PEBBLE: A second order correct Bootstrap method in logistic regression

DEBRAJ DAS[1,a] and PRIYAM DAS[2,b]

[1] Department of Mathematics, Indian Institute of Technology, Bombay, India, [a] debrajdas@math.iitb.ac.in
[2] Department of Biostatistics, Virginia Commonwealth University, Richmond, USA, [b] dasp4@vcu.edu

Logistic regression is often used in many different fields, such as clinical trials, biomedical surveys, marketing, and banking, to predict the probability of a binary outcome. In this paper we propose a novel Bootstrap technique for approximating the distribution of the maximum likelihood estimator (MLE) of the regression parameter vector. Improved inference performance is obtained over the traditional normal approximation via establishing second order correctness. The main challenge in establishing second order correctness remains in the fact that the response variable being binary, the MLE may have a lattice structure resulting in non-existence of formal Edgeworth expansion. In order to achieve the second order correctness, a smoothing technique developed in Lahiri (1993) is adopted to define the original and Bootstrapped studentized pivots. Second order results are also extended to the one-dimensional smooth functions of the regression parameter e.g., odds ratio and success probability. Simulation experiments are performed to evaluate the finite-sample properties of the proposed Bootstrap method. The proposed methodology is demonstrated with an application on real dataset in healthcare operations.

AMS Subject Classification: Primary 62J12; Secondary 62E20.

Keywords: Lattice; logistic regression; perturbation Bootstrap; smoothing; SOC

1. Introduction

Logistic regression is one of the most widely used regression techniques for predicting binary outcomes. The use of the 'logit' function as a statistical tool dates back to Berkson (1944), followed by Cox (1958), who popularized it in the field of regression. Following those seminal works, numerous applications of logistic regression can be found in different fields, from banking sectors to epidemiology, clinical trials, biomedical surveys, among others (Hosmer et al., 2013). The logistic regression model is defined as follows. Suppose that y denotes the binary response variable and the value of y depends on $\boldsymbol{x} = (x_1, \ldots, x_p)'$, a p-dimensional covariate. In logistic regression, typically the logarithm of the odds ratio corresponding to success (i.e. the event $\{y = 1\}$) is modeled as a linear function of the covariates. The odds ratio for the event $\{y = 1\}$ is given by $[P(y = 1)][1 - P(y = 1)]^{-1}$. Then the underlying model of the logistic regression is given by

$$\operatorname{logit}(P(y=1)) