SSC Annual Meeting, May 2008
Proceedings of the Survey Methods Section

CONSTRAINED INFERENCE IN GENERALIZED LINEAR AND MIXED MODELS

Karelyn Davis[1] , Chul Gyu Park[2], Sanjoy Sinha[2]

ABSTRACT

In the past half-century, statisticians have recognized the improvement in efficiency of many inference problems as a result of implementing the prior ordering of parameters or restrictions in the analysis. As it is often the case that observations are not normally distributed and are sometimes observed in a cluster, generalized linear models (GLMs) or generalized linear mixed models (GLMMs) are employed. The paper extends estimation and hypothesis testing methods for such models under inequality constraints using the Gradient Projection (GP) algorithm. Results of simulation studies and applications are also discussed.

KEY WORDS: Constrained inference; Generalized linear models; Gradient projection; Linear models; Maximum likelihood; Mixed models

RÉSUMÉ

Au cours des dernières années, les statisticiens ont reconnu l'amélioration en précision des problèmes d'inférence lorsqu'on tient compte de l'ordre des paramètres ou de restrictions dans l'analyse. Souvent les données n'ont pas une distribution normale ou se retrouve en grappes, alors les modèles linéaires généralisés (GLM) ou linéaires généralisés mixtes (GLMM) sont utilisés. L'article examinera des méthodes d'estimation et de tests d'hypothèses de tels modèles sous des contraintes d'inégalité en appliquant l'algorithme de projection du gradient (GP). Les résultats de simulations et d'applications seront aussi discutés.

MOTS CLÉS : Inférence sous contraintes; maximum de vraisemblance; modèles linéaires; modèles mixtes; modèles linéaires généralisés; projection du gradient;

1. INTRODUCTION AND MOTIVATION

Statistical modeling and analysis techniques for observational and experimental data often require methods to address a constrained parameter environment. Problems of such type may originate from various fields of study: an educator may wish to determine if levels of distraction varying from none to excessive during an examination result in scores in the reverse order of magnitude; a sociologist may examine if people in low, middle and high socio-economic groups possess low, middle and high knowledge of current events; and a National Hockey League (NHL) owner may be interested in determining whether selecting players with a high ranking in the Entry Draft will lead to improved team performance (Daniel, 1990 and Dawson and Magee, 2001). Hypotheses of this nature are referred to as ordered alternatives and are studied in the general area of order restricted, or constrained, statistical inference.

The implementation of constraints in statistical analysis has been studied under various names, including one-sided testing, isotonic regression or restricted analysis. Advantages of using such constraints are that the restrictions are often natural, and allow for additional estimation and hypothesis tests; inference with constraints is often more efficient than unrestricted counterparts which ignore the constraints; and restricted ML estimation has also been shown to obtain consistent estimates of parameters in most cases. On the other hand, such constraints require additional algorithms which may be complex or inefficient in terms of computing time to implement; and algorithms are usually proposed for specific cases. The reader is asked to refer to the books by Silvapulle and Sen (2005) and Robertson, Wright and Dykstra (1988) for further details.

1 Karelyn Davis, Statistics Canada, Ottawa ON, Canada K1A 0T6 & School of Mathematics and Statistics, Carleton University, Ottawa ON Canada K1S 5B6 (Karelyn.Davis@statcan.gc.ca);
2 Chul Gyu Park and Sanjoy Sinha, School of Mathematics and Statistics, Carleton University, Ottawa ON Canada K1S 5B6

In a linear model (LM), the mean function is of a linear form $\mu_{ij} = \mu + \alpha_i + \beta_j$, where μ, α_i and β_j are unknown constants for which interest lies in their estimation. If either the α_i's or β_j's are random variables as opposed to fixed constants, the model is termed a linear mixed model (LMM). Normality of the random α_i's is also often assumed. However, in many statistical applications, it is not the case that the mean of an observation is a linear combination of parameters nor that data are normally distributed. To model such non-normal data, generalized linear models (GLMs) and generalized linear mixed models (GLMMs) are often employed.

In particular, GLMMs are of importance to current statistical problems. Breslow and Clayton (1993) describe numerous applications including modeling of longitudinal data, overdispersion, spatial aggregation, etc. In effect, models are built to accommodate correlated data or to consider levels of a factor as selected from a population of levels in order to make inference to that population (McCulloch and Searle, 2001). Nevertheless, the addition of random effects into a generalized linear model complicates procedures for estimating the model parameters. McCulloch and Searle (2001) review a number of existing methods for analyzing mixed models. In essence, maximum likelihood estimation of parameters is preferred; however it is difficult and computationally intensive.

Our motivation is to extend constrained inference to analyze problems of the following type:

- Estimation of the effect of increasing age levels or ordered income levels on the probability of being a smoker
- Observations are observed longitudinally or in clusters (e.g. schools, neighbourhoods)

Hence, we are motivated to develop constrained inference techniques in GLMs and GLMMs for analyzing such nonlinear data.

Constrained inferences have been considered in many papers, including Dykstra (1983), El Barmi and Dykstra (1994, 1995), El Barmi and Johnson (2006) and Dardanoni and Forcina (1998). However, these papers have focused on inferences under the normal or multinomial setting. In the context of linear mixed models, relatively few authors have proposed algorithms for maximum likelihood (ML) estimation under inequality (Shi et al.(2005), Zheng et al.(2005)) or equality constraints (Kim and Taylor(1995)). Nevertheless, many of these algorithms concern a specific cone type of constraint (i.e. $A\boldsymbol{\beta} \leq \mathbf{0}$). In contrast, Jamshidian (2004) used the Gradient Projection (GP) algorithm for equality and inequality constraints in a general likelihood function with missing values and provided an example of this estimation method for linear mixed models. Given the general nature of this procedure, both in terms of linear inequality constraints and general likelihood function, we are motivated to apply these results to the nonlinear mixed model context. Furthermore, for hypothesis testing, Silvapulle (1994) considered the likelihood ratio test (LRT) for the GLM problem.

The paper outlines an innovative method of maximum likelihood estimation for GLMs and GLMMs under linear inequality constraints, and assesses some relevant properties of the estimators. In addition, extensions to constrained likelihood ratio tests are discussed. The paper is organized as follows: Section 2 describes the models under study and background on the GP algorithm and its application to ML estimation and hypothesis testing. Section 3 offers results of a simulation study which investigates the properties of the constrained versus unconstrained estimators. Section 4 provides a summary and suggestions for future research.

2. MAXIMUM LIKELIHOOD ESTIMATION WITH INEQUALITY CONSTRAINTS

In the case of a generalized linear model, we note that μ_i, the mean of the response variable, y_i, is related to the explanatory variables through the link function g(μ_i) = $x_i^t\boldsymbol{\beta}$. Assuming $y_i \sim$ *indep.* $f_{Y_i}(y_i)$, i = 1,2,…,n with $f_{Y_i}(y_i) = \exp\{[y_i\theta_i - b(\theta_i)]/\tau^2 - c(y_i,\tau)\}$, the log-likelihood function becomes

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)]/\tau^2 - \sum_{i=1}^{n}c(y_i,\tau).$$

We note that the conditional mean of y_i is related to the natural or canonical parameter θ_i by $\mu_i = \partial b(\theta_i)/\partial\theta_i$. Furthermore, it is often the case that the canonical parameter θ_i is a linear function of the predictors and parameters as θ_i = g(μ_i) = $x_i^t\boldsymbol{\beta}$. For example, in the case of a Bernoulli distribution, we have θ_i = g(μ_i)= $\log(\mu_i/(1-\mu_i))$.

The ML estimating equations for $\boldsymbol{\beta}$ may then be derived as:

$$\mathbf{X^t W \Delta (y - \mu) = 0},$$

where $\mathbf{W} = diag(w_i) = diag\left[v(\mu_i)(g_\mu^2(\mu_i))^{-1}\right]$ and $\boldsymbol{\Delta} = diag(g_\mu(\mu_i))$, with $v(\mu_i) = \partial^2 b(\theta_i)/\partial\theta_i^2$ and $g_\mu(\mu_i)$ is the derivative of g(μ_i) with respect to μ_i. For further details, the reader is referred to McCulloch and Searle (2001) or McCullagh and Nelder (1989).

In the case of a generalized linear mixed model, we may augment the link function to account for the clustering and/or longitudinal aspect of the data with g(μ_i) = $x_i^t\boldsymbol{\beta} + z_i^t\mathbf{u}$ where $\mathbf{u}$ is the vector of random effects which accounts for such correlation or overdispersion. Then, as before, we consider the conditional distribution $y_i \mid \mathbf{u} \sim indep.\ f_{Y_i|u}(y_i \mid \mathbf{u})$, with $f_{Y_i|u} = \exp\{[y_i\theta_i - b(\theta_i)]/\tau^2 - c(y_i,\tau)\}$ and $\mathbf{u} \sim f_u(\mathbf{u} \mid \Sigma)$. For such models, we consider the marginal likelihood function, which is obtained by integrating over the random effects,

$$L(\boldsymbol{\beta},\tau,\Sigma) = \int f_{y_i|u}(y_i \mid \mathbf{u})\ f_u(\mathbf{u} \mid \Sigma)\ d\mathbf{u},$$

However, many authors have noted that these high-dimensional integrals are difficult to calculate (see McCulloch and Searle (2001)). Numerous methods have been proposed in the literature to alleviate these computational difficulties by either approximating the integrals or the integrand under study. While penalized quasi-likelihood and generalized estimating equations have many practical advantages, these methods do not have the optimal properties of statistical efficiency as those of maximum likelihood estimation and may lead to inconsistent estimates in some cases. Thus, for both constrained and unconstrained ML estimation, we consider evaluating the integrals using numerical quadrature methods.

In many practical situations, such as the binary and Poisson regression models, the dispersion parameter τ is fixed at unity. For the simulation study described in Section 3, we assumed τ = 1. However, in some situations, such as the gamma distribution, an estimating equation for τ may be determined, using a similar technique. Algorithms for computing the constrained ML estimates of $\boldsymbol{\beta}$ and Σ are detailed in the next section.

2.1. Gradient Projection Algorithm

Consider inequality constraints of the form $A\boldsymbol{\beta} \le \mathbf{c}$, where A is an r x p matrix of full rank $r \le p$, thus the constrained parameter space is $\Omega = \{\boldsymbol{\beta} : A\boldsymbol{\beta} \le \mathbf{c}\}$. The method proposed by Jamshidian (2004) implements the gradient projection (GP) algorithm which searches among all active sets (constraints which hold with equality) to determine the optimal solution. This method is globally convergent, and finds a solution to maximize the log-likelihood function $l(\boldsymbol{\beta})$ or $l(\boldsymbol{\beta},\Sigma)$ subject to

$$a_i^t\boldsymbol{\beta} = c_i \quad i \in I_1$$
$$a_i^t\boldsymbol{\beta} \le c_i \quad i \in I_2,$$

where I_2 is the index set of the rows of A pertaining to inequality constraints, and I_1 is the index set for the rows of A corresponding to equality constraints. The algorithm begins with an initial working set of active constraints, denoted $\mathcal{W}$. This set includes indexes of the constraints in I_1, if any, and may include indexes from I_2. Let $\overline{A}$ be an $\overline{m}$ x p matrix whose rows consist of a_i^t for all $i \in \mathcal{W}$ and let $\bar{\mathbf{c}}$ be the corresponding vector of c_i's. Moreover, define the generalized gradient of $l(\boldsymbol{\beta})$ or $l(\boldsymbol{\beta},\Sigma)$ in the matrix of $\mathbf{W}^*$ as $\tilde{s}(\boldsymbol{\beta}) = \mathbf{W}^{*-1}s(\boldsymbol{\beta})$ where $s(\boldsymbol{\beta})$ denotes the gradient vector and $\mathbf{W}^*$ is the estimate of the information matrix defined on the next page.

