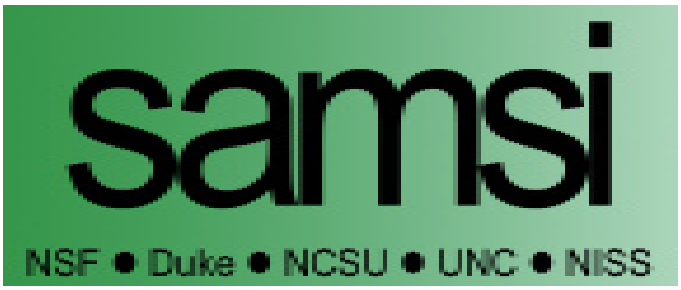


On Bayesian Analysis of Generalized Linear Models: A New Perspective

Sourish Das* and Dipak K.Dey

Technical Report #2007-8
October 17, 2007

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

On Bayesian Analysis of Generalized Linear Models: A New Perspective

Sourish Das*and Dipak K.Dey

Abstract

We present a new Bayesian approach for analyzing Generalized Linear Models (GLM). We elicit prior on canonical parameters, since it is much easier for expert, instead of parameters of interest, that is regression parameters. Under the full rank assumption of design matrix (X), we show that the various elicited prior on canonical parameters induce a class of priors on regression parameters, which includes many commonly used priors. We further show that posterior mode is the Bayes estimator under a particular loss function. This new method helps us to find the closed form estimator for the regression parameters of a GLM. We further broaden the horizon of simple diagnostic techniques of linear model, like Cook's distance under GLM frame work. We found that this new Generalized Cook's distance to be invariant to

*Sourish Das is a PhD student (email: sourish.das@uconn.edu) and Dipak K. Dey (email: dipak.dey@uconn.edu) is Professor and head of Department of Statistics at University of Connecticut at Storrs, CT. 06269 - 4120. This work is partially completed while both the authors are visiting SAMSI, Research Triangle Park, NC-27709. The authors sincerely thank Dr. David Dunson of NIEHS, for his comments, which results much improved version of the paper.

the choice of prior. We also develop various other model diagnostic methods. We present the performance of this new method on two different real data sets. First we apply the method on popular data set of Finney (1947); we considered this data set as test bed for our new method. Then we present the performance of the method for second data set. In order to investigate the efficacy of a new treatment a very current investigation on Osteoarthritis of knee was performed during the summer of 2006 and pain score was collected after 90 days of treatment. Bio-marker information on TNFα cytokines was also recorded as covariates information.

KEY WORDS: Diagnostics; Generalized Cook's Distance; Outliers; Prior Elicitation.

1 INTRODUCTION

Generalized Linear Models (GLM) have been used for modeling various kinds of data consisting of exponential family response with covariates. Typical examples include those for binomial and Poisson response data. A regression model determines the structure of the covariate information, where a link function specify the relationship between the regression model and expected values of the observation. However, simple linear model with normal error structure is a special case of GLM.

Usual Bayesian analysis of such data requires specification of prior for the regression parameters used in the model. Uniform priors, Jeffrey's prior or diffuse priors are used very commonly as conventional noninformative priors. However, eliciting prior directly on regression parameters are really hard. But often it is much easier to elicit prior on canonical parameters. In this paper we show that under full rank assumption of

design matrix, the elicited prior on canonical parameters will induce a proper prior on the regression parameters. Then one can carry the usual Bayesian analysis based on the induced prior.

Identifying atypical observation for discordancy in the statistical analysis is an important practical problem. The inference from any model are based on the statistical inductive jump from the observed data. The validity of that inference would be expected to be poor, if the parameter estimates of the model are dependent on few irregular observations, resulting from isolated events. In addition, when these parameters have scientific interpretations, presence of unidentified influential observations can mislead the scientist from the other sciences. Moreover, endeavor to reveal these observations provides wisdom about the robustness of the fitted model to the given data set. In this paper, we present straight forward diagnostic technique for GLM.

The format of the paper is as follows. In section 2, we present an overview of GLM and discuss the new method of fitting GLM by eliciting prior on canonical parameters. In section 3, we introduce the residuals of GLM. We also discuss the concept of generalized Cook's distance to detect the influential observations and some new model diagnostic technique. In section 4, we discuss the model selection criteria for GLM. In section 5, we presented the application of the methodology for two different data sets and section 6, concludes the paper with brief discussions.

2 GENERALIZED LINEAR MODELS.

The class of GLM, introduced by Nelder and Wedderburn (1972), enjoyed a great deal of interest from statistical researchers and practitioners. GLM is defined by assuming that $y_1, y_2, ...y_n$ are independent observations, where y_i has the density from the natural exponential family

$$f(y_i \mid \theta_i) = exp\big\{(\theta_i y_i - \psi(\theta_i)) + c(y_i)\big\}, \tag{2.1}$$

where $i = 1, 2, ..., n$. The density in (2.1) is parameterized by the canonical parameter θ_i. The $\psi(\cdot)$, and $c(\cdot)$ are known functions and the θ_i's are related to the regression coefficients by the link function

$$\theta_i = g(\eta_i), \tag{2.2}$$

where $i = 1, 2, ..., n$; and

$$\eta_i = x_i'\beta \tag{2.3}$$

is the systematic component of GLM. In (2.3) $x_i' = (x_{i1}, x_{i2}, ..., x_{ip})$ is a $1 \times p$ vector denoting the i^{th} row of $n \times p$ matrix of covariates X, $\beta = (\beta_1, ..., \beta_p)'$ is a p vector of regression coefficient, and $g(\cdot)$ is monotonic differentiable function. The model given by (2.1), (2.2) and (2.3) is called GLM. The Normal, logistic, binomial and Poisson regression models are special cases of the GLM; see McCaullagh and Nelder (1989) for more details.

It follows from Diaconis and Ylvisker (1979) that the conjugate prior distribution for θ_i is

$$\pi(\theta_i) = Kexp\{m\mu_0\theta_i - m\psi(\theta_i)\}, \tag{2.4}$$

where the normalizing constant $K = K(m, \mu_0)$ is chosen such that $\pi(\theta_i)$ is a proper density function. Here $\pi(\theta_i)$ is a two parameter natural exponential family of densities for θ_i with

the natural parameters m and μ_0. It follows immediately from (2.4) that the posterior distribution of θ_i is

$$\pi(\theta_i \mid y_i) = K\Big(\mu_0 + 1, \frac{y_i + m\mu_0}{1+m}\Big) exp\{(y_i + m\mu_0)\theta_i - (1+m)\psi(\theta_i)\}. \tag{2.5}$$

Since link function in (2.2) can be considered as a simple monotonic transformation of θ_i, then it follows from Das and Dey (2006) that the Jacobian of transformation from θ_i to η_i is $J\Big(\frac{\theta_i}{\eta_i}\Big) = g'(\eta_i)$. Hence we have the posterior density for η_i as

$$\pi_1(\eta_i \mid y_i) = K\Big(\mu_0 + 1, \frac{y_i + m\mu_0}{1+m}\Big) exp\{(y_i + m\mu_0)g(\eta_i) - (1+m)\psi(g(\eta_i))\}g'(\eta_i), \tag{2.6}$$

where $K\Big(\mu_0 + 1, \frac{y_i+m\mu_0}{1+m}\Big)$ is the normalizing constant. Since the posterior distribution of η_i is given by (2.6) and $g(\cdot)$ is monotonic twice differentiable function and $g'(\eta_i) \neq 0$, then a little algebra shows that solution of the following equation

$$\psi'(g(\eta_i)) = \frac{y_i + m\mu_0}{1+m} + \frac{g''(\eta_i)}{[g'(\eta_i)]^2} \cdot \frac{1}{(1+m)} \tag{2.7}$$

is the posterior mode of η_i, which we denote as $\hat{\eta}_i = h(y_i, m, \mu_0)$. Note that $h(\cdot)$ is a function of the data y_i and hyper parameters m and μ_0. From now onwards, for simplicity, we denote the posterior mode of the response as $\hat{\eta}_i = h(y_i)$. In vector notation, we write the same as $\hat{\eta} = h(\mathbf{y})$. The n-dimensional vector $h(\mathbf{y})$ can be uniquely decomposed as

$$h(\mathbf{y}) = \mathbf{P}h(\mathbf{y}) + (\mathbf{I} - \mathbf{P})h(\mathbf{y}), \tag{2.8}$$

where $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'}$ is of the full rank. Note that $\mathbf{P}$ is the linear transformation matrix representing the orthogonal projection from n-dimensional space $\mathcal{R}^n$ onto estimation space $\mathcal{C}(\mathbf{X})$, while $(\mathbf{I} - \mathbf{P})$ represents the orthogonal projection of $\mathcal{R}^n$ onto the error space $\mathcal{C}^{\perp}(\mathbf{X})$. Therefore,

$$\mathbf{P}h(\mathbf{y}) = X(X'X)^{-1}X'h(\mathbf{y}) = X\hat{\beta},$$

where, $\hat{\beta} = (X'X)^{-1}X'h(\mathbf{y})$ is the unique least-square estimator of β. Though, Das and Dey (2006) proposed this estimator for the regression parameters of GLM, they did not discuss the least-square property of the estimator. Also note that from decision theoretic point of view one can argue that η_i is a generalized maximum likelihood estimator (GMLE). Using the invariance property of GMLE, $\hat{\beta} = (X'X)^{-1}X'\hat{\eta}$ is also GMLE of the regression parameter β. Hald (1999) discussed, why Fisher argued in favor of maximum likelihood estimator, especially in the context of its invariance property and criticize the posterior mode strongly because of its lack of invariance and unbiasedness property. Note that Griffith, Hill and Pope (1987) found, for the nonlinear models, the MLE to have significant bias for small sample. However for large sample both the MLE and posterior modes are asymptotically equivalent and consistent estimators.

Since for the small sample both posterior mode and MLE have bias for nonlinear models, we would take this opportunity to explore the optimal property of posterior mode. The following lemma shows that under particular loss the posterior mode is Bayes estimator.

Lemma 1. *Suppose the posterior expected loss of an action a, when the posterior distribution is* $\pi(\theta \mid x)$, *is*

$$\rho(\theta, a) = \int_{\Theta} L(\theta, a) dF^{\pi(\theta|x)}(\theta),$$

where the loss function is $L(\theta, a) = log\left(\frac{\pi(a|x)}{\pi(\theta|x)}\right)$. *Then the Bayes rule is*

$$\delta^{\pi}(x) = \underset{\theta \in\ \Theta}{\operatorname{argmax}}\, \pi(\theta \mid x),$$

which is the mode of posterior distribution of $\pi(\theta \mid x)$.

Proof. The value of a which minimizes the $\rho(\theta, a)$ can be found by differentiating it

with respect to a, and setting it equal to zero. Assuming all integrals are finite, the result is

$$\begin{aligned} 0 &= \frac{\partial}{\partial a}\left[\int_{\Theta} log\left(\frac{\pi(a \mid x)}{\pi(\theta \mid x)}\right) dF^{\pi(\theta|x)}(\theta)\right] \\ &= \frac{\partial}{\partial a}\left[\int_{\Theta} \left\{log\pi(a \mid x) - log\pi(\theta \mid x)\right\} dF^{\pi(\theta|x)}(\theta)\right] \\ &= \frac{\partial}{\partial a}\ log\pi(a \mid x). \end{aligned}$$

Hence solving for a gives the required results. □

Please note that the posterior expected loss in Lemma 1, $\rho(\theta, a) = E\left[log\left(\frac{\pi(a|x)}{\pi(\theta|x)}\right)\right]$, can be interpreted as Kullback-Leiblar divergence of the posterior distribution evaluated under action a from the true posterior distribution of unknown parameter θ. Hence posterior expected loss or Kullback-Leiblar divergence is minimum, if we choose our action as posterior mode.

Issues regarding the prior elicitation: Now we consider one very important issue about this new procedure. According to this method we elicit prior on the canonical parameters θ which then induce prior on η; now the question is: Does the prior on η (which is of n-dimension) induces prior on β ? Note that β has $p(< n)$-dimension. The answer to this question is yes and we provide a formal proof in the following lemma.

Lemma 2. Suppose $\pi(\eta)$ is a prior distribution on η, where $\eta = (\eta_1, ..., \eta_n)'$. If we consider $\beta = (X'X)^{-1}X'\eta$, where β is $p-$dimensional vector. Then under the full-rank assumption of X, $\pi(.)$ induces prior on β.

Proof. We consider $\beta = A\eta$ (where $A = (X'X)^{-1}X'$) as a p-dimensional linear transformation of η. We define another set of linear transformation as $\xi = C\eta$ of $(n -$

$p)-$dimension. The joint density of β and ξ is

$$\pi(\beta, \xi) = \pi(\eta) \times J\Big(\frac{\eta}{\beta, \xi}\Big), \tag{2.9}$$

where $J\Big(\frac{\eta}{\beta,\ \xi}\Big)$ is the Jacobian of transformation. Now we integrate out ξ from (2.9) to get,

$$\pi(\beta) = \int_{\Xi} \pi(\beta, \xi) d\xi.$$

Hence $\pi(\beta)$ is the prior distribution on β, which is being induced by the prior distribution of η. $\square$

Immediately this new result, motivates us to investigate what prior would be induced on the regression parameter (at least for basic intercept model), if we elicit prior on canonical parameters. Consider simple logistic regression model with intercept parameter only. We can show that if we elicit Beta distribution as prior on the p, that is $p \sim Beta(a, b)$, then $log\Big(\frac{p}{1-p}\Big) = \beta_0 \sim Skew - Logistic(a, b)$, with the density as $\pi(\beta_0) = \frac{1}{B(a,b)} \frac{e^{a\beta_0}}{(1+e^{\beta_0})^{a+b}}$, where the range of the intercept parameter β_0 is real line $\mathbb{R}$. Hence if we assume the prior distribution over p is uniform distribution *i.e.*, $U(0, 1)$, then the induced prior on β_0 follows standard logistic distribution with location parameter 0 and scale parameter 1. Note that if we elicit symmetric Beta distribution $Beta(a, a)$ as prior on p, then the induced prior on β_0 is $Symmetric - Logistic(a)$ with density as $\pi(\beta_0) = \frac{1}{B(a,a)} \frac{e^{a\beta_0}}{(1+e^{\beta_0})^{2a}},\ -\infty < \beta_0 < \infty$. Hence if we elicit Jeffrey's prior on p, which is $Beta(0.5, 0.5)$ then the induced prior on β_0 is $Symmetric - Logistic(0.5)$.

It is common in Bayesian analysis, to assign normal prior on the regression parameters with a very large variance, so that the prior distribution will be as much non-informative as possible. The following result provides a prior distribution on p which is induced by normal distribution on the regression parameters.

Result 1. *For logistic regression if we elicit the normal distribution on the regression parameters* β *as* $N(\beta_0, \sigma^2(X'X)^{-1})$, *then that would induce logistic-normal distribution on* p_i, *independent of* i, *with location parameter* $x_i'\beta_0$ *and scale parameter* σ^2 *uniquely and vice-versa. Note that* $P(y_i = 1) = p_i$ *and* $P(y_i = 0) = 1 - p_i \ \forall i = 1(1)n$.

