Journal of the Korean
Statistical Society
Vol. 24, No. 1, 1995

Statistical Estimation for Generalized Logit Model of Nominal Type with Bootstrap Method †

JoongJae Cho[1] and JeongHye Han[2]

ABSTRACT

The generalized logit model of nominal type with random regressors is studied for bootstrapping. In particular, asymptotic normality and consistency of bootstrap model estimators are derived. It is shown that the bootstrap approximation to the distribution of the maximum likelihood estimators is valid for almost all sample sequences.

KEYWORDS : Generalized logit model, Method of maximum likelihood, Information matrix, Newton-Raphson method, Bootstrap method

1. INTRODUCTION

Some asymptotic theory for applications of Efron's(1979) bootstrap to the generalized logit model of nominal type with random regressors case, which

† This paper was supported (in part) by NON-DIRECTED RESEARCH FUND, Korea Research Foundation, 1992

[1]Department of Statistics, Chungbuk National University, Cheongju, 360-763, Korea
[2]Department of Statistics, Chungbuk National University, Cheongju, 360-763, Korea

is frequently used in modelling the polytomous response relationship, will be developed. McCullagh and Nelder(1989) gives a good overview of the context. See also Fahrmeir and Kaufmann(1985) for consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. The bootstrap approximation to the distribution of the least squares estimators in linear models context is studied by Freedman(1981), and the less metrical version of the same results can be found in Beran(1984) with more examples. A nice review of bootstrap methods developed so far is given by Hinkley(1988), where numerous references are cited. Hosmer and Lemeshow(1989) considered the generalized logit model with some applications. The bootstrap results for some generalized linear models have been studied by Lee(1990), and Lee, Kim, Sohn and Jeong(1992).

In section 2 we introduce the generalized logit model with some asymptotic properties of the maximum likelihood estimators. In particular, asymptotic normality and strong consistency of model estimators are studied. Section 3 gives bootstrapping generalized logit model of nominal type with random regressors. That is, we present the ways of bootstrapping and its validity for almost all sample sequences.

2. MODEL AND STATISTICAL ESTIMATION

2.1. Model and the Method of Estimation

Logistic regression is most frequently employed to model the relationship between a dichotomous (or binary) outcome variable and a set of covariates, but with a few modifications it may also be used when the outcome variable is polytomous. The extension of the model and methods for a binary outcome variable to a polytomous outcome variable is easily illustrated when the outcome variable has $(m+1)$ categories. In developing models for a polytomous outcome variable we need to be aware of its measurement scale. Most applications, and hence the focus of the material in this paper involve a nominal scaled outcome variable. Methods are available for modeling an ordinal scale outcome variable but we will not present them. Assume that the categories of the outcome variable, Y, are coded $0, 1, 2, \cdots, m$ and that a set of covariates $X_0, X_1, X_2, \cdots, X_p$ are continuous (interval scale) random variables. The generalized logit model of nominal type specifies the relationship between the response variable Y and covariates $\mathbf{X}_i$'s as follows ($X_0 \equiv 1$) :

$$\log \frac{P(Y_i = j \mid x_{i1}, x_{i2}, \cdots, x_{ip})}{P(Y_i = 0 \mid x_{i1}, x_{i2}, \cdots, x_{ip})} = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \cdots + \beta_{jp}x_{ip} = \mathbf{x}_i^t\boldsymbol{\beta}_j, \tag{2.1}$$

where $\mathbf{x}_i^t = (1, x_{i1}, x_{i2}, \cdots, x_{ip})$ is the i-th vector of explanatory variables and $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \cdots, \beta_{jp})^t$ is a vector of parameters of interest ($i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, m$).

Let us introduce formally our models mentioned above. Suppose that $(Y_i, \mathbf{X}_i^t), i = 1, 2, \cdots, n$ are independent and identically distributed random vectors such that $\mathbf{X}_i = (X_{i0}, X_{i1}, X_{i2}, \cdots, X_{ip})^t$ is a $(p+1)$-variate random vector with common distribution function $G(\cdot)$, which is unknown, and there exists a positive number M such that $\|\mathbf{X}_i\| \leq M$ with probability one, where $\|\cdot\|$ is the Euclidean norm. In convenience, the reponse variables are coded as follows: if $Y_i = j$ then $Y_{ij} = 1$ and $Y_{ij'} = 0$ ($j \neq j'$; $j, j' = 0, 1 \cdots, m$).

Now, we turn to the problem of statistical estimation for our model. A number of estimators for $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t, \cdots, \boldsymbol{\beta}_m^t)^t$ are available. However, in this paper we consider the maximum likelihood estimator (MLE) only.

Given $\mathbf{X}_i = \mathbf{x}_i$, the conditional log-likelihood for a single observation is given by

$$\begin{aligned} l_i \equiv l(\boldsymbol{\beta}, G; y_i, \mathbf{x}_i) &= \log[P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \cdots, Y_{im} = y_{im} \mid \mathbf{X}_i = \mathbf{x}_i)] \\ &= \log[\prod_{j=0}^{m} \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)^{y_{ij}}] \\ &= \sum_{j=1}^{m} y_{ij}\mathbf{x}_i^t\boldsymbol{\beta}_j - \log(1 + \sum_{j=1}^{m} \exp(\mathbf{x}_i^t\boldsymbol{\beta}_j)), \end{aligned} \tag{2.2}$$

where $y_{ij} = 0$ or 1 which satisfies $\sum_{j=0}^{m} y_{ij} = 1$, and for $j = 1, 2, \cdots, m$

$$\begin{aligned} \pi_{i0}(\boldsymbol{\beta}_o : \mathbf{x}_i) &= P(Y_i = 0 \mid \mathbf{X}_i = \mathbf{x}_i) = \frac{1}{1 + \sum_{l=1}^{m} \exp(\mathbf{x}_i^t\boldsymbol{\beta}_l)} \\ \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i) &= P(Y_i = j \mid \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^t\boldsymbol{\beta}_j)}{1 + \sum_{l=1}^{m} \exp(\mathbf{x}_i^t\boldsymbol{\beta}_l)}. \end{aligned} \tag{2.3}$$

Let $\frac{\partial l_i}{\partial \boldsymbol{\beta}_j}$ be a vector consisting of the first derivative of $l(\boldsymbol{\beta}, G : y_i, \mathbf{x}_i)$ with respect to $\boldsymbol{\beta}_j$, then it can be easily seen, using chain rule, that

$$\frac{\partial l_i}{\partial \boldsymbol{\beta}_j} = \mathbf{x}_i[y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)],$$

since it follows from $\frac{\partial l_i}{\partial \beta_{jk}} = \mathbf{x}_{ik}[y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)]$ $(j = 1, 2, \cdots, m;\ k = 0, 1, 2, \cdots, p)$.

Hence, for the log-likelihood function l for our random sample $(Y_i, \mathbf{X}_i^t)$, $i = 1, 2, \cdots, n$, we obtain the first derivative $\frac{\partial l}{\partial \boldsymbol{\beta}}$ as follows :

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}_1} \\ \frac{\partial l}{\partial \boldsymbol{\beta}_2} \\ \vdots \\ \frac{\partial l}{\partial \boldsymbol{\beta}_m} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \frac{\partial l_i}{\partial \boldsymbol{\beta}_1} \\ \sum_{i=1}^n \frac{\partial l_i}{\partial \boldsymbol{\beta}_2} \\ \vdots \\ \sum_{i=1}^n \frac{\partial l_i}{\partial \boldsymbol{\beta}_m} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n \mathbf{X}_i(Y_{i1} - \pi_{i1}(\boldsymbol{\beta}_1 : \mathbf{X}_i)) \\ \sum_{i=1}^n \mathbf{X}_i(Y_{i2} - \pi_{i2}(\boldsymbol{\beta}_2 : \mathbf{X}_i)) \\ \vdots \\ \sum_{i=1}^n \mathbf{X}_i(Y_{im} - \pi_{im}(\boldsymbol{\beta}_m : \mathbf{X}_i)) \end{pmatrix}_{m(p+1)\times 1.} \tag{2.4}$$

Therefore, the solution $\hat{\boldsymbol{\beta}}_n$ of the likelihood equation

$$\frac{\partial l}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\sum_{i=1}^n l(\boldsymbol{\beta}_n, G : Y_i, \mathbf{X}_i)\right) = \mathbf{0},$$

that is,

$$\sum_{i=1}^n \mathbf{X}_{ik}\,(Y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)) = \mathbf{0}, \qquad k = 1, 2, \cdots, m, \tag{2.5}$$

which can be obtained by an iterative method (see McCullagh and Nelder(1989)) is asymptotically efficient, where $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{1n}, \hat{\boldsymbol{\beta}}_{2n}, \cdots, \hat{\boldsymbol{\beta}}_{mn})$; $\hat{\boldsymbol{\beta}}_{jn}^t = (\hat{\beta}_{j0}, \hat{\beta}_{j1}, \cdots, \hat{\beta}_{jp})$, $j = 1, 2, \cdots, m$.

The first derivative $\frac{\partial l}{\partial \beta}$ does not depend on G, so the efficient score function for estimating $\boldsymbol{\beta}$ when G is unknown is the same as that when G is known. Since above equations are non-linear, it is impossible to solve them with explicit form. We will find MLE $\hat{\boldsymbol{\beta}}_n$ by using a iterative procedure, such as Newton-Raphson Method. In general, some typical algorithms for non-linear equations are introduced in McCullagh and Nelder(1989). From the first derivative $\frac{\partial l_i}{\partial \beta_j}$ we obtain the second derivative matrix $\nabla^2 l_i$(the Hessian matrix) of order $m(p+$

1) $\times\, m(p+1)$ as follows :

$$\nabla^2 l_i = \left(\frac{\partial^2 l_i}{\partial\beta_{jk}\partial\beta_{j'k'}}\right),\ (j,j' = 1,2,\cdots,m;\ k,k' = 0,1,2,\cdots,p). \tag{2.6}$$

Also, each component of the matrix is given by

$$\begin{aligned}\frac{\partial^2 l_i}{\partial\beta_{jk'}\partial\beta_{jk}} &= \frac{\partial}{\partial\beta_{jk'}}(\frac{\partial l_i}{\partial\beta_{jk}}) \\ &= -x_{ik}x_{ik'}\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)[1-\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l_i}{\partial\beta_{j'k'}\partial\beta_{jk}} &= \frac{\partial}{\partial\beta_{j'k'}}(\frac{\partial l_i}{\partial\beta_{jk}}) \\ &= x_{ik}x_{ik'}\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{x}_i)\pi_{ij'}(\boldsymbol{\beta}_{j'} : \mathbf{x}_i)\end{aligned}$$

$(j \neq j';\ j,j' = 1,2,\cdots,m;\ k,k' = 0,1,2,\cdots,p)$. Hence, the information matrix $I(\boldsymbol{\beta})$ regarding β, denoted by $I(\boldsymbol{\beta})$, is given by

$$I(\boldsymbol{\beta}) = \begin{pmatrix} I(\boldsymbol{\beta})_{11} & & & \\ I(\boldsymbol{\beta})_{21} & I(\boldsymbol{\beta})_{22} & & Symmetric \\ \vdots & \vdots & \ddots & \\ I(\boldsymbol{\beta})_{m1} & I(\boldsymbol{\beta})_{m2} & \cdots & I(\boldsymbol{\beta})_{mm} \end{pmatrix}_{m(p+1)\times m(p+1)}, \tag{2.7}$$

where $I(\boldsymbol{\beta})_{jj} = E[\mathbf{X}_i\mathbf{X}_i^t\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)(1-\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i))]$ is $(p+1)\times(p+1)$-matrix, and $I(\boldsymbol{\beta})_{jj'} = -E[\mathbf{X}_i\mathbf{X}_i^t\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)\pi_{ij'}(\boldsymbol{\beta}_{j'} : \mathbf{X}_i)] = I(\boldsymbol{\beta})_{j'j}$ is $(p+1)\times(p+1)$-matrix $(j \neq j';\ j,j' = 1,2,\cdots,m)$. Also, these two matrices $I(\boldsymbol{\beta}_{jj})$ and $I(\boldsymbol{\beta}_{jj'})$ are represented as follows :

$$I(\boldsymbol{\beta})_{jj} = E[\mathbf{X}_i\mathbf{X}_i^t(Y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)^2]$$

$$I(\boldsymbol{\beta})_{jj'} = E[\mathbf{X}_i\mathbf{X}_i^t(Y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i))(Y_{ij'} - \pi_{ij'}(\boldsymbol{\beta}_{j'} : \mathbf{X}_i))].$$

The reason is due to the following facts(Lehmann ; 1983) :

$$-E(\frac{\partial^2 l_i}{\partial\beta_{jk'}\partial\beta_{jk}}) = E[(\frac{\partial l_i}{\partial\beta_{jk}})(\frac{\partial l_i}{\partial\beta_{jk'}})]$$
$$-E(\frac{\partial^2 l_i}{\partial\beta_{j'k'}\partial\beta_{jk}}) = E[(\frac{\partial l_i}{\partial\beta_{j'k'}})(\frac{\partial l_i}{\partial\beta_{jk}})]$$

$(j \neq j'; j, j' = 1, 2, \cdots, m; k, k' = 0, 1, 2, \cdots, p)$. Out of necessity, we cite the following important fact(Fahrmeir and Kasufmann ;1985). Assume that a solution $\hat{\boldsymbol{\beta}}_n$ corresponding to the equation (2.5) is the maximum likelihood estimator of $\boldsymbol{\beta}$ for our model.

