BAYESIAN LOGISTIC REGRESSION USING GAUSSIAN NAÏVE BAYES

J. F. Ojo, R.O.Olanrewaju, & S.A.Folorunsho
Department of Statistics,
University of Ibadan, Ibadan, Nigeria.
Email: jfunminiyiojo@yahoo.co.uk, rasakiolawale@gmail.com, serifatf005@gmail.com,

ABSTRACT

This study describes the approach of Gaussian Naïve Bayes (GNB) as a prior distribution classifier in a two-class (dichotomous) classification of the posterior probability of the dependent variable $P(Y_i, i = 0,1/X_i)$ in a Bayesian logistic regression. This approach establishes the procedure for parameter estimation of Bayesian logistic regression when we could not ascertained whether the prior distribution is informative or non-informative. The Newton-Raphson iterative procedure was used in estimating the vector parameters because there was no closed-form solution due to non-linearity of the logistic function. This study was applied to four set of panaceas drugs on diarrhea treatment for babies less than a year old (Nigeria Demographic Health Survey (NDHS, 2013)). It was noted that the standard errors of parameters estimated via Bayesian logistic regression using the GNB were lower than that of standard errors of parameters estimated via the Classical Logistic Regression (CLR) using the Maximum Likelihood Estimation (MLE), which makes Bayesian logistic regression via GNB better than CLR.

Keywords: Gaussian Naïve Bayes, Bayesian Logistic Regression, Maximum Likelihood Estimation, Posterior Distribution, Prior Distribution.

INTRODUCTION

Prior distribution is a key part of a Bayesian framework that represents the information about some uncertain Parameter (s) that when combined with the probability distribution of the data yields the posterior distribution which is used for future references and decisions involving those parameter(s) (Andrew, 2002). Andrew (2006) noted that some key issues coming up on prior distribution are what information is going into the prior distribution and the properties of the resulting posterior distribution. There are different strategies used in assigning prior distributions to different parameter(s) in a model. These strategies collectively refer to the information going into prior distribution. This information, in turn, is bifurcated into: Non-informative (objective) and Informative (subjective) priors (Dana *et al.* 2010). Non-informative priors are used in a setting where scientific objectivity is paramount.

In other words, Non-informative priors are used in a Bayesian framework to deal with complex multi-dimensional models or multi-dimensional posterior probability distributions. Andrew (2006) affirmed the point made by (Zeng, 2002) that non-informative priors are restricted to distributions such as Inverse-gamma, uniform distribution, Walshart distribution and Jeffrey's rule. However, Andrew (2006) claimed that non-informative prior distribution seemed to be positively biased whenever the lower limit of the range of these distributions made mentioned are less than four. Based on this, some Bayesian analysts concluded that non-informative priors might be misleading, diffuse and vague (Robert *et al.* 1996), (Mu Zhu, 2004) and (Ulrich, 2012).

The informative priors, on the other hand, avow that analysis is based on something more than the data in hand whose importance to the parameters of interest is modeled through the likelihood. The point we are making is that Informative priors make use of the pre-knowledge of distribution that the data follows. Mark *et al.* (2011) affirmed that informative priors are based on the pre-knowledge about the substantive problem based on the data along with elicited expert opinion if possible, to construct a prior distribution that properly reflects the researcher's beliefs on about the unknown parameter(s). They were of the notion that informative priors may seemly over subjected and unscientific. Now, the informative prior is also sub-divided into conjugate and non-conjugate priors. The former arises when the likelihood times the prior distribution produces a recognizable posterior kernel of the same form with the prior distribution whilst reverse is for the later. In conclusion, there seems to be no clear-cut whether the objective approach is better-off the subjective or vice versa.

LITERATURE REVIEW

Logistic regression model has been widely used in modeling, *inter alia,* biostatistics data, epidemiological data, and biometric data. Greenland (2006) opined that parameters in logistic regression analysis that are usually carried out using the classical approach via Maximum likelihood Estimation (MLE) are founded on the bases of randomization and random sampling such that its usage in observational studies is questioned. He advocated and urged that Bayesian logistic regression is more appropriate for studies wherein the procedures used in generating the samples and data do not follow randomization. In a similar vein, Wioletta (2015), in his work, juxtaposed the credible interval derived from the posterior means and the mean gotten from the descriptive statistics to conclude that Bayesian confidence regions yielded information on the range of the changes of estimated parameters with probability of 0.95.