Beginning with an initial point $\boldsymbol{\beta}_r$ that satisfies $\overline{A}\boldsymbol{\beta} \le \bar{\mathbf{c}}$, the algorithm proceeds as follows:

1. Compute $\mathbf{d} = P_{\mathbf{W}^*}\,\tilde{s}(\boldsymbol{\beta}_r)$, where $P_{\mathbf{W}^*} = I - \mathbf{W}^{*-1}\,\overline{A}^t(\overline{A}\mathbf{W}^{*-1}\,\overline{A}^t)^{-1}\overline{A}$.
2. If $\mathbf{d} = \mathbf{0}$, compute the Lagrange multipliers $\boldsymbol{\lambda} = (\overline{A}\mathbf{W}^{*-1}\,\overline{A}^t)^{-1}\overline{A}\,\tilde{s}(\boldsymbol{\beta}_r)$.

a. If $\lambda_i \geq 0$ for all $i \in \mathcal{W} \cap I_2$, stop. The current point satisfies the Kuhn-Tucker necessary conditions.

b. If there is at least one $\lambda_i < 0$ for $i \in \mathcal{W} \cap I_2$, determine the index corresponding to the smallest such λ_i and delete the index from $\mathcal{W}$. Modify $\overline{A}$ and $\overline{\mathbf{c}}$ by dropping a row from each accordingly and go to Step 1.

3. If $\mathbf{d} = \mathbf{0}$, obtain $\alpha_1 = \arg\max_{\alpha}\{\alpha : \boldsymbol{\beta} + \alpha\,\mathbf{d}$ is feasible$\}$. Then search for $\alpha_2 = \arg\max_{\alpha}\{\, l(\boldsymbol{\beta} + \alpha\,\mathbf{d})\text{: } 0 \leq \alpha \leq \alpha_1 \}$. Set $\widetilde{\boldsymbol{\beta}}_r = \boldsymbol{\beta}_r + \alpha_2\,\mathbf{d}$. Add indexes of new coordinates, if any, of $\boldsymbol{\beta}_r$ that are newly on the boundary to the working set $\mathcal{W}$. Modify $\overline{A}$ and $\overline{\mathbf{c}}$ by adding additional rows.

4. Replace $\boldsymbol{\beta}_r$ by $\widetilde{\boldsymbol{\beta}}_r$ and go to Step 1, continuing until convergence.

For GLMs, we note the information matrix to be (See McCulloch and Searle (2001, p. 144):

$$\mathbf{I}(\boldsymbol{\beta}, \tau^2) = \begin{bmatrix} \mathbf{I}(\beta) & \mathbf{I}(\beta, \tau^2) \\ \mathbf{I}(\beta, \tau^2)^t & \mathbf{I}(\tau^2) \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau^2}\mathbf{X}^t\mathbf{W}\mathbf{X} & \mathbf{0} \\ \mathbf{0}^t & \mathbf{I}(\tau^2) \end{bmatrix},$$

while for the GLMM case,

$$\mathbf{I}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} E\left[\partial U(\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}^t \mid y_i\right] - \sum_{i=1}^{n} E\left[U(\boldsymbol{\gamma})U(\boldsymbol{\gamma})^t \mid y_i\right] + \sum_{i=1}^{n} E\left[U(\boldsymbol{\gamma}) \mid y_i\right] E\left[U(\boldsymbol{\gamma}) \mid y_i\right]^{\mathrm{t}},$$

where $U(\boldsymbol{\gamma}) = \partial\, l(\boldsymbol{\gamma} \mid y_i) / \partial \boldsymbol{\gamma}$, $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \tau^2, \sigma^2)^{\mathrm{t}}$.

2.2. Likelihood Ratio Tests

Once the maximum likelihood estimate is obtained, various hypothesis tests may be considered. First, define

H_{o}: $A\boldsymbol{\beta} = \mathbf{c}$

H_1: $A\boldsymbol{\beta} \leq \mathbf{c}$

H_2: $\boldsymbol{\beta}$ unconstrained.

Then, we may derive likelihood ratio tests (LRT) for the two hypotheses, as defined by Silvapulle and Sen (2005, p. 28):

(T1) H_{o} versus $H_1 - H_o$

(T2) H_1 versus $H_2 - H_1$ (Goodness-of-fit test)

The hypothesis test (T2) is termed a goodness-of-fit test for the constrained parameter space under study, and would usually be performed before (T1). Further, consider

$$T_{01} = \left[\sup_{A\boldsymbol{\beta} \leq \mathbf{0}} l(\boldsymbol{\beta}) - \sup_{A\boldsymbol{\beta} = \mathbf{0}} l(\boldsymbol{\beta})\right] \text{ for (T1*)} \quad \text{and} \quad T_{12} = \left[\sup_{\boldsymbol{\beta} \in R^p} l(\boldsymbol{\beta}) - \sup_{A\boldsymbol{\beta} \leq \mathbf{0}} l(\boldsymbol{\beta})\right] \text{ for (T2*)},$$

where (T1*) and (T2*) represent (T1) and (T2) with $\mathbf{c} = \mathbf{0}$. Silvapulle (1994) showed for the GLM case, that for (T1*):

$$T_{01} = \frac{1}{\hat{\tau}^2}\left[\min_{A\boldsymbol{\beta} = \mathbf{0}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathbf{X}^t \mathbf{W}(\boldsymbol{\beta}_0)\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \min_{A\boldsymbol{\beta} \leq \mathbf{0}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathbf{X}^t \mathbf{W}(\boldsymbol{\beta}_0)\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] + o_p(1),$$

where $\hat{\boldsymbol{\beta}}$ represents the unconstrained global MLE, and $\boldsymbol{\beta}_0$ represents the true value under H_{o}.

The asymptotic null distribution of the LRT follows a *chi-bar-square* distribution as follows:

$$\lim_{n \to \infty} P_{\boldsymbol{\beta}_0}(T_{01} \geq t_{01} \mid H_0) = \sum_{i=0}^{r} w_i(r, \hat{\tau}^2 A(\mathbf{X}^t\mathbf{W}(\boldsymbol{\beta}_0)\mathbf{X})^{-1} A^t) P(\chi_i^2 \geq t_{01}).$$

For the goodness-of-fit test (T2*), as in Silvapulle and Sen (2005), we let H_1 be satisfied, with $\boldsymbol{\beta}_1$ denoting the true value in H_1 that also belongs to H_{o}. Then we may show

$$T_{12} = \frac{1}{\hat{\tau}^2}\left[\min_{A\boldsymbol{\beta}\le\mathbf{0}} (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^t \mathbf{X}^t\mathbf{W}(\boldsymbol{\beta}_1)\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right] + o_p(1).$$

The asymptotic least favourable null distribution of the LRT is also chi-bar-square as

$$\lim_{n\to\infty} P_{\boldsymbol{\beta}_1}(T_{12} \ge t_{12} \mid \boldsymbol{\beta} = \mathbf{0}) = \sum_{i=0}^{r} w_{r-i}(r, \hat{\tau}^2 A(\mathbf{X}^t\mathbf{W}(\boldsymbol{\beta}_1)\mathbf{X})^{-1}A^t) P(\chi_i^2 \ge t_{12}),$$

where $\boldsymbol{\beta} = \mathbf{0}$ is the least favourable null value.

As noted earlier, the chi-bar-square distributions are a weighted sum of chi-square distributions. The weights, $w_i(r, \mathbf{V})$, represent the probability that the least squares projection of *r*-dimensional normal observations from $N(\mathbf{0}, \mathbf{V})$ onto the orthant cone has exactly *i* positive component values. For a given information matrix, $\mathbf{W}^*$, these weights may be estimated via simulation after replacing $\boldsymbol{\beta}_1$ (or $\boldsymbol{\beta}_0$) and τ by their unconstrained estimators (see Silvapulle and Sen (2005), sections 3.5 and 3.6).

3. SIMULATION STUDY

A simulation study was conducted to assess the bias and mean squared error (MSE) of the constrained and unconstrained estimates using the aforementioned methods. The constraints of interest and associated matrix are as follows:

$$\begin{array}{l} \beta_0 + \beta_1 \le \quad c_1 \\ \beta_0 - \beta_1 \ge \ -c_2 \end{array} \quad \Rightarrow \quad A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Both Bernoulli and Poisson models for the GLM and GLMM case are considered. First, $n = 100$ observations were generated from Bernoulli(π_i) where $\pi_i = \exp(x_i^t\boldsymbol{\beta})/(1+\exp(x_i^t\boldsymbol{\beta}))$. Similarly, $n = 100$ observations were generated from Poisson (λ_i) where $\lambda_i = \exp(x_i^t\boldsymbol{\beta})$. The results for these cases are displayed in Tables 1 and 2 respectively. Furthermore, for the GLMM case, we assumed $u_i \sim$ i.i.d. $N(0, \sigma^2)$. In the Bernoulli GLMM, $\sigma^2 = 1$ and $n = 200$ clusters each of size $k = 4$ were generated from Bernoulli(π_{ij}) with $\pi_{ij} = \exp(x_{ij}^t\boldsymbol{\beta} + u_i)/(1+\exp(x_{ij}^t\boldsymbol{\beta} + u_i))$. For the Poisson GLMM, $\sigma^2 = 0.25$, and $n = 200$ clusters each with size $k = 4$ were generated from Poisson (λ_{ij}) where $\lambda_{ij} = \exp(x_{ij}^t\boldsymbol{\beta} + u_i)$. The GLMM results are displayed in Tables 3 and 4, respectively. Note that for both Bernoulli and Poisson, we have $\tau = 1$.

Moreover, for the GLMM case, as there are no restrictions placed on the σ^2 parameter, the augmented matrices are needed

$$A^* = [A \mid \mathbf{0}], \quad \boldsymbol{\gamma} = (\boldsymbol{\beta}, \sigma^2)^t.$$

Also, $\mathbf{X} \sim U(a_1, a_2)$, where for Bernoulli we have $a_1 = -4.5$, $a_2 = -2$ and for Poisson, $a_1 = 1$, $a_2 = 5$. In addition, for Bernoulli we consider constraint values $c_1 = 5$, $c_2 = -2$ whereas for Poisson, $c_1 = 1$, $c_2 = 0.5$.

The unconstrained estimation was performed using the software package R and associated functions *glm()* and *glmmML()*. For each simulation, three constraint cases were considered: (a) those beta values on the vertex point (e.g. $\boldsymbol{\beta} = (3.50, 1.50)^t$), (b) values on the boundary of the constraints or (c) values within the constraint cone. The empirical bias and MSE were then computed for each parameter case, in both the unconstrained and constrained settings. The simulation results are provided on the following page.

Table 1 – Bernoulli GLM Model

Case	Parameter	Constrained Bias	Constrained MSE	Unconstrained Bias	Unconstrained MSE
(a)	$\beta_0 = 3.50$	-0.0318	0.0139	0.0962	2.4688
	$\beta_1 = 1.50$	-0.0650	0.0150	0.0361	0.5891
(b)	$\beta_0 = 3.00$	0.1055	0.0547	0.1075	1.1440
	$\beta_1 = 1.00$	0.0255	0.0091	0.0374	0.1067
(b)	$\beta_0 = 3.90$	-0.2072	0.1481	0.0873	1.2059
	$\beta_1 = 1.10$	-0.0642	0.0163	0.0255	0.1051
(c)	$\beta_0 = 3.00$	0.1149	0.1104	0.0188	1.2435
	$\beta_1 = 0.70$	0.0296	0.0144	0.0032	0.1060
(c)	$\beta_0 = 3.90$	-0.2090	0.1384	0.0350	0.6443
	$\beta_1 = 0.90$	-0.0623	0.0133	0.0106	0.0552

