International Journal of Mathematical Archive-8(5), 2017, 95-114
Available online through www.ijma.info ISSN 2229 – 5046

ASYMPTOTIC PROPERTIES OF THE LOGISTIC AND MULTINOMIAL LOGISTIC REGRESSION MODELS AND ITS APPLICATIONS ON SIMULATION APPROACH

K. NAGENDRA KUMAR*[1], B. MUNISWAMY[2]

**Department of Statistics,
Andhra University, Visakhapatnam-530 003, Andhra Pradesh, India.**

(Received On: 19-05-17; Revised & Accepted On: 05-06-17)

ABSTRACT

Logistic regression is extensively used as a ideal model for the analysis of binary data with the areas of applications including physical, medical and social sciences. In this paper simulation studies are used to estimate the effect of varying sample size, is increases. To assess the accuracy of the estimated parameters and variance components of the logistic and multinomial logistic regression model. As well as the maximum likelihood procedure for the estimation of its parameters are introduced in detail. The essential assumptions are underlying conventional results on the properties of maximum likelihood estimate of the stochastic which determines the behaviour of the observable facts investigated, is known to lie within a specified parameter family of probability distribution. The outcome of the simulation studies are performs we, in performance of the consistency and normality of the Maximum Likelihood Estimation for various sample sizes.

Keywords: *Logistic Regression, Maximum Likelihood Estimation, Multinomial Logistic Regression Model, Consistency and Normality.*

1. INTRODUCTION

Regression analysis is one of the most useful and the most frequently used statistical methods (David Collett, 2002). regression analysis is a form of predictive modeling technique which investigates the relationship between a dependent and one or more predictor variables. Among the different regression models, logistic regression plays a particular role. However, the basic concept of the linear regression model is quantifying the effect of several explanatory variables on one dependent continuous variable. For situations where the dependent variable is qualitative, however, other methods have been developed. One of the method is logistic regression model, which specifically covers the case of binary or dichotomous response (Givens, G.H. and Hoeting. J.A, 2005).

The statistical analysis of dichotomous outcome variable is frequently interpreted with the use of logistic regression methods. The multiple logistic regression model is a commonly applied procedure for describing the relationship between a dichotomous outcome variable such as presence or absence of disease, and a number of independent variables known as potential risk factors.

Logistic regression analysis is a statistical modeling method for analyzing categorical outcome variable. This statistical model describes the relationship between a categorical response variable and a set of explanatory variables. The response variable in logistic regression model is usually dichotomous, but more than two response options can be modelled using multinomial or polytomous logistic regression model.

2. LITERATURE REVIEW

Regression analysis is one of the most useful and the most frequently used statistical methods (Efron and Tibsirani, 1993). The aim of the regression methods is to describe the relationship between a response variable and one or more explanatory variables. Among the different regression models, logistic regression plays a particular role. The basic concept, however, is universal. The linear regression model is, under certain conditions, in many circumstances a valuable tool for quantifying the effects of several explanatory variables on one dependent continuous variable. For situations where the dependent variable is qualitative, however, other methods have been developed. These methods are the least squares and the discriminent function analysis.

Corresponding Author: K. Nagendra Kumar[1]*
Department of Statistics, Andhra University, Visakhapatnam-530 003, Andhra Pradesh, India.

The discriminent approach to estimation of the coefficients is of historical importance as popularized by (Cox, D. R., and Snell, E. J, 1989). (J.Wilde, 2008) compared the two methods when the model is dichotomous and concluded that the discriminate function was sensitive to the assumption of normality. In particular, the estimation of the coefficient for the non normal distributed variables are biased away from zero, when the coefficient function will over estimate the magnitude of the dichotomous independent variable the discriminant function will overestimate the magnitude of the coefficient.

One of these is the logistic regression model, which specifically covers the case of a binary (dichotomous) response. Cramer (2003) discussed an overview of the development of the logistic regression model. He identifies three sources that had a profound impact on the model: applied mathematics, experimental statistics, and economic theory. Agresti (2002) also provided details of the development on logistic regression in different areas. He states, "Sir David R. Cox introduced many statisticians to logistic regression through his 1958 article and 1970 book, The Analysis of Binary Data." However, logistic regression is widely used as a popular model for the analysis of binary data with the areas of applications including physical, biomedical, and behavioral sciences. For example, Cornfield (1962) presented the preliminary results from the Framingham Study. The purpose of the study was to find the roles of risk factors of cholesterol levels (low versus high values) and blood pressure (low versus high values) in the development of coronary heart disease (yes or no) in the population of the town.

According to (D.W.Hosmer and S.Lemeshow, 1989), the fact concerning the interpretability of the coefficients is the fundamental reason why logistic regression has proven such a powerful analytic tool for epidemiologic research. At least, this argumentation holds whenever the explanatory variable X are quantitative. (Hosmer, D. and Lemeshow, 2000) Investigate the asymptotic properties of various discrete and qualitative response models and provided conditions under which the MLE has its usual asymptotic properties, that is, the p-vector β of coefficients of linear combinations (X, β) has to be estimated from a finite sample of n observations. The method of analysis of generalized linear models can be used since logistic models are sub-category (P.McCullagh and J.A.Nelder, 1989).

The asymptotic normality of the maximum likelihood in logistic regression models are also found in (B. Munsiwamy and Shibru Temesgen Wakweya, 2011). (Anthony. N, 2014) presents regularity conditions for a multinomial response model when the logit link is used. (L.Nordberg, 1980) presents regularity conditions that assure asymptotic normality for the logit link in binomial response models and further verifies that his conditions are equivalent to those of (L.McFadden,1974). (C.Gourienx and A.Monfort,1981) discuss the asymptotic distribution of the MLE for constructing confidence intervals and conducting tests of hypotheses. (C.Gourienx, 1981) prove that the MLE is asymptotically normal in this setting as long as certain regularity conditions are satisfied.

3. MULTINOMIAL LOGISTIC REGRESSION

Polytomous Logistic Regression is the extension for the (binary) logistic regression when the categorical dependent outcome has more than two levels. One possible way to handle such situations is to split the categorical response variable in several ways and apply binary logistic regression to each dichotomous variable. However, this will result in several different analysis for only one categorical response. A more structured approach is to formulate one model for the categorical response by means of so called generalized logits. Suppose that Y has k categories and the probability for category i is given by

$$P(Y=i) = p_i, \qquad i=1, 2,\ldots,k$$

Then the k generalized logits are defined by

$$\log it(Y=i) = \ln\left(\frac{P_i}{1-(P_1+P_2+\ldots..P_{k-1})}\right) = \ln\left(\frac{P_i}{P_k}\right) \tag{3.1}$$

This means that the generalized logits relate the probabilities p_i for the categories I = 1, 2,….k-1 to the reference category k. For m categories the general polytomous logistic regression model becomes

$$\log it(Y=i) = \gamma_{i0} + \gamma_{i1}X_1 + \gamma_{i2}X_2 + \qquad \ldots.. + \gamma_{im}X_m, i=1,2,\ldots..k-1 \tag{3.2}$$

Note that the polytomous logistic model is given by k-1 equations if Y has k categories and that we have one logistic coefficient γ_{ij} for each category or covariate combination.

If a sample belongs to a special class Y_i with probability p_i it has odds $\left(\frac{p_i}{1-p_i}\right)$

The vector x_i consist out of the data for sample i , odds have the range $(0, \infty)$

We also make some additional assumptions.

The response Y_i is Bernoulli distribution such that

$$P(y_i = 1 | x_i) = p$$

The dichotomous logistic regression principle is described by a linear predictor

$$\ln\left(\frac{p_i}{1-p_i}\right) = \gamma_0 + x_i^T\gamma..... \quad (3.3)$$

The linear predictor is described by assume $\eta_i = \gamma_0 + x_i^T\gamma$ under conventional notation

$$\eta_i = [1 ... x_i]\begin{bmatrix}\gamma_0 \\ \gamma\end{bmatrix}$$

Using the definitions and assumptions stated above. We can define the equation for the dichotomous logistic regression

$$P(y_i = 0 | x) + P(y_i = 1 | x) = 1.... \quad (3.4)$$

We can show, for the logistic regression equation of $P(y_i = 0 | x)$

$$\ln\left(\frac{P(y_i = 1 | x)}{1 - P(y_i = 0 | x)}\right) = X^T\gamma$$

$$\Rightarrow \quad \ln\left(\frac{1 - P(y_i = 0 | x)}{P(y_i = 0 | x)}\right) = X^T\gamma$$

$$\Rightarrow \quad \ln\left(\frac{1}{P(y_i = 0 | x)} - 1\right) = X^T\gamma$$

$$\Rightarrow \quad \frac{1}{P(y_i = 0 | x)} - 1 = e^{X^T\gamma}$$

$$\Rightarrow \quad \frac{1}{P(y_i = 0 | x)} = e^{X^T\gamma} + 1$$

$$\Rightarrow \quad P(y_i = 0 | x) = \frac{1}{e^{X^T\gamma} + 1}$$

Now $P(y_i = 1 | x) = 1 - P(y_i = 0 | x)$

$$= 1 - \frac{1}{e^{X^T\gamma} + 1}$$

$$= \frac{e^{X^T\gamma}}{e^{X^T\gamma} + 1}$$

Now

$$\ln\left(\frac{P(y_i = 1 | x)}{P(y_i = 0 | x)}\right)$$

$$\Rightarrow \quad \ln\left(\frac{e^{X^T\gamma}}{e^{X^T\gamma} + 1} \Big/ \frac{1}{e^{X^T\gamma} + 1}\right)$$

$$\Rightarrow \quad \ln\left(\frac{e^{X^T\gamma}}{e^{X^T\gamma} + 1} . e^{X^T\gamma} + 1\right)$$

$$\Rightarrow \quad \ln(e^{X^T\gamma}) = X^T\gamma$$

$$\ln\left(\frac{P(y_i=1|x_i^T)}{P(y_i=0|x_i^T)}\right)=X_i^T\gamma_1 \ldots.. \quad (3.5)$$

Similarly $\ln\left(\frac{P(y_i=2|x_i^T)}{P(y_i=0|x_i^T)}\right)=X_i^T\gamma_2 \ldots..$ (3.6)