Also, that informative prior do result in significant reduction of highest posterior density region compared to non-informative prior. On his part, Tom (2005) designed learning algorithms based on understanding of probability distribution of discrete values for a dependent variable given some set of conditional independent variables. His work was mainly on conditional independence of covariates on dependent variable given an interaction variable using Baye's rule. However, his work failed to model a two-class of dependent variable to some set of covariates. Chia-chung (2004) assumed Gaussian distribution as an informative prior in Bayesian Inference in Binomial Logistic Regression as a case study of the 2002 Taipei Mayoral Election. He narrowed his work down to fitting a binary logistic regression model with the observed data using lay down Gaussian distribution as an informative prior without challenging the theoretical buildup. Hee *et al.* (2013) came up with a theoretical framework called Polya-Gamma Gibbs Sampler for Bayesian logistic regression.According to them, Polya-Gamma Gibbs Sampler was uniformly ergodic for Bayesian probit regression. In a similar vein, Bosabella et al. (2014) used "vague informative prior" in Bayesian Logistic Regression Analysis of the Association of Intimate Partner Violence and Modern Contraceptive use in the Philippines. Like Bosabella *et al.* (2014), Andrea *et al.* (2014) in their work came up with penalized likelihood with Gaussian distribution, log-normal distribution or Log-F as the informative prior distribution in estimating parameters in the posterior distribution of the logistic regression. They claimed that the three prior distributions are symmetric, and unimodal, and that log-normal prior was equivalent to the Odd Ratio (OR) scale while that of the log-F priors are more flexible than normal priors and useful when priors are directional. Based on these non-conformities and aforementioned short falls in both the non-informative and informative, and that no clear distinction whether one is better than the other, we are of the impression that a robust naïve prior distribution needs to be incorporated to neutralize the hiatus of range of value limitation in non-informative priors, and the task of the researcher having the pre-knowledge of the source of the data and parameter(s) to be estimated, a full range of prior distribution that can accommodate all real number values needed to be proposed to abridge the lacuna between informative and non-informative prior. So, a theoretical framework of Gaussian distribution as a naïve prior distribution using Baye's rule, that is Gaussian Bayes Naïve (GNB) will be used in this research work.

METHODOLOGY

Classical Logistic Regression

Logistic regression is an approach for constraining a discrete response variable to a well-defined vector of a discrete or continuous covariates, that is, $f: X \to Y \ or \ P_r(Y/X)$, where Y_i, the response variable is a member of the exponential family gotten from the binomial distribution. For a two-class classification problem of the response variable (Binary Logistic Regression), $Y_i, i = 0,1$ the conditional probability mass function of Y will be either $P(Y = 1/X)$ or $P(Y = 0/X)$, (Brian, 2012).

Suppose $Y_i, i = 0,1$ denoted a set of responses of a binary outcome variable of Y, and X_i are the vector of the corresponding covariates of a specified dimension say "p". Then the binary logistic regression model from Generalized Linear Model (GLM) can be specified as:

$$\eta_i = Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \qquad (1)$$

$$X' = (X_1, X_2, \cdots, X_p),\ \beta' = (\beta_1, \beta_2, \cdots \beta_p)$$

whereX is the design matrix of $p_x p$, β is a 1 by p regression coefficients

The link function $g(u_i)$ which transforms the response variable $\pi_i = E(Y_i)$ to linear predictor

$$g(\pi_i) = \eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \quad (2)$$

The link function of Bernoulli distribution known to be $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$

So, Inverting the link function,

$$\pi_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p) \quad (3)$$

$\frac{\pi_i}{1-\pi_i} = e^{\eta_i}$, $\pi_i = (1-\pi_i)e^{\eta_i}$, $\pi_i = e^{\eta_i} - \pi_i e^{\eta_i}$ $\pi_i + \pi_i e^{\eta_i} = e^{\eta_i}$, $\pi_i = \frac{e^{\eta_i}}{1-e^{\eta_i}}$

Since, $\eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$

$$\pi_i = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}} \text{ for } i = 0,1 \quad (4)$$