Table 2 – Poisson GLM Model

Case	Parameter	Constrained Bias	Constrained MSE	Unconstrained Bias	Unconstrained MSE
(a)	$\beta_0 = 0.25$	0.0072	0.0039	-0.0113	0.0083
	$\beta_1 = 0.75$	-0.0784	0.0137	0.0049	0.0261
(b)	$\beta_0 = 1.30$	-0.0786	0.0220	-0.0025	0.0398
	$\beta_1 = -0.30$	0.0210	0.0032	-0.0017	0.0049
(b)	$\beta_0 = 0.15$	0.0428	0.0056	-0.0063	0.0148
	$\beta_1 = 0.65$	-0.0095	0.0004	0.0014	0.0010
(c)	$\beta_0 = 1.30$	-0.0527	0.0349	0.0137	0.0615
	$\beta_1 = -0.45$	0.0127	0.0059	-0.0083	0.0090
(c)	$\beta_0 = 0.35$	-0.0276	0.0026	-0.0001	0.0115
	$\beta_1 = 0.65$	0.0067	0.0002	0.0002	0.0007

Table 3 – Bernoulli GLMM Model

Case	Parameter	Constrained Bias	Constrained MSE	Unconstrained Bias	Unconstrained MSE
(a)	$\beta_0 = 3.50$	-0.0102	0.0016	-0.0554	0.3509
	$\beta_1 = 1.50$	-0.0216	0.0017	-0.0261	0.0849
	$\sigma^2 = 1$	0.0009	0.0036	-0.0317	0.0699
(b)	$\beta_0 = 2.50$	0.1309	0.0687	-0.0449	0.1555
	$\beta_1 = 0.50$	0.0321	0.0062	-0.0167	0.0134
	$\sigma^2 = 1$	0.0256	0.0387	-0.0195	0.0811
(b)	$\beta_0 = 4.00$	-0.2118	0.1202	-0.0684	0.1989
	$\beta_1 = 1.00$	-0.0613	0.0103	-0.0201	0.0169
	$\sigma^2 = 1$	-0.0561	0.0457	-0.0259	0.0828
(c)	$\beta_0 = 2.65$	0.1224	0.0382	-0.0558	0.1513
	$\beta_1 = 0.60$	0.0317	0.0052	-0.0172	0.0126
	$\sigma^2 = 1$	-0.0587	0.0382	-0.1735	0.0911
(c)	$\beta_0 = 4.00$	-0.2184	0.1296	-0.0677	0.2299
	$\beta_1 = 0.90$	-0.0629	0.0107	-0.0208	0.0182
	$\sigma^2 = 1$	-0.0627	0.0593	-0.0271	0.0906

Table 4 – Poisson GLMM Model

Case	Parameter	Constrained Bias	Constrained MSE	Unconstrained Bias	Unconstrained MSE
(a)	$\beta_0 = 0.25$	0.0400	0.0028	0.0541	0.0041
	$\beta_1 = 0.75$	-0.0682	0.0071	-0.0689	0.0073
	$\sigma^2 = 0.25$	0.0065	0.0008	0.0056	0.0011
(b)	$\beta_0 = 2.00$	-0.0397	0.0033	0.0681	0.0158
	$\beta_1 = -1.00$	0.0335	0.0023	-0.0074	0.0026
	$\sigma^2 = 0.25$	0.0101	0.0010	0.0485	0.0050
(b)	$\beta_0 = -0.75$	0.1433	0.0367	0.1135	0.0417
	$\beta_1 = -0.25$	-0.0349	0.0037	-0.0291	0.0042
	$\sigma^2 = 0.25$	0.0315	0.0105	0.0569	0.0160
(c)	$\beta_0 = -0.65$	0.1276	0.0329	0.0987	0.0384
	$\beta_1 = -0.25$	-0.0288	0.0030	-0.0216	0.0036
	$\sigma^2 = 0.25$	0.0386	0.0096	0.0530	0.0125
(c)	$\beta_0 = -0.55$	0.1169	0.0326	0.1064	0.0352
	$\beta_1 = -0.25$	-0.0285	0.0032	-0.0257	0.0035
	$\sigma^2 = 0.25$	0.0488	0.0106	0.0527	0.0112