Proof : We know $\beta = (X'X)^{-1}X'\eta = A\eta$, where $A = (X'X)^{-1}X'$ such that (i) $AA' = (X'X)^{-1}$ and (ii) $AX = I$. Hence $A\eta \sim N(AX\beta_0, \sigma^2 AA')$ which implies $\eta \sim N(X\beta_0, \sigma^2 I)$. For simplicity $\eta_i \sim N(\mu_i, \sigma^2)$, where $\mu_i = x_i'\beta_0$ and σ^2 is generally known and large. Now consider the link $\eta_i = log\Big(\frac{p_i}{1-p_i}\Big)$, where the Jacobian of transformation is $J\Big(\frac{\eta_i}{p_i}\Big) = \frac{1}{p_i(1-p_i)}$. Hence the prior distribution on p_i is logistic-normal distribution with density

$$\pi(p_i) = \frac{exp\{-\frac{1}{2\sigma^2}\Big(log\big(\frac{p_i}{1-p_i}\big) - \mu_i\Big)^2\}}{\sigma\sqrt{2\pi}p_i(1-p_i)}, \qquad 0 < p_i < 1 \quad \forall\ i = 1(1)n,$$

with location parameter $x_i'\beta_0$ and scale parameters σ^2. $\square$

The result discussed above motivates us to see the behavior of induced prior on p when $\sigma^2 \to \infty$. We can write the density $\pi(p_i)$ as

$$\pi(p_i)p_i(1-p_i) = \phi\Bigg(\frac{log\big(\frac{p_i}{1-p_i}\big) - x_i'\beta_0}{\sigma}\Bigg),$$

where $log\big(\frac{p_i}{1-p_i}\big) \sim N(x_i'\beta_0, \sigma^2)$. By Taylor's series expansion we can show $\phi(x) = \phi(0) + x\phi'(\xi)$, where $\xi \in (0, x)$; or $\xi \in (-x, 0)$ for $x > 0$. As $\sigma^2 \to \infty$, $x\phi'(\xi) \to 0$. Therefore, $Lim_{\sigma^2 \to \infty}\pi(p_i) = cp_i^{-1}(1-p_i)^{-1}$, where c is $(\sqrt{2\pi})^{-1}$. Hence the induced prior for p_i is proportional to that of Novic and Hall's prior (1965), *i.e.*, $\pi(p_i) \propto p_i^{-1}(1-p_i)^{-1}$. We summarize this in the following result.

Result 2. If we elicit a diffuse prior on β as $\pi(\beta) = \Big\{N(\beta_0, \sigma^2(X'X)^{-1}) \mid Lim_{\sigma^2 \to \infty}\pi(\beta) = \pi(\beta) \propto 1\Big\}$,then that would induce Novic and Hall's (1965) non-informative improper prior $\pi(p_i) \propto p_i^{-1}(1-p_i)^{-1}$ on p_i.

Developing analogous proof, we get the following results for Poisson regression model.

Result 3. *For Poisson regression model with log-link, if we elicit the normal distribution on the regression parameters* β *as* $N(\beta_0, \sigma^2(X'X)^{-1})$, *then that would induce log-normal distribution on* λ_i, *independent of* i, *with location parameter* $x_i'\beta_0$ *and scale parameter* σ^2 *uniquely and vice-versa. Note that* $log(\lambda_i) = x_i'\beta$ *and* $y_i \sim Poisson(\lambda_i)\ \forall i = 1(1)n$.

Result 4. If we elicit a diffuse prior on β as $\pi(\beta) = \Big\{N(\beta_0, \sigma^2(X'X)^{-1}) \mid Lim_{\sigma^2 \to \infty}\pi(\beta) = \pi(\beta) \propto 1\Big\}$, then that would induce non-informative improper prior $\pi(\lambda_i) \propto \frac{1}{\lambda_i}$ on λ_i, which is non-informative improper prior for scale parameters.

Result 5. If we elicit $Gamma(a, b)$ on the λ_i parameter on the intercept only Poisson regression model then the induced prior on the intercept parameter β_0 follows *Log-Gamma*(a, b) with density as $\pi(\beta_0) = \frac{exp\{a\beta_0 - exp\{\beta_0/b\}\}}{\Gamma(a)b^a}, \quad -\infty < \beta_0 < \infty$.

We summarize these result in the Table 1. In order to find the standard error of the estimator, Das and Dey (2006) used Bayesian CLT to find credible sets for the regression parameters. However, in our method we already showed that elicited prior on canonical parameters would induce prior on regression parameters. Analogously we can say that the posterior on canonical parameters would induce posterior distribution on the regression parameters. Since we are considering conjugate prior for canonical parameters, it is very easy to sample from the posterior. Hence we can sample β using Monte-Carlo technique following three steps.

Step 1. Suppose we are at the r^{th} iteration. Generate sample $\theta_i^{(r)}$ from $\pi(\theta_i \mid y_i)$ in (2.5) for $i = 1, ..., n$.

Step 2. Calculate $\eta_i^{(r)} = g(\theta_i^{(r)})$ for $i = 1, ..., n$ or $\eta^{(r)} = (\eta_1^{(r)}, \eta_2^{(r)}, ..., \eta_n^{(r)})'$.

Step 3. Calculate $\beta^{(r)} = (X'X)^{-1}X'\eta^{(r)}$, $r = 1, 2, ..., N$; where N is the simulation size.

Once we have the posterior samples of β, $\{\beta^{(r)} \mid r = 1, ..., N\}$; we can calculate the credible set and all other summary statistics for β. Note that here the samples $\{\beta^{(r)} \mid r = 1, ..., N\}$ are not dependent like MCMC samples.

Note that since we are considering conjugate prior for the natural exponential family, we can generate θ_i from (2.5) very easily. However one might try some other family where the posterior distribution of θ_i does not belong to known distribution. In that case one might use Metropolis-Hasting to generate sample from the posterior of θ_i, which is beyond the scope of this paper. Now we discuss some examples for the better understanding of the GLM.

2.1 Normal model with identity link.

In this subsection, we present the performance of the method for normal model, since normal distribution can be used as test bed. For the normal model, we assume, $y_1, ..., y_n$ are n independent observations from $N(\mu_i, \sigma^2)$ with identity link as $\mu_i = \eta_i = x_i'\beta$. We assume noninformative improper prior for the μ_i as $\pi(\mu_i) = 1$ for all i. Therefore the posterior distribution of μ_i given y_i is $N(y_i, \sigma^2)$. Hence, the posterior distribution of η_i given y_i is $N(y_i, \sigma^2)$ which implies the posterior mode for η_i is $h(y_i) = y_i$. Inevitably our new method gives us simple ordinary least square estimator $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$ for the regression parameter β which is expected.

Next we discuss the Log-Normal Regression model with log link. Though the Bayesian analysis of Log-Normal regression model was introduced by Zellner (1971); this new Bayesian method for GLM set-up encourages us to revisit the model once again from

a new perspective.

2.2 Log-Normal Regression Model

Suppose $Y_i \sim Log\ Normal(\nu_i, \sigma^2)$, $i = 1, 2, ..., n$. Therefore the likelihood of Y_i is

$$f(y_i) = \frac{1}{y_i\sqrt{2\pi\sigma^2}} exp\Big\{ -\frac{1}{2\sigma^2}(logy_i - \nu_i)^2 \Big\}, \qquad 0 < y_i < \infty,$$

with $\mu_i = e^{\nu_i}$ as the median and the link function as $\eta_i = log\mu_i$, where $i = 1, 2, ..., n$. The systematic component is $\eta_i = x_i'\beta$. We note that the unknown parameter σ^2 is actually a nuisance parameter, since our primary interest of parameters are the regression coefficients in the systematic component part.