Bayesian Logistic Regression using Gaussian Naïve Bayes

The Gaussian Naïve Bayes (GNB) classifier for the conditional probability for $P(Y/X)$ is adopted. Let the response variable $Y_i, i = 0,1$ be a Boolean governed by a Bernoulli distributed with parameter π *for* $P(Y=1/X)$ and $1-\pi$ *for* $P(Y=0/X)$ for each covariates X_i, $P(X_i/Y=y_k) \square (u_i, \sigma_i)$ for X_i *and* X_j that are conditional independently given Y such that $i \neq j$ for each i and j such that the standard deviations vary but do not depend on Y(Brian, 2012). The conditional probability (posterior distribution) from the GNB assumptions is derived using Baye's rule as follows:

$$P(Y_i/X) = \frac{P(Y_i).P(X/Y_i)}{\sum_{i=1}^{n} P(Y_i).P(X/Y_i)} \quad (5)$$

For $Y_i, i = 0,1$; $P(Y=1/X) = \frac{P(Y=1).P(X/Y=1)}{P(Y=1).P(X/Y=1) + P(Y=0).P(X/Y=0)}$ (6)

Dividing both the numerator and denominator by $P(Y=1).P(X/Y=1)$

$$P(Y=1/X)=\frac{\dfrac{P(Y=1).P(X/Y=1)}{P(Y=1).P(X/Y=1)}}{\dfrac{P(Y=1)P(X/Y=1)+P(Y=0)P(X/Y=0)}{P(Y=1).P(X/Y=1)}} \quad (7)$$

$$P(Y=1/X)=\frac{1}{\dfrac{P(Y=1)P(X/Y=1)}{P(Y=1)P(X/Y=1)}+\dfrac{P(Y=0).P(X/Y=0)}{P(Y=1)P(X/Y=1)}} \quad (8)$$

$$P(Y=1/X)=\frac{1}{1+\dfrac{P(Y=0).P(X/Y=0)}{P(Y=1).P(X/Y=1)}} \quad (9)$$

$$P(Y=1/X)=\frac{1}{1+e^{\ln\left[\frac{(P(Y=0).P(X/Y=0))}{P(Y=1).P(X/Y=1)}\right]}} \quad (10)$$

$$P(Y=1/X)=\frac{1}{1+e^{\ln\left[\frac{P(Y=0)}{P(Y=1)}x\frac{P(X/Y=0)}{P(X/Y=1)}\right]}} \quad (11)$$

$$P(Y=1/X)=\frac{1}{1+e^{\left[\ln\frac{P(Y=0)}{P(Y=1)}+\sum_i\frac{\ln P(X_i/Y=0)}{\ln P(X_i/Y=1)}\right]}} \quad (12)$$

$$P(Y=1/X)=\frac{1}{1+e\left[\ln\dfrac{1-\pi}{\pi}+\sum_i\ln\dfrac{P(X_i/Y=0)}{P(X_i/Y=1)}\right]} \quad (13)$$

It is to be noted that $\ln\left(\frac{1-\pi}{\pi}\right)$ is the inverse of the link function.

The final expression $P(Y=1)$ is in terms of the binomial parameter π. Considering the summation in the denominator in equation (13) and given our assumption of GNB that $P(X_i/Y=y_k)$ is Gaussian, we expand as follows:

$$\sum_i\ln\frac{P(X_i/Y=0)}{P(X_i/Y=1)}=\sum_i\ln\left[\frac{\dfrac{1}{\sqrt{2\pi\sigma^2}}\varepsilon xp^{\left[\frac{-(X_i-\pi_{i0})^2}{2\sigma_i^2}\right]}}{\dfrac{1}{\sqrt{2\pi\sigma^2}}\varepsilon xp^{\left[\frac{-(X_i-\pi_{i1})^2}{2\sigma_i^2}\right]}}\right]$$

$$= \sum_i \ln \frac{1}{\sqrt{2\pi\sigma^2}} . \sqrt{2\pi\sigma^2} \frac{\varepsilon xp^{\left[\frac{-(X_i-\pi_{i0})^2}{2\sigma_i^2}\right]}}{\varepsilon xp^{\left[\frac{-(X_i-\pi_{i0})^2}{2\sigma_i^2}\right]}}$$

$$= \sum_i \ln \varepsilon xp^{\left[\frac{\left[(X_i-\pi_{i1})^2-(X_i-\pi_{i0})^2\right]}{2\sigma_i}\right]}$$

$$= \sum_i \frac{\left(X_i^2 - 2X_i\pi_{i1} + \pi_{i1}^2\right) - \left(X_i^2 - 2X_i\pi_{i0} + \pi_{i0}^2\right)}{2\sigma_i^2}$$

$$= \sum_i \frac{\left(X_i^2 - 2X_i\pi_{i1} + \pi_{i1}^2 - X_{i0}^2 + 2X_i\pi_{i0} - \pi_{i0}^2\right)}{2\sigma_i^2}$$

$$= \sum_i \left(\frac{2X_i\left(\pi_{i0} - \pi_{i1}\right) + \pi_{i1}^2 - \pi_{i0}^2}{2\sigma_i^2} \right)$$

$$= \sum_i \left(\frac{\left(\pi_{i0} - \pi_{i1}\right) X_i}{\sigma_i^2} + \frac{\pi_{i1}^2}{2\sigma_i^2} \right) \qquad (14)$$

Equation (14) is called linear weighted sum of the $X_i's$. Substituting equation (14) back in equation (13).

$$P\left(Y=1/X\right) = \frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi} + \sum_i \left(\frac{\pi_{i0}-\pi_{i1}}{\sigma_i^2} X_i + \frac{\pi_{i1}^2 - \pi_{i0}^2}{2\sigma_i^2}\right)\right)}$$

For $w_i = \frac{\pi_{i0}-\pi_{i1}}{\sigma_i^2}$, $\quad w_0 = \ln\frac{1-\pi}{\pi} + \sum_i \frac{\pi_{i1}^2 - \pi_{i0}^2}{2\sigma_i^2}$

Then,

$$P\left(Y=1/X\right) = \frac{1}{1+\exp\left(w_0 + \sum_{i=1}^{n} w_i X_i\right)},$$

where $w_1, w_2, \cdots, w_n$ are weights, then

$$P\left(Y=1/X\right) = \sigma\left(W^T X_i\right)$$

$$P\left(Y=0/X\right) = 1 - \sigma\left(W^T X_i\right)$$

Where $\sigma(.)$ is the logistic sigmoid function defined by

$$\sigma(a) = \frac{1}{1+\exp(-a)} \qquad (15)$$

Parameter Estimation for the Bayesian Logistic Regression using Gaussian Naïve Bayes

From the log of the conditional likelihood of

$$\ln P(Y_i / X_i, W) = \sum_i^n Y_i \ln P(Y_i = 1/ X_i) + (1 - Y_i)\ln(Y_i = 0 / X_i, W)$$

$$= \sum_i^n Y_i \ln \sigma\left(W^T X_i\right) + (1 - Y_i)\ln\left(1 - \left(\sigma\left(W^T X_i\right)\right)\right)$$

$$= \sum_i^n \ln \frac{\sigma\left(W^T X_i\right)}{1 - \sigma\left(W^T X_i\right)} + \ln\left(1 - \sigma\left(W^T X_i\right)\right)$$

$$\ell(w; x) = \ell(w; x_i) = \ln P(Y / X_i, w) = \sum_i^n \frac{\ln \sigma\left(w^T X_i\right)}{1 - \sigma\left(w^T X_i\right)} + \ln\left(1 - \sigma\left(w^T X_i\right)\right)$$

$$= \sum_i^n \left\{ \frac{x_i \theta_i(w) - b(\theta_i(w))}{a_i(\phi)} + c(y_i, \theta) \right\}$$

By chain rule,

$$\frac{\partial \ell_i}{\partial W_i} = \frac{\partial \ell_i}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \sigma_i} \cdot \frac{\partial \sigma_i}{\partial w_i} \qquad (16)$$

From the canonical form of thebernoulli distribution, that is

$$\Pr(x_i, \pi) = \pi^{x_i}\left(1 - \pi^{1 - x_i}\right)$$

Re-writing in an exponential form

$$f(x_i, \pi_i) = \exp\left\{\log\left\{\pi^{x_i}\left(1 - \pi^{1 - x_i}\right)\right\}\right\}$$

$$f(x_i, \pi_i) = \exp\left\{x_i \log \pi_i + (1 - x_i)\log(1 - \pi_i)\right\}$$

$$= \exp\left\{x_i \log \pi_i + \log(1-\pi_i) - x_i \log(1-\pi_i)\right\}$$

$$= \exp\left\{x_i \log \pi_i - x_i \log(1-\pi_i) + \log(1-\pi_i)\right\}$$

$$= \exp\left\{\log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right\}$$

