Fast QLB algorithm and hypothesis tests in logistic model for ophthalmologic bilateral correlated data

Yi-Qi LIN[†,§], Yu-Shun ZHANG[†§], Guo-Liang TIAN[*,] and Chang-Xing MA[‡]

[†] *Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China*

[‡] *Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, USA*

[§] Equally contributed

SMMR version

12 June 2018

Outline

1. Introduction

In ophthalmologic (or otolaryngologic) studies, data collection not only includes the information from each subject but also contains the observations from each of two eyes (or two ears).

Although the data from different subjects are independent, the observations from the two eyes (or ears) are generally positively correlated. Therefore, any statistical models and methods that ignore this feature of dependency could lead to incorrect conclusions.

1.1 Donner's model

Suppose that we want to compare n different individuals on some ocular disease. Let $Z_{ik} = 1$ if the k-th eye of the i-th individual is diseased, and $Z_{ik} = 0$ otherwise for $i = 1, \ldots, n$ and $k = 1, 2$. The parametric model proposed by Donner (1989) is

$$\begin{cases} \Pr(Z_{ik} = 1) & = & \pi_i, & i = 1, \ldots, n, & k = 1, 2, \\ \text{Corr}(Z_{i1}, Z_{i2}) & = & \rho, & 0 \leqslant \rho \leqslant 1, \end{cases}$$

where π_i denotes the disease probability for the i-th individual and ρ is assumed to be the common correlation coefficient between two eyes of the same individual across all patients. It is easy to derive that the disease probabilities for none, one or both eyes are

$$p_{i0} = (1-\pi_i)(\rho\pi_i - \pi_i + 1),\ p_{i1} = 2\pi_i(1-\rho)(1-\pi_i) \text{ and } p_{i2} = \pi_i^2 + \rho\pi_i(1-\pi_i), \tag{1.1}$$

respectively.

1.1 Donner's model

We define $Y_i = 0, 1, 2$ if the i-th individual has none, one or both defective eyes.

Log-likelihood function

$$\begin{aligned}\ell(\boldsymbol{\pi}, \rho) \quad = \quad & \sum_{i=1}^{n} \Big\{ I(Y_i = 0) \log \left[(1 - \pi_i)(\rho\pi_i - \pi_i + 1) \right] \\ & + I(Y_i = 1) \log \left[2\pi_i (1 - \rho)(1 - \pi_i) \right] \\ & + I(Y_i = 2) \log \left[\pi_i^2 + \rho\pi_i (1 - \pi_i) \right] \Big\}, \end{aligned} \tag{1.2}$$

where $I(\cdot)$ denotes the indicator function.

1.2 Model formulation

To our best knowledge, there is few paper exploring the relationship between the disease probability/rate and some covariates (e.g., age, weight and so on) under Donner's model.

Therefore, in this article, we focus on investigating the relationship between the disease probability and covariates via the logistic regression.

1.2 Model formulation

Let $\boldsymbol{x}_{(i)} = (1, x_{i1}, \ldots, x_{i,q-1})^\top$ be a vector of covariates of interest for $i = 1, \ldots, n$. To link the disease probability π_i with these covariates, we consider the following logistic model:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}, \tag{1.3}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{q-1})^\top$ is the vector of unknown regression coefficients. From (1.3), it is easy to obtain

$$\pi_i = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} = \frac{1}{1 + \mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \quad \text{and} \quad 1 - \pi_i = \frac{1}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}, \tag{1.4}$$

so that the log-likelihood function (1.2) can be rewritten as:

1.2 Model formulation

Log-likelihood function in logistic model

$$
\begin{aligned}
\ell(\boldsymbol{\beta}, \rho) \quad = \quad & \sum_{i=1}^{n} \Bigg\{ I(Y_i = 0) \log \left[\frac{1}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \left(\frac{\rho - 1}{1 + \mathrm{e}^{-\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} + 1 \right) \right] \\
& + I(Y_i = 1) \log \left[\frac{2(1-\rho)}{1 + \mathrm{e}^{-\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \cdot \frac{1}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \right] \\
& + I(Y_i = 2) \log \left[\left(\frac{1}{1 + \mathrm{e}^{-\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \right)^2 + \frac{\rho}{1 + \mathrm{e}^{-\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \cdot \frac{1}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^{\top} \boldsymbol{\beta}}} \right] \Bigg\} \qquad (1.5)
\end{aligned}
$$

2.1 MLEs of parameters via a new MM algorithm

Given $\boldsymbol{\pi}$, the MLE of ρ can be obtained by solving the solution to the following equation

$$0 = \frac{\partial \ell(\boldsymbol{\pi}, \rho)}{\partial \rho} = \sum_{i=1}^{n} \left[\frac{\pi_i I(Y_i = 0)}{\pi_i(\rho - 1) + 1} + \frac{I(Y_i = 1)}{\rho - 1} + \frac{(1 - \pi_i) I(Y_i = 2)}{\pi_i + \rho(1 - \pi_i)} \right].$$

That is, given $\boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}$ and the t-th approximation $\rho^{(t)}$, the $(t+1)$-th approximation of the MLE $\hat{\rho}$ can be calculated via the Newton–Raphson iteration:

$$\rho^{(t+1)} = \rho^{(t)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho^2} \right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho}. \tag{2.1}$$

2.1 MLEs of parameters via a new MM algorithm

Given ρ, to calculate the MLEs of $\boldsymbol{\beta}$ in (1.5), we have noted that the well-known NR algorithm and the Fisher scoring algorithms are not available for the following reasons:

(i) The log-likelihood function $\ell(\boldsymbol{\beta}, \rho)$ in (1.5) is not concave with respect to $\boldsymbol{\beta}$, and it is too complicated.

(ii) It requires tedious calculations of the inverse Hessian matrix at each iteration, especially when sample size n is quite large.

(iii) It is sensitive to the initial values, it does not converge if a poor initial value is chosen.

(iv) The likelihood function does not necessarily increase at each iteration for the NR algorithm.

2.1 MLEs of parameters via a new MM algorithm

To overcome those difficulties, we utilize the AD approach of Tian, Huang & Xu (2018) to propose a new MM algorithm by constructing a surrogate function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$, which minorizes $\ell(\boldsymbol{\beta}, \rho)$ in (1.5).
To this end, we first introduce the following important inequality:

AD approach

$$h(\boldsymbol{\alpha}^\top \boldsymbol{z}) \geqslant \sum_{i=1}^{n} \frac{\alpha_i z_i^{(t)}}{\boldsymbol{\alpha}^\top \boldsymbol{z}^{(t)}} h\left(\frac{\boldsymbol{\alpha}^\top \boldsymbol{z}^{(t)}}{z_i^{(t)}} z_i\right), \tag{2.2}$$

where $h(\cdot)$ is an arbitrary concave function, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$, $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$ and $\boldsymbol{z}^{(t)} = (z_1^{(t)}, \ldots, z_n^{(t)})^\top$ are three positive vectors.

The inequality (2.2) is just a variant of the discrete version of Jensen's inequality: $h\left(\sum_{i=1}^{n} p_i z_i\right) \geqslant \sum_{i=1}^{n} p_i h(z_i)$, where $\{p_i\}_{i=1}^{n}$ are probability weights.

2.1 MLEs of parameters via a new MM algorithm

For the purpose of brevity, we define

$$A_i = I(Y_i = 1 \text{ or } 2), B_i = I(Y_i = 0 \text{ or } 1), C_i = I(Y_i = 2) \; and \; D_i = I(Y_i = 0),$$

so that we can re-express (1.2) as

$$\begin{aligned}
\ell(\boldsymbol{\pi}, \rho) &= \sum_{i=1}^{n} \Big\{ A_i \log(\pi_i) + B_i \log(1-\pi_i) + C_i \log[\pi_i + \rho(1-\pi_i)] \\
&\quad + D_i \log[\rho\pi_i + (1-\pi_i)] + c_{i1}(\rho) \Big\} \\
&\overset{(2.7)}{\geqslant} \sum_{i=1}^{n} \Bigg\{ A_i \log(\pi_i) + B_i \log(1-\pi_i) + \frac{C_i \pi_i^{(t)}}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})}{\pi_i^{(t)}} \pi_i \right] \\
&\quad + \frac{C_i \rho(1-\pi_i^{(t)})}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})}{1-\pi_i^{(t)}} (1-\pi_i) \right] \\
&\quad + \frac{D_i \rho \pi_i^{(t)}}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}{\pi_i^{(t)}} \pi_i \right]
\end{aligned}$$

2.1 MLEs of parameters via a new MM algorithm

$$
\begin{aligned}
&+ \frac{D_i(1-\pi_i^{(t)})}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}{1-\pi_i^{(t)}}(1-\pi_i)\right] + c_{i1}(\rho)\Bigg\} \\
= &\sum_{i=1}^{n}\Bigg\{\left[A_i + \frac{C_i\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} + \frac{D_i\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}\right]\log(\pi_i) \\
&+ \left[B_i + \frac{C_i\rho(1-\pi_i^{(t)})}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} + \frac{D_i(1-\pi_i^{(t)})}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}\right]\log(1-\pi_i) + c_{i2}(\rho)\Bigg\} \\
\hat{=} &\; Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}, \rho),
\end{aligned}
$$

where ρ is treated as a fixing constant, $\pi_i^{(t)}$ denotes the t-th approximation of π_i and $\{c_{i1}(\rho), c_{i2}(\rho)\}$ are functions of ρ but not depending on π_i for $i = 1, \ldots, n$.

