International Journal of Engineering & Technology, 7 (4) (2018) 4911-4917

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.24842
Research paper

Approximation techniques for maximizing likelihood functions of generalized linear mixed models for binary response data

Julius Remi Dare [1] *, Dawud Adebayo Agunbiade [2], Oludolapo Kehinde Famurewa [1], Olumide Sunday Adesina [2], Dayo Funmi Adedotun [2], Olatunde Iyaniwura [3]

[1] Department of Mathematical Sciences, Kings University, Odeomu osogbo, Nigeria
[2] Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-iwoye, Nigeria
**Corresponding author E-mail: jr.dare@kingsuniversity.edu.ng*

Abstract
Evaluating Maximum likelihood estimates in Generalized Linear Mixed Models (GLMMs) has been a serious challenge due to some integral complexities encountered in maximizing its likelihood functions. It is computationally difficult to establish analytical solutions for the integrals. In view of this, approximation techniques would be needed. In this paper, various approximation techniques were examined including Laplace approximation (LA), Penalized Quasi likelihood (PQL) and Adaptive Gauss-Hermite Quadrature (AGQ) techniques. The performances of these methods were evaluated through both simulated and real-life data in medicine. The simulation results showed that the Adaptive Gauss-Hermit Quadrature approach produced better estimates when compared with PQL and LA estimation techniques based on some model selection criteria.

Keywords: *Generalized Linear Mixed Models; Adaptive Gauss-Hermit Quadrature; Likelihood Function; Binary Response; Medicine.*

1. Introduction

Collection of data without any reference to standard assumption of independence across observations is currently posing a serious challenge in statistical modeling. This kind of data may include data with repeated observations on a particular subject over time or observations which are clustered in one way on the other. Example is classes in the same school or households in the same neighborhood). Due to computational advancement in statistical modeling, analytical techniques have been modified to handle such data complex structures. Binary data is a statistical data type described by binary variables, which can take only two possible values. Binary response is a data type that assumes 0 and 1 qualitative values. To make predictions on such data; correlation within observations on the same subject area must be taken into consideration for better regression parameters.

Lee *et al.* [1] extended the linear mixed model to handle outcomes with non-normal distributions. Breslow [2] made a submission that in Generalized Linear Mixed Models (GLMMs), the model should be specified such that subject-specific random effects, which are often normally independently distributed, should be considered. By so doing, Breslow [2] mentioned that the second order structure or correlation between subjects in the same cluster can be described and accounted for. When the outcome is binary, GLMMs regression coefficient is estimated on the random effect, and as such, has a subject-specific interpretation. Bolker et al.[3] identified that the rate at which data are collected in which the standard assumption of independence between observations is not met. By implication, data may comprise of multiple observations on a subject over time or subjects which are clustered in some way. Among the authors that identified the flexibility of GLMMs in handling non-normal data is Casals et. al [4] and McCulloch & Searle [5] .These authors encounter difficulty mainly in the area of parameter estimation because there was no analytic solution for the maximization of the marginal likelihood. However, in recent times, there are some proposed methods for estimating parameters of GLMMs and in-built packages in R software, by R Core team [6].

Maximum likelihood Estimation (MLE) is commonly used in GLMMs for parameter estimation. However, this method of estimation has to do with high-dimensional integrals that pose some analytical complexity, particularly when the response variable is not normally distributed. To handle this problem, numerical approximation methods are needed, of which their estimations can either be classical or Bayesian. Gamerman [7], Christensen & Waagepetersen [8], Zhao *et. al* [9], Fong *et. al* [10], and Adesina *et. al* [11] estimated parameters of GLMMs using both Bayesian and Frequentist approach. Efron [12] , Adesina *et. al* [13] and Brooks et. al [12], estimated parameters of GLMMs with Bayesian technique.

This study provides analytical methods in estimating parameters of GLMMs, and to investigate the suitability of Adaptive Gauss-Hermite quadrature estimation technique in fitting GLMMs with binary responses over other exiting GLMM estimation techniques. Over the years, the application of GLMMs in medical analysis has increased to solve the problem of dependence across data when modeling binary or count data. In this study, the methods proposed are evaluated with R software using both real from medical domain and simulated data. The remaining part of this paper is sectioned as follows; various methods of estimation of Linear models, Generalized Linear models and GLMMs was expounded in section 2, while in section 3, the results obtained were presented, and section 4 contains the summary and discussion respectively.

2. Methods

2.1 Classical Linear Models

Classical linear models are ordinary models in linear regression. They are often accepted by many due to their effectiveness in handling regression problems. The one dimensional version is defined as

$$y = \beta_o + \beta_1 x + e \quad (1)$$

In this setting, x is often called the predictor and y the response variable where
β_o and β_1 are the regression parameters and e is the error term. For one predictor variable X, this notation is often sufficient.
When more predictors are introduced a matrix based setup is more appropriate as it will ease computational and notational effort. Let p be number of parameters β_1,…….. β_p and n the number of observations, then, we can rewrite the expanded version of equation (1) for one observation as

$$y_i = \beta_1 + \beta_2 x_{i2} +, \dots, + \beta_p x_{ip} + e_i \quad (2)$$

In matrix notation, we can rewrite the vectored version as

$$y = X\beta + e \quad (3)$$

If we assume there is an intercept in X such that $x_{i1} = 1$, the corresponding
Matrix notation for the elements Y, X, β and e is

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_p \\ 1 & x_{22} & x_{23} & & x_{1p} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \quad (4)$$

Y is the response vector, where β the predictor vector and X is in computational application often referred to as the model matrix.
.
Considering the model defined in (4), we can further investigate the 'noise' parameter e_i. We often make an assumption that the parameters e_i are i.i.d. with $e_i \sim N(0, \sigma)$. Therefore in equation (3), it still holds that E(β_i) = β_i . Moreover this leads to $y_i \sim N(\mu, \sigma^2)$ and Cov$(y_{i,} y_j) = 0$ for $i \neq j$. This can easily be extended to the matrix notation of the model by assuming that $E(e_i) = 0$, Var(e_i) = σ^2 .
Y$\sim N(\mu, \varepsilon)$ to be multivariate-normal (ε is in this case the diagonal matrix with value σ^2). In this case, Maximum Likelihood Estimation (MLE) is used. For the LM, the MLE coincides with that of the minimized least squares under the above assumption.

2.2 Generalized Linear Models

Linear Models provide excellent solutions to many applications across all of science; they are however bound to two constraints as two assumptions are made:
i. The components of Y $\sim(\mu, \sigma^2)$ and the variance σ^2 is constant
ii. $E(Y) = \mu$ and $\mu = X\beta$

In nature where many things are normally distributed due to the central limit theorem, LMs are therefore often sufficient. In actuarial application however, Y is often not normally distributed with non-trivial connection between the predictors and the response. Generalized linear models (GLMs) provide a solution for this. GLMs happen to be an extension of classical linear model which allows the mean of any given population to depend on a linear predictor using a nonlinear link function. GLM and application is sufficiently discussed by Jong and Heller [15].

Components of Generalized linear models are as follows;

i. Y as a response variable has a distribution in the exponential family with a probability function of the form:

$$f\left(y_{i,}; \theta_i\right) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c\left(y_{i,}\phi\right)\right) \quad (5)$$

for suitable choice of $a_i \neq 0$, b and c. Note: φ is the scale parameter while natural parameter is represented by θ .For the exponential family. Here, $\phi > 0$ is the dispersion parameter; o $\in \theta$ is the parameter of the given distribution and $\theta \subset R$ is an open set containing θ.

$$E(y_i) = \mu_i = b^{'}(\theta_i), \quad \text{var}(Y_i) = b^{"}(\theta_i) a_i(\phi) \quad (6)$$

Distribution that are members in the exponential family are flexible and can fit continuous, count and binary data adequately.

ii. Considering the random variable $Y_1, \dots, Y_n$, the linear part is defined as $\eta_i = X_i\beta$,
$i = 1, \dots n$ with some vector parameters $\beta = (\beta_1, \dots, \beta_p$ and covariate $X = \left(x_{i1, \dots,} x_{ip}\right)$ related to observations Y_i.

iii. A nonlinear link function g illustrate how the expected response $\mu_i = E(Y_i)$ is a function of the linear predictor $g(\mu_i) = \eta_i$, $i = 1, \dots, n$

The linear predictor η connects to observable $E(Y)$ through a link. In linear regression models, this is always the identity link (g(μ_i) = μ_i = η_i) When using Gamma or Poisson distributions. This may be less useful as these distributions only have values on the positive line R^+. For example, we may require that the mean needs to be strictly positive (as in claim counts and severity). Hence, links that only take positive values may be more appropriate.

2.3 The Generalized Mixed Models

Generalized linear mixed model is considered as being extension of generalized linear models which include both the fixed and random effect is express as follows:

$$y = X_i\beta + Z_i\gamma + e \quad (7)$$

Methods for estimating parameters in statistical analyses play a vital role in determining the fit of any model. Both fixed-effect and random-effect parameters are estimate by maximum likelihood procedures. However, this ML procedure tends to break down more complex cases in both LMMs and GLMMs. To solve this problem, one has to integrate likelihoods over all random effects. To illustrate, a direct computation of the likelihood;

$$L = \int f_y/u\left(\frac{y}{u}\right) fu(u)du \quad (8)$$

Or individually,

$$L = \int \prod_{i-1}^{n} \frac{fy_i}{u\left(\frac{y_i}{u}\right)} fu(u)du \quad (9)$$

For functions of the exponential family, this leads to a likelihood equation of the form

$$L = \int \prod_{i}^{n} exp\left[\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i;\phi)\right] \times \frac{1}{\sigma_u\sqrt{2\pi}} exp\left[\frac{-(x-\mu)^2}{2\sigma_u^2}\right] du \quad (10)$$

with corresponding log-likelihood

$$L = log\left[\int \prod_{i}^{n} exp\left[\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i;\phi)\right]\right] \times \frac{1}{\sigma_u\sqrt{2\pi}} exp\left[\frac{-(x-\mu)^2}{2\sigma_u^2}\right] du \quad (11)$$

For this equation, no analytical solution can be given. Therefore, numerical approximation must be used to estimate the maximum likelihood.

2.3.1 Estimation Techniques for GLMMs

Penalized Quasi-Likelihood Estimation Method:

Variance components are estimated with quasj-likelihood methods when there is to pre-knowledge of the distribution. Approximation for vector response data y_i is given by

$$y \approx \mu_i + \varepsilon_i = E(y_i/b) + \epsilon_i \quad (12)$$

$$= h(x_i'\beta + Z_i'b) + \epsilon_i \quad (13)$$

Where β =fixed vector parameter

$$= x_i'\beta + Z_i'b \quad (14)$$

Using Taylor expansion in (14)

$$y_i \approx h\left(x_i^{'} + Z_i^{'}b\right) + h'\left(x_i^{'}\hat{\beta} + Z_i^{'}\hat{b}\right)x_i^{'}\left(\beta - \hat{\beta}\right) + h'\left(x_i^{'}\hat{\beta} + Z_i^{'}\hat{b}\right)Z_i^{'}\left(\beta - \hat{\beta}\right) + \varepsilon_i \quad (15)$$

$$= \mu_i + V(\mu_i)x_i^{'}\left(\beta - \hat{\beta}\right) + V(\hat{\mu})Z_i^{'}\left(\beta - \hat{\beta}\right) + \varepsilon_i \quad (16)$$

Equation (16) can also be written as

$$y_i^* = V_i^{-'}(y_i - \mu_i) + x_i\hat{\beta} + Z_i\hat{b}_i = x_i\beta + Z_i b_i + \varepsilon_i^* \quad (17)$$

Where $y_i^* = pseudo - response$
Conditional variance for y_i^* is given by

$$var(y_i/b) = a_i'(\emptyset)V(\mu_i) \quad (18)$$

$\emptyset$ = Dispersion parameter

Laplace Approximation

The method uses a Taylor expression of an exponential form. It approximates integrals

$$\int e^{h(u)}\, du \quad (19)$$

Where $u = q - dimensinal\ vector$ and $h(u) = sufficiency\ smooth\ function$

We define the second order Taylor expansion for h in u_o as

$$h(u) \approx h(u_o) + \frac{1}{2}(u - u_o)^{'}h''(u_o)(u - u_o) \quad (20)$$

From (20) using Laplace approximate function

$$\int e^{h(u)}du \approx exp\left[h(u_o)\right](2\pi)^{\frac{q}{2}} |-h''(u_o)|^{\frac{-1}{2}} \quad (21)$$

In order to approximate the likelihood function, we consider the Laplace approximation

$$L = log\int fy/u(y/u)\, fu(u)du = log\int exp\left[logfy/u(y/u)\right] + logfu(u)du \quad (22)$$

Where
$h(u) = logfy/u(y/u) + logfu(u)$
we assume
$u \sim N(0, D)$ where u =univariate Normal distribution

$$\log(fu) = \frac{1}{2}u'D^{-1}u - \frac{q}{2}\log(2\pi) - \frac{1}{2}log|D \quad (23)$$

$$\frac{\partial logfu}{\partial u} = -D^{-1}u$$

$$\frac{\partial^2 logfu}{\partial u \partial u|} = -D^{-1} \quad (24)$$

Using chain rule on the exponential family, we have

$$\frac{\partial logfy/u(y/u)}{\partial u} = \frac{1}{\varphi}\sum_i \left(y_i\frac{\partial\theta_i}{\partial u} - \frac{\partial b(\theta_i)\partial\theta_i}{\partial\theta_i\partial u}\right)$$

$$= \frac{1}{\varphi}\sum_i (y_i - \mu_i)\frac{1}{v(\mu_i)}\frac{1}{g^{"}(u_i)Z_i'} \quad (25)$$

To find μ_o we solve

$$\frac{\partial h(u)}{\partial u} = \frac{1}{\varphi}Z'W\Delta(y - \mu) - D^{-1} = 0 \quad (26)$$

$$\frac{\partial^2 h(u)}{\partial u \partial u'} = \frac{\partial}{\partial u'}\left(\frac{1}{\varphi}Z'W\Delta(y - \mu) - D^{-1}u\right) = \frac{1}{\varphi}\left(-Z^{'}W\Delta\frac{\partial\mu}{\partial u'} + Z^{'}\frac{\partial W}{\partial u'}(y - \mu) - D^{-1}\right) \quad (27)$$

For calculative convenience, we choose to ignore the second term (20)

$$\frac{\partial^2 h(u)}{\partial u \partial u'} = -\frac{1}{\varphi}(Z'WZD + I)D^{-1} \quad (28)$$

Putting (20) and (30), we have

$$L \approx logfy/u(y/u_o) - \frac{1}{2}u_o^{'}D^{-1}u_o - \frac{1}{2}\log\left|\left(\frac{1}{\varphi}Z'WZD + I\right)D^{-1}\right| \quad (29)$$

$$\frac{\partial l}{\partial\beta} = \frac{\partial logfy/u(y/u_o)}{\partial\beta} + \frac{\partial}{\partial\beta}\frac{1}{2}\log|\frac{Z'WZD}{\varphi} \approx \frac{1}{\varphi}X'W\Delta(y - \mu) \quad (30)$$

W changes with respect to β and gives an estimate of β and u by solving the equation

$$\frac{1}{\varphi}X'W\Delta(y - \mu) = 0 \quad (31)$$

$$\frac{1}{\varphi}Z'W\Delta(y - \mu) = D^{-1}u \quad (32)$$

Adaptive Gauss-Hermite quadrature methods

The adaptive version of the AGQ uses a Gaussian approach, in which a Gaussian function replaces the factor $\exp(-Z^2)$ with suitable changes in the weights and approximation points. We follow the outline and adapt it to the case of the GLMM. The goal is to approximate $\int(g(t))dt$ by transformation. To achieve this, we have:

$$\int f(t)\phi(t,\mu,\sigma)dt \tag{33}$$

We approximate the likelihood by picking optimal subdivisions to evaluate the integrand. Adaptive GHQ brings in information from an initial fit to increase precision. This requires a transformation of the sampling nodes according to the
Transformation from $exp[Z_i]$ to $\phi(t:\mu,\sigma)$ which equals

$$t_i = \mu + \sqrt{2\sigma Z_i} \tag{34}$$

Moreover as we want to sample the integral in the region of g(t), we can define μ as the mode of g(t) and $\sigma = 1/\sqrt{j}$ for

$$j = -\frac{\partial^2}{\partial t^2}\log\big(g(t)\big)/_{t=\mu} \tag{35}$$

$$\text{define } h(t) = \frac{g(t)}{\phi(t;\mu,\sigma)} \tag{36}$$

then we can rewrite the integral for g(t) as

$$\int g(t)dt = \int h(t)\phi(t,\mu,\sigma))dt \tag{37}$$

which after applying the transformed Gauss-Hermite quadrature equals

$$\int g(t)dt = \sqrt{2\sigma}\sum_{i=1}^{Q} w_i^* g\left(\mu + \sqrt{2\sigma Z_i}\right) \tag{38}$$
$$for\ w_i^* = w_i \exp(Z_i)$$

In the case of the GLMMs, the implementation of a single random effect can be seen as clustered into different groups. Every cluster i has a random effect which is distributed as $u_i \sim N(0,\sigma^2)$. Thus, the posterior mode of u_i needs to be determined, and this depends on the factors $\beta, \phi\ and\ \sigma$ and u_i. We replace these by the current estimate (and in the first step a well-chosen value) $\beta^*, \phi^*\ and\ \sigma^*$. Then using these estimates, define u_i which maximizes

$$f(y_i/u_i)f(u_i r\sigma^*) \propto f(u_i/y_i) \tag{39}$$

Thus we can use u_i as the mode for u_i and use the above Gauss-Hermite quadrature to approximate

$$\int fY/u\left(yi/u_i\right) fu\left(u_i\right)du_i \approx \sum_{i=1}^{Q} w_i^* \left(\prod_{j=1}^{n} fY/u(y_{ij}Z_i^*)\right) \tag{40}$$

in which u_i is the size of the cluster i, y_{ij} is the j-th element of cluster i and we have the adaptive weights

$$W_i^* = \sqrt{2\sigma_i} w_i \exp\left(Z_i^2\right)\phi\left(Z_i^*;0,1\right) \tag{41}$$

$$Z_i^* = \partial_i + \sqrt{2\sigma} Z_i \tag{42}$$

where σ_i is the approximation for $\sigma^{-1}u_i \sim N(0,1)$. Multiplication of this sum leads to the approximated maximum likelihood. This method leads to new current best estimates until convergence.

2.4 Simulation Study

In order to evaluate the strength of the three statistical estimation techniques with varying sample sizes. Simulation study was carried out to identify how the three underlying models performed at different sizes (that is, 20 to 1000. Random numbers were generated using a discrete uniform distribution ranging in the range Response variable was simulated at different n along with two predictors using (0, 1), (0, 1.5) (0, 1) and (0, 3) respectively, to generate predictors in the fitting of GLMMs with simulated binary responses. Different GLMMs parameter estimation methods were used to produce estimates with different sample sizes, model selection criteria was used to identify the best model. Coefficients and random intercept variances as estimated via PQL, LA and GHQ for each simulation scenario are generated with R with lme4 package in R by Bate et. al (2015).

2.5 Data Description

The data set consists of health data containing 1500 patients using some of the Health facilities in Ogun State, Nigeria for the period of July 2016 to July 2017. The data was collected with the aim of modelling and identifying which predictor impacts significantly on the response variable. Response variable in this work is follow-up, while predictors are Sex, Ages of patients, and Number of diagnosis respectively for the period of visits. For the period of observation, it either a patient is on follow-up or not as advised by the physician. The analysis centres on medical follow-up effects on Patients. The model parameters were estimated using Penalized Quasi-Likelihood (PQL), Gaussian Hermite quadrature (GHQ) and Laplace Approximation techniques (LA).

3. Results

The following table represents the descriptive statistics of the simulated data using R software package.

Table 1: Descriptive Statistics of Simulated Data

Desc.Stat	Y	X_1	X_2	X_3	X_4
Min.	0.0000	1.4047e-04	0.0688	0.1584	0.0531
Max.	1.0000	8.8124e-01	1.4285	1.9003	2.4090
Sum	20.000	3.9962+01	73.0204	101.9712	2.3559
Median	0.0000	3.1773e-01	0.7518	1.0222	1.2599
Mean	0.2000	3.9622e-01	0.7302	1.0197	1.1980
SE.mean	0.8020	3.2340e-02	0.0462	0.0520	0.0715
C.I mean	0.0797	6.4171e-02	0.0917	0.10330	0.1420
Var.	0.1616	1.0459e-01	0.2139	0.2710	0.5122
Std.dev	0.4020	3.2340e-01	0.4624	0.5262	0.7158
Coef.var	2.0107	8.1333e-01	0.6333	0.5105	0.5973
Skewness	1.3889	0.2249	0.0212	-0.1329	-0.1115
Kurtosis	-0.0668	-1.7106	-1.5156	-1.3983	-1.3563

Table 2: Descriptive Statistics for Real-Life Data

	Follow-up	Sex	Age	Ndiagnosis	Bgroup	Gnotype	Status
Mean	0.18466	0.46000	30.02067	2.681333	1.6560	1.6560	0.2106
Standard Error	0.01002	0.01287	0.471299	0.051015	2.2952e-02	2.2952e-02	0.9105
Median	0.00000	0.00000	35.00000	2.000000	1.0000	1.0000	0.0000
Mode	0.00000	0.00000	37.0000	1.000000	1.0000	1.0000	0.0000
Standard Deviation	0.38815	0.49856	18.25332	1.975784	7.9010e-01	7.9010e-01	0.40791
Sample Variance	0.15066	0.24857	333.1837	3.903721	8.8890e-01	8.8890e-01	0.1664
Kurtosis	0.64781	-1.97680	-1.36130	4.242080	0.8244	0.8244	0.00972
Skewness	1.62694	0.16067	-0.12010	1.748514	1.2974	1.2974	1.4174
Range	1.00000	1.00000	77.00000	14.00000	3.0000	3.0000	1.0000
Minimum	0.00000	0.00000	0.000000	1.00000	1.0000	1.0000	0.00000
Maximum	1.00000	1.00000	77.00000	15.0000	4.0000	4.0000	1.00000
Sum	277.000	690.000	45031.0	4022.0	2.4843e+03	2.4843e+03	316.0000
Count	1500	1500	1500	1500	1500	1500	1500

For the real life data, from Table 2, the response variable is the follow-up status of patients accounting for 1500 users of the various health facilities within the space of twelve months. The covariates/predictor variables include sex, age, number of diagnoses, blood group, genotype and smoking status of individual patients. The Table also indicates descriptive statistics of both response and predictor variables with their corresponding properties. Considering the properties of the response variables, it shows lack of normality which justifies the fitting through generalized linear mixed models.

Table 3: Simulation Results for Penalized Likelihood Model

n	AIC	BIC	DF
20	410.46	426.09	15
30	108.31	116	25
40	109.54	116	35
50	314.82	330.45	95
100	269.35	284.98	95
200	503.50	523.29	195
400	1246.47	1270.42	395
500	1384.07	1409.36	495
1000	2799.44	2828.87	995

Table 4: Simulation Results for Laplace Approximation Model

n	AIC	BIC	Loglik	Deviance	DF
20	21.20	36.90	-4.60	9.2	94
30	53.4	71.50	-20.70	41.40	144
40	63.60	71.50	-25.80	51.60	194
50	79.30	100.40	-36.60	67.30	244
100	162.00	187.30	-75.00	150.00	494
200	278.20	307.70	-133.10	266.20	994
400	609.10	642.70	-298.50	597.10	1994
500	734.90	769.80	-361.40	722.90	2494
1000	1534.70	1573.80	-761.40	1522.70	4994

Table 5: Simulation Results for Adaptive Gauss-Hermite Quadrature Model

n	AIC	BIC	Loglik	Deviance	DF
20	29.00	44.60	-8.50	17.0	94
30	50.50	68.6	-19.20	38.50	144
40	65.40	85.20	-26.70	53.40	194
50	74.00	95.10	-31.00	62.00	244
100	152.20	177.50	-70.10	140.20	494
200	256.20	285.60	-122.10	244.20	994
400	497.60	531.20	-242.80	485.60	1994
500	637.60	672.60	-312.80	625.60	2494
1000	1276.70	1315.80	-632.30	1264.70	4994

Table 6: A) Estimates of Penalized Likelihood Model for Health Data

	Estimate	Std.error	Z-value	Pr(>IZI)
Intercept	-1.5694	0.2863	-5.4800	4.25e-08***
Sex	-0.1166	0.1220	-0.9560	0.3390
Age	0.03095	0.00381	8.1070	5.20e-16***
Ndiagnosis	0.01257	0.02977	0.4300	0.6670
Bgroup	-0.3457	0.0866	-5.1081	3.2431
Gnotype	-0.0814	0.0697	-1.1681	0.2431
Sstatus	-0.03506	0.1473	-0.2381	0.8127

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '.

Table 6: B) Model Performance of Penalized Likelihood Model

AIC	BIC	Null deviance	R. deviance	Df
1398	1435.15	935.45	829.97	1498

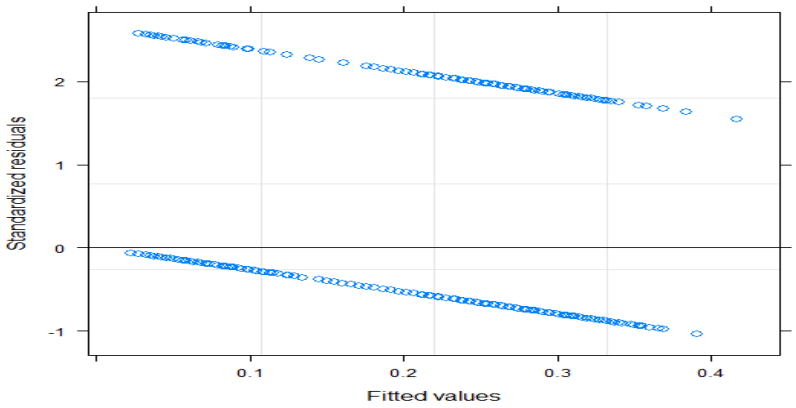

Fig. 1: Standard Residual Plot for Penalized Likelihood Model of Health Data.

