IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 9, 2022, accepted January 25, 2022, date of publication January 27, 2022, date of current version February 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3147231

Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in *Ga*GLM

BENCHAO WANG[1,2], PAN QIN[1], AND HONG GU[1]
[1]Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
[2]Department of Administration and Supervision, Liaoning Police College, Dalian 116036, China

Corresponding author: Pan Qin (qp112cn@dlut.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities of China under Grant DUT21YG118, and in part by the National Natural Science Foundation of China under Grant U1560102, Grant 61502074, and Grant 61633006.

ABSTRACT The Gamma distribution based generalized linear model (*Ga*GLM) is a kind of statistical model feasible for the positive value of a non-stationary stochastic system, in which the location and the scale are regressed by the corresponding explanatory variables. This paper theoretically investigates the asymptotic properties of maximum likelihood estimates (MLE) of *Ga*GLM, which can benefit the further interval estimates, hypothesis tests and stochastic control design. First, the score function and the Fisher information matrix for *Ga*GLM are derived. Then, the Lyapunov condition is derived to ensure the asymptotic normality of the score function normalized by the Fisher information matrix. Based on this condition, the asymptotic normality of the MLE of *Ga*GLM is proven. Finally, a numerical example is given to testify the asymptotic properties obtained in the research. The numerical results indicate that the MLE of *Ga*GLM converged to a normal distribution as the number of sample measurements increased.

INDEX TERMS Gamma distribution, Gamma regression, consistency and asymptotic normality, central limit theorem, maximum likelihood estimator.

I. INTRODUCTION

The generalized linear model (GLM) expands the general linear model so that a dependent variable is linearly related to the factors and covariates via a specified link function [1]. Moreover, the model allows the dependent variable to keep the attribute of actually applied data, such as integer literal, positive and asymmetric, not belong to a normal distribution. It covers widely used statistical models, such as logistic regression models for binary distributed responses, Poisson regression models for count data and Gamma regression models for positive real data.

As a family of moderate skewness and continuous phenomena distributions, the Gamma distribution is a useful model in many areas of statistics when the normal distribution is not appropriate. In the Gamma distribution-based approach, the system output Z can be assumed to be a subject $Z \sim Ga(\alpha, \beta)$, where $Ga(\alpha, \beta)$ is a Gamma distribution with the shape parameter α and the rate parameter β governing its probability density function shape. This distribution was first introduced [2] and subsequently studied in detail [3]. In some special cases, the Gamma distribution reduces to the exponential distribution as $\alpha = 1$ and $\beta = 1/\lambda$ for $\lambda > 0$, the Erlang distribution as $\alpha = n$ and the χ^2 distribution as $\alpha = n/2$ and $\beta = 1/2$.

Because of the flexibility of the relationship to many other distributions, the Gamma distribution can be a suitable alternative for modelling such kinds of the positive-valued dependent variable. The Gamma distribution-based models have been applied in many areas, such as medical science [4], [5], biology [6], economics [7], [8], forest science [9] and education [10]. Considering the ubiquitous heteroscedasticity of actually applied data, as a member of the well-known GLM, the Gamma distribution based generalized linear model (*Ga*GLM) is more widely used when α and β are both dependent variables. However, it should be noted that the *Ga*GLM does not belong to the exponential family of distributions based GLM [11]. Therefore, a baseable asymptotic theory for *Ga*GLM is established.

This research investigates the theoretical aspects of maximum likelihood estimator (MLE) for *Ga*GLM. Because *Ga*GLM is a model with two equations (4) being respectively parameterized for α and β, the estimation procedure could be relatively complex. In statistics, several

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

expectation-maximization (EM) type algorithms have been developed for the Gamma distribution inference, where β was assumed to be a latent variable [12]. However, those algorithms were developed by fixing β as constant. If β is parameterized as regression models, the EM algorithm would be extremely computationally involved. Thus, instead of using EM algorithms for latent variable, MLE for GaGLM by using the Fisher scoring algorithm is performed [13]. To this end, the score function and the Fisher information matrix are derived for *Ga*GLM. Furthermore, we obtain the condition to assure the positive definiteness of the Fisher information matrix.

The consistency and the asymptotic normality explaining the efficiency of the estimators have been widely investigated in system identification and statistics [14], [15]. The consistency of MLE for *Ga*GLM can be proved by using the same approach for GLM [15]. To verify the asymptotic normality of MLE, the asymptotic normality of the normalized score function is necessary. GLM was developed for the exponential family, whose moment generating functions are exponential functions of the sufficient statistics. Based on the uniform moment generating function, the asymptotic normality of the normalized score function was proved for GLM.

To investigate asymptotic properties of MLE for parameters occurring in *Ga*GLM, we need to prove the consistency and asymptotic normality of MLE by central limit theorems. Compared with commonly used Lindeberg condition, Lyapunov condition is stronger in proving asymptotic properties. First, we derived the score functions normalized by the Fisher information matrix for the Lyapunov condition, which ensure the asymptotic normality of the normalized score functions [16]. Based on this result, the asymptotic normality of MLE for *Ga*GLM is finally proved. These results can dramatically facilitate the hypothesis testing, the construction of interval estimates, and stochastic control design for the non-stationary stochastic system [17], [18].

The rest of this paper is organized as follows. The concept of *Ga*GLM and maximum likelihood estimation are introduced in Section II. Section III gives the assumptions of asymptotic properties of the MLE in *Ga*GLM, including the proof of related lemmas and theorems. Results of a simulation study are reported in Section IV. Concluding comments are presented in Section V.

II. PROBLEM STATEMENT

In this section, we briefly review *Ga*GLM, including its structure and numerical method of MLE.

A. MODEL AND ESTIMATION

Suppose that we observe realizations of a positive real random variable Z, and we believe that Z has a specified positive continuous distribution.

Let $D_n = \{(Z_i, \boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, \dots, n\}$ be independent random vectors defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $i = 1, \dots, n$, the response variable Z_i is generated from the following process:

$$Z_i \sim Ga(\alpha_i, \beta_i), \tag{1}$$

where $Ga(\alpha_i, \beta_i)$ denotes the Gamma distribution with positive shape parameter α_i and rate parameter β_i. The probability density function is

$$f(Z_i|\alpha_i, \beta_i) = \begin{cases} \dfrac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} z_i^{\alpha_i - 1} e^{-\beta_i z_i}, & \text{when } z > 0 \\ 0, & \text{when } z \le 0, \end{cases}$$

where $\Gamma(\cdot)$ is the Gamma function. The mean and variance of the random variable Z_i are given by

$$E(Z_i) = \frac{\alpha_i}{\beta_i} \tag{2}$$

and

$$Var(Z_i) = \frac{\alpha_i}{\beta_i^2}. \tag{3}$$

Then, we can develop *Ga*GLM by regressing explanatory variable $\boldsymbol{\omega}_i = (\boldsymbol{x}_i^\top, \boldsymbol{y}_i^\top)^\top$ with $\boldsymbol{x}_i \in \mathbb{R}^p$ to α_i and $\boldsymbol{y}_i \in \mathbb{R}^q$ to β_i as follows:

$$\begin{cases} \alpha_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{a}) \\ \beta_i = \exp(\boldsymbol{y}_i^\top \boldsymbol{b}), \end{cases} \tag{4}$$

where $\boldsymbol{a} = (a_1, \dots, a_p)^\top$ and $\boldsymbol{b} = (b_1, \dots, b_q)^\top$ denote the regression parameter vectors for α_i and β_i respectively, and $\bullet^\top$ denotes the transpose of $\bullet$. Further, $\boldsymbol{\theta} = (\boldsymbol{a}^\top, \boldsymbol{b}^\top)^\top$ is any parameter in an admissible set $K_{\boldsymbol{\theta}} \subset \mathbb{R}^{p+q}$. For the observations $z_1, z_2, \dots, z_n$, the log-likelihood $l(\boldsymbol{\theta})$ derived from the *Ga*GLM can be written as

