mathematics

Article

Bivariate Continuous Negatively Correlated Proportional Models with Applications in Schizophrenia Research

Yuan Sun [1,†], Guoliang Tian [2,†], Shuixia Guo [3,4], Lianjie Shu [5] and Chi Zhang [6,*]

1 Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China; 11849457@mail.sustech.edu.cn
2 Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, China; tiangl@sustech.edu.cn
3 MOE-LCSM, School of Mathematics and Statistics, Hunan Normal University, Changsha 410081, China; guoshuixia75@163.com
4 Key Laboratory of Applied Statistics and Data Science, Hunan Normal University, Changsha 410081, China
5 Faculty of Business, University of Macau, Macau, China; ljshu@um.edu.mo
6 College of Economics, Shenzhen University, Shenzhen 518055, China
* Correspondence: chizhang@szu.edu.cn
† These authors contributed equally to this work.

Abstract: Bivariate continuous negatively correlated proportional data defined in the unit square $(0,1)^2$ often appear in many different disciplines, such as medical studies, clinical trials and so on. To model this type of data, the paper proposes two new bivariate continuous distributions (i.e., *negatively correlated proportional inverse Gaussian* (NPIG) and *negatively correlated proportional gamma* (NPGA) distributions) for the first time and provides corresponding distributional properties. Two mean regression models are further developed for data with covariates. The *normalized expectation–maximization* (N-EM) algorithm and the gradient descent algorithm are combined to obtain the maximum likelihood estimates of parameters of interest. Simulations studies are conducted, and a data set of cortical thickness for schizophrenia is used to illustrate the proposed methods. According to our analysis between patients and controls of cortical thickness in typical mutual inhibitory brain regions, we verified the compensatory of cortical thickness in patients with schizophrenia and found its negative correlation with age.

Keywords: bivariate NPGA models; bivariate NPIG models; cortical thickness; N-EM algorithm; proportional data

check for updates

Citation: Sun, Y.; Tian, G.; Guo, S.; Shu, L.; Zhang, C. Bivariate Continuous Negatively Correlated Proportional Models with Applications in Schizophrenia Research. *Mathematics* **2022**, *10*, 353. https://doi.org/10.3390/math10030353

Academic Editors: Min Wang, Haijun Gong, Liucang Wu and Songfeng Zheng

Received: 15 December 2021
Accepted: 21 January 2022
Published: 24 January 2022

1. Introduction

In many aspects, experimental results or measurements are reported in the form of ratios, scores, proportions or percentages, which is frequently encountered in sociology, psychology, epidemiology and clinical trials. The characteristic of the data is that they are continuously valued within the unit interval $(0,1)$; thus, models focusing on this limited range are worthwhile. Researchers have developed different strategies for modeling such kinds of data. First, the beta distribution and beta regression models have been exhaustively studied by many authors, including [1–3]. Kieschnick and McCullough [4] summarized and compared different regression models for proportional data in the open interval. Next, the simplex distribution investigated by Zhang and Qiu [5] can also be utilized to model such continuous proportional data, and they further pointed out the simplex regression model is more robust than the beta model. By mimicking the construction of beta distributions with gamma variates, Lijoiu et al. [6] proposed a so-called normalized *inverse Gaussian* (IG) distribution by substituting the gamma variates with IG variates, as a new tool for modeling univariate proportional data. Later, Liu et al. [7] renamed it as the *proportional inverse Gaussian* (PIG) distribution and set up regression models. Due to the diversity and dimension enlarger of data, we need to generalize the univariate continuous proportional models to multivariate cases. Wang and Tu [8] considered the semiparametric tests for multigroup proportional data in a closed interval $[0, 1]$.

From the perspective of data structure, the multi-dimensional data limited in unit intervals can be divided into compositional data and multivariate proportional data according to their domains. For compositional data, which often appear in various fields, such as biology, medicine and economics, the summation of all components of data values equals one, also known as structure relative numbers reflecting the composition of objects. Thus, the corresponding models fitting for the compositional data are defined in the open hyperplane $\mathbb{T}_m = \{\boldsymbol{x} = (x_1, \ldots, x_m)^\top : x_j > 0,\ j = 1, \ldots, m,\ \mathbf{1}_m^\top \boldsymbol{x} = 1\}$. Due to the constraint of $\mathbf{1}_m^\top \boldsymbol{x} = 1$, it leads to certain negative correlations between any two dimensions of compositional data. One of the well-known distributions is the Dirichlet distribution, which can be regarded as a generalization of the beta distribution to more than two components. It was first used to fit two compositional biological data in [9]. Campbell and Mosimann [10] considered a Dirichlet regression model by linking the parameters to a set of covariates via a polynomial function, and the models with applications to the analysis of psychiatric data are investigated in [11]. By the way, the beta distribution could be regarded as a two-dimensional Dirichlet distribution, and a beta variate X and its complement $1 - X$ are also negatively correlated. Other research on related models can be found in recent literature [12,13].

For multivariate proportional data, it appears that each component of the data is valued between 0 and 1 with no direct constraint among components. The corresponding models for this type of data are defined in the unit cubic $(0, 1)^m = \{\boldsymbol{x} = (x_1, \ldots, x_m)^\top : 0 < x_j < 1,\ j = 1, \ldots, m\}$ without restriction $\mathbf{1}_m^\top \boldsymbol{x} = 1$. There are many ways to construct appropriate models, such as beta distribution with copula linking functions. Cepeda-Cuervo et al. [14] defined a bivariate beta regression model from copulas and considered the Bayesian approach, in which the correlation could be positive or negative. Petterle et al. [15] proposed a multivariate generalized linear mixed model for modeling continuous bounded variables in the interval $(0, 1)$. Sun et al. [16] proposed a linear *stochastic representation* (SR) to construct multivariate positively correlated continuous models based on IG and gamma distributions, named as multivariate PIG and *proportional gamma* (PGA) distributions, respectively, which can only fit positively correlated continuous proportional data.

The cortical thickness of schizophrenia data used in [16] shows high correlations and compensation behaviors related to disease severity among different brain regions. Further, we find that a large number of negative covariant region pairs may occur in patients if the changes of compensations are reduced. This indicates the observations of negatively correlated regions in cortical thickness are of great significance for the study of schizophrenia and its prognosis. Motivated by the construction technique in multivariate PIG and PGA distributions, we will propose models to capture the negative correlation among components for multivariate proportional data. To the best of our knowledge, work considering the negative correlation of multivariate proportion data is quite scarce. Here, we focus on the bivariate situations; thus, the proposed models are expected to provide efficient tools in modeling negatively correlated proportional data.

By combining the construction of multivariate PIG/PGA distributions and the negative correlation structure in beta/Dirichlet distributions, we define a new random vector $\mathbf{x} = (X_1, X_2)^\top \in (0, 1)^2$ via the following SR:

$$X_1 = \frac{Y_1}{Y_0 + Y_1} \quad \text{and} \quad X_2 = 1 - \frac{Y_2}{Y_0 + Y_2} = \frac{Y_0}{Y_0 + Y_2}, \tag{1}$$

where $\{Y_j\}_{j=0}^2$ are independent random variables with the same support $\mathbb{R}_+$, and each Y_j can follow any same continuous distribution family but with possibly different parameters. In the following, for each Y_j ($j = 0, 1, 2$), we applied the IG and gamma distributions to construct bivariate *negatively correlated PIG* (NPIG) and *negatively correlated proportional gamma* (NPGA) distributions.

The rest of the paper is organized as follows. In Sections 2 and 3, the bivariate NPIG and NPGA distributions are, respectively, proposed and related distributional properties (e.g., moments, joint densities) are provided. Moreover, the *normalized expectation–maximization* (N-EM) facilitated by the one-step gradient descent algorithms are established for calculating the *maximum likelihood* (ML) estimations of parameters of interest. In Section 4, simulations for the proposed methods are performed. A data set on the cortical thickness of schizophrenia is used to illustrate the proposed methods in Section 5. Finally, a discussion is provided in Section 6. Some technical details are put in the Appendices A and B, and others are shown in the Supplementary Material.

2. Bivariate Negatively Correlated PIG Models

First, we propose a new bivariate NPIG distribution based on equi-dispersed IG distributions and develop the corresponding NPIG mean regression model. The N-EM algorithms for calculating the ML estimators of parameters are also provided.

2.1. Bivariate NPIG Distribution

The IG distribution with location parameter a (>0) and shape parameter b (>0), denoted by $Y \sim \mathrm{IG}(a, b)$, if it has the *probability density function* (pdf)

$$f_{\mathrm{IG}}(y|a,b) = \sqrt{\frac{b}{2\pi}}\, y^{-\frac{3}{2}} \exp\left[-\frac{b(y-a)^2}{2a^2 y}\right], \quad y > 0.$$

According to the results of [17], we have $E(Y) = a$ and $\mathrm{Var}(Y) = a^3/b$. By setting $b = a^2$, the general IG distribution reduces to the equi-dispersed $\mathrm{IG}(a, a^2)$ as its mean equals the variance.

By adopting three independent equi-dispersed IG variates $Y_j \overset{\mathrm{ind}}{\sim} \mathrm{IG}(\mu_j, \mu_j^2)$ with $\mu_j > 0$ for $j = 0, 1, 2$, the random vector defined by (1) is said to follow a bivariate NPIG distribution, denoted by $\mathbf{x} = (X_1, X_2)^\top \sim \mathrm{NPIG}_2(\boldsymbol{\mu})$ with $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)^\top$. Since the *moment generating functions* (MGF) of Y_j is $M_{Y_j}(t) = \exp[\mu_j(1-\sqrt{1-2t}\,)]$, the expectations, variances and the covariance are computed based on (A1)–(A5) as

$$E(X_1) = \frac{\mu_1}{\mu_0+\mu_1} \triangleq \theta_1 \in (0,1), \tag{2}$$

$$E(X_2) = \frac{\mu_0}{\mu_0+\mu_2} \triangleq \theta_2 \in (0,1), \tag{3}$$

$$\mathrm{Var}(X_j) = \mu_0\mu_j \mathrm{e}^{\mu_0+\mu_j}\Gamma(-2, \mu_0+\mu_j), \quad j = 1, 2,$$

$$\mathrm{Cov}(X_1, X_2) = -\mu_1\mu_2 \mathrm{e}^{\mu_0+\mu_1+\mu_2} \int_1^\infty \int_1^\infty \mathrm{e}^{-\mu_1 t}\mathrm{e}^{-\mu_2 s}\left[\mathrm{e}^{-\mu_0\sqrt{t^2+s^2-1}} - \mathrm{e}^{-\mu_0(t+s-1)}\right] \mathrm{d}t\,\mathrm{d}s,$$

where $\Gamma(-2, \mu_0+\mu_j) = \int_{\mu_0+\mu_j}^\infty t^{-3}\mathrm{e}^{-t}\,\mathrm{d}t$ is the incomplete gamma function. According to the numerical experiments in [16], the correlation coefficient is limited in the open interval $(-1, 0)$. The joint pdf of the bivariate NPIG distribution is derived as

$$f_{\mathrm{NPIG}_2}(\boldsymbol{x}|\boldsymbol{\mu}) = \frac{\prod_{j=0}^2 \mu_j \exp(\mu_j)}{(2\pi)^{\frac{3}{2}}\left[x_1^3 x_2(1-x_2)^3(1-x_1)\right]^{\frac{1}{2}}} \int_0^\infty h(s|\boldsymbol{x},\boldsymbol{\mu})\,\mathrm{d}s, \quad \boldsymbol{x} = (x_1, x_2)^\top \in (0,1)^2,$$

where

$$h(s|\boldsymbol{x},\boldsymbol{\mu}) = s^{-\frac{5}{2}} \exp\left\{-\frac{1}{2}\left[s\cdot a(\boldsymbol{x}) + \frac{1}{s}\cdot b(\boldsymbol{x},\boldsymbol{\mu})\right]\right\},$$

$$a(\boldsymbol{x}) = 1 + \frac{x_1}{1-x_1} + \frac{1-x_2}{x_2} \quad \text{and} \quad b(\boldsymbol{x},\boldsymbol{\mu}) = \mu_0^2 + \frac{1-x_1}{x_1}\mu_1^2 + \frac{x_2}{1-x_2}\mu_2^2.$$