From these results, it is evident that in many cases, the constrained estimates have larger bias than the unconstrained counterparts. However, the MSE for the constrained estimates is considerably less for all cases under study. In particular, when a given beta moves closer to the boundary points, the bias increases and MSE decreases for the constrained over the unconstrained estimates. Moreover, when comparing the two GLMM cases, we note a similar pattern of smaller MSE in the variance component analysis, despite no *a priori* constraints present. Hence, the estimation of the variance component is affected by imposing constraints on the regression parameters.

4. CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

In this paper, we have described a method for calculating the ML estimates in generalized linear and generalized linear mixed models under constraints. Such methods have wide application in many statistical and non-statistical disciplines. The gradient projection algorithm is implemented to calculate maximum likelihood estimates under linear inequality constraints and resulting estimates exhibit smaller MSE than unconstrained counterparts. While larger bias and smaller MSE are typical properties of constrained estimators, development of methods for constrained likelihood ratio tests for the GLMM case is of interest. Power comparisons will be performed and an application to the Statistics Canada Youth Smoking Survey 2002 is planned. Future work would include extending these methods to constrained problems of general variance structures, and incomplete or survey sampling data.

REFERENCES

Breslow, N.E. and Clayton, D.G. (1993). "Approximate inference in generalized linear mixed models". *Journal of the American Statistical Association*, **88**, 9-25.

Daniel, Wayne W. (1990) *Applied Nonparametric Statistics* (2nd edition). Pacific Grove: Duxbury.

Dardanoni, V. and Forcina, A. (1998). "A unified approach to likelihood inference in stochastic orderings in a nonparametric context". *Journal of the American Statistical Association*, **93**, 1112-1123.

Dawson, D. and Magee, L. (2001). "The National Hockey League Entry Draft, 1969-1995: An Application of a Weighted Pool-Adjacent Violators Algorithm". *The American Statistician*, **55**(3), 194-199.

Dykstra, R. (1983). "An algorithm for restricted least squares regression". *Journal of the American Statistical Association,* **78**, 837-842.

El Barmi, H. and Dykstra, R. (1994). "Restricted multinomial MLE based upon Fenchel duality". *Statistics and Probability Letters*. **21**, 121-130.

El Barmi, H. and Dykstra, R. (1995). "Testing for and against a set of linear inequality constraints in a multinomial setting". *The Canadian Journal of Statistics*, **23**, 131-143.

El Barmi, H. and Johnson, M. (2006). "A unified approach to testing for and against a set of linear inequality constraints in the product multinomial setting". *Journal of Multivariate Analysis*, **97**, 1894-1912.

Jamshidian, M. (2004). "On algorithms for restricted maximum likelihood estimation". *Computational Statistics & Data Analysis,* **45**, 137-157.

Kim, D.K., and Taylor, J.M. (1995). "The Restricted EM Algorithm for Maximum Likelihood Estimation Under Linear Restrictions on the Parameters". *Journal of the American Statistical Association*, **90**, 708-716.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd Ed.). New York: Wiley.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd Ed.). London: Chapman & Hall.

McCulloch, C.E. and Searle, S.E. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Chichester, U.K.: John Wiley.

Shi, N.Z., Zheng, S.R., Guo, J. (2005). "The restricted EM algorithm under inequality restrictions on the parameters". *Journal of Multivariate Analysis*, **92**, 53-76.

Silvapulle, M.J. (1994). “On tests against one-sided hypotheses in some generalized linear models”. *Biometrics*, **50**, 853-858.

Silvapulle, M.J. and Sen, P.K. (2005) *Constrained Statistical Inference*. New Jersey: John Wiley.

Zheng, S.R., Shi, N.Z., Guo, J. (2005). “The restricted EM algorithm under linear inequalities in a linear model with missing data”. *Science in China Series A – Mathematics*, **48**, 819-828.