (n + 2m_1)} - \frac{nm_2 \theta_1 \theta_2}{(n + m_1)(n + m_1 + m_2)} \tag{13}$$

$$\text{Var}(\hat{\theta}_2) = \frac{n\theta_2}{(n + m_2)^2} - \frac{n^2 m_2 \theta_2^2}{(n + m_2)^2 (n + 2m_2)} - \frac{nm_1 \theta_1 \theta_2}{(n + m_2)(n + m_1 + m_2)} \tag{14}$$

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = \frac{n\theta_1 \theta_2}{n + m_1 + m_2} - \frac{n^2 \theta_1 \theta_2}{(n + m_2)(n + m_2)} \tag{15}$$

3.2. Properties of the BGPD

1: If $(X, Y) \sim BGPD$, then X and Y are stochastically independent if and only if $m_1 = m_2 = 0$.

2: If $(X, Y) \sim BGPD$, then X and Y cannot be perfectly correlated.

Proof. The quality of ρ^2 , given in (10), to unity, will imply that $(1 - m_1 \theta_1 - m_2 \theta_2)^2 = 0$. This is not true unless $m_1 \theta_1 + m_2 \theta_2 = 1$, which is a contradiction to the condition enforced by the strict inequality $0 < m_1 \theta_1 + m_2 \theta_2 < 1$.

3: If $(X_i, Y_i), i = 1, 2, \dots, n$ is a random sample taken from the BGPD family, the joint probability distribution of the sums $Z_1 = X_1 + X_2 + \dots + X_n$ and $Z_2 = Y_1 + Y_2 + \dots + Y_n$ is also a GDPD and its probability function takes the form

$$P(Z_1 = z_1, Z_2 = z_2) = \frac{n(n + m_1 z_1 + m_2 z_2)^{z_1 + z_2 - 1}}{z_1! z_2!} \theta_1^{z_1} \theta_2^{z_2} \exp[1(\theta_1 + \theta_2)(n + m_1 z_1 + m_2 z_2)] \tag{17}$$

4: The marginal distribution of X is the generalized Poisson distribution given by Consul and Jain [8], where their λ_1 and λ_2 are taken as θ_1 and $m_1 \theta_1$ respectively. Thus, the marginal probability distribution of X and Y are given respectively in (18) and (19):

$$P(X = x) = (1 + m_1 x)^{x-1} (x!)^{-1} (\theta_1 e^{-\theta_1})^x / e^{\theta_1} \tag{18}$$

Similarly

$$P(Y = y) = (1 + m_1 y)^{y-1} (y!)^{-1} (\theta_2 e^{-\theta_2})^y / e^{\theta_2} \tag{19}$$

5: The regression equation of X on Y is given by

$$E(Y/x) = \begin{cases} \theta_2(1-m_2\theta_2)^{-1} & x = 0 \\ \theta_2(1-m_2\theta_2)^{-1} + \theta_2m_1(1-m_2\theta_2)^{-1}x & x \geq 0 \end{cases}$$

3.3. Score Test on the Hypothesis of Independence ($H_0:m = 0$)

In the subsequent analyses we shall assume that $m_1 = m_2 = m$. Let $(x_i, y_i), i = 1, 2, \dots, n$ be an srs. The likelihood is given by:

$$L = \prod_{i=1}^n P(x_i, y_i),$$

The log-likelihood function is:

$$\begin{aligned} \ell = \log L = \sum_{i=1}^n [(x_i + y_i - 1) \log(1 + m(x_i + y_i))] \\ + n\bar{x} \log \theta_1 + n\bar{y} \log \theta_2 - (\theta_1 + \theta_2) \sum_{i=1}^n [1 + m(x_i + y_i)] \end{aligned} \quad (20)$$

Moreover denote the sample means by: $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$.

Differentiating the log-likelihood function (20) with respect to m , the score function $\left. \frac{\partial \ell}{\partial m} \right|_{m=0}$ is given by:

$$U = \left. \frac{\partial \ell}{\partial m} \right|_{m=0} = \sum_{j=1}^n (x_j + y_j)(x_j + y_j - 1) - n(\bar{x} + \bar{y})(\theta_1 + \theta_2)$$

Replacing $(\theta_1 + \theta_2)$ with their MLE under $H_0; m = 0$, we have

$$U = \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + 2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) - n(\bar{x} + \bar{y}) \quad (21)$$

Under $H_0 = m = 0$, from (21) we have:

$$E[U] = 0$$

Since

$$\begin{aligned} u &= nS_x^2 + nS_y^2 - n(\bar{x} + \bar{y}) \\ &= n(S_x^2 - \bar{x}) + n(S_y^2 - \bar{y}) \\ E[u] &= n\mu_{20} + n\mu_{02} - n(\mu'_{10} + \mu'_{01}) \end{aligned}$$

And under $H_0 = m = 0$

$$\begin{aligned} \mu_{20} &= \theta_1, \mu_{02} = \theta_2 \\ \mu'_{10} &= \theta_1, \mu'_{01} = \theta_2 \end{aligned}$$

Therefore, $E[u] = 0$, and the variance of u under the null hypothesis is:

$$\begin{aligned} \text{Var}(u) = V_0 &= n^2 \left[(S^2x - \mu_{20})^2 \right] + n^2 E \left[(S^2y - \mu_{02})^2 \right] + 2ncov(\bar{x}, S^2x) \\ \text{Var}(u) &= 2n(\theta_1^2 + \theta_2^2) \end{aligned}$$

Hence

$$\begin{aligned}
 z' &= \frac{u^2}{\text{var}(u)} \sim \chi_{(1)}^2 \\
 &= n^2 \left[S^2x + S^2y - \bar{x} - \bar{y} \right] / 2n(\theta_1^2 + \theta_2^2) \\
 &= \frac{n \left[(S^2x - \bar{x}) + (S^2y - \bar{y}) \right] 2}{2(\bar{x}^2 + \bar{y}^2)}
 \end{aligned}
 \tag{21}$$

as $n \rightarrow \infty$ the distribution of the test statistic (21) follows Chi-square distribution with one degree of freedom.

To evaluate the asymptotic power of the score test we need to find the non-null asymptotic distribution of u .

When the null hypothesis does not hold, the expected value of, denoted by $M_1 = E[u | m \neq 0]$, is such that:

$$\frac{M_1}{n} = (\mu_{20} + \mu_{02}) - (\mu'_0 + \mu'_{01})
 \tag{22}$$

Moreover, denoting the variance of the test statistics u under the non-null distribution by

$$V_1 = \text{var}(u | m \neq 0), \text{ then}$$

$$V_1 = nD, \text{ where,}$$

$$\begin{aligned}
 D &= (\mu_{40} - \mu_{20}^2) + (\mu_{20} - 2\mu_{30}) + (\mu_{04} - \mu_{02}^2) + (\mu_{02} - 2\mu_{03}) \\
 &\quad + 2[\mu_{22} - \mu_{12} - \mu_{21} - \mu_{11}]
 \end{aligned}
 \tag{23}$$

An expression for the estimated sample size needed to verify the hypothesis of independence for given type I error rate and power $1-\beta$ is shown to be:

$$n = \frac{\left[Z_\beta \sqrt{D} + Z_\alpha \sqrt{2(\theta_1^2 + \theta_2^2)} \right]^2}{M_1^2}
 \tag{24}$$

For example, in (24) given 5% type error rate $Z_\alpha = 1.64$, and power 80%, $Z_\beta = 0.84$.

In the Appendix we give the algebraic expressions of the elements of D as functions of the population parameters.

3.4. Confidence Interval on the Ratio of BGPD Means

1) Delta Method

Let $R = \mu'_{10} / \mu'_{01}$ denote the ratio of the mean of X to the mean of Y , The point estimator $\hat{R} = \bar{x} / \bar{y}$. From Kendall [7], we have to the first order of approximation:

$$\text{Var}(\hat{R}) = n^{-1} \left[\mu_{20} / (\mu'_{01})^2 + \mu_{02} (\mu'_{10} / \mu'_{01})^2 + 2\mu_{11} (\mu'_{10} / \mu'_{01}) \right]
 \tag{25}$$

The $(1 - \alpha)$ 100% confidence interval on R is therefore given by:

$$\hat{R} \pm Z_{\alpha/2} \text{Var}(\hat{R})^{1/2}
 \tag{26}$$

2) Feiller's Interval

Let $\delta = \bar{x} - R\bar{y}$

$$\text{Var}(\delta) = \frac{1}{n} [\mu_{20} - 2R\mu_{11} + R^2\mu_{02}]$$

Hence, the

$$1 - \alpha = P_r \left[\frac{\delta}{\sqrt{V(\delta)}} < Z_{\alpha/2} \right]$$

$$1 - \alpha = P_r [\alpha^2 \leq V(\alpha) Z_{\alpha/2}^2] = P_r [(\bar{x} - R\bar{y})^2 \leq V(\alpha) Z_{\alpha/2}^2]$$

A point estimator of $\text{Var}(\delta)$ is given by:

$$\hat{V} = n^{-1} [S_x^2 - 2R\hat{\rho}S_xS_y + R^2S_y^2]$$

Therefore,

$$1 - \alpha = P_r [aR^2 - Rb + c]$$

where

$$a = \bar{y}^2 - Z_{\alpha/2}^2 S_y^2 / n$$

$$b = 2(\bar{x}\bar{y} + Z_{\alpha/2}^2 \hat{\rho} S_x S_y / n)$$

and

$$c = \bar{x}^2 - Z_{\alpha/2}^2 S_x^2 / n$$

Solving the quadratic in R we get

$$R \in (A, B), \text{ where}$$

$$A \equiv \frac{1}{2a} [-b + \sqrt{b^2 - 4ac}] \quad (27)$$

$$B \equiv \frac{1}{2a} [-b - \sqrt{b^2 - 4ac}] \quad (28)$$

In the data analysis section, we shall compare the above intervals to the non-parametric bootstrap confidence interval.

4. Application: Modeling Metabolic Syndrome in Spousal Pairs Using GDPD

The metabolic syndrome is the co-aggregation of hypertension, impaired glucose tolerance, dyslipidemia, and abdominal obesity and is associated with an increased risk of total and cardiovascular mortality in adults [9] [10]. Genetics as well as environmental influences have been implicated in obesity and several cardiovascular risk factors [11] [12]. Family is one of the most important factors affecting metabolic risk factors in children, in that family displays an interaction between genetic and shared environmental factors [13] [14]. Recent research showed that childhood and adolescent overweight has been increasing in Asian countries due to urbanization and economic development. For example, over the

past 10 years, the rates of overweight among Korean children and adolescents aged 5 - 20 years have doubled [15], which may ultimately cause an increase in adverse cardiovascular outcomes. Globally, the prevalence of the metabolic syndrome is high among obese children and adolescents, and increases with increasing obesity [16]. The widely used definition of metabolic syndrome is that of the World Health Organization [17] [18]. The components of each definition and criteria for making the diagnosis of the metabolic syndrome are summarized in **Table 1**.

The abbreviations of the medical official bodies given in **Table 1** are:

IDF \equiv International Diabetes Federation, WHO \equiv World Health Organization, EGIR \equiv European Group for the Study of Insulin Resistance, and NCEP \equiv National Cholesterol Education Program.

Table 1. Definition of the components of metabolic syndrome [Journal of the Royal Society of Medicine Vol. 99 September 2006].

Component	IDF		WHO		EGIR		NCEP	
	M	F	M	F	M	F	M	F
Central Obesity	≥ 102	≥ 88	≥ 102	≥ 88	≥ 94	≥ 80	≥ 102	≥ 88
Raised TG	≥ 1.7	≥ 1.7	≥ 1.7	≥ 1.7	≥ 2.0	≥ 2.0	≥ 1.7	≥ 1.7
Low HDL	< 1.03	< 1.29	≤ 9	≤ 1	≤ 1	≤ 1	≤ 1.03	≤ 1.29
Hypertension	$\geq 130/85$	$\geq 130/85$	$\geq 140/90$	$\geq 140/90$	$\geq 140/90$	$\geq 140/90$	$\geq 130/85$	$\geq 130/85$
Fasting Glucose	≥ 5.6	≥ 5.6	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1	≥ 6.1

The Metabolic syndrome data:

Based on a random sample of 4000 spousal pairs, the female and male counts are cross classified in **Table 2** and **Table 3**.

The test independence had a value of $U = 7.027$, with a corresponding p -value = $2.11e-12$.

Therefore the hypothesis of independence is not supported by the metabolic

Table 2. Cross classification of the counts of components of MS in spousal pairs.

		Female count					Total
		0.00	1.00	2.00	3.00	4.00	
Male count	0.00	2337	0	0	0	0	2337
	1.00	1228	0	0	0	0	1228
	2.00	0	360	0	0	0	360
	3.00	0	0	44	12	0	56
	4.00	0	0	0	7	12	19
Total		3565	360	44	19	12	4000

Table 3. Summary statistics of counts.

	N	Mean	Std. Deviation	Variance
Male count	4000	0.5480	0.75421	0.569
Female count	4000	0.1382	0.45353	0.206

syndrome data.

To fit the model to the data we shall use the method of moments to estimate the three population parameters. We shall use the notations μ_{10}^* & μ_{01}^* to respectively denotes the sample means of x & y , while μ_{11}^* denotes the sample covariance between x & y .

$$\theta_1^* = \frac{\mu_{10}^*}{1 + m^* (\mu_{10}^* + \mu_{01}^*)} \quad (29)$$

$$\theta_2^* = \frac{\mu_{01}^*}{1 + m^* (\mu_{10}^* + \mu_{01}^*)} \quad (30)$$

$$m^{*2} (\mu_{10}^* + \mu_{01}^*) \mu_{10}^* \mu_{01}^* + 2m^* \mu_{10}^* \mu_{01}^* - \mu_{11}^* = 0 \quad (31)$$

From Equations (29)-(31), the admissible moment estimator for m is.

$$m^* = \frac{\left[(\mu_{10}^* \mu_{01}^*)^2 + (\mu_{10}^* + \mu_{01}^*) \mu_{11}^* \mu_{10}^* \mu_{01}^* \right]^{1/2} - \mu_{10}^* \mu_{01}^*}{(\mu_{10}^* + \mu_{01}^*) \mu_{10}^* \mu_{01}^*} \quad (32)$$

The moments estimators of θ_1 & θ_2 are given respectively by θ_1^* & θ_2^* in Equations (33) and (34).

$$\theta_1^* = \frac{\mu_{10}^{*2} \mu_{01}^*}{\left[(\mu_{10}^* \mu_{01}^*)^2 + (\mu_{10}^* + \mu_{01}^*) \mu_{11}^* \mu_{10}^* \mu_{01}^* \right]^{1/2}} \quad (33)$$

$$\theta_2^* = \frac{\mu_{01}^{*2} \mu_{10}^*}{\left[(\mu_{10}^* \mu_{01}^*)^2 + (\mu_{10}^* + \mu_{01}^*) \mu_{11}^* \mu_{10}^* \mu_{01}^* \right]^{1/2}} \quad (34)$$

From the metabolic syndrome data we have, for females, $\theta_1^* = 0.294$, and for male, $\theta_2^* = 0.074$, and $m^* = 1.262$. The familiar chi-square goodness of fit, with 4 degrees of freedom had a p-value = 0.141. This shows that the model gives a reasonable fit to the data.

The following is a short R code from which we evaluate the nonparametric confidence interval and assess the large sample distribution of the ratio of correlated Poisson means.

```
R-CODE for bootstrapping the confidence interval on the ratio of means
x<-data$Male_count
y<-data$Female_count
df<-data.frame(x,y)
library(boot)
boot_df=boot(df,function(i,d)mean(i[d] [1])/mean(i[d] [2]),R=999)
```

boot_df

Bootstrap Statistics :

	original	bias	std. error
ratio	3.964	0.0163	0.148

It is clear from **Figure 1** that the quantile plot of the sample ratio of means follows a normal distribution (**Table 4**).

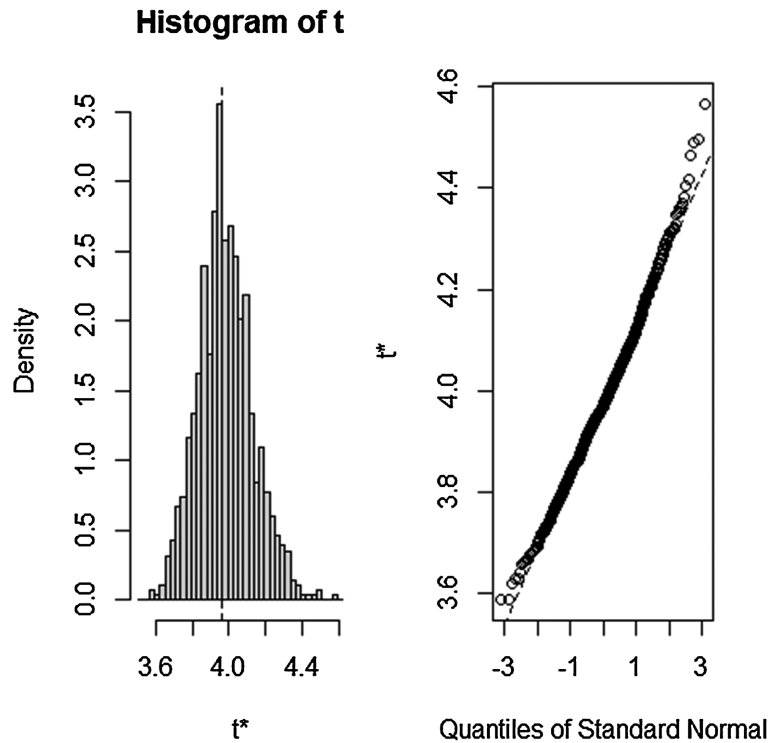


Figure 1. Histogram and normal quantile plot for 1000 bootstrap replicates of the estimated ratio of means.

Table 4. The upper and lower limits of the 95% confidence interval on the ratio of means for the three methods (delta, Feiller’s, and bootstrap).

Limits	Delta method	Feiller’s	Bootstrap
Lower limit	3.417	3.467	3.674
Upper limit	4.514	4.575	5.254

5. Concluding Remarks

In this article, we discussed some of the interesting applications of a bivariate generalized Poisson distribution. The model has three parameters and has provided satisfactory fit to the metabolic syndrome data collected from spousal pairs. The common parameter (m) captures overdispersion and accurately accounts for the strength of the positive correlation between the two variables.

It is worth noting that this family of models, which is a member of the Lagrange family of bivariate distributions, is suitable for many applications, and in particular

in situation where queuing theory is applicable. For example, in large tertiary hospital outpatients departments, the patients usually wait for admission form queues. If we have two types of queues (e.g. males and females), then the random variables would represent the numbers of admitted patients during a busy period of the admission office (single server queue). We also note that the marginal probability distributions have been applied to a variety of data from genomics, and case fatality due to COVID [19].