2.1 MLEs of parameters via a new MM algorithm

Let

$$a_i(\pi_i^{(t)}, \rho) = A_i + \frac{C_i \pi_i^{(t)}}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} + \frac{D_i \rho \pi_i^{(t)}}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})} \quad \text{and}$$

$$b_i(\pi_i^{(t)}, \rho) = B_i + \frac{C_i \rho(1-\pi_i^{(t)})}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} + \frac{D_i (1-\pi_i^{(t)})}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}, \quad i = 1, \ldots, n,$$

we have

$$Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}, \rho) = \sum_{i=1}^{n} \left\{ a_i(\pi_i^{(t)}, \rho) \log(\pi_i) + b_i(\pi_i^{(t)}, \rho) \log(1-\pi_i) + c_{i2}(\rho) \right\}. \quad (2.3)$$

2.1 MLEs of parameters via a new MM algorithm

Furthermore, we have noted that

$$\begin{aligned} & a_i(\pi_i^{(t)}, \rho) + b_i(\pi_i^{(t)}, \rho) \\ = \quad & A_i + B_i + C_i + D_i = 2[I(Y_i = 0) + I(Y_i = 1) + I(Y_i = 2)] = 2. \end{aligned} \tag{2.4}$$

From (1.4), by substituting

$$\pi_i = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \quad \text{and} \quad \pi_i^{(t)} = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}^{(t)}}}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}^{(t)}}}$$

into (2.3), we obtain

$$\begin{aligned} & Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho) \\ = \quad & \sum_{i=1}^{n} \left\{ a_i(\pi_i^{(t)}, \rho) \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta} - \left[a_i(\pi_i^{(t)}, \rho) + b_i(\pi_i^{(t)}, \rho) \right] \log(1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) + c_{i2}(\rho) \right\} \\ \overset{(2.4)}{=} \quad & \sum_{i=1}^{n} \left\{ a_i(\pi_i^{(t)}, \rho) \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta} - 2 \log(1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) + c_{i2}(\rho) \right\}. \end{aligned} \tag{2.5}$$

2.1 MLEs of parameters via a new MM algorithm

Obviously, $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ satisfies the following conditions:

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho) &\leqslant \ell(\boldsymbol{\beta}, \rho), \qquad \forall\, \boldsymbol{\beta}, \boldsymbol{\beta}^{(t)} \in \mathbb{R}^q \quad \text{and} \\ Q(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}, \rho) &= \ell(\boldsymbol{\beta}^{(t)}, \rho). \end{aligned}$$

That is, $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ minorizes $\ell(\boldsymbol{\beta}, \rho)$ at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$. By the MM principle (Becker *et al.*, 1997; Lange *et al.*, 2000; Hunter & Lange, 2004), given $(\boldsymbol{\beta}^{(t)}, \rho^{(t)})$, the $(t+1)$-th approximation of the MLE $\hat{\boldsymbol{\beta}}$ can be calculated by

$$\boldsymbol{\beta}^{(t+1)} = \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^q} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho^{(t)}). \tag{2.6}$$

2.1 MLEs of parameters via a new MM algorithm

We noted that $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ becomes a logistic regression model.

Although the calculation of $\boldsymbol{\beta}^{(t+1)}$ in (2.6) can be implemented by the NR algorithm (equivalently the Fisher scoring algorithm for the logistic regression setting), we want to employ the QLB algorithm (Böhning & Lindsay, 1988) with monotonic convergence since the NR algorithm cannot guarantee the increase of $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ at each iteration.

2.2 MLEs of parameters via a fast QLB algorithm

The QLB iteration is defined by

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + \boldsymbol{B}^{-1}\nabla Q(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}, \rho^{(t)}) \\ &= \boldsymbol{\beta}^{(t)} + 2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top[\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \end{aligned} \tag{2.7}$$

where $\boldsymbol{B}_{q\times q} = (1/2)\boldsymbol{X}^\top\boldsymbol{X}$, $\boldsymbol{X}_{n\times q} = (\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)})^\top$,
$\boldsymbol{y}^{(t)} = (a_1(\pi_1^{(t)}, \rho^{(t)}), \ldots, a_n(\pi_n^{(t)}, \rho^{(t)}))^\top$.

2.2 MLEs of parameters via a fast QLB algorithm

In addition, Tian, Tang & Liu (2012) proposes a novel 'shrinkage parameter' approach to accelerate the QLB algorithm while maintaining its simplicity and stability (i.e., monotonic increase in the target function). The fast QLB iteration is defined by

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + 4(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top [\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \tag{2.8}$$

which is double faster than (2.7), as shown by Tian, Tang & Liu (2012).

2.2 MLEs of parameters via a fast QLB algorithm

By combining (2.8) with (2.1), we summarize the proposed algorithm as follows:

$$\begin{cases} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + 4(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top [\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \\ \rho^{(t+1)} &= \rho^{(t)} - \left[\dfrac{\partial^2 \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho^2} \right]^{-1} \dfrac{\partial \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho}. \end{cases} \tag{2.9}$$

2.3 Simulation results

To evaluate the accuracy of point estimates and confidence intervals, we consider two cases for the dimension: $q = 2$ and $q = 4$. Parameter configurations are set as follows: The sample size $n = 100$, 200, 400; $\rho = 0.3$, 0.5, 0.7;

(a) When $q = 2$, $\boldsymbol{\beta} = (-1, 2)^\top$, $\boldsymbol{x}_{(i)} = (1, x_{i1})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, \sigma_0^2)$ with $\sigma_0^2 = 1.5 \times 10^{-3}$;

(b) When $q = 4$, $\boldsymbol{\beta} = (-1, 2, -1, 2)^\top$, $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$, $\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

2.3 Simulation results

First, for $i = 1, \ldots, n$, we generate Y_i from the following finite discrete distribution:

Y_i	0	1	2
$\Pr(Y_i = y_i)$	p_{i0}	p_{i1}	p_{i2}

denoted by $Y_i \sim \text{FDiscrete}_3((0,1,2)^\top, \boldsymbol{p}_i)$, where $\boldsymbol{p}_i = (p_{i0}, p_{i1}, p_{i2})^\top$ is defined by (1.1).

2.3 Simulation results

Second, we calculate the MLEs of $(\boldsymbol{\beta}^\top, \rho)^\top$ via the fast QLB algorithm (2.9) and the corresponding standard errors.

Third, we independently repeat this process 10,000 times. The resultant average bias (denoted by Bias; i.e., average MLE minus true value of the parameter), the mean square error (denoted by MSE; i.e., Bias^2 + (standard deviation)^2), the coverage probability (denoted by CP) and the average width of confidence interval (denoted by CIW) are reported in the following tables.

As we can see, all Bias, MSEs and coverage probabilities are satisfactory.

2.3 Simulation results

Table 1 Parameter estimates based on 10,000 replications for Case (a)

	ρ	0.3			0.5			0.7		
	n	100	200	400	100	200	400	100	200	400
$\hat{\beta}_0$	Bias	−.0189	−.0120	−.0050	−.0235	−.0146	−.0094	−.0344	−.0113	−.0032
	MSE	.1994	.1032	.0536	.2600	.1399	.0683	.3414	.1624	.0682
	CP	.9552	.9482	.9494	.9535	.9535	.9463	.9551	.9479	.9522
	CIW	1.7443	1.2559	.9089	1.9659	1.4539	1.0087	2.2741	1.5544	1.0233
$\hat{\beta}_1$	Bias	.0441	.0291	.0110	.0555	.0333	.0192	.0760	.0204	.0047
	MSE	.9900	.5387	.2933	1.4741	.7716	.3623	1.9044	.8607	.4020
	CP	.9521	.9511	.9510	.9515	.9523	.9496	.9552	.9498	.9504
	CIW	3.8462	2.8736	2.1278	4.6610	3.4020	2.3229	5.3592	3.5943	2.4844
$\hat{\rho}$	Bias	−.0108	−.0044	−.0023	−.0077	−.0045	−.0019	−.0057	−.0023	−.0011
	MSE	.0098	.0048	.0024	.0081	.0040	.0019	.0055	.0027	.0013
	CP	.9409	.9446	.9462	.9402	.9473	.9520	.9409	.9438	.9484
	CIW	.3799	.2683	.1896	.3470	.2445	.1726	.2862	.2016	.1428

2.3 Simulation results

Table 2 Parameter estimates based on 10,000 replications for Case (b)

	ρ	0.3			0.5			0.7		
	n	100	200	400	100	200	400	100	200	400
$\hat{\beta}_0$	Bias	−.0648	−.0682	−.0132	−.0518	−.0255	−.0078	−.0436	−.0167	−.0136
	MSE	1.4208	.5655	.3024	1.8714	.7086	.3662	2.3333	.9540	.4613
	CP	.9483	.9485	.9506	.9494	.9486	.9482	.9510	.9544	.9507
	CIW	4.5808	2.8697	2.1330	5.1962	3.2369	2.3615	5.8134	3.8079	2.6294
$\hat{\beta}_1$	Bias	.0930	.0324	.0345	.0757	.0346	.0212	.1181	.0393	.0251
	MSE	1.0767	.5590	.3054	1.6137	.9241	.4025	1.9436	.8725	.4512
	CP	.9493	.9494	.9479	.9497	.9437	.9504	.9500	.9479	.9504
	CIW	3.9332	2.9019	2.1368	4.8028	3.6298	2.4716	5.2545	3.5782	2.6173
$\hat{\beta}_2$	Bias	−.0061	−.0053	−.0204	−.0253	−.0173	−.0225	−.0741	−.0387	−.0096
	MSE	3.0323	1.4664	.7467	3.5625	1.7114	.8559	5.4095	2.2990	.9219
	CP	.9512	.9476	.9471	.9478	.9483	.9527	.9514	.9503	.9498
	CIW	6.6641	4.6945	3.3705	7.1563	5.0258	3.6211	8.8206	5.8564	3.7432
$\hat{\beta}_3$	Bias	.1016	.1934	.0281	.1063	.0599	.0298	.0908	.0570	.0236
	MSE	6.3117	2.0239	1.2900	4.8338	2.8539	1.4032	9.5762	3.7430	1.7111
	CP	.9502	.9537	.9468	.9491	.9514	.9515	.9489	.9497	.9491
	CIW	9.6467	5.2832	4.3639	8.3203	6.5245	4.6293	11.6542	7.4092	5.0665
$\hat{\rho}$	Bias	−.0199	−.0094	−.0040	−.0147	−.0075	−.0037	−.0107	−.0049	−.0025
	MSE	.0104	.0051	.0024	.0085	.0041	.0019	.0057	.0027	.0014
	CP	.9370	.9396	.9494	.9389	.9465	.9494	.9441	.9499	.9462
	CIW	.3834	.2697	0.1896	.3508	.2450	.1728	.2911	.2034	.1430

3. Hypothesis Tests

In this section, we are interested in testing the following null hypothesis

$$H_0: \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}_m, \tag{3.1}$$

where $\boldsymbol{C}$ is an $m \times q$ matrix with rank $(\boldsymbol{C}) = r$ and $r < q$.

