行政院國家科學委員會專題研究計畫　成果報告

非線性模式中使用應變數訊息的迴歸校正估計方法

計畫類別：個別型計畫
計畫編號：NSC91-2118-M-032-009-
執行期間：91年08月01日至92年07月31日
執行單位：淡江大學數學系

計畫主持人：黃逸輝

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 92年10月31日

Simple estimation in logistic regression when covariates are subject to measurement errors

STEVE Y. H. HUANG

Tamkang University

Abstract: This paper studies estimation in functional logistic regression when covariate are subject to measurement errors. We introduce the estimating equation come from estimating a weighted score function. Due to the weighting, the derivation of the consistent estimator and its variance estimator becomes simple and easy. The modification to accommodate other parametric error assumption seems straightforward. Simulation study with the comparison to the sufficient estimator in Stefanski & Carroll (85) is provided. Further simulation study reveals that the estimation without extra information is possible by a natural extension of the estimating equations.

1 Introduction

Logistic regression is an useful method for analyzing relationship between covariates and binary response variable. For independent observations (Y_i, X_i), it assumes that

$$P(Y_i = 1 \mid X_i) = F(\beta_0 + \beta_1 X_i) \triangleq (1 + exp(-\beta_0 - \beta_1 X_i))^{-1}, i = 1, \ldots, n. \tag{1.1}$$

It may happen that in stead of observing the true X_i, one can only observed its surrogates W_i, where W_i equals X_i plus the measurement error, i.e., $W_i = X_i + \delta_i$. Many examples exhibit such logistic regression with additive measurement error can be found in Carroll et al. (1995).

When the X_i is treated as i.i.d. random variables, we say we have a structural model, and we say we have a functional model if X_i is treated as fixed but unknown constant or parameter. In either case, the logistic regression analysis using pairs (Y_i, W_i), which produces the naive estimators, does not give correct inference about β_0 and β_1. In structural model, perhaps the easiest situation is when both X_i and δ_i are normally distributed, then the following approximation holds.

$$E(Y_i \mid W_i) = E(F(\beta_0 + \beta_1 X_i) \mid W)) \approx E(\Phi(c(\beta_0 + \beta_1 X_i) \mid W_i)$$

$$= \Phi(c(\beta_0^* + \beta_1^* W_i)) \approx F(\beta_0^* + \beta_1^* W_i),$$

where β_0^*, β_1^* are functions of β_0, β_1, σ^2 and moments of X. The logistic regression still holds (only approximately) with parameters changed. Beside the normal assumption, another easy method called Regression Calibration is applicable if some validation data are available, but may still needs some approximation and makes the resultant estimator being only nearly consistent. In functional model, the problem becomes more troublesome since the unknown parameters are more than the sample size.

When the measurement error's variance is known or estimable, Stefanski and Carroll (1985) proposed three estimatior, a bias-adjusted estimator and two appropriate estimator for normally distributed measurement error. The bias-adjusted estimator and one of the other two estimator, which in fact is a kind of Regression Calibration, developed their consistency based on the assumption of errors diminish as samples size increases. The remaining estimator which is derived from the sufficient score does not assume small error but normality on measurement errors. In this paper, we propose an estimation method needing the same assumption as sufficient score, but the author believe that the method can be easily modified to accommodate other parametric assumption of error, and can extend to the case when error's variance is unknown.

In section 2 we motivate our method and propose a class of estimators. Section 3 discusses how to choose the weight function. Section 4 contains a simulation study when σ^2 is estimated. Section 5 discusses the possibility of extending the proposed method to the case when σ^2 is unknown, and provide some simulation study when σ^2 is unknown. A final conclusion is given in section 6.

2 Weighted score and estimation

Hereafter we treat X_i as unknown fixed parameter. When no measurement errors presents, the score function for β_0 and β_1 in (1.1) is

$$\sum_{i=1}^{n}(Y_i - \frac{1}{1+e^{-\beta_0-\beta_1 X_i}}) \begin{pmatrix} 1 \\ X_i \end{pmatrix} \qquad (2,1)$$

In (2.1), the score function involve the conditional mean of Y, which is a fraction and not wield to handle, and hence not easy to find any corrected score function to replace it (Nakaruma, 1990). Here we invoke a weighted score function to make the unwieldy term disappeared. For example, one can consider the function

$$\sum_{i=1}^{n}(1+e^{-\beta_0-\beta_1 X_i})(Y_i - \frac{1}{1+e^{-\beta_0-\beta_1 X_i}}) \begin{pmatrix} 1 \\ X_i \end{pmatrix} = \sum_{i=1}^{n}(Y_i - 1 + Y_i e^{-\beta_0-\beta_1 X_i}) \begin{pmatrix} 1 \\ X_i \end{pmatrix}.$$

Apparently, this weighted score contains no fraction terms and seems easier to handle. However the weighting is not unique, for example one can replace $\frac{1}{1+exp(-\beta_0-\beta_1 X_i)}$ by $\frac{exp(\beta_0+\beta_1 X_i)}{1+exp(\beta_0+\beta_1 X_i)}$ and consider another weighted score function

$$\sum_{i=1}^{n}(1+e^{\beta_0+\beta_1 X_i})(Y_i - \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}})\begin{pmatrix}1\\X_i\end{pmatrix} = \sum_{i=1}^{n}(Y_i+(Y_i-1)e^{\beta_0+\beta_1 X_i})\begin{pmatrix}1\\X_i\end{pmatrix}.$$

In general, the conditional mean can be represented as $\frac{exp(\beta_0+(\beta_1-k)X_i)}{exp(-kX)+exp(\beta_0+(\beta_1-k)X_i)}$ with any constant k. A class of weighted score function is thus define as

$$\sum_{i=1}^{n} S_i = \sum_{i=1}^{n}(e^{-kX}+e^{\beta_0+(\beta_1-k)X_i})(Y_i - \frac{e^{\beta_0+(\beta_1-k)X_i}}{e^{-kX}+e^{\beta_0+(\beta_1-k)X_i}})\begin{pmatrix}1\\X_i\end{pmatrix}.$$

$$= \sum_{i=1}^{n}\begin{pmatrix} Y_i e^{-kX_i}+(Y_i-1)e^{\beta_0+(\beta_1-k)X_i} \\ X_i Y e^{-kX_i}+X_i(Y_i-1)e^{\beta_0+(\beta_1-k)X_i}\end{pmatrix}, \qquad (2.2)$$

where k can be any constant. If there is parametric assumption about δ, then the estimations of every terms in (2.2) by functions of (Y_i, W_i) become possible and easy. If we assume that $\delta_i \sim N(0, \sigma^2)$, then the unbiased estimate of every terms in (2.2) are ready by observing that

$$E(e^{kW-\frac{1}{2}k^2\sigma^2}) = e^{kX}, \quad E(e^{kW-\frac{1}{2}k^2\sigma^2}(W-k\sigma^2)) = Xe^{kX} \qquad (2.3).$$

The first equality come from the moment generating function of normal distribution, and the second are derived by differentiating the first equation with respect to k. Consequently, replace every terms in (2.2) with their "estimates", we have the estimating function

$$\sum_{i=1}^{n} T_i \equiv \sum_{i=1}^{n}\begin{pmatrix} Y_i e^{-kW_i-0.5k^2\sigma^2}+(Y_i-1)e^{\beta_0+(\beta_1-k)W_i-0.5(\beta_1-k)^2\sigma^2} \\ Ye^{-kW_i-0.5k^2\sigma^2}(W_i+k^2\sigma^2)+(Y_i-1)e^{\beta_0-(\beta_1-k)W_i-0.5(\beta_1-k)^2\sigma^2}(W_i-(\beta_1-k)\sigma^2)\end{pmatrix}. \qquad (2.4)$$

From the surrogate assumption, that is Y_i is independent of W_i (conditional on X_i). It is easy to shown that the $ET_i = ES_i$, and hence $\sum_{i=1}^{n} T_i$ is a mean zero estimating function for any k. Denote any solution of (2.4) by $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$ if there is one. Then from the first row of (2.4), it can be shown that

$$e^{\hat{\beta}_0} = \frac{\sum(Y_i e^{-kW-0.5k^2\sigma^2})}{\sum(1-Y)e^{(\hat{\beta}_1-k)W-0.5(\hat{\beta}_1-k)^2\sigma^2}}.$$