From the perspective of practice, usually, we would like to have intuitive interpretations of population means. Therefore, we re-parametrize the bivariate NPIG distribution in terms of the parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^\top$ according to (2) and (3) by making the following one-to-one mapping

$$\mu_0 = \theta_0, \quad \mu_1 = \theta_0\theta_1/(1-\theta_1) \quad \text{and} \quad \mu_2 = \theta_0/\theta_2 - \theta_0.$$

The pdf of the re-parameterized bivariate NPIG distribution, denoted by $\mathbf{x} \sim \mathrm{NPIG}_2(\boldsymbol{\theta})$, is

$$f_{\mathrm{NPIG}_2}(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\theta_0^3 \cdot \frac{\theta_1}{1-\theta_1} \cdot \frac{1-\theta_2}{\theta_2} \exp\left(\theta_0 + \theta_0 \cdot \frac{\theta_1}{1-\theta_1} + \theta_0 \cdot \frac{1-\theta_2}{\theta_2}\right)}{(2\pi)^{\frac{3}{2}} \left[x_1^3 x_2 (1-x_2)^3 (1-x_1)\right]^{\frac{1}{2}}} \int_0^\infty h_1(s|\boldsymbol{x}, \boldsymbol{\theta})\, \mathrm{d}s,$$

where $\boldsymbol{x} \in (0,1)^2$,

$$\begin{aligned} h_1(s|\boldsymbol{x},\boldsymbol{\theta}) &= s^{-\frac{5}{2}} \exp\left\{-\frac{1}{2}\left[s \cdot a(\boldsymbol{x}) + \frac{1}{s} \cdot b_1(\boldsymbol{x},\boldsymbol{\theta})\right]\right\} \quad \text{and} \\ b_1(\boldsymbol{x},\boldsymbol{\theta}) &= \theta_0^2 \left[1 + \frac{(1-x_1)\theta_1^2}{x_1(1-\theta_1)^2} + \frac{x_2(1-\theta_2)^2}{(1-x_2)\theta_2^2}\right]. \end{aligned} \tag{4}$$

Figure 1 plots the bivariate NPIG distribution $\mathrm{NPIG}_2(\boldsymbol{\theta})$ with two sets of different values of parameters. We note that a larger value of θ_0 makes the distribution more concentrated, and it also influences the number of modes. When θ_0 is large enough, the change in the values of $(\theta_1, \theta_2)^\top$ affects the location of modes and the skewness of distributions. Thus, it is appropriate to regard θ_0 as the dispersion parameter and θ_1, θ_2 as the two location parameters. Sometimes, while the distributions are dense and unimodal, the modes are very different from the expectations.

2.2. ML Estimation of Parameters via the N-EM Algorithm

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{iid}}{\sim} \mathrm{NPIG}_2(\boldsymbol{\theta})$ and $Y_{\mathrm{obs}_1} = \{\boldsymbol{x}_i\}_{i=1}^n$ denote the observed data, where $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$ is the realization of $\mathbf{x}_i = (X_{i1}, X_{i2})^\top$. The log-likelihood function of the parameter vector $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \ell_1(\boldsymbol{\theta}|Y_{\mathrm{obs}_1}) &= 3n\log\theta_0 + n\theta_0 + n\frac{\theta_0\theta_1}{1-\theta_1} + n\log\frac{\theta_1}{1-\theta_1} + n\frac{\theta_0(1-\theta_2)}{\theta_2} + n\log\frac{1-\theta_2}{\theta_2} \\ &\quad + \sum_{i=1}^n \log\left[\int_0^\infty h_1(s|\boldsymbol{x}_i,\boldsymbol{\theta})\,\mathrm{d}s\right] + c_1, \end{aligned} \tag{5}$$

where c_1 is a constant free from the parameter vector $\boldsymbol{\theta}$. Due to the existence of the intractable integrals in (5), neither the Newton–Raphson nor the Fisher scoring algorithm is attainable in dealing with the above expression. Instead, we adopt the N-EM algorithm, which is composed of three steps:

N-step: Establish the following normalized density function based on $h_1(\cdot|\boldsymbol{x}_i,\boldsymbol{\theta})$ as

$$g_1(s|\boldsymbol{x}_i,\boldsymbol{\theta}) \triangleq \frac{h_1(s|\boldsymbol{x}_i,\boldsymbol{\theta})}{\int_0^\infty h_1(t|\boldsymbol{x}_i,\boldsymbol{\theta})\,\mathrm{d}t}, \quad s > 0,$$

so that $g_1(s|\boldsymbol{x}_i,\boldsymbol{\theta}^{(t)})$ is also a valid pdf defined on $(0,\infty)$, where $\boldsymbol{\theta}^{(t)}$ denotes the t-th approximation of $\hat{\boldsymbol{\theta}}$.

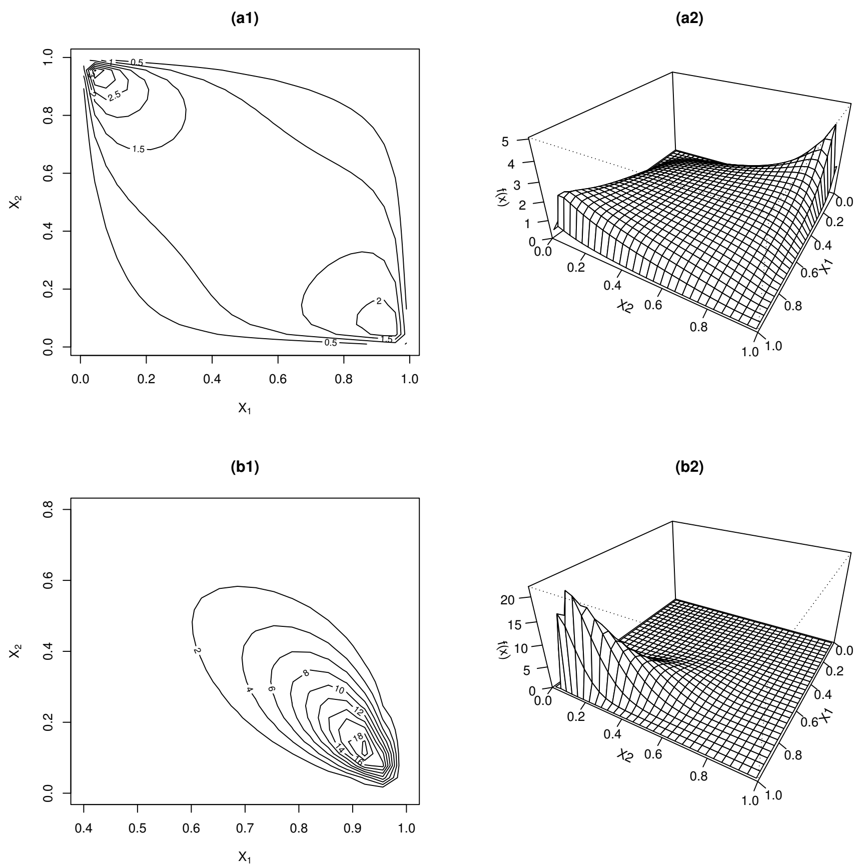

Figure 1. The contour plots and 3D perspectives of the bivariate NPIG distribution $\mathrm{NPIG}_2(\boldsymbol{\theta})$ with different values of parameters: (**a1**,**a2**) $\boldsymbol{\theta} = (0.5, 0.5, 0.5)^\top$; (**b1**,**b2**) $\boldsymbol{\theta} = (1.5, 0.8, 0.3)^\top$.

E-step: Construct a surrogate Q-function by utilizing the integral version of Jensen's inequality as

$$\begin{aligned} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \quad &= \quad 3n\log\theta_0 + n\theta_0 + n\frac{\theta_0\theta_1}{1-\theta_1} + n\log\frac{\theta_1}{1-\theta_1} + n\frac{\theta_0(1-\theta_2)}{\theta_2} \\ &\quad + n\log\frac{1-\theta_2}{\theta_2} - \frac{1}{2}\sum_{i=1}^{n}\Big[B_1(\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)}) \cdot b_1(\boldsymbol{x}_i, \boldsymbol{\theta})\Big] + c_1^{(t)}, \end{aligned} \tag{6}$$

where

$$B_1(\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)}) \triangleq \int_0^\infty s^{-1} \cdot g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}s,$$

$b_1(\boldsymbol{x}, \boldsymbol{\theta})$ is defined by (4), and $c_1^{(t)}$ is a constant not depending on $\boldsymbol{\theta}$. It can be proven that $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ satisfies

$$Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \leqslant \ell_1(\boldsymbol{\theta}|Y_{\mathrm{obs}_1}) \quad \text{and} \quad Q_1(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) = \ell_1(\boldsymbol{\theta}^{(t)}|Y_{\mathrm{obs}_1}),$$

indicating that it minorizes $\ell_1(\boldsymbol{\theta}|Y_{\mathrm{obs}_1})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$.

M-step: Maximize $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$ and obtain

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}_+\times(0,1)^2} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}).$$

However, it is difficult to obtain the unique explicit expression of $\boldsymbol{\theta}^{(t+1)}$ in the M-step. Instead, it is recommended to separate the estimation procedures into two parts:

M-step-1: Given $\{\theta_1^{(t)}, \theta_2^{(t)}\}$, by solving $\partial Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\theta_0 = 0$, we have the $(t+1)$-th approximation for $\hat{\theta}_0$ as

$$\theta_0^{(t+1)} = \frac{T_1^{(t)} + \sqrt{[T_1^{(t)}]^2 + 12nT_2^{(t)}}}{2T_2^{(t)}}, \tag{7}$$

where

$$\begin{aligned} T_1^{(t)} &= n\left(1 + \frac{\theta_1^{(t)}}{1-\theta_1^{(t)}} + \frac{1-\theta_2^{(t)}}{\theta_2^{(t)}}\right) \quad \text{and} \\ T_2^{(t)} &= \sum_{i=1}^{n} B_1(\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})\left\{1 + \frac{(1-x_{i1})[\theta_1^{(t)}]^2}{x_{i1}[1-\theta_1^{(t)}]^2} + \frac{x_{i2}[1-\theta_2^{(t)}]^2}{(1-x_{i2})[\theta_2^{(t)}]^2}\right\}. \end{aligned}$$

M-step-2: The iteration for $\boldsymbol{\theta}_{-0} \triangleq (\theta_1, \theta_2)^\top$ is obtained by adopting the gradient descent algorithm as

$$\boldsymbol{\theta}_{-0}^{(t+1)} = \boldsymbol{\theta}_{-0}^{(t)} + s_1^{(t)} \nabla G_1(\boldsymbol{\theta}_{-0}^{(t)}|\boldsymbol{\theta}^{(t)}), \tag{8}$$

where

$$\nabla G_1(\boldsymbol{\theta}_{-0}|\boldsymbol{\theta}^{(t)}) = \frac{\partial Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\theta}_{-0}} = \left(\frac{\partial Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\theta_1}, \frac{\partial Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\theta_2}\right)^\top,$$

and $s_1^{(t)}$ is the step size at the t-th iteration of the algorithm, determined by

$$s_1^{(t)} = \frac{|[\boldsymbol{\theta}_{-0}^{(t)} - \boldsymbol{\theta}_{-0}^{(t-1)}]^\top[\nabla G_1(\boldsymbol{\theta}_{-0}^{(t)}|\boldsymbol{\theta}^{(t)}) - \nabla G_1(\boldsymbol{\theta}_{-0}^{(t-1)}|\boldsymbol{\theta}^{(t-1)})]|}{\|\nabla G_1(\boldsymbol{\theta}_{-0}^{(t)}|\boldsymbol{\theta}^{(t)}) - \nabla G_1(\boldsymbol{\theta}_{-0}^{(t-1)}|\boldsymbol{\theta}^{(t-1)})\|^2}.$$

The stopping rule of the above loops under the proposed N-EM embedded with the gradient descent algorithm is controlled by

$$\max\left\{|\ell_1(\boldsymbol{\theta}^{(t+1)}|Y_{\text{obs}_1}) - \ell_1(\boldsymbol{\theta}^{(t)}|Y_{\text{obs}_1})|,\ \|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\|_\infty\right\} \leqslant \delta,$$

where δ is a pre-determined precision. The details of constructing the N-EM algorithm are shown in Appendix B.1, and other relevant calculations are given in Supplementary Material A.1 and A.2. Finally, the ML estimates of $(\theta_0, \theta_1, \theta_2)$ can be obtained by combining (7) and (8) when the algorithm stops.

2.3. Bivariate NPIG Mean Regression Model

We extend the re-parametrized $\text{NPIG}_2(\boldsymbol{\theta})$ distribution to the corresponding regression model for investigating the relationship between the mean vector $(\theta_1, \theta_2)^\top$ with a set of covariates. The logit link function is adopted for $\theta_j \in (0,1)$ with $j = 1, 2$, then the resulting model can be formulated as

$$\begin{cases} \mathbf{x}_i = (X_{i1}, X_{i2})^\top \overset{\text{ind}}{\sim} \text{NPIG}_2(\theta_0, \theta_{i1}, \theta_{i2}), \quad i = 1, \ldots, n, \\ \log\left(\dfrac{\theta_{ij}}{1-\theta_{ij}}\right) = \boldsymbol{w}_i^\top \boldsymbol{\alpha}_j \quad \text{or} \quad \theta_{ij} = \dfrac{\exp(\boldsymbol{w}_i^\top \boldsymbol{\alpha}_j)}{1 + \exp(\boldsymbol{w}_i^\top \boldsymbol{\alpha}_j)}, \quad j = 1, 2, \end{cases} \tag{9}$$

where $\boldsymbol{w}_i = (1, w_{i1}, \ldots, w_{iq})^\top$ is the vector of covariates associated with the i-th subject, and $\boldsymbol{\alpha}_j = (\alpha_{0j}, \alpha_{1j}, \ldots, \alpha_{qj})^\top$ is the $(q+1)$-vector of unknown regression coefficients. The log-likelihood function of the new parameter vector $\boldsymbol{\vartheta} = (\theta_0, \boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$ for the regression model given the observed data $Y_{\text{obs}_2} = \{\boldsymbol{x}_i, \boldsymbol{w}_i\}_{i=1}^n$ is written as

$$\begin{aligned}\ell_2(\boldsymbol{\vartheta}|Y_{\text{obs}_2}) &= 3n\log\theta_0 + n\theta_0 + \sum_{i=1}^n \Big[\theta_0\exp(\boldsymbol{w}_i^\top\boldsymbol{\alpha}_1) + \boldsymbol{w}_i^\top\boldsymbol{\alpha}_1 + \theta_0\exp(-\boldsymbol{w}_i^\top\boldsymbol{\alpha}_2) - \boldsymbol{w}_i^\top\boldsymbol{\alpha}_2\Big] \\ &\quad + \sum_{i=1}^n \log\left[\int_0^\infty h_2(s|\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta})\,\mathrm{d}s\right] + c_2,\end{aligned}$$

where c_2 is a constant free from the parameter vector $\boldsymbol{\vartheta}$,

$$\begin{aligned}h_2(s|\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}) &= s^{-\frac{5}{2}}\exp\left\{-\frac{1}{2}\left[s\cdot a(\boldsymbol{x}_i) + \frac{1}{s}\cdot b_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta})\right]\right\} \quad \text{and} \\ b_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}) &= \theta_0^2 + \frac{\theta_0^2(1-x_{i1})}{x_{i1}}\exp(2\boldsymbol{w}_i^\top\boldsymbol{\alpha}_1) + \frac{\theta_0^2 x_{i2}}{1-x_{i2}}\exp(-2\boldsymbol{w}_i^\top\boldsymbol{\alpha}_2).\end{aligned}$$

Similar to the construction of $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, we can obtain

$$\begin{aligned}Q_2(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(t)}) &= 3n\log\theta_0 + n\theta_0 + \sum_{i=1}^n \Big[\theta_0\exp(\boldsymbol{w}_i^\top\boldsymbol{\alpha}_1) + \boldsymbol{w}_i^\top\boldsymbol{\alpha}_1 + \theta_0\exp(-\boldsymbol{w}_i^\top\boldsymbol{\alpha}_2) - \boldsymbol{w}_i^\top\boldsymbol{\alpha}_2\Big] \\ &\quad - \frac{1}{2}\sum_{i=1}^n \Big[B_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}^{(t)})\cdot b_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta})\Big] + c_2^{(t)},\end{aligned}$$

where $c_2^{(t)}$ is a constant, $\boldsymbol{\vartheta}^{(t)}$ denotes the t-th approximation of the ML estimator $\hat{\boldsymbol{\vartheta}}$ and

$$B_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}^{(t)}) \triangleq \int_0^\infty \frac{s^{-1}\cdot h_2(s|\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}^{(t)})}{\int_0^\infty h_2(t|\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}^{(t)})\,\mathrm{d}t}\,\mathrm{d}s.$$

The procedure of obtaining the ML estimators of $\boldsymbol{\vartheta}$ is similar to that in Section 2.2. First, for given $\{\boldsymbol{\alpha}_1^{(t)}, \boldsymbol{\alpha}_2^{(t)}\}$, we set $\partial Q_2(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(t)})/\partial\theta_0 = 0$ and find the positive root to obtain the $(t+1)$-th approximation for $\hat{\theta}_0$, which is given by

$$\theta_0^{(t+1)} = \frac{T_3^{(t)} + \sqrt{\big[T_3^{(t)}\big]^2 + 12nT_4^{(t)}}}{2T_4^{(t)}} \tag{10}$$

with

$$\begin{aligned}T_3^{(t)} &= n + \sum_{i=1}^n \Big[\exp(\boldsymbol{w}_i^\top\boldsymbol{\alpha}_1^{(t)}) + \exp(-\boldsymbol{w}_i^\top\boldsymbol{\alpha}_2^{(t)})\Big] \quad \text{and} \\ T_4^{(t)} &= \sum_{i=1}^n B_2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\vartheta}^{(t)})\left[1 + \frac{1-x_{i1}}{x_{i1}}\exp(2\boldsymbol{w}_i^\top\boldsymbol{\alpha}_1^{(t)}) + \frac{x_{i2}}{1-x_{i2}}\exp(-2\boldsymbol{w}_i^\top\boldsymbol{\alpha}_2^{(t)})\right].\end{aligned}$$

Moreover, to obtain the ML estimator of $\boldsymbol{\vartheta}_{-0} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$, we first define

$$\nabla G_2(\boldsymbol{\vartheta}_{-0}|\boldsymbol{\vartheta}^{(t)}) \triangleq \frac{\partial Q_2(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(t)})}{\partial\boldsymbol{\vartheta}_{-0}} = \left(\frac{\partial Q_2(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(t)})}{\partial\boldsymbol{\alpha}_1^\top}, \frac{\partial Q_2(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(t)})}{\partial\boldsymbol{\alpha}_2^\top}\right)^\top.$$

Using the one-step gradient descent algorithm, we have the iteration

$$\boldsymbol{\vartheta}_{-0}^{(t+1)} = \boldsymbol{\vartheta}_{-0}^{(t)} + s_2^{(t)}\nabla G_2(\boldsymbol{\vartheta}_{-0}^{(t)}|\boldsymbol{\vartheta}^{(t)}), \tag{11}$$

where the step size $s_2^{(t)}$ is defined by

$$s_2^{(t)} = \frac{|[\boldsymbol{\vartheta}_{-0}^{(t)} - \boldsymbol{\vartheta}_{-0}^{(t-1)}]^\top [\nabla G_2(\boldsymbol{\vartheta}_{-0}^{(t)}|\boldsymbol{\vartheta}^{(t)}) - \nabla G_2(\boldsymbol{\vartheta}_{-0}^{(t-1)}|\boldsymbol{\vartheta}^{(t-1)})]|}{\|\nabla G_2(\boldsymbol{\vartheta}_{-0}^{(t)}|\boldsymbol{\vartheta}^{(t)}) - \nabla G_2(\boldsymbol{\vartheta}_{-0}^{(t-1)}|\boldsymbol{\vartheta}^{(t-1)})\|^2}.$$

By combining (10) with (11), we could obtain the ML estimates of $\boldsymbol{\vartheta}$.

3. Bivariate Negatively Correlated PGA Models

To provide other candidates for flexibly modeling the above-mentioned negatively correlated continuous proportional data, in this section, we propose a new bivariate NPGA distribution based on equi-dispersed gamma distributions (see the first paragraph in Section 3.1) and develop a bivariate NPGA mean regression model.

3.1. Bivariate NPGA Distribution

Let $Y \sim \text{Gamma}(a, 1)$, then it is an equi-dispersed gamma distribution with $E(Y) = \text{Var}(Y)$, and its pdf is $f_{\text{GA}}(y|a) = y^{a-1}\mathrm{e}^{-y}/\Gamma(a)$, $y > 0$. Let $\{Y_j\}_{j=0}^2 \stackrel{\text{ind}}{\sim} \text{Gamma}(\lambda_j, 1)$ with $\lambda_j > 0$ for $j = 0, 1, 2$ be three independent equi-dispersed gamma variates, then the random vector defined by (1) is said to follow a bivariate NPGA distribution, denoted by $\mathbf{x} = (X_1, X_2)^\top \sim \text{NPGA}_2(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)^\top$. The MGF of Y_j, in this case, is $M_{Y_j}(t) = (1-t)^{-\lambda_j}$, with $t < 1$, from (A1)–(A5), we have

$$E(X_1) = \frac{\lambda_1}{\lambda_0 + \lambda_1} \triangleq \phi_1 \in (0, 1), \tag{12}$$

$$E(X_2) = \frac{\lambda_0}{\lambda_0 + \lambda_2} \triangleq \phi_2 \in (0, 1), \tag{13}$$

$$\text{Var}(X_j) = \frac{\lambda_0 \lambda_j}{(\lambda_0 + \lambda_j)^2 (1 + \lambda_0 + \lambda_j)}, \quad j = 1, 2,$$

$$\text{Cov}(X_1, X_2) = -\lambda_1 \lambda_2 \int_1^\infty \int_1^\infty t^{-\lambda_1 - 1} s^{-\lambda_2 - 1} \Big[(t + s - 1)^{-\lambda_0} - (ts)^{-\lambda_0}\Big]\, \mathrm{d}t\, \mathrm{d}s.$$

The correlation coefficient takes values within $(-1, 0)$ as well. The pdf of $\mathbf{x} \sim \text{NPGA}_2(\boldsymbol{\lambda})$ is

$$f_{\text{NPGA}_2}(\boldsymbol{x}|\boldsymbol{\lambda}) = \frac{x_1^{\lambda_1 - 1}(1 - x_2)^{\lambda_2 - 1}\Gamma(\lambda_+)}{x_2^{\lambda_2 + 1}(1 - x_1)^{\lambda_1 + 1}\prod_{j=0}^2 \Gamma(\lambda_j)} \left(1 + \frac{x_1}{1 - x_1} + \frac{1 - x_2}{x_2}\right)^{-\lambda_+},$$

where $\boldsymbol{x} = (x_1, x_2)^\top \in (0, 1)^2$ is the realization of $\mathbf{x}$ and $\lambda_+ = \sum_{j=0}^2 \lambda_j$.