In Ma & Liu (2017), they tested the equality of disease rates in several groups, which is just a special case of (3.1) with $\boldsymbol{C} = (0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$.

3.1 The likelihood ratio test (LRT)

The likelihood ratio statistic is given by

$$T_L = 2[\ell(\hat{\boldsymbol{\beta}}, \hat{\rho}) - \ell(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\rho}_{H_0})],$$

where $(\hat{\boldsymbol{\beta}}, \hat{\rho})$ are calculated via the fast QLB algorithm (2.9), denoting the unconstrained MLEs of $(\boldsymbol{\beta}, \rho)$; while $(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\rho}_{H_0})$ are the constrained MLEs of $(\boldsymbol{\beta}, \rho)$ under H_0, which can be obtained by

$$\boldsymbol{\beta}_{H_0}^{(t+1)} = \arg\max_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}|\boldsymbol{\beta}_{H_0}^{(t)}, \rho_{H_0}^{(t)}) \quad \text{subject to} \quad \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}, \quad \text{and} \quad (3.2)$$

$$\rho_{H_0}^{(t+1)} = \rho_{H_0}^{(t)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho^2}\right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho}.$$

3.1 The likelihood ratio test (LRT)

Since $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ defined in (2.5) is a nonlinear concave function, the constrained optimization (3.2) can be solved by many effective methods such as the active-set or interior-point methods. In this paper, we employed the built-in fmincon function in MATLAB.

Under the null hypothesis H_0, T_L is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

3.2 The Wald test

Let $\boldsymbol{A}_{m\times(q+1)} = [\boldsymbol{C}, \boldsymbol{0}]$ and $\boldsymbol{\theta}_{(q+1)\times 1} = (\boldsymbol{\beta}^\top, \rho)^\top$, then the null hypothesis H_0 becomes $\boldsymbol{A\theta} = [\boldsymbol{C}, \boldsymbol{0}](\boldsymbol{\beta}^\top, \rho)^\top = \boldsymbol{0}$. The Wald statistic is given by

$$T_W = (\boldsymbol{A}\hat{\boldsymbol{\theta}})^\top [\boldsymbol{A}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{A}^\top]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\rho})^\top$ denote the unconstrained MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{I}(\boldsymbol{\theta})$ denotes the Fisher information matrix given by (A.2).

T_W is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

3.3 The score test

The score test statistic T_S^2 is given by

$$T_S^2 = [\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{H_0})]^\top \boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{H_0}) \boldsymbol{s}(\hat{\boldsymbol{\theta}}_{H_0}),$$

where $\hat{\boldsymbol{\theta}}_{H_0} = (\hat{\boldsymbol{\beta}}_{H_0}^\top, \hat{\rho}_{H_0})^\top$ denote the constrained MLEs of $\boldsymbol{\theta}$ under H_0, and

$$\boldsymbol{s}(\boldsymbol{\theta}) \hat{=} \left(\frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_0}, \ldots, \frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_{q-1}}, 0 \right)^\top,$$

which is given by (A.1).

T_S^2 is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

3.3 Simulation results: type I error rate

The LRT, the Wald test and the score test are developed for testing

$$H_0: \boldsymbol{C\beta} = \mathbf{0} \quad \text{against} \quad H_1: \boldsymbol{C\beta} \neq \mathbf{0}. \tag{3.3}$$

To compare the type I error rates for the three tests, we consider two cases for the dimension: $q = 2$ and $q = 4$. Parameter configurations are set as follows: The sample size $n = 100$, 200, 400; $\rho = 0.3, 0.5, 0.7$;

The results are based on 10,000 times simulation.

3.3 Simulation results: type I error rate

(A1) When $q = 2$, $\boldsymbol{C} = (1, -1)$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ so that (3.3) becomes H_0: $\beta_0 = \beta_1$. We set $\beta_0 = \beta_1 = 0$, $\boldsymbol{x}_{(i)} = (1, x_{i1})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, \sigma_0^2)$ with $\sigma_0^2 = 1.5 \times 10^{-3}$;

(B1) When $q = 4$,

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top,$$

so that (3.3) becomes $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3$. We set $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$, $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$, $\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

3.3 Simulation results: type I error rate

Table 3 The empirical type I error rates of three test statistics (T_L, T_W, T_S^2) for testing (4.1)