Denote $C_{1n} = \sum(Y_i e^{-kW-0.5k^2\sigma^2})/n, C_{2n}(s) = \sum(1-Y)e^{(s-k)W-0.5(s-k)^2\sigma^2}/n$. Replace $e^{\hat{\beta}_0}$ by $C_{1n}/C_{2n}(\hat{\beta}_1)$ in the 2nd row of (2.4). We know that $\hat{\beta}_1$ is the root of the equation $\psi_n(s) = 0$, where

$$\psi_n(s) = \frac{1}{n}\sum_{i=1}^{n}(Ye^{-kW_i-0.5k^2\sigma^2}(W_i + k^2\sigma^2) + \frac{C_{1n}}{C_{2n}(s)}(Y_i - 1)e^{-(s-k)W_i-0.5(s-k)^2\sigma^2}(W_i - (s-k)\sigma^2)). \tag{2.5}$$

Hence the calculation become simple by solving $\psi(s) = 0$. Let $\hat{\beta}_1$ be the root, then $\hat{\beta}_0$ is known to be $\log \frac{C_{1n}}{C_{2n}(\hat{\beta}_1)}$. Consistence of $\hat{\boldsymbol{\beta}}$ can be proved if some conditions on $x_i's$ are provided.

Condition: Let $M_n(s) = \frac{1}{n}\sum_{i=1}^{n} e^{sX_i}$. We assume there is a function $M(s)$ such that $M_n(s) \to M(s)$, $\forall s \in R$, and $\frac{d^k M_n(s)}{ds^k}(= \frac{1}{n}\sum X_i^k e^{sX_i}) \to \frac{d^k M(s)}{ds^k}$ for $k = 1, 2$, as $n \to \infty$.

Theorem 1. Suppose the above condition holds, then for any fixed constant k, the equation $\psi_n(s) = 0$ determining a root that converge to β_1 almost surely.

pf: see appendix.

As a consequence of theorem 1, $\hat{\beta}_0$ converge to β_0 almost surely, too. Let $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, and the estimators derived from solving $\sum_{i=1}^{n} T_i = \mathbf{0}$ be $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$. Apparently, different choice of k corresponds to different estimates $\hat{\boldsymbol{\beta}}$, and hence their variance are also dependent on k.

To calculate the asymptotic variance of $\hat{\beta}_0$ and $\hat{\beta}_1$, we adopt the sandwich method (Appendix A, Carroll, et al., 1995). Denote T_i by $T_i = (T_{1i} + T_{2i}, T_{3i} + T_{4i})'$, where $T_{1i} =$

$Y_ie^{-kW_i-0.5k^2\sigma^2}$, $T_{2i}=(Y_i-1)e^{\beta_0+(\beta_1-k)W_i-0.5(\beta_1-k)^2\sigma^2}$, $T_{3i}=Ye^{-kW_i-0.5k^2\sigma^2}(W_i+k^2\sigma^2)$, and $T_{4i}=(Y_i-1)e^{-kW_i-0.5k^2\sigma^2}(W_i-(\beta_1-k)\sigma^2)$. By the definition of $\hat{\boldsymbol{\beta}}$, we knows that

$$\sum_{i=1}^{n}\begin{pmatrix} T_{1i}+T_{2i} \\ T_{3i}+T_{4i} \end{pmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}=\mathbf{0},$$

and hence under some regular condition, we have

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'=(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial T_i}{\partial\boldsymbol{\beta}})^{-1}\frac{1}{n}\sum_{i=1}^{n}T_i+O(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^2,$$

and

$$Var(\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})')=(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial T_i}{\partial\boldsymbol{\beta}})^{-1}(\frac{1}{n}\sum_{i=1}^{n}ET_iT_i')(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial T_i}{\partial\boldsymbol{\beta}})^{-1'}+o(1).$$

Hence one can evaluate the matrix

$$(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial T_i}{\partial\boldsymbol{\beta}})^{-1}(\frac{1}{n}\sum_{i=1}^{n}T_iT_i')(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial T_i}{\partial\boldsymbol{\beta}})^{-1'} \tag{2.6}$$

at $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}$ as an estimates of covariance matrix of $\hat{\boldsymbol{\beta}}$.

A reasonable way to choose k is to minimize certain function of the estimated covariance matrix. For example, one can choose one k to minimize the variance of $\hat{\beta}_0$ and another k to minimize the estimated variance of $\hat{\beta}_1$. However doing this needs much computation since there are infinitely many choices of k. If little lost in efficiency is tolerable, we suggest to use the method in the following section to determine a suitable value of k.

3 Determination of constant k

The right-hand side of (2.4) is an estimate of weighted score function (2.2), where the weights in (2.2) are $e^{-kX}+e^{\beta_0+\beta_1X}$. We believe that (2.4) will be more close to the (unweighted) score function, which is the optimal estimation function when no errors present, if the weights can be choose as the same as possible. Hence we are led to seek k that can minimize the function

$$\sum_{i=1}^{n}(\frac{e^{-kX}+e^{\beta_0+\beta_1X}}{\sum_{i=1}^{n}(e^{-kX}+e^{\beta_0+\beta_1X})}-\frac{1}{n})^2. \tag{2.7}$$

This function depends on the unknown $X's$. In order to have practical results, we make normal assumptions and utilize some approximation as follows.

Denote the quantity $e^{-kX_i} + e^{\beta_0+\beta_1 X_i}$ by η_i and define $\bar{\eta}$ as $\frac{\sum_{i=1}^n \eta_i}{n}$, then to minimize (2.7) is equivalent to minimize $\frac{1}{n}\sum_{i=1}^n (\frac{\eta_i}{\bar{\eta}_i} - 1)^2$. This quantity is approximately equal to

$$E(\frac{\eta_i}{E(\eta_i)} - 1)^2 = \frac{E\eta_i^2}{(E\eta_i)^2} - 1. \tag{2.8}$$

If k is the minimizer of (2.8), then from differentiating the right hand side of (2.8), we know that k satisfy

$$E(\eta)E(\eta\eta^{'}) = E(\eta^2)E(\eta^{'}), \text{ where } \eta^{'} = \frac{d\eta}{dk}. \tag{2.9}$$

Assume that X is standard normal, then we can solve this equation by numerical method easily.

Note that equation for determining k are only suitable for standard normal. When X is distributed as $N(\mu_X, \sigma_x^2)$, one can rewrite the model

$$E(Y \mid X_i) = \frac{1}{e^{-\beta_0-\beta_1}} \text{ as } E(Y \mid X_i) = \frac{1}{e^{-\beta_0^*-\beta_1^*\frac{(X-\mu_x)}{\sigma_X}}},$$

where $\beta_0^* = \beta_0 + \beta_1\frac{\mu}{\sigma_X}, \beta_1^* = \sigma_X\beta_1$, In conclusion, we suggest the following procedure for estimation.

1. Initially set k to 0, and solve (2.5) for initial estimates of β_1, and hence β_0.

2. Use W_i and the knowledge of σ^2 to estimate the mean and standard deviation of X.

3. Rewrite the logistic model in terms of standardized X, and use current estimates of β_0,β_1,μ_X, and σ_X^2 to calculate estimates of β_0^* and β_1^*.

4. Solve (2.9) for k corresponds to current estimates of β_0^* and β_1^*.

5. Divide k by standard deviation of X, and solve (2.5) to update the estimates of β_1 and β_0. Go back to 3 until estimates converge.