Now for some constant $\tau > 0$, we assume the conditional prior distribution of $[\mu_i \mid \sigma^2] \sim$ *Log-Normal*$(\mu, \tau\sigma^2)$, and the marginal prior distribution of $[\sigma^2] \sim$ *Inv-Gamma*(a, b). It follows that the conditional prior distribution of $[\nu_i \mid \sigma^2] \sim N(\mu, \tau\sigma^2)$. Thus the conditional posterior distribution of $[\nu_i \mid \sigma^2, y_i]$ is

$$\pi(\nu_i \mid \sigma^2, y_i) \sim N\bigg(\frac{\mu + \tau logy_i}{\tau + 1}, \frac{\tau}{\tau + 1}\sigma^2\bigg),$$

and the marginal posterior distribution of $[\sigma^2 \mid y_i]$

$$\pi(\sigma^2 \mid y_i) \sim Inv\ Gamma\big(a_1, b_1\big),$$

where $a_1 = a + \frac{1}{2}$ and $b_1 = b + \frac{(logy_i - \mu)^2}{2(\tau+1)}$. Therefore we can easily show that the marginal posterior distribution of $[\nu_i \mid y_i]$ follows a Student's t-distribution with $2a$ degrees of freedom where location parameter is $\mu_1 = \frac{\mu + \tau logy_i}{\tau+1}$ and scale parameter is $\frac{\tau}{\tau+1}\sigma_1^2$ with $2a$ degrees of freedom. Hence the marginal posterior distribution of η_i given y_i is

$$\pi(\eta_i \mid y_i) \sim t_{2a}\bigg(\eta_i \mid \mu_1, \frac{\tau}{\tau + 1}\sigma_1^2\bigg),$$

where $\mu_1 = \frac{\mu+\tau log y_i}{\tau+1}$. Note that $\sigma_1^2 = \frac{b_1}{a_1}$, where $Var(\eta_i \mid y_i) = \frac{a}{a-2}\frac{\tau}{\tau+1}\frac{b_1}{a_1}$. It follows immediately that the posterior mode of η_i is

$$\hat{\eta}_i = h(y_i) = \frac{\mu + \tau\ log(y_i)}{\tau + 1}, \qquad \forall\ i = 1(1)\ n.$$

Hence the required estimator for β is

$$\hat{\beta} = (X'X)^{-1}X'\hat{\eta} = (X'X)^{-1}X'\left(\frac{\mu + \tau\ log(\mathbf{y})}{\tau + 1}\right),$$

where $\mathbf{y} = (y_1, y_2, ..., y_n)'$ and the logarithm is defined component wise.

Next we discuss the binomial and Poisson regression models.

2.3 Binomial and Poisson response model.

The binomial response with the logit link is also known as logistic regression model. In this model, we assume, $y_1, ..., y_n$ are n independent observations from $Binomial(m_i, p_i)$, where $log\left(\frac{p_i}{1-p_i}\right) = \theta_i = \eta_i = x_i'\beta$, $i = 1, ..., n$. Considering the conjugate prior for p_i as $Beta(a, b)$, the solution of the equation (2.7) gives the posterior mode of η_i. For our case, $g(\eta_i) = \eta_i$, and thus $g'(\eta_i) = 1$ and $g''(\eta_i) = 0$ and thus we have $\frac{m_i e^{\eta_i}}{1+e^{\eta_i}} = \frac{(y_i+a)m_i}{(m_i+a+b)}$. Hence the posterior mode of η_i is $\hat{\eta}_i = log\left(\frac{y_i+a}{m_i+b-y_i}\right)$ and the required estimator for β is $\hat{\beta} = (X'X)^{-1}X'log\left(\frac{\mathbf{y}+a}{b+m-\mathbf{y}}\right)$, where $\mathbf{y} = (y_1, y_2, ..., y_n)'$ and the logarithm is defined component wise. Note that if the response variable has the Bernoulli distribution, then m_i is equal to unity for all i.

Following Das and Dey (2006), we can show the estimator of β for the Poisson response model with log-link is $\hat{\beta} = (X'X)^{-1}X'log\left(\frac{\mathbf{y}+a}{1+\frac{1}{b}}\right)$, where $\mathbf{y} = (y_1, y_2, ..., y_n)'$ and the logarithm is defined component wise. Note that we choose conjugate prior $Gamma(a, b)$ for the rate parameters λ_i, where $Y_i \sim Poisson(\lambda_i), \quad i = 1, ..., n$.

In the next section, we discuss the important issues of appropriate residuals to be used for model diagnostic purpose and different diagnostic techniques of GLM.

3 DETECTING ATYPICAL OBSERVATIONS IN GLM.

Before we draw the inference, appropriate diagnostic method of detecting atypical observations, like outliers and influential points, is the natural succeeding step to critically asses the fit of GLM to the data. Any data from large studies often suffers from "atypical" observations - "atypical" from the point of view of outlying responses (y), and that of extreme points in the design space ($\mathbf{X}$). Now identifying the influential observation is different from the study of outliers. Because, the fact that an observation is an outlier does not necessarily imply that this observation substantially affects the parameter estimates of the assumed model. However, in general it will affect the variance of the estimates. In this section we present a new diagnostic method of identifying the observations which are not in agreement with the fitted GLM.

The classical literature on the model diagnostics of linear model is rich, and the pertinent references are Andrew and Pregibon (1978), Belsley et al., (1980), Draper and John (1981) and Cook and Weisberg (1982). Pregibon (1981) extended that study to logistic regression model. The main idea of this approach is to delete dubious observation and measure the change in the important features of the model, such as estimated parameter values, because of this deletion. In the context of simple linear model, Chaloner and Brant (1988) developed influence diagnostics from a Bayesian viewpoint. In order

to define the latent data residual for binary response regression model, Albert and Chib (1995) used the notion of tolerance random variable. A different divergence measure approach based on MCMC technique for detecting influential observations from Bayesian paradigm was developed by Peng and Dey (1995), Weiss (1996) for GLM. Recently Miller and Stewart (2007), showed that posterior variance of the log-likelihood from particular observation is a measure of that observation's local influence. They obtained this result by considering the Kullback-Leibler divergence between baseline and case-weight perturbed posteriors, with local influence being the curvature of the divergence evaluated at baseline posterior. There are two distinct Bayesian approaches for the determination of outliers and influential observations. The first is a formal Bayesian method based on posterior distributions. The second approach to assessing influence is with regards to predictive distribution, where the influence of an observation is studied by noting the changes in Bayes, pseudo-Bayes and posterior Bayes factors. The pertinent references are Pettit and Young (1990) and Gelfand, Dey and Chang (1992). Recently, Marshall and Spiegelhalter (2007) developed a simulation based approach for identifying outliers in Bayesian hierarchical models.