For the purpose of modeling the population means in (12) and (13) directly, we also make a one-to-one transformation among parameter vectors $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2)^\top$ and $\boldsymbol{\lambda}$ by

$$\lambda_0 = \phi_0, \quad \lambda_1 = \phi_0\phi_1/(1 - \phi_1) \quad \text{and} \quad \lambda_2 = \phi_0/\phi_2 - \phi_0.$$

The pdf of re-parameterized bivariate NPGA distribution, denoted by $\mathbf{x} \sim \text{NPGA}_2(\boldsymbol{\phi})$, is

$$\begin{aligned} f_{\text{NPGA}_2}(\boldsymbol{x}|\boldsymbol{\phi}) &= \frac{x_1^{\frac{\phi_0\phi_1}{1-\phi_1} - 1}(1 - x_2)^{\frac{\phi_0}{\phi_2} - \phi_0 - 1}\Gamma\left(\frac{\phi_0\phi_1}{1-\phi_1} + \frac{\phi_0}{\phi_2}\right)}{x_2^{\frac{\phi_0}{\phi_2} - \phi_0 + 1}(1 - x_1)^{\frac{\phi_0\phi_1}{1-\phi_1} + 1}\Gamma(\phi_0)\Gamma\left(\frac{\phi_0\phi_1}{1-\phi_1}\right)\Gamma\left(\frac{\phi_0}{\phi_2} - \phi_0\right)} \\ &\quad \times \left(1 + \frac{x_1}{1 - x_1} + \frac{1 - x_2}{x_2}\right)^{-\frac{\phi_0\phi_1}{1-\phi_1} - \frac{\phi_0}{\phi_2}}. \end{aligned}$$

Figure 2 plots the bivariate NPGA distribution $\mathrm{NPGA}_2(\boldsymbol{\phi})$ with two sets of different values of parameters. Similar to those findings in Figure 1, ϕ_0 is regarded as the dispersion parameter and $(\phi_1, \phi_2)^\top$ is the location vector.

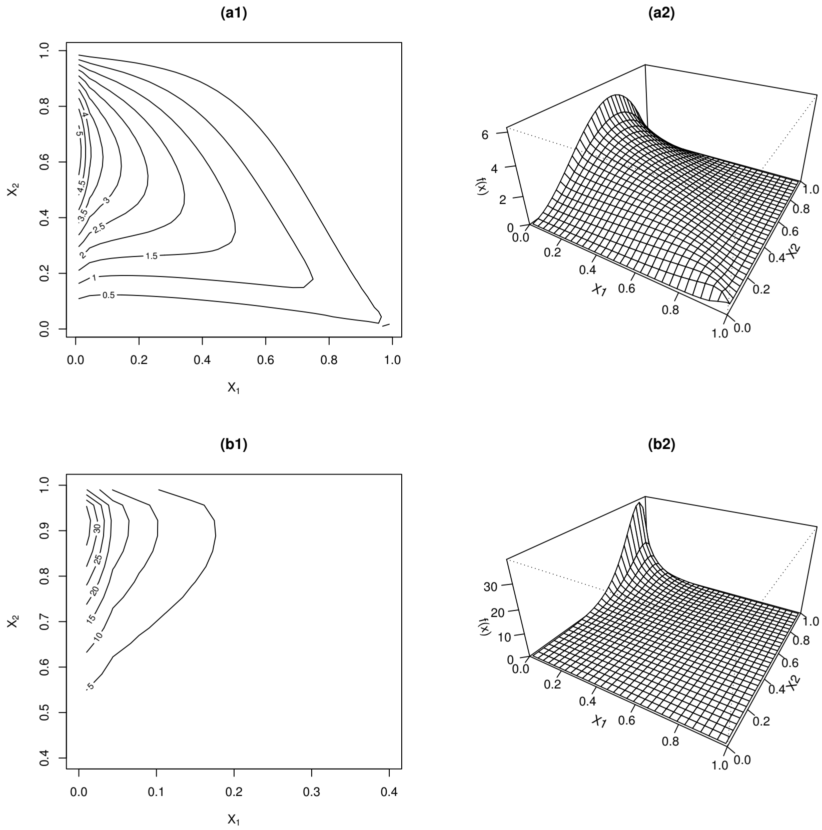

Figure 2. The contour plots and 3D perspectives of the bivariate NPIG distribution $\mathrm{NPGA}_2(\boldsymbol{\phi})$ with different values of parameters: (**a1**,**a2**) $\boldsymbol{\phi} = (2, 0.3, 0.5)^\top$; (**b1**,**b2**) $\boldsymbol{\phi} = (6, 0.1, 0.8)^\top$.

3.2. ML Estimation of Parameters via the Gradient Descent Algorithm

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{iid}}{\sim} \mathrm{NPGA}_2(\boldsymbol{\phi})$ and $Y_{\text{obs}_3} = \{\boldsymbol{x}_i\}_{i=1}^n$ denote the observed data, where $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$ is the realization of $\mathbf{x}_i = (X_{i1}, X_{i2})^\top$. The log-likelihood function of the parameter vector $\boldsymbol{\phi}$ is given by

$$\begin{aligned}
\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3}) &= n\left[\log\Gamma\left(\frac{\phi_0\phi_1}{1-\phi_1} + \frac{\phi_0}{\phi_2}\right) - \log\Gamma(\phi_0) - \log\Gamma\left(\frac{\phi_0\phi_1}{1-\phi_1}\right) - \log\Gamma\left(\frac{\phi_0}{\phi_2} - \phi_0\right)\right] \\
&\quad + \frac{\phi_0\phi_1}{1-\phi_1}\sum_{i=1}^n \log\frac{x_{i1}}{1-x_{i1}} + \left(\frac{\phi_0}{\phi_2} - \phi_0\right)\sum_{i=1}^n \log\frac{1-x_{i2}}{x_{i2}} \\
&\quad - \left(\frac{\phi_0\phi_1}{1-\phi_1} + \frac{\phi_0}{\phi_2}\right)\sum_{i=1}^n \log\left(1 + \frac{x_{i1}}{1-x_{i1}} + \frac{1-x_{i2}}{x_{i2}}\right) + c_3,
\end{aligned}$$

where c_3 is a constant free from the parameter vector $\boldsymbol{\phi}$. Then, we adopt the gradient descent algorithm directly to find the ML estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ by setting

$$\nabla\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3}) \triangleq \frac{\partial\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3})}{\partial\boldsymbol{\phi}} = \left(\frac{\partial\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3})}{\partial\phi_0}, \frac{\partial\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3})}{\partial\phi_1}, \frac{\partial\ell_3(\boldsymbol{\phi}|Y_{\text{obs}_3})}{\partial\phi_2}\right)^\top.$$

Thus, the $(t+1)$-th estimation is given by

$$\boldsymbol{\phi}^{(t+1)} = \boldsymbol{\phi}^{(t)} + s_3^{(t)}\nabla\ell_3(\boldsymbol{\phi}^{(t)}|Y_{\text{obs}_3}), \tag{14}$$

where the step size at the t-th iteration is

$$s_3^{(t)} = \frac{|[\boldsymbol{\phi}^{(t)} - \boldsymbol{\phi}^{(t-1)}]^\top [\nabla \ell_3(\boldsymbol{\phi}^{(t)}|Y_{\text{obs}_3}) - \nabla \ell_3(\boldsymbol{\phi}^{(t-1)}|Y_{\text{obs}_3})]|}{\|\nabla \ell_3(\boldsymbol{\phi}^{(t)}|Y_{\text{obs}_3}) - \nabla \ell_3(\boldsymbol{\phi}^{(t-1)}|Y_{\text{obs}_3})\|^2}.$$

We also provide another method in Appendix B.2 with the N-EM algorithm applied, which results in the same iteration shown in (14).

3.3. Bivariate NPGA Mean Regression Model

The bivariate NPGA mean regression model is formulated in a similar way as

$$\begin{cases} \mathbf{x}_i = (X_{i1}, X_{i2})^\top \overset{\text{ind}}{\sim} \text{NPGA}_2(\phi_0, \phi_{i1}, \phi_{i2}), \quad i = 1, \ldots, n, \\ \log\left(\dfrac{\phi_{ij}}{1 - \phi_{ij}}\right) = \boldsymbol{v}_i^\top \boldsymbol{\beta}_j, \quad \text{or} \quad \phi_{ij} = \dfrac{\exp(\boldsymbol{v}_i^\top \boldsymbol{\beta}_j)}{1 + \exp(\boldsymbol{v}_i^\top \boldsymbol{\beta}_j)}, \quad j = 1, 2, \end{cases} \tag{15}$$

where $\boldsymbol{v}_i = (1, v_{i1}, \ldots, v_{iq})^\top$ is the vector of covariates associated with the i-th subject, and $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{qj})^\top$ is the $(q+1)$-vector of unknown regression coefficients. The gradient descent algorithm still works for finding the ML estimators of $\boldsymbol{\varphi} = (\phi_0, \boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ in the NPGA mean regression model, which is similar to that stated in Section 3.2.

4. Simulation Experiments

For all above bivariate NPIG- and NPGA-related models, although no explicit expressions for the ML estimators of parameters, the bootstrap method is an efficient tool to approximately calculate the standard errors and the *confidence intervals* (CIs) for them, while the details of the bootstrap procedure are omitted due to its routines. Based on it, we conduct several numerical experiments in the section to investigate the performances of the above-proposed estimation methods for the bivariate NPIG and NPGA distributions with their corresponding mean regression models. We use R software to design parallel computing on 64 CPUs of the windows system for the time-consuming simulations. The impacts of other computational aspects on the simulations are not considered.

4.1. Experiment for NPIG Models

Firstly, for the bivariate NPIG distribution, we choose the parameter configurations as follows: sample size is $n = 30, 50, 100, 300$; true values of $(\theta_0, \theta_1, \theta_2)$ are set as $(1.2, 0.3, 0.8)$, $(0.5, 0.5, 0.6)$ and $(0.2, 0.8, 0.1)$, corresponding to a low, moderate and high correlations. In the regression model, the corresponding sample size is $n = 50, 200, 350, 500$; $\theta_0 = 0.6$, $\boldsymbol{\alpha}_1 = (1.2, 0.8, -0.5, 0.5)^\top$, $\boldsymbol{\alpha}_2 = (1.5, -2, 0.7, -0.5)^\top$; the covariates are $\boldsymbol{w}_i = (1, w_{i1}, w_{i2}, w_{i3})^\top$, with $w_{i1} \overset{\text{iid}}{\sim} \text{Unif}(-1, 1)$, w_{i2} is randomly chosen from $\{0.2, 0.4, 0.6, 0.8\}$ and $w_{i3} \overset{\text{iid}}{\sim} \text{Poisson}(3)$ for $i = 1, \ldots, n$. For a given sample size n, experimental data $\{\boldsymbol{x}_i\}_{i=1}^n$ are i.i.d. sampled from $\text{NPIG}_2(\theta_0, \theta_1, \theta_2)$ or each $\boldsymbol{x}_i$ is generated from $\text{NPIG}_2(\theta_0, \theta_{i1}, \theta_{i2})$ according to the regression model specified by (9), where SR (1) based on three IG variates can facilitate the sample generation. Parameters of interest are $(\theta_0, \theta_1, \theta_2)$ and $(\theta_0, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$, respectively.

For each generated sample group, parameters are estimated by the proposed N-EM embedded with the gradient descent algorithm, and the whole process is repeated K times. The value of K is chosen as 1000 and 500 for the distribution and regression model, respectively. To better express the quantitative values on evaluating the estimation accuracy, we use a general symbol ψ to denote each component of parameters to be estimated, and ψ_0 is its true value. The obtained ML estimate for ψ in each loop is denoted by $\hat{\psi}^{(k)}$, and the number of iterations is recorded as t_k to the converged algorithm, where $k = 1, \ldots, K$.

The *averaged ML estimate* (Ave-MLE), *standard deviation* (Std) and *mean squared error* (MSE) for the estimator $\hat{\psi}$ and the *averaged iterative number* (it.no) are, respectively, computed as

$$
\begin{aligned}
\text{Ave-MLE}(\hat{\psi}) &= \frac{1}{K}\sum_{k=1}^{K}\hat{\psi}^{(k)}, \\
\text{Std}(\hat{\psi}) &= \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\hat{\psi}^{(k)}-\frac{1}{K}\sum_{k=1}^{K}\hat{\psi}^{(k)}\right)^2}, \\
\text{MSE}(\hat{\psi}) &= \left(\frac{1}{K}\sum_{k=1}^{K}\hat{\psi}^{(k)}-\psi_0\right)^2+\frac{1}{K-1}\sum_{k=1}^{K}\left(\hat{\psi}^{(k)}-\frac{1}{K}\sum_{k=1}^{K}\hat{\psi}^{(k)}\right)^2, \\
\text{it.no} &= \frac{1}{K}\sum_{k=1}^{K}t_k.
\end{aligned}
$$

The simulated results for $(\theta_0,\theta_1,\theta_2)$ of the bivariate NPIG distribution are summarized in Table 1. The simulated results for $(\theta_0,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)$ of the bivariate NPIG regression model are listed in Table 2. From the results, it is easy to find that the estimates of the parameters are well provided and are much closer to their true values as the sample size increases; more specifically, the estimation stability and accuracy are both improved, as indicated by the decreasing values of Stds and MSEs. The population correlation coefficient and the averaged estimated value calculated with the ML estimates of parameters are also presented, which shows the relationship is completely depicted.

