第45卷第6期 中国科学技术大学学报 Vol. 45, No. 6
2015年6月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Jun. 2015

Article ID: 0253-2778(2015)06-0449-11

Sequential shrinkage estimation in generalized linear models with measurement errors

LU Haibo[1,2], WANG Zhanfeng[1], WU Yaohua[1]

(1. *Dept. of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei* 230026, *China*; 2. *School of Mathematical Sciences, Xinjiang Normal University, Urumqi* 830054, *China*)

Abstract: A sequential shrinkage estimation method was developed to determine a minimum sample size under which both of the variable selection and the parameter estimation with a pre-specified accuracy were achieved for the generalized linear model with measurement errors. Asymptotic properties of the proposed sequential estimation method, such as the coverage probability of the sequential confidence set and the efficiency of the minimum sample size, were studied. Simulation studies were conducted and the results show that the proposed method can save a large number of samples compared to the traditional sequential sampling method. Finally a diabetes data set was used as an example.

Key words: generalized linear model; sequential sampling; adaptive shrinkage estimate; stopping rule; confidence set

CLC number: O212.1 **Document code**: A doi: 10.3969/j.issn.0253-2778.2015.06.004

2010 Mathematics Subject Classification: Primary 62F12; Secondary 62E20

Citation: Lu Haibo, Wang Zhanfeng, Wu Yaohua. Sequential shrinkage estimation in generalized linear models with measurement errors[J]. Journal of University of Science and Technology of China, 2015, 45(6): 449-459, 496.

广义线性模型中带有测量误差的序贯压缩估计

鲁海波[1,2],王占锋[1],吴耀华[1]

(1.中国科学技术大学管理学院统计与金融系,安徽合肥 230026;2.新疆师范大学数学科学学院,新疆乌鲁木齐 830054)

摘要:对带测量误差的广义线性模型提出一种序贯压缩估计方法来确定最小样本量,使得在此最小样本量下所提方法可以选择有效变量,同时还可以获得给定精度下的回归参数估计.也研究了所提方法的渐进性质,包括序贯置信域的覆盖概率、最小样本量的效率等.模拟研究表明基于序贯压缩估计的抽样方法比传统的序贯抽样方法能够节省大量的样本.最后,用所提方法来分析一个糖尿病数据集.

关键词:广义线性模型;序贯抽样;自适应压缩估计;停时;置信域

Received: 2014-12-08; **Revised**: 2015-04-13

Foundation item: Supported by NNSF of China (11231010, 11471302), the Fundamental Research Funds for the Central Universities (WK2040000010).

Biography: LU Haibo, male, born in 1980, master. Research field: sequential analysis and survival analysis. E-mail: lhb1980@mail.ustc.edu.cn

Corresponding author: WANG Zhanfeng, PhD/ associate Prof. E-mail: zfw@ustc.edu.cn

0 Introduction

Generalized linear model (GLM) is widely applied to analyzing the relationship between the target variable (response) and the explanatory variables in many fields of modern science such as econometrics, medical science, sociology and so on. Especially when the discrete responses, for example, attribute data and enumeration data, are analyzed, GLM brings a number of powerful capabilities. For studies on GLM please refer to Refs. [1-4] and references therein. Suppose pairs of response and covariates, $(y_i, x_i)\in R\times R^p$, $i=1, 2, \cdots$, are sampled from GLM with a link function μ as follows,

$$\left.\begin{aligned} E[y_i \mid x_i] &= \mu(x_i^{\mathrm{T}}\beta_0), \\ \mathrm{Var}[y_i \mid x_i] &= v(x_i^{\mathrm{T}}\beta_0) \end{aligned}\right\} \tag{1}$$

where β_0 is a p-dimensional vector of the unknown regression coefficients, $\mu(\cdot)$ and $v(\cdot)$ are two known functions, and the first derivative of $\mu(t)$ for all t is positive. In practical applications, among the p explanatory variables only a few ones denoted by effective variables in Ref. [5] have contributions to the response. That is, only p_0 ($p_0<p$) components of β_0 are non-zero.

Under a given sample size (not a random variable), many methods can be used to select the effective variables and obtain estimates of the parameters, for instance, LASSO[6], LARS[7], and SCAD[8]. However, the number of samples needed to identify the effective variables and simultaneously estimate their coefficients under a pre-specified estimation accuracy is an important issue, especially when the cost of the sampling is considered in Biology and Epidemiology.

It is well-known that the sequential method can save samples by some "early stopping" sampling. For the linear regression model, Wang et al.[5] proposed a sequential shrinkage estimate method to identify the effective variables and attain the parameter estimates of a presupposed accuracy with the minimum number of samples. For GLM, a sequential method has been proposed in Ref. [9] without distinguishing the effective variables, and Lu et al.[10] developed a sequential approach to determining a minimum sample size under which the effective variables and their estimates are obtained with a pre-set precision. There is an assumption in Refs. [5, 10] that the explanatory variables can be observed without measurement errors. However, in practical applications the true values of the designs x_i's may not be measured frequently due to the inadequate accuracy of the measuring tools or some operating mechanism. Instead, the variable z_i consisting of x_i and the measurement error is observed. Chang[11] showed that in a linear regression model, the measurement errors in the covariates generally do not affect the consistency of least squares estimator under the assumption that the measurement errors have mean 0 conditional on the regressors. But he did not consider the issue of sample size. For GLM with measurement errors presented in explanatory variables, in this paper we propose a sequential procedure to construct a fixed size confidence set of the effective variables based on an adaptive shrinkage estimate (ASE). According to this procedure, the effective variables can be efficiently identified with the minimum sample size and simultaneously their corresponding regression coefficients are estimated with the required precision. Moveover, we show that the proposed sequential approach is asymptotically consistent and efficient in the sense of Ref. [12].

In the rest of this paper, an adaptive shrinkage estimate as well as its asymptotic properties will be established in Section 1 with measurement errors appearing in the covariates. In Section 2, the sequential sampling method based on the ASE is proposed with the stopping rule and the fixed size confidence set constructed. Simulation study is presented in Section 3 to illustrate the performance of the proposed sequential sampling method based on the ASE compared to the traditional method based on the maximum quasi-likelihood estimate (MQLE). A diabetes data set is used as a real example in the last section and details of the proofs of the main theorems are given in Appendix.

1 ASE with errors in covariates

Let $\{(y_i, x_i), i=1,2,\cdots,n\}$ be i. i. d. samples generated from model (1). Define $\hat{\delta}_n$ as the MQLE of β_0 in model (1), that is, $\hat{\delta}_n$ satisfies the score equation

$$\ln^*(\hat{\delta}_n) \equiv \sum_{i=1}^{n} \dot{\mu}(x_i^{\mathrm{T}}\hat{\delta}_n) w(x_i^{\mathrm{T}}\hat{\delta}_n)[y_i - \mu(x_i^{\mathrm{T}}\hat{\delta}_n)]x_i = 0 \quad (2)$$

where $\mu(t)$ denotes the link function of model (1) and $w(t)=v^{-1}(t)$. For more information about GLM and MQLE, please refer to Refs. [3,13]. However, the true covariates x_i's in Eq. (2) can not be observed due to contamination by measurement errors. Instead, $z_i = x_i + \xi_i$, is measured, where ξ_i is measurement error term. Based on z_i, we can obtain the quasi-likelihood equation with measurement errors in the covariates

$$\ln(\beta) \equiv \sum_{i=1}^{n} \dot{\mu}(z_i^{\mathrm{T}}\beta) w(z_i^{\mathrm{T}}\beta)[y_i - \mu(z_i^{\mathrm{T}}\beta)]z_i = 0 \quad (3)$$

Let $\tilde{\beta}_n$ be a solution to the estimating function (3), that is, $\ln(\tilde{\beta}_n)=0$. We call $\tilde{\beta}_n$ an MQLE of β_0 with measurement errors. Suppose ξ_i is independent of x_i and z_i, $i=1,\cdots,n$.

