Robust Parameter Design via Bayesian Generalized Linear models

Jian-Jun Wang

Department of Management Engineering
Nanjing University of Science and Technology
Nanjing, China
Email: wjj19770818@163.com

Yi-Zhong Ma

Department of Management Engineering
Nanjing University of Science and Technology
Nanjing, China
Email: yzma-2004@163.com

Abstract—Generalized linear models (GLM) are discussed in this paper, which are used widely in the field of robust parameter design involving non-normal response variables. As for the estimation problems such as data over-dispersion which exist generally in robust parameter design, the Markov chain Monte Carlo (MCMC) approach based on adaptive rejection metropolis sampling algorithm is brought forward to simulate dynamically the Markov chain of the parameter's posterior distribution of the GLM. Furthermore, the parameters' Bayesian estimation and significant factors of the GLM will be given when relative objective Jeffreys' prior distribution is used for the parameters of the GLM. Practical industrial experiment data is utilized to simulate and analyze the Bayesian GLM by the SAS software. The results demonstrate that the Bayesian GLM performs more reliable and valid in parameter robust estimation and significant factors identification than the conventional GLM.

Keywords— MCMC simulation; Jeffreys' prior; ARMS sampling; robust parameter design; non-normal response

I. INTRODUCTION

Robust parameter design has widely been used at the development phase of product quality characteristics and processes optimization since latest three decades. The robust design method aims to find settings of the design factors so that quality characteristics are least sensitive to variation of noise factors. Although the philosophy of robust parameter design is of great value, the experimental designs and data-analysis methods proposed by Taguchi are under considerable debate. A panel discussion which was edited by Nair gave a good overview for this controversy [1]. As applications of robust parameter design are extended into many new application areas, experimenters and quality engineers have more frequently encountered non-normal response variables such as binomial responses (proportion defective data), Poisson responses (count number of defects) or exponential responses (time to failure data). For such responses involving non-normal distribution, traditional approach is to employ data transformations. However, single data transformation may fail to satisfy all required properties such as normality, constant variance and simple linear model form.

Nelder and Lee firstly proposed the use of generalized linear models (GLM) as an alternative to Taguchi's approach [2]. In order to overcome the shortcomings of data-

transformation approach, the GLM also has been used as one useful alternative to data transformation. Myers and Montgomery [3] once compared the GLM to an ordinary least squares model based on data transformation, the outcomes demonstrate a better model may be possible with the GLM. Lewis, Montgomery and Myers [4] also presented the several real examples of designed experiments with non-normal responses, giving an explicit comparison of the potential benefits of the GLM relative to the conventional approach based on data transformation.

Although the GLM is an excellent alternative to the transformation approach in robust parameter design, there still exist some estimation problems for the conventional GLM. Wu and Hamada [5] proposed if there are too many zeroes in Poisson or binomial count data, the outcome will be infinite via the method of maximum likelihood estimation. Moreover, if the over-dispersion phenomenon exists in the model with binomial or Poisson distributions, the identification of significant factors in robust design will be affected badly by the maximum likelihood estimation. In this paper, we restrict our research focuses to the estimable problems of the GLM on some special occasions. Wu and Hamada [5] once put forward to resolve above problems via Bayesian method. However, they also proposed that it would be difficult for the GLM to use Gibbs sampling from full condition. They once indicated to use an adaptive rejection sampling (ARS) method from full condition [6], but it will be restricted by the aspects of selecting the right link function in the GLM. Therefore, they failed to put forward better suggestions and explicit approach for the Bayesian analytic methods of the GLM. With the stepwise perfection of Bayesian analytic theory of the GLM and the recent advance of MCMC simulative technology, especially the improved adaptive rejection metropolis sampling (ARMS) method which was proposed by Gilks, Best and Tan [7], it will be feasible and practical to use Bayesian analytic method for the GLM via the MCMC approach. Our purpose in this paper is to present the practical approach of the Bayesian GLM in robust parameter design, giving an explicit comparison of the potential benefits of the Bayesian GLM relative to the conventional GLM as for the problems of estimation which exist generally in robust parameter design. We hope to further illustrate the benefits in terms of improved identification ability of significant factors that can be obtained with the

This paper was supported by National Natural Science Foundation of China, Grant No70672088

Bayesian GLM and encourage more widespread use of this approach.

II. THE GENERALIZED LINEAR MODEL

The GLM, introduced by Nelder and Wedderburn, is assumed to possess a probability distribution of the exponential form. The normal, binomial, Poisson, exponential, gamma, inverse normal, and negative binominal distributions are all members of the exponential family. Therefore, the experimenters or quality engineers can select flexibly a response distribution of the exponential family.

The probability density of the response Y for continuous response variables, or the probability function for discrete response, can be expressed as

$$f(y) = \exp\{a^{-1}(\phi)(y\theta - b(\theta)) + c(y, \phi)\} \quad (1)$$

The GLM consists of three components: (1) the response distribution $f(y)$, which is a member of the exponential family; (2) the linear predictor $\eta_i = X_i\beta$, which is a linear combination of the design variables X_i , including main effects and interactive effects; (3) the link function $g(\mu_i) = \eta_i$, which relates the natural mean μ_i for the specified exponential family member to the linear predictor. Parameter estimation in the GLM is typically performed with the method of maximum likelihood. At present, there are several software packages that support the computation of GLM, including SAS and S-PLUS.

III. BAYESIAN ANALYSIS OF THE GENERALIZED LINEAR MODELS

Generalized linear models have proved suitable for modeling various kinds of data consisting of exponential family response variables. Bayesian analysis of such data requires specification of a prior for the regression parameters in the model. Many statisticians favor non-informative priors because they appear to be more objective. However, it is unrealistic to expect that non-informative priors represent total ignorance about the parameter of interest. In some case, non-informative priors can lead to improper posteriors (non-integrable posterior density). Ibrahim and Laud [8] once pointed out the uniform priors for the GLM can lead to improper posterior distribution. They also gave two theorems that support the use of Jeffreys' priors for the GLM. Jeffreys' prior has been used as the reference prior in Bayesian analysis of the GLM, which is obtained by applying Jeffreys' rule. In this article, we choose Jeffreys' prior as the regression parameters of the GLM and resulting in posterior distribution of the GLM.

Suppose $y_1, \dots, y_n$ are independent observations for the GLM. For the binomial and Poisson models, ϕ is intrinsically equal to 1, whereas for the normal, gamma, and inverse Gaussian model, ϕ may be unknown. Now suppose the scale parameter ϕ is known. The Fisher information matrix of β is given by

$$\mathbf{I}(\beta) = \phi^{-1}(\mathbf{X}'\mathbf{W}\mathbf{V}(\beta)\Delta^2(\beta)\mathbf{X}) \quad (2)$$

where $\mathbf{W}$ is a $n \times n$ diagonal matrix with i th diagonal element w_i . Moreover, $\mathbf{V}(\beta)$ and $\Delta(\beta)$ are $n \times n$ diagonal matrices with i th diagonal elements

$$v_i = v(x_i'\beta) = d^2b(\theta_i) / d\theta_i^2 \quad (3)$$

$$\delta_i = \delta(x_i'\beta) = d\theta_i / d\eta_i \quad (4)$$

respectively. The matrix $\Delta(\beta)$ is the adjustment for the link function and equals the identity matrix when a canonical link is used. Jeffreys's prior for β assuming ϕ is known is now given by

$$p(\beta) \propto |\mathbf{X}'\mathbf{W}\mathbf{V}(\beta)\Delta^2(\beta)\mathbf{X}|^{1/2} \quad (5)$$

Since the likelihood function of β is

$$L(\beta | \mathbf{X}, y, \phi) = \prod_{i=1}^n p(y_i | \theta_i, \phi) \quad (6)$$

The posterior density of β is given by

$$\pi(\beta | \mathbf{X}, y, \phi, w) \propto \exp\left\{\sum_{i=1}^n \left\{\phi^{-1}w_i(y_i\theta_i - b(\theta_i))\right\}\right\} \times |\mathbf{X}'\mathbf{W}\mathbf{V}(\beta)\Delta^2(\beta)\mathbf{X}|^{1/2} \quad (7)$$

For some GLM, the scale parameter ϕ may not be known.

Jeffreys's joint prior for (β, ϕ) is given by

$$p(\beta, \phi) \propto |I(\beta, \phi)|^{1/2} \quad (8)$$

where $I(\beta, \phi)$ is the Fisher information matrix of (β, ϕ) . It is clear from equation (1) that $I(\beta, \phi)$ is block diagonal in β and ϕ . Thus Jeffreys' prior for (β, ϕ) is given by

$$p(\beta, \phi) \propto \phi^{-p/2} |\mathbf{X}'\mathbf{W}\mathbf{V}(\beta)\Delta^2(\beta)\mathbf{X}|^{1/2} \times \left(-\sum_{i=1}^n \left\{2w_i\phi^{-3} \left(\frac{db(\theta_i)}{d\theta_i}\theta_i - b(\theta_i)\right) + E\left(\frac{\partial^2 c(y_i, \phi)}{\partial \phi^2}\right)\right\}\right)^{1/2} \quad (9)$$

Giving the parameter θ_i and unknown scale parameter ϕ , according to the equation (1), the likelihood function of (β, ϕ) is given by

$$L(\beta, \phi | \mathbf{X}, y, w) = \prod_{i=1}^n p(y_i | \theta_i, \phi, w) \quad (10)$$

In terms of mutual effect of the likelihood function and Jeffreys' prior, the joint posterior density of (β, ϕ) is given by

$$\begin{aligned} \pi(\beta, \phi | \mathbf{X}, y, w) &\propto \prod_{i=1}^n p(y_i | \theta_i, \phi) \phi^{-p/2} \times |\mathbf{X}' \mathbf{W} \mathbf{V}(\beta) \Delta^2(\beta) \mathbf{X}|^{1/2} \\ &\times \left(-\sum_{i=1}^n \left\{ 2w_i \phi^{-3} \left(\frac{db(\theta_i)}{d\theta_i} \theta_i - b(\theta_i) \right) \right. \right. \\ &\left. \left. + E \left(\frac{\partial^2 c(y_i, \phi)}{\partial \phi^2} \right) \right\} \right)^{1/2} \end{aligned} \quad (11)$$

Considering the complicated form of above equation (7) and equation (11), it is very difficult to compute the accurate estimation of the parameters of the posterior distributions by means of the numerical calculation method. In this article, we attempt to work out the parameters of the posterior density in the GLM via the MCMC simulative method.

IV. MARKOV CHAIN MONTE CARLO METHOD

In moderately complex models, posterior densities are too difficult to work with directly. The MCMC method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest.

Let $\theta = (\theta_1, \dots, \theta_k)'$ be the parameter vector. Let $L(D|\theta)$ be the likelihood function, where D is the observed data. Let $p(\theta)$ be the prior distribution. The full conditional distribution of $[\theta_i | \theta_j, i \neq j, D]$ is proportional to the joint distribution, this is

$$\pi(\theta_i | \theta_j, i \neq j, D) \propto L(D|\theta) p(\theta) \quad (12)$$

For instance, the one-dimensional conditional distribution of θ_1 given $\theta_j = \theta_j^*, 2 \leq j \leq k$, is computed as

$$\begin{aligned} \pi(\theta_1 | \theta_j = \theta_j^*, 2 \leq j \leq k, D) \\ = L(D | (\theta = (\theta_1, \theta_2^*, \dots, \theta_k^*)')) \\ \times P(\theta = (\theta_1, \theta_2^*, \dots, \theta_k^*)) \end{aligned} \quad (13)$$

Suppose having a set of arbitrary starting values $\theta^{(0)} = \{\theta_1^{(0)}, \dots, \theta_k^{(0)}\}$, on the condition of the prior distribution $p(\theta)$ and known data information $L(D|\theta)$, using the ARMS algorithm of Gilk, Best and Tan [7], we can work out the sampling of the posterior distribution as follow: draw $\theta_1^{(1)}$ from $\pi(\theta_1 | \theta_2^{(0)}, \dots, \theta_k^{(0)})$; draw $\theta_2^{(1)}$ from $\pi(\theta_2^{(1)} | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$; ...; draw $\theta_k^{(1)}$ from $\pi(\theta_k^{(1)} | \theta_1^{(1)}, \dots, \theta_{k-1}^{(1)})$. This completes the transition from $\theta^{(0)}$ to $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$. After n iteration, we will have $\theta^{(n)} = (\theta_1^{(n)}, \dots, \theta_k^{(n)})$. Finally, we will get $\theta^{(1)} \dots \theta^{(k)}$. According to the ergodic theorem of the Markov chain, the marginal distribution of $\theta^{(n)}$ is guaranteed to converge to the stationary distribution when $n \rightarrow \infty$. Note that we should

