Parameter Estimation in Generalized Linear Models through Modified Maximum Likelihood

Oral, Evrim
Louisiana State University Health Sciences Center SPH, Biostatistics Program
1615 Poydras Street Suite:1400
New Orleans, LA 70112, USA
E-mail: eoral@lsuhsc.edu

1. Introduction

For most of the generalized linear models (GLM), the maximum likelihood (ML) equations involve nonlinear functions of the parameters; thus, they are intractable. Solving these equations by iterations can be problematic for reasons of convergence to wrong values, extremely slow convergence, or non-convergence of the iterations. To alleviate these difficulties, Tiku and Vaughan (1997) and Oral (2005, 2006a) derived modified maximum likelihood (MML) estimators for logit, log-linear and proportional odds models, respectively. In this study, we generalize the estimators given in Tiku and Vaughan (1997) and Oral (2005, 2006a), and provide explicit solutions that is applicable for all GLMs which use canonical link functions. By using real life data sets, we show that the derived estimators are fully efficient for large sample sizes and highly efficient for small samples. We also study the robustness properties of these estimators via simulations.

2. Modified Maximum Likelihood Methodology

The method of MML estimation was originated by Tiku (1967, 1978, 1980) and has been used extensively in literature (Tan and Tabatabai, 1988; Tiku and Suresh, 1992; Oral and Gunay, 2004, Oral 2006b; Oral and Kadilar, 2011). The methodology of MML is employed in situations where the ML estimation is intractable. There are three steps to apply the method: (i) express the likelihood equations in terms of the order statistics, (ii) replace the intractable functions by their linear approximations such that the differences between the two converge to zero as n tends to infinity, and (iii) solve the resulting equations. The solutions, called MML estimators (MMLEs), have closed forms, and are therefore easy to compute. A rigorous proof is available in Vaughan and Tiku (2000) for the fact that, under some regularity conditions, MMLEs have exactly the same asymptotic properties as ML estimators (MLEs), and for small n values they are known to be essentially as efficient as MLEs.

To highlight the methodology, consider the family of skewed distributions

$$f_X(x;\mu,\sigma)=\frac{b}{\sigma}\frac{\exp[-(x-\mu)/\sigma]}{\{1+\exp[-(x-\mu)/\sigma]\}^{b+1}},\ -\infty<x<\infty; \tag{2.1}$$

where b is the shape parameter. Note that $E(X)=\mu+\sigma[\psi(b)-\psi(1)]$ and $V(X)=\sigma^2[\psi'(b)+\psi'(1)]$, where $\psi(x)=\Gamma'(x)/\Gamma(x)$ is the digamma function and $\psi'(x)$ is its derivative. For $b<1$, $b=1$ and $b>1$, (2.1) represents negatively skewed, symmetric and positively skewed distributions, respectively. Given a random sample $X_1,X_2,...,X_n$ from (2.1), we want to estimate the parameters μ and σ. The MLE of μ and σ are the solutions of the likelihood equations

$$\frac{\partial \ln L}{\partial\mu}=\frac{n}{\sigma}-\frac{(b+1)}{\sigma}\sum_{i=1}^{n}\frac{\exp[-(x_i-\mu)/\sigma]}{1+\exp[-(x_i-\mu)/\sigma]}=0, \tag{2.2}$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n}[(x_i-\mu)/\sigma] - \frac{(b+1)}{\sigma}\sum_{i=1}^{n}\frac{[(x_i-\mu)/\sigma]\exp[-(x_i-\mu)/\sigma]}{1+\exp[-(x_i-\mu)/\sigma]} = 0, \quad (2.3)$$

that include nonlinear functions of the parameters, and need to be solved iteratively. However, solving these equations by iterations can be problematic for the reasons given above. MML methodology proceeds as follows:

Since complete sums are invariant to ordering, the likelihood equations (2.2)-(2.3) can be re-written in terms of the ordered statistics $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ as

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0, \quad (2.4)$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0, \quad (2.5)$$

where $g(z_{(i)}) = \exp(-z_{(i)})/(1+\exp(-z_{(i)}))$, $z_{(i)} = (x_{(i)}-\mu)/\sigma$. Linearizing the intractable function $g(z_{(i)})$ around $t_{(i)} = E(z_{(i)})$, $1 \le i \le n$ gives $g(z_{(i)}) \cong \alpha_i - \beta_i z_{(i)}$, where

$$\alpha_i = [1+\exp(t_{(i)}) + t_{(i)}\exp(t_{(i)})]/[1+\exp(t_{(i)})]^2, \quad \beta_i = \exp(t_{(i)})/[1+\exp(t_{(i)})]^2,$$

and $t_{(i)}$ is determined from the equation $t_{(i)} = -\ln(q_i^{-1/b} - 1)$, $q_i = i/(n+1)$. By incorporating the linearized function $g(z_{(i)})$ into the likelihood equations (2.4)-(2.5), the modified likelihood equations are obtained as

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n}(\alpha_i - \beta_i z_{(i)}) = 0, \quad (2.6)$$

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)}(\alpha_i - \beta_i z_{(i)}) = 0. \quad (2.7)$$

The solutions of (2.6)-(2.7) are the MMLEs which are explicit functions of the observations as given below

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \left(B + \sqrt{B^2 + 4nC}\right)\Big/\left(2\sqrt{n(n-1)}\right),$$

where

$$K = \sum_{i=1}^{n} \beta_i x_{(i)} \Big/ m, \quad m = \sum_{i=1}^{n} \beta_i, \quad D = \sum_{i=1}^{n} \Delta_i \Big/ m, \quad \Delta_i = (1/(b+1)) - \alpha_i,$$

$$B = (b+1)\sum_{i=1}^{n} \Delta_i (x_{(i)} - K), \quad C = (b+1)\sum_{i=1}^{n} \beta_i (x_{(i)} - K)^2.$$

Remark: In practice, the shape parameter b in (2.1) may not be known. However, one can easily determine the value of the shape parameter by constructing several Q–Q plots with the observed values. The Q–Q plot that most closely approximates a straight line would be deemed the most appropriate.

3. Modified Maximum Likelihood Estimators for Generalized Linear Models

For a generalized linear model, where Y is the outcome and X_j ($j=1,...p$) are the explanatory variables, the random sample $Y_1, Y_2,..., Y_n$ has a distribution in the exponential family

$$f(y_i, \theta_i, \phi) = \exp\{(y_i\theta_i - b(\theta_i))/a(\phi) + c(y_i, \phi)\}, \quad 1 \le i \le n, \tag{3.1}$$

for some specific functions a, b and c. Suppose that ϕ is known, so that (3.1) is an exponential family model with a canonical parameter θ_i ($1 \le i \le n$). If we let

$$E(Y_i) = \mu_i \quad \text{and} \quad g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij},$$

the likelihood equations for $j=1,...p$ can be written as

$$\sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)x_{ij}}{Var(Y_i)} g'(\mu_i)^{-1} \right] = 0,$$

or, equivalently

$$\sum_{i=1}^{n} \left[\frac{(y_i - b'(\theta_i))x_{ij}}{a(\phi)b''(\theta_i)g'(b'(\theta_i))} \right] = 0, \text{ for } j=1,...p. \tag{3.2}$$

When the canonical link function is used, equations (3.2) become

$$\sum_{i=1}^{n} \left[\frac{(y_i - b'(\theta_i))}{a(\phi)} x_{ij} \right] = 0, \text{ for } j=1,...p, \tag{3.3}$$

and do not have explicit solutions. To derive the MMLEs, we assume the mean of the outcome depends on a single explanatory covariate X, that is, $\theta_i = \gamma_0 + \gamma_1 x_i$, $1 \le i \le n$. The likelihood equations for estimating γ_0 and γ_1 are written in terms of the ordered statistics $\theta_{(i)} = \gamma_0 + \gamma_1 x_{[i]}$ as

$$\frac{\partial \ln L}{\partial \gamma_0} = \sum_{i=1}^{n} (y_{[i]} - b'(\theta_{(i)})) = 0$$

$$\frac{\partial \ln L}{\partial \gamma_1} = \sum_{i=1}^{n} [(y_{[i]} - b'(\theta_{(i)}))x_{[i]}] = 0$$

where $(y_{[i]}, x_{[i]})$ pair is the (y_i, x_i) observation (concomitant) which corresponds to $\theta_{(i)}$, $1 \le i \le n$. By using the procedure described in Section 2, we obtain the MMLEs as follows:

$$\hat{\gamma}_0 = \Delta - (\hat{\gamma}_1 \bar{x}_a), \text{ and } \hat{\gamma}_1 = \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_a) \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_a)^2, \tag{3.4}$$

where

$$\Delta = \delta/m\,,\ \ \delta = \sum_{i=1}^{n}\delta_i\,,\ \ \delta_i = y_{[i]} - \alpha_i\,,\ \ m = \sum_{i=1}^{n}\beta_i\,,\ \ \bar{x}_a = \left(\sum_{i=1}^{n}\beta_i x_{[i]}\right)\Big/m\,, \tag{3.5}$$

$$\alpha_i = b'(t_{(i)}) - t_{(i)}\beta_i\,,\ \ \beta_i = b''(t_{(i)})\,,\ \ t_{(i)} = E(\theta_{(i)})\,. \tag{3.6}$$