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \log(f(Z_i|\boldsymbol{\omega}_i, \boldsymbol{\theta})) \\ &= \sum_{i=1}^n \big(\alpha_i \log \beta_i - \log \Gamma(\alpha_i) + \alpha_i \log z_i \\ &\qquad - \log z_i - \beta_i z_i\big) \\ &= \sum_{i=1}^n \Big(e^{\boldsymbol{x}_i^\top \boldsymbol{a}} \boldsymbol{y}_i^\top \boldsymbol{b} - \log \Gamma(e^{\boldsymbol{x}_i^\top \boldsymbol{a}}) \\ &\qquad + e^{\boldsymbol{x}_i^\top \boldsymbol{a}} \log z_i - \log z_i - e^{\boldsymbol{y}_i^\top \boldsymbol{b}} z_i\Big), \end{aligned} \tag{5}$$

where log denotes the natural logarithm. Then, $\boldsymbol{\theta}$ can be estimated by

$$\widehat{\boldsymbol{\theta}} = \arg\max\ l_n(\boldsymbol{\theta}). \tag{6}$$

According to (5), the first three order derivative of $l_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is continuous and finite for all $\boldsymbol{\theta} \in K_{\boldsymbol{\theta}}$. This condition ensures the existence of the Taylor expansion, the finite variance of the derivatives of $l_n(\boldsymbol{\theta})$. Thus, MLE can be obtained by the scoring method [19], in which the score function can be obtained by

$$\begin{aligned} \boldsymbol{s}_n(\boldsymbol{\theta}) &= \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \left(\boldsymbol{s}_{\boldsymbol{a}}^{\top}(\boldsymbol{\theta}), \boldsymbol{s}_{\boldsymbol{b}}^{\top}(\boldsymbol{\theta})\right)^{\top} \\ &= \left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^{\top}} \quad \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^{\top}}\right]^{\top}, \end{aligned} \tag{7}$$

and the Fisher information matrix can be obtained by

$$\begin{aligned} \boldsymbol{F}_n(\boldsymbol{\theta}) &= -E\left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}\right] \\ &= -E\begin{bmatrix} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^{\top}} & \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^{\top}} \\ \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^{\top}} & \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^{\top}} \end{bmatrix}. \end{aligned} \tag{8}$$

With the score function and the Fisher information matrix, (6) can be iteratively solved by using the generalized Newton-Raphson (NR) method, so-called Fisher's scoring (FS) algorithm [19] as the following

$$\widehat{\boldsymbol{\theta}}_{(new)} = \widehat{\boldsymbol{\theta}}_{(old)} + \boldsymbol{F}_n^{-1}(\widehat{\boldsymbol{\theta}}_{(old)})\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_{(old)}). \tag{9}$$

In what follows, the score function and the Fisher information matrix are derived for *Ga*GLM. Furthermore, the condition that ensures the positive definiteness of $\boldsymbol{F}_n(\boldsymbol{\theta})$ obtained in Corollary 1.

In statistics, the asymptotic properties, mainly including the consistency and asymptotic normality, are often used to evaluate the efficiency of estimators [20]. Another important role of $\boldsymbol{s}_n(\boldsymbol{\theta})$ and $\boldsymbol{F}_n(\boldsymbol{\theta})$ is to prove the asymptotic properties. If the first three order derivates of $l_n(\boldsymbol{\theta})$ with respect of $\boldsymbol{\theta}$ exist, the consistency, i.e. $\widehat{\boldsymbol{\theta}}$ converging in probability to the true coefficients $\boldsymbol{\theta}_0$, can be proved under a generalized framework [21]. However, the asymptotic convergence of the covariance matrix for *Ga*GLM cannot be proved by using the generalized approach in [21]. To tackle this problem, we first prove the asymptotic normality of the normalized score function $\boldsymbol{F}_n^{-\top/2}(\boldsymbol{\theta})\boldsymbol{s}_n(\boldsymbol{\theta})$ motivated by [15]. Note that [15] dealt with the exponential family-based models, whose moment generating function is the exponential function of the sufficient statistics. [15] used such moment generating function to prove the asymptotic normality of the normalized score function. However, there is not an asymptotic theory of MLE to Gamma distribution, where the approach in [15] cannot be extended to the *Ga*GLM. Furthermore, the elements constructing $\boldsymbol{F}_n^{-\top/2}(\boldsymbol{\theta})\boldsymbol{s}_n(\boldsymbol{\theta})$ cannot be expected to be identically distributed. Therefore, by investigating the Lyapunov condition and Taylor expansion, we can prove the asymptotic normality of *Ga*GLM's MLE. In what follows, we first derive the score function and the Fisher information matrix of MLE of *Ga*GLM.

III. SCORE FUNCTION AND FISHER INFORMATION MATRIX FOR *Ga*GLM

For deriving the score function and the Fisher information matrix, the log-likelihood function of $\boldsymbol{\theta}$ is formulated from (6). The score function (7) can be represented as follows.

Lemma 1 (Component-Wise Score Function for GaGLM): The two components of the score function (7) are obtained by

$$\begin{aligned} \boldsymbol{s}_{\boldsymbol{a}}(\boldsymbol{\theta}) &= \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \\ &= (s_1(\boldsymbol{\theta}), \ldots, s_p(\boldsymbol{\theta}))^{\top} \\ &= \sum_{i=1}^{n} (\log \beta_i - \psi_0(\alpha_i) + \log z_i)\, \boldsymbol{x}_i \\ &= \sum_{i=1}^{n} (\alpha_i \log \beta_i - \psi_0(\alpha_i)\alpha_i + \alpha_i \log z_i)\, \boldsymbol{x}_i \end{aligned} \tag{10}$$

and

$$\boldsymbol{s}_{\boldsymbol{b}}(\boldsymbol{\theta}) = \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} = \textstyle\sum_{i=1}^{n} (\alpha_i - \beta_i z_i)\, \boldsymbol{y}_i, \tag{11}$$

where $\psi_0(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ is digamma function, which can be seen in Equation (20).

Proof Lemma 1: The following derivatives can be directly derived from (4).

$$\begin{cases} \dfrac{\partial \alpha_i}{\partial \boldsymbol{a}} = \alpha_i \boldsymbol{x}_i \\ \dfrac{\partial \beta_i}{\partial \boldsymbol{b}} = \beta_i \boldsymbol{y}_i. \end{cases} \tag{12}$$

Then, derivatives of the log-likelihood function (4) are straightforwardly obtained. ■

The Fisher information matrix will be derived via the Hessian matrix follows

We define Hessian matrix $\mathcal{H}_n(\boldsymbol{\theta})$ as follows to establish the Fisher information matrix $\boldsymbol{F}_n(\boldsymbol{\theta})$ shown below.

$$\begin{aligned} \mathcal{H}_n(\boldsymbol{\theta}) &= \begin{bmatrix} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^{\top}} & \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^{\top}} \\ \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^{\top}} & \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^{\top}} \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{h}_{p,p}(\boldsymbol{\theta}) & \boldsymbol{h}_{p,q}(\boldsymbol{\theta}) \\ \boldsymbol{h}_{q,p}(\boldsymbol{\theta}) & \boldsymbol{h}_{q,q}(\boldsymbol{\theta}) \end{bmatrix}, \end{aligned} \tag{13}$$

with entries:

$$\begin{aligned} \boldsymbol{h}_{p,p}(\boldsymbol{\theta}) &= \frac{\partial^2 l_n(\theta)}{\partial \boldsymbol{a} \partial \boldsymbol{a}^{\top}} \\ &= \sum_{i=1}^{n} \alpha_i \left[\log \beta_i + \log z_i - \psi_0(\alpha_i) - \alpha_i \psi_1(\alpha_i)\right] \boldsymbol{x}_i \boldsymbol{x}_i^{\top} \end{aligned} \tag{14}$$

where $\psi_1(\alpha) = \psi'(\alpha)$ is trigamma function, which can be seen in Equation (21).

$$\begin{aligned} \boldsymbol{h}_{q,q}(\boldsymbol{\theta}) &= \frac{\partial^2 l_n(\theta)}{\partial \boldsymbol{b} \partial \boldsymbol{b}^{\top}} \\ &= -\sum_{i=1}^{n} \beta_i z_i \boldsymbol{y}_i \boldsymbol{y}_i^{\top} \end{aligned} \tag{15}$$

and

$$\boldsymbol{h}_{p,q}(\boldsymbol{\theta}) = \boldsymbol{h}_{q,p}^{\top}(\boldsymbol{\theta}) = \frac{\partial^2 l_n(\theta)}{\partial \boldsymbol{a} \partial \boldsymbol{b}^{\top}} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i \boldsymbol{y}_i^{\top} \quad (16)$$

Now set

$$\boldsymbol{H}_n(\boldsymbol{\theta}) = -\mathcal{H}(\boldsymbol{\theta}). \quad (17)$$

In order to derive the Fisher information matrix and prove asymptotic normality, the following Lemmas 2 and 3 are necessary.

Lemma 2: If $Z \sim Ga(\alpha, \beta)$ then

$$E(\log Z) = \psi_0(\alpha) - \log \beta \quad (18)$$

and

$$E(\log^k Z) \le C(\alpha, \beta, k), \quad k = 1, 2, \ldots, n \quad (19)$$

where C is positive constant depending on α, β and k, and $k > 0$ is any finite positive integer.