ρ	n	Case (A1)			Case (B1)		
		T_L	T_W	T_S^2	T_L	T_W	T_S^2
0.3	100	.0517	.0479	.0498	.0586	.0491	.0513
	200	.0479	.0463	.0466	.0554	.0428	.0476
	400	.0472	.0464	.0468	.0532	.0500	.0519
0.5	100	.0535	.0487	.0511	.0610	.0468	.0531
	200	.0528	.0506	.0522	.0545	.0483	.0510
	400	.0484	.0475	.0481	.0517	.0488	.0496
0.7	100	.0535	.0498	.0521	.0543	.0393	.0472
	200	.0502	.0480	.0494	.0553	.0454	.0492
	400	.0526	.0519	.0523	.0503	.0466	.0485

3.4 Simulation results: empirical power

Next, to evaluate the power performance of the three tests, we consider two cases for the regression coefficients, where Case (A2): $\boldsymbol{\beta} = (-2, 2, -1, 1)^\top$ and Case (B2): $\boldsymbol{\beta} = (-1, 2, -1, 2)^\top$. Other parameter configurations are set as follows: The sample size $n = 100$, 200, 400; $\rho = 0.3, 0.5, 0.7$; $q = 4$,

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top,$$

so that (3.3) becomes $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3$ against H_1: At least two of $(\beta_0, \beta_1, \beta_2, \beta_3)$ are not equal; $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$,
$\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$,
$\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

The results are based on 10,000 times simulation.

3.4 Simulation results: empirical power

Table 4 The empirical powers of three test statistics (T_L, T_W, T_S^2) for testing (4.1)

		Case (A2)			Case (B2)		
ρ	n	T_L	T_W	T_S^2	T_L	T_W	T_S^2
0.3	100	.5325	.4996	.5136	.5319	.4935	.5065
	200	.8119	.7927	.7983	.8413	.8156	.8254
	400	.9585	.9547	.9565	.9569	.9551	.9557
0.5	100	.3845	.3341	.3604	.3962	.3539	.3714
	200	.5734	.5511	.5611	.6010	.5801	.5883
	400	.9328	.9280	.9304	.9328	.9291	.9289
0.7	100	.2657	.2053	.2376	.2768	.2333	.2562
	200	.6325	.5960	.6167	.6293	.5994	.6156
	400	.8899	.8818	.8863	.8993	.8946	.8962

3.4 Simulation results: empirical power

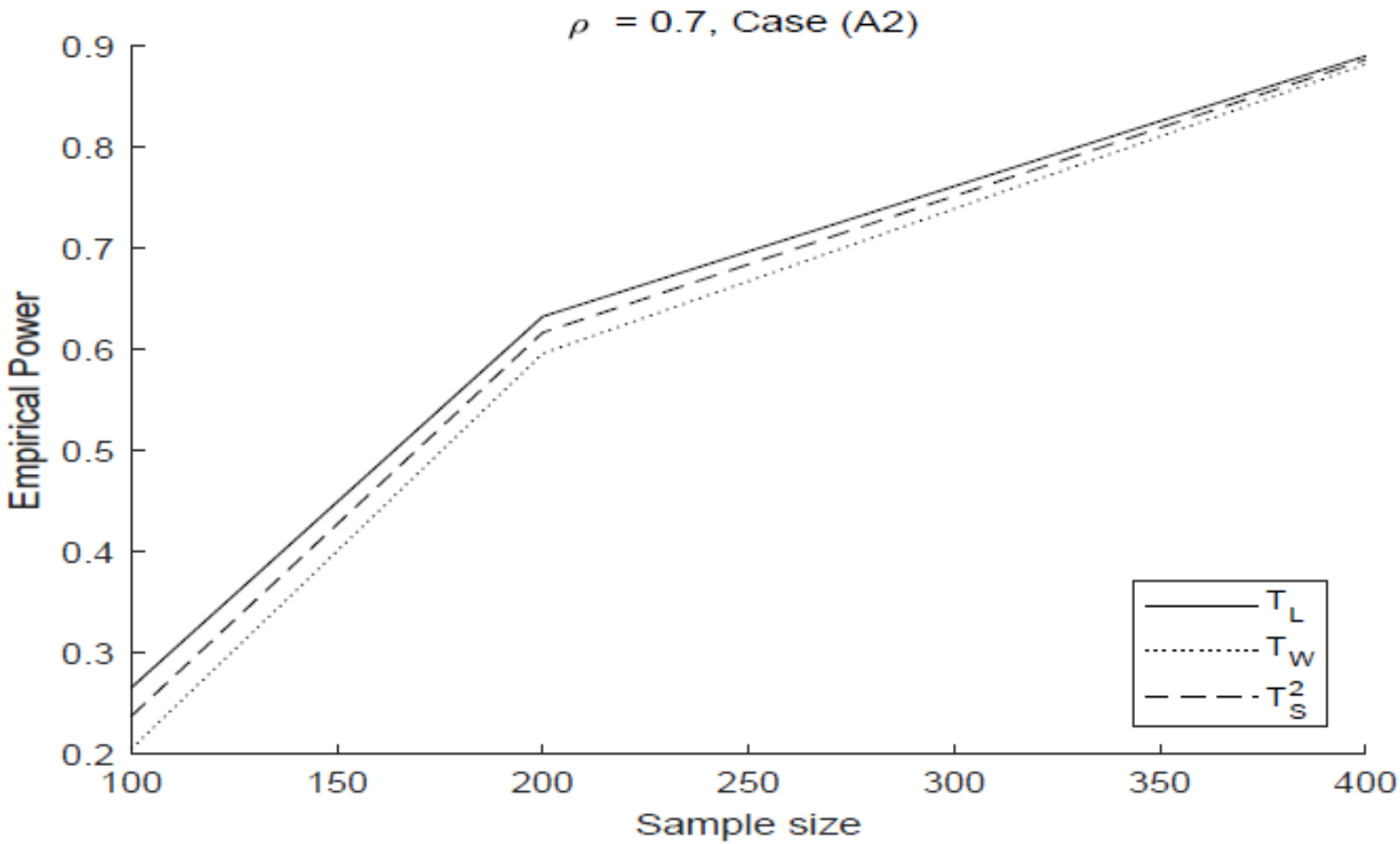

Figure 1. The empirical powers of three test statistics (T_L, T_W, T_S^2) for testing (4.1).

3.4 Simulation results: empirical power

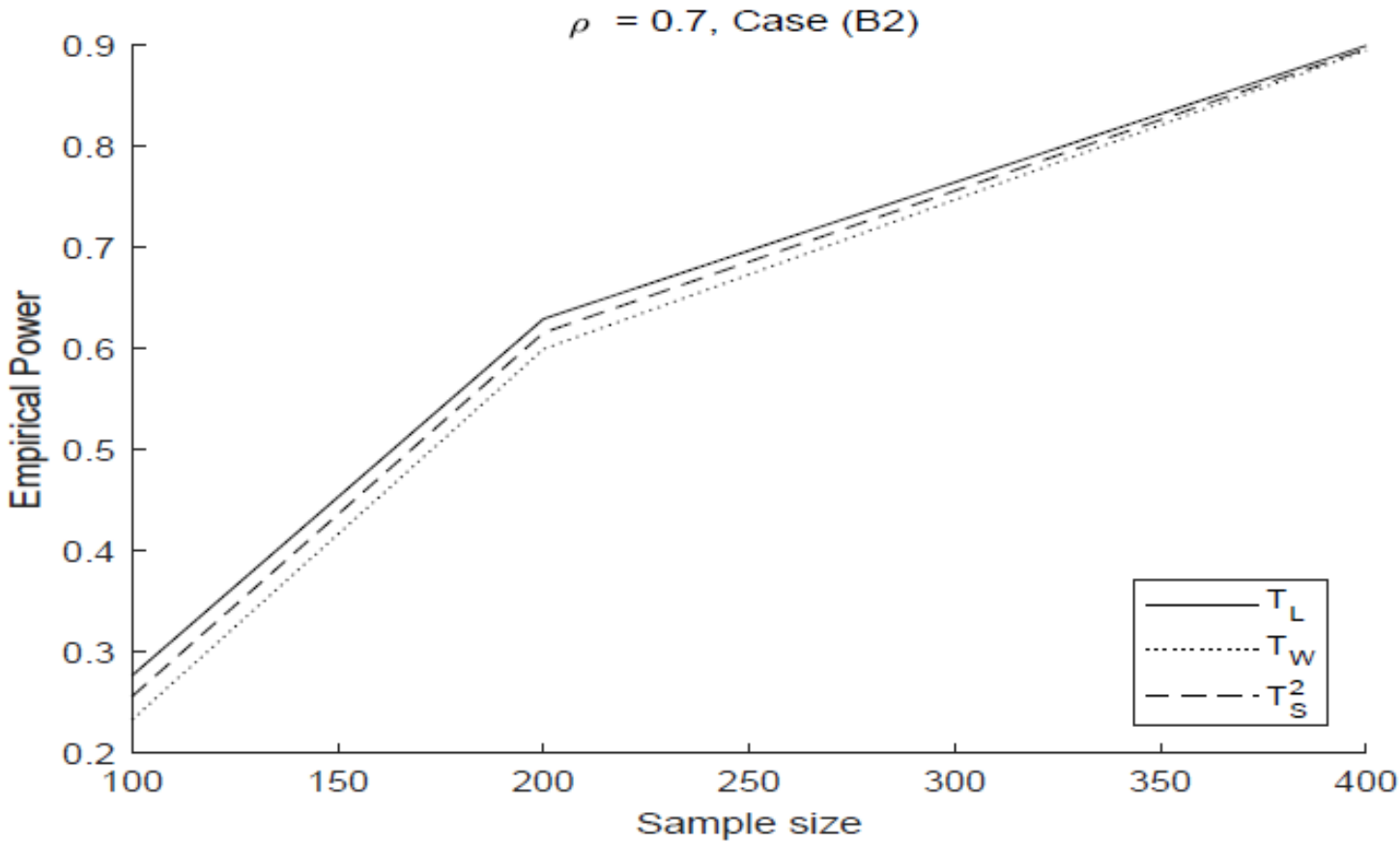

Figure 2. The empirical powers of three test statistics (T_L, T_W, T_S^2) for testing (4.1).