check the convergence of all parameters, and not just those of interest, before proceeding to make any inference. Therefore, we must make use of the diagnostic statistics (Gelman and Rubin diagnostic statistics) or the diagnostic tools (visual trace plots) to assess the convergence of the posterior distribution of all parameters before making statistical inference. At present, we can easily use all kinds of software (SAS, WINBUGS or S-PLUS) to test the convergence of all parameters. Further, we can select the suitable Burn-in period via testing the convergence of the Markov chain. Finally, we can get the estimation value of all parameters from the posterior distribution and posterior quantities (mean, mode and posterior intervals) of interest.

V. EXAMPLE

This example comes from [5]. We will re-analyze the example to illustrate the advantages of the Bayesian GLM in robust design. In order to identify significant factors and reduce the numbers of solder defects, the experimenters research a 2_{IV}^{7-3} fractional factorial design including seven factors with two levels: A (preheat temperature), B (flux density), C (conveyor speed), D (Preheat condition), E (cooling time), F (ultrasonic solder beater) and G (solder pot temperature). The experiments explore six potential major interact effects of AB, AC, AD, BC, BD and CD besides seven main effects. The response y denotes the numbers of solder defects in some circuit board. As for every combination of factor level, the experimenters record the defect numbers of three circuit boards respectively.

In order to compare with [5], we adopt the same saturated model including seven main effects (A-G), CE and its alias effects, ABC and its alias effects. In this paper, we select the Jeffreys' prior for all parameters to ensure the objectivity. At the same time, we choose the simulative initial value of the conventional GLM without considering data over-dispersion. Throwing away the initial iteration 2000 times (Burn-in period) to ensure the convergence of all parameters, and running the 20000 times iteration to simulative analyze the GLM. In order to reduce sample autocorrelation, we need to set the thin = 2 in the SAS program. We employ the SAS program and MCMC simulative technology for the GLM to obtain the diagnostic plots and parameter estimation of every effect. Limited to the space, this paper only give the sampling trace plots, sampling autocorrelation plots and posterior density plots of the effects A and C as showed in Fig.1 and Fig.2.

From above diagnostic plots, we can intuitively judge the convergence and the autocorrelation of every parameter. Moreover, we can get the posterior estimation of every parameter as showed in Table I. Wu and Hamada [5] proposed the principle of identifying the significant factors via the Bayesian posterior intervals. If the zero is not included by the posterior intervals of some factor, we can conclude that the factor is significant, and vice versa. According to the principle, we can find that the main effects of A, B, C, G and interactive effects of AC, BD are significant factors. The result of the Bayesian GLM is accordance with the result of the

conventional GLM considering the over-dispersion in [5], but it is very different from the result of the conventional GLM without considering the over-dispersion as showed in Table II. Moreover, there exists more less standard error in the Bayesian GLM than the GLM with considering over-dispersion in [5]. Therefore, the Bayesian GLM possesses more advantages in identifying the significant factors and estimating precisely the parameter than the conventional GLM. According to those signs of the parameter estimation, we will set the optimum level of those significant factors to A, B, C, G₊. As for those insignificant factors D, E and F, we will set the suitable level in terms of other conditions such as the economic cost or the operability.

Figure 1. Trace plot, autocorrelation plot and posterior density plot of the parameter A after 20000 iteration

Figure 2. Trace plot, autocorrelation plot and posterior density plot of the parameter C after 20000 iteration