Proof Lemma 2: Before the proof of this lemma, we should recall the Euler's gamma function $\Gamma(\alpha)$ and digamma function $\psi_0(\alpha)$ for $\alpha > 0$ defined as

$$\psi_0(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \quad \mathit{with}\ \Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u}\, du. \quad (20)$$

For basic properties of these functions see [22]. Polygamma functions ψ_n, such as trigamma, tetragamma and pentagamma functions when $n = 1, 2, 3$, are defined to be n-order derivatives of ψ_0 function, that is,

$$\psi_n(\alpha) = \psi_0^{(n)}(\alpha), \quad n = 1, 2, \ldots \quad (21)$$

The following integral and series representations are valid for $z > 0$ and $n = 1, 2, 3, \ldots$:

$$(-1)^{n-1}\psi_n(\alpha) = \int_0^{\infty} \frac{t^n e^{-\alpha t}}{1 - e^{-t}}\, dt = n! \sum_{k=0}^{\infty} \frac{1}{(\alpha + k)^{n+1}} \quad (\alpha > 0), \quad (22)$$

which are monotonically increasing and continuous function in $\alpha > 0$ [23]. Then, We can get $\frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)}$ by polygamma functions with $k = 1, 2, \ldots, n$. When $n = 1$, we can get

$$\begin{aligned}\psi_1(\alpha) &= \psi_0'(\alpha) \\ &= \left(\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)' \\ &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \left(\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)^2 \\ &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \psi_0^2(\alpha),\end{aligned}$$

then

$$\frac{\Gamma''(\alpha)}{\Gamma(\alpha)} = \psi_1(\alpha) + \psi_0^2(\alpha). \quad (23)$$

When $n = 2$, we can get

$$\begin{aligned}\psi_2(\alpha) &= \psi_1'(\alpha) \\ &= \left(\frac{\Gamma''(\alpha)}{\Gamma(\alpha)}\right)' - (\psi_0^2(\alpha))' \\ &= \frac{\Gamma'''(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma''(\alpha)\Gamma'(\alpha)}{\Gamma^2(\alpha)} - 2\psi_0(\alpha)\psi'(\alpha) \\ &= \frac{\Gamma'''(\alpha)}{\Gamma(\alpha)} - \left(\psi_1(\alpha) + \psi_0^2(\alpha)\right)\psi_0(\alpha) - 2\psi_0(\alpha)\psi_1(\alpha) \\ &= \frac{\Gamma'''(\alpha)}{\Gamma(\alpha)} - \psi_0^3(\alpha) - 3\psi_0(\alpha)\psi_1(\alpha).\end{aligned}$$

In the following,

$$\frac{\Gamma'''(\alpha)}{\Gamma(\alpha)} = \psi_2(\alpha) + \psi_0^3(\alpha) + 3\psi_0(\alpha)\psi_1(\alpha) \quad (24)$$

Continuous derivative of (20) function as (23) and (24), we can get $\frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)}$ function.

$$\frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)} = f_{k\Gamma}\left(\psi_0(\alpha), \psi_1(\alpha), \psi_2(\alpha), \ldots, \psi_{k-1}(\alpha)\right), \quad (25)$$

where $f_{k\Gamma}(\psi_0(\alpha), \psi_1(\alpha), \psi_2(\alpha), \ldots, \psi_{k-1}(\alpha))$ is a finite j-order polynomial combination function of the $\psi_i(\alpha)$ functions with $i < k$ and $j \le k$.

Let $Z \sim Ga(\alpha, \beta)$, we have

$$\int_0^{\infty} \frac{1}{\Gamma(\alpha)} \beta^{\alpha} z^{\alpha-1} e^{-\beta z} dz = 1. \quad (26)$$

Multiply at each side of equation (26) by $\Gamma(\alpha)$, we can get

$$\Gamma(\alpha) = \int_0^{\infty} \beta^{\alpha} z^{\alpha-1} e^{-\beta z} dz. \quad (27)$$

Then, take the k-order derivative with respect to α of both sides, that

$$\Gamma^{(k)}(\alpha) = \textstyle\sum_{i=0}^{k} C_k^i \int_0^{\infty} \beta^{\alpha} z^{\alpha-1} e^{-\beta z} \log^{k-i} \beta \log^i z dz. \quad (28)$$

Divided by $\Gamma(\alpha)$ at both sides of (28), we obtain

$$\begin{aligned}\frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)} &= \sum_{i=0}^{k} C_k^i \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \beta^{\alpha} z^{\alpha-1} e^{-\beta z} \log^{k-i} \beta \log^i z dz \\ &= \sum_{i=0}^{k} C_k^i \log^{k-i} \beta E(\log^i Z). \quad (29)\end{aligned}$$

For $k = 1$,

$$\begin{aligned}\psi_0(\alpha) &= \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \log \beta z^{\alpha-1} e^{-\beta z} dz \\ &\quad + \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \beta^{\alpha} z^{\alpha-1} \log z e^{-\beta z} dz \\ &= \log \beta + E(\log Z) \quad (30)\end{aligned}$$

then

$$E(\log Z) = \psi_0(\alpha) - \log \beta. \quad (31)$$

Let $f_k(\bullet)$ denote finite j-order polynomial of $\log \beta$ by $j \le k$ and linear combination of $E(\log^i Z)$ with

$i = 1, 2, \ldots, k-1$ function. Combining (25) and (27), we can get

$$\begin{aligned} E(\log^k Z) &= \frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)} + f_k\big(\log\beta, E(\log Z), E(\log^2 Z), \ldots, E(\log^{k-1} Z)\big) \\ &= f_k\big(\log\beta, E(\log Z), E(\log^2 Z), \ldots, E(\log^{k-1} Z)\big) \\ &\quad + f_{k\Gamma}\left(\psi_0(\alpha), \psi_1(\alpha), \psi_2(\alpha), \ldots, \psi_{k-1}(\alpha)\right) \\ &\le C(\alpha, \beta, k) \end{aligned} \tag{32}$$

■

Lemma 3: If $Z \sim Ga(\alpha, \beta)$, the kth moment of Z is limited as

$$EZ^k \le C(\alpha, \beta, k), \quad k = 1, 2, \ldots, n \tag{33}$$

where C is positive constant depending on α, β and k, and $k > 0$ is any finite positive integer.

Proof Lemma 3:

$$\begin{aligned} EZ^k &= \int_0^\infty \frac{1}{\Gamma(\alpha)} \beta^\alpha z^{k+\alpha-1} e^{-\beta z} dz \\ &= \int_0^\infty \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \frac{\beta^{-k}}{\Gamma(k+\alpha)} \beta^{k+\alpha} z^{k+\alpha-1} e^{-\beta z} dz \\ &= \frac{\Gamma(k+\alpha)\beta^{-k}}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\Gamma(k+\alpha)} \beta^{k+\alpha} z^{k+\alpha-1} e^{-\beta z} dz \\ &= \frac{\Gamma(k+\alpha)\beta^{-k}}{\Gamma(\alpha)} \\ &\le C(\alpha, \beta, k) \end{aligned} \tag{34}$$

■

Theorem 1 (Fisher Information Matrix for GaGLM): The components of the Fisher information matrix are obtained as follows:

$$E\left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}^\top}\right] = \sum_{i=1}^n (\alpha_i \log\beta_i - \psi_0(\alpha_i)\alpha_i + \alpha_i \log z_i)\, \boldsymbol{x}_i \boldsymbol{x}_i^\top \tag{35}$$

$$E\left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^\top}\right] = \sum_{i=1}^n \alpha_i \boldsymbol{y}_i \boldsymbol{y}_i^\top \tag{36}$$

$$E\left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{a}} \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{b}^\top}\right] = -\sum_{i=1}^n \alpha_i \boldsymbol{x}_i \boldsymbol{y}_i^\top \tag{37}$$

Proof Theorem 1: Under the assumptions of mild general regularity, we have $\boldsymbol{F}_n(\boldsymbol{\theta}) = E\boldsymbol{H}_n(\boldsymbol{\theta})$ by [24], and $\boldsymbol{F}_n(\boldsymbol{\theta})$ is positive semi-definite matrix [25]. Thus, using Lemmas 1, 2, equations (8) and (13), the Fisher information matrix can be straightforward computed as follows

$$\begin{aligned} \boldsymbol{F}_n(\boldsymbol{\theta}) &= E\boldsymbol{H}_n(\boldsymbol{\theta}) \\ &= -\begin{bmatrix} E\boldsymbol{h}_{p,p}(\boldsymbol{\theta}) & E\boldsymbol{h}_{p,q}(\boldsymbol{\theta}) \\ E\boldsymbol{h}_{q,p}(\boldsymbol{\theta}) & E\boldsymbol{h}_{q,q}(\boldsymbol{\theta}) \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{f}_{p,p}(\boldsymbol{\theta}) & \boldsymbol{f}_{p,q}(\boldsymbol{\theta}) \\ \boldsymbol{f}_{q,p}(\boldsymbol{\theta}) & \boldsymbol{f}_{q,q}(\boldsymbol{\theta}) \end{bmatrix}. \end{aligned}$$

According to (18) in lemma 2, they are expressed respectively as follows

$$\begin{aligned} \boldsymbol{f}_{p,p} &= -E\left[\boldsymbol{h}_{p,p}(\boldsymbol{\theta})\right] \\ &= -E\left[\sum_{i=1}^n \alpha_i\big(\log\beta_i + \log z_i - \psi_0(\alpha_i) - \alpha_i\psi_1(\alpha_i)\big)\boldsymbol{x}_i\boldsymbol{x}_i^\top\right] \\ &= \sum_{i=1}^n \alpha_i^2 \psi_1(\alpha_i) \boldsymbol{x}_i \boldsymbol{x}_i^\top, \end{aligned} \tag{38}$$

$$\begin{aligned} \boldsymbol{f}_{q,q} &= -E\left[\boldsymbol{h}_{q,q}(\boldsymbol{\theta})\right] \\ &= -E\sum_{i=1}^n \beta_i z_i \boldsymbol{y}_i \boldsymbol{y}_i^\top \\ &= \sum_{i=1}^n \alpha_i \boldsymbol{y}_i \boldsymbol{y}_i^\top \end{aligned} \tag{39}$$

and

$$\begin{aligned} \boldsymbol{f}_{p,q}(\boldsymbol{\theta}) = \boldsymbol{f}_{q,p}^\top(\boldsymbol{\theta}) &= -E\left[\boldsymbol{h}_{p,q}(\boldsymbol{\theta})\right] \\ &= -E\sum_{i=1}^n \alpha_i \boldsymbol{x}_i \boldsymbol{y}_i^\top \\ &= -\sum_{i=1}^n \alpha_i \boldsymbol{x}_i \boldsymbol{y}_i^\top. \end{aligned} \tag{40}$$

Then, we can get

$$\begin{aligned} \boldsymbol{F}_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \begin{bmatrix} \alpha_i^2\psi_1(\alpha_i) & -\alpha_i \\ -\alpha_i & \alpha_i \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_i\boldsymbol{x}_i^\top & \boldsymbol{x}_i\boldsymbol{y}_i^\top \\ \boldsymbol{y}_i\boldsymbol{x}_i^\top & \boldsymbol{y}_i\boldsymbol{y}_i^\top \end{bmatrix} \\ &= \sum_{i=1}^n \alpha_i \begin{bmatrix} \alpha_i\psi_1(\alpha_i) & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_i^\top & \boldsymbol{y}_i^\top \end{bmatrix}. \end{aligned} \tag{41}$$

■

Corollary 1 (Definiteness of Fisher Information Matrix for GaGLM): If $\sum_{i=1}^n \boldsymbol{\omega}_i\boldsymbol{\omega}_i^\top$ is of full rank, with $\boldsymbol{\omega}_i = (\boldsymbol{x}_i^\top, \boldsymbol{y}_i^\top)^\top$ denoted in Section II, the Fisher information matrix $\boldsymbol{F}_n(\boldsymbol{\theta})$ is positive-definite.

Proof Corollary 1: To prove the positive character of the Fisher information matrix, we need derive the range of $\alpha\psi_1(\alpha)$. From the equation (22) as $n = 1$, we can get following inequality

$$\begin{aligned} \psi_1(\alpha) &= \sum_{k=0}^\infty \frac{1}{(\alpha+k)^2} \\ &> \sum_{k=0}^\infty \frac{1}{(\alpha+k)(\alpha+k+1)} \\ &= \sum_{k=0}^\infty \left(\frac{1}{\alpha+k} - \frac{1}{\alpha+k+1}\right) \\ &= \frac{1}{\alpha} - \frac{1}{\alpha+1} + \frac{1}{\alpha+1} - \frac{1}{\alpha+2} + \frac{1}{\alpha+2} - \ldots \\ &= \frac{1}{\alpha}. \end{aligned} \tag{42}$$

Then, we can get

$$\alpha_i \begin{vmatrix} \alpha_i\psi_1(\alpha_i) & -1 \\ -1 & 1 \end{vmatrix} = \alpha_i(\alpha_i\psi_1(\alpha_i) - 1) > 0. \tag{43}$$

If $\sum_{i=1}^{n} \boldsymbol{\omega}_i\boldsymbol{\omega}_i^\top$ is of full rank, the Fisher information matrix $\boldsymbol{F}_n(\boldsymbol{\theta})$ is positive-definite.

■

IV. ASYMPTOTIC THEORY FOR THE MAXIMUM LIKELIHOOD ESTIMATOR IN *Ga*GLM

Under the mild assumptions, the asymptotic properties of the MLE was proved in GLM for canonical link functions [15]. These asymptotic conditions can be applied to prove similar results for *Ga*GLM as well as noncanonical.

To normalize the score function, we introduce the Cholesky square root matrix for positive definite matrix $\boldsymbol{F}$, such that $\boldsymbol{F}^{1/2}(\boldsymbol{F}^{1/2})^\top = \boldsymbol{F}$. We set $\boldsymbol{F}^{1/2}$ denotes the unique lower triangular matrix with positive diagonal elements. For convenience, set $\boldsymbol{F}^{\top/2} := (\boldsymbol{F}^{1/2})^\top$, $\boldsymbol{F}^{-1/2} := (\boldsymbol{F}^{1/2})^{-1}$ and $\boldsymbol{F}^{-\top/2} := (\boldsymbol{F}^{\top/2})^{-1}$. For convenience, we drop the argument $\boldsymbol{\theta}_0$ in $\boldsymbol{s}_n(\boldsymbol{\theta}_0)$, $\boldsymbol{s}_{ni}(\boldsymbol{\theta}_0)$, $\boldsymbol{F}_n(\boldsymbol{\theta}_0)$ $E_{\boldsymbol{\theta}_0}$ etc. and write $\boldsymbol{s}_n$, $\boldsymbol{s}_{ni}$, $\boldsymbol{F}_n$, E etc. C_i for $i = 1, 2, \ldots$ will further denote constants, with or without subindices. The same C's represent different constants in different formula.

Let $\|\cdot\|$ denote the spectral norm of square matrices. The spectral norm of a real-valued matrix $\boldsymbol{F}$ is given by

$$\begin{aligned} \|\boldsymbol{F}\| &= \left(\lambda_{\max}(\boldsymbol{F}\boldsymbol{F}^\top)\right)^{1/2} \\ &= \sup_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{F}\boldsymbol{u}\|_2, \end{aligned} \tag{44}$$

where $\|\cdot\|_2$ denotes the L^2- norm of vectors. The maximal (minimal) eigenvalue of a square matrix $\boldsymbol{F}$ will be further denoted by $\lambda_{\max}(\boldsymbol{F})$ ($\lambda_{\min}(\boldsymbol{F})$). For $\varepsilon > 0$, a neighborhood of the unknown true parameter $\boldsymbol{\theta}_0$ can denote by

$$N_n(\varepsilon) = \left\{\boldsymbol{\theta} : \left\|\boldsymbol{F}_n^{\top/2} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\| \le \varepsilon\right\}. \tag{45}$$

In this paper, let's make the following assumptions.

(A1)

$$\lambda_{\min}(\boldsymbol{F}_n) \ge \frac{n}{C} \quad \forall\, n \ge 1, \tag{46}$$

where C is a positive constant.

(A2) $\{\boldsymbol{x}_n, n \ge 1\} \subset K_x$,$\{\boldsymbol{y}_n, n \ge 1\} \subset K_y$, where $K_x \subset \mathbb{R}^p$ and $K_y \subset \mathbb{R}^q$ are compact sets.

(A3) $K_{\boldsymbol{\theta}} \subset \mathbb{R}^{p+q}$ is an open set, and $\boldsymbol{\theta}_0$ is an interior point of the set $K_{\boldsymbol{\theta}}$.

Furthermore, Assumption (A1) means that $\lambda_{\min}(\boldsymbol{F}_n)$ and n are the same order infinity, which is used to prove Lemmas 4 and 5. Assumption (A2) implies what we deal with are compact regressors. If $\boldsymbol{\theta}$ lies on the boundary of parameter space $K_{\boldsymbol{\theta}}$, the statements of Theorem 2 do not valid anymore.

Based on the assumptions above, we need to prove two preliminary Lemmas 4 and 5 for asymptotic properties of MLE $\widehat{\boldsymbol{\theta}}$ first.

Lemma 4: Under the assumptions (A1)-(A3), there is

$$\boldsymbol{F}_n^{-\top/2}\boldsymbol{s}_n \xrightarrow{D} N_P(\boldsymbol{0}, \boldsymbol{I}_{p+q}) \quad as\ n \to \infty, \tag{47}$$

where $N_P(\boldsymbol{0}, \boldsymbol{I}_{p+q})$ is a $(p+q)$-dimensional normal distribution with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_{p+q}$.