In this case we have taken the group y_i=0 as reference category

Using the definition

$$P(y_i=0|x_i^T)+P(y_i=1|x_i^T)+P(y_i=2|x_i^T)=1$$

Now $\frac{P(y_i=1|x_i^T)}{P(y_i=0|x_i^T)}=e^{X_i^T\gamma_1} \ldots.$ (3.7)

and $\frac{P(y_i=2|x_i^T)}{P(y_i=0|x_i^T)}=e^{X_i^T\gamma_2} \ldots..$ (3.8)

We can also show that

$$
\begin{aligned}
l\,\mathrm{n}\left[\frac{p(y_i=2|x_i^T)}{p(y_i=1|x_i^T)}\right] &= \log\left[p(y_i=2|x_i^T)\right]-\log\left[p(y_i=1|x_i^T)\right]\\
&=\ln\left[\frac{p(y_i=2|x_i^T)\,p(y_i=0|x_i^T)}{p(y_i=0|x_i^T)}\right]-\ln\left[\frac{p(y_i=1|x_i^T)\,p(y_i=0|x_i^T)}{p(y_i=0/x_i^T)}\right]\\
&=\ln\left[\frac{p(y_i=2|x_i^T)}{p(y_i=0|x_i^T)}\right]+\ln\left[\frac{p(y_i=0|x_i^T)}{p(y_i=0|x_i^T)}\right]-\ln\left[\frac{p(y_i=1|x_i^T)}{p(y_i=0|x_i^T)}\right]-\ln\left[\frac{p(y_i=0|x_i^T)}{p(y_i=0|x_i^T)}\right]\\
&=\ln\left[\frac{p(y_i=2|x_i^T)}{p(y_i=0|x_i^T)}\right]-\ln\left[\frac{p(y_i=1|x_i^T)}{p(y_i=0|x_i^T)}\right]\\
&=\ln\left[e^{X_i^T\gamma_2}\right]-\ln\left[e^{X_i^T\gamma_1}\right]\\
&=X_i^T\gamma_2-X_i^T\gamma_1\\
&=X_i^T\left[\gamma_2-\gamma_1\right]
\end{aligned}
$$

Thus $\frac{p(y_i=2|x_i^T)}{p(y_i=1|x_i^T)}=e^{X_i^T[\gamma_2-\gamma_1]}$ (3.9)

Now defined all preliminary equations we can, now derive the logit functions of each individual outcome category

$$
\begin{aligned}
&\frac{p(y_i=2|x_i^T)}{p(y_i=1|x_i^T)}=\frac{1-p(y_i=1|x_i^T)-p(y_i=0|x_i^T)}{p(y_i=1|x_i^T)}=e^{X_i^T[\gamma_2-\gamma_1]}\\
&\Rightarrow\frac{1}{p(y_i=1|x_i^T)}-\frac{p(y_i=1|x_i^T)}{p(y_i=1|x_i^T)}-\frac{p(y_i=0|x_i^T)}{p(y_i=1|x_i^T)}=e^{X_i^T[\gamma_2-\gamma_1]}\\
&\Rightarrow\frac{1}{p(y_i=1|x_i^T)}-1-\frac{p(y_i=0|x_i^T)}{p(y_i=1|x_i^T)}=e^{X_i^T[\gamma_2-\gamma_1]}\\
&\Rightarrow\frac{1}{p(y_i=1|x_i^T)}-1-\frac{1}{e^{X_i^T\gamma_1}}=e^{X_i^T[\gamma_2-\gamma_1]}\\
&\Rightarrow\frac{1}{p(y_i=1|x_i^T)}=1+\frac{1}{e^{X_i^T\gamma_1}}+e^{X_i^T[\gamma_2-\gamma_1]}\\
&\qquad\qquad=\frac{e^{X_i^T\gamma_1}+1+e^{X_i^T\gamma_1}\ e^{X_i^T[\gamma_2-\gamma_1]}}{e^{X_i^T\gamma_1}}
\end{aligned}
$$

Then we get $=\dfrac{e^{X_i^T\gamma_1}+1+e^{X_i^T\gamma_1+X_i^T\gamma_2-X_i^T\gamma_1}}{e^{X_i^T\gamma_1}}$

$$=\frac{e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}+1}{e^{X_i^T\gamma_1}}$$

$$\therefore\ p(y_i=1\,|\,x_i^T)=\frac{e^{X_i^T\gamma_1}}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}\ldots.. \tag{3.10}$$

For $p(y_i=0\,|\,x_i^T)$

Now $\dfrac{p(y_i=1\,|\,x_i^T)}{p(y_i=0\,|\,x_i^T)}=e^{X_i^T\gamma_1}$

$$\Rightarrow \frac{1}{p(y_i=0\,|\,x_i^T)}=\frac{e^{X_i^T\gamma_1}}{p(y_i=1\,|\,x_i^T)}$$

$$=\frac{e^{X_i^T\gamma_1}}{\dfrac{e^{X_i^T\gamma_1}}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}}$$

$$=\frac{e^{X_i^T\gamma_1}(1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2})}{e^{X_i^T\gamma_1}}$$

$$\frac{1}{p(y_i=0\,|\,x_i^T)}=1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}$$

$$\therefore\ p(y_i=0\,|\,x_i^T)=\frac{1}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}$$

Finally for $p(y_i=2\,|\,x_i^T)$

$$p(y_i=0\,|\,x_i^T)=1-p(y_i=1\,|\,x_i^T)-p(y_i=0\,|\,x_i^T)$$

$$=1-\frac{e^{X_i^T\gamma_1}}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}-\frac{1}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}$$

$$=\frac{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}-e^{X_i^T\gamma_1}-1}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}}$$

$$p(y_i=2\,|\,x_i^T)=\frac{e^{X_i^T\gamma_2}}{1+e^{X_i^T\gamma_1}+e^{X_i^T\gamma_2}} \tag{3.11}$$

Hence we have found the logit functions for all outcomes categories when using the group y_i=0 as reference category. As noted earlier on we have constructed a multinomial logistic regression model with reference category.

The additional advantage of this type of modeling is that the model is not over parameterized. But intuitively it indicates that if we know the regression coefficient vectors $\gamma_1,\gamma_2,\ldots..,\gamma_{g-1}$.

We have it is already parameterized due to the other regression coefficient vector.

For convenient notation and reasons that becomes clear later, we define the log it functions for each category outcome variable as

$$\mu_{is}=p(y_i=s\,|\,x_i^T)=\frac{e^{\gamma_0+X_i^T\gamma}}{\sum_{t=1}^{g}e^{\gamma_0+X_i^T\gamma}} \tag{3.12}$$

4. FITTING THIS MODEL

There is no analytical way to fit the model to the observed data, as we do not have a closed expression. To solve this problem we are going to make use of the likelihood function of the model. Recall that the probabilities of the category outcome are Bernoulli distributed. If random variables are independently distributed.

We can state that

$$
\begin{aligned}
L(\theta) &= P(y_1, y_2, \ldots\ldots, y_n \mid \theta) \\
&= P(y_1 / \theta), P(y_2 \mid \theta), \ldots\ldots P(y_n \mid \theta) \\
&= \prod_{i=1}^{n} f_\theta(y_i)
\end{aligned}
$$

Hence we are trying to maximize the conditional probability of the observed data given the parameters. We are trying to find the parameters which best explains the outcomes given the model structure.

In case of Bernoulli variables and our logistic regression principle is

$$L(\beta) = \prod_{i=1}^{n} \prod_{s=1}^{g} \mu_{is}^{y_{is}} \tag{4.1}$$

To maximize this expression we are going to make use of the log-likelihood function which has its maxima at exactly the same parameters values.

$$
\begin{aligned}
l(\beta) = \ln(L(\beta)) &= \ln\left(\prod_{i=1}^{n} \prod_{s=1}^{g} \mu_{is}^{y_{is}} \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{g} \ln\left(\mu_{is}^{y_{is}} \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{g} y_{is} \ln\left(\mu_{is} \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{g} y_{is} \ln\left[\frac{e^{X_i^T \gamma_s}}{\sum_{t=1}^{g} e^{X_i^T \gamma_t}} \right] \\
&= \sum_{i=1}^{n} \sum_{s=1}^{g} y_{is} \left[\ln\left(e^{X_i^T \beta_s} \right) - \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right] \\
&= \sum_{i=1}^{n} \sum_{s=1}^{g} y_{is} \left[X_i^T \gamma_s - \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right]
\end{aligned}
$$

To find the gradient of this log-likelihood function with respect to the beta coefficients we can show the procedure without loss of generality for one partial derivative.

$$\frac{\partial}{\partial \gamma_{kh}} l(\gamma) = \frac{\partial}{\partial \gamma_{kh}} \left[= \sum_{i=1}^{n} \sum_{s=1}^{g} y_{is} \left[X_i^T \gamma_s - \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right] \right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_{kh}} \sum_{s=1}^{g} y_{is} \left[X_i^T \gamma_s - \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right] \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_{kh}} \left(y_{i1} + y_{i2} + \ldots\ldots + y_{ig} \right) \left(X_i^T \gamma_s - \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right) \\
&= \sum_{i=1}^{n} y_{ih} \left(\frac{\partial}{\partial \gamma_{kh}} \left(X_i^T \gamma_s \right) - \frac{\partial}{\partial \gamma_{kh}} \ln\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) \right)
\end{aligned}
$$

$$= \sum_{i=1}^{n} y_{ih} X_{ik} - \frac{1}{\sum_{t=1}^{g} e^{X_i^T \gamma_t}} e^{X_i^T \gamma_h} \frac{\partial}{\partial \gamma_{kh}} \left(e^{X_i^T \gamma_t} \right)$$

$$= \sum_{i=1}^{n} y_{ih} \ X_{ik} - \left[\frac{e^{X_i^T \gamma_h}}{\sum_{t=1}^{g} e^{X_i^T \gamma_t}} \right] X_{ik}$$

$$= \sum_{i=1}^{n} \left(y_{ih} \ X_{ik} - \mu_{ih} X_{ik} \right)$$

$$= \sum_{i=1}^{n} \left(y_{ih} - \mu_{ih} \right) X_{ik}$$

$$\therefore \frac{\partial}{\partial \gamma_{kh}} l(\gamma) = \sum_{i=1}^{n} \left(y_{ih} - \mu_{ih} \right) X_{ik} \qquad (4.2)$$

The Matrix form is

$$X^* = \begin{pmatrix} X & 0 & \cdots & 0 \\ 0 & X & \cdots & 0 \\ 0 & 0 & \cdots & X \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_g \end{pmatrix}$$

Using these definitions we can define the gradient vector as

$$\frac{\partial}{\partial \beta} l(\gamma) = X^T (Y - \mu) \qquad (4.3)$$

Second derivative with respective ks

$$\frac{\partial}{\partial \gamma_{ks}} \left(\frac{\partial}{\partial \gamma_{kh}} \right) = \frac{\partial^2}{\partial \gamma_{ks} \partial \gamma_{kh}}$$

$$= \frac{\partial}{\partial \gamma_{ks}} \left(\sum_{i=1}^{n} X_{ik} \left(y_{ih} - \mu_{ih} \right) \right)$$

$$= \frac{\partial}{\partial \gamma_{ks}} \left(\sum_{i=1}^{n} \left(y_{ih} \ X_{ik} - X_{ik} \ \mu_{ih} \right) \right)$$

$$= \frac{\partial}{\partial \gamma_{ks}} \sum_{i=1}^{n} y_{ih} \ X_{ik} - \frac{\partial}{\partial \gamma_{ks}} \sum_{i=1}^{n} X_{ik} \ \mu_{ih}$$

$$= -\frac{\partial}{\partial \gamma_{ks}} \sum_{i=1}^{n} X_{ik} \ \mu_{ih}$$

$$= -\sum_{i=1}^{n} X_{ik} \ \frac{\partial}{\partial \gamma_{ks}} (\mu_{ih})$$

$$= -\sum_{i=1}^{n} X_{ik} \ \frac{\partial}{\partial \gamma_{ks}} \left(\frac{e^{X_i^T \gamma_h}}{\sum_{t=1}^{g} e^{X_i^T \gamma_t}} \right)$$

$$= -\sum_{i=1}^{n} X_{ik} \ e^{X_i^T \gamma_h}$$

Now
$$= \frac{\partial}{\partial \gamma_{kh}} \left(\sum_{i=1}^{n} X_{ik} \left(y_{ih} - \mu_{ih} \right) \right)$$

$$= \frac{\partial}{\partial \gamma_{kh}} \left(\sum_{i=1}^{n} \left(X_{ik} y_{ih} - X_{ik}\, \mu_{ih} \right) \right)$$

$$= \frac{\partial}{\partial \gamma_{kh}} \sum_{i=1}^{n} X_{ik} y_{ih} - \frac{\partial}{\partial \gamma_{kh}} \sum_{i=1}^{n} X_{ik}\, \mu_{ih}$$

$$= \frac{\partial}{\partial \gamma_{kh}} \sum_{i=1}^{n} X_{ik} y_{ih} - \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_{kh}} \left(X_{ik} \mu_{ih} \right)$$

$$= 0 - \sum_{i=1}^{n} X_{ik} \frac{\partial}{\partial \gamma_{kh}} \left[\frac{e^{X_i^T \gamma_h}}{\sum_{t=1}^{g} e^{X_i^T \gamma_t}} \right]$$

$$= -\sum_{i=1}^{n} X_{ik} \left[\frac{e^{X_i^T \gamma_h} \left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) X_{ik} - \sum_{t=1}^{g} e^{X_i^T \gamma_t} e^{X_i^T \gamma_h} X_{ik}}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right]$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \left[\frac{e^{X_i^T \gamma_h} \left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) - e^{X_i^T \gamma_h} e^{X_i^T \gamma_h}}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right]$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \left[\frac{e^{X_i^T \gamma_h} \left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right) - \left(e^{X_i^T \gamma_h} \right)^2}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right]$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \left[\frac{e^{X_i^T \gamma_h} \sum_{t=1}^{g} e^{X_i^T \gamma_t}}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right] - \left[\frac{\left(e^{X_i^T \gamma_h} \right)^2}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right]$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \left[\frac{e^{X_i^T \gamma_h}}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)} \right] - \left[\frac{\left(e^{X_i^T \gamma_h} \right)^2}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)^2} \right]$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \mu_{ih} - \mu_{ih}^2$$

Where
$$\mu_{ih} = \frac{e^{X_i^T \gamma_h}}{\left(\sum_{t=1}^{g} e^{X_i^T \gamma_t} \right)}$$

$$= -\sum_{i=1}^{n} X_{ik}^2 \mu_{ih} (1 - \mu_{ih}) \qquad (4.4)$$

We know the second order partial derivatives for all cases and we can construct the Hessian conveniently by matrix operations. The Hessian is defined as

$$\frac{\partial^2 l(\gamma)}{\partial\gamma_{ks}\partial\gamma_{kh}} = -\left(X^*\right)^T ZX^* \tag{4.5}$$

$where\ the\ ng \times ng\ matrix\ Z\ is\ given\ by:$

$$Z = \begin{pmatrix} Z^{11} & Z^{12} & ..... & Z^{1g} \\ Z^{21} & Z^{22} & ..... & Z^{2g} \\ \vdots & \vdots & \ddots & \vdots \\ Z^{g1} & Z^{g2} & ..... & Z^{gg} \end{pmatrix}$$

Now, we are going to show that $\frac{\partial^2 l(\gamma)}{\partial\gamma_{ks}\partial\gamma_{kh}}$ is negative semi-definite for any $\gamma \in \mathbb{R}_{p+1}$ we have,

$$w^T l''(\gamma) w = w^T X^T Z\ X\ w$$

$$= -\sum_{i=1}^{n}\left(X_i^T w\right)^2 diag\left(\mu_i(1-\mu_i)\right) \tag{4.6}$$

As $diag\left(\mu_i(1-\mu_i)\right)$ is always positive, we can see that $w^T l''(\gamma) w \le 0 \quad \forall \ w \in \mathbb{R}_{p+1}$ and all $\gamma \in \mathbb{R}_{p+1}$.

Since $l''(\gamma)$ is negative semi-definite, the log-likelihood, l is a concave function of of the parameter γ; several optimization techniques are available for finding the maximizing parameters (see,for example,Mak,1993; Givens and Hoeting,2005).

We use the Newton-Raption, we use $\gamma^{(t)}$, the current estimate of γ, to calculate $\mu^{(t)}$ and $Z^{(t)}$. The new estimate of γ is

$$\gamma^{(t+1)} = \gamma^{(t)} + \left(X^{(T)} Z^{(t)} X\right)^{-1} X^{(T)}(y - \mu^{(t)}). \tag{4.7}$$

This process is required until the estimates stop changing, that is, until $\gamma^{(t+1)}$ is sufficiently close to $\gamma^{(t)}$, then we say that the Newton-Raption method converges.

To better understand what ensures convergence, we must carefully analyze the errors at successive steps.

5. THE SIMULATION RESULTS FOR CONSISTENCY AND NORMALITY OF THE BINARY LOGISTIC REGRESSION

The binary logistic regression model: Consistency of the Maximum Likelihood estimators:

Now the standard Monte Carlo simulation, the fixed sample performance of consistency of the maximum likelihood estimatiors of the logistic regression model. In our simulation study, we consider four explanatory variables x_1, x_2, x_3 and x_4, which are fixed and the binary response variable y, which is treated as a random variable in the logistic model. For fixed values of the intercept parameter γ_0 and four other parameters are $\gamma_1, \gamma_2, \gamma_3$ and γ_4 our aim is to compare the performance of the values of parameters and their standard errors when sample size increases.

For fixed values of $\gamma_0 = 0.8, \gamma_1 = 1.2, \gamma_2 = 1.1, \gamma_3 = 0.35$ and $\gamma_4 = 0.5$ the logistic regression becomes

$$\mu(x) = \frac{e^{0.8+1.2\ x_1+1.1\ x_2+0.35\ x_3+0.5\ x_4}}{1+e^{0.8+1.2\ x_1+1.1\ x_2+0.35\ x_3+0.5\ x_4}} \tag{5.1}$$

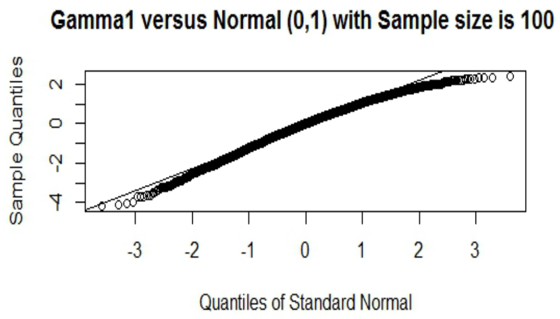
In this simulation, we consider sample sizes of n=100, 200, 300 and 400 and generate 2000 independent sets of random samples for each different sample size.

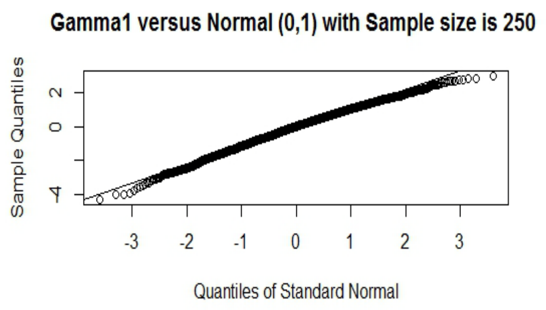
For each set of random sample with a particular sample size, we estimate γ_0 $\gamma_1, \gamma_2, \gamma_3$ and γ_4 and their standard errors based on the logistic regression model. The final estimates and standard errors of γ_0 $\gamma_1, \gamma_2, \gamma_3$ and γ_4 are the average of 2000 estimates of γ_0 $\gamma_1, \gamma_2, \gamma_3$ and γ_4 for that particular sample size. The following table gives the results of the simulation study for different sample size.

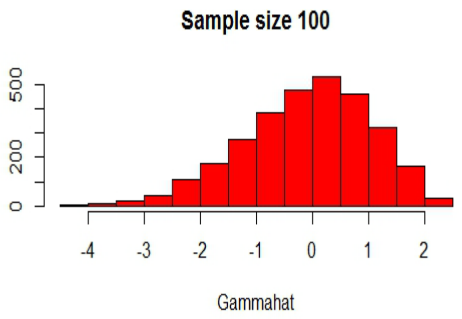
Table-1: Estimated parameter values and their standard errors using the logistic regression model for different sample sizes of 100, 200, 300, 400.

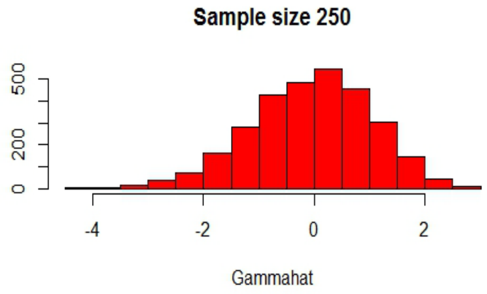
Parameters	n=100		n=200		n=300		n=400	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\gamma}_0$	1.3209	0.0120	0.7670	0.0141	0.6230	0.0075	0.6180	0.0061
$\hat{\gamma}_1$	1.3874	0.0529	1.0735	0.0357	1.0279	0.0079	1.0188	0.0064
$\hat{\gamma}_2$	1.0495	0.0238	1.4631	0.0241	0.9193	0.0076	0.9242	0.0066
$\hat{\gamma}_3$	0.3869	0.0133	0.3728	0.0213	0.3570	0.0076	0.3514	0.0063
$\hat{\gamma}_4$	0.0986	0.0170	0.0321	0.0152	0.0610	0.0075	0.0410	0.0061

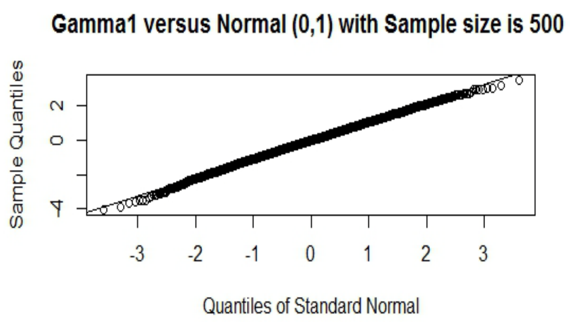
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_1$ (Simulation size = 3000)

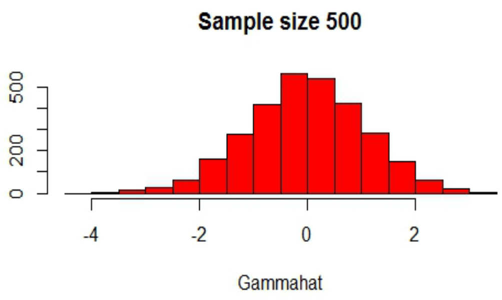


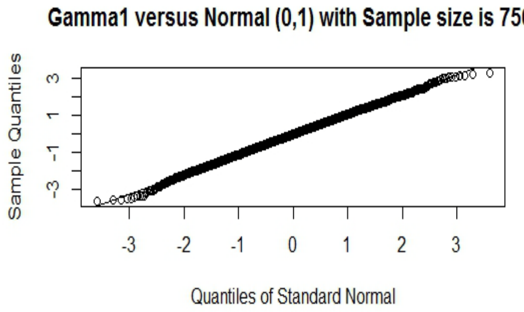


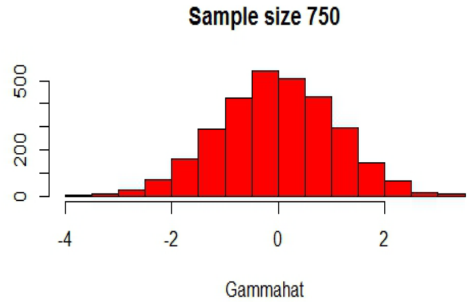




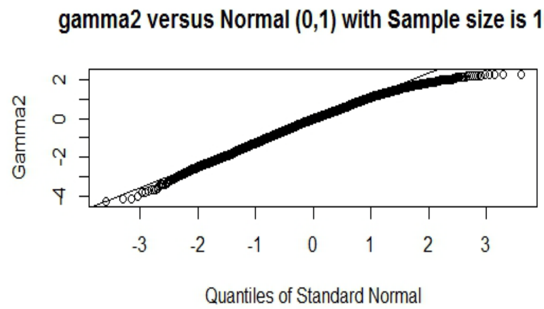


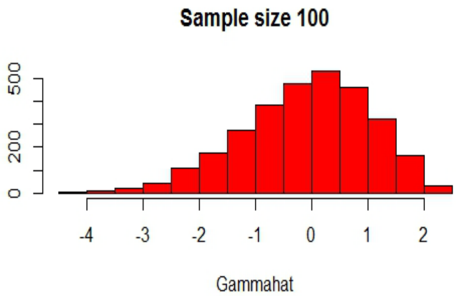


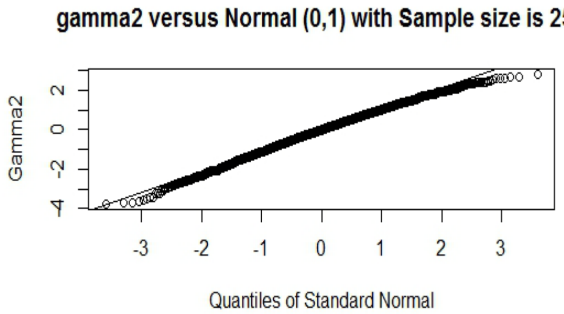


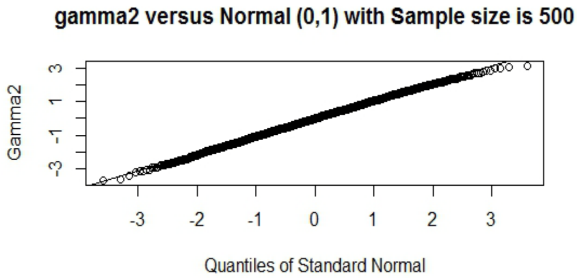


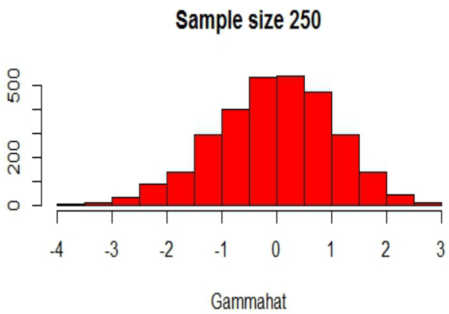
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_2$ (Simulation size =3000)

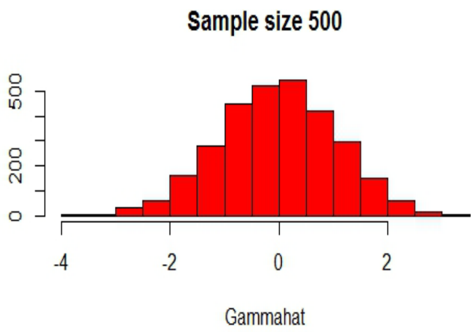


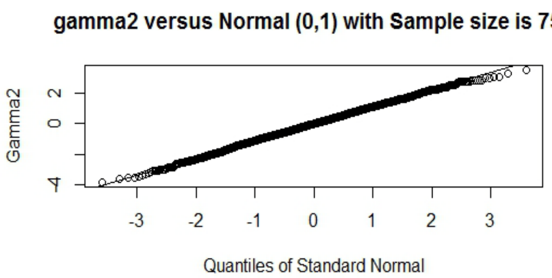


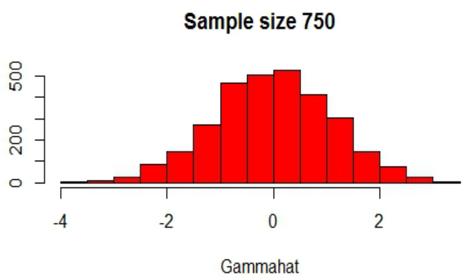




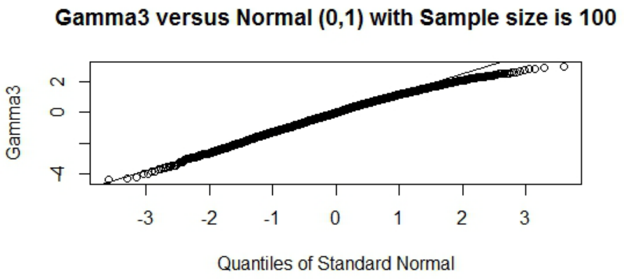


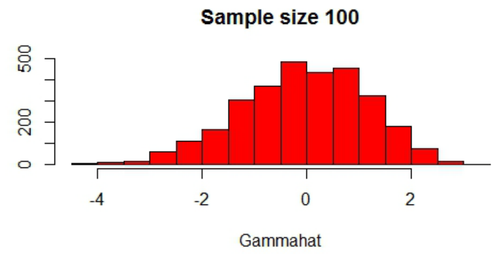


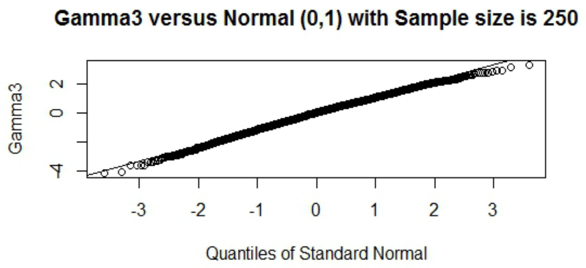


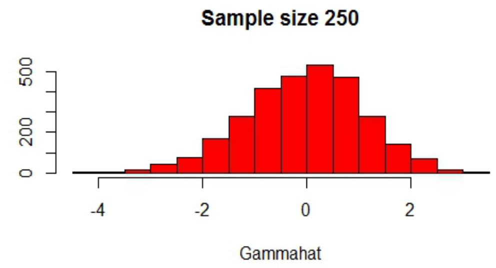


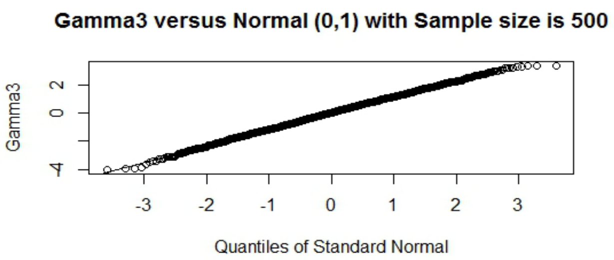
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_3$ (Simulation size =3000)

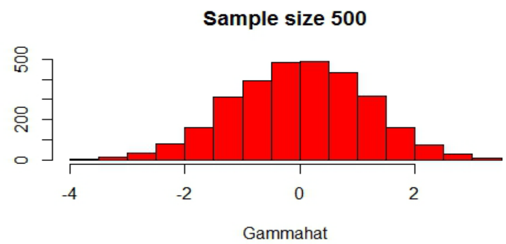


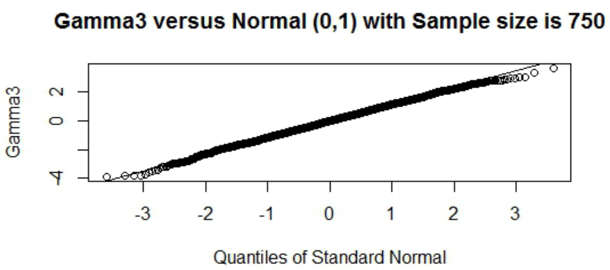


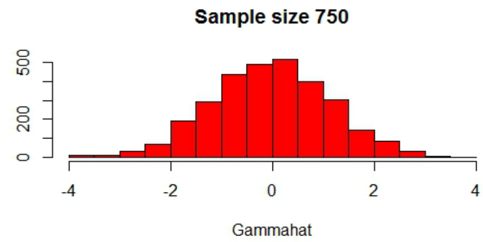




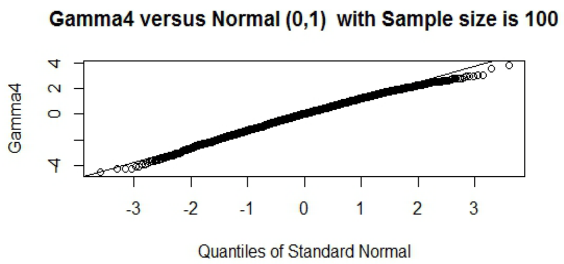


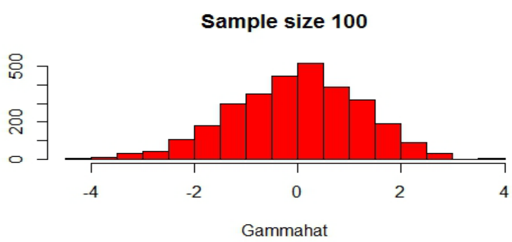


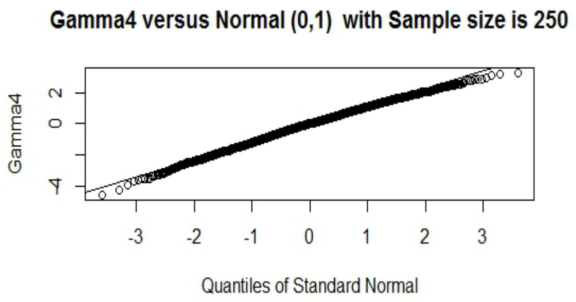


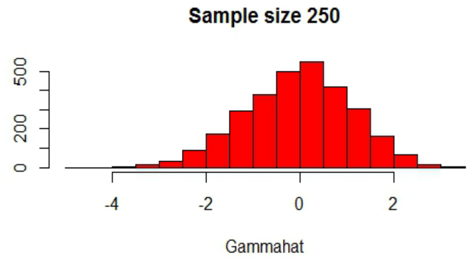


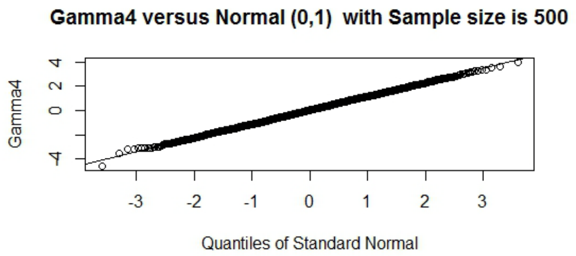
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_4$ (Simulation size =3000)

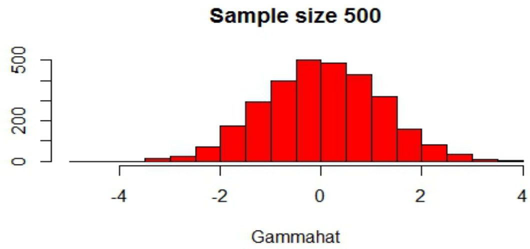


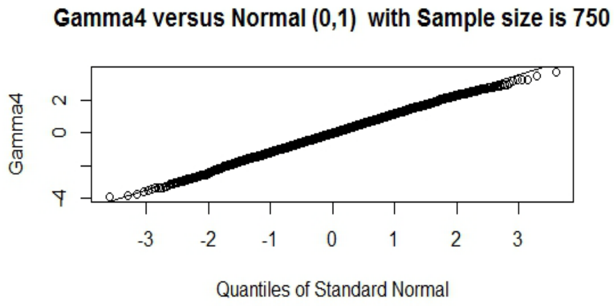


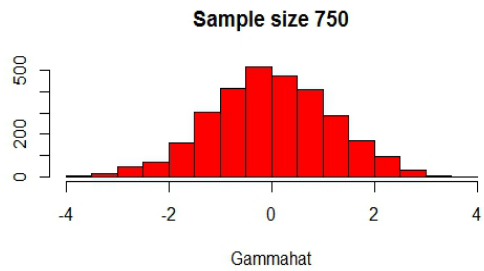












6. THE MULTINOMIAL LOGISTIC REGRESSION MODEL: CONSISTENCY OF THE ML ESTIMATORS

Here, we have to demonstrate the consistency of the maximum likelihood estimators for the multinomial logistic regression modal via standard Monte Carlo simulation. Now, we consider the outcome variable Y is random and has three categories, this is, Y takes values coded as 1,2 and 3. Here we assume that there are four explanatory variables w_1, w_2, w_3 *and* w_4 in the model, where each of them is a vector and take two possible values coded as 0 or 1. If we treat the last category of the outcome variable as the baseline, then the multinomial logistic regression model can be written as

$$\ln\left[\frac{P(Y=1|w_1,w_2,w_3,w_4)}{P(Y=3|w_1,w_2,w_3,w_4)}\right]=\gamma_{01}+\gamma_{11}w_1+\gamma_{12}w_2+\gamma_{13}w_3+\gamma_{14}w_4 \tag{6.1}$$

$$\ln\left[\frac{P(Y=2|w_1,w_2,w_3,w_4)}{P(Y=3|w_1,w_2,w_3,w_4)}\right]=\gamma_{02}+\gamma_{21}w_1+\gamma_{22}w_2+\gamma_{12}w_3+\gamma_{24}w_4 \tag{6.2}$$

Under these models, the response probabilities are

$$P(Y=1|w_1,w_2,w_3,w_4)=\frac{e^{\gamma_{01}+\gamma_{11}w_1+\gamma_{12}w_2+\gamma_{13}w_3+\gamma_{14}w_4}}{1+e^{\gamma_{01}+\gamma_{11}w_1+\gamma_{12}w_2+\gamma_{13}w_3+\gamma_{14}w_4}+e^{\gamma_{02}+\gamma_{21}w_1+\gamma_{22}w_2+\gamma_{12}w_3+\gamma_{24}w_4}} \tag{6.3}$$

$$P(Y=2|w_1,w_2,w_3,w_4)=\frac{e^{\gamma_{02}+\gamma_{21}w_1+\gamma_{22}w_2+\gamma_{12}w_3+\gamma_{24}w_4}}{1+e^{\gamma_{01}+\gamma_{11}w_1+\gamma_{12}w_2+\gamma_{13}w_3+\gamma_{14}w_4}+e^{\gamma_{02}+\gamma_{21}w_1+\gamma_{22}w_2+\gamma_{12}w_3+\gamma_{24}w_4}} \tag{6.4}$$

$$P(Y=3|w_1,w_2,w_3,w_4)=\frac{1}{1+e^{\gamma_{01}+\gamma_{11}w_1+\gamma_{12}w_2+\gamma_{13}w_3+\gamma_{14}w_4}+e^{\gamma_{02}+\gamma_{21}w_1+\gamma_{22}w_2+\gamma_{12}w_3+\gamma_{24}w_4}} \tag{6.5}$$

From the above models, we estimate the unknown parameters $\gamma_{01}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14}$, $\gamma_{02}, \gamma_{21}, \gamma_{22}, \gamma_{12}$ and γ_{24} . The purpose is to show that if the numbers of observations $(y_i, w_{1i}, w_{2i}, w_{3i}, w_{4i})$, $i = 1, 2, \ldots, n$ increases, then the estimates of the parameters converge to their true values. Now, we simulate the values of the outcome and explanatory variables.

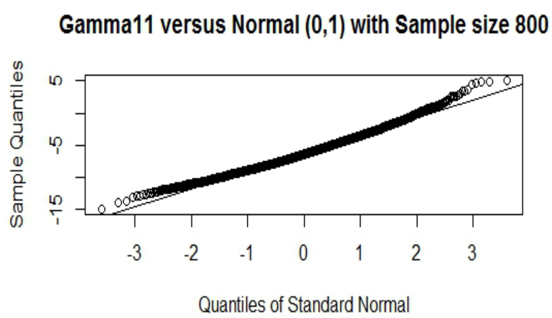
As the explanatory variables are fixed, the variable w_1, w_2, w_3 and w_4 are created based on the binomial distribution for arbitrary number of sample size. Once the variable w_1, w_2, w_3 and w_4 are in hand, we have a tendency to calculate probabilities for the outcome variable based on the above equations (6.3),(6.4) and (6.5). These probabilities are used to simulate the data for Y from the multinomial distribution as Y exceeds more than two categories (actually, in this case it would be trinomial since Y has only three categories). For standard Monte Carlo simulation, we consider sample sizes of n = 200, 500 and 1000.

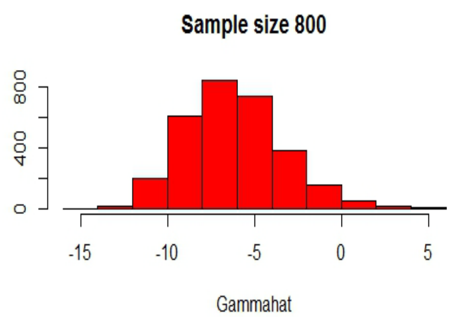
For the arbitrary fixed values of $\gamma_{01} = 0.3$, $\gamma_{11} = 0.7$, $\gamma_{12} = 1.5$, $\gamma_{13} = -0.4$, $\gamma_{14} = 1.3$, $\gamma_{20} = 1.1$, $\gamma_{21} = 1.6, \gamma_{22} = 0.8, \gamma_{23} = -0.5$ *and* $\gamma_{24} = 0.2$, we generate 3000 independent sets of random samples for each different sample sizes. Then we estimate $\gamma_{01}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14}$, $\gamma_{02}, \gamma_{21}, \gamma_{22}, \gamma_{12}$ and γ_{24} based on the average of 3000 estimates of $\gamma_{01}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14}$, $\gamma_{02}, \gamma_{21}, \gamma_{22}, \gamma_{12}$ and γ_{24}, which are estimated from the simultaneously fitted multinomial logistic regression model and their standard error for each estimated parameter. The results of simulation study are provided in the following table.

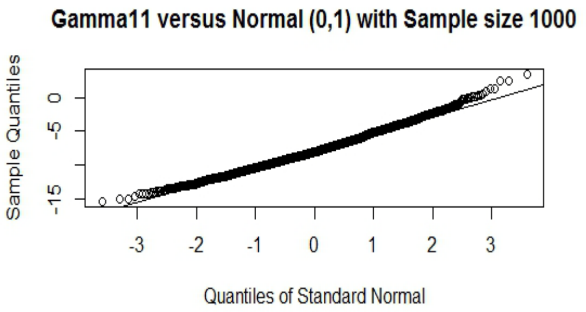
Table-2: Estimated parameter values and their standard errors using the multinomial logistic regression model for different sample sizes of 800,1000 and 1200.

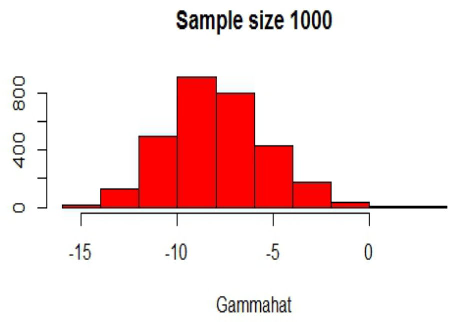
Estimated Parameter	n=800		n=1000		n=1200	
	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\gamma}_{01}$	0.8962	0.0090	0.9692	0.0073	0.8135	0.0070
$\hat{\gamma}_{11}$	-0.0558	0.0084	-0.0485	0.0070	0.0912	0.0061
$\hat{\gamma}_{12}$	0.1019	0.0061	0.0940	0.0054	0.0311	0.0051
$\hat{\gamma}_{13}$	0.0415	0.0062	-0.1733	0.0054	-0.0422	0.0049
$\hat{\gamma}_{14}$	0.0705	0.0061	0.0081	0.0053	-0.0053	0.0049
$\hat{\gamma}_{20}$	1.5327	0.0088	1.6113	0.0071	1.4723	0.0068
$\hat{\gamma}_{21}$	-0.0326	0.0081	-0.0831	0.0068	0.0080	0.0059
$\hat{\gamma}_{22}$	0.0146	0.0060	0.0399	0.0054	-0.0451	0.0049
$\hat{\gamma}_{23}$	-0.0492	0.0061	-0.0904	0.0052	0.0123	0.0048
$\hat{\gamma}_{24}$	0.1027	0.0060	0.0294	0.0052	0.0240	0.0048

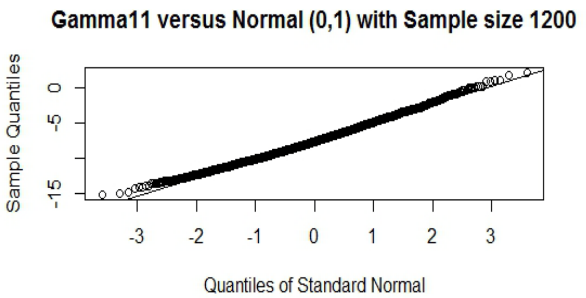
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{11}$ (Simulation size =3000)

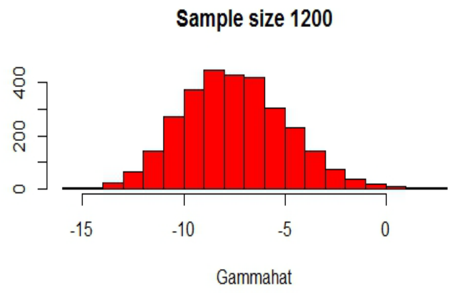




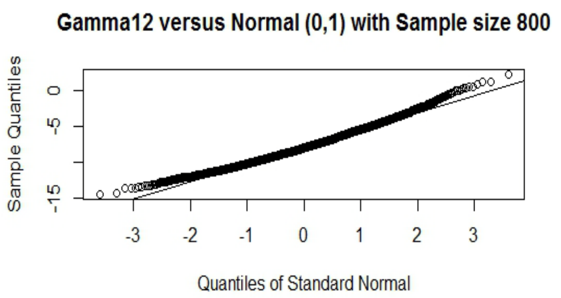


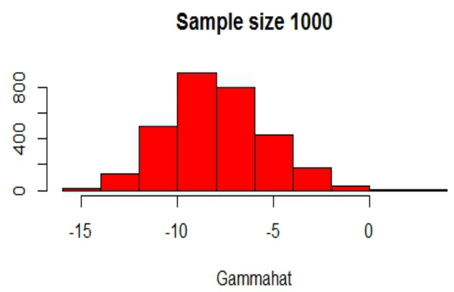


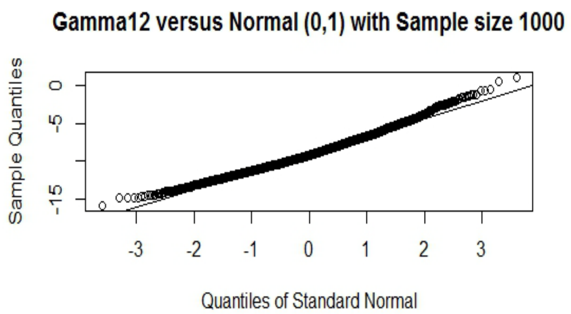


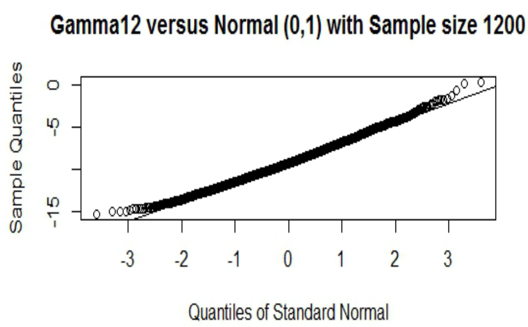


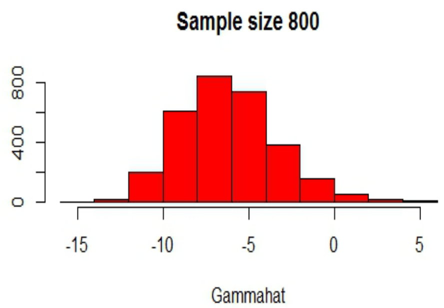
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{12}$ (Simulation size =3000)

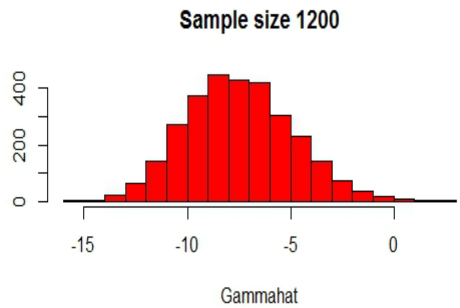




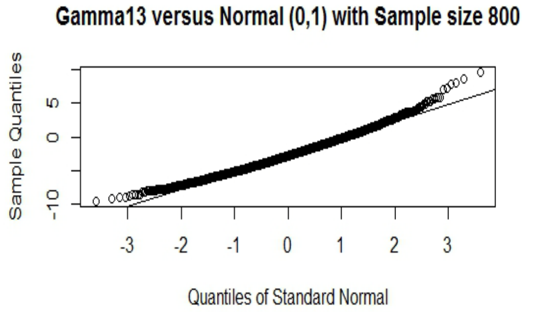


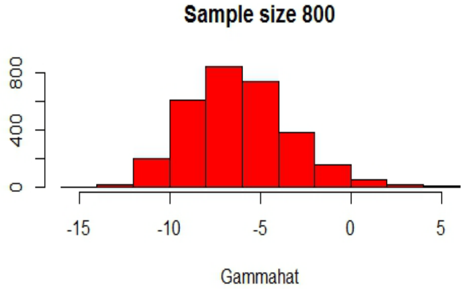


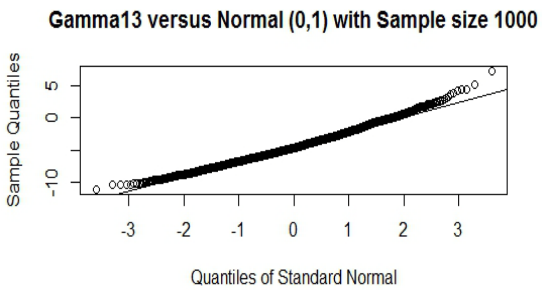


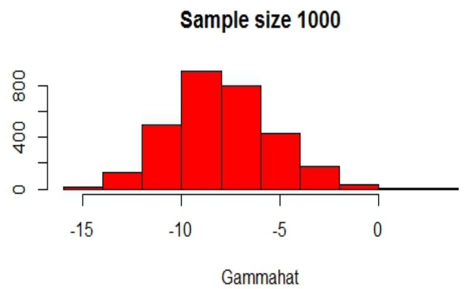


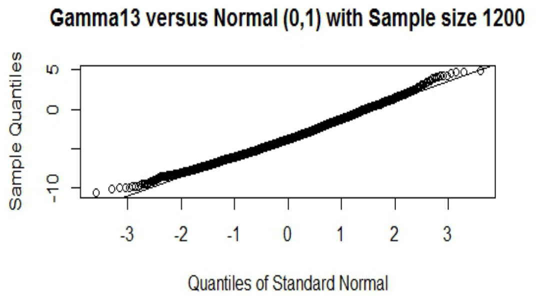
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{13}$ (Simulation size =3000)