Set $\varepsilon_i = E[y_i] - \mu(x_i^{\mathrm{T}}\beta)$, $i=1,\cdots,n$, to be independent error terms. In this paper, we need the following assumptions:

(Ⅰ) $\sup_i \| z_i \| < \infty$, and $E[|\varepsilon_i|^p] \leqslant \infty$ for $p>2$.

(Ⅱ) $\lambda_{\min}(\sum_{i=1}^{n} x_i^{\mathrm{T}} x_i) \to \infty$ a. s. and

$$\lim_{n\to\infty}[\lambda_{\min}(\sum_{i=1}^{n} x_i^{\mathrm{T}} x_i)/\ln(\lambda_{\max}(\sum_{i=1}^{n} x_i^{\mathrm{T}} x_i))] = \infty$$

a. s..

(Ⅲ) $\lim_{n\to\infty}(\sum_{i=1}^{n} z_i[\dot{\mu}(z_i^{\mathrm{T}}\beta_n)]^2 w(z_i^{\mathrm{T}}\beta_n) z_i^{\mathrm{T}})/n = \Sigma$, where Σ is a positive matrix.

(Ⅳ) There exists a sequence of constants $\{a_n\}$ with $1 \leqslant a_n \to \infty$ such that $\sum_{n=1}^{\infty} E(\| \xi_n \|^2 / a_n)^{\nu_n} < \infty$ for some constants $0 \leqslant \nu_n \leqslant 1/2$.

The conditions (Ⅰ), (Ⅱ) and (Ⅳ) are the same as those in Ref. [11], and under these conditions, Theorem 1 in Ref. [11] implies that for some $\eta>0$

$$n^{\frac{1}{2}-\eta}(\tilde{\beta}_n - \beta_0) = o(1), \text{ a. s. as } n \to \infty \quad (4)$$

Assume $\kappa \equiv \kappa(n)$ is a non-random function of n such that for some $0<\delta<\frac{1}{2}$ and $\gamma>0$,

$$n^{\frac{1}{2}}\kappa \to 0 \text{ and } n^{\frac{1}{2}+\gamma\delta}\kappa \to \infty, \text{ as } n \to \infty \quad (5)$$

For instance, we can take $\kappa = n^{-\theta}$ with $\theta \in (\frac{1}{2}, \frac{1}{2}+\gamma\delta)$.

Define $\kappa_{nj} = \kappa\, |\tilde{\delta}_{nj}|^{-\gamma}$, where $\tilde{\beta}_{nj}$ is the jth component of $\tilde{\beta}_n$. With presuming $\infty \times 0 = 0$, we have from(4) and (5),

$$n^{1/2}\kappa_{nj} \to 0 \times I(\beta_{0j} \neq 0) + \infty \times I(\beta_{0j} = 0) \quad \text{a. s. as } n \to \infty \quad (6)$$

where $I(\cdot)$ denotes the indicator function. That is, for each j, the indicator $I_{nj}(\epsilon) = I(\sqrt{n}\kappa_{nj} < \epsilon)$ can be used to identify whether the jth component of β_0 is significantly apart from zero by at least a pre-specified positive constant ϵ. It also implies that $\hat{p}_0 \equiv \hat{p}_0(n) \equiv \sum_{i=1}^{p} I_{nj}(\epsilon)$ can be used as an estimator of p_0. It can be proved from Lemma 1 in Appendix that the index $I_{nj}(\epsilon)$ almost surely converges to $I(\beta_{0j} \neq 0)$, which easily implies that $\hat{p}_0$ almost surely converges to p_0 and $E[\hat{p}_0]$ converges to p_0, as n tends to ∞.

Similar to Ref. [5], define $\hat{\beta}_n = I_n(\epsilon)\tilde{\beta}_n$ as an ASE of β_0, where

$$I_n(\epsilon) = \mathrm{diag}\{I_{n1}(\epsilon), \cdots, I_{np}(\epsilon)\} \quad (7)$$

is a $p \times p$ diagonal matrix. The strong consistency and asymptotic distribution of $\hat{\beta}_n$ are presented in the following theorem:

Theorem 1 Assume that $(y_i, x_i), i=1,\cdots,n$, are pairs of responses and covariates generated from model (1) and x_i satisfies $z_i = x_i + \xi_i$. Under (Ⅰ), (Ⅱ) and (Ⅳ), for any small $\epsilon \geqslant 0$, as $n \to \infty$ we have with probability one,

(ⅰ) $\hat{\beta}_n \to I_0\beta_0 = \beta_0$, with convergence rate;

(ⅱ) $\| \hat{\beta}_n - \beta_0 \| = O([\ln n/n]^{1/2})$;

(ⅲ) $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{L} N(0, I_0\Sigma^{-1}I_0)$ as $n\to\infty$;

where $\|\cdot\|$ denotes Euclidean norm and $I_0=\mathrm{diag}(I(\beta_{01}\neq 0),\cdots,I(\beta_{0p}\neq 0))$ is a $p\times p$ diagonal matrix and Σ is a positive matrix defined in (Ⅲ).

The proof of Theorem 1 is given in Appendix.

2 Sequential sampling strategy with errors

Sequential sampling methods[14-15], are usually applied to determine sample size when samples are observed sequentially or no proper fixed sample size procedure can be used any more. In this situation, it is reasonable to consider how many samples can be saved using sequential procedures. Thus, we focus on the efficiency of the sequential method based on the ASE instead of the MQLE. Similar to Ref. [5], the sequential fixed size confidence set estimation and the expected sample size are employed to illustrate the performance of our method.

In order to construct the confidence set for β_0 we need to study the asymptotic properties of the ASE, i. e. $\hat{\beta}_n$, under random sample size. The property of the uniform continuity in probability[16-17] is proved to be a sufficient condition such that the randomly stopped sequence has the same asymptotic distribution as that of the fixed sample size estimate. Lemma 2 of Appendix shows that the sequence $\sqrt{n}(\hat{\beta}_n-\beta_0), n=1,2,\cdots$, has the uniform continuity in probability property, which indicates the following theorem holds.