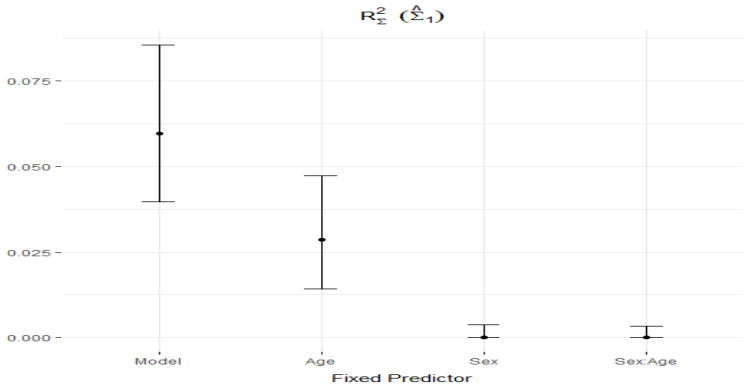

Fig. 2: Visualize Standardized Effect Sizes and Model R Squared for Penalized Likelihood Model of Health Data.

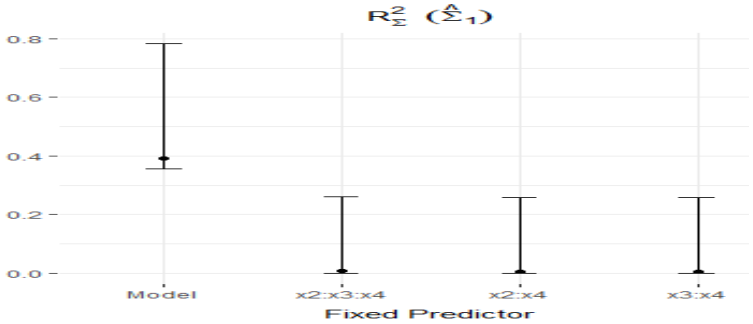

Fig. 3: Visualize Standardized Effect Sizes and Model R Squared of Simulated Data for Penalized Likelihood Model with $n = 20$.

Table 7: A) Coefficient of Predictors Using LA Model

	Coef	s.e(coef)	Z-value	Pr(>IZI)
Intercept	-1.5694	0.2863	-5.4800	4.25e-08***
Sex	-0.1166	0.1220	-0.9560	0.3390
Age	0.03095	0.00381	8.1070	5.20e-16***
Ndiagnosis	0.01257	0.02977	0.4300	0.6670
Bgroup	-0.3457	0.0866	-5.1081	3.2431
Gnotype	-0.0814	0.0697	-1.1681	0.2431
Sstatus	-0.03506	0.1473	-0.2381	0.8127

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 7: B) Model Performance of Laplace Approximation Model for Health Data

AIC	BIC	Null deviance	Deviance	df
1398	1435.15	935.45	829.97	1498

Table 8: A) Coefficient of Predictors Using Gauss-Hermite Quadrature Model for Health Data

	Coef	s.e(coef)	Z-value	Pr(>IZI)
Intercept	-1.19124	0.3280	-3.6310	0.00028***
Sex	-0.1367	0.1411	-0.9690	0.3326
Age	0.0389	0.00434	8.9700	<2e-16***
Ndiagnosis	0.0181	0.0346	0.5220	0.6015
Bgroup	-0.4662	0.0805	-5.7910	7.02e-09***
Gnotype	-0.1060	0.0797	-1.3310	0.1833
Sstatus	-0.0507	0.1710	-0.2970	0.7668

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 8: B) Model Performance of Gauss-Hermite Quadrature Model for Health Data

AIC	BIC	Loglit	Deviance	df
1318.7	1361.2	-651.3	1302.7	1491

Table 9: Summary of Model Performances for PQL, LA, Glmm TMB and GHQ Using Real Life Health Data Provided

Models	AIC	BIC	Loglik	Deviance	DF
PQL	1398.00	1435.15	-	935.45	1498
LA	1400.00	1442.50	-692.00	1384.00	1491
GHQ	1318.70	1361.20	-651.20	1302.70	1491

4. Discussion

This study presented and explored methods of estimating parameters of Generalized Linear Mixed Models; the descriptive statistics of the simulation was presented in Table 1. Using the simulated data, the performances of the underlying models were assessed with model selection criteria, and results from Table 3, Table 4, and Table 5 shows that that Gauss- Hermite quadrature is superior to the two other models (Penalized Likelihood Model, and Laplace Approximation Model) based on model selection criteria.

Estimation was further carried out using From Gauss Hermite Quadrature (glmmGHQ) model given in Table 8, having established its potency as shown in Table 9. It was observed that the number of diagnoses account for more follow-up status of individual patients. By implication, it shows that every increase in follow-up status, there is an increase of 0.03 factors in the number of diagnoses of individual patients. It was also observed that for every increase in the follow-up status, there is a decrease in the smoking status of individual patients by a factor of 0.1. By implication patients with follow-up status tend to avoid smoking which accounts for the decrease. It was also observed that for every increase in the follow-up status, there is an increase in ages of patients by a factor of 0.4. By implication, older patients tend to have more diagnoses than other age categories. Therefore, Gauss Hermite Quadrature and family, is hereby recommend in fitting Generalized Linear Mixed Models using binary response data.

References

[1] Y. Lee and J. A. Nelder, Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika, (2001)88, 987–1006. https://doi.org/10.1093/biomet/88.4.987.

[2] N.E. Breslow, Extra-poisson variation in log-linear models. *Applied Statistics*, 33, (1984), 38–44. https://doi.org/10.2307/2347661.

[3] B.M. Bolker, C.J. Brooks, S.W. Clark, Geange, J. R, M.H.H. Poulsen, Stevens, J.S. White, Generalized linear mixed models: a practical guide for ecology and evolution, Trends in Ecology & Evolution, 24, (2009),127–135. https://doi.org/10.1016/j.tree.2008.10.008.

[4] M. Casals, K. Langohr, J.L. Carrasco1, L. Ronneg, Parameter Estimation of Poisson Generalized Linear Mixed Models Based on Three Different Statistical Principles: a Simulation Study. SORT 39 (2), (2015), 1-28.

[5] C. E. McCulloch S.R. Searle Generalized, Linear, and Mixed Models, New York: Willey, (2001).

[6] R Core Team, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org, (2018).

[7] D.Gamerman, Sampling from the Posterior Distribution in Generalized Linear Mixed Models. *Statistics and Computing*, 7, (1997), 57-68. https://doi.org/10.1023/A:1018509429360.

[8] O.F. Christensen, R. Waagepetersen, Bayesian Prediction of Spatial Count Data Using Generalized Linear Mixed Models, *JSTOR Biometrics*, 58 (2) (2002), 280-286.

[9] Y. Zhao, J. Staudenmayer, B.A. Coull, M.P. Wand, General Design Bayesian Generalized Linear Mixed Models. Mathematics and statistics online, Statist. Sci. Volume 21 (1), (2006), 35-51. https://doi.org/10.1214/088342306000000015.

[10] Y. Fong, H. Rue, J. Wakefield, Bayesian inference for generalized linear mixed models. *Biostatistics*, 11, (2010) 397–412. https://doi.org/10.1093/biostatistics/kxp053.

[11] O.S. Adesina, D.A. Agunbiade, O.S. Osundina, Bayesian Regression Model for Counts in Scholarship, *Journal of Mathematical Theory and Modeling*. 7, 9, (2017) 46-57.

[12] B. Efron, B. Double Exponential Families and Their Use in Generalised Linear Regression. *Journal of the American Statistical Association*, 81(395), (1986) 709-721: URL https://doi.org/10.1080/01621459.1986.10478327.

[13] O.S. Adesina, T.O. Olatayo, O.O. Agboola, P.E. Oguntunde, Bayesian Dirichet Process Prior for Count Data. *International Journal of Mechanical Engineering and Technology (IJMET), 9, 12, (2018) 630-646.*

[14] M.E. Brooks, K. Kristensen, K.J. Benthem, A. Magnusson, C.W. Berg, A. Nielsen, H.J. Skaug, M. Maechler, M. Bolker, Modelling Zero-Inflated Count Data With glmmTMB. (2017) bioRxiv preprint bioRxiv: 132753; https://doi.org/10.1101/132753.

[15] P. Jong, G.Z. Heller, Generalized Linear Models for Insurance Data, (1st ed., Cambridge University Press, New York. (2008) ISBN-13 978-0-511-38677-0). https://doi.org/10.1017/CBO9780511755408.