Lemma 1. Suppose there does not exist a nonzero vector $\mathbf{a}_j^t = (a_{j0}, a_{j1}, \cdots, a_{jp})^t$ and a real number c_j $(j = 1, 2, \cdots, m)$ such that for $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, m$

$$\mathbf{a}_j^t \mathbf{X}_i [Y_{ij} - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)] = c_j.$$

Then

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \xrightarrow{w.p.1} \mathbf{0} \tag{2.8}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I^{-1}(\boldsymbol{\beta})). \tag{2.9}$$

Of course, maximum likelihood estimator $\hat{\boldsymbol{\beta}}_n$ will be derived through the iterative formula

$$\boldsymbol{\beta}_n^{(k)} = \boldsymbol{\beta}_n^{(k-1)} + \frac{1}{n}[H_j(\boldsymbol{\beta}_n^{(k-1)} ; \mathbf{x})]^{-1} \frac{\partial l}{\partial \boldsymbol{\beta}} \mid_{\beta = \beta_n^{(k-1)}} , \tag{2.10}$$

where the superscript on $\boldsymbol{\beta}$ indicates the iteration number$(j = 1, 2)$ and two matrices $H_j(\hat{\boldsymbol{\beta}}_{nn}; \mathbf{X})$, $j = 1, 2$ are reasonable estimators of information matrix $I(\boldsymbol{\beta}_n)$ which are to be discussed in the following section 2.2.

2.2. Some Asymptotic Properties

When maximum likelihood estimators of the coefficients in a nonlinear model such as the logistic model are obtained, there are a number of asymptotically equivalent covariance matrix estimators that can be used. These covariance matrix estimators are typically associated with different computer algorithms to solve the likelihood equation (2.5). There are at least two natural ways of estimating the matrix $I(\boldsymbol{\beta})$ which inverse is related to the asymptotic covariance matrix of MLE $\hat{\boldsymbol{\beta}}_n$. The two estimators considered will be denoted by $H_j(\hat{\boldsymbol{\beta}}_n : \mathbf{X})$ $(j = 1, 2)$.

First, taking into consideration the matrix $I(\boldsymbol{\beta})$, we have two reasonable estimation matrices $H_k(\hat{\boldsymbol{\beta}}_n : \mathbf{X}), k = 1,2$ for the matrix $I(\boldsymbol{\beta})$, which are given by

$$H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X}) = \begin{pmatrix} H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{11} & & & \\ H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{21} & H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{22} & & Symmetric \\ \vdots & \vdots & \ddots & \\ H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{m1} & H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{m2} & \cdots & H_k(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{mm} \end{pmatrix} \quad (2.11)$$

for $j \neq j'; j, j' = 1, 2, \cdots, m$, where some notations as follows.

$$\hat{\pi}_{ij} = \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i)$$

$$H_1(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{jj} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t\hat{\pi}_{ij}(1-\hat{\pi}_{ij})$$

$$H_1(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{jj'} = -\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t\hat{\pi}_{ij}\hat{\pi}_{ij'}$$

$$H_2(\hat{\boldsymbol{\beta}}_n;\mathbf{X})_{jj} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t(Y_{ij}-\hat{\pi}_{ij})^2$$

$$H_2(\hat{\boldsymbol{\beta}}_n:\mathbf{X})_{jj'} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t(Y_{ij}-\hat{\pi}_{ij})(Y_{ij'}-\hat{\pi}_{ij'})$$

$$\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{in}^t, \hat{\boldsymbol{\beta}}_{2n}^t, \cdots, \hat{\boldsymbol{\beta}}_{mn}^t)^t,\ \hat{\boldsymbol{\beta}}_{jn}^t = (\hat{\beta}_{j0}, \hat{\beta}_{j1}, \cdots, \hat{\beta}_{jp}),\ j = 1, 2, \cdots, m.$$

Theorem 1. Under the assumptions of Lemma 1, we obtain the following two statements; For each $i = 1, 2$, as $n \to \infty$

$$H_i(\hat{\boldsymbol{\beta}}_n : \mathbf{X}) \xrightarrow{w.p.1} I(\boldsymbol{\beta}) \quad (2.12)$$

$$\sqrt{n}H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n : \mathbf{X})(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I_{m(p+1)\times m(p+1)}), \quad (2.13)$$

where $H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n : \mathbf{X})[H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n : \mathbf{X})]^t = H_i(\hat{\boldsymbol{\beta}}_n : \mathbf{X})$.

Proof. Consider the j-th matrix $H_1(\boldsymbol{\beta};\mathbf{X})_{jj}$ diagonally of the matrix $H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})$, denoted by $H_1(\boldsymbol{\beta};\mathbf{X})_{jj} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)(1-\pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i))$ for $j = 1, 2, \cdots, m$. For the sake of convenience, we decompose the difference matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj} - I(\boldsymbol{\beta})_{jj}]$ for $j = 1, 2, \cdots, m$ as follows :

$$H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj} - I(\boldsymbol{\beta})_{jj} = [H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj} - H_1(\boldsymbol{\beta} : \mathbf{X})_{jj}] + [H_1(\boldsymbol{\beta} : \mathbf{X})_{jj} - I(\boldsymbol{\beta})_{jj}]$$

Fortunately, the second term in the right side of above equation converges to the zero matrix, denoted by $\mathbf{0}_{(p+1)\times(p+1)}$ by the strong law of the large numbers(SLLN). And by Taylor's expansion of the multi-variable function, the first term is

$$\begin{aligned}
[H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj}-H_1(\boldsymbol\beta:\mathbf{X})_{jj}]_{l,k} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{il}\mathbf{X}_{ik}[\hat\pi_{ij}(1-\hat\pi_{ij})-\pi_{ij}(1-\pi_{ij})] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{il}\mathbf{X}_{ik}[\sum_{l'\neq j}^{m}(\hat{\boldsymbol\beta}_{l'n}-\boldsymbol\beta_{l'})^t \\
&\quad \mathbf{X}_i(-\tilde\pi_{ij}\tilde\pi_{il'}+2\tilde\pi_{ij}^2\tilde\pi_{il'}) \\
&\quad +(\hat{\boldsymbol\beta}_{jn}-\boldsymbol\beta_j)^t\mathbf{X}_i(\tilde\pi_{ij}-3\tilde\pi_{ij}^2+2\tilde\pi_{ij}^3)],
\end{aligned}$$

since the (l,k)-th element of the matrix $H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj}$ can be expanded around $\boldsymbol\beta$, where $\hat\pi_{ij}=\pi_{ij}(\hat{\boldsymbol\beta}_{jn}:\mathbf{X}_i)$, $\pi_{ij}=\pi_{ij}(\boldsymbol\beta_{jn}:\mathbf{X}_i)$, $\tilde\pi_{ij}=\pi_{ij}(\tilde{\boldsymbol\beta}_{jn}:\mathbf{X}_i)$, and $\tilde{\boldsymbol\beta}_{jn}$ which satisfies $\|\tilde{\boldsymbol\beta}_{jn}-\boldsymbol\beta_j\|\leq\|\hat{\boldsymbol\beta}_{jn}-\boldsymbol\beta_j\|$. Since the vector $\hat{\boldsymbol\beta}_n-\boldsymbol\beta$ converges to $\mathbf{0}_{m(p+1)\times 1}$ by Lemma 1, $\hat{\boldsymbol\beta}_{jn}-\boldsymbol\beta_j$ converges to zero vector $\mathbf{0}_{(p+1)\times 1}$ $(j=1,2,\cdots,m)$. The (l,k)-th element of the difference matrix $[H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj}-H_1(\boldsymbol\beta:\mathbf{X})_{jj}]$ converges to 0 since all the column vectors are bounded from above except the row vectors $(\hat{\boldsymbol\beta}_{jn}-\boldsymbol\beta_j)^t$, $(j=1,2,\cdots,m)$. Hence, the first matrix $[H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj}-H_1(\boldsymbol\beta:\mathbf{X})_{jj}]$ converges to $\mathbf{0}_{(p+1)\times(p+1)}$ and the matrix $H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj}$ converges to $I(\boldsymbol\beta)_{jj}$ with probability one $(j=1,2,\cdots,m)$.