Theorem 2 Suppose that the conditions (Ⅰ)～(Ⅳ) are satisfied, and let $N(t)$ be a positive integer-valued random variable for which $N(t)/t$ converges to 1 in probability as $t\to\infty$. Then

$$\sqrt{N(t)}(\hat{\beta}_{N(t)}-\beta_0) \xrightarrow{L} N(0, I_0\Sigma^{-1}I_0),\ \text{as } t\to\infty \tag{8}$$

We employ Theorem 2 to construct a sequential sampling procedure to determine the sample size under which the effective variables are identified and simultaneously their coefficients are estimated with a pre-specified accuracy. Let $Y_n=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ and $Z_n=(z_1,z_2,\cdots,z_n)$ be $n\times 1$ and $p\times n$ matrices of the observations, respectively. And Z_n satisfies $Z_n=X_n+\xi_n$ in which X_n is the true design and ξ_n is the measurement error. Following Ref. [5], conditional on the samples given up to the current stage, there exists an orthonormal matrix O_n satisfying $O_n^{\mathrm{T}}O_n=I_p$, and $(\hat{\beta}_{n1}^{\mathrm{T}},\hat{\beta}_{n2}^{\mathrm{T}})^{\mathrm{T}}=O_n\hat{\beta}_n$, where $(\hat{\beta}_{n1}^{\mathrm{T}},\hat{\beta}_{n2}^{\mathrm{T}})^{\mathrm{T}}$ is the rearranged order of $\hat{\beta}_n$ components such that all of the indicators I_{nj}'s relevant to $\hat{\beta}_{n1}$ are 1 and those relevant to $\hat{\beta}_{n2}$ are 0.

Denote

$$\sum_{i=1}^{n} z_i[\dot{\mu}(z_i^{\mathrm{T}}\hat{\beta}_n)]^2 w(z_i^{\mathrm{T}}\hat{\beta}_n)z_i^{\mathrm{T}} = (W^{1/2}Z_n)(W^{1/2}Z_n)^{\mathrm{T}},$$

where

$$W=\mathrm{diag}\{[\dot{\mu}(z_i^{\mathrm{T}}\hat{\beta}_n)]^2 w(z_i^{\mathrm{T}}\hat{\beta}_n),\ i=1,2,\cdots,n\}.$$

Partition the matrix $(O_nW^{1/2}Z_n)(O_nW^{1/2}Z_n)^{\mathrm{T}}$ according to the first $\hat{p}_0$ non-zero components of $O_n\hat{\beta}_n$ such that

$$(O_nW^{1/2}Z_n)(O_nW^{1/2}Z_n)^{\mathrm{T}} = \begin{pmatrix} \Sigma_{11}(n)_{\hat{p}_0\times\hat{p}_0} & \Sigma_{12}(n)_{\hat{p}_0\times(p_0-\hat{p}_0)} \\ \Sigma_{21}(n)_{(p-\hat{p}_0)\times\hat{p}_0} & \Sigma_{22}(n)_{(p-\hat{p}_0)\times(p-\hat{p}_0)} \end{pmatrix}.$$

Then by simple matrix computation, we have

$$O_nI_n(\epsilon)((W^{1/2}Z_n)(W^{1/2}Z_n)^{\mathrm{T}})^{-1}I_n(\epsilon)O_n^{\mathrm{T}} = O_nI_n(\epsilon)O_n^{\mathrm{T}}((O_nW^{1/2}Z_n)(O_nW^{1/2}Z_n)^{\mathrm{T}})^{-1}O_nI_n(\epsilon)O_n^{\mathrm{T}} = \begin{pmatrix} \tilde{\Sigma}_{11}^{-1}(n) & 0 \\ 0 & 0 \end{pmatrix} \tag{9}$$

where

$$\tilde{\Sigma}_{11}^{-1}(n) = \Sigma_{11}^{-1}(n)+\Sigma_{11}^{-1}(n)\Sigma_{12}(n)\Sigma_{22.1}^{-1}(n)\Sigma_{21}(n)\Sigma_{11}^{-1}(n),$$

$$\Sigma_{22.1}^{-1}(n)=\Sigma_{22}(n)-\Sigma_{21}(n)\Sigma_{11}^{-1}(n)\Sigma_{12}(n).$$

Let M^- denote a general inverse matrix M. It follows that

$$(U-\hat{\beta}_n)^{\mathrm{T}}(I_n(\epsilon)((W^{1/2}Z_n)(W^{1/2}X_n)^{\mathrm{T}})^{-1}I_n(\epsilon))^-\cdot (U-\hat{\beta}_n) = (O_nU-O_n\hat{\beta}_n)^{\mathrm{T}}[O_nI_n(\epsilon)O_n^{\mathrm{T}}\cdot$$

$$((O_nW^{1/2}Z_n)(O_nW^{1/2}Z_n)^{\mathrm{T}})^{-1}O_nI_n(\epsilon)O_n^{\mathrm{T}}]^{-}\cdot (O_nU-O_n\hat{\beta}_n)= (U_{n1}-\hat{\beta}_{n1})^{\mathrm{T}}\tilde{\Sigma}_{11}(n)(U_{n1}-\hat{\beta}_{n1}) \quad (10)$$

where $U=(u_1,u_2,\cdots,u_p)^{\mathrm{T}}\in R^p$ and U_{n1} is a sub-vector of U corresponding to $\hat{\beta}_{n1}$. Theorem 1 implies that as $N\to\infty$,

$$N(\hat{\beta}_N-\beta_0)^{\mathrm{T}}(I_N(\epsilon)(W^{1/2}Z_N(W^{1/2}Z_N)^{\mathrm{T}})^{-1}I_N(\epsilon))^{-}\cdot (\hat{\beta}_N-\beta_0)= N(\hat{\beta}_{N1}-\beta_{01})^{\mathrm{T}}\tilde{\Sigma}_{11}(N)(\hat{\beta}_{N1}-\beta_{01})\xrightarrow{L}\chi^2(p_0) \quad (11)$$

However, the true p_0 is unknown and has to be estimated based on the observations. Suppose C_k denote the first k observations $\{(y_i,z_i):i=1,\cdots,k\}$. We can use $\hat{p}_0(k)=\sum_{j=1}^{p}I_{kj}(\epsilon)$ to estimate p_0 with a known positive ϵ based on C_k. Let $\{a_k\}^2\in R$ be a constant satisfying the conditional probability $P(\chi^2_{\hat{p}_0(k)}\leqslant a_k{}^2\mid C_k)=1-\alpha$ for a given α. Then a stopping rule N_d is defined as

$$N=N_d\equiv\inf\{k:\ k\geqslant n_0 \text{ and } \nu_k\leqslant\frac{d^2}{a_k^2}\} \quad (12)$$

where ν_k is the maximum eigenvalue of $kI_k(\epsilon)(\sum_{i=1}^{k}[\dot{\mu}(z_i^{\mathrm{T}}\hat{\beta}_k)]^2w(z_i^{\mathrm{T}}\hat{\beta}_k)z_iz_i^{\mathrm{T}})^{-1}I_k(\epsilon)$, and d is a pre-specified precision of the confidence set. A new observation is collected at a time until the stopping criterion defined in Eq. (12) is satisfied. Then when the stopping rule holds, using N samples, a confidence set of β_0 is constructed as follows,

$$R_N=\{U\in R^p:\ \frac{S_N}{N}\leqslant\frac{d^2}{\nu_N} \text{ and } u_j=0 \text{ for } I_{Nj}(\epsilon)=0, 1\leqslant j\leqslant p\} \quad (13)$$

where $S_N=(U_{N1}-\hat{\beta}_{N1})^{\mathrm{T}}\tilde{\Sigma}_{11}(N)(U_{N1}-\hat{\beta}_{N1})$. It is easy to show that the maximum axis of the ellipsoid defined by R_N, that is the equation

$$(N^{-1}\nu_N/d^2)(U_{N1}-\hat{\beta}_{N1})^{\mathrm{T}}\tilde{\Sigma}_{11}(N)\{(U_{N1}-\hat{\beta}_{N1})\}=1,$$

is $2d$, which is the pre-specified accuracy of estimation of β_0.