3.4 Simulation results: empirical power

From Figures 1 and 2, we can see that the performance of the LRT is the best among the three tests for all parameter configurations; and the score test performs better than the Wald test.

4. Real Data Set Analysis

In this section, we analyze the ophthalmologic data set presented in Ma & Liu (2017) by employing the proposed methods.

The study was from a cross-sectional, population-based sample in Iran to assess the prevalence of avoidable blindness (Rajavi *et al*., 2011). Nearly 3000 patients were examined and blindness is assessed for seven age groups as shown in the first five columns of Table 5.

4. Real Data Set Analysis

Table 5 The ophthalmologic data in Iran (Rajavi *et al.*, 2011)

Age in year	Number of patients	Blindness			MLE of π_i
		None	Unilateral	Bilaterial	
50–54	989	964	23	2	0.015
55–59	566	541	17	8	0.026
60–64	491	469	18	4	0.026
65–69	278	257	16	5	0.045
70–74	277	242	32	3	0.074
75–79	166	127	30	9	0.145
80+	143	104	29	10	0.171

4. Real Data Set Analysis

From (2.3), the MLEs of $\{\pi_i\}_{i=1}^n$ can be iteratively calculated via the proposed MM algorithm:

$$\pi_i^{(t+1)} = \frac{a_i(\pi_i^{(t)}, \rho^{(t)})}{2}, \quad i = 1, \ldots, n.$$

These results are displayed in the last column of Table 5, exhibiting a potential trend that $\{\hat{\pi}_i\}_{i=1}^n$ are increasing with the ages of the patient groups.

4. Real Data Set Analysis

Ma & Liu (2017) performed several hypothesis tests and found out that the disease rates in different age groups are different, however, they fail to reveal the exact relationship between ages and disease rates.

4. Real Data Set Analysis

To further explore the relationship between ages and disease rates, we consider

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + x_{i1}\beta_1,$$

where x_{i1} denotes the age covariate for patient i.

In this example, we set the ages in different groups as their middle points. By the fast QLB algorithm (2.9), we can calculate the MLEs and confidence intervals of parameters.

Finally, we are interested in testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$, which is corresponding to $\boldsymbol{C} = (0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ in (3.1).

4. Real Data Set Analysis

Table 6 MLEs and confidence intervals of parameters for the ophthalmologic data in Iran

Parameter	MLE	Standard error	95% CI
β_0	−6.9885	0.3525	[−7.6795, −6.2975]
β_1	0.0688	0.0056	[0.0577, 0.0798]
ρ	0.2733	0.0406	[0.1937, 0.3530]

Table 7 Values of three statistics T_L, T_W, T_S^2 and their p-values

Method	T_L	T_W	T_S^2
Statistic	151.7455	149.0259	169.1516
p-value	7.2018e-35	2.1336e-34	1.8305e-28

4. Real Data Set Analysis

From Table 6, $\hat{\rho} = 0.2733$ indicates a positive correlation between eyes from the same patient, while $\hat{\beta}_1 = 0.0688$ reveals a positive relationship between disease rate and age. The estimation results coincide with the common sense that older people have higher chance to suffer from eye disease.

The values of three statistics T_L, T_W, T_S^2 and their p-values are listed in Table 7. Since three p-values are almost equal to zero, the null hypothesis H_0 must be rejected with a strong evidence. This is consistent to the conclusion from the CI of β_1, which excludes zero.

5. Discussion

Based on the model of Donner (1989), in this paper, we introduce the logistic regression model to explore the relationship between ophthalmologic disease rate and covariates such as patients' ages, weights, gender and so on. To avoid direct optimization on the complicated, non-concave log-likelihood $\ell(\boldsymbol{\beta}, \rho)$, we successfully construct a minorizing function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ by utilizing Jensen's inequality, where the Q function enjoys a simple and concave form of the log-likelihood function in the original logistic regression. In Section 2, we combine the fast QLB algorithm with the coordinate ascend method to calculate the unconstrained MLEs of the parameters of interest. The confidence intervals of parameters can be derived by estimating standard errors via the inverse Fisher information matrix.

5. Discussion

We have developed three large-sample tests (i.e., the LRT, Wald test, and score test) for testing if covariates have a significant impact on the disease probability. Simulations were implemented to assess the performance of the proposed fast QLB algorithm and three testing approaches. Simulation results showed that all three tests can satisfactorily control the type I error rates regardless of sample size, or parameter configurations. Moreover, the LRT has the highest power among the three tests and its power increases steadily with the sample sizes. Therefore, the LRT is highly recommended.

5. Discussion

In addition to logit transformation, there are many other link functions (e.g., probit, log-log, complementary log-log) connecting covariates to disease rate. One of our future works is to introduce other link functions in generalized linear Donner's model. Second, when the prior information about the ophthalmologic disease rates are available, we could employ the Bayesian method to generate posterior samples via MCMC. Finally, when the dimension of β is very large, we might use the LASSO approach to select important variables.

Acknowledgements

G.L. Tian's research was supported by a grant from the National Natural Science Foundation of China (No. 11771199).