On the other hand, define $H_1(\boldsymbol\beta:\mathbf{X})_{jj'}=-\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^t\pi_{ij}(\boldsymbol\beta_j:\mathbf{X})\pi_{ij'}(\boldsymbol\beta_{j'}:\mathbf{X})$ for $j\neq j'$; $j,j'=1,2,\cdots,m$ and consider the following decomposed matrix:

$$H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj'}-I(\boldsymbol\beta)_{jj'}=[H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj'}-H_1(\boldsymbol\beta:\mathbf{X})_{jj'}]+[H_1(\boldsymbol\beta:\mathbf{X})_{jj'}-I(\boldsymbol\beta)_{jj'}].$$

Then, the second matrix in the right side converges to $\mathbf{0}_{(p+1)\times(p+1)}$ by SLLN and the (l,k)-th element of the first can be expanded by Taylor's expansion as below :

$$\begin{aligned}
&[H_1(\hat{\boldsymbol\beta}_n:\mathbf{X})_{jj'}-H_1(\boldsymbol\beta_n:\mathbf{X})_{jj'}]_{l,k} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{il}\mathbf{X}_{ik}[\hat\pi_{ij}\hat\pi_{ij'}-\pi_{ij}\pi_{ij'}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{il}\mathbf{X}_{ik}\{\sum_{j\neq l'\neq j'}^{m}(\hat{\boldsymbol\beta}_{l'n}-\boldsymbol\beta_{l'})^t\mathbf{X}_i(-2\tilde\pi_{ij}\tilde\pi_{ij'}\tilde\pi_{il'})
\end{aligned}$$

$$+(\hat{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_j)^t \mathbf{X}_i(\tilde{\pi}_{ij}\tilde{\pi}_{ij'} - 2\tilde{\pi}_{ij}^2\tilde{\pi}_{ij'})$$
$$+(\hat{\boldsymbol{\beta}}_{j'n} - \boldsymbol{\beta}_{j'})^t \mathbf{X}_i(\tilde{\pi}_{ij}\tilde{\pi}_{ij'} - 2\tilde{\pi}_{ij}\tilde{\pi}_{ij'}^2)\},$$

where $\hat{\pi}_{ij} = \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i)$, $\pi_{ij} = \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i)$, $\tilde{\pi}_{ij} = \pi_{ij}(\tilde{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i)$,and $\tilde{\boldsymbol{\beta}}_{jn}$ which satisfies $\|\tilde{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_j\| \le \|\hat{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_j\|$ for $j = 1, 2, \cdots, m$. And the vector $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ converges to $\mathbf{0}_{m(p+1)\times 1}$ vector as n converges to infinity. All the (l, k)-th elements of $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'} - H_1(\boldsymbol{\beta} : \mathbf{X})_{jj'}]$ converge to 0 since the column vectors are bounded from above except the row vectors $(\hat{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_j)^t$, $(j = 1, 2, \cdots, m)$. Hence, the first matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'} - H_1(\boldsymbol{\beta} : \mathbf{X})_{jj'}]$ converges to the matrix $\mathbf{0}_{(p+1)\times(p+1)}$,and the matrix $H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'}$ converges to the matrix $I(\boldsymbol{\beta})_{jj'}$ with probability one ($j \ne j'$; $j, j' = 1, 2, \cdots, m$). Consequently, we obtain the following result by above fact, Lemma 1, and Slutsky's Theorem :

$$\sqrt{n} H_1^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n : \mathbf{X})(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I_{m(p+1)\times m(p+1)}).$$

Similarly, the results related to the estimation matrix $H_2(\hat{\boldsymbol{\beta}}_n : \mathbf{X})$ would be proved. □

3. BOOTSTRAP APPROXIMATION

3.1. Bootstrap Algorithm

Suppose $(Y_i, \mathbf{X}_i^t)$, $i = 1, 2, \cdots, n$ are independent and identically distributed random vectors. Then, bootstrap estimators would be considered by the empirical c.d.f $\hat{G}_n(\cdot)$ computed from the covariates $\mathbf{X}_i$, $i = 1, 2, \cdots, n$ and MLE $\hat{\boldsymbol{\beta}}_n$. The bootstrap algorithm goes as follows:

- Step 1 : From the original sample $(Y_i, \mathbf{X}_i)$, $i = 1, 2, \cdots, n$, compute the MLE $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ where $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{1n}^t, \hat{\boldsymbol{\beta}}_{2n}^t, \cdots, \hat{\boldsymbol{\beta}}_{mn}^t)$; $\hat{\boldsymbol{\beta}}_{jn}^t = (\hat{\beta}_{j0}, \hat{\beta}_{j1}, \cdots, \hat{\beta}_{jp})$, and then construct an empirical distribution function, $\hat{G}_n(\cdot)$ based on the covariates $\mathbf{X}_i$'s.

- Step 2 : Choose $\mathbf{X}^* = \left[\binom{Y_1^*}{\mathbf{X}_1^*}, \binom{Y_2^*}{\mathbf{X}_2^*}, \cdots \binom{Y_n^*}{\mathbf{X}_n^*}\right]$ from the fitted model $P_{\hat{\theta}_n, n}$, where $\hat{\theta}_n = (\hat{G}_n, \hat{\beta}_n)$ as follows. Let $\mathbf{X}_i^*, i = 1, 2, \cdots, n$ be a bootstrap sample of size n from $\hat{G}_n(\cdot)$, that is, a simple random sample with replacement from $(X_1, X_2, \cdots, X_n)$. Then given $\mathbf{X}_i^* = \mathbf{x}_i^*$, choose Y_{ij}^* having the following distribution function $(i = 1, 2, \cdots, n;\ j = 0, 1, 2, \cdots, m)$:

$$P(Y_{i1}^* = y_{i1}^*, Y_{i2}^* = y_{i2}^*, \cdots, Y_{im}^* = y_{im}^*, \mid \mathbf{X}_i^* = \mathbf{x}_i^*) = \prod_{j=0}^{m} [\pi_{ij}(\hat{\beta}_{jn} : \mathbf{X}_i^*)]^{y_{ij}}, \tag{3.1}$$

where $y_{ij}^* = 0$ or 1 which satisfies $\sum_{j=0}^{m} y_{ij}^* = 1$, and

$$\pi_{ij}(\hat{\beta}_{jn} : \mathbf{X}_i^*) = \frac{\exp(\mathbf{X}_i^{*t} \hat{\beta}_{jn})}{1 + \sum_{l=1}^{m} \exp(\mathbf{X}_i^{*t} \hat{\beta}_{ln})}$$

which satisfies $\sum_{j=0}^{m} \pi_{ij}(\hat{\beta}_{jn} : \mathbf{X}_i^*) = 1$.