Proof Lemma 4: Derived from Cramer-Wald [20], we only need to prove that a linear combination $\boldsymbol{u}^\top\boldsymbol{F}_n^{-1/2}\boldsymbol{s}_n$ converges in distribution to $N(0, \boldsymbol{u}^\top\boldsymbol{u})$ for any vector $\boldsymbol{u} \in \mathbb{R}^{p+q}(\boldsymbol{u} \ne \boldsymbol{0})$. Without loss of generality, we set $\|\boldsymbol{u}\| = 1$. Then, let

$$\begin{aligned} \boldsymbol{s}_n(\boldsymbol{\theta}) &= \left(\boldsymbol{s}_{\boldsymbol{a}}^\top(\boldsymbol{\theta}), \boldsymbol{s}_{\boldsymbol{b}}^\top(\boldsymbol{\theta})\right)^\top \\ &= \left(s_1(\boldsymbol{\theta}), \ldots, s_p(\boldsymbol{\theta}), s_{p+1}(\boldsymbol{\theta}), \ldots, s_{p+q}(\boldsymbol{\theta})\right)^\top, \end{aligned} \tag{48}$$

where

$$s_r(\boldsymbol{\theta}) := \frac{\partial l_n(\boldsymbol{\theta})}{\partial a_r} = \sum_{i=1}^{n} s_{r,i}(\boldsymbol{\theta})$$

with

$$s_{r,i}(\boldsymbol{\theta}) = (\alpha_i \log\beta_i - \psi_0(\alpha_i)\alpha_i + \alpha_i \log z_i)\, x_{ir} \tag{49}$$

for $r = 1, \ldots, p$, and $\psi_0(\cdot)$ is digamma function (seen in Equation (20)).

$$s_{p+r}(\boldsymbol{\theta}) := \frac{\partial l_n(\boldsymbol{\theta})}{\partial b_r} = \sum_{i=1}^{n} s_{p+r,i}(\boldsymbol{\theta})$$

with

$$s_{p+r,i}(\boldsymbol{\theta}) = (\alpha_i - \beta_i z_i)\, y_{ir} \tag{50}$$

for $r = 1, \ldots, q$.

Now observe that $\boldsymbol{s}_n$ can be written as a sum of independent random vectors, namely

$$\boldsymbol{s}_n = \sum_{i=1}^{n} \boldsymbol{s}_{ni}, \tag{51}$$

where $\boldsymbol{s}_{ni} = (s_{1,i}, \ldots, s_{p,i}, s_{p+1,i}, \ldots, s_{p+q,i})^\top$ with $s_{r,i} := s_{r,i}(\boldsymbol{\theta}_0)$ defined in (49) and (50) for $r = 1, 2, \ldots, p+q$ and $i = 1, \ldots, n$, respectively. Further, define independent random variables ξ_{in} by $\xi_{in} := \boldsymbol{u}^\top\boldsymbol{F}_n^{-1/2}\boldsymbol{s}_{ni}$. Since $E(\xi_{in}) = 0$ and $Var(\sum_{i=1}^{n}\xi_{in}) = 1$, it is enough to show that the Lyapunov condition is satisfied, i.e.

$$L_s := \sum_{i=1}^{n} E|\xi_{in}|^s \xrightarrow{n\to\infty} 0, \quad \exists\, s > 2. \tag{52}$$

Let $s = 3$ (see p. 393, e.g., Hoffmann [26]). Noticing that $\left\|\boldsymbol{F}_n^{-1/2}\right\|^2 = 1/\lambda_{\min}(\boldsymbol{F})$, it follows from (A1) that

$$\begin{aligned} L_3 &\le \sum_{i=1}^{n} E\left(\left\|\boldsymbol{u}^\top\right\|^3 \left\|\boldsymbol{F}_n^{-1/2}\right\|^3 \|\boldsymbol{s}_{ni}\|^3\right) \\ &\le \frac{C}{n^{3/2}} \sum_{i=1}^{n} E\,\|\boldsymbol{s}_{ni}\|^3 \\ &\le \frac{C}{\sqrt{n}} \max_{i=1,\ldots,n} E\,\|\boldsymbol{s}_{ni}\|^3. \end{aligned} \tag{53}$$

Using an extension of the c_r-inequality given by

$$E\left|\sum_{i=1}^{n}\zeta_i\right|^k \leq n^{k-1}\sum_{i=1}^{n}E\left|\zeta_i\right|^k \quad (k>1, k\in\mathbb{R}), \tag{54}$$

to n arbitrary random variables $\zeta_1,\ldots,\zeta_n$ (see p.58, e.g., Petrov [27]) yields that

$$\begin{aligned} E\left\|s_{ni}\right\|^3 &\leq (p+q)^2\left(E|s_{1,i}|^3+E|s_{2,i}|^3+\ldots+E|s_{p+q,i}|^3\right)\\ &\leq C\left(E|s_{1,i}|^3+E|s_{2,i}|^3+\ldots+E|s_{p+q,i}|^3\right). \end{aligned} \tag{55}$$

Thus, it remains to establish that $\max_{i=1,\ldots,n}E|s_{p,i}|$ is uniformly bounded in n for $r=1,\ldots,p+q$. This will be shown for case $r=1,\ldots,p$ and $r=p+1,\ldots,p+q$. The remaining cases can be treated similarly. Without loss of generality, set $r=p$ and $r=p+q$ respectively. Using Lemma 2 and formula (54), we have

$$\begin{aligned} &\max_{i=1,\ldots,n}E|s_{p,i}|^3\\ &=\max_{i=1,\ldots,n}E\left|x_{ip}\left(\alpha_i\log\beta_i-\psi_0(\alpha_i)\alpha_i+\alpha_i\log Z_i\right)\right|^3\\ &=\max_{i=1,\ldots,n}E\left(|x_{ip}|^3|\alpha_i\log\beta_i-\psi_0(\alpha_i)\alpha_i+\alpha_i\log Z_i|^3\right)\\ &\leq C\max_{x\in K_x,y\in K_y}\|x\|^3\left[E\left|\exp(x^{\top}a)y^{\top}b\right|^3\right.\\ &\qquad+E\left|\psi_0(\exp(x^{\top}a))\exp(x^{\top}a)\right|^3\\ &\qquad\left.+E\left|\exp(x^{\top}a)\log Z_i\right|^3\right]\\ &\leq C_1(\theta_0)+C_2(\theta_0)\max_{x\in K_x,y\in K_y}E\left|\log Z_i\right|^3\\ &\leq C_1(\theta_0)+C_2(\theta_0)\max_{x\in K_x,y\in K_y}\sqrt{E(\log^6 Z_i)}\\ &\leq C_3(\theta_0), \end{aligned} \tag{56}$$

where $Z_i \sim Ga(\alpha_i,\beta_i)$ for $i=1,\ldots,n$ and $Z\sim Ga(\exp(x^{\top}a),\exp(y^{\top}b))$, and

$$\begin{aligned} &\max_{i=1,\ldots,n}E|s_{p+q,i}|^3\\ &=\max_{i=1,\ldots,n}E\left(y_{iq}(\alpha_i-\beta_iZ_i)\right)\\ &\leq\max_{i=1,\ldots,n}E\left(|y_{iq}|^3\cdot|\alpha_i-\beta_iZ_i|^3\right)\\ &\leq\max_{x\in K_x,y\in K_y}\|y\|^3E|\exp(x_i^{\top}a)-\exp(y_i^{\top}b)Z_i|^3\\ &\leq C\max_{x\in K_x,y\in K_y}\|y\|^3\left(E|\exp(x_i^{\top}a)|^3+E|\exp(y_i^{\top}b)Z_i|^3\right)\\ &\leq C_1(\theta_0)+C_2(\theta_0)\max_{x\in K_x,y\in K_y}EY_i^3\\ &\leq C_3(\theta_0). \end{aligned} \tag{57}$$

We can get

$$L_3\xrightarrow{n\to\infty}0, \tag{58}$$

then

$$F_n^{-\top/2}s_n\xrightarrow{D}N_P(0,I_{p+q}) \tag{59}$$

■

Lemma 5: Under the assumptions (A1)-(A3)

$$\max_{\theta\in N_n(\varepsilon)}\left\|V_n(\theta)-I_{p+q}\right\|\xrightarrow{P}0,\quad\forall\,\varepsilon>0, \tag{60}$$

where $V_n(\theta):=F_n^{-1/2}H_n(\theta)F_n^{-\top/2}$ for $n=1,2,\ldots$

Proof Lemma 5: We have a.s.

$$\begin{aligned} \left\|V_n(\theta)-I_{p+q}\right\| &=\left\|F_n^{-1/2}(H_n(\theta)-F_n)F_n^{-\top/2}\right\|\\ &\leq\frac{1}{\lambda_{\min}(F_n)}\|H_n(\theta)-F_n\|\\ &\leq\frac{C}{n}\|H_n(\theta)-F_n\| \end{aligned} \tag{61}$$

Thus, conditions

$$\max_{\theta\in N_n(\varepsilon)}\left\|\frac{1}{n}(H_n(\theta)-EH_n(\theta))\right\|\xrightarrow{p}0 \tag{62}$$

and

$$\max_{\theta\in N_n(\varepsilon)}\left\|\frac{1}{n}(EH_n(\theta)-F_n)\right\|\xrightarrow{p}0 \tag{63}$$

imply (60).