If we view k, the solution of (2.9), as a function of β_0 and β_1, then an approximation of this function is a linear function of β_0 and β_1. A regression analysis show that the approximation

$$k \approx 0.088 + 0.138\beta_0 + 0.423\beta_1, \quad \text{for } \beta_0 < 0, \beta_1 < 0 \tag{2.10}$$

is fairly good with R-square 0.966. As an alternative of solution to "(2.9)=0", one can use this regression function as a convenient choice of k

4 Simulation Study

We conducted some simulation analysis to study the performance of proposed estimator. We adopt the setting used in Stefanski & Carroll (1985), and set

$$E(Y_i \mid X_i) = \frac{1}{1 + e^{-1.4-1.4X_i}}, \quad \begin{pmatrix} W_{i1} \\ W_{i2} \end{pmatrix} = X_i + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix},$$

where X_i are drew from $N(0, 0.1)$ or $\frac{\chi_1^2 - 1}{\sqrt{2}}\sqrt{0.1}$, and δ_{ij} are from $N(0, \frac{0.1}{3})$ or contaminated normal, which is $N(0, \frac{0.1}{3})$ with probability 0.9 and $N(0, \frac{2.5}{3})$ with probability 0.1. Let $W_i = \frac{W_{i1}+W_{i2}}{2}$, then W_i is a surrogate of X_i and the measurement error variance can be estimated by the sample variance of $\frac{W_{i1}-W_{i2}}{2}$.

We record the estimates of β_1 only. The subscript "naive" means naive estimator; "$k = \beta$" means that k is set at β in the weights; " Opt-k" means that k is chosen as the root of "(2.9) = 0"; " k-reg" means that k is derived by (2.10) with β estimated by $\hat{\beta}_{k=\beta}$; "suff" means the one step sufficient estimator in Stefanski & Carroll (85).

From table 1 to table 4 we can see that naive estimator is severely biased in most cases. The performance of $\hat{\beta}_{Opt-k}$ and $\hat{\beta}_{k-reg}$ differ little. This means that the approximation in (2.10) are quite good. The proposed estimator $\hat{\beta}_{Opt-k}$ and $\hat{\beta}_{k-reg}$ are less efficiency than $\hat{\beta}_{suff}$ which is optimal in some structural model, but the lost is very minor. In the case of contaminated error with $n = 1500$, our estimator is superior than the sufficient estimator.

Table 1. $X \sim \frac{\chi_1^2 - 1}{\sqrt{2}}\sqrt{0.1}, \quad \delta \sim N(0, \frac{0.1}{3})$

N=300	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.171	-0.104	0.088	0.077	-0.025
Std. Dev	0.432	0.673	0.530	0.496	0.488
Mse	0.216	0.462	0.287	0.251	0.238

N=600	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.136	0.013	0.075	0.066	0.015
Std. Dev	0.281	0.412	0.339	0.322	0.315
Mse	0.097	0.170	0.120	0.107	0.099

N=1,500	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.175	-0.022	0.008	0.000	-0.034
Std. Dev	0.176	0.245	0.219	0.208	0.198
Mse	0.062	0.061	0.048	0.043	0.040

Table 2. $X \sim N(0, 0.1), \quad \delta \sim N(0, \frac{0.1}{3})$

N=300	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.157	0.069	0.084	0.094	0.032
Std. Dev	0.433	0.574	0.522	0.520	0.503
Mse	0.212	0.333	0.278	0.279	0.254

N=600	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.209	0.002	0.005	0.012	-0.033
Std. Dev	0.309	0.413	0.371	0.369	0.358
Mse	0.139	0.170	0.137	0.136	0.129

N=1,500	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suf-f}$
Bias	-0.213	-0.003	-0.004	-0.001	-0.038
Std. Dev	0.198	0.266	0.240	0.239	0.231
Mse	0.084	0.070	0.058	0.057	0.054

Table 3. $X \sim \frac{\chi_1^2-1}{\sqrt{2}}\sqrt{0.1}, \quad \delta \sim 0.9N(0, \frac{0.1}{3}) + 0.1N(0, \frac{2.5}{3})$

N=300	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.385	-0.228	0.146	0.233	-0.065
Std. Dev	0.361	1.344	0.528	0.600	0.472
Mse	0.277	1.840	0.297	0.411	0.225

N=600	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.423	-0.377	0.032	0.078	-0.125
Std. Dev	0.253	1.069	0.334	0.384	0.325
Mse	0.242	1.274	0.116	0.152	0.120