Table 1. ML estimate, Std and MSE for $(\theta_0,\theta_1,\theta_2)$ in bivariate NPIG distribution.

$(\theta_0,\theta_1,\theta_2)=(1.2,0.3,0.8),\ \rho=-0.2586$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n=30$			$n=50$		
θ_0	1.358641	0.442545	0.221013	1.316643	0.362304	0.144870
θ_1	0.298424	0.036809	0.001357	0.298486	0.028556	0.000818
θ_2	0.799534	0.031722	0.001007	0.801276	0.023760	0.000566
	it.no = 231, $\hat{\rho}=-0.2569$			it.no = 224, $\hat{\rho}=-0.2564$		
	$n=100$			$n=300$		
θ_0	1.243133	0.237282	0.058163	1.209933	0.121216	0.014792
θ_1	0.300323	0.020431	0.000418	0.300539	0.012035	0.000145
θ_2	0.799605	0.016497	0.000272	0.799794	0.009520	0.000091
	it.no = 213, $\hat{\rho}=-0.2586$			it.no = 197, $\hat{\rho}=-0.2588$		
$(\theta_0,\theta_1,\theta_2)=(0.5,0.5,0.6),\ \rho=-0.4390$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n=30$			$n=50$		
θ_0	0.566086	0.174171	0.034703	0.544802	0.128736	0.018580
θ_1	0.498231	0.048484	0.002354	0.499272	0.037687	0.001421
θ_2	0.601595	0.046221	0.002139	0.600750	0.035776	0.001280
	it.no = 172, $\hat{\rho}=-0.4366$			it.no = 167, $\hat{\rho}=-0.4379$		
	$n=100$			$n=300$		
θ_0	0.519031	0.085486	0.007670	0.504517	0.049257	0.002447
θ_1	0.500203	0.026148	0.000684	0.500225	0.015722	0.000247
θ_2	0.600492	0.025049	0.000628	0.599723	0.014972	0.000224
	it.no = 157, $\hat{\rho}=-0.4386$			it.no = 145, $\hat{\rho}=-0.4391$		

Table 1. *Cont.*

$(\theta_0, \theta_1, \theta_2) = (0.2, 0.8, 0.1),\ \rho = -0.7321$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n = 30$			$n = 50$		
θ_0	0.220128	0.055619	0.003499	0.211129	0.041009	0.001806
θ_1	0.800242	0.031396	0.000986	0.800302	0.022899	0.000524
θ_2	0.099971	0.018400	0.000339	0.099890	0.013603	0.000185
	it.no = 119, $\hat{\rho} = -0.7350$			it.no = 114, $\hat{\rho} = -0.7338$		
	$n = 100$			$n = 300$		
θ_0	0.205324	0.027086	0.000762	0.201148	0.015389	0.000238
θ_1	0.799645	0.016570	0.000275	0.800039	0.009393	0.000088
θ_2	0.100182	0.009942	0.000099	0.100087	0.005523	0.000031
	it.no = 107, $\hat{\rho} = -0.7326$			it.no = 97, $\hat{\rho} = -0.7323$		

Table 2. ML estimate, Std and MSE for $(\theta_0, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$ in bivariate NPIG regression model.

$\theta_0 = 0.6,\ \boldsymbol{\alpha}_1 = (1.2, 0.8, -0.5, 0.5)^\top,\ \boldsymbol{\alpha}_2 = (1.5, -2, 0.7, -0.5)^\top$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n = 50$			$n = 200$		
θ_0	0.616064	0.098125	0.009887	0.594675	0.054724	0.003023
α_{01}	1.197065	0.182278	0.033234	1.202689	0.051280	0.002637
α_{11}	0.791274	0.159555	0.025534	0.797376	0.050753	0.002583
α_{21}	−0.503090	0.172935	0.029916	−0.501269	0.052572	0.002765
α_{31}	0.496402	0.052044	0.002721	0.495028	0.022855	0.000547
α_{02}	1.493490	0.172693	0.029865	1.493448	0.059934	0.003635
α_{12}	−1.994148	0.159105	0.025349	−1.994163	0.055346	0.003097
α_{22}	0.693314	0.189516	0.035961	0.695733	0.044827	0.002028
α_{32}	−0.497199	0.051602	0.002671	−0.496026	0.023453	0.000566
	it.no = 141			it.no = 68		
	$n = 350$			$n = 500$		
θ_0	0.594409	0.040014	0.001632	0.592610	0.034314	0.001232
α_{01}	1.199199	0.032275	0.001042	1.200771	0.021140	0.000447
α_{11}	0.795187	0.036466	0.001353	0.798587	0.025180	0.000636
α_{21}	−0.500220	0.028680	0.000823	−0.500038	0.021437	0.000460
α_{31}	0.499380	0.019778	0.000392	0.496742	0.018138	0.000340
α_{02}	1.494840	0.035134	0.001261	1.495448	0.025697	0.000681
α_{12}	−1.997357	0.033092	0.001102	−1.997702	0.025642	0.000663
α_{22}	0.698411	0.028526	0.000816	0.697747	0.020047	0.000407
α_{32}	−0.498473	0.019025	0.000364	−0.497432	0.015664	0.000252
	it.no = 55			it.no = 49		

4.2. Experiments for NPGA Models

For the bivariate NPGA distribution, the parameter settings are similar with those for the NPIG models. The choice of sample size n is the same. True values of (ϕ_0, ϕ_1, ϕ_2) are chosen as $(2, 0.9, 0.2)$ and $(5, 0.4, 0.6)$ for the NPGA distribution. In the NPGA regression model, $\phi_0 = 1.2$, $\boldsymbol{\beta}_1 = (-0.9, 1.4, -0.5)^\top$, $\boldsymbol{\beta}_2 = (0.5, -0.2, -0.8)^\top$; the covariates are $\boldsymbol{v}_i = (1, v_{i1}, v_{i2})^\top$, with $v_{i1} \overset{\text{iid}}{\sim} \text{Unif}(0,1)$, and v_{i2} is randomly sampled from $\{-0.5, -0.2, 0.3, 0.6\}$ for $i = 1, \ldots, n$. Data $\{\boldsymbol{x}_i\}_{i=1}^n$ are generated from $\text{NPGA}_2(\phi_0, \phi_1, \phi_2)$ or $\text{NPGA}_2(\phi_0, \phi_{i1}, \phi_{i2})$ according to the model specified by (15). To assess the estimation performances on parameters of interest (ϕ_0, ϕ_1, ϕ_2) and $(\phi_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, we still adopt the measurements introduced in Section 4.1 for comparisons.

Tables 3 and 4 summarize the results of simulation studies for the bivariate NPGA distribution and the corresponding mean regression model. The averaged ML estimates are provided, as well as the Stds and MSEs of the estimators. It is also observed that the estimation performance is satisfactory. The values of iterative numbers indicate that the computational efficiency and convergence rate are good. All averaged estimated values of the correlation calculated with the ML estimates of parameters are close to the population correlation coefficients.

Table 3. ML estimate, Std and MSE for (ϕ_0, ϕ_1, ϕ_2) in bivariate NPGA distribution.

$(\phi_0,\phi_1,\phi_2) = (2, 0.9, 0.2),\ \rho = -0.8083$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n = 30$			$n = 50$		
ϕ_0	1.997760	0.026205	0.000692	1.998313	0.023109	0.000537
ϕ_1	0.898923	0.010484	0.000111	0.899695	0.008393	0.000071
ϕ_2	0.202299	0.019508	0.000386	0.200620	0.015605	0.000244
	it.no = 27, $\hat{\rho} = -0.8064$			it.no = 27, $\hat{\rho} = -0.8078$		
	$n = 100$			$n = 300$		
ϕ_0	1.999660	0.024147	0.000583	1.998969	0.020451	0.000419
ϕ_1	0.900010	0.005360	0.000029	0.899674	0.003072	0.000010
ϕ_2	0.201871	0.010523	0.000114	0.201171	0.006473	0.000043
	it.no = 26, $\hat{\rho} = -0.8072$			it.no = 27, $\hat{\rho} = -0.8074$		
$(\phi_0,\phi_1,\phi_2) = (5, 0.4, 0.6),\ \rho = -0.4057$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n = 30$			$n = 50$		
ϕ_0	5.001001	0.036186	0.001310	4.998937	0.032888	0.001083
ϕ_1	0.402014	0.026780	0.000721	0.402845	0.020853	0.000443
ϕ_2	0.598549	0.026008	0.000679	0.598827	0.020339	0.000415
	it.no = 21, $\hat{\rho} = -0.4069$			it.no = 21, $\hat{\rho} = -0.4074$		
	$n = 100$			$n = 300$		
ϕ_0	4.996352	0.042099	0.001786	5.001158	0.038537	0.001486
ϕ_1	0.400435	0.015158	0.000230	0.399487	0.008230	0.000068
ϕ_2	0.599306	0.015034	0.000227	0.600232	0.007946	0.000063
	it.no = 21, $\hat{\rho} = -0.4061$			it.no = 16, $\hat{\rho} = -0.4053$		

Table 4. ML estimate, Std and MSE for $(\phi_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ in bivariate NPGA regression model.

$\phi_0 = 1.2,\ \boldsymbol{\beta}_1 = (-0.9, 1.4, -0.5)^\top,\ \boldsymbol{\beta}_2 = (0.5, -0.2, -0.8)^\top$						
Parameter	Ave-MLE	Std	MSE	Ave-MLE	Std	MSE
	$n = 50$			$n = 200$		
ϕ_0	1.273002	0.200704	0.045612	1.227011	0.092474	0.009281
β_{01}	−0.863261	0.274594	0.076752	−0.860950	0.137262	0.020366
β_{11}	1.374930	0.482249	0.233192	1.341559	0.226420	0.054681
β_{21}	−0.489769	0.324336	0.105299	−0.481630	0.147419	0.022070
β_{02}	0.437918	0.290704	0.088363	0.4730029	0.140378	0.020435
β_{12}	−0.146745	0.492742	0.245631	−0.177881	0.230247	0.053503
β_{22}	−0.784241	0.340255	0.116021	−0.773613	0.157350	0.025455
		it.no = 51			it.no = 44	
	$n = 350$			$n = 500$		
ϕ_0	1.218954	0.072089	0.005556	1.210845	0.028854	0.000950
β_{01}	−0.867234	0.106863	0.012493	−0.867112	0.042890	0.002921
β_{11}	1.362347	0.170824	0.030598	1.353711	0.068964	0.006899
β_{21}	−0.486299	0.120704	0.014757	−0.486765	0.048496	0.002527
β_{02}	0.467544	0.101781	0.011413	0.476607	0.042584	0.002361
β_{12}	−0.168965	0.166065	0.028541	−0.174730	0.069102	0.005414
β_{22}	−0.772744	0.117103	0.014456	−0.777010	0.049669	0.002996
		it.no = 42			it.no = 36	

4.3. Numerical Study on Means, Variances, Covariances and Correlations

In this subsection, we provide some numerical studies on the means, variances, covariances and correlations. For the bivariate NPIG distribution, we choose the values of mean parameters as $(\theta_1, \theta_2) = (0.5, 0.6), (0.3, 0.8), (0.8, 0.1)$, combined with the value of θ_0 being $0.2, 0.4, 0.6, 0.8, 1, 3, 5, 10, 15$ and 20, respectively. The expectations, variances for two components, the covariances, and the correlation coefficients between them are presented in Table 5. For the bivariate NPGA distribution, we choose the values of the mean parameters as $(\phi_1, \phi_2) = (0.9, 0.2), (0.4, 0.7), (0.2, 0.2)$, combined with the value of ϕ_0 being the same as that of θ_0. The corresponding properties are summarized in Table 6. Note that the expectation for each component is just θ_i or ϕ_i for $i = 1, 2$ in the two distributions. "Mean1" indicates the expectation for the first component, and "Mean2" indicates the expectation for the second component. Variances and covariances are computed based on the derived formulae in the Sections 2.1 and 3.1, respectively, and "Var1" indicates the variance for the first component, "Var2" indicates the variance for the second component and "Cov" indicates the covariance between the two components. The correlation coefficient indicated by "Coef" is calculated according to its definition.