As an example, for the logistic regression model, since $\theta_i = \ln[\mu_i/(1-\mu_i)]$, $b'(\theta_i) = \exp(\theta_i)/(1+\exp(\theta_i))$, and $b''(\theta_i) = \exp(\theta_i)/(1+\exp(\theta_i))^2$, α_i and β_i ($1 \le i \le n$) values in (3.6) become

$$\alpha_i = \left[\exp(t_{(i)})/(1+\exp(t_{(i)}))\right] - t_{(i)}\beta_i \ \text{ and } \ \beta_i = \exp(t_{(i)})/(1+\exp(t_{(i)}))^2\,. \tag{3.7}$$

As another example, for the log-linear model, $\theta_i = \ln(\mu_i)$, $b'(\theta_i) = b''(\theta_i) = \exp(\theta_i)$; thus, α_i and β_i ($1 \le i \le n$) values in (3.6) become

$$\alpha_i = \exp(t_{(i)})(1 - t_{(i)}) \ \text{ and } \ \beta_i = \exp(t_{(i)})\,.$$

Note that, the initial values of $t_{(i)}$ can be taken as $t_{(i)} = \tilde{\gamma}_0 + \tilde{\gamma}_1 x_{[i]}$ where $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are the least squares estimators

$$\tilde{\gamma}_0 = \bar{y} - \tilde{\gamma}_1\bar{x} \ \text{ and } \ \tilde{\gamma}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i\Big/(x_i - \bar{x})^2\,.$$

After the first iteration, the estimators needed to be revised by replacing $t_{(i)}$ by $t_{(i)} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{[i]}$ from (3.4). This process may be repeated a few times to obtain the final estimates of $\hat{\gamma}_0$ and $\hat{\gamma}_1$.

Examples: To illustrate the methodology, we first analyze the coronary heart disease data given on page 3 of Hosmer and Lemeshow (1989). This data represents the values of coronary heart disease (CHD) status Y (0 or 1), and the corresponding values of the age X of 100 subjects. We also consider the data given on page 82 of Agresti (1996), which is from a study of nesting horseshoe crabs where the response Y is the number of satellites that each female crab has, and the corresponding values of the covariate X is the carapace width of 173 crabs. Both studies investigate the relationship between Y and X. For both data, we calculated the MML estimates from equations (3.4)-(3.6). The calculations given in Table 1 are completely consistent with those given in Hosmer and Lemeshow (1989) and Agresti (1996).

Table 1 Estimates for CHD and horseshoe crabs data

	CHD Data		*Horseshoe crabs Data*	
	Coefficient	Estimate	Coefficient	Estimate
ML	γ_0	-5.310	γ_0	-3.3048
	γ_1	0.111	γ_1	0.1640
	Coefficient	Estimate	Coefficient	Estimate
MML	γ_0	-5.309	γ_0	-3.3047
	γ_1	0.111	γ_1	0.1640

4. Robustness Properties of the Estimators

From a practical point of view, it is very important for an estimator to have efficiency robustness. Such an estimator is fully efficient (or nearly so) for an assumed model and maintains high efficiency for plausible alternatives. In practice, specifically outliers are a frequently encountered problem in GLM; thus, in this section we search the robustness properties of the estimators given in (3.4)-(3.6) with respect to the outliers.

For illustration, we consider the log-linear model and perform a Monte-Carlo study as follows: We assume that $\gamma_0 = -1$, and for three different values of γ_1 (0.0, 0.5, and 1.0), we generate (n-r) of the $X_1, X_2, ..., X_n$ observations from the Normal distribution with parameter σ, and the remaining r (we don't know which) from the Normal distribution with parameter $d\sigma$ (d is a positive constant). We calculate r from the formula $r = [0.1n + 0.5]$. We assume that $\mu = 0, \sigma = 1$ without loss of generality and consider the models below:

a) (n-r) come from N(0,1) and r come from N(0,1) (No outliers),
b) (n-r) come from N(0,1) and r come from N(0,1.5),
c) (n-r) come from N(0,1) and r come from N(0,2),
d) (n-r) come from N(0,1) and r come from N(0, 4).

The model (a) above is the model without outliers and is given for sake of comparisons. Note that for each model, after generating the X values, we calculated $\theta_i = \gamma_0 + \gamma_1 x_i$ and $\mu_i = \exp(\theta_i)$ for $1 \le i \le n$ to generate Y_i values from Poisson(μ_i). The values obtained from [100000/n] iterations are given in Table 2. As can be seen from the table, the biases in the estimates are negligible for all models. The variances $V(\hat{\gamma}_1)$ are almost the same for a given n for the models (a), (b), (c) and (d). We conclude that the MMLEs (3.4)-(3.6) are fairly robust to outliers.