If $Y_{ij}^* = 1$ and $Y_{ij'}^* = 0$ for $j \neq j'; j, j' = 1, 2, \cdots, m$, then we denote it by $Y_i^* = j$.

- Step 3 : From the bootstrap sample $\mathbf{X}^* = \left[\binom{Y_1^*}{\mathbf{X}_1^*}, \binom{Y_2^*}{\mathbf{X}_2^*}, \cdots \binom{Y_n^*}{\mathbf{X}_n^*}\right]$, compute the bootstrap MLE $\hat{\boldsymbol{\beta}}_n^*$. That is, $\hat{\boldsymbol{\beta}}_n^*$ is the solution of the following non-linear equation :

$$\sum_{i=1}^{n} \mathbf{X}_i^* [Y_{ij}^* - \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i^*)] = \mathbf{0}, \quad (j = 1, 2, \cdots, m). \tag{3.2}$$

3.2. Some properties of bootstrap estimators

Define the log likelihood function as $l^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} l(\boldsymbol{\beta} : Y_i^*, \mathbf{X}_i^*)$, and the first derivative $\dot{l}^*(\boldsymbol{\beta}) = \frac{\partial l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, and the second derivative $\ddot{l}^*(\boldsymbol{\beta}) = \frac{\partial^2 l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t}$ in bootstrap version. Let $N_\epsilon(\boldsymbol{\beta})$ be a $m(p+1)$-dimensional open ball in $R^{m(p+1)}$ centered at $\boldsymbol{\beta}$ with positive radius ϵ for convenience's sake. After proving three lemmas we will present Theorem 2.

Lemma 2. As $n \to \infty$, we obtain

$$\frac{1}{\sqrt{n}} \dot{l}^*(\hat{\boldsymbol{\beta}}_n) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I(\boldsymbol{\beta})). \tag{3.3}$$

Proof. For any vector of constants $\mathbf{a} = (\mathbf{a}_1^t, \mathbf{a}_2^t, \cdots, \mathbf{a}_m^t)^t$, $\mathbf{a}_j^t = (a_{j1}, a_{j2}, \cdots, a_{jp+1})$; $j = 1, 2, \cdots, m$, we get

$$\frac{1}{\sqrt{n}}\mathbf{a}^t \dot{l}^*(\hat{\boldsymbol{\beta}}_n) = \sum_{i=1}^n U_{in}^*,$$

where $U_{in}^* = \frac{1}{\sqrt{n}}\sum_{j=1}^m \mathbf{a}_j^t \mathbf{X}_i^*[Y_{ij}^* - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)]$, $(i = 1, 2, \cdots, n)$. Then we obtain the conditional expectation $E(U_{in}^*)$ and the conditional variance $Var(U_{in}^*)$ given $\mathbf{X}$ and Y as follows :

$$\begin{aligned}
E(U_{in}^*) \equiv E(U_{in}^*|\mathbf{X}, Y) &= E[E(U_{in}^*|\mathbf{X}_i^*)] \\
&= \frac{1}{\sqrt{n}} E\left[\sum_{j=1}^m \mathbf{a}_j^t \mathbf{X}_i^* E\left[[Y_{ij}^* - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)] \mid \mathbf{X}_i^*\right]\right] \\
&= \frac{1}{\sqrt{n}} E\left[\sum_{j=1}^m \mathbf{a}_j^t \mathbf{X}_i^* \left[P(Y_i^* = j \mid \mathbf{X}_i^*) - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)\right]\right] \\
&= \mathbf{0},
\end{aligned}$$

$$\begin{aligned}
& Var(U_{in}^*) \\
&\equiv Var(U_{in}^*|\mathbf{X}, Y) = E(U_{in}^{*2}) = E\left[E(U_{in}^{*2}|\mathbf{X}_i^*)\right] \\
&= E\left[E\left[\frac{1}{n}\sum_{j=1}^m \sum_{j'=1}^m \mathbf{a}_j^t \mathbf{X}_i^*[Y_{ij}^* - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)][Y_{ij'}^* - \pi_{ij'}(\hat{\boldsymbol{\beta}}_{j'n} : \mathbf{X}_i^*)]\mathbf{X}_i^{*t}\mathbf{a}_{j'}|\mathbf{X}_i^*\right]\right] \\
&= \frac{1}{n} E\left\{\sum_{j=1}^m \mathbf{a}_j^t \mathbf{X}_i^* \mathbf{X}_i^{*t} \mathbf{a}_j \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)[1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)]\right. \\
&\quad \left. - \sum_{j=1}^m \sum_{j' \neq j}^m \mathbf{a}_j^t \mathbf{X}_i^* \mathbf{X}_i^{*t} \mathbf{a}_{j'} \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*) \pi_{ij'}(\hat{\boldsymbol{\beta}}_{j'n} : \mathbf{X}_i^*)\right\} \\
&= \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^n \left[\sum_{j=1}^m \mathbf{a}_j^t \mathbf{X}_i \mathbf{X}_i^t \mathbf{a}_j \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^t)[1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^t])\right]\right. \\
&\quad \left. - \frac{1}{n}\sum_{i=1}^n \left[\sum_{j=1}^m \sum_{j' \neq j}^m \mathbf{a}_j^t \mathbf{X}_i \mathbf{X}_i^t \mathbf{a}_{j'} \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^t) \pi_{ij'}(\hat{\boldsymbol{\beta}}_{j'n} : \mathbf{X}_i^t)\right]\right] \\
&= \frac{1}{n}\sum_{j=1}^m \sum_{j'=1}^m \mathbf{a}_j^t H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'} \mathbf{a}_{j'} \\
&= \frac{1}{n}(\mathbf{a}_1^t, \mathbf{a}_2^t, \cdots, \mathbf{a}_m^t) H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})(\mathbf{a}_1^t, \mathbf{a}_2^t, \cdots, \mathbf{a}_m^t)^t
\end{aligned}$$

$$= \frac{1}{n}\mathbf{a}^t H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})\mathbf{a},$$

since $\mathbf{X}_i^*$, $i = 1, 2, \cdots, n$'s are independent and identically distributed with empirical c.d.f. $\hat{G}_n$ conditionally on the original sample $\mathbf{X}_i$. Furthermore, $E(|U_{in}^*|^3)$ is bounded from above for any $\boldsymbol{\beta}_n(\boldsymbol{\beta}_n \in N_\epsilon(\boldsymbol{\beta}))$ as follows :

$$E(|U_{in}^*|^3) \leq \frac{m\sqrt{m}}{n\sqrt{n}} M^3 \|\mathbf{a}\|^3.$$

And as $n \to \infty$, since $\mathbf{a}^t H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})\mathbf{a} \overset{w.p.1}{\to} \mathbf{a}^t I(\boldsymbol{\beta})\mathbf{a}$, we obtain the following relationship :

$$\sum_{i=1}^{n} E(|U_{in}^*|^3)/S_n^3 \leq \frac{m\sqrt{m}M^3\|\mathbf{a}\|^3}{\sqrt{n}} \frac{1}{[\mathbf{a}^t H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})\mathbf{a}]^{\frac{3}{2}}} \longrightarrow \mathbf{0},$$

where $S_n^2 = \sum_{i=1}^n Var(U_{in}^*) = \mathbf{a}^t H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})\mathbf{a}$. Consequently, the Lyapounov's condition for the triangular array is checked, which completes the proof of Lemma 2. □

Lemma 3. As $n \to \infty$, we obtain

$$\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n \xrightarrow{p} \mathbf{0}_{m(p+1)\times 1}. \tag{3.4}$$

Proof. Note that $\dot{l}^*(\hat{\boldsymbol{\beta}}_n^*)$ can be expanded around $\hat{\boldsymbol{\beta}}_n$ as follows:

$$\dot{l}^*(\hat{\boldsymbol{\beta}}_n^*) = \dot{l}^*(\hat{\boldsymbol{\beta}}_n) + (\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)^t \ddot{l}^*(\tilde{\boldsymbol{\beta}}_n^*), \tag{3.5}$$

where $\tilde{\boldsymbol{\beta}}_n^*$ lies on the line segment joining $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_n^*$. Then the equation (3.5) can be rearranged as

$$(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)^t [-\frac{1}{n}\ddot{l}^*(\tilde{\boldsymbol{\beta}}_n^*)] = \frac{1}{n}\dot{l}^*(\hat{\boldsymbol{\beta}}_n),$$

since $\dot{l}^*(\hat{\boldsymbol{\beta}}_n^*) = 0$ by the previous assumption. Note that
(i) $\frac{1}{n}\dot{l}^*(\hat{\boldsymbol{\beta}}_n) \to \mathbf{0}$ as $n \to \infty$ by Lemma 2 ;
(ii) $-\frac{1}{n}\ddot{l}^*(\tilde{\boldsymbol{\beta}}_n^*)$ tends to a positive definite matrix with probability 1, since $E(\mathbf{X}_i\mathbf{X}_i^t)$ exists and $\ddot{b}(\theta_1, \theta_2, \cdots, \theta_m) = \frac{\partial^2 b(\theta_1, \theta_2, \cdots, \theta_m)}{\partial\theta_j \partial\theta_{j'}}$ for $j, j' = 1, 2, \cdots, m$ is bounded from above such that $b(\theta_1, \theta_2, \cdots, \theta_m) = \log(1 + \sum_{j=1}^m e^{\theta_j})$, and

$\theta_j = \mathbf{X}_i^{*t}\beta_j;\ j = 1, 2, \cdots, m,\ \|\mathbf{X}_i^*\| \le M,\ \beta \in B$(Lehmann;1983).
From (i) and (ii), the result follows. □

Lemma 4. As $n \to \infty$, we obtain

$$-\frac{1}{n}\ddot{l}^*(\tilde{\boldsymbol{\beta}}_n^*) \xrightarrow{p} I(\boldsymbol{\beta})_{m(p+1)\times m(p+1)}. \tag{3.6}$$

Proof. Note that $-\frac{1}{n}\ddot{l}^*(\boldsymbol{\beta}) = H_1(\boldsymbol{\beta} : \mathbf{X}^*)$. Then the (l,k)-th element of difference matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*) - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})]_{jj}$ is represented as $\frac{1}{n}\sum_{i=1}^n W_{in}^*$ $(j = 1, 2, \cdots, m;\ l, k = 0, 1, 2, \cdots, p)$, where $W_{in}^* = W_{in}^*(l,k) = \mathbf{X}_{il}^*\mathbf{X}_{ik}^*\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)(1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)) - E[\mathbf{X}_{il}^*\mathbf{X}_{ik}^*\ \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)(1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*))]$. And we define $W_{in} = W_{in}(l,k) = \mathbf{X}_{il}\mathbf{X}_{ik}\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i)(1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i))$, $\overline{W}_n = \frac{1}{n}\sum_{i=1}^n W_{in}$. For some constant K, the following inequality holds with probability one :

$$\begin{aligned}
Var(W_{in}^*) &= Var(W_{in}^*|\mathbf{X}, Y) = E(W_{in}^{*2}) - [E(W_{in}^*)]^2 \\
&= E[\mathbf{X}_{il}^*\mathbf{X}_{ik}^*\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)(1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*))]^2 - [\overline{W}_n]^2 \\
&= \frac{1}{n}\sum_{i=1}^n [\mathbf{X}_{il}\mathbf{X}_{ik}\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i)(1 - \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i))]^2 - [\overline{W}_n]^2 \\
&= \frac{1}{n}\sum_{i=1}^n (W_{in} - \overline{W}_n)^2 \le K.
\end{aligned}$$

Therefore, as $n \to \infty$, $Var(\frac{1}{n}\sum_{i=1}^n W_{in}^*)$ converges to 0. Since all the (l,k)-th elements of the matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj}]$ converge to zero 0 by the weak law of large numbers(WLLN), the matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj} - H(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj}]$ converges to $\mathbf{0}_{(p+1)\times(p+1)}$ matrix $(j = 1, 2, \cdots, m)$. Now, we write

$$\begin{aligned}
&H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - I(\boldsymbol{\beta})_{jj} \\
&= [H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj}] \\
&\quad + [H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj}] + [H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj} - I(\boldsymbol{\beta})_{jj}].
\end{aligned}$$

The first term tends to a matrix of zeros by applying Taylor's expansion (see the proof of Theorem 1), Lemma 3, and $\tilde{\boldsymbol{\beta}}_n^*$ which satisfies $\|\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\| \le \|\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\|$. And the third converges to the zero matrix $\mathbf{0}_{(p+1)\times(p+1)}$ by Theorem 1. So $H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj}$ converges to $I(\boldsymbol{\beta})_{jj}$ $(j = 1, 2, \cdots, m)$. The (l,k)-th element of the matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'}]$ is given by $\frac{1}{n}\sum_{i=1}^n M_{in}^*$ $(j \neq j',\ j, j' = 1, 2, \cdots, m;\ l, k = 0, 1, 2, \cdots, p)$, where

$$\begin{aligned} &M_{in}^* \\ &= M_{in}^*(l,k) \\ &= -\mathbf{X}_{il}^*\mathbf{X}_{ik}^*\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)\pi_{ij'}(\hat{\boldsymbol{\beta}}_{j'n} : \mathbf{X}_i^*) + E[\mathbf{X}_{il}^*\mathbf{X}_{ik}^*\pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)\pi_{ij'}(\hat{\boldsymbol{\beta}}_{j'n} : \mathbf{X}_i^*)]. \end{aligned}$$