Finally, it is worth noting that with the not so complicated structure of the model, it can be shown that it belongs to the class of bivariate of generalized linear models where inclusion of covariates will make this model attractive alternative in many applications.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Winkelmann, R. (2008) *Econometric Analysis of Count Data*. Springer Science & Business Media.
- [2] Cameron, A.C. and Trivedi, P.K. (2013) *Regression Analysis of Count Data*. 2nd Edition, Cambridge University Press. <https://doi.org/10.1017/cbo9781139013567>
- [3] Tzougas, G. and di Cerchiaro, A.P. (2021) Bivariate Mixed Poisson Regression Models with Varying Dispersion. *North American Actuarial Journal*, **27**, 211-241. <https://doi.org/10.1080/10920277.2021.1978850>
- [4] Shenton, L.R. and Consul, P.C. (1973) On Bivariate Lagrange and Borel-Tanner Distributions and Their Use in Queuing Theory. *Sankhya*, **35**, 229-236.
- [5] Jain, G.C. and Singh, N. (1975) On Bivariate Power Series Distributions Associated with Lagrange Expansion. *Journal of the American Statistical Association*, **70**, 951-954. <https://doi.org/10.1080/01621459.1975.10480329>
- [6] Shoukri, M.M. (1982) On the Generalization and Estimation for the Double Poisson Distribution. *Trabajos de Estadística y de Investigación Operativa*, **33**, 97-109. <https://doi.org/10.1007/bf02888625>
- [7] Kendall, M. and Ord, K. (1987) *Advanced Theory of Statistics*. 5th Edition, Griffen and Company.
- [8] Consul, P.C. and Jain, G.C. (1973) A Generalization of the Poisson Distribution. *Technometrics*, **15**, 791-799. <https://doi.org/10.1080/00401706.1973.10489112>
- [9] Laaksonen, D.E. (2002) Metabolic Syndrome and Development of Diabetes Mellitus: Application and Validation of Recently Suggested Definitions of the Metabolic Syndrome in a Prospective Cohort Study. *American Journal of Epidemiology*, **156**, 1070-1077. <https://doi.org/10.1093/aje/kwf145>
- [10] Lakka, H. (2002) The Metabolic Syndrome and Total and Cardiovascular Disease Mortality in Middle-Aged Men. *Journal of the American Medical Association*, **288**, 2709-2716. <https://doi.org/10.1001/jama.288.21.2709>
- [11] Carmelli, D., Cardon, L.R. and Fabsitz, R. (1994) Clustering of Hypertension, Diabetes, and Obesity in Adult Male Twins: Same Genes or Same Environments? *American Journal of Human Genetics*, **55**, 566-573.
- [12] Edwards, K.L., Newman, B., Mayer, E., Selby, J.V., Krauss, R.M. and Austin, M.A.

- (1997) Heritability of Factors of the Insulin Resistance Syndrome in Women Twins. *Genetic Epidemiology*, **14**, 241-253. [https://doi.org/10.1002/\(sici\)1098-2272\(1997\)14:3<241::aid-gepi3>3.0.co;2-8](https://doi.org/10.1002/(sici)1098-2272(1997)14:3<241::aid-gepi3>3.0.co;2-8)
- [13] Hong, Y., Pederson, N.L., *et al.* (1997) Genetic and Environmental Architecture of the Features of Insulin Resistance Syndrome. *American Journal of Human Genetics*, **60**, 143-152.
- [14] Lee, K.E., Klein, B.E. and Klein, R. (2003) Familial Aggregation of Components of the Multiple Metabolic Syndrome in the Framingham Heart and Offspring Cohorts: Genetic Analysis Workshop Problem. *BMC Genetics*, **4**, Article No. S94. <https://doi.org/10.1186/1471-2156-4-s1-s94>
- [15] Park, H.S., Park, J.Y. and Cho, S. (2006) Familial Aggregation of the Metabolic Syndrome in Korean Families with Adolescents. *Atherosclerosis*, **186**, 215-221. <https://doi.org/10.1016/j.atherosclerosis.2005.07.019>
- [16] Borch-Johnson, K. (2007) The Metabolic Syndrome in a Global Perspective. *Danish Medical Bulletin*, **54**, 157-159.
- [17] Cameron, A.J., Shaw, J.E. and Zimmet, P.Z. (2004) The Metabolic Syndrome: Prevalence in Worldwide Populations. *Endocrinology and Metabolism Clinics of North America*, **33**, 351-375. <https://doi.org/10.1016/j.ecl.2004.03.005>
- [18] World Health Organization (1999) Definition, Diagnosis, and Classification of Diabetes Mellitus and Its Complications. Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus.
- [19] Al-Gahtani, S., Shoukri, M. and Al-Eid, M. (2021) Predictors of the Aggregate of COVID-19 Cases and Its Case-Fatality: A Global Investigation Involving 120 Countries. *Open Journal of Statistics*, **11**, 259-277. <https://doi.org/10.4236/ojs.2021.112014>

Appendix: Higher Order Cross-Product Moments

On using the recurrence relationship in (6) we can show that: Denote

$$\begin{aligned}\Delta &= (1 - m\theta_1 - m\theta_2) \\ \mu_{11} &= \frac{m\theta_1\theta_2(1 - m\theta_1 - m\theta_2)}{(1 - m\theta_1 - m\theta_2)^3} \\ \mu_{12} &= \frac{1}{\Delta^4} \left[m\theta_1\theta_2^2(1 - m\theta_2) - m\theta_1\theta_2(1 - m\theta_1)(1 - m\theta_2) \right. \\ &\quad \left. + m\theta_1\theta_2(1 - m\theta_1)^2 + 2m^3\theta_1^2\theta_2^2 \right] \\ &\quad + \frac{1}{\Delta^5} \left[3m^4\theta_1^2\theta_2^3 + 3m^2\theta_1\theta_2^2(1 - m\theta_1)^2 + 3m^3\theta_1^2\theta_2^2(1 - m\theta_2) \right. \\ &\quad \left. + 3m\theta_1\theta_2(1 - m\theta_1)^2(1 - m\theta_2) \right] \\ \mu_{22} &= E \left[(x - \mu'_{10})^2 (y - \mu'_{01})^2 \right] \\ &= \frac{\theta_1}{1 - m\theta_1 - m\theta_2} \left\{ \left[(1 - m\theta_2) \frac{\partial \mu_{12}}{\partial \theta_1} + m\theta_2 \frac{\partial \mu_{12}}{\partial \theta_2} \right] \right. \\ &\quad \left. + \mu_{02} \left[\frac{(1 - m\theta_2)^2}{(1 - m\theta_1 - m\theta_2)^2} + \frac{m^2\theta_1\theta_2}{(1 - m\theta_1 - m\theta_2)^2} \right] \right. \\ &\quad \left. + 2\mu_{11} \left[\frac{m\theta_2(1 - m\theta_2)}{(1 - m\theta_1 - m\theta_2)^2} + \frac{m\theta_2(1 - m\theta_1)}{(1 - m\theta_1 - m\theta_2)^2} \right] \right\}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \mu_{12}}{\partial \theta_1} &= \frac{1}{\Delta^4} \left[m\theta_2^2(1 - m\theta_2) - m\theta_2(1 - m\theta_2)(1 - 2m\theta_1) \right. \\ &\quad \left. + m\theta_2 \left[(1 - m\theta_1)^2 - 2m^2\theta_1\theta_2(1 - m\theta_1) + 4m^3\theta_1^2\theta_2^2 \right] \right. \\ &\quad \left. + \frac{1}{\Delta^5} \left[4m^2\theta_1^2\theta_2(1 - m\theta_2) - 4m^2\theta_1\theta_2(1 - m\theta_1) \times (1 - m\theta_2) \right. \right. \\ &\quad \left. \left. + 4m^2\theta_1\theta_2(1 - m\theta_1)^2 + 8m^4\theta_1^2\theta_2^2 + 6m^4\theta_1\theta_2^3 + 3m^2\theta_2^2(1 - m\theta_1)^2 \right. \right. \\ &\quad \left. \left. - 6m^3\theta_1\theta_2^2(1 - m\theta_1) + 6m^3\theta_1\theta_2^2(1 - m\theta_2) + 3m\theta_2(1 - m\theta_1)^2(1 - m\theta_2) \right. \right. \\ &\quad \left. \left. - 6m^2\theta_1\theta_2(1 - m\theta_1)(1 - m\theta_2) \right] + \frac{5m}{\Delta^6} \left[3m^4\theta_1^2\theta_2^3 + 3m^2\theta_1\theta_2^2(1 - m\theta_1)^2 \right. \right. \\ &\quad \left. \left. + 3m^3\theta_1^2\theta_2^2(1 - m\theta_2) + 3m\theta_1\theta_2(1 - m\theta_1)^2(1 - m\theta_2) \right] \right]\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \mu_{12}}{\partial \theta_2} &= \frac{1}{\Delta^4} \left[2m\theta_1\theta_2(1 - m\theta_2) - m^2\theta_1\theta_2^2 - m\theta_1(1 - m\theta_1)(1 - m\theta_2) \right. \\ &\quad \left. + m^2\theta_1\theta_2(1 - m\theta_1) + m\theta_1(1 - m\theta_1)^2 + 4m^3\theta_1^2\theta_2 + 4m^3\theta_1\theta_2^2 \right] \\ &\quad + \frac{1}{\Delta^5} \left[6m^4\theta_1\theta_2^3 + 3m^2\theta_2^2(1 - m\theta_1)^2 - 6m^3\theta_2^2\theta_1(1 - m\theta_1) \right. \\ &\quad \left. + 6m^3\theta_1\theta_2^2(1 - m\theta_2) + 3m\theta_2(1 - m\theta_1)^2(1 - m\theta_2) \right]\end{aligned}$$