Table 5. Means, variances, covariances and correlations for the bivariate NPIG distribution.

θ_0	Mean1	Mean2	Var1	Var2	Cov	Coef
0.2	0.5	0.6	0.095957	0.095424	−0.041228	−0.430854
0.4	0.5	0.6	0.080300	0.081386	−0.035332	−0.437049
0.6	0.5	0.6	0.069668	0.071562	−0.031100	−0.440454
0.8	0.5	0.6	0.061796	0.064130	−0.027863	−0.442600
1	0.5	0.6	0.055664	0.058245	−0.025285	−0.444056
3	0.5	0.6	0.028704	0.031267	−0.013426	−0.448160
5	0.5	0.6	0.019542	0.021624	−0.009222	−0.448594
10	0.5	0.6	0.010927	0.012291	−0.005197	−0.448441
15	0.5	0.6	0.007596	0.008602	−0.003623	−0.448208
20	0.5	0.6	0.005823	0.006620	−0.002782	−0.448038
0.2	0.3	0.8	0.085745	0.066704	−0.019270	−0.254800
0.4	0.3	0.8	0.074236	0.058458	−0.016938	−0.257123
0.6	0.3	0.8	0.065981	0.052423	−0.015181	−0.258120
0.8	0.3	0.8	0.059627	0.047708	−0.013789	−0.258536
1	0.3	0.8	0.054527	0.043879	−0.012652	−0.258650
3	0.3	0.8	0.030294	0.025118	−0.007072	−0.256390
5	0.3	0.8	0.021249	0.017844	−0.004950	−0.254195
10	0.3	0.8	0.012261	0.010441	−0.002841	−0.251100
15	0.3	0.8	0.008637	0.007400	−0.001995	−0.249545
20	0.3	0.8	0.006671	0.005735	−0.001538	−0.248617
0.2	0.8	0.1	0.047708	0.020039	−0.022637	−0.732127
0.4	0.8	0.1	0.035625	0.013569	−0.016702	−0.759669
0.6	0.8	0.1	0.028700	0.010334	−0.013357	−0.775590
0.8	0.8	0.1	0.024122	0.008365	−0.011170	−0.786306
1	0.8	0.1	0.020844	0.007035	−0.009616	−0.794114
3	0.8	0.1	0.008965	0.002734	−0.004079	−0.823825
5	0.8	0.1	0.005735	0.001700	−0.002599	−0.832436
10	0.8	0.1	0.003022	0.000874	−0.001365	−0.839907
15	0.8	0.1	0.002052	0.000588	−0.000926	−0.842639
20	0.8	0.1	0.001554	0.000443	−0.000701	−0.844056

Table 6. Means, variances, covariances and correlations for the bivariate NPGA distribution.

ϕ_0	Mean1	Mean2	Var1	Var2	Cov	Coef
0.2	0.9	0.2	0.030000	0.080000	−0.031973	−0.652641
0.4	0.9	0.2	0.018000	0.053333	−0.022113	−0.713705
0.6	0.9	0.2	0.012857	0.040000	−0.016892	−0.744877
0.8	0.9	0.2	0.010000	0.032000	−0.013669	−0.764106
1	0.9	0.2	0.008182	0.026667	−0.011480	−0.777230
3	0.9	0.2	0.002903	0.010000	−0.004421	−0.820416
5	0.9	0.2	0.001765	0.006154	−0.002738	−0.830996
10	0.9	0.2	0.000891	0.003137	−0.001404	−0.839489
15	0.9	0.2	0.000596	0.002105	−0.000944	−0.842438
20	0.9	0.2	0.000448	0.001584	−0.000711	−0.843936
0.2	0.4	0.7	0.180000	0.163333	−0.052559	−0.306532
0.4	0.4	0.7	0.144000	0.133636	−0.045743	−0.329747
0.6	0.4	0.7	0.120000	0.113077	−0.039804	−0.341705
0.8	0.4	0.7	0.102857	0.098000	−0.034977	−0.348383
1	0.4	0.7	0.090000	0.086471	−0.031078	−0.352290
3	0.4	0.7	0.040000	0.039730	−0.014237	−0.357122
5	0.4	0.7	0.025714	0.025789	−0.009141	−0.354974
10	0.4	0.7	0.013585	0.013738	−0.004805	−0.351719
15	0.4	0.7	0.009231	0.009363	−0.003256	−0.350212
20	0.4	0.7	0.006990	0.007101	−0.002462	−0.349366
0.2	0.2	0.2	0.128000	0.080000	−0.026476	−0.261637
0.4	0.2	0.2	0.106667	0.053333	−0.022288	−0.295506
0.6	0.2	0.2	0.091429	0.040000	−0.019121	−0.316183
0.8	0.2	0.2	0.080000	0.032000	−0.016708	−0.330219
1	0.2	0.2	0.071111	0.026667	−0.014822	−0.340368
3	0.2	0.2	0.033684	0.010000	−0.006910	−0.376481
5	0.2	0.2	0.022069	0.006154	−0.004494	−0.385587
10	0.2	0.2	0.011852	0.003137	−0.002395	−0.392746
15	0.2	0.2	0.008101	0.002105	−0.001632	−0.395167
20	0.2	0.2	0.006154	0.001584	−0.001238	−0.396378

5. Applications

We obtain the cortical thickness of 41 patients with schizophrenia and 40 healthy controls from [18]. Structural magnetic resonance imaging scans obtained from the participants were processed using Freesurfer. Cortical thickness was parcellated by the Destrieux atlas [19] to provide 148 brain regions and estimated by the standard procedures described in [20]. Regional Ethics Committees (Nottinghamshire & Derbyshire) approved the study and all participants provided written informed consent. We aim to analyze the negative co-varying pairs of regions for investigating the influence of schizophrenia on the cortical thickness between controls and patients. The negative correlation pairs among 148 dimensions of data based on Pearson correlation coefficients are shown in Figure 3. The locations of squares marked with red circles are our following examples in subsections. The descriptions of used data are given in the Supplementary Material.

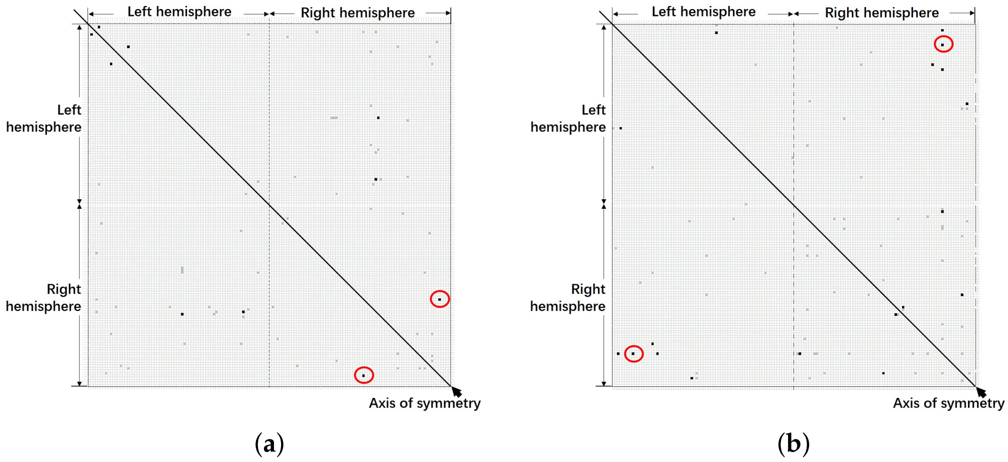

Figure 3. Negative correlations between the thickness of 74 different sulco-gyral cortical units in each hemisphere of (**a**) patients; (**b**) controls. (Each square represent a negative correlation of corresponding units under the Spearman significance test, where the p-values of black ones are $p < 0.01$ and gray ones are $0.01 \leqslant p < 0.05$).

5.1. Lateral and Suborbital Sulcus

We take the thickness difference of the horizontal ramus of the anterior segment of the lateral sulcus (X_1) and suborbital sulcus (X_2) in the right hemisphere as $\mathbf{x} = (X_1, X_2)^\top$. Based on the significant, negative correlation between X_1 and X_2 in patients and a positive correlation in controls, we fit the patient and control groups data into the four different distributions, where bivariate PIG and PGA models are derived from [16]. The 95% CIs and Stds of *parameters* (Par.) are calculated by bootstrap re-samplings.

The results are shown in Table 7. We note that the bivariate PGA and NPGA distributions perform better under the model selection criterion. The ML estimates of the mean parameters in two distributions between the patient and control groups fall on the boundary of the corresponding parameters' confidence intervals of the other group, respectively. This implies the different cortical thinness between the two groups. Although the two regions inhibited each other, their thicknesses in the patients were significantly reduced compared with the control group. With the weakening of the compensatory behaviors of the patients' cortical thickness in these areas, the negatively correlated pair different from the control group was produced, which is consistent with the clinical manifestations of changes in the cerebral cortex of schizophrenia.

Table 7. *ML estimates* (MLEs), stds and CIs for the thickness of X_1 and X_2 (Section 5.1) between controls and patients in two distributions with model selection criterion AIC and BIC.

Par.	**Controls**			**Patients**		
	MLE	**CI**	**Std**	**MLE**	**CI**	**Std**
	Bivariate PIG distribution			Bivariate NPIG distribution		
θ_0	0.4296	[0.2744, 0.7503]	0.1242	0.4837	[0.3158, 0.8444]	0.1370
θ_1	0.4664	[0.3821, 0.5473]	0.0416	0.4249	[0.3504, 0.5042]	0.0399
θ_2	**0.5029**	**[0.4213, 0.5877]**	0.0427	**0.4209**	**[0.3438, 0.4923]**	0.0378
	AIC = 18.0480; BIC = 23.1147			AIC = 1.4008; BIC = 6.5415		
	Bivariate PGA distribution			Bivariate NPGA distribution		
ϕ_0	1.4540	[1.0527, 2.0820]	0.2606	1.5477	[1.2360, 1.8403]	0.1418
ϕ_1	**0.5071**	**[0.4317, 0.5812]**	0.0373	**0.4294**	**[0.3659, 0.5023]**	0.0359
ϕ_2	**0.5161**	**[0.4454, 0.5881]**	0.0383	**0.4059**	**[0.3394, 0.4676]**	0.0319
	AIC = 2.0847; BIC = 7.1514			**AIC = −9.9734; BIC = −4.8327**		

To further study the thickness changes in patients, we introduce two common covariates for the two groups: w_{i1}, v_{i1} are the logarithm transformation of the age in years, and w_{i2}, v_{i2} are gender (male=0, female=1). Based on (9) and (15), we have the following regression models:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \alpha_{j0} + \alpha_{j1}w_{i1} + \alpha_{j2}w_{i2} \quad \text{and} \quad \log\left(\frac{\phi_{ij}}{1-\phi_{ij}}\right) = \beta_{j0} + \beta_{j1}v_{i1} + \beta_{j2}v_{i2},$$

where $i = 1, \ldots, n$, and $j = 1, 2$. Table 8 listed the results by fitting the data of patients and controls with the four corresponding regression models. From the 95% bootstrap CIs, we know that there are significantly negative relationships between $\mathbf{x}$ and log(age) only in controls. The other difference is focused on the influence of gender to X_2, which is significantly positive to X_2 in controls and indicates irrelevant to X_2 in patients. Based on the results, we see that the mutual inhibition between X_1 and X_2 is mainly due to the opposite compensation of X_1 with gender and X_2 with age in patients.