Note that there are $p-\hat{p}_0$ axes relevant to zero components of $\hat{\beta}_N$ in R_N, and other $\hat{p}_0$ components make up a degenerate ellipsoid confidence set, called $R_N^{\hat{p}_0}$, which is the projection of R_N into the $\hat{p}_0$-dimensional space spanned by the axes with the non-zero components of $\hat{\beta}_N$. In other words, the proposed sequential estimation procedure focuses on the effective variables while ignoring the non-effective ones. The crucial difference between the proposed method and others is its ability save a large number of samples. Properties of the sequential procedure and its confidence set of β_0 are summarized below, whose proof can be seen in Appendix.

Theorem 3 Assume that the conditions (Ⅰ)～(Ⅳ) are satisfied, and let N be the stopping time as defined in Eq. (12). Then (ⅰ) $\lim_{d\to0}\frac{d^2N}{a^2\nu}=1$ almost surely; (ⅱ) $\lim_{d\to0}P(\beta_0\in R_N)=1-\alpha$; (ⅲ) $\lim_{d\to0}\frac{d^2E(N)}{a^2\nu}=1$; (ⅳ) $\lim_{d\to0}\hat{p}_0(N)=p_0$, almost surely, and $\lim_{d\to0}E(\hat{p}_0(N))=p_0$; where ν is the maximum eigenvalue of matrix $I_0\Sigma^{-1}I_0$.

From Theorem 3, result (ⅱ) shows that the sequential confidence set has the coverage probability asymptotically equal to the nominated value $1-\alpha$, and (ⅲ) illustrates that the ratio of the expected sample size of the sequential procedure to the best (unknown) sample size goes to 1.

In applications, ϵ is an unknown tuning parameter, and needs to be determined by using some model selection criteria such as the Akaike information criterion (AIC), Bayesian information criterion (BIC) and a general cross validation (GCV) method. In this paper, BIC is employed in our numerical study for illustration purposes. The precision index d is chosen by users based on the practical need.

3 Simulation

Performance of the sequential adaptive estimation with measurement errors is evaluated by numerical simulation studies in this section. We use the stopping time (final sample size) and the coverage probability of final fixed size confidence set to compare the proposed method based on and denoted by ASE with the classical sequential approach based on and denoted by MQLE. To illustrate the power of the proposed method, we use just the p_0 effective variables to build a GLM model while assuming that the p_0 effective variables are known in advance, denoted by MQLE_{p_0}. Thus, MQLE_{p_0} is apparently the most efficient method.

Under the fixed design, the synthesized data sets for the model (1) are generated as follows: except the intercept term the other designs with errors ξ_n's are generated from $z_n = x_n + \xi_n$, where the true designs x_n's follow a standard multivariate normal distribution with mean 0 and identity covariance matrix and the measurement errors ξ_n's follow a multivariate normal distribution with mean 0 and covariance matrix as $2/[\sqrt{n}(\ln(n)^{1.8}]I_p$. The response y_i is independently drawn from the poisson regression model for each $i \geqslant 1$. The true values of the parameters are chosen as $\beta_0 = (-1.5, 2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$ with 8 non-effective variables or $(-1.2, 1.2, 1.5, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$ with 7 non-effective variables.

We choose $\gamma = 1$, $\delta = 0.45$ and $\eta = 0.75$ in analyzing simulated data. Different precisions of confidence ellipsoid d are taken as $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ When applying the ASE method, the regularization parameter ϵ is determined by BIC criterion,

$$\text{BIC} = -2l + \ln(n) \cdot df,$$

where l is the log-likelihood function of the samples $\{y_i, i = 1, 2 \cdots, n\}$ and df is the number of the non-zero components in β.

Tab. 1 states the results of the sequential sampling estimation for Poisson regression, where values of the final sample size N (stopping time), $\kappa = d^2 N/(a^2\nu)$ and empirical coverage probability CP of the 95% confidence set R_N are presented. The κ is very close to 1 for all three cases: MQLE_{p_0}, ASE and MQLE, and the empirical coverage probability CP approaches the nominal 95% as d decreases, as stated in Theorem 2. However, the sample sizes N of MQLE are much larger than those of the other two cases, and the sample size of ASE is very close to those of MQLE_{p_0}. Especially for $d = 0.1$, the ratios of the sample size N of MQLE and ASE are 1961.188/783.538 ≈ 2.503 and 3211.775/1401.204 ≈ 2.292 when the non-zero components of β_0 are $(-1.5, 2.0)$ and $(-1.2, 1.2, 1.5)$, respectively. In conclusion, when measurement errors occurs in the covariates, the proposed ASE is more efficient

Tab. 1 Results of sequential sampling method based on ASE, MQLE with all variables and MQLE with only p_0 non-zero variables

	d	MQLE_{p_0} N	κ	CP	ASE N	κ	CP	MQLE N	κ	CP
$\beta = (-1.5, 2.0, 0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$	0.6	19.554(2.689)	1.044	0.952	42.422(10.994)	1.116	0.980	104.190(7.036)	1.015	0.950
	0.5	28.934(3.697)	1.027	0.952	57.514(12.232)	1.075	0.966	125.350(2.447)	1.080	0.966
	0.4	46.808(4.106)	1.018	0.946	77.104(15.264)	1.084	0.974	201.122(8.838)	1.006	0.950
	0.3	90.268(6.110)	1.009	0.958	102.180(14.704)	1.019	0.950	280.331(9.351)	1.110	0.972
	0.2	182.256(8.632)	1.004	0.944	266.038(31.932)	1.012	0.946	546.462(11.591)	1.003	0.944
	0.1	824.934(0.789)	1.246	0.968	783.538(18.241)	1.000	0.938	1 961.188(25.382)	1.000	0.952
$\beta = (-1.2, 1.2, 1.5, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$	0.6	39.960(2.020)	1.060	0.961	70.166(9.155)	1.044	0.956	181.620(5.436)	1.050	0.960
	0.5	46.000(2.587)	1.056	0.980	92.112(11.135)	1.030	0.948	241.200(18.919)	1.013	0.940
	0.4	65.320(3.961)	1.016	0.945	129.494(11.306)	1.013	0.950	324.840(2.244)	1.075	0.953
	0.3	122.300(7.274)	1.006	0.960	207.564(8.004)	1.082	0.962	403.340(13.516)	1.005	0.920
	0.2	273.900(12.097)	1.004	0.947	386.386(15.283)	1.003	0.946	837.000(16.887)	1.021	0.960
	0.1	1 434.780(8.512)	1.000	0.971	1 401.204(10.676)	1.016	0.954	3 211.775(42.623)	1.000	0.951

【Note】 ① $\kappa_1 = d^2 N/(a^2\nu)$.

② CP is the empirical coverage probability of 95% confidence ellipsoid region R_N.

Tab. 2 Power of variable identification and estimation of nonzero components (−1.5, 2.0) under sequential sampling method based on ASE and MQLE

	ASE				MQLE			
d	Num_{ic}	Num_c	β_{01}	β_{02}	Num_{ic}	Num_c	β_{01}	β_{02}
0.6	0.132	8	−1.323(0.550)	1.969(0.131)	—	—	−1.498(0.127)	1.996(0.053)
0.5	0.084	8	−1.407(0.457)	1.987(0.110)	—	—	−1.506(0.109)	2.001(0.041)
0.4	0.058	8	−1.422(0.376)	1.985(0.087)	—	—	−1.499(0.088)	1.999(0.034)
0.3	0.022	8	−1.463(0.249)	1.992(0.060)	—	—	−1.508(0.062)	2.002(0.021)
0.2	0.016	8	−1.477(0.205)	1.996(0.046)	—	—	−1.499(0.043)	2.000(0.015)
0.1	0.00	8	−1.500(0.040)	2.000(0.014)	—	—	−1.500(0.023)	2.000(0.008)

【Note】 Num_c and Num_{ic} are the average number of zero components in β correctly identified and nonzero components incorrectly estimated as zero values, respectively.

Tab. 3 Power of variable identification and estimation of nonzero components (−1.2, 1.2, 1.5) under sequential sampling method based on ASE and MQLE

	ASE					MQLE				
d	Num_{ic}	Num_c	β_{01}	β_{02}	β_{03}	Num_{ic}	Num_c	β_{01}	β_{02}	β_{03}
0.6	0.014	6.956	−1.205(0.238)	1.192(0.110)	1.507(0.073)	—	—	−1.199(0.123)	1.195(0.057)	1.497(0.040)
0.5	0.004	6.984	−1.223(0.186)	1.204(0.067)	1.505(0.060)	—	—	−1.205(0.124)	1.205(0.045)	1.495(0.047)
0.4	0.002	6.996	−1.210(0.145)	1.201(0.049)	1.502(0.145)	—	—	−1.198(0.073)	1.195(0.032)	1.500(0.030)
0.3	0.002	7	−1.205(0.109)	1.201(0.038)	1.501(0.029)	—	—	−1.202(0.068)	1.200(0.027)	1.500(0.022)
0.2	0.0	7	−1.200(0.068)	1.200(0.024)	1.500(0.018)	—	—	−1.207(0.040)	1.201(0.018)	1.500(0.014)
0.1	0.0	7	−1.201(0.033)	1.200(0.011)	1.500(0.008)	—	—	−1.202(0.024)	1.202(0.008)	1.501(0.008)

【Note】 Num_c and Num_{ic} are the average number of zero components in β correctly identified and nonzero components incorrectly estimated as zero values, respectively.

than MQLE, while remaining efficient with MQLE_{p_0}.

Tabs. 2 and 3 report the power of the proposed method for identify the effective variables and their estimates of the regression coefficients for Poisson regression when the non-zero components of β_0 are (−1.5, 2.0) and (−1.2, 1.2, 1.5), respectively. We can see that number of incorrectly identified zero variables (Num_{ic}) using ASE is almost close to 0, and the number of correctly identified zero variables (Num_c) are all very close to the true number of effective variables (8 and 7). These results suggest that $\hat{p}_0(N)$ is a good estimator of p_0 under the sequential sampling method based on ASE. The MQLE procedure does not identify the effective variables, so Num_c and Num_{ic} are not available. In addition, the estimates of parameters of effective variables are all very close to the true values.

4 A real example

We apply the sequential ASE method to a diabetes data set provided in Ref. [18], and compare the performance of the proposed method with the traditional MQLE method. There are 381 subjects which have complete observations in total 403 samples of the data set, and 8 explanatory variables: total cholesterol (Chol), stabilized glucose(Stab. glu), high density lipoprotein(Hdl), cholestarol/HDL ratio(Ratio), age(Age), gender (Gender), body mass index(BMI), and waist/hip ratio (WHR). The response variable is glycosolated hemoglobin(Glyhb) which is usually selected as a primary measure of diabetes. Due to the precision requirement of measurement tools and skills of the researchers, variables such as Stab. glu and Hdl exhibit measurement errors except Age and Gender. It is well-known that diabetes (Glyhb) has a significant relationship with Stab. glu[5]. So we remove Stab. glu from the explanatory factors and focus on the relationship between Glyhb and other variables in this paper.

Tabs. 4 and 5 list the results of the sequential estimation for the diabetes data set. In Tab. 4, the sample sizes based on ASE and MQLE are presented. The coefficients estimates are given in Tab. 5. It is shown that the traditional sequential sampling based on MQLE can not be stopped even if the sample size reaches the number of total samples 381, that is, the final sample size N is larger than 381 no matter what the value d takes from 0.6 to 0.2. But the sample size for the sequential sampling method based on ASE ranges between 23 and 127 when the value of d varies from 0.6 to 0.2. From the lower panel, it follows that both of ASE and MQLE select Age as an effective variable. In conclusion, the sequential sampling based on the ASE method is much more efficient and can save a lot of samples compared to the traditional sequential sampling based on MQLE.

Tab. 4 Results of sequential estimation (sample size estimates) for two methods: ASE and MQLE based on the diabetes data

d	ASE		LSE	
	N	κ	N	κ
0.6	23	1.692	381	0.671
0.5	23	1.175	—	—
0.4	34	1.013	—	—
0.3	101	1.702	—	—
0.2	127	1.008	—	—

【Note】 ① $\kappa_1=d^2N/(a^2\nu)$.

② Stopping rule is not satisfied although sample size reaches total 381.

Tab. 5 Results of sequential estimation (coefficient estimates) for two methods: ASE and MQLE based on the diabetes data

method	d	Chol	Hdl	Ratio	Age	Gender	WHR	BMI
ASE	0.6	0	0	0	0.240	0	0	0
	0.5	0	0	0	0.240	0	0	0
	0.4	0	0	0	0.321	0	0	0
	0.3	0	0	0	0.309	0	0	0
	0.2	0	0	0	0.335	0	0	0
LSE	0.4	0.029	0.080	0.312	0.256	0.011	0.071	0.064

5 Conclusion

When measurement errors are presented in explanatory variables, we develop a sequential sampling procedure to construct the fixed size confidence set for the effective parameters based on ASE. According to this sequential procedure, the effective coefficients can be efficiently identified with the minimum sample size and their corresponding regression coefficients estimated simultaneously with the required precision. The consistency and asymptotic normality of ASE have been proved by using a last time method. Meanwhile, we prove that the proposed sequential procedure is asymptotically optimal in the sense of Ref. [12]. Results of simulation study implies that the proposed method can save a large number of samples compared to the traditional sequential sampling method. Application of the diabetes data set shows that the proposed method based on ASE is more efficient than the traditional method with MQLE. However, the design of covariates is fixed in this paper, hence we will investigate the sequential sampling method in the case of covariates with random design in our future work.

Appendix

First of all, we define a last time random variable, L_η, as follows,

$$L_\eta = \sup\{n\geqslant 1: -(\beta-\beta_0)^{\mathrm T} l_n(\beta) < 0,\ \exists \beta\in\partial\mathscr{B}\} = \sup\{n\geqslant 1: (\beta-\beta_0)^{\mathrm T} l_n(\beta) \geqslant 0,\ \exists \beta\in\partial\mathscr{B}\} \quad (\mathrm{A1})$$

where $\mathscr{B}=(\mathscr{B}_\eta)=\{\beta: \|\beta-\beta_0\|\leqslant\eta\}$ for some $\eta\geqslant 0$ and $\partial\mathscr{B}$ denotes the boundary of $\mathscr{B}$.

Lemma 1 Assume that $(y_i, z_i), i=1,\cdots,n$, are pairs of responses and true designs satisfying Eq. (1) with random error ε_i, which has mean 0 and variance $0<\sigma^2<\infty$. Then for any small $\epsilon>0$, we have $I_{nj}(\epsilon)\to I(\beta_{0j}\neq 0)$, a. s.. In addition, $\lim_{n\to\infty}\hat p_0=p_0$ a. s. and $\lim_{n\to\infty}E\hat p_0=p_0$ a. s..

Similar to Ref. [5, Theorem 1], Lemma 1 holds.

Proof of Theorem 1 By definition of $\hat\beta_n$ for any given $\zeta>0$ and $\epsilon>0$, we have

$$P(\sup_{k\geqslant n}\|\hat{\beta}_k-I_0\beta_0\|>\zeta)=$$
$$P\left(\sup_{k\geqslant n}\|\tilde{\beta}_k-\beta_0\|>\frac{\zeta}{2}\right)+$$
$$P(\sup_{k\geqslant n}\|\beta_0\|\cdot\|I_k(\epsilon)-I_0\|>\zeta)\quad(A2)$$

According to Lemma 1 and Ref. [11, Theorem 1], the consistency of $\tilde{\beta}_n$ can be easily obtained, thus similar to the proof of Ref. [5, Theorem 2], the consistency of $\hat{\beta}_n$ is proved. For simplisity we only consider the situation of the canonical link function, then following Eq. (3) and by the mean-value theorem, it can be shown that

$$0=\frac{1}{n}\sum_{i=1}^{n}z_i[y_i-\mu(x_i\beta_0)]-$$
$$\frac{1}{n}\sum_{i=1}^{n}[\mu(z_i^{\mathrm{T}}\beta_0)-\mu(x_i^{\mathrm{T}}\beta_0)]-$$
$$\frac{1}{n}\sum_{i=1}^{n}[\mu(z_i^{\mathrm{T}}\tilde{\beta}_n)-\mu(z_i^{\mathrm{T}}\beta_0)]=$$
$$\frac{1}{n}\sum_{i=1}^{n}z_i\varepsilon_i-$$
$$\frac{1}{n}\sum_{i=1}^{n}[\dot{\mu}(x_i^*\beta_0)z_i\beta_0^{\mathrm{T}}\xi_i+\dot{\mu}(z_i^{\mathrm{T}}\beta_n^*)z_iz_i^{\mathrm{T}}(\tilde{\beta}_n-\beta_0)]\quad(A3)$$

where x_i^* is in the line segment of x_i and z_i for each i, and β_n^* is in the line segment of $\tilde{\beta}_n$ and β_0. Eq. (A3) implies that

$$\tilde{\beta}_n-\beta_0=(\sum_{i=1}^{n}\dot{\mu}(z_i^{\mathrm{T}}\beta_n^*)z_iz_i^{\mathrm{T}})^{-1}\cdot$$
$$[\sum_{i=1}^{n}z_i\varepsilon_i-\sum_{i=1}^{n}\dot{\mu}(x_i^*\beta_0)z_i\beta_0^{\mathrm{T}}\xi_i]=$$
$$(\sum_{i=1}^{n}\dot{\mu}(z_i^{\mathrm{T}}\beta_n^*)z_iz_i^{\mathrm{T}})^{-1}\cdot$$
$$[\sum_{i=1}^{n}z_i\varepsilon_i-\sum_{i=1}^{n}\dot{\mu}(x_i^*\beta_0)(x_i+\xi_i)\beta_0^{\mathrm{T}}\xi_i]=$$
$$(\sum_{i=1}^{n}\dot{\mu}(z_i^{\mathrm{T}}\beta_n^*)z_iz_i^{\mathrm{T}})^{-1}\cdot$$
$$[\sum_{i=1}^{n}z_i\varepsilon_i-\sum_{i=1}^{n}\dot{\mu}(x_i^*\beta_0)x_i\beta_0^{\mathrm{T}}\xi_i^{\mathrm{T}}-$$
$$\sum_{i=1}^{n}\dot{\mu}(x_i^*\beta_0)\xi_i\beta_0^{\mathrm{T}}\xi_i^{\mathrm{T}}]\quad(A4)$$

provided that the inverse of $\sum_{i=1}^{n}\dot{\mu}(z_i^{\mathrm{T}}\beta_n^*)z_iz_i^{\mathrm{T}}$ exists when n is large. If $n>L_\rho$, then $\tilde{\beta}_n$ exists and is in $\mathscr{B}_\rho$. It is proved in Ref. [11] that under (Ⅳ) with $a_n=O(\sqrt{n})$, with probability one

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i\to 0,\ \frac{1}{n}\sum_{i=1}^{n}\|\xi_i\|^2\to 0\quad(A5)$$

Therefore, it follows from the assumption that $\sup_i\|z_i\|\leqslant\infty$ almost surely, and by continuity properties of $\dot{\mu}$, we have on the event $\{n>L_\rho\}$

$$\tilde{\beta}_n-\beta_0=O(1)(\sum_{i=1}^{n}z_iz_i^{\mathrm{T}})^{-1}\sum_{i=1}^{n}z_i\varepsilon_i\quad(A6)$$

By integrability of L_ρ, we have $\lim_{n\to\infty}P(n<L_\rho)=0$. Note that the right-hand side of Eq. (A6) is similar to that in Ref. [19]. Then, by applying the results of Ref. [19] to (A6), we obtain the convergence rate of $\tilde{\beta}_n$; that is, with probability one

$$\|\tilde{\beta}_n-\beta_0\|=O\left(\left[\frac{\ln\{\lambda_{\max}(\sum_{i=1}^{n}z_iz_i^{\mathrm{T}})\}}{\lambda_{\min}(\sum_{i=1}^{n}z_iz_i^{\mathrm{T}})}\right]^{\frac{1}{2}}\right)=$$
$$O([\ln n/n]^{1/2})\quad(A7)$$

Just like Ref. [5, Theorem 2], the proofs of Theorem 1(ii) and (iii) are completed. □

Lemma 2 Under the conditions of Theorem 1, the sequence of random variables, $\{\sqrt{n}(\hat{\beta}_n-\beta_0),\ n=1,2,\cdots\}$ is uniformly continuous in probability.