$$\begin{aligned}
 & -6m^2\theta_1\theta_2(1-m\theta_1)(1-m\theta_2)] + \frac{5m}{\Delta^6} [3m^4\theta_1^3\theta_2^3 + 3m^2\theta_1\theta_2^2(1-m\theta_1)^2 \\
 & + 3m^2\theta_1^2\theta_2^2(1-m\theta_2) + 3m\theta_1\theta_2(1-m\theta_1)^2(1-m\theta_2)] \\
 \mu_{30} &= \frac{m^3\theta_1^2\theta_2 + \theta_1(1-m\theta_2)^3}{\Delta^4} + \frac{3m^3\theta_1^3\theta_2 + 3m\theta_1^2(1-m\theta_2)^2}{\Delta^5}
 \end{aligned}$$

And from the recurrence relation, we have:

$$\mu_{40} = \frac{\theta_1(1-m\theta_2)}{\Delta} \cdot \frac{\partial\mu_{30}}{\partial\theta_1} + \frac{m\theta_1\theta_2}{\Delta} \cdot \frac{\partial\mu_{30}}{\partial\theta_2} + 3\mu_{20} \left[\frac{(1-m\theta_2)^2}{\Delta^2} + \frac{m^2\theta_1\theta_2}{\Delta^2} \right]$$

The μ_{40} depends on the:

$$\frac{\partial\mu_{30}}{\partial\theta_1} \quad \& \quad \frac{\partial\mu_{30}}{\partial\theta_2}$$

We can show that:

$$\begin{aligned}
 \frac{\partial\mu_{30}}{\partial\theta_1} &= \frac{1}{\Delta^4} [2m^2\theta_1\theta_2 + (1-m\theta_2)^3] \\
 &+ \frac{1}{\Delta^5} [13m^3\theta_1^2\theta_2 + 6m\theta_1(1-m\theta_2)^2 + 4m\theta_1(1-m\theta_2)^3] \\
 &+ \frac{1}{\Delta^6} [15m^4\theta_1^3\theta_2 + 15m^2\theta_1^2(1-m\theta_2)^2] \\
 \frac{\partial\mu_{30}}{\partial\theta_2} &= \frac{1}{\Delta^4} [m^2\theta_1^2 - 3m\theta_1(1-m\theta_2)^2] \\
 &+ \frac{1}{\Delta^5} [3m^3\theta_1^3 - 6m^2\theta_1^2(1-m\theta_2) + 4m^3\theta_1^2\theta_2 + 4m\theta_1(1-m\theta_2)^3] \\
 &+ \frac{1}{\Delta^6} [15m^4\theta_1^3\theta_2 + 15m^2\theta_1^2(1-m\theta_2)^2]
 \end{aligned}$$

It should also be noted that μ_{03} and μ_{04} can be obtained on exchanging θ_1 by θ_2 in μ_{30} and μ_{40} and visa-versa.