Table 2 Simulation results for models (a)-(d)

		(a): No Outlier		(b): d=1.5	
γ_1	n	Bias($\hat{\gamma}_1$)	V($\hat{\gamma}_1$)	Bias($\hat{\gamma}_1$)	V($\hat{\gamma}_1$)
0	30	0.011	0.119	0.012	0.118
	50	0.004	0.062	0.006	0.062
	100	0.005	0.029	0.004	0.029
0.5	30	0.014	0.112	0.014	0.115
	50	0.007	0.057	0.004	0.058
	100	0.002	0.026	0.002	0.026
1.0	30	0.020	0.098	0.019	0.099
	50	0.019	0.047	0.007	0.047
	100	0.004	0.020	0.004	0.019
		(c): d=2.0		(d): d=4.0	
γ_1	n	Bias($\hat{\gamma}_1$)	V($\hat{\gamma}_1$)	Bias($\hat{\gamma}_1$)	V($\hat{\gamma}_1$)
0	30	0.013	0.120	0.009	0.122
	50	0.002	0.064	0.008	0.065
	100	0.005	0.029	0.001	0.028
0.5	30	0.012	0.113	0.010	0.117
	50	0.003	0.057	0.005	0.055
	100	0.000	0.025	0.007	0.024
1.0	30	0.017	0.098	0.019	0.098
	50	0.005	0.045	0.009	0.044
	100	0.005	0.017	0.005	0.016

5. Concluding Remarks and Future Work

The values of the parameters in GLM are usually obtained by ML estimation, which require iterative computational procedures. There are many iterative methods for ML estimation in the generalized linear models, of which the Newton-Raphson and Fisher scoring methods are among the most widely used ones. Using iterative methods, however, can be problematic (Vaughan, 2002; Tiku and Vaughan, 1997). In this study, we generalize the estimators given in Tiku and Vaughan (1997) and Oral (2005, 2006a), and provide explicit solutions that is applicable for all GLMs which use canonical link functions. We also study the robustness properties of these estimators via simulations. We are currently working on generalizing the method to multivariable situations and also to the case where $a(\phi)$ in (3.1) is also a parameter.

REFERENCES

Agresti, A., 1996, Categorical Data Analysis, John Wiley and Sons, New York, 558p.

Hosmer, D. W. and Lemeshow, S., 1989, Applied Logistic Regression, John Wiley and Sons, New York, 373p.

Tan, W.Y., Tabatabai, M.A., 1988, A modified Winsorized regression procedure for linear models, Journal of Statistical Computation and Simulation 30, 299-313.

Tiku, M.L., 1967, Estimating the mean and standard deviation from a censored normal sample, Biometrika 54, 155-165.

Tiku, M.L., 1978, Linear regression model with censored observations, Communications in Statistics: Theory and Methods A7, 1219-1232.

Tiku, M.L., 1980, Robustness of MML estimators based on censored samples and robust test statistics, Journal of Statistical Planning and Inference 4, 123-143.

Tiku, M.L., Suresh, R.P., 1992, A new method of estimation for location and scale parameters. Journal of Statistical Planning and Inference 30, 281-292.

Tiku, M.L., Vaughan, D.C., 1997, Logistic and non-logistic density functions in binary regression with non-stochastic covariates. Biometrical Journal 39, 883-898.

Oral, E., 2005, Parameter estimation in poisson regression via modified maximum likelihood method, 2005 Proceedings of the American Statistical Association, Biometrics Section [CD-ROM], Alexandria, VA: ASA, 347-355.

Oral, E., 2006a, Modified maximum likelihood estimation of the proportional odds model, 2006 Proceedings of the American Statistical Association, Biometrics Section [CD-ROM], Alexandria, VA: ASA, 307-314.

Oral, E., 2006b, Binary regression with stochastic covariates, Communications in Statistics: Theory and Methods 35, 1429-1447.

Oral E. and Gunay S., 2004, Stochastic Covariates in Binary Regression, Hacettepe Journal of Mathematics and Statistics, 33, 97-109.

Oral E. and Kadilar C., 2011, Robust ratio-type Estimators in Simple Random Sampling, article in press, Journal of the Korean Statistical Society, doi: 10.1016/j.jkss.2011.04.001.

Vaughan D. C. and Tiku M. L. (2000) Estimation and hypothesis testing for nonnormal bivariate distribution with applications, Mathematical and Computer Modeling 32, 53-67.

Vaughan, D. C., 2002, The generalized secant hyperbolic distribution and its properties, Communications in Statistics-Theory and Methods, 31, 219-238.