To prove (62), it is sufficient to establish that the (j,k)-element of the random matrix $(H_n(\theta)-EH_n(\theta))/n$ converges to zero in probability, i.e.

$$\max_{\theta\in N_n(\varepsilon)}\left|\frac{h_{j,k}(\theta)-Eh_{j,k}(\theta)}{n}\right|\xrightarrow{p}0 \tag{64}$$

There are three different types of entries (14), (15) and (16) in the Hessian matrix. We will show the convergence of formula (64) in the cases of $1\leq j,k\leq p$. It is similar to treat the remaining cases. In order to avoid generality, let $j=p$ and $k=p$, then

$$\max_{\theta\in N_n(\varepsilon)}\frac{1}{n}\left|h_{p,p}(\theta)-Eh_{p,p}(\theta)\right|\xrightarrow{p}0. \tag{65}$$

We have the following bounds:

$$\begin{aligned} &\max_{\theta\in N_n(\varepsilon)}\frac{1}{n}\left|h_{p,p}(\theta)-Eh_{p,p}(\theta)\right|\\ &=\max_{\theta\in N_n(\varepsilon)}\frac{1}{n}\left|\sum_{i=1}^{n}x_{ip}x_{ip}e^{x_i^{\top}a}(\log(Z_i)-E\log(Z_i))\right|\\ &\leq\max_{\theta\in N_n(\varepsilon)}\max_{x\in K_x,y\in K_y}C\sum_{i=1}^{n}\left|\frac{\log(Z_i)-E\log(Z_i)}{n}\right|\\ &=\max_{\theta\in N_n(\varepsilon)}\max_{x\in K_x,y\in K_y}CG_n \end{aligned} \tag{66}$$

From the law of large numbers and standard arguments, we can get $G_n\to 0$ in probability as $n\to\infty$. It remains to show

$$\max_{\theta\in N_n(\varepsilon)}\left|\frac{1}{n}(f_{p,p}(\theta)-f_{p,p}(\theta_0))\right|$$

TABLE 1. Average estimate, standard deviation, estimated MSE, MAE and p-value of $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$ and $\hat{b}_2$ for a GaGLM model on the basis of 100 replications.

Parameter	True value	n	Mean	MSE	MAE	p-value	Normality test
a_0	-1	50	-0.9868	0.2660	0.2158	0.0123	NO
		100	-0.9798	0.1717	0.1344	0.2690	YES
		200	-0.9972	0.1139	0.0902	0.9717	YES
		500	-0.9975	0.0744	0.0577	0.0141	NO
		1000	-0.9969	0.0507	0.0411	0.0064	NO
		2000	-1.0011	0.0316	0.0249	0.8457	YES
a_1	1	50	0.9561	0.4433	0.3659	0.7720	YES
		100	0.9446	0.2622	0.2095	0.6501	YES
		200	1.0040	0.1716	0.1346	0.0473	NO
		500	0.9905	0.1354	0.1110	0.1552	YES
		1000	1.0114	0.0844	0.0672	0.3956	YES
		2000	1.0009	0.0635	0.0488	0.3657	YES
a_2	3	50	2.9553	0.4785	0.3861	0.8913	YES
		100	2.9721	0.2598	0.2129	0.1981	YES
		200	3.0030	0.1837	0.1367	0.7403	YES
		500	2.9869	0.1232	0.0997	0.5261	YES
		1000	2.9869	0.0888	0.0689	0.7357	YES
		2000	2.9785	0.0634	0.0520	0.0964	YES
b_1	1	50	1.0854	0.5467	0.4576	0.0487	NO
		100	1.0024	0.3859	0.3081	0.0176	NO
		200	1.0014	0.2281	0.1828	0.3867	YES
		500	0.9973	0.1647	0.1297	0.6182	YES
		1000	0.9922	0.1122	0.0911	0.0717	YES
		2000	0.9852	0.0748	0.0586	0.8467	YES
b_2	3	50	3.0246	0.5903	0.4588	0.7750	YES
		100	3.0225	0.3148	0.2526	0.2990	YES
		200	3.0388	0.2618	0.2072	0.5040	YES
		500	2.9603	0.1706	0.1345	0.2272	YES
		1000	2.9592	0.1230	0.1006	0.4994	YES
		2000	2.9848	0.0933	0.0715	0.4488	YES

$$
\begin{aligned}
&= \max_{\theta\in N_n(\varepsilon)} \left| \frac{1}{n}\sum_{i=1}^{n} x_{ip}x_{ip}\left(e^{2\mathbf{x}_i^\top \boldsymbol{a}}\psi_1(e^{\mathbf{x}_i^\top \boldsymbol{a}}) \right.\right.\\
&\qquad\qquad \left.\left. -e^{2\mathbf{x}_i^\top \boldsymbol{a}_0}\psi_1(e^{\mathbf{x}_i^\top \boldsymbol{a}_0})\right)\right| \\
&\le \max_{\theta\in N_n(\varepsilon)} \frac{1}{n}\sum_{i=1}^{n} \left|x_{ip}x_{ip}\right| \cdot \left|e^{2\mathbf{x}_i^\top \boldsymbol{a}}\psi_1(e^{\mathbf{x}_i^\top \boldsymbol{a}}) \right.\\
&\qquad\qquad \left. -e^{2\mathbf{x}_i^\top \boldsymbol{a}_0}\psi_1(e^{\mathbf{x}_i^\top \boldsymbol{a}_0})\right| \\
&\le \frac{C}{n}\sum_{i=1}^{n} \left|x_{ip}x_{ip}\right| \max_{\theta\in N_n(\varepsilon)} \max_{\boldsymbol{x}\in K_x,\boldsymbol{y}\in K_y} \left|e^{2\mathbf{x}^\top \boldsymbol{a}}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}}) \right.\\
&\qquad\qquad \left. -e^{2\mathbf{x}^\top \boldsymbol{a}_0}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}_0})\right| \\
&\le C_1 \max_{\theta\in N_n(\varepsilon)} \max_{\boldsymbol{x}\in K_x,\boldsymbol{y}\in K_y} \left|e^{2\mathbf{x}^\top \boldsymbol{a}}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}}) \right.\\
&\qquad\qquad \left. -e^{2\mathbf{x}^\top \boldsymbol{a}_0}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}_0})\right| \\
&=: CG_{2n}
\end{aligned} \tag{67}
$$

The continuity in $\boldsymbol{\theta}$ of the function $\max_{\boldsymbol{x}\in K_x,\boldsymbol{y}\in K_y}\left|e^{2\mathbf{x}^\top \boldsymbol{a}}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}}) - e^{2\mathbf{x}^\top \boldsymbol{a}_0}\psi_1(e^{\mathbf{x}^\top \boldsymbol{a}_0})\right|$ with value zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields that G_{2n} converges to 0 in probability as $n \to \infty$. ■

Theorem 2: Under the assumptions (A1)-(A3), the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ can be obtained as the following

$$\boldsymbol{F}_n^{\top/2} \cdot (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N_P(\mathbf{0}, \boldsymbol{I}_{p+q}) \tag{68}$$

Proof Theorem 2: With the mean value theorem, we have

$$\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{F}_n(\boldsymbol{\theta}_0 + \tau(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)) \cdot (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{69}$$

for $0 < \tau < 1$ and $s_n(\widehat{\boldsymbol{\theta}}) = 0$. By pre-multiplying $\boldsymbol{F}_n^{\top/2}$ and integrating with respect to τ on [0, 1], we have

$$\boldsymbol{F}_n^{-1/2}\boldsymbol{s}_n = \left[\int_0^1 \boldsymbol{V}_n(\boldsymbol{\theta}_0 + \tau(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))d\tau\right] \cdot \boldsymbol{F}_n^{\top/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{70}$$

Meanwhile, Lemma 5 implies that

$$\int_0^1 \boldsymbol{V}_n(\boldsymbol{\theta}_0 + \tau(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))d\tau \xrightarrow{p} \boldsymbol{I}_{p+q} \tag{71}$$

By using Lemma 4 and the continuous mapping theorem [28], the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ can be proved. ■

V. SIMULATION STUDY

In this section, we will provide some simulation experiments to illustrate our asymptotic theory and stability results.