Summarizing the results in Tables 7 and 8, we think the causes of these features, such as the effect of drug dose on different genders or the variable changes of brain regions with the durations, still need to be further explored.

Table 8. *ML estimates* (MLEs), stds and CIs for the thickness of X_1 and X_2 (Section 5.1) between controls and patients in two regression models with selection criterion AIC and BIC.

Par.	**Controls**			**Patients**		
	MLE	**CI**	**Std**	**MLE**	**CI**	**Std**
	Bivariate PIG mean regression			Bivariate NPIG mean regression		
θ_0	0.5588	[0.3899, 1.0599]	0.1789	0.5355	[0.4019, 1.0499]	0.1579
α_{10}	4.9412	[1.7141, 8.1186]	1.6109	2.6217	[−1.0682, 6.7279]	1.8809
α_{11}	**−1.5045**	**[−2.4762, −0.6009]**	0.4686	**−0.8339**	**[−2.0480, 0.2186]**	0.5428
α_{12}	**0.6536**	**[0.0220, 1.4045]**	0.3506	**−0.0466**	**[−0.7188, 0.7475]**	0.3654
α_{20}	4.0251	[0.7731, 7.6577]	1.6832	0.4262	[−3.5984, 4.4486]	1.8909
α_{21}	**−1.1637**	**[−2.2148, −0.2369]**	0.4891	**−0.2414**	**[−1.3743, 0.8807]**	0.5476
α_{22}	0.1555	[−0.4933, 0.8664]	0.3469	0.3114	[−0.5136, 0.9808]	0.3816
	AIC = 15.0705; BIC = 26.8926			AIC = 4.6480; BIC = 16.6430		
	Bivariate PGA mean regression			Bivariate NPGA mean regression		
ϕ_0	1.5535	[1.1976, 2.2985]	0.2744	1.6295	[1.2301, 2.5405]	0.3288
β_{10}	6.1568	[4.3330, 7.6466]	0.7389	1.4482	[−1.7922, 5.0879]	1.7278
β_{11}	**−1.7962**	**[−2.2712, −1.2685]**	0.2251	**−0.4871**	**[−1.5415, 0.4178]**	0.4990
β_{12}	**0.4822**	**[0.0032, 1.0710]**	0.2857	**−0.1693**	**[−0.7979, 0.4999]**	0.3374
β_{20}	4.8569	[3.3534, 6.5856]	0.7110	1.8815	[−1.3326, 4.7364]	1.5787
β_{21}	**−1.3976**	**[−1.9201, −0.9069]**	0.2127	**−0.6746**	**[−1.5283, 0.2427]**	0.4539
β_{22}	0.2720	[−0.2365, 0.9072]	0.2990	0.3070	[−0.2977, 0.8459]	0.3052
	AIC = 4.2536; BIC = 16.0757			**AIC = −6.7527; BIC = 5.2424**		

5.2. Cingulate Gyrus and Lateral Occipito-Temporal Sulcus

In this subsection, we analyze regions in different hemispheres. The left posterior-dorsal part of the cingulate gyrus (X_1) and right lateral occipito-temporal sulcus (X_2) are taken as $\mathbf{x} = (X_1, X_2)^\top$. Based on the significant negative correlation between X_1 and X_2 in controls and a positive correlation in patients, we fit the data of patients and controls into the four distributions, respectively. Similar to Section 5.1, we also consider covariates in four corresponding mean regression models. The correlation information from the samples implies that the data are not related to gender, so we only consider one covariate log(age). The results are summarized in Tables 9 and 10.

Based on the selection criterion, bivariate PGA and NPGA distributions and models show better performance. The mean cortical thickness differences between the two groups are significant. Similar to the results of medical research, the thicknesses in patients were consistently smaller than those of the controls. In the mean regression models, the influences of log(age) to $\mathbf{x}$ are quite similar in the two groups, which is obviously different from the results shown in the previous subsection. The slight difference between the two groups is the influence of log(age) to X_2, which is not significant in controls.

Combining the results in the two tables, we find the thickness difference between the two groups due to the loss of compensatory behaviors in the patients and raise a reasonable doubt that the duration of patients may cause the loss.

Table 9. *ML estimates* (MLEs), stds and CIs for the thickness of X_1 and X_2 (Section 5.2) between controls and patients in two distributions with model selection criterion AIC and BIC.

Par.	Patients			Controls		
	MLE	**CI**	**Std**	**MLE**	**CI**	**Std**
	Bivariate PIG distribution			Bivariate NPIG distribution		
θ_0	0.6307	[0.4001, 1.1844]	0.1903	1.6761	[1.1946, 2.6281]	0.3711
θ_1	**0.4010**	**[0.3243, 0.4790]**	0.0396	**0.5031**	**[0.4390, 0.5676]**	0.0325
θ_2	**0.4552**	**[0.3859, 0.5331]**	0.0386	**0.5215**	**[0.4614, 0.5843]**	0.0320
	AIC = 4.1243; BIC = 9.2650			AIC = −29.0117; BIC = −23.9451		
	Bivariate PGA distribution			Bivariate NPGA distribution		
ϕ_0	1.9787	[1.5283, 2.8760]	0.3586	3.0165	[2.8431, 3.1387]	0.0711
ϕ_1	**0.3967**	**[0.3249, 0.4524]**	0.0334	**0.5000**	**[0.4531, 0.5639]**	0.0267
ϕ_2	**0.4650**	**[0.4009, 0.5221]**	0.0334	**0.5374**	**[0.4771, 0.5886]**	0.0273
	AIC = −12.8517; BIC = −7.7109			**AIC = −43.0450; BIC = −37.9783**		

Table 10. *ML estimates* (MLEs), stds and CIs for the thickness of X_1 and X_2 (Section 5.2) between controls and patients in two regression models with selection criterion AIC and BIC.

Par.	Patients			Controls		
	MLE	**CI**	**Std**	**MLE**	**CI**	**Std**
	Bivariate PIG mean regression			Bivariate NPIG mean regression		
θ_0	0.7825	[0.5220, 1.3723]	0.2194	1.8085	[1.3285, 2.8110]	0.3876
α_{10}	3.4074	[0.3196, 6.1716]	1.5023	3.2295	[0.3851, 6.4295]	1.5459
α_{11}	**−1.0921**	**[−1.9158, −0.2041]**	0.4349	**−0.9194**	**[−1.8404, −0.0923]**	0.4459
α_{20}	3.6225	[0.4132, 6.5227]	1.5104	−0.1836	[−2.8446, 2.9085]	1.4255
α_{21}	**−1.0925**	**[−1.9570, −0.1878]**	0.4373	**0.0806**	**[−0.8328, 0.8209]**	0.4091
	AIC = 0.0722; BIC = 8.6400			AIC = −31.1357; BIC = −22.6913		
	Bivariate PGA mean regression			Bivariate NPGA mean regression		
ϕ_0	2.1331	[1.6301, 3.0572]	0.3632	3.2817	[2.5071, 4.8420]	0.6160
β_{10}	3.0007	[1.0820, 4.6351]	0.7655	3.2619	[0.4007, 6.0649]	1.4288
β_{11}	**−0.9823**	**[−1.4382, −0.4542]**	0.2243	**−0.9366**	**[−1.7524, −0.1013]**	0.4134
β_{20}	2.6646	[0.8734, 4.2967]	0.8151	−0.0168	[−2.9359, 2.9117]	1.5148
β_{21}	**−0.8051**	**[−1.3183, −0.3098]**	0.2397	**0.0486**	**[−0.8062, 0.8667]**	0.4337
	AIC = −16.1259; BIC = −7.5581			**AIC = −46.5825; BIC = −38.1381**		

6. Conclusions, Limitations, and Future Research

In this paper, we proposed models that fit bivariate negatively correlated continuous proportional data for the first time. Based on the equal-dispersed IG distribution and the gamma distribution with a single parameter, we developed the bivariate NPIG and NPGA distributions. Models with covariates are also considered by formulating the mean regression models based on the two new distributions. Moreover, we provide efficient methods for parameter estimations of the four different models, respectively. The N-EM algorithm aided by the gradient descent algorithms based on Jensen's inequality is used to overcome the difficulties in calculating ML estimates of parameters. For readers interested in algorithms, we recommend reading [21,22]. In Section 5, we used two different criteria to evaluate the models. We study the negative correlation pairs that increase with the decrease in compensation behaviors, and the information obtained from the main research is consistent with our previous findings with the same dataset [23]. Moreover, we propose the hypotheses of the causes of them based on the results, which needs further medical exploration. According to our analysis of the cortical thickness of schizophrenic patients and the control group, we verified the compensatory nature of cortical thickness in schizophrenic patients and found that it was negatively correlated with age. If you want to use the original data and R code of this article for your research, please contact the corresponding author by email. In addition, the use of original data should be agreed with the data collection team.

There are other topics worthy of further research beyond this paper. We only considered the mean regression models for the proposed distributions and did not consider the mode regressions as there are no closed forms for their modes. Similarly, there are quantile regressions. To better interpret the data, we hope to explore the mode regression models and have already constructed a new model with an explicit expression of the mode. The construction structure is $1/(1+Y)$ similar to (1). Moreover, linear constructions, such as SR (1), to set models with arbitrary positive or negative correlations are difficult to achieve. We consider changing independent $\{Y_j\}_{j=1}^{2}$ to a bivariate correlated vector $\mathbf{y} = (Y_1, Y_2)^\top$ and then the correlation structure between components based on the construction (1) more flexible. Moreover, the Copula method may be one feasible way, or mixture models could be considered by combining PIG with NPIG and PGA with NPGA. Finally, the exact tests in the bivariate NPIG and NPGA models for one sample and multiple samples are also our interests. They can help us research the significance of differences.

Supplementary Materials: The following are available at https://www.mdpi.com/article/10.3390/math10030353/s1, A.1: Solution to $\nabla Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbf{0}_3$, A.2: Calculations for $\nabla G_1(\boldsymbol{\theta}_{-0}|\boldsymbol{\theta}^{(t)})$, A.3: Calculations for $\nabla G_2(\boldsymbol{\vartheta}_{-0}|\boldsymbol{\vartheta}^{(t)})$, A.4: Calculations for $\nabla \ell_3(\boldsymbol{\phi}|Y_{\mathrm{obs}_3})$, A.5: Calculations for $\nabla Q_3(\boldsymbol{\phi}|\boldsymbol{\phi}^{(t)})$, A.6: Calculations for $\nabla \ell_4(\boldsymbol{\varphi}|Y_{\mathrm{obs}_4})$, B.1: Lateral and suborbital sulcus, B.2: Cingulate gyrus and lateral occipito-temporal sulcus.

Author Contributions: Y.S. contributed to the methodology, data curation, visualization and writing—original draft preparation; G.T. contributed to the conceptualization and supervision; S.G. contributed to the data curation; L.S. contributed to the validation; C.Z. contributed to the software, visualization and writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant numbers 11771199, 11801380, 12071124 and 12171225. Prof. Shu's work was funded in part by the Science and Technology Development Fund, Macau SAR (File No. FDCT/0033/2020/A1), and by the Research Committee under the grant MYRG2020-00081-FBA.

Institutional Review Board Statement: For ethics recruitment of participants and data collection has been described previously (and was approved by National Research Ethics Committee, Nottinghamshire (NHS REC Ref: 10/H0406/49).

Informed Consent Statement: All participants provided written informed consent.

Data Availability Statement: The use of original data should be agreed with the data collection team.