Proof With probability one, we have that

$$\sqrt{n}(\hat{\beta}_n-\beta_0)=$$
$$\sqrt{n}I_n(\epsilon)(\tilde{\beta}_n-\beta_0)+\sqrt{n}(I_n(\epsilon)-I_0)\beta_0\approx$$
$$\Delta_1(n)+\Delta_2(n)\quad(A8)$$

Following Ref. [17, Example 1.8], we can show that the sequence $\{\sqrt{n}(\tilde{\beta}_n-\beta_0),\ n=1,2,\cdots\}$ is uniformly continuious in probability and that for $\eta>0$,

$$P(\sup_{0\leqslant k\leqslant n\tau}|\sqrt{n+k}(\tilde{\beta}_{n+k}-\beta_0)-\sqrt{n}(\tilde{\beta}_n-\beta_0)|>\eta/2)<\eta/2\quad(A9)$$

This implies that

$$P(\sup_{0\leqslant k\leqslant n\tau}|\Delta_1(n+k)-\Delta_1(n)|>\eta)=$$
$$P(\sup_{0\leqslant k\leqslant n\tau}|\sqrt{n+k}I_{n+k}(\epsilon)(\tilde{\beta}_{n+k}-\beta_0)-$$
$$\sqrt{n}I_n(\epsilon)(\tilde{\beta}_n-\beta_0)|>\eta)\leqslant$$
$$P(\sup_{0\leqslant k\leqslant n\tau}|(I_{n+k}(\epsilon)-I_n(\epsilon))\sqrt{n}(\tilde{\beta}_n-\beta_0)|>\eta/2)+$$
$$P(\sup_{0\leqslant k\leqslant n\tau}|I_{n+k}(\epsilon)(\sqrt{n+k}(\tilde{\beta}_{n+k}-\beta_0)-$$

$$\sqrt{n}(\tilde{\beta}_n-\beta_0))|>\eta/2) \tag{A10}$$

From Lemma 1 and (A9), we have

$$P(\sup_{0\leqslant k\leqslant n\tau}|\Delta_1(n+k)-\Delta_1(n)|>\eta)< \eta/2+\eta/2=\eta \tag{A11}$$

Thus if we prove that $\Delta_2(n)$ converges to 0 almost surely as n tends to infinity, then the proof is completed. Note that $\Delta_{2j}(n)=0$ as $\beta_{0j}=0$ for all j. Therefore, we need to prove that for $\beta_{0j}\neq 0$,

$$P(\sup_{n\leqslant k}|\Delta_{2j}(k)|>\eta)<\eta \tag{A12}$$

By the definition of $\Delta_2(n)$, we have

$$\begin{aligned}
&P(\sup_{n\leqslant k}|\Delta_{2j}(k)|>\eta)=\\
&P(\sup_{n\leqslant k}\sqrt{k}\,|\beta_{0j}(I(\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}<\epsilon)-1)|>\eta)=\\
&P(\sup_{n\leqslant k}\sqrt{k}\,|I(\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}<\epsilon)-1|>c\eta,\\
&\sup_{n\leqslant k}\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}<\epsilon)+\\
&P(\sup_{n\leqslant k}\sqrt{k}\,|I(\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}<\epsilon)-1|>c\eta,\\
&\sup_{n\leqslant k}\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}\geqslant\epsilon)\leqslant\\
&P(\sup_{n\leqslant k}\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}\geqslant\epsilon)
\end{aligned} \tag{A13}$$

According to the strong consistency of $\tilde{\beta}_n$, it implies that $P(\sup_{n\leqslant k}|\tilde{\beta}_{kj}-\beta_0|\geqslant\eta_{0j})<\eta$ for large enough n. Hence,

$$\begin{aligned}
&P(\sup_{n\leqslant k}\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}\geqslant\epsilon)\leqslant\\
&P(\sup_{n\leqslant k}|\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}\geqslant\epsilon,\sup_{n\leqslant k}|\tilde{\beta}_{kj}-\beta_{0j}|>\eta)+\\
&P(\sup_{n\leqslant k}|\sqrt{k}\kappa\,|\tilde{\beta}_{kj}|^{-\gamma}\geqslant\epsilon,\sup_{n\leqslant k}|\tilde{\beta}_{kj}-\beta_{0j}|\leqslant\eta)\leqslant\\
&\eta+P(\sup_{n\leqslant k}(\sqrt{k}\kappa\,|\beta_{0j}|-\eta)^{-\gamma}\geqslant\epsilon)=\\
&\eta+P(\sup_{n\leqslant k}(\sqrt{k}\kappa\geqslant c\epsilon)
\end{aligned} \tag{A14}$$

Since $\sqrt{n}\kappa\rightarrow 0$ by definition of κ, then from (A13) and (A14) we get the inequality (A12). Combining (A11) and (A12) implies Lemma 2. □

Lemma 3 Let a be a constant such that $P(\chi_{p_0}{}^2\leqslant a^2)=1-\alpha$. Assume the conditions of Lemma 2 are satisfied, and let a_n be a sequence of positive constants satisfying the conditional probability $P(\chi^2_{p_0}\leqslant a_n^2\mid C_n)=1-\alpha$, where C_n denotes the set of first n observations $\{(y_i,x_i):i=1,\cdots,n\}$, then $a_n\rightarrow a$ as $n\rightarrow\infty$.

For the proof of Lemma 3 please see that of Ref. [5, Lemma 3].

Proof of Theorem 3 Let

$$f(n)=\frac{na^2\nu}{a_n^2\nu_n},\ t=\frac{a^2\nu}{d^2} \tag{A15}$$

Then Eq. (12) can be rewritten as

$$N=\min\{k:k\geqslant n_0 \text{ and } f(k)/t\geqslant 1\}.$$

Since $I_n(\epsilon)$ and $(Z_n(\dot{\mu}(z_i^{\mathrm{T}}\hat{\beta}_n)^2 w(z_i^{\mathrm{T}}\hat{\beta}_n))Z_n^{\mathrm{T}})/n$ converge almost surely to I_0 and Σ, respectively. Then ν_n converges to ν with probability one as $n\rightarrow\infty$. By using Ref. [12, Lemma 1],

$$1=\lim_{t\rightarrow\infty}\frac{f(N)}{t}=\lim_{d\rightarrow 0}\frac{d^2N}{\nu_N a_N^2}=\lim_{d\rightarrow 0}\frac{d^2N}{\nu a^2}\text{ a. s.} \tag{A16}$$

which implies (i). By the definition of R_n, we have

$$\begin{aligned}
&P(\beta=(\beta_1,\cdots,\beta_p)^{\mathrm{T}}\in R_N)=\\
&P\{(\beta_{N1}-\hat{\beta}_{N1})^{\mathrm{T}}\tilde{\Sigma}_{11}(N)(\beta_{N1}-\hat{\beta}_{N1})\leqslant\\
&\frac{d^2N}{\nu_N}\text{ and }\beta_j=0\text{ for }I_{nj}(\epsilon)=0\}
\end{aligned} \tag{A17}$$

where β_{N1} is sub-vector of β corresponding to $\hat{\beta}_{N1}$. From (A16), $\frac{d^2N}{\nu_N}\rightarrow a^2$ and $N/t\rightarrow 1$ with probability one as $t\rightarrow\infty$, then by Lemma 3, (ii) follows.