TABLE I. POSTERIOR ESTIMATE STATISTICS OF THE BAYESIAN GLM

Parameter	Estimate	Standard Error	Posterior Intervals		Posterior Intervals	
			0.5%	99.5%	2.5%	97.5%
Intercept	3.070	0.034	2.969	3.1648	2.9909	3.1453
A	-0.114	0.034	-0.212	-0.0158	-0.193	-0.0368
B	-0.136	0.034	-0.302	-0.2324	-0.212	-0.0602
C	0.416	0.034	0.3197	0.5148	0.3391	0.4956
D	-0.058	0.034	-0.154	0.0406	-0.135	0.0198
E	-0.095	0.034	-0.190	0.0010	-0.171	-0.0173
F	-0.014	0.034	-0.111	0.0802	-0.091	0.0616
G	-0.382	0.034	-0.481	-0.2876	-0.460	-0.3051
AB	0.021	0.034	-0.074	0.1178	-0.054	0.0999
AC	-0.176	0.034	-0.274	-0.0810	-0.253	-0.1002
AD	-0.096	0.034	-0.192	0.0045	-0.174	-0.0186
BC	-0.079	0.034	-0.174	0.0147	-0.158	-0.0015
BD	-0.300	0.034	-0.397	-0.2047	-0.378	-0.2222
CD	0.047	0.034	-0.047	0.1406	-0.029	0.1243
CE	0.005	0.034	-0.089	0.1007	-0.072	0.0819
ABC	-0.015	0.034	-0.116	0.0796	-0.093	0.0607

TABLE II. ESTIMATE STATISTICS OF THE CONVENTIONAL GLM

Parameter	Estimate	Standard Error	Wald 97.5% confidence limits		P
Intercept	3.0706	0.0346	2.993	3.148	<.0001
A	-0.1141	0.0346	-0.192	-0.037	0.0010
B	-0.1362	0.0346	-0.214	-0.059	<.0001
C	0.4167	0.0346	0.3391	0.4942	<.0001
D	-0.0589	0.0346	-0.1365	0.0186	0.0885
E	-0.0956	0.0346	-0.1732	-0.0180	0.0057
F	-0.0146	0.0346	-0.0922	0.0630	0.6732
G	-0.3822	0.0346	-0.4598	-0.3047	<.0001
AB	0.0215	0.0346	-0.0561	0.0991	0.5347
AC	-0.1760	0.0346	-0.2536	-0.0984	<.0001
AD	-0.0961	0.0346	-0.1737	-0.0185	0.0055
BC	-0.0797	0.0346	-0.1573	-0.0022	0.0212
BD	-0.2998	0.0346	-0.3773	-0.2222	<.0001
CD	0.0472	0.0346	-0.0303	0.2148	0.1722
CE	0.0051	0.0346	-0.0725	0.0827	0.8826
ABC	-0.0154	0.0346	-0.0930	0.0622	0.6560

VI. CONCLUSION

In this paper, as for some estimation problems including the data over-dispersion in the conventional GLM, we propose a new method of robust parameter design via the Bayesian GLM to improve the ability of identifying the significant factors. Compared with the conventional GLM without considering the over-dispersion, the Bayesian GLM possesses more advantages in the aspects of identifying the significant factors and the parameter estimation. Moreover, In view of the GLM involving many usual exponential distributions, the Bayesian analytical method of the GLM will be applicable to other fields such as engineering science, biometrics and management science.

ACKNOWLEDGEMENTS

This paper was supported by the National Natural Science Foundation of China (NSFC). Grant Number is 70672088.

REFERENCES

- [1] V. N. NAIR,. "Taguchi's Parameter Design: A Panel Discussion". Technometrics, vol.30.pp.127-161,1992.
- [2] J.A. NELDER and Y.LEE "Generalized Linear Models for the Analysis of Taguchi-Type Experiments". Applied Stochastic Models and Data Analysis,vol.7, pp.107-120,1991.
- [3] R.H. MYERS and D.C. MONTGOMERY "A Tutorial on Generalized Linear Models".Journal of Quality Technology,Vol.29, pp. 274-291,1997.
- [4] S.L. LEWIS, D.C.MONTGOMERY and R.H. MYERS "Examples of Designed Experiments With Nonnormal Responses." Journal of Quality Technology,Vol.33, pp.265-278,2001.
- [5] C.F.J.Wu and M. Hamada "Experiments: Planning, Analysis and Parameter Design Optimization". John Wiley, New York, NY,2000.
- [6] W.R.Gilks and P.Wild "Adaptive Rejection Sampling for Gibbs Sampling with Gibbs Sampling". Applied Statistics,Vol.41, pp.337-348,1992.
- [7] W.R.Gilks, N.G.Best and K.K.C.Tan " Adaptive Rejection Metropolis Sampling with Gibbs Sampling". Applied Statistics,Vol.44, pp.452-472,1995.
- [8] J.G.Ibrahim and P. W. Laud. "On Bayesian Analysis of Generalized Linear Models Using Jeffreys'Prior".Journal of the American Statistical Association,Vol.86,pp.981-986,1991.