N=1,500	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.453	-0.255	-0.027	-0.007	-0.161
Std. Dev	0.137	0.642	0.213	0.208	0.177
Mse	0.224	0.473	0.046	0.043	0.057

Table 4. $X \sim N(0, \frac{0.1}{3})$, $\delta \sim 0.9N(0, \frac{0.1}{3}) + 0.1N(0, \frac{2.5}{3})$

N=300	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.499	0.017	0.030	0.095	-0.161
Std. Dev	0.375	0.849	0.611	0.631	0.526
Mse	0.389	0.719	0.373	0.405	0.302

N=600	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.496	-0.039	0.020	0.064	-0.154
Std. Dev	0.260	0.684	0.412	0.432	0.366
Mse	0.313	0.467	0.175	0.190	0.157

N=1,500	$\hat{\beta}_{Naive}$	$\hat{\beta}_{k=\beta}$	$\hat{\beta}_{Opt-k}$	$\hat{\beta}_{k-reg}$	$\hat{\beta}_{suff}$
Bias	-0.506	-0.071	-0.023	0.009	-0.172
Std. Dev	0.172	0.404	0.270	0.279	0.240
Mse	0.286	0.168	0.073	0.078	0.087

5 Estimation when information of σ^2 is not available

It is possible to extend the estimation method to the situation when no extra information is available. Recall that the weighted score function for β_0, β_1 in (2.2) are

$$\sum_{i=1}^{n} S_i = \sum_{i=1}^{n}(e^{-kX_i} + e^{\beta_0+(\beta_1-k)X_i})(Y_i - \frac{e^{\beta_0+(\beta_1-k)X_i}}{e^{-kX_i} + e^{\beta_0+(\beta_1-k)X_i}})\begin{pmatrix} 1 \\ X_i \end{pmatrix}$$

A natural extension of this equation is

$$\sum_{i=1}^{n}(e^{-kX_i} + e^{\beta_0+(\beta_1-k)X_i})(Y_i - \frac{e^{\beta_0+(\beta_1-k)X_i}}{e^{-kX_i} + e^{\beta_0+(\beta_1-k)X_i}})\begin{pmatrix} 1 \\ X_i \\ X_i^2 \end{pmatrix}$$

$$= \sum_{i=1}^{n} \begin{pmatrix} Y_i e^{-kX_i} + (Y_i - 1)e^{\beta_0+(\beta_1-k)X_i} \\ X_i Y e^{-kX_i} + X_i(Y_i - 1)e^{\beta_0+(\beta_1-k)X_i} \\ X_i^2 Y e^{-kX_i} + X_i^2(Y_i - 1)e^{\beta_0+(\beta_1-k)X_i^2} \end{pmatrix}.$$

Since $E(e^{kW_i - \frac{k^2\sigma^2}{2}}) = e^{kX}$, hence $\frac{d^2}{dk^2} e^{kW_i - \frac{k^2\sigma^2}{2}} = [W_i^2 - 2W_i k\sigma^2 + k^2\sigma^4 - \sigma^2]e^{kW_i - \frac{1}{2}k^2\sigma^2}$ has expectation $X^2 e^{kX}$. Combine this with (2.3), we found that

$$\sum_{i=1}^{n} \begin{pmatrix} Y_i e^{-kW_i - 0.5k^2\sigma^2} + (Y_i - 1)e^{\beta_0+(\beta_1-k)W_i - 0.5(\beta_1-k)^2\sigma^2} \\ Y e^{-kW_i - 0.5k^2\sigma^2}(W_i + k^2\sigma^2) + (Y_i - 1)e^{\beta_0-(\beta_1-k)W_i - 0.5(\beta_1-k)^2\sigma^2}(W_i - (\beta_1 - k)\sigma^2) \\ Y(W^2 + 2Wk\sigma^2 + k^2\sigma^4 - \sigma^2)e^{-kw - \frac{1}{2}k2\sigma^2} \\ +(Y_i - 1)(W_i^2 - 2W_i(\beta_1 - k)\sigma^2 + (\beta_1 - k)^2\sigma^4 - \sigma^2)e^{\beta_0+(\beta_1-k)W_i - \frac{1}{2}(\beta_1-k)^2\sigma^2} \end{pmatrix} \tag{5.1}$$

is zero-unbiased estimating function for β_0, β_1 and σ^2. Hence it is possible to derive consistent estimators of β_0, β_1 and σ^2 from equating (5.1) to 0. However a numerical method is needed in the derivation. If k is set at $\frac{\beta_1}{2}$ then the equation "(5.1) = 0" can be simplified to the following equations.