where,

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right),\ b(\pi_i) = \log(1+e_i^{\theta}),\ a_i(\phi) = 1$$

$$w_i = 1,\quad \phi = 1,\quad c(y_i,\theta) = 1,\quad \pi = \frac{e^{\theta}}{1+e^{\theta}}$$

$$so,\quad \mu_i = b^{'}(\theta_i) = \frac{e^{\theta}}{1+e^{\theta}} = \pi_i$$

$$\text{var}(x_i) = a_i(\phi).b^{''}(\theta_i) = \pi_i(1-\pi_i)$$

So, equation (15) equals,

$$\frac{\partial \ell_i}{\partial W_j} = \frac{\partial \ell_i}{\partial \pi_i}.\frac{\partial \pi_i}{\partial \mu_i}.\frac{\partial \mu_i}{\partial \sigma_i}.\frac{\partial \sigma_i}{\partial w_i}$$

$$= \frac{x_i - b^{'}(\pi_i)}{a_i(\phi)}.\frac{1}{b^{''}(\pi_i)}.\frac{1}{g^{'}(\mu_i)}.x_{ij}$$

$$=\frac{\sum_{i}^{n}(x_i-\mu_i)}{1}.1.\frac{1}{g^{'}(\mu_i)}.x_{ij}$$

$$let\ d_i=g^{'}(\mu_i)=\frac{\partial\sigma_i}{\partial\mu_i};u_i=\frac{1}{\mathrm{var}(x_i)d_i^2}=\frac{1}{a_i(\phi)b^{"}(\pi_i)d_i^2}$$

$$\frac{\partial\ell}{\partial W_j}=\sum_i\frac{\partial\ell_i}{\partial W_j}=\sum_i\left(x_i-\mu_i\right)d_i x_{ij}=0,\qquad j=1\cdots p$$

Where $\mu_i\ and\ d_i$ depend on w.

The solution has no closed-form, which collaborate with equation (1) that gives the cross entropy error function and it has no close form solution to maximize the likelihood with respect to W. Since, there is no close-form solution due to the non-linearity of the logistic sigmoid function in equation (15). The parameter in the maximum likelihood can be estimated by iterative technique of Newton-Raphson iterative optimization method which uses a local quadratic approximation of the log-likelihood function.

We want to solve $f(w)=0$. We need to find $w^{(new)}$ satisfying $f(w^{(new)})=0$ which require $\left|f^{'}(w^{new})\right|>0$

A Taylor Series expansion of $f^{'}(w^{new})$ gives

$$0=f(w^{new})=f(w^{old})+f^{'}(w^{old})(w^{new}-w^{old})+\cdots$$

Which implies,

$$w^{(new)}\approx w^{(old)}-\frac{f(w)}{f^{'}(w)}$$

For w^{old} close to w^{new}. This implies the iteration of

$$w^{(new)}=w^{(old)}-\frac{f(w^{(old)})}{f^{'}(w^{(old)})}$$

or in a case of multiple vector where "f" is the score vector, $\frac{\partial \ell}{\partial w}$

$$w^{new} = w^{old} - \left[\left(\frac{\partial^2 \ell}{\partial w \partial w^T} \right)_{w=w^{new}} \right]^{-1} \left(\frac{\partial \ell}{\partial w} \right)_{w=w^{new}}$$

$$W^{(new)} = W^{(old)} - \left[\nabla^2 \sigma\left(W^{(old)}\right)\right]^{-1} \nabla f\left(W^{(old)}\right) \qquad (17)$$

$$\nabla E(W) = \sum_{i}^{n} \left(W^T X_n - Y^{'}\right) X_n$$

$$= X^T XW - X^{'}Y^{'}$$

also,

$$\nabla^2 E(W) = \sum_{i=1}^{n} X_n X_n^{\ T}$$

$$= X^T X$$

Substituting $\nabla^2 E(W)$ into (17)

$$W^{(new)} = W^{(old)} - \left(X^T X\right)^2 \left\{X^T XW^{(old)} - X^T Y\right\}$$

$$= \left(X^T X\right)^{-1} X^T Y \qquad (18)$$

Where $W^{(new)}$ converges

EXPERIMENTAL WORK

The data considered in this research work was the panaceas used for diarrhea treatment on babies less than a year old in Nigeria from a survey conducted by the Nigeria Demographic Health Survey (NDHS, 2013) to determine whether the four set of panacea drugs are effective in curing diarrhea among babies or not. The NDHS sample was designed to represent Government Hospitals in each of the 36 states. The covariates include the oral rehydration, home solution, antibiotics pills/syrups and Zinc used on 500 babies suffering from diarrhea.

POSTERIOR SUMMARIES

Table1: Empirical Mean and Standard Deviation for Classical Logit and Posterior Moments

	Classical Logit			Posterior Moments (Naïve)		
Variables (panaceas)	**Mean**	**SD**	**S.E**	**Mean**	**SD**	**S.E**
(Intercept)	-2.9131	0.2979	0.0103	-2.9217	0.1534	0.0025
Oral rehydration	0.2230	0.0962	0.0037	0.2238	0.0852	0.0010
Home solution	0.0832	0.0683	0.0020	0.0800	0.0509	0.0005
Antibiotics pills/syrups	1.1819	0.3104	0.0118	1.1797	0.3000	0.0030
Zinc	-1.4475	0.6528	0.0452	-1.8539	0.5630	0.0065

From table1 above, diarrhea as the dependent variable and oral rehydration, home solution, antibiotics pills/syrups and Zinc as the independent variables were evaluated for both the Bayesian and Classical logistic regression. The positive signs of means of Oral rehydration, Home solution, and Antibiotics pills/syrups in both Bayesian logistic regression and classical logistic regression unfolds that there is a positive contribution or relationshipin curing diarrhea among the babies while Zinc effect on diarrhea on the babies are inconclusive. The estimated means of the two approaches are very close but that of Bayesian logistic regression using GNB were smaller. Also, noted were the standard errors of the Bayesian logistic regression using GNB that were smaller than that of logistic regression which indicated a greater stability of the parameters estimated via Bayesian logistic regression using GNB.

Table 2: Quantiles of Posterior Distribution

Variables	2.5%	25%	50%	75%	97.5%
(Intercept)	-3.4200	-3.0979	-2.9234	-2.7378	-2.4311
Oral rehydration	0.0356	0.1594	0.2239	0.2852	0.4092
Home solution	0.0246	0.0458	0.0816	0.1161	0.1760
Antibiotics pills/syrups	0.5831	0.9831	1.1785	1.3824	1.7615
Zinc	-3.5213	-2.1800	-1.7313	-1.3970	-0.8944

From table 2, the quantiles indicated that parameters are mostly around 0.2230, 0.0832, 1.1819 and -1.4475 for Oral rehydration, Home solution, Antibiotics spills/syrups and Zinc with a 2.5% probability taking a value below 0.0356, 0.0246, 0.5831 and -3.5213 or a value above 0.4092, 0.1760, 1.7615 and -0.8944 respectively.

While:

$$P(Y=1/X)=\sigma\left(W^T X_i\right)$$

$$=\frac{1}{1+\exp\left(-2.9217+0.2238X_1+0.0800X_2+1.1797X_3-1.8539X_4\right)}$$

Then,

$$P(Y=0/X)=1-\sigma\left(W^T X_i\right)$$

$$=1-\frac{1}{1+\exp\left(-2.9217+0.2238X_1+0.0800X_2+1.1797X_3-1.8539X_4\right)}$$

Table 3: Predictive Probabilities

Variables	Oral rehydration	Home solution	Antibiotics pills/syrups	Zinc
Logistic Sigmoid $\sigma(a_i)$	0.7055	0.7134	0.5911	0.5205

From table 3 above, it was deduced that Home solution has been the most contributing panaceas among the four panaceas followed by Oral rehydration, Antibiotics pills/syrups and Zinc.

CONCLUSION

It was found that Gaussian Naïve Bayes is appropriate for prior distribution for Bayesian logistic regression when we are not sure of the two forms of priors (informative and non-informative). The parameters for Bayesian approach using GNB were smaller than that of classical logistic regression. Also, the standard errors for the parameters for Bayesian logistic regression using GNB were lower than the standard errors of the classical approach using MLE. This makes the Bayesian logistic regression using GNB to be more preferred.

REFERENCES

Andrea Discacciati and Nicola Orsini, (2014). Approximate Bayesian logistic Regression via Penalized Likelihood Estimation with Data Augmentation. Unit of Biostatistics and Unit of Nutritional Epidemiology Institute of Environmental Medicine, Karolinska.1-24

Andrew Gelman (2002). Prior Distribution.*Encyclopedia of Environmetrics*. Vol.3, 1634- 1637. ISBN 0471899976.

Andrew Gelman. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *International Society for Bayesian Analysis.* Vol. 1(3), 515-533

Brian Kulis (2012). Lecture note on Bayesian Logistic Regression. CSE 788.04: Topic in Machine Learning.

Chia-chung Tsai, (2004). Bayesian Inference in Binomial Logistic Regression: A Case Study of the 2002 Taipei Mayoral Election. *Taiwan Political Science.*103-123.

Dana L. Kelly, Robert W. Youngblood, and Kurt G. Vedros (2010): Minimally Informative Prior Distributions for PSA. *Idaho National* *laboratory, Idaho falls,* ID USA.1-9

Greenland, S. (2006). Bayesian Perspectives for Epidemiological Research: Foundations and Basic Methods. *International Journal of Epidemiology,*Vol.3(5), 765-775.

Hee Min Choi and James P. Hobert, (2013). The Polya-Gamma Gibbs Sampler for Bayesian Logistic Regression is Uniformly Ergodic. *AMS 2000 subject classifications*. Primary 60J27; Secondary 62F15.1-13

Mark E. Glickman and David A. van Dyk (2000). Basic Bayesian Methods, *Method in Molecular Biology*, Vol.404: 1-22

Mu Zhu (2004).The Counter-intuitive Non-informative Prior for the Bernoulli Family. *Journal of Statistics Education.* Vol.12, No.2, 1-10.

Rosebella B. Montes and Eduardo L. Cruz (2014): Bayesian Logistic Regression Analysis of the Association of Intimate Partner Violence and Modern Contraceptive Use in the Philippines. *Asian Journal of Social Sciences & Humanities* Vol. 3(2) ISSN: 2186-8492

Robert E. Kass and Larry Wasserman (1996). The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association.* Vol.91, No.435,1-28

Tom M. Mitchell: Machine Learning 10-701:Naïve Bayes and Logistic Regression. Center for Automated Learning and Discovery Carnegie Mellon University September 27, 2005.

Ulrich K. Muller (2012). Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models. *Journal* *of Monetary Economics.* Vol. 9, 581-597

Wioletta Grzenda, (2015).The Advantages of Bayesian Methods over Classical Methods in the Context of Credible Intervals. *Information Systems in Management.* Vol. 4 (1), 53-63.

Zeng.X.L. and Zaslavsky A.M. (2002). Single Observation Unbiased Priors. *Annals of Statistics*, Vol.3, 1345-1375.

BIOGRAPHY

J. F. Ojo

Lectures at the Department of Statistics, University of Ibadan, Nigeria. His research interest encompasses statistical inference, statistical computing and time series analysis. Dr. Ojo has to his credit a number of refereed journal articles and conference papers within and outside Nigeria. He is a member of Nigerian Statistical Association (NSA) and Science Association of Nigeria (SAN). He is a fellow of Royal Statistical Society, London.

R. O. Olanrewaju

OLANREWAJU Rasaki Olawale is a Course Facilitator for the Distance Learning Centre (DLC), University of Ibadan on Statistical courses. He holds a Professional Diploma in Statistics (P.D.S) (Distinction classification); Bachelor of Science (First Class honor) and Master of Science (Proceed to Ph.D. classification) both in Statistics from University of Ibadan. He is a member of the Nigeria Statistical Society (NSS). To his credit are some reputable articles and conference papers.

Folorunso, Serifat Adedamola is a Teaching Assistant in the Department of Statistics as well as a course facilitator under the University distance learning program. She holds diploma certificate in Statistics from Federal School of Statistics, Ajibode Ibadan in 2004 with a Distinction, a B.Sc. (Statistics) from University of Agriculture, Abeokuta , a M.Sc (Statistics) from University of Ibadan and currently a Ph.D student in the Department of Statistics, University of Ibadan. She is a member of the Nigeria Statistical Society (NSS). To his credit are some reputable articles and conference papers.

Reference to this paper should be made as J. F. Ojo, R. O. Olanrewaju, & S.A. Folorunsho (2017), Bayesian Logistic Regression using Gaussian Naïve Bayes. *J. of Medical and Applied Biosciences,* Vol. 9, No. 2, Pp. 52-69