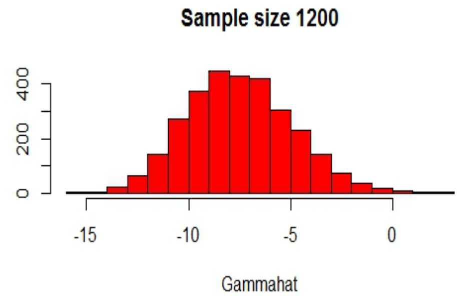




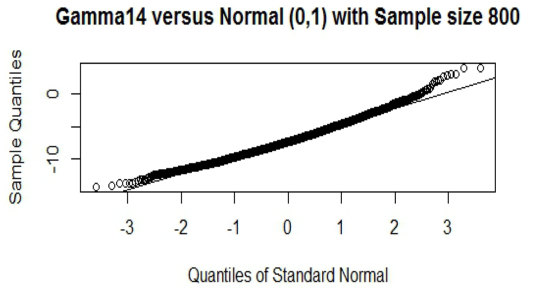


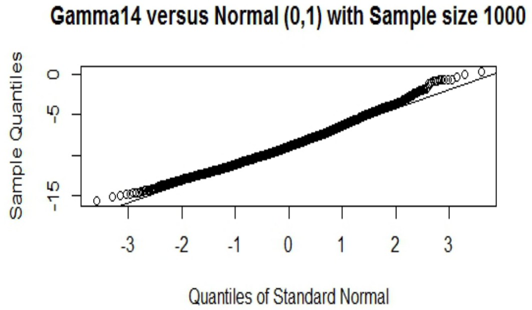


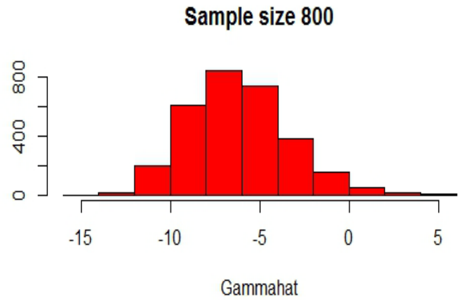


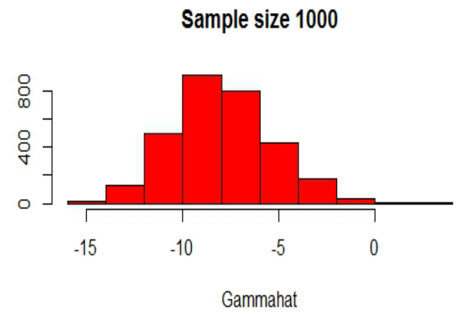


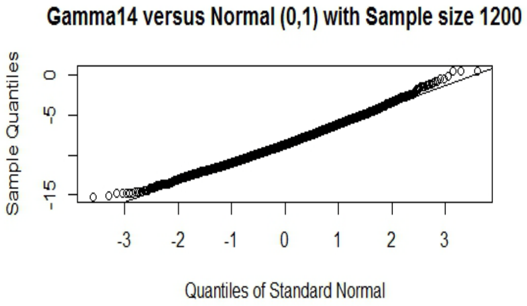
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{14}$ (Simulation size =3000)

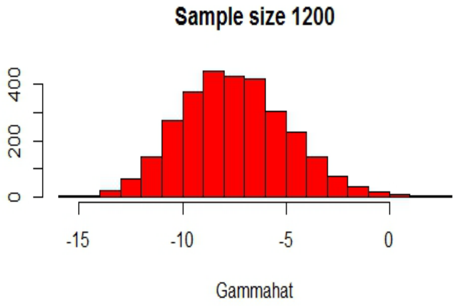




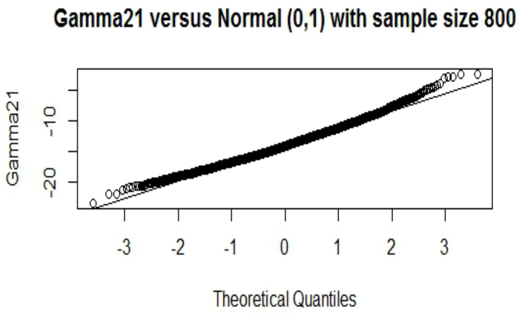


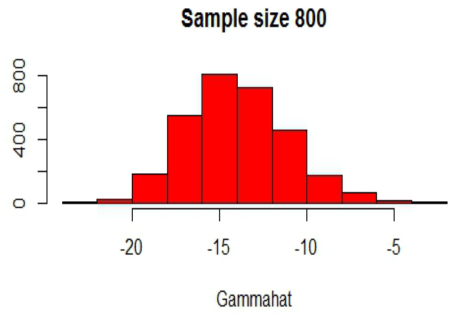


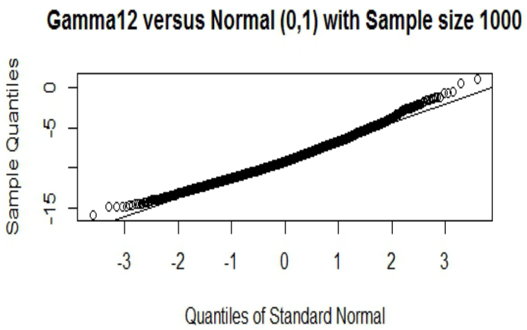


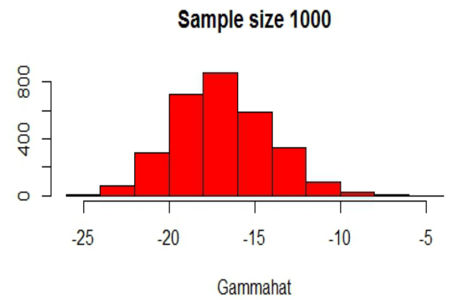


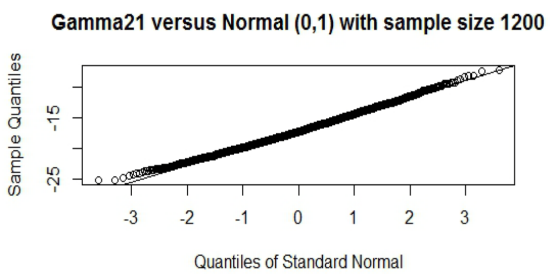
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{21}$ (Simulation size =3000)

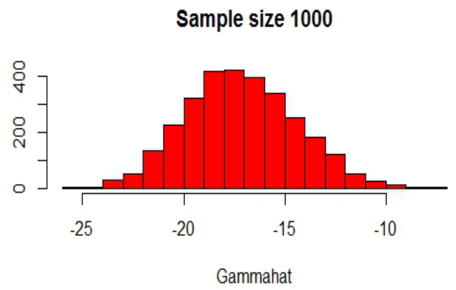




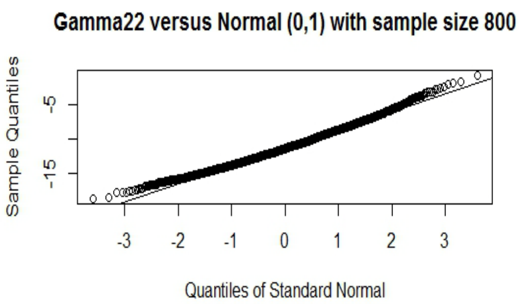


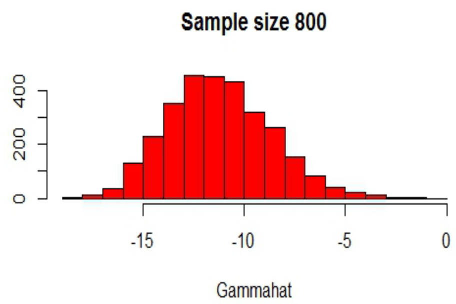


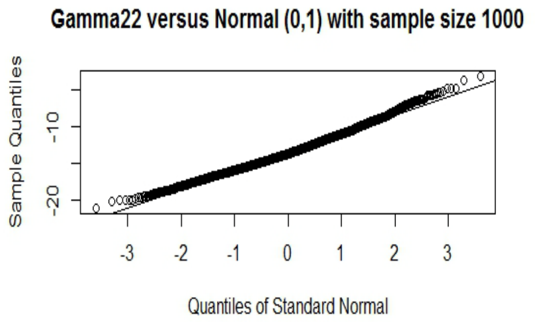


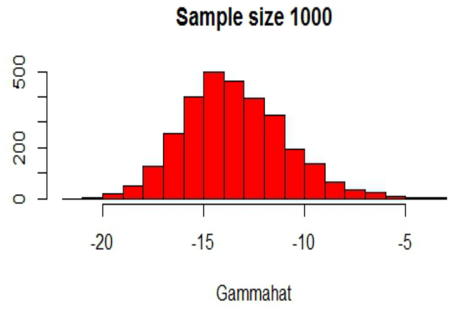


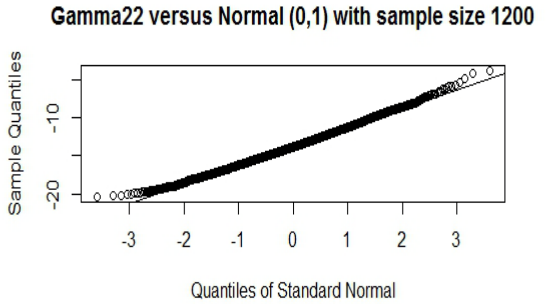
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{22}$ (Simulation size =3000)