$$\begin{aligned} e^{\beta_0} &= \frac{\sum_{i=1}^{n} Y_i e^{-\frac{\beta_1}{2}W}}{\sum_{i=1}^{n}(1 - Y)e^{\frac{\beta_1}{2}W}} \\ \sigma^2 &= \frac{1}{\beta_1} \frac{\sum_{i=1}^{n} W^2(1 - Y)e^{\beta_0+\frac{\beta_1}{2}W} - \sum_{i=1}^{n} YW^2 e^{-\frac{\beta_1}{2}W}}{\sum_{i=1}^{n} YWe^{-\frac{\beta_1}{2}W} + \sum_{i=1}^{n}(1 - Y)We^{\beta_0+\frac{\beta_1}{2}W}} \end{aligned}$$

and β_1 satisfy

$$\sum_{i=1}^{n} Y(W + \frac{\beta_1}{2}\sigma^2)e^{-\frac{\beta_1}{2}W} = \sum_{i=1}^{n}(1 - Y)e^{\beta_0+\frac{\beta_1}{2}}(W - \frac{\beta_1}{2}\sigma^2).$$

We have conducted some simulations to see if the above equations really yields consistent estimators. We adopt the same setting in the previous section but discard the case when X is normal because of severe multiple root problems. We also add the case when X is distributed as $1 - Uniform(0, 1)^2$.

Note that the performance of sufficient estimator was also recorded to see how much efficiency is lost when extra information is not available.

From these tables, we see that the bias and the variance shrink as the sample size increases. Both distribution of X demonstrate such phenomenon. They indicate that estimation without extra information in functional logistic model is possible.

Table 5. $X \sim \frac{(\chi_1^2-1)}{\sqrt{2}}\sqrt{0.1}, \quad \sigma^2 = 0.0167$

N=300	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	-0.013	-0.16	-0.014	-0.015	-0.012	0.108
Std. Dev	0.150	0.399	0.150	0.449	0.150	0.620
Estimated variance			$\hat{\sigma}^2 = 0.02384$ (0.0204)			

· The parenthesis follows $\hat{\sigma}^2$ is the standard deviation of $\hat{\sigma}^2$

Table 6. $X \sim \frac{(\chi_1^2-1)}{\sqrt{2}}\sqrt{0.1}, \quad \sigma^2 = 0.0167$

N=600	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	-0.012	-0.175	-0.014	-0.029	-0.013	0.051
Std. Dev	0.105	0.295	0.106	0.333	0.106	0.451
Estimated variance			$\hat{\sigma}^2 = 0.021$ (0.0183)			

Table 7. $X \sim \frac{(\chi_1^2-1)}{\sqrt{2}}\sqrt{0.1}, \quad \sigma^2 = 0.0167$

N=1500	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	0.004	-0.177	0.002	-0.036	0.002	0.017
Std. Dev	0.063	0.179	0.063	0.202	0.064	0.272
Estimated variance			$\hat{\sigma}^2 = 0.0186$ (0.0129)			

Table 8. $X \sim (1 - U(0,1)^2)$, $\sigma^2 = 0.0167$

N=300	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	0.110	-0.185	-0.036	0.025	-0.108	0.125
Std. Dev	0.305	0.392	0.357	0.467	0.658	0.938

Estimated variance $\hat{\sigma}^2 = 0.021$ (0.022)

Table 9. $X \sim (1 - U(0,1)^2)$, $\sigma^2 = 0.0167$

N=600	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	0.157	-0.217	0.020	-0.019	-0.022	0.040
Std. Dev	0.215	0.283	0.249	0.333	0.473	0.675