From (i), to prove (iii), it suffices to show that

$$\{d^2N:\ d\in(0,1)\} \tag{A18}$$

is uniformly integrable. By definition of L_η, we have

$$N=NI_{\{N>L_\eta\}}+NI_{\{N\leqslant L_\eta\}} \tag{A19}$$

On $\{n\geqslant L_\eta\}$, and by (A1) again, we have

$$\lambda_{\min}(\sum_{i=1}^n[\dot{\mu}(z_i^{\mathrm{T}}\hat{\beta}_n)]^2 w(z_i^{\mathrm{T}}\hat{\beta}_n)z_iz_i^{\mathrm{T}})\geqslant \lambda_{\min}(\sum_{i=1}^n g_{\min}\dot{\mu}_{\min}z_iz_i^{\mathrm{T}}) \tag{A20}$$

where $g_{\min}=\inf_i g(z_i)>0$ and $\dot{\mu}_{\min}=\inf_i\dot{\mu}_{\min}(z_i)>0$. For each $i\in N$, let

$$M_i=g_{\min}\dot{\mu}_{\min}z_iz_i^{\mathrm{T}}-E[g_{\min}\dot{\mu}_{\min}z_iz_i^{\mathrm{T}}] \tag{A21}$$

Define

$$L_1=\sup\Big\{n\geqslant 1:v^{\mathrm{T}}\sum_{i=1}^n M_iv\leqslant\frac{-n\rho^*}{2},\ \text{for some } v\in\mathbb{R}^p \text{ and } \|v\|=1\Big\} \tag{A22}$$

where $\rho^* = \rho\, g_{\min}\dot{\mu}_{\min}$. By Ref. [13, Lemma 3.1], it can be shown that $EL_1 < \infty$. Let $L = \max(L_\eta, L_1)$. Then, by definitions of L_η and L_1, on $\{n > L\}$,

$$\lambda_{\min}(\tilde{\Sigma}_n) \geqslant \lambda_{\min}(\sum_{i=1}^{n} g_{\min}\dot{\mu}_{\min} z_i z_i^{\mathrm{T}}) \geqslant$$

$$\lambda_{\min}(\sum_{i=1}^{n} M_i) + \lambda_{\min}(\sum_{i=1}^{n} E[g_{\min}\dot{\mu}_{\min} z_i z_i^{\mathrm{T}}]) \geqslant$$

$$-\frac{n\rho^*}{2} + n\rho^* = \frac{n\rho^*}{2} \tag{A23}$$

Therefore, it follows from Eq. (A19), for $d \in (0,1)$

$$d^2 N \leqslant d^2 \times \left(\left[\frac{2a^2}{\rho^* d^2}\right] + 1\right) + L_\eta = \frac{2a^2}{\rho^* d + L_\eta} \tag{A24}$$

where notation $[t]$ denotes the largest integer less than t for any $t \in \mathbb{R}$. Note that RHS of (A24) does not depend on d. Hence, $EL_\eta < \infty$ implies that $\{d^2 N: d \in (0,1)\}$ is uniformly integrable. Thus, the proof of (ⅲ) is completed. To prove conclusion (ⅳ) we denote that the integer part of $\frac{\nu a^2}{d^2}$ by N_0, which is not a random variable. From (ⅰ), we have that N/N_0 almost surely converges to 1 as d tends to 0, and

$$\hat{p}_0(N) - p_0 = \hat{p}_0(N) - \hat{p}_0(N_0) + \hat{p}_0(N_0) - p_0 =$$

$$\sum_{i=1}^{p} \{I(\sqrt{N}\kappa\, |\tilde{\beta}_{Nj}|^{-\gamma} < \epsilon) - I(\sqrt{N}\kappa\, |\tilde{\beta}_{N_0 j}|^{-\gamma} < \epsilon)\} +$$

$$\hat{p}_0(N_0) - p_0 \tag{A25}$$

We know that $\hat{p}_0(N_0) - p_0$ almost surely converges to 0 as d tends to 0. Thence, to satisfy (ⅳ) we only need to prove that $I(\sqrt{N}\kappa\,|\tilde{\beta}_{Nj}|^{-\gamma} < \epsilon) - I(\sqrt{N_0}\kappa|\tilde{\beta}_{N_0 j}|^{-\gamma} < \epsilon)$ almost surely converges to 0, that is, we need to validate that $\tilde{\beta}_N - \tilde{\beta}_{N_0}$ almost surely converges to 0 as d tends to 0. Without loss of generality, take $d = 1/k$ for some integer k and denote $N = N(1/k)$, $N_0 = N_0(1/k)$, then

$$P(\sup_{n\leqslant k} |\tilde{\beta}_{N(\frac{1}{k})} - \tilde{\beta}_{N_0(\frac{1}{k})}| > \eta) =$$

$$P\Big(\sup_{n\leqslant k} |\tilde{\beta}_{N(\frac{1}{k})} - \tilde{\beta}_{N_0(\frac{1}{k})}| > \eta,$$

$$\sup_{n\leqslant k} |N\left(\frac{1}{k}\right) - N_0\left(\frac{1}{k}\right)| \leqslant \xi\Big) +$$

$$P\Big(\sup_{n\leqslant k} |\tilde{\beta}_{N(\frac{1}{k})} - \tilde{\beta}_{N_0(\frac{1}{k})}| > \eta,$$

$$\sup_{n\leqslant k} |N\left(\frac{1}{k}\right) - N_0\left(\frac{1}{k}\right)| > \xi\Big) \leqslant$$

$$P\Big(\sup_{n\leqslant k} |\tilde{\beta}_{N(\frac{1}{k})} - \tilde{\beta}_{N_0(\frac{1}{k})}| > \eta,$$

$$N_0\left(\frac{1}{k}\right) - \xi \leqslant N\left(\frac{1}{k}\right) \leqslant N_0\left(\frac{1}{k}\right) + \xi,$$

$$\text{for all } k \geqslant n\Big) +$$

$$P\Big(\sup_{n\leqslant k} |N\left(\frac{1}{k}\right) - N_0\left(\frac{1}{k}\right)| > \xi\Big) \leqslant$$

$$P(\sup_{n\leqslant k} \sup_{N_0(\frac{1}{k}) - \xi \leqslant L \leqslant N_0(\frac{1}{k}) + \xi} |\tilde{\beta}_L - \tilde{\beta}_{N_0(\frac{1}{k})}| > \eta) +$$

$$\eta/2 \leqslant$$

$$P(\sup_{n\leqslant k} \sup_{N_0(\frac{1}{k}) - \xi \leqslant L \leqslant N_0(\frac{1}{k}) + \xi} |\tilde{\beta}_L - \beta_0| > \eta/2) +$$

$$P(\sup_{n\leqslant k} \sup_{N_0(\frac{1}{k}) - \xi \leqslant L \leqslant N_0(\frac{1}{k}) + \xi} |\tilde{\beta}_{N_0(\frac{1}{k})} - \beta_0| > \eta/2) +$$

$$\eta/2 \leqslant$$

$$\eta/4 + \eta/4 + \eta/2 = \eta \to 0 \tag{A26}$$

It is clear that $\{I_{N(d)j}(\epsilon): d \in (0,1)\}$, for each $j = 1, \cdots, p$, is uniformly integrable. Then the last part of (ⅳ) of Theorem 3 follows from the dominated convergence theorem. □

References

[1] Nelder J A, Wedderburn R W M. Generalized linear models[J]. Journal of the Royal Statistical Society, Series A, 1972, 135(3): 370-384.

[2] Fahrmeir L, Kaufmann H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models[J]. The Annals of Statistics, 1985, 13(1): 342-368.

[3] McCullagh P, Nelder J A. Generalized Linear Models [M]. 2nd ed. New York: Chapman & Hall, 1989.

[4] Chen K, Hu I, Ying Z. Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs[J]. The Annals of Statistics, 1999, 27(4): 1 155-1 163.

[5] Wang Z F, Chang Y C I. Sequential estimate for linear regression models with uncertain number of effective variables[J]. Metrika, 2013, 76:949-978.

[6] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society, Series B, 1996, 58(1): 267-288.

[7] Efron B, Hastie T, Johnstone I, et al. Least angle regression[J]. The Annals of Statistics, 2004, 32(2): 407-499.

（下转第 496 页）