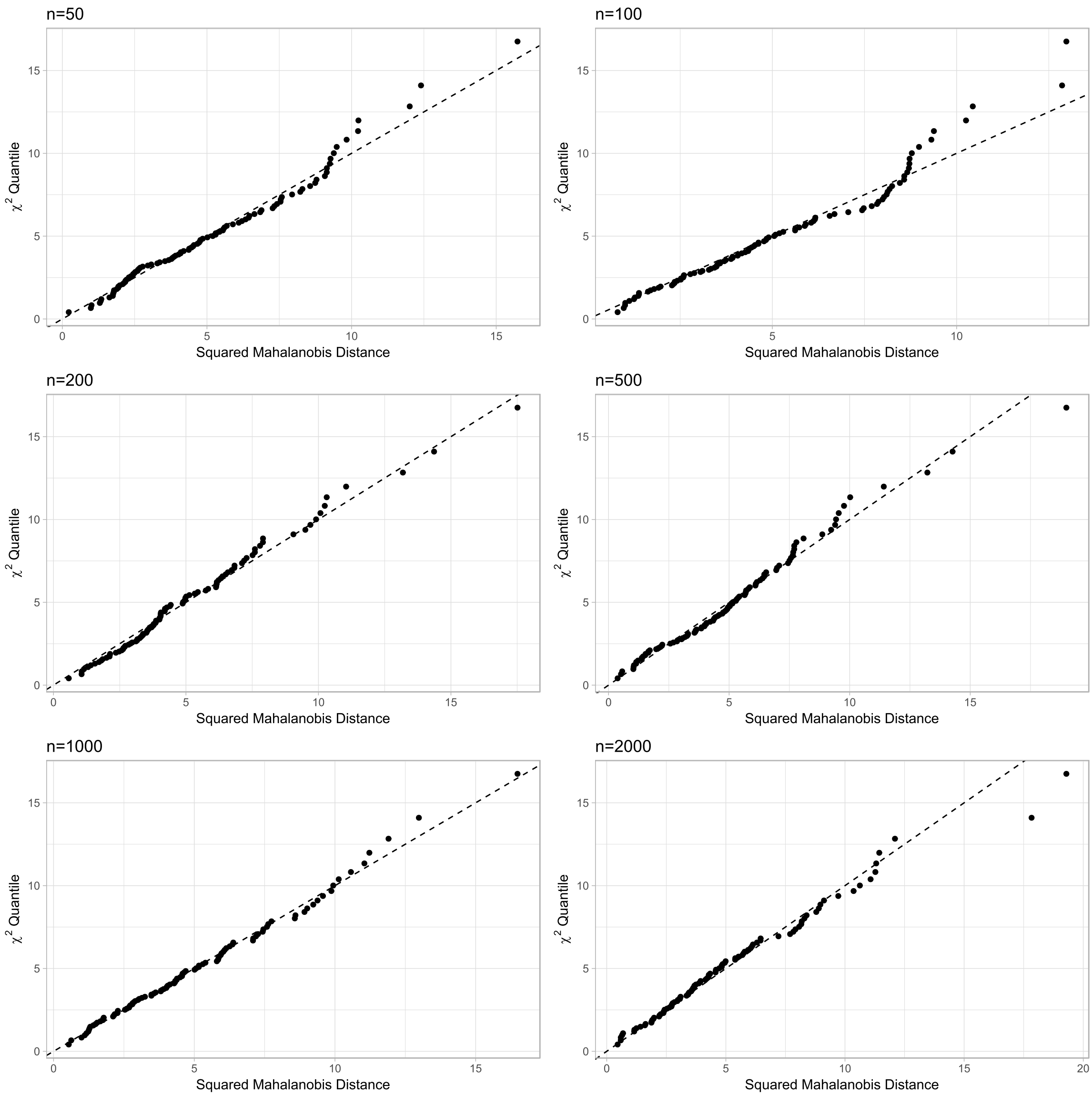


FIGURE 1. **Normal QQ-plots of centered and normalized ML estimators for a *Ga*GLM model based on 100 replications.**

A. FISHER's SCORING METHOD

The FS method is an efficient iterative algorithm for attempting to find the roots of a function $s_n(\boldsymbol{\theta})$ by choosing a starting value $\widehat{\boldsymbol{\theta}}_{(0)}$. The method for the score function is finding an iterative solution to the likelihood equations. As the modification of the NR method, the FS algorithm is an iterative method for finding the roots of a differentiable function that generates a sequence of estimates which usually come increasingly close to the optimal solution. The iteration is

$$\widehat{\boldsymbol{\theta}}_{(j+1)} = \widehat{\boldsymbol{\theta}}_{(j)} + \boldsymbol{F}_n^{-1}(\widehat{\boldsymbol{\theta}}_{(j)})\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_{(j)}), \tag{72}$$

which is the jth iteration of the FS algorithm based on the observed Fisher information (OFI) matrix for estimating the parameters in the *Ga*GLM.

B. RESULTS FOR ASYMPTOTIC THEORY

The numerical simulation based on Theorem 2 was conducted for the verification of the asymptotic properties of MLE of *Ga*GLM. To this end, the coefficients of *Ga*GLM were estimated from various data sets independently generated by the same system. Then, the distributions of the estimated coefficients were compared with the normal distributions. Such experiments were repeated for different numbers of measurements n sized as $\{50, 100, 200, 500, 1000, 2000\}$. In this way, the relation between the convergence and n can be investigated.

A simple model with intercept and covariate 1, x_1, x_2, y_1 and y_2 were considered for the linear predictors $\boldsymbol{\eta}_i(\boldsymbol{\theta})'s$, i.e., $\eta_{\alpha i}(\boldsymbol{\theta}) = a_0 + a_1x_{i1} + a_2x_{i2}$ and $\eta_{\beta i}(\boldsymbol{\theta}) = b_1y_{i1} + b_2y_{i2}$ for $i = 1, \ldots, n$. The values of the covariate x_1, x_2, y_1 and

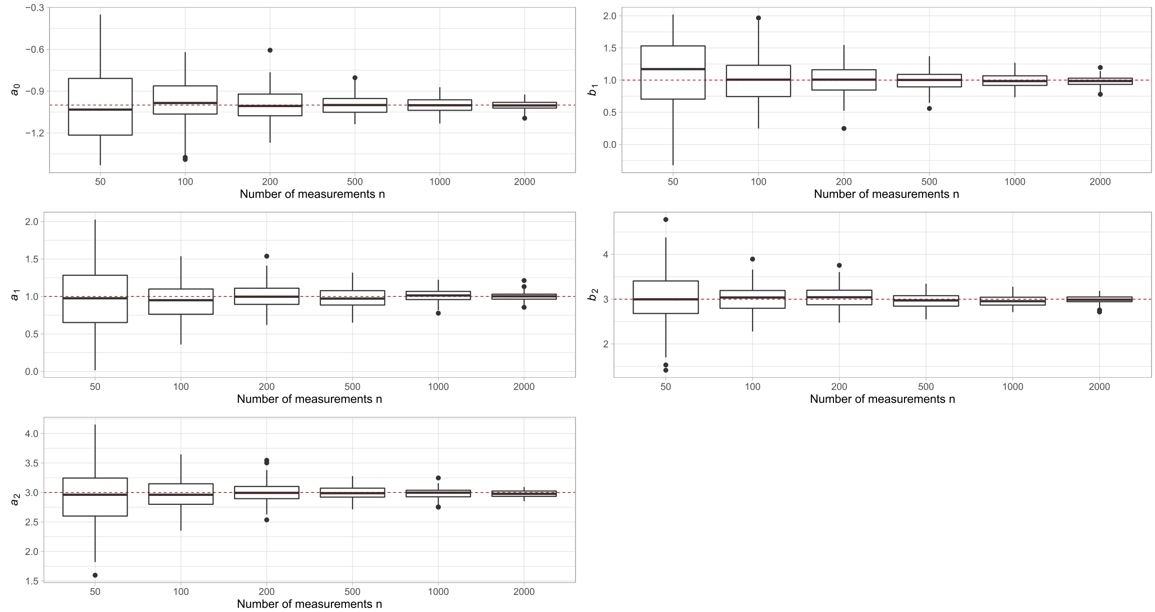


FIGURE 2. Box plots of estimated coefficients $\widehat{\theta}$.

y_2 were chosen equally spaced between -1 and 1. Further, for distinguishing the effects of different size parameters on the results, we examined set $a_0 = -1$, $a_1 = 1$, $a_2 = 3$, $b_1 = 1$ and $b_2 = 3$. Since we are also interested in the case when *Ga*GLM does not satisfactorily fit the count regression data. For each combination of sample size n, setting, we simulated 100 samples of responses Z_i's, i.e., $Z_i \sim Ga(\exp(a_0 + a_1x_{i1} + a_2x_{i2}), \exp(b_1y_{i1} + b_2y_{i2}))$ for $i = 1, \ldots, n$.

Henze-Zirkler test [29]was used for normality test of $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$ and $\hat{b}_2$ respectively, in which p-value larger than 0.05 indicates the data fitting well to normal distributions. We computed the average estimate, the estimated mean squared error (MSE) and the mean absolute error (MAE) to indicate the convergence status with the increase of sample numbers in 100 replications for each considered case, shown in Table 1. Simulation results reveal that the average estimate of each parameter close to the true value roughly as the sample size n increases. With n increase, the truncation error in the iterative process affects the estimation accuracy. The MSE and MAE decrease strictly as the number of samples increasing, demonstrate similar patterns.