Acknowledgments: The authors wish to express their sincere appreciation to all those who made suggestions for improvements to this paper. We acknowledge the Centre for Translational Neuroimaging in Mental Health, University of Nottingham (Peter F Liddle, Lena Palaniyappan) and the Sir Peter Mansfield Imaging Centre, University of Nottingham (Penny Gowland) who provided the imaging data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Some Properties of New Distributions

Similar to [16], when the (MGF) of Y_0 and Y_j ($j = 1, 2$) exist, we obtain the expectation and variance of X_j with the SR (1) as follows:

$$E(X_1) = -\int_0^\infty \frac{\mathrm{d}M_{Y_1}(-t)}{\mathrm{d}t} M_{Y_0}(-t)\,\mathrm{d}t, \tag{A1}$$

$$E(X_2) = 1 + \int_0^\infty \frac{\mathrm{d}M_{Y_2}(-t)}{\mathrm{d}t} M_{Y_0}(-t)\,\mathrm{d}t, \tag{A2}$$

$$\mathrm{Var}(X_1) = \int_0^\infty t \cdot \frac{\mathrm{d}^2 M_{Y_1}(-t)}{\mathrm{d}t^2} M_{Y_0}(-t)\,\mathrm{d}t - [E(X_1)]^2 \quad \text{and} \tag{A3}$$

$$\mathrm{Var}(X_2) = \int_0^\infty t \cdot \frac{\mathrm{d}^2 M_{Y_0}(-t)}{\mathrm{d}t^2} M_{Y_2}(-t)\,\mathrm{d}t - [E(X_2)]^2, \tag{A4}$$

respectively, where $M_Y(t)$ denotes the MGF of Y. The covariance of X_1 and X_2 is given by

$$\mathrm{Cov}(X_1, X_2) = -\int_0^\infty \int_0^\infty \frac{\mathrm{d}M_{Y_1}(-t)}{\mathrm{d}t} \cdot \frac{\mathrm{d}M_{Y_2}(-s)}{\mathrm{d}s} \cdot \Delta(t,s)\,\mathrm{d}t\,\mathrm{d}s, \tag{A5}$$

where $\Delta(t,s) = M_{Y_0}(-t-s) - M_{Y_0}(-t) \cdot M_{Y_0}(-s)$. It is easy to verify that $\mathrm{Cov}(X_1, X_2) \leqslant 0$.

Appendix B. The Construction of the N-EM Algorithm

Appendix B.1. ML Estimation of Parameters in the Bivariate NPIG Distribution

We develop the N-EM algorithm by introducing the integral version of Jensen's inequality:

$$H\left[\int_{\mathbb{X}} \tau(x) \cdot g(x)\,\mathrm{d}x\right] \geqslant \int_{\mathbb{X}} H[\tau(x)] \cdot g(x)\,\mathrm{d}x, \tag{A6}$$

where $H(\cdot)$ is a concave function, $\tau(\cdot)$ is a real-valued function and $g(\cdot)$ is a pdf defined on $\mathbb{X} \subseteq \mathbb{R}$ [7]. Then, we have

$$\begin{aligned}
\log\left[\int_0^\infty h(s|\boldsymbol{x}_i, \boldsymbol{\theta})\,\mathrm{d}s\right] &= \log\left[\int_0^\infty \frac{h_1(s|\boldsymbol{x}_i, \boldsymbol{\theta})}{g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})} \cdot g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}s\right] \\
&\overset{\text{(A6)}}{\geqslant} \int_0^\infty \log\left[\frac{h_1(s|\boldsymbol{x}_i, \boldsymbol{\theta})}{g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})}\right] \cdot g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}s \\
&= c_{i1}^{(t)} + \int_0^\infty \log[h_1(s|\boldsymbol{x}_i, \boldsymbol{\theta})] \cdot g_1(s|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}s \\
&= c_{i2}^{(t)} - \frac{1}{2} B_1(\boldsymbol{x}_i, \boldsymbol{\theta}^{(t)}) \cdot b_1(\boldsymbol{x}_i, \boldsymbol{\theta}),
\end{aligned} \tag{A7}$$

where $\{c_{ik}^{(t)}\}_{k=1}^2$ are constants free from $\boldsymbol{\theta}$. Based on (A7), we derive the surrogate function $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ shown in (6). By the MM principle [24–26], given $\boldsymbol{\theta}^{(t)}$, the $(t+1)$-th approximation is updated by $\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta} \in \mathbb{R}_+^3} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$. Obviously, $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ minorizes $\ell_1(\boldsymbol{\theta}|Y_{\mathrm{obs}_1})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$.

Appendix B.2. ML Estimation of Parameters in the Bivariate NPGA Distribution

To apply the N-EM algorithm, we need to construct the surrogate function $Q_3(\boldsymbol{\phi}|\boldsymbol{\phi}^{(t)})$. By using Jensen's inequality, we obtain

$$\log\Gamma(a) \geqslant \int_0^\infty \log\left[\frac{s^{a-1}\mathrm{e}^{-s}}{g_2(s|a^{(t)})}\right]\cdot g_2(s|a^{(t)})\,\mathrm{d}s = c^{(t)} + a\frac{\Gamma'(a^{(t)})}{\Gamma(a^{(t)})},$$

where $g_2(s|a) = s^{a-1}\mathrm{e}^{-s}/\Gamma(a)$ is the pdf of Gamma$(a,1)$, $c^{(t)}$ is a constant and $\Gamma'(a^{(t)}) = \frac{\mathrm{d}\Gamma(a)}{\mathrm{d}a}\big|_{a=a^{(t)}}$. In addition, by the supporting hyperplane inequality, we have

$$-\log\Gamma(a) \geqslant 1 - \log\Gamma(a^{(t)}) - \frac{\Gamma(a)}{\Gamma(a^{(t)})}.$$

With the two inequalities we obtained, the surrogate function is

$$\begin{aligned} Q_3(\boldsymbol{\phi}|\boldsymbol{\phi}^{(t)}) \quad = \quad & n\left\{\left(\frac{\phi_0\phi_1}{1-\phi_1}+\frac{\phi_0}{\phi_2}\right)\frac{\Gamma'\left(\frac{\phi_0^{(t)}\phi_1^{(t)}}{1-\phi_1^{(t)}}+\frac{\phi_0^{(t)}}{\phi_2^{(t)}}\right)}{\Gamma\left(\frac{\phi_0^{(t)}\phi_1^{(t)}}{1-\phi_1^{(t)}}+\frac{\phi_0^{(t)}}{\phi_2^{(t)}}\right)} - \frac{\Gamma(\phi_0)}{\Gamma(\phi_0^{(t)})} - \frac{\Gamma\left(\frac{\phi_0\phi_1}{1-\phi_1}\right)}{\Gamma\left(\frac{\phi_0^{(t)}\phi_1^{(t)}}{1-\phi_1^{(t)}}\right)}\right. \\ & \left. - \frac{\Gamma\left(\frac{\phi_0}{\phi_2}-\phi_0\right)}{\Gamma\left(\frac{\phi_0^{(t)}}{\phi_2^{(t)}}-\phi_0^{(t)}\right)}\right\} + \frac{\phi_0\phi_1}{1-\phi_1}\sum_{i=1}^n\log\left(\frac{x_{i1}}{1-x_{i1}}\right) + \left(\frac{\phi_0}{\phi_2}-\phi_0\right)\sum_{i=1}^n\log\left(\frac{1-x_{i2}}{x_{i2}}\right) \\ & - \left(\frac{\phi_0\phi_1}{1-\phi_1}+\frac{\phi_0}{\phi_2}\right)\sum_{i=1}^n\log\left(1+\frac{x_{i1}}{1-x_{i1}}+\frac{1-x_{i2}}{x_{i2}}\right) + c_3^{(t)}, \end{aligned}$$

where $c_3^{(t)}$ is a constant.

References

1. Ferrari, S.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
2. Simas, A.B.; Barreto-Souza, W.; Rocha, A.V. Improved estimators for a general class of beta regression models. *Comput. Stat. Data Anal.* **2010**, *54*, 348–366. [CrossRef]
3. Ferrari, S.L.P.; Pinheiro, E.C. Improved likelihood inference in beta regression. *J. Stat. Comput. Simul.* **2011**, *81*, 431–443. [CrossRef]
4. Kieschnick, R.; McCullough, B.D. Regression analysis of variates observed on $(0,1)$: Percentages, proportions and fractions. *Stat. Model.* **2003**, *3*, 193–213. [CrossRef]
5. Zhang, P.; Qiu, Z.G. Regression analysis of proportional data using simplex distribution. *Sci. Sin. Math.* **2014**, *44*, 89–104. (In Chinese) [CrossRef]
6. Lijoi, A.; Mena, R.H.; Prünster, I. Hierarchical mixture modeling with normalized inverse–Gaussian priors. *J. Am. Stat. Assoc.* **2005**, *100*, 1278–1291. [CrossRef]
7. Liu, P.Y.; Tian, G.L.; Yuen, K.C.; Zhang, C.; Tang, M.L. Proportional inverse Gaussian distribution: A new tool for analyzing continuous proportional data. *Aust. N. Z. J. Stat.* 2021, *in press*. [CrossRef]
8. Wang, C.; Tu, D. A bootstrap semiparametric homogeneity test for the distributions of multigroup proportional data, with applications to analysis of quality of life outcomes in clinical trials. *Stat. Med.* **2020**, *39*, 1715–1731. [CrossRef]
9. Connor, R.J.; Mosimann, J.E. Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Am. Stat. Assoc.* **1969**, *64*, 194–206. [CrossRef]
10. Campbell, G.; Mosimann, J.E. Multivariate methods for proportional shape. *ASA Proc. Sect. Stat. Graph.* **1987**, *1*, 10–17.
11. Gueorguieva, R.; Rosenheck, R.; Zelterman, D. Dirichlet component regression and its applications to psychiatric data. *Comput. Stat. Data Anal.* **2008**, *52*, 5344–5355. [CrossRef] [PubMed]
12. Zhang, B. On Compositional Data Modeling and Its Biomedical Applications. Ph.D. Dissertation, Columbia University, New York, NY, USA, 2013.
13. Morais, J.; Thomas-Agnan, C.; Simioni, M. Using compositional and Dirichlet models for market share regression. *J. Appl. Stat.* **2018**, *45*, 1670–1689. [CrossRef]
14. Cepeda-Cuervo, E.; Achcar, J.A.; Lopera, L.G. Bivariate beta regression models: joint modeling of the mean, dispersion and association parameters. *J. Appl. Stat.* **2014**, *41*, 677–687. [CrossRef]

15. Petterle, R.R.; Laureano, H.A.; da Silva, G.P.; Bonat, W.H. Multivariate generalized linear mixed models for continuous bounded outcomes: Analyzing the body fat percentage data. *Stat. Methods Med. Res.* **2021**, *30*, 2619–2633. [CrossRef]
16. Sun, Y.; Tian, G.L.; Guo, S.X.; Liu, P.Y. New models for analyzing positively correlated continuous proportional data: A cortical thickness study for schizophrenia. *Biometrics* 2021, (*submitted*).
17. Tweedie, M.C.K. Statistical properties of inverse Gaussian distributions. I. *Ann. Math. Stat.* **2018**, *28*, 362–377. [CrossRef]
18. Palaniyappan, L.; Liddle, P.F. Diagnostic discontinuity in psychosis: A combined study of cortical gyrification and functional connectivity. *Schizophr. Bull.* **2014**, *40*, 675–684. [CrossRef]
19. Destrieux, C.; Fischl, B.; Dale, A.; Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **2010**, *53*, 1–15. [CrossRef]
20. Fischl, B.; Dale, A.M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11050–11055. [CrossRef]
21. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2007.
22. Lange, K. *MM Optimization Algorithms*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2016.
23. Guo, S.X.; Palaniyappan, L.; Liddle, F.P.; Feng, J.F. Dynamic cerebral reorganization in the pathophysiology of schizophrenia: A MRI-derived cortical thickness study. *Psychol. Med.* **2016**, *46*, 2201–2214.
24. Lange, K.; Hunter, D.R.; Yang, I. Optimization transfer using surrogate objective functions (with discussions). *J. Comput. Graph. Stat.* **2000**, *9*, 1–20.
25. Hunter, D.R.; Lange, K. A tutorial on MM algorithms. *Am. Stat.* **2004**, *58*, 30–37. [CrossRef]
26. Tian, G.L.; Huang, X.F.; Xu, J.F. An assembly and decomposition approach for constructing separable minorizing functions in a class of MM algorithms. *Stat. Sin.* **2004**, *29*, 961–982. [CrossRef]