Similarly, we can easily prove that $Var(M_{in}^*)$ is bounded from above and $Var(\frac{1}{n}\sum_{i=1}^n M_{in}^*)$ converges to 0 as $n \to \infty$. Since the (l,k)-th element of difference matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'}]$ tends to 0 by WLLN, the matrix $[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'}]$ converges to a matrix of zeros $\mathbf{0}_{(p+1)\times(p+1)}$. Now, we decompose $[H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj'} - I(\boldsymbol{\beta})_{jj'}]$ into three terms as follows:

$$\begin{aligned} &[H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj'} - I(\boldsymbol{\beta})_{jj'}] \\ &= [H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj'}] \\ &\quad + [H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj'}] + [H_1(\hat{\boldsymbol{\beta}} : \mathbf{X})_{jj'} - I(\boldsymbol{\beta})_{jj'}]. \end{aligned}$$

The first term tends to zero matrix $\mathbf{0}_{(p+1)\times(p+1)}$ by Taylor's expansion (see the proof of Theorem 1), Lemma 3, and $\tilde{\boldsymbol{\beta}}_n^*$ which satisfies $\|\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\| \leq \hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\|$. $H_1(\tilde{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj'}$ converges to $I(\boldsymbol{\beta})_{jj'}$ in conditional probability $(j \neq j', j, j' = 1, 2, \cdots, m)$ since the third tends to zero matrix $\mathbf{0}_{(p+1)\times(p+1)}$ by Theorem 1. Therefore, $-\frac{1}{n}\ddot{l}^*(\tilde{\boldsymbol{\beta}}_n^*)$ converges to information matrix $I(\boldsymbol{\beta})$ in conditional probability. □

Theorem 2. In addition to the assumptions of Lemma 1, suppose that there exists $E(\mathbf{X}_i\mathbf{X}_i^t)$ which is positive definite. Along almost all sample sequences, given the original sample, as $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I^{-1}(\boldsymbol{\beta})) \tag{3.7}$$

$$H_i(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*) \xrightarrow{p} I(\boldsymbol{\beta})_{m(p+1)\times m(p+1)} \tag{3.8}$$

$$\sqrt{n}H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I_{m(p+1)\times m(p+1)}), \quad (i = 1, 2) \tag{3.9}$$

where the starred items imply the bootstrap versions of the original ones, and for $i = 1, 2$ $H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)[H_i^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)]^t = H_i(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)$.

Proof. Consider the fact (3.6) to prove the result (3.7) again. That is,

$$\dot{l^*}(\hat{\boldsymbol{\beta}}_n^*) = \dot{l^*}(\hat{\boldsymbol{\beta}}_n) + (\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)^t \ddot{l^*}(\tilde{\boldsymbol{\beta}}_n^*),$$

where $\|\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\| \leq \|\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\|$. The following equation is now easy to obtain

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)^t[-\frac{1}{n}\ddot{l^*}(\tilde{\boldsymbol{\beta}}_n^*)] = \frac{1}{\sqrt{n}}\dot{l^*}(\hat{\boldsymbol{\beta}}_n)$$

by using $\dot{l^*}(\hat{\boldsymbol{\beta}}_n^*) = \mathbf{0}$. Consequently, we obtain that the result (3.7) follows from Lemma 2, Lemma 4, and Slustky's Theorem.

To prove (weak) consistency of $H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)$, define $H_1(\boldsymbol{\beta} : \mathbf{X})_{jj} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^t\ P_{ij}(1-P_{ij})$ where $P_{ij} = \pi_{ij}(\boldsymbol{\beta}_j : \mathbf{X}_i) = \frac{\exp(\mathbf{X}_i^t\boldsymbol{\beta}_j)}{1+\sum_{l=1}^m \exp(\mathbf{X}_i^t\boldsymbol{\beta}_l)}$. Then, we can decompose it as follows :

$$\begin{aligned}
&H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - I(\boldsymbol{\beta})_{jj} \\
&= [H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj}] \\
&\quad +[H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj}] + [H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X})_{jj} - I(\boldsymbol{\beta})_{jj}].
\end{aligned}$$

By Taylor's expansion, the (l,k)-th element of the first matrix is as follows :

$$\begin{aligned}
[H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj}]_{l,k} &= \frac{1}{n}\sum_{i=1}^n \mathbf{X}_{il}^*\mathbf{X}_{ik}^*[\hat{\pi}_{ij}(1-\hat{\pi}_{ij}) - \pi_{ij}(1-\pi_{ij})] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbf{X}_{il}^*\mathbf{X}_{ik}^*[\sum_{l'\neq j}^m (\hat{\boldsymbol{\beta}}_{l'n}^* - \hat{\boldsymbol{\beta}}_{l'n})^t \\
&\qquad \mathbf{X}_i^*(-\tilde{\pi}_{ij}\tilde{\pi}_{il'} + 2\tilde{\pi}_{ij}^2\tilde{\pi}_{il'}) \\
&\qquad +(\hat{\boldsymbol{\beta}}_{jn}^* - \hat{\boldsymbol{\beta}}_{jn})^t\mathbf{X}_i^*(\tilde{\pi}_{ij} - 3\tilde{\pi}_{ij}^2 + 2\tilde{\pi}_{ij}^3)],
\end{aligned}$$

where $\hat{\pi}_{ij} = \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn}^* : \mathbf{X}_i^*)$, $\pi_{ij} = \pi_{ij}(\hat{\boldsymbol{\beta}}_{jn} : \mathbf{X}_i^*)$, $\tilde{\pi}_{ij} = \pi_{ij}(\tilde{\boldsymbol{\beta}}_{jn}^* : \mathbf{X}_i^*)$, and $\tilde{\boldsymbol{\beta}}_{jn}^*$ which satisfies $\|\tilde{\boldsymbol{\beta}}_{jn}^* - \hat{\boldsymbol{\beta}}_{jn}\| \leq \|\hat{\boldsymbol{\beta}}_{jn}^* - \hat{\boldsymbol{\beta}}_{jn}\|$. The difference vector $\hat{\boldsymbol{\beta}}_{jn}^* - \hat{\boldsymbol{\beta}}_{jn}$ converges to zero vector, $\mathbf{0}_{(p+1)\times 1}$, $(j = 1, 2, \cdots, m)$ since the vector $\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n$ tends to $\mathbf{0}_{m(p+1)\times 1}$. Hence all the column vectors except the row vectors $(\hat{\boldsymbol{\beta}}_{jn}^* - \hat{\boldsymbol{\beta}}_{jn})^t$ $(j = 1, 2, \cdots, m)$ are bounded from above. That is, it follows that all the (l,k)-th elements of matrix $[H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj}]$ tend to 0. Consequently, we conclude that the matrix $[H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol{\beta}}_n : \mathbf{X}^*)_{jj}]$ converges to a matrix of zeros $\mathbf{0}_{(p+1)\times(p+1)}$.

To prove the second matrix $[H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj}]$ converges to a matrix of zeros $\mathbf{0}_{(p+1)\times(p+1)}$ is the same procedure as the proof of Lemma 4. The third $[H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj} - I(\boldsymbol\beta)_{jj}]$ converges to zero matrix $\mathbf{0}_{(p+1)\times(p+1)}$ by Theorem 1. Therefore, the bootstrap estimation matrix $H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj}$ converges to the matrix $I(\boldsymbol\beta)_{jj}$ in conditional probability ($j = 1,2,\ldots,m$). Now, we define $H_1(\boldsymbol\beta : \mathbf{X})_{jj'} = -\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^t P_{ij}P_{ij'}$ ($j \neq j'; j,j' = 1,2,\cdots,m$) where $P_{ij} = \pi_{ij}(\boldsymbol\beta_j : \mathbf{X}_i) = \frac{\exp(\mathbf{X}_i^t\boldsymbol\beta_j)}{1+\sum_{l=1}^m \exp(\mathbf{X}_i^t\boldsymbol\beta_l)}$. Consider the following decomposed matrix :

$$\begin{aligned}
&H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'} - I(\boldsymbol\beta)_{jj'}\\
&= [H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'}]\\
&\quad +[H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj'}] + [H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj'} - I(\boldsymbol\beta)_{jj'}].
\end{aligned}$$

The (l,k)-th element of the first term would be expanded as follows :

$$\begin{aligned}
&[H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'}]_{l,k}\\
=\;& -\frac{1}{n}\sum_{i=1}^n \mathbf{X}^*_{il}\mathbf{X}^*_{ik}[\hat\pi_{ij}\hat\pi_{ij'} - \pi_{ij}\pi_{ij'}]\\
=\;& -\frac{1}{n}\sum_{i=1}^n \mathbf{X}^*_{il}\mathbf{X}^*_{ik}[\sum_{j\neq l'\neq j'}^m (\hat{\boldsymbol\beta}^*_{l'n} - \hat{\boldsymbol\beta}_{l'n})^t\mathbf{X}^*_i(-2\tilde\pi_{ij}\tilde\pi_{ij}\tilde\pi_{il'})\\
&+(\hat{\boldsymbol\beta}^{*}_{jn} - \hat{\boldsymbol\beta}_{jn})^t\mathbf{X}^*_i(\tilde\pi_{ij}\tilde\pi_{ij'} - 2\tilde\pi^2_{ij}\tilde\pi_{ij'}) + (\hat{\boldsymbol\beta}^*_{j'n} - \hat{\boldsymbol\beta}_{j'n})^t\mathbf{X}^*_i(\tilde\pi_{ij}\tilde\pi_{ij'} - 2\tilde\pi_{ij}\tilde\pi^2_{ij'})],
\end{aligned}$$

where $\hat\pi_{ij} = \pi_{ij}(\hat{\boldsymbol\beta}^*_{jn} : \mathbf{X}^*_i)$, $\pi_{ij} = \pi_{ij}(\hat{\boldsymbol\beta}_{jn} : \mathbf{X}^*_i)$, $\tilde\pi_{ij} = \pi_{ij}(\tilde{\boldsymbol\beta}^*_{jn} : \mathbf{X}^*_i)$, $\tilde{\boldsymbol\beta}^*_{jn}$ which satisfies $\|\tilde{\boldsymbol\beta}^*_{jn} - \hat{\boldsymbol\beta}_{ln}\| \leq \|\hat{\boldsymbol\beta}^*_{jn} - \hat{\boldsymbol\beta}_{jn}\|$ ($j = 1,2,\ldots,m$). Fortunately, the first term converges to $\mathbf{0}_{(p+1)\times(p+1)}$ matrix since $\hat{\boldsymbol\beta}^*_n - \hat{\boldsymbol\beta}_n$ converges to a vector of zeros $\mathbf{0}_{m(p+1)\times 1}$. That is, all the column vectors except row vectors $(\hat{\boldsymbol\beta}^*_{jn} - \hat{\boldsymbol\beta}_{jn})^t$ ($j = 1,2,\cdots,m$) are bounded from above. So all the (l,k)-th elements of matrix $[H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'}]$ tend to 0 as $n \to \infty$. Hence, the matrix $[H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'}]$ converges to a matrix of zeros $\mathbf{0}_{(p+1)\times(p+1)}$. To show the second matrix $[H_1(\hat{\boldsymbol\beta}_n : \mathbf{X}^*)_{jj'} - H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj'}]$ converges to a matrix of zeros, $\mathbf{0}_{(p+1)\times(p+1)}$ is the same as Lemma 4. The third difference matrix $[H_1(\hat{\boldsymbol\beta}_n : \mathbf{X})_{jj'} - I(\hat{\boldsymbol\beta})_{jj'}]$ converges to zero matrix $\mathbf{0}_{(p+1)\times(p+1)}$ by Theorem 1. Hence the matrix $H_1(\hat{\boldsymbol\beta}^*_n : \mathbf{X}^*)_{jj'}$ converges to the matrix $I(\boldsymbol\beta)_{jj'}$ ($j \neq j'$; $j,j' = 1,2,\ldots,m$). Consequently, it follows that the bootstrap estimation matrix $H_1(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)$ converges to the information matrix $I(\boldsymbol{\beta})$.

On the other hand, by above two facts and Slutsky's Theorem, we obtain the following result (3.9).

$$\sqrt{n} H_1^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \xrightarrow{d} N_{m(p+1)}(\mathbf{0}, I_{m(p+1)\times m(p+1)})$$

For the case $H_2(\hat{\boldsymbol{\beta}}_n^* : \mathbf{X}^*)$, it suffices to apply all the similar procedures routinely. □

ACKNOWLEDGEMENT

The detailed and constructive comments of referees were very useful in making the paper more readable. Their efforts are greatly appreciated.

REFERENCES

(1) Beran, R.J.(1984). Bootstrap Methods in Statistics, *Jahresbericht der Deutschen Mathematiker-Vereiningung,* **86** , 14-30.

(2) Efron, B.(1979). Bootstrap Methods, *Annals of Statistics* , **7**, 1-26.

(3) Fahrmeir and Kaufmann, H.(1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear models, *Annals of Statistics*, **13**, 342-368.

(4) Freedman, D.A.(1981). Bootstrapping Regression Models, *Annals of Statistics*, **9** , 1218-1228.

(5) Hinkley, D.(1988). The Bootstrap, *Journal of the Royal Statistical Society B*, **151**, 338-354.

(6) Hosmer, D.W.and Lemeshow, S.(1989). *Applied Logistic Regression,* John Wiley and Sons, Inc..

(7) Lee, K.W.(1990). Bootstrapping Logistic Regression Models with Random Regressors, *Communications in Statistics: Theory and Methods* , **19 (7)**, 2527-2539.

(8) Lee, K.W., Kim, C.R.,Sohn, K.T.and Jeong, K.M.(1992). Bootstrapping Generalized Linear Models with Random Regressors, *Journal of the Korean Statistical Society* , **21 (1)** , 70-79.

(9) Lehmann, E.L.(1983). *Theory of Point Estimation,* ; John Wiley and Sons, New York.

(10) McCullagh, P. and Nelder, J.A.(1989). *Generalized Linear Models,* 2nd ed., ; Chapman and Hall, London.