It is important to discuss the issues of appropriate residuals to be considered for the diagnostic purpose. Residuals for simple linear model is well defined and very straight forward. The concepts of residuals can be defined and illustrated at different levels of GLM. We can define residuals at systematic component level and simultaneously we can define it at random component level. Equation (2.8) motivates us to decompose the n-dimension vector η as

$$\eta = P\eta + (I - P)\eta, \tag{3.10}$$

where $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T = X\beta$. Now using Bayesian CLT, the posterior distribution of $P\eta$ is approximately $N(P\hat{\eta}, PI(\hat{\eta})^{-1}P^T)$ and the posterior distribution of $(I-P)\eta$ is approximately $N((I-P)\hat{\eta}, (I-P)I(\hat{\eta})^{-1}(I-P)^T)$. Note that $I(\hat{\eta})^{-1}$ is the inverse of observed generalized Fisher's Information. Therefore, we define the residuals as $\epsilon_i = (1-p_{ii})\eta_i$ and hence the posterior mean of ϵ_i is the observed residual $\hat{\epsilon}_i = (1-p_{ii})h(y_i)$. Here we define the residual at systematic component. Similarly, the residuals can also be defined at random component level. The popular form of this type of residual is $\chi = (y_i - \hat{\mu}_i)/\{V(\hat{\mu}_i)\}^{\frac{1}{2}}$, where $V(\mu)$ is the function relating the variance to the mean of y and $\hat{\mu}_i$ is the generalized maximum likelihood estimate (or posterior mode) of the i^{th} mean as fitted to the regression model; see McCullough and Nelder (1989).

3.1 Detecting Outliers using Highest Residual Density Interval

Generally it is to be expected that the posterior distribution of residuals at both random component and systematic levels of the GLM have some approximate normal distributions with means $E(\hat{\epsilon}) = 0$ and $E(\hat{\epsilon}\hat{\epsilon}') = \sigma^2(\mathbf{I}-\mathbf{P})$. Now we define $\hat{R}(\hat{\pi}_\alpha) = \{\hat{\epsilon}_i : \pi(\hat{\epsilon}_i) \geq \pi_\alpha;\ \forall\ i = 1, ..., n\}$, where $\hat{R}(\hat{\pi}_\alpha)$ is the highest marginal density set of the Bayesian residuals at systematic levels. If the residual $\hat{\epsilon}_i \not\subseteq \hat{R}(\hat{\pi}_\alpha)$, then we detect it as an outlier. Chen, Shao and Ibrahim (2001) discussed the algorithm for computation of $100 \times (1-\alpha)\%$ HPD interval if we just sample the (generally MCMC sample) from the posterior distribution. Now we can use the same algorithm to compute the set $\hat{R}(\hat{\pi}_\alpha)$.

Next we discuss some diagnostic tools for GLM and discuss the issues related to them. In GLM, the influence of case i on the estimation of β can be measured by the distance which calibrates $\hat{\beta} - \hat{\beta}_{(i)}$ by comparison to the confidence contour for β. This simple idea

motivates us to the concept of *Influence Functions*. This concept is well known for simple linear model.

3.2 Detecting Influential Observations

A class of measures of influence of i^{th} observation is based on the idea of the influence function introduced by Hampel (1974). Following the motivation from Chatterjee and Hadi (1986) the generalized sample influence curve ($GSIC$) can be defined as

$$\begin{aligned} GSIC_i &= (n-1)(X^TX)^{-1}x_i^T(h(y_i) - x_i\hat{\beta}_{(i)}) \\ &= (n-1)(X^TX)^{-1}x_i^T\frac{\hat{\epsilon}_i}{1-p_{ii}}, \end{aligned} \tag{3.11}$$

where $\hat{\beta}_{(i)} = (X_{(i)}^TX_{(i)})^{-1}X_{(i)}^Th(\mathbf{y}_{(i)})$ is the estimate of β when the i^{th} observation is omitted. Also note that $\hat{\eta} = h(\mathbf{y})$ is the posterior mode of the systematic component η, $\tilde{\eta} = Ph(\mathbf{y}) = P\hat{\eta} = X\hat{\beta}$ is the estimated systematic component at design space and ϵ-residual is defined in the previous section as $\hat{\epsilon} = \hat{\eta} - \tilde{\eta} = (I-P)\hat{\eta}$.

The generalized sensitivity curve (GSC) is obtained as $GSC_i = n(X^TX)^{-1}x_i^T\frac{\hat{\epsilon}_i}{1-p_{ii}}$. Now one can easily show that

$$IF_i = \hat{\beta} - \hat{\beta}_{(i)} = (X^TX)^{-1}x_i^T\frac{\hat{\epsilon}_i}{1-p_{ii}}, \qquad i = 1, ..., n. \tag{3.12}$$

Clearly, $GSIC_i$ and GSC_i are equivalent and can be interpreted as proportional to the distance between $\hat{\beta}$ and $\hat{\beta}_{(i)}$. Since IF_i is a vector, we must normalize it so that observation can be ordered. The class of norms which are location and/or scale invariant is given by

$$D_i(M;c) = \frac{(IF_i)^TM(IF_i)}{c} \tag{3.13}$$

for appropriate choice of M and c. A large value of $D_i(M;c)$ presents the evidence that the i^{th} observation has strong influence on the estimated coefficients relative to M and c.

3.3 Generalized Cook's Distance

If we use the generalized sample influence curve ($GSIC$) to approximate the influence function, then from (3.13) we obtain the Generalized Cook's distance as

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (X^T X)(\hat{\beta} - \hat{\beta}_{(i)})}{p\hat{\sigma}^2}, \tag{3.14}$$

where $\hat{\sigma}^2$ is defined in following way. Equation (2.8) motivates us the orthogonal normalized sum of squares decomposition as

$$\hat{\eta}^T\hat{\eta} = \hat{\eta}^T P\hat{\eta} + \hat{\eta}^T(I-P)\hat{\eta}, \tag{3.15}$$

where we define these sum of squares as: $GSST = \hat{\eta}^T\hat{\eta} = h(\mathbf{y})^T h(\mathbf{y})$ as the generalized total sum of squares, $GSSR = \hat{\eta}^T P\hat{\eta} = h(\mathbf{y})^T P h(\mathbf{y})$ as the generalized model sum of squares and $GSSE = \hat{\eta}^T(I-P)\hat{\eta} = h(\mathbf{y})^T(I-P)h(\mathbf{y})$ as the generalized error sum of squares. For normal model with identity link in section 2.1, the posterior mode is $h(\mathbf{y}) = \mathbf{y}$; hence simple linear model becomes a particular case of this generalized ANOVA decomposition. Consequently, after some simple algebra, we can show that

$$\begin{aligned} &E(GSSE) = E(\hat{\eta}^T(I-P)\hat{\eta}) \simeq \sigma^2(n-p) \\ &or \quad E\Big[\frac{GSSE}{(n-p)}\Big] = E\Big[\frac{\hat{\eta}^T(I-P)\hat{\eta}}{(n-p)}\Big] \simeq \sigma^2. \end{aligned} \tag{3.16}$$

Hence, from (3.16) we define $\hat{\sigma}^2$ for (3.14) as $\hat{\sigma}^2 = \frac{\hat{\eta}^T(I-P)\hat{\eta}}{(n-p)} = \frac{GSSE}{(n-p)} = GMSE$. Note that by $GMSE$ we mean generalized mean square error. Here we must annotate the generalized Cook's distance as the scaled Euclidean distance between the two vectors of fitted value of GLM when fitting is done by including or excluding the i^{th} observation.

In the next section, we develop some simple idea of model selection criterion.

4 MODEL SELECTION CRITERIA FOR GLM

One of the fundamental mission of any scientific investigation is, given the evidence, select a statistical model from a set of potential models. In order to arbitrate the principle behind the series of observations, which is often linked directly to a statistical model predicting those observations, correct model selection is an important scientific problem that we discussed in this section. There are endless number of possible models that could have produced the data. How can one even initiate to decide on the appropriate model? One of the conventional approach is first to choose an appropriate class of models with proper scientific justification. Then choose the model from that class which is well-suited to the evidence or data. A good model selection technique will maintain the harmony between goodness of fit and complexity of the model. Generally complex models are more able to adapt the fit of the data. However, these models often face the lack of interpretability.