Estimated variance $\hat{\sigma}^2 = 0.019$ (0.019)

Table 10 $X \sim (1 - U(0,1)^2)$, $\sigma^2 = 0.0167$

N=1500	$\hat{\beta}_{Naive}$		$\hat{\beta}_{suff}$		$\hat{\beta}_{unkn}$	
	β_0	β_1	β_0	β_1	β_0	β_1
Bias	0.144	-0.215	0.007	-0.016	-0.026	0.030
Std. Dev	0.129	0.175	0.150	0.206	0.286	0.408

Estimated variance $\hat{\sigma}^2 = 0.017$ (0.014)

6 Conclusion

A simple estimation from estimating the weighted score function is proposed. The derivation of the estimator is simple in theory and calculation. This method, the estimation of weighted score function, can be easily modified to adjust for other parametric distribution of error, while the sufficient estimator needs sufficient statistics of X available. We believe that the proposed method is more applicable than the sufficient estimator with minor lost in efficiency

The natural extension of weighted score in section 5 show that the estimation without knowledge of σ^2 is possible. The simulation result also support this conjecture. However, a rigorous proof is not established now, and we should pursue it in the future.

7 Appendix

Proof of Theorem 1: From the condition above Theorem 1. We knows that $\frac{1}{n}\sum_{i=1}^{n} e^{sX} \rightarrow M(S)$, and hence it is also true that there exists functions $Q(s)$ such that $\frac{1}{n}\sum_{i=1}^{n} \frac{e^{sX}}{1+e^{-\beta_0-\beta_1 X}} \rightarrow Q(s)$. From the condition, it is also true that

$$\frac{1}{n}\sum_{i=1}^{n} Xe^{sX} \rightarrow M^{'}(s), \quad \frac{1}{n}\sum_{i=1}^{n} \frac{Xe^{sX}}{1+e^{-\beta_0-\beta_1 X}} \rightarrow Q^{'}(s)$$

Let

$$\psi(s) = Q^{'}(-k) + \frac{Q(-k)}{M(s-k) - Q(s-k)}[Q^{'}(s-k) - M^{'}(s-k)]$$

Apparently β_1 is a root of $\psi(s) = 0$. Now, we should prove that there is only one root for $\psi(s) = 0$. Let $H(s) = M(s) - Q(s)$ then $\psi(s) = Q^{'}(-k) - Q(-k)\frac{d\log H(b)}{db}\mid_{b=s-k}$. To show that $\psi(s)$ is a monotone function, It is enough to show that $\frac{d\log H(b)}{db}$ is a monotone function. Consider the derivative of $\frac{d\log H(b)}{db}$, $\frac{d^2\log H(b)}{db^2} = \frac{H^{''}(b)H(b)-H^{'}(b)^2}{H(b)^2}$. The numerator equals

$$\lim_{n\rightarrow\infty}[\frac{1}{n}\sum_{i=1}^{n} \frac{X^2e^{bX}}{1+e^{\beta_0+\beta_1 X}} \frac{1}{n}\sum_{i=1}^{n} \frac{e^{bX}}{1+e^{\beta_0+\beta_1 X}} - (\frac{1}{n}\frac{Xe^{bX}}{1+e^{\beta_0+\beta_1 X}})^2]$$

The term in the bracket is always positive. Hence $\frac{d^2\log H(b)}{db^2}$ is positive which implies that $\frac{d\log H(b)}{db}$ is a monotone function. Hence we conclude that β_1 is the only zero-crossing of the monotone function $\psi(s)$. A straightforward computation show that $\psi_n(s) \rightarrow \psi(s)$ almost surely, hence the theorem follows.

References

[1] Carroll, R. J., Rupper, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, London.

[2] Stefanski, L. A., and Carroll, R. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716

[3] Stefanski, L. A., and Carroll, R. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics* **13**, 1335-1351

[4] Nakamura, T. (1990). Corrected Scores Function for Errors-in-Variables Models: Methodology and Application to Generalized Linear Models. *Biometrika*, 77, 127-137.