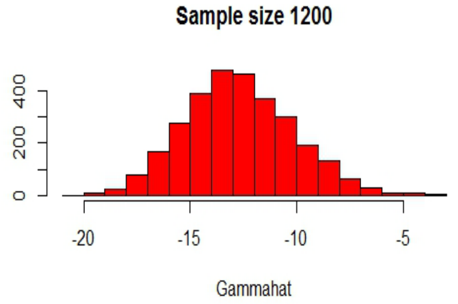




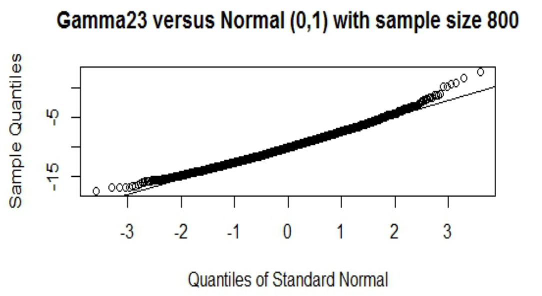


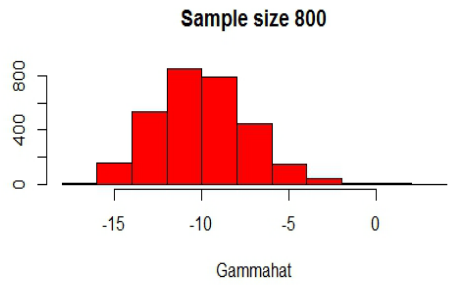


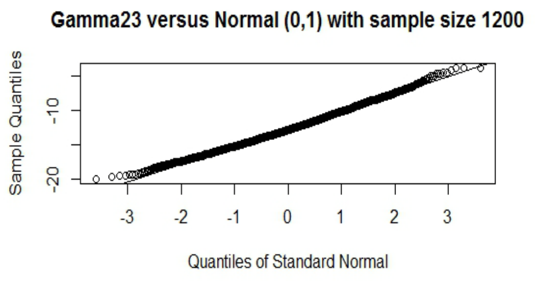


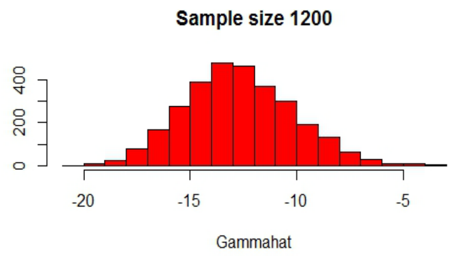


Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{23}$ (Simulation size =3000)

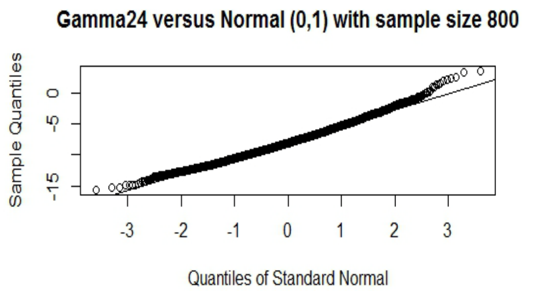


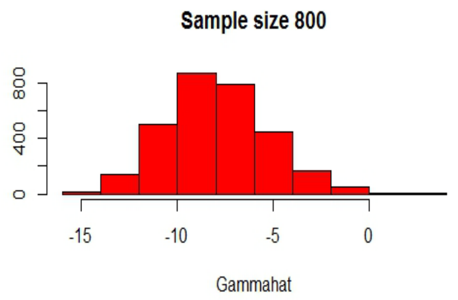


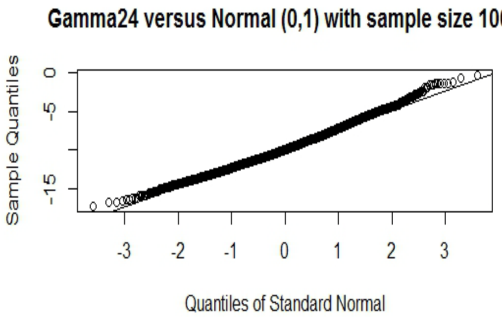


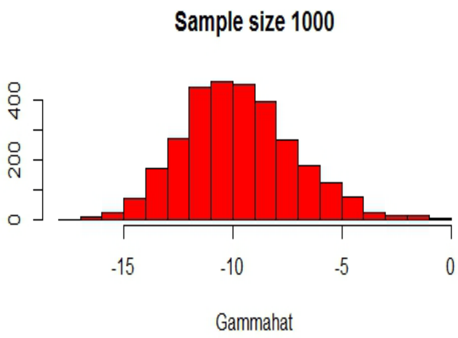


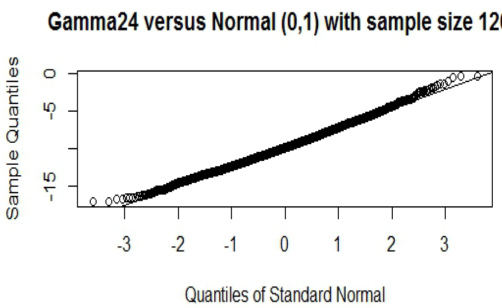
Monte Carlo Simulation of finite sample behavior for normality of the parameter $\hat{\gamma}_{24}$ (Simulation size =3000)

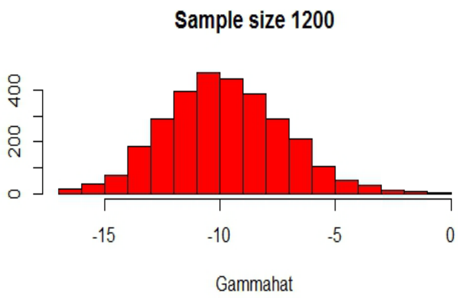












7. NORMALITY OF THE ML ESTIMATORS

Under some assumptions that permits among several analytical properties, the utilization of the delta method, the central limit theorem holds. We have a tendency to conducted a simulation study via the software system package R. We have to show, however the properties of an estimator are affected by changing conditions such as its sample size and therefore the value of the underlying parameters. Using it in practice, we have a tendency to illustrate the sensitivity of the QQ-plots, we show that;

$$\sqrt{N}\left(\gamma_{mle}-\gamma\right)\rightarrow N\left(0,\frac{1}{I\left(\gamma_{mle}\right)}\right)$$

Where

$$I(\gamma)=-E_{\gamma}\begin{pmatrix} \frac{\partial^2 \log l}{\partial\gamma_0^2} & \frac{\partial \log l}{\partial\gamma_0\partial\gamma_1} & \frac{\partial \log l}{\partial\gamma_0\partial\gamma_2} & \frac{\partial \log l}{\partial\gamma_0\partial\gamma_3} & \frac{\partial \log l}{\partial\gamma_0\partial\gamma_4} \\ \frac{\partial \log l}{\partial\gamma_1\partial\gamma_0} & \frac{\partial^2 \log l}{\partial\gamma_1^2} & \frac{\partial \log l}{\partial\gamma_1\partial\gamma_2} & \frac{\partial \log l}{\partial\gamma_1\partial\gamma_3} & \frac{\partial \log l}{\partial\gamma_1\partial\gamma_4} \\ \frac{\partial \log l}{\partial\gamma_2\partial\gamma_0} & \frac{\partial \log l}{\partial\gamma_2\partial\gamma_1} & \frac{\partial^2 \log l}{\partial\gamma_2^2} & \frac{\partial \log l}{\partial\gamma_2\partial\gamma_3} & \frac{\partial \log l}{\partial\gamma_3\partial\gamma_4} \\ \frac{\partial \log l}{\partial\gamma_3\partial\gamma_0} & \frac{\partial \log l}{\partial\gamma_3\partial\gamma_1} & \frac{\partial \log l}{\partial\gamma_3\partial\gamma_2} & \frac{\partial^2 \log l}{\partial\gamma_3^2} & \frac{\partial \log l}{\partial\gamma_3\partial\gamma_4} \\ \frac{\partial \log l}{\partial\gamma_4\partial\gamma_0} & \frac{\partial \log l}{\partial\gamma_4\partial\gamma_1} & \frac{\partial \log l}{\partial\gamma_4\partial\gamma_2} & \frac{\partial \log l}{\partial\gamma_4\partial\gamma_3} & \frac{\partial^2 \log l}{\partial\gamma_4^2} \end{pmatrix}$$

For different sample sizes n=800,n=1000 and n=1200 we calculate the equation 11 and repeat it 3000 times. The results are presented in the Figures 1,2,3,4 and 5 through the Quantile –Quantile normal plots for $\hat{\gamma}$.

A Quantile-Quantile normal graph, plots the quantiles of the data set against the theoretical quantiles of the standard normal distribution. If the data set seems to be a sample from a normal population, then the points can fall roughly along the line. The computation results indicates that the distribution of parameters approximates normal distribution as sample size n is increases.

8. CONCLUSION

In this study shows that the asymptotic properties of the maximum likelihood estimates of the logistic and multinomial logistic regression can be obtained by some transformation of the regularity conditions of the linear regression model. The results for both binary and multinomial logistic regression indicate that the simulation study performs well in showing the consistency of the maximum likelihood estimators for parameters of the models. We also believe that the computation results illustrate the distribution of parameters approximates normal distribution as sample size increases. However, it takes relatively huge set of data for such results in the case of multinomial logistic regression models. Consequently it is important to verify the model assumptions before applying these results in a real life situation.

REFERENCES

1. Agresti,A.(2002).Categorical Data Analysis. John Wiley & Sons, Inc.
2. Anthony. N, On the Estimation and Properties of Logistic Regression Parameters. Jomo Kenyatta University of Agriculture & Technology, PP 57-68, 2014.
3. Amemiya,T, Advanced Econometric. Havard University Press, Cambridge, 1985.
4. B. Munsiwamy and Shibru Temesgen Wakweya, Asymptotic properties of estimates for the parameters in the Logistic Regression Model.Vol.11, No.1, 2011.
5. C.Gourienx and A.Monfort. Asymptotic properties of the maximum likelihood estimation in dichotomous logit models. Journal of Econometric 17, 83-97, 1981.
6. Cramer,J.S.(2003). Logit Models from Economics and Other Fields. Cambridge University Press.
7. Cornfild,J.(1962).Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure.Federation Proc., 21,58-61.
8. Cox,D.R.,(1970). Analysis of Binary Data.London:Chapman & Hall.
9. Cox,D.R., and Snell,E.J.(1989).Analysis of Binary Data.London: Chapman & Hall.
10. David Collett, Modeling Binary Data. Chapman & Hall/CRC,New York,USA, second edition, 2002.
11. D.W.Hosmer and S.Lemeshow. Applied Logistic Regerssion, NewYork, Wiley, 1989.
12. Efron and Tibsirani,R.J.(1993).A introduction to the Bootstrap. Journal of the American Statistical Association, 89,463-478.
13. Givens, G.H. and Hoeting, J.A.Computational Statistics, John Wiley & Sons, Inc., 2005.
14. Hosmer, D. and Lemeshow, S. Applied Logistic Regression. Second Edition, JohnWiley & Sons Inc., New York, 2000.
15. J.Wilde.A note on gmm estimator of probit models with endogenous regression, 2008
16. L.McFadden, Conditional Logit analysis of qualitative choice behavior.1974.
17. L.Nordberg. asymptotic normality of maximum likelihood estimators based on Independent unequal distributed observations in exponential family models. 1980.
18. P.McCullagh and J.A.Nelder, Generalized linear models, 1989.

Source of support: Nil, Conflict of interest: None Declared.