We also test the normality of each parameter with estimating result by 100 replications. Due to the same range of randomly generated samples, when the number of samples is limited, i.e. $n = 50, 100$ and 200, the estimation of the smaller parameters will be affected by the bigger parameters. As the number of samples increases, when the number of samples reaches $n = 2000$, the estimated value tends to be stable and presents a normal distribution.

A normal quantile-quantile (QQ) plots for the empirical distribution of multi-normal components are illustrated in figure 1. When the sample size is $n = 50$ and 100, there are more outliers in the multivariate normal QQ plots. When the sample size increases to $n = 200$ and 500, the QQ plots tend to be stable, when $n = 1000$ and 2000, the QQ plots are normally distributed.

Figure 2 illustrates the convergence of parameters estimation by different size of samples. With the increase of samples, the mean value of each estimated parameter gradually approaches the real value, and the fluctuation range gradually decreases, that is, the variance decreases. Therefore, it indicates that the parameter estimation value converges to the actual value, and the maximum likelihood estimate is consistent. Among them, it can be seen from the figure that due to the difference in parameter size, the bigger parameter is easier to converge to the real value, which has the significance of estimation.

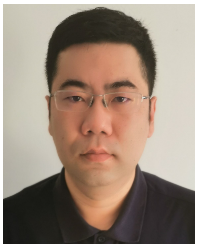
VI. CONCLUSION

*Ga*GLM is a kind of generalized regression model for investigating positive real data. This research obtained several theoretical results of MLE for *Ga*GLM. The score function and Fisher information matrix were derived. Then, the Lyapunov conditions were obtained to ensure the asymptotic normality of the score function normalized by the Fisher information matrix. Consequently, the asymptotic normality of MLE for *Ga*GLM was proved.

In the derivation process, we also discussed the range of the logarithmic k-order expectation $E(\log^k Z)$ when Z obey the Gamma distribution. And we proved the inequality holds on the trigamma function ψ_1 that $\alpha\psi_1(\alpha) > 1$ where $\alpha > 0$. Moreover, the simulation study illustrates that the normal approximation is satisfactory for moderate and large sample sizes. Finally, with the established asymptotic theory, we can further benefit interval estimates [30], hypothesis tests [31]and stochastic control design [32]in a theoretical basis.

REFERENCES

[1] J. A. Nelder and R. W. Wedderburn, "Generalized linear models," *J. Roy. Stat. Soc., A, Gen.*, vol. 135, no. 3, pp. 370–382, Apr. 1972.
[2] E. W. Stacy, "A generalization of gamma distribution," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1187–1192, 1962.
[3] S. K. Agarwal and S. L. Kalla, "A generalized gamma distribution and its application in reliabilty," *Commun. Statist.-Theory Methods*, vol. 25, no. 1, pp. 201–210, 1996.
[4] C. R. M. de Andrade, E. L. C. Silva, M. D. F. B. da Matta, M. B. Castier, M. L. G. Rosa, and M. D. B. Gomes, "Influence of a family history of type 2 diabetes, demographic and clinical data on carotid intima-media thickness in patients with type 1 diabetes: A cross-sectional study," *Cardiovascular Diabetol.*, vol. 13, no. 1, p. 87, May 2014.
[5] S. Mandal, R. A. Belaghi, A. Mahmoudi, and M. Aminnejad, "Stein-type shrinkage estimators in gamma regression model with application to prostate cancer data," *Statist. Med.*, vol. 38, no. 22, pp. 4310–4322, Jul. 2019.
[6] S. Ecker, V. Pancaldi, D. Rico, and A. Valencia, "Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia," *Genome Med.*, vol. 7, no. 1, pp. 1–12, 2015.
[7] K. Summun, N. M. Khan, Z. Jannoo, Y. Sunecher, and I. Veerasamy, "An assessment of the determinants of mauritian automobile insurance claims using negative binomial and gamma regression models," *J. Statist. Manage. Syst.*, vol. 21, no. 5, pp. 725–740, Aug. 2018.
[8] G. O. Agogo, "A zero-augmented generalized gamma regression calibration to adjust for covariate measurement error: A case of an episodically consumed dietary intake," *Biometrical J.*, vol. 59, no. 1, pp. 94–109, Jan. 2017.
[9] E. I. George, "The variable selection problem," *J. Amer. Statist. Assoc.*, vol. 95, no. 452, pp. 1304–1308, 2000.
[10] S. A. Culpepper, "Studying individual differences in predictability with gamma regression and nonlinear multilevel models," *Multivariate Behav. Res.*, vol. 45, no. 1, pp. 153–185, Jan. 2010.
[11] P. McCullagh and J. Nelder, *Generalized Linear Models II*. London, U.K.: Chapman & Hall, 1989.
[12] M. Kazeminia and M. Mehrjoo, "A new method for maximum likelihood parameter estimation of gamma-gamma distribution," *J. Lightw. Technol.*, vol. 31, no. 9, pp. 1347–1353, May 1, 2013.
[13] T. J. Ypma, "Historical development of the Newton–Raphson method," *SIAM Rev.*, vol. 37, no. 4, pp. 531–551, 1995.
[14] J. Guo, L. Y. Wang, G. Yin, Y. Zhao, and J.-F. Zhang, "Asymptotically efficient identification of FIR systems with quantized observations and general quantized inputs," *Automatica*, vol. 57, pp. 113–122, Jul. 2015.
[15] L. Fahrmeir and H. Kaufmann, "Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models," *Ann. Statist.*, vol. 13, no. 1, pp. 342–368, 1986.
[16] R. Ash and C. Doleans-Dade, *Probability and Measure Theory*. New York, NY, USA: Academic, 2000.
[17] A. Leucht, "Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency," *Bernoulli*, vol. 18, no. 2, pp. 552–585, 2012.
[18] Y. Zeng, H. Li, N. M. Schweppe, R. S. Hawley, and W. D. Gilliland, "Statistical analysis of nondisjunction assays in drosophila," *Genetics*, vol. 186, no. 2, pp. 505–513, Oct. 2010.
[19] R. I. Jennrich and P. F. Sampson, "Newton–Raphson and related algorithms for maximum likelihood variance component estimation," *Technometrics*, vol. 18, no. 1, pp. 11–17, 1976.
[20] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton Univ. Press, 1946.
[21] K.-Y. Liang, "The asymptotic efficiency of conditional likelihood methods," *Biometrika*, vol. 71, no. 2, pp. 305–313, 1984.
[22] T.-Y. Hwang and P.-H. Huang, "On new moment estimation of parameters of the gamma distribution using its characterization," *Ann. Inst. Statist. Math.*, vol. 54, no. 4, pp. 840–847, 2002.
[23] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[24] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. London, U.K.: Academic, 1979.
[25] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, NY, USA: Springer-Verlag, 2006.
[26] J. Hoffmann-Jorgensen, *Probability With a View Toward Statistics*. New York, NY, USA: Chapman & Hall, 1994.
[27] V. V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Variables*. Oxford, U.K.: Clarendon, 1995.
[28] P. Billingsley, *Convergence of Probability Measures*. Hoboken, NJ, USA: Wiley, 2013.
[29] N. Henze and B. Zirkler, "A class of invariant consistent tests for multivariate normality," *Commun. Statist., Theory Methods*, vol. 19, no. 10, pp. 3595–3617, 1990.
[30] Z. Yan and N. Wang, "Statistical analysis based on adaptive progressive hybrid censored sample from alpha power generalized exponential distribution," *IEEE Access*, vol. 8, pp. 54691–54697, 2020.
[31] C. Xu, "Detection test for periodic signals revisited against various stochastic models," *IEEE Access*, vol. 7, pp. 92203–92209, 2019.
[32] S. Luo, F. Deng, and X. Yu, "Unified stability analysis for ito stochastic systems: From almost surely asymptotic to finite-time convergence," *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 406–412, Jan. 2022.



BENCHAO WANG received the M.Eng. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2012, where he is currently pursuing the Ph. D. degree with the Faculty of Electronic Information and Electrical Engineering.

He is currently a Teacher with the Liaoning Police College, Dalian. His current research interests include statistical modeling and stochastic control theory.

PAN QIN received the Ph.D. degree in information science from Kyushu University, Fukuoka, Japan, in 2008.

He was a Research Fellow with the Faculty of Mathematics, Institute of Mathematics for Industry, Kyushu University, from 2008 to 2013. He is currently an Associate Professor with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China. His current research interests include statistical modeling and stochastic control theory.

HONG GU received the B.Sc. and M.Sc. degrees from the Shenyang University of Technology, Shenyang, China, in 1982 and 1984, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1990.

Since 1996, he has been a Full Professor with the Dalian University of Technology, Dalian, China. His current research interests include machine learning, big data, and bioinformatics.