As we discussed in the previous paragraph that we choose the appropriate model from a scientifically justified class of models. In this paper we discuss some criteria of selecting model from the class of GLM. The orthogonal sum normalized decomposition presented at (3.15), motivates us to discuss a simple idea of *Generalized* R^2, which is analogous to the idea of selecting model based on R^2 criterion for the class of simple linear models. We can define *Generalized* R^2 as

$$GR^2 = \frac{GSSR}{GSST} = \frac{|| \, Ph(\mathbf{y}) \, ||^2}{|| \, h(\mathbf{y}) \, ||^2}. \tag{4.17}$$

Clearly for normal model with identity link as discussed in section 2.1, the posterior mode is $h(\mathbf{y}) = \mathbf{y}$, and usual R^2 of simple linear model is a special case of *Generalized* R^2. In addition to *Generalized* R^2, we also explain the notion of *Generalized Predicted Residual*

Sum of Square($GPRESS$), which can be defined as

$$GPRESS = \sum_{i=1}^{n} (\hat{\eta}_i - \hat{\hat{\eta}}_{i(i)})^2 = \sum_{i=1}^{n} \Big(\frac{\hat{\epsilon}_i}{1 - p_{ii}} \Big)^2, \tag{4.18}$$

with the same notation as explained in the previous sections.

Next, we see the performance of these methods in real data set.

5 EXAMPLES

In this section, we present the performance of our proposed methods, developed in the previous sections, for two real data sets. First we apply the method on popular data set of Finney (1947), where a carefully controlled experiments in human physiology has been demonstrated. Then we exhibit the performance of the method for second data set, where a very current investigation on Osteoarthritis of knee was performed and pain score was collected after 90 days of treatment. Bio-marker information on TNFα cytokineses was also recorded as covariates information.

5.1 Application to Finney data: an *example* of Logistic Regression.

As an illustration of the standard output from this new method of fit, and diagnostics we use the data set from Finney (1947), as test bed of our method. The detail analysis can be found in Finney (1947), Pregibon (1981) and Peng and Dey (1995). However we present a summary report of the analysis and diagnostics in this paper. We fit the logistic

model to the data as,

$$log\Big(\frac{p}{1-p}\Big) = \beta_1 + \beta_2 log(Rate) + \beta_3 log(Volume).$$

Following findings are worth to mention in this paper. The ordered component of ϵ residuals defined at systematic component level and χ residuals defined at random component level are plotted against standard normal quantiles in Figure 1. Clearly, four observations, 4^{th}, 18^{th}, 19^{th} and 24^{th} are not well fit by the model. Both χ-residuals and ϵ-residuals for these observations deviate from the straight line configuration of the others.

We compute the Generalized Cook's distance for each observation and the index plot of generalized Cook's distance in Figure 2, clearly shows that 32^{nd} observation has significant influence on the analysis. However, other outliers do not have any major influential effects on the regression fit. Hence 32^{nd} observation is most influential, even if it is not an outlier. Pregibon (1981) said this about this 32^{nd} observation:

> " A very careful reader may have noticed a discrepancy between the data used here and those given in the Finney paper. The problem is that he reports $\text{RATE}_{32} = 0.03$ but $log_{10}(10\times\text{RATE}_{32}) = 0.48(= log_{10}(10 \times 0.30))$. Judging from his plot of the data, RATE_{32} should be 0.30 rather than 0.03, but the latter value is the one that we used. Will this appreciably change the results of our analysis?..."

So possibly that was a typographical error in the Finney's data. However, we used that value 0.03, (which should be 0.30 for RATE_{32}) as it is there in the Finney's data, for our analysis. Interestingly enough, Figure 2 is detecting that typographical error of Finney's Data. One must remember in this regard that Generalized Cook's Distance is

invariant to the choice of prior.

Following Peng and Dey (1995), Hellinger distances, Kullback-Leibler divergence, and L_1 divergence measure for each observations of the Finney data were computed. All three divergence measure indicates that first five most influential observations are in the same order; i.e., 4^{th}, 18^{th}, 19^{th}, 32^{nd} and 24^{th} are the most influential observations of the data set. We must note that 4^{th}, 18^{th}, 19^{th}, and 24^{th} observations are detected as outliers by our technique, and 32^{nd} observation is detected as the most influential observation of the data set. Divergence measure techniques can identify all those atypical values in the data set. However, these measures cannot classify an atypical observation from outlier to influential observation. In Finney's data we have only two exposure variables $Volume$ and $Rate$. If we consider all possible combinations, then we have 4 different models all together. The four different models are as follows:

$$
\begin{aligned}
M1: \quad & log\Big(\frac{p}{1-p}\Big) = \beta_0, \\
M2: \quad & log\Big(\frac{p}{1-p}\Big) = \beta_0 + \beta_1 log(Volume), \\
M3: \quad & log\Big(\frac{p}{1-p}\Big) = \beta_0 + \beta_2 log(Rate), \\
M4: & log\Big(\frac{p}{1-p}\Big) = \beta_0 + \beta_1 log(Volume) + \beta_2 log(Rate).
\end{aligned}
$$

We use GR^2 and $GPRESS$ as model selection criteria. The one of the interesting feature is that GR^2 is invariant to the choice of the prior distribution. However $GPRESS$ is very sensitive to the choice of prior. So if any one wants to use the $GPRESS$ as the model selection criterion then the investigator must determine the prior before hand and use the same prior for all the models. However for GR^2, investigators need not worry about the choice of prior distribution. Irrespective of the selection criteria, the order of the model

from the best to worst as M4, M2, M3 and M1 respectively. In summary we can say GR^2 has some comparative advantages over $GPRESS$, as because this is invariant to choice of prior distribution. In addition to that GR^2 has simple interpretation of the amount of the variability explained by the model compare to that of total variability exist in the data.

Next we discuss the performance of these new methods for a recent study on Osteoarthritis of knee.

5.2 Osteoarthritis Data

Osteoarthritis (OA) is very common form of inflammatory joint disease characterized by articular cartilage degradation with an accompanying peri-articular bone response. OA affects nearly 21 million people in United States, accounting for 25% visits to primary care physicians. Clinical manifestations of OA of knee are pain in and around the joint, stiffness of the joint after rest, crepitation on motion, limited joint motion etc.

In recent years, the gum resin extracted from an ancient herb Boswellia (*Bosqwellia serrata*) has gained good attention as a potent anti-inflammatory, anti-arthritic and analgesic agent. 5-Loxin® is a novel *Boswellia serrata* extract (US Patent 2004/0073060A1) which confers a significant improvement in paw inflammation in albino Wister rats. Pertinent reference is Roy, *et.al.* (2005). Cell based *in vitro* studies and *in vivo* experiments conducted in Sprague Dawley rats demonstrate that 5-Loxin® potentially inhibits the pro-inflammatory cytokines such as TNFα. Importantly, acute and dose-dependent subchronic safety studies in rats demonstrate that 5-Loxin® does not exhibit toxic manifes-

tation.

The first human clinical trial on 5-Loxin® was performed at ASR Academy of Medical Science, Eluru, Andhrapradesh, India during 2006. A total of 70 patients were selected for the trial following the IRB protocol of the inclusion-exclusion criterion set by American College of Rheumatology. All participants submitted the written consent to participate in the study.The participants were randomly distributed into three groups; Placebo (n=23), 5 Loxin® low dose (100 mg/day) (n=24), and 5 Loxin® (250 mg/day) (n=23) high dose group. After three months of treatment, all participants were assessed for pain score and physical ability at Day 90. Various parameters of serum biochemistry, hematology and urine analysis were also carried out. In this article we present the analysis that reveals the relation between VAS (Visual Analogue Scale) score for pain, treatment and the bio-markers measured for the pro-inflammatory cytokines (measured in log-scale), named TNFα.

We assume VAS score for pain as response and has a log-normal distribution. We considered the treatment (2 (low/high) doses of 5 Loxin® and Placebo) and the Bio-markers for TNFα cytokines measured in log-scale as covariates with an intercepts parameter. We denote y_i as the VAS score for i^{th} patient, x_{i1}= indicator variable corresponding to Low dose group (5 Loxin® 100 mg/day), x_{i2}= indicator variable corresponding to High dose group (5 Loxin® 250 mg/day), x_{i3}= log(Bio-markers for TNFα cytokines) for the i^{th} patient. So altogether we denote $x_i' = (1, x_{i1}, x_{i2}, x_{i4})_{4\times 1}$ as the covariates observed for the i^{th} patient. We assume, $y_i \sim$ *log-normal*(ν_i, σ^2) with $\mu_i = e^{\nu_i}$ as the median and the link function as $\eta_i = log(\mu_i)$, where $i = 1, 2, ..., 70$. The systematic component is $\eta_i = x_i'\beta$, where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is the regression coefficient vector. We must note that

the unknown parameter σ^2 is actually a nuisance parameter and our primary interest of parameters are the regression coefficients in the systematic component part.

We recall one thing that according to FDA guideline for Bayesian Clinical trial, an FDA advisory panel and the investigator(s) must identify the prior information in an agreement meeting. Generally eliciting the prior directly on the parameters (endpoints) of interest in a clinical trial is always very controversial. However, it is easier to reach at the agreement if we elicit prior on nuisance parameter or canonical parameters. In other words, it is advisable to elicit prior on the parameter which is not the primary interest of the study. Our proposed method exactly solve that purpose. We elicit the prior on the canonical parameters and nuisance parameters as $[\nu_{i1} \mid \sigma^2] \sim Normal(\mu, \tau\sigma^2), \quad \tau > 0$ and $[\sigma^2] \sim Inv\ Gamma(a, b)$, then following the method as we discussed in section(2.2) we fit the model to the data. We use R software for our purpose of analysis.

We present the estimate of the regression coefficients, corresponding standard deviation and the 95% Credible Interval based on Monte-Carlo method in Table 2. We also perform the analysis based on as what described by Das and Dey (2006). We found the estimates of the regression parameters, based on Das and Dey's estimator, almost coincide with our Monte-Carlo estimates. However, the credible interval based on Bayesian CLT as recommended by Das and Dey (2006) is much wider than our Monte-Carlo 95% credible interval. We conclude from Table 2 as following. The 95% Credible Interval for both Low Dose and High Dose are far away from the value 0, which lead us to believe that the treatment has statistically significant effect in reducing the pain. In addition, we can conclude that High Dose is statistically significantly better than Low Dose, because 95% credible interval for High dose fall entirely below the 95% credible interval for Low dose.

Interestingly our analysis shows that the TNFα cytokines does not have any statistically significant effect over pain score. We still allow that covariate in our final model and the reason for that we substantiate from Roy, *et.al.* (2005).

The residual density plot and index plot of Generalized Cook's distance is presented in Figure 3. The residual density plot is fairly normal and we found that 26^{th}, 27^{th} and 51^{st} observations are outliers, where 26^{th} and 27^{th} patients were treated under high dose group and 51^{st} patient was assigned in placebo group. We also observed that three most influential observations, as detected by Generalized Cook's distance, are 51^{st}, 26^{th} and 27^{th} observations. Interestingly, all these three patients were also detected as outliers. We also note that observation from 52^{nd} and 63^{rd} patients have large influential effect.

6 Discussion

One major contribution of this method is to elicit the prior on either canonical parameters or nuisance parameters, avoiding the controversy of eliciting the prior on parameters of interest. In this paper we introduced the generalized version of regular diagnostic technique like Cook's distances or sample influence curve for GLM and simple idea of detecting outliers using highest marginal residual density interval. The concepts of residuals we introduced in section 3, can be improved in many ways using Bayesian asymptotic techniques. We can extend the notion of generalized Cook's distance using the simple concepts of influence function in many directions. We can broaden this view comfortably to the direction of developing generalized adaption of modified Cook's distance, DFFITS or Andrew-Pregibon type statistics etc. One must note that *Generalized* Cook's distance or

Generalized R^2 type model selection techniques are location and/or scale invariant. In consequence these measure are invariant to the choice of hyper parameters; which means they are independent of choice of prior distributions. Since Generalized Cook's Distance is invariant to the choice of prior, therefore this invariance property of Generalized Cook's Distance makes it more useful in detecting the influential observation. In Finney's data example the Figure 2 about index plots of Generalized Cook's distance exposed the 32^{nd} observation as a big influential observation, which is actually a typographical error from Finney's paper of 1947. However, this event elevates two things. Generalized Cook's distance is a good measure of detecting the influential observations and it shows that this new method is very sensitive to influential observations. Therefore we strongly recommend that one must check all the diagnostic for the atypical observations before drawing the inference based on the regression fit using this new method of GLM. The graphical presentation of these simple diagnostic techniques, presented in this paper, has a charm of its own for the broader scientific community beyond statistics. This is certainly an important feature of these diagnostic techniques presented in this paper.

References

[1] Albert, J. H., and Chib, S. (1995). "Bayesian Residual Analysis for Binary Response Regression Models". *Biometrika*, 82, 747–759.

[2] Andrew, D.F., and Pregibon, D. (1978). "Finding the Outliers that matter". *Journal of the Royal Statistical Society, Series B*, 40, 85–94.

[3] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.

[4] Belsley, D. A., Kuh. W., and Welsch., R. E. (1980). *Regression Diagnostics*. New York: John Wiley.

[5] Chatterjee, S., and Hadi, A. S. (1986). "Influential Observations, High Leverage Points, and Outliers in Linear Regression". *Statistical Science*, 1, 379–416.

[6] Chaloner, K., and Brant, R. (1988). "A Bayesian Approach to Outlier Detection and Residual Analysis". *Biometrika*, 75, 651–659.

[7] Cook, R.D., and Weisberg, S. (1982). *Residual and Influence in Regression*. London: Chapman & Hall.

[8] Das, S., and Dey, D. (2006). "On Bayesian Analysis of Generalized Linear Models using Jacobian technique". *The American Statistician*,60, 264–268.

[9] Diaconis, P., and Ylvisaker. D. (1979). "Conjugate Priors for Exponential Families". *The Annals of Statistics*, 17, 269–281.

[10] Draper, N.R., and John, J.A. (1981). "Influential Observations and Outliers in Regression". *Technometrics*, 23, 21–26.

[11] Finney, D.J. (1947). "The Estimation from Individual Records of the Relationship Between Dose and Quantal Response". *Biometrika*,34, 320–334.

[12] Gelfand, A.E., Dey. D.K., and Chang,H. (1992) "Model Determination Using Predictive Distributions with Implementation via Sampling-based Methods". *Bayesian Statistics*, (Bernardo,J. et.al., ed.), 4, 147 – 167.

[13] Griffiths, W.E., Hill, R.C., and Pope, P.J. (1987) "Small Sample Properties of Probit Model Estimator". *Journal of the American Statistical Association*,82, 929–937.

[14] Hald, A. (1999) "On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares". *Statistical Science*,14, 214–222.

[15] Hampel, F.R. (1974). "The Influence Curve and Its Role in Robust Estimation". *Journal of the American Statistical Association*,19, 431–453.

[16] Laird, N.M., and Ware, J. H. (1982). "Random Effects Models for Longitudinal Data". *Biometrics*,38,963–974.

[17] Liang, K.Y., and Zeger, S., (1986) "Longitudinal Data Analysis using Generalized Linear Models". *Biometrika*,73, 13–22.

[18] Marshall, E. C., and Spigelhalter, D.J. (2007) "Identifying Outliers in Bayesian Hierarchical models: A Simulation - based Approach". *Bayesian Analysis*, 2, 409–444.

[19] McCullough, P., and Nelder, J.A. (1989) *Generalized Linear Models*. London: Chapman and Hall.

[20] Miller, R.B., and Stewart, W.S. (2007). "Assessment of Locally Influential Observations in Bayesian Models". *Bayesian Analysis*, 2, 365–384.

[21] Nelder, J.A., and Wedderburn, R.W.M. (1972) "Generalized Linear Models". *Journal of the Royal Statistical Society, Series A*, 135, 370–384.

[22] Novic, M.R., and Hall, W.J. (1965) "A Bayesian Inference Procedure". *Journal of American Statistical Association*,60, 1104 – 1117.

[23] Pettit, L.I., and Young, K. D. S. (1990) "Measuring the Effect of Observations on Bayes Factors". *Biometrika*, 77, 455–466.

[24] Peng, F. and Dey, D. (1995). "Bayesian analysis of Outlier Problems using divergence measures". *The Candian Journal of Statistics*, 23, 199–213.

[25] Pregibon, D. (1981) "Logistic Regression Diagnostic" *Annals of Statist*,9, 705–724.

[26] Roy, S., Khanna, S., Shah, H., Rink, C., Phillips, C., Pressus, H., Subbaraju, G. V., Trimuttulu, G., Krishnaraju, A. V., Bagchi, M., Bagchi, D., and Sen, C.K. (2005) "Human Genome Screen to Identify the Genetic Basis of the Anti-Inflammatory effects of *Boswellia* in microvascular endothelial cells". *DNA Cell Biol*, 24, 244–255.

[27] U.S. Food and Drug Administration (2006) "Draft Guidence for Industry and FDA stuff: Guidence for the Use of Bayesian Statistics in Medical Device Clinical Trials", Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health (CDRH), Division of Biostatistics, Office of Surveillance and Biometrics.

[28] Weiss, R. (1996) "An Approach to Bayesian Sensitivity Analysis". *Journal of the Royal Statistical Society, Series B*, 58, 739–750.

[29] Williams, D.A. (1987) "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions". *Applied Statistics*, 36, 181 –191.

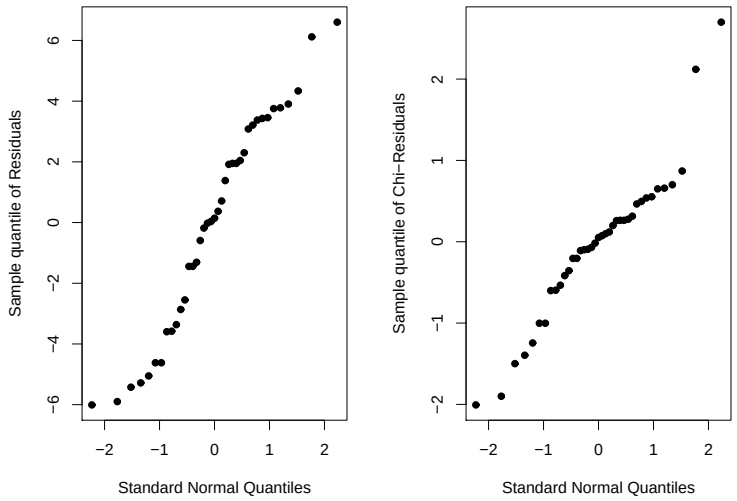
[30] Zellner, A. (1971). "Bayesian and Non-Bayesian Analysis of Log-Normal Distribution and Log-Normal Regression". *Journal of the American Statistical Association*, 66, 327–330.

Table 1: Relation between elicited and induced prior for some standard models

Model	θ	prior on θ	prior on β
Logistic regression			
with intercept only	p_i	$Beta(a, b)$	Skew-Logistic(a,b)
same as above	p_i	$Beta(a, a)$	Symmetric-Logistic (a,b)
same as above	p_i	$Beta(1, 1)$	Standard-Logistic(0,1)
Regular Logistic Model	p_i	*logit-Norm*$(x_i'\beta_0, \sigma^2)$	$\beta \sim N(\beta_0, \sigma^2(X'X)^{-1})$
Poisson regression			
with intercept only	λ_i	$Gamma(a, b)$	Log-Gamma(a,b)
Regular Poisson Model	λ_i	*log-Norm*$(x_i'\beta_0, \sigma^2)$	$\beta \sim N(\beta_0, \sigma^2(X'X)^{-1})$

Table 2: Bayesian Monte Carlo Analysis of Osteoarthritis Study

	Intercept	Low Dose	High Dose	Bio-Marker (TNFα)
	β_0	β_1	β_2	β_3
Estimates	3.550	-0.6228	-1.0650	0.0179
sd	0.43377	0.08026	0.09436	0.08820
95% CI	(2.724, 4.392)	(-0.775, -0.466)	(-1.248, -0.886)	(-0.155, 0.187)



cc

Figure 1: *Normal probability plot of Finney's data, the ordered component of ϵ residual vs standard normal quantiles and the ordered component of χ residuals vs standard normal quantiles.*

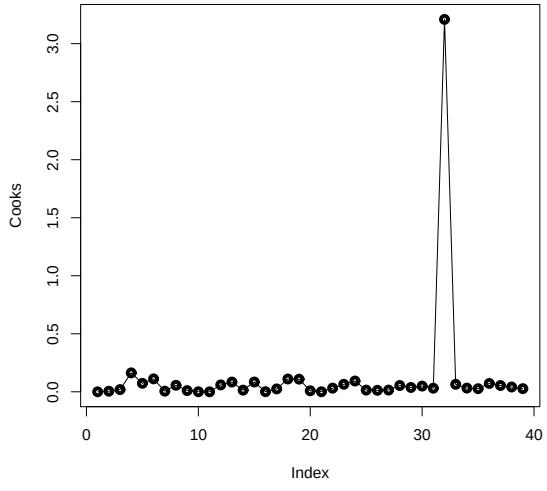


Figure 2: I*ndex plots of the Generalized Cook's Distance for Finney's Data.*

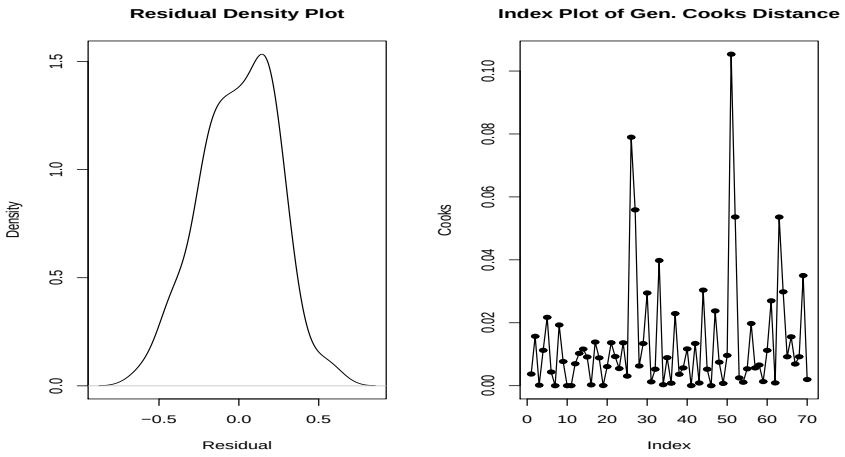


Figure 3: *Model Diagnostics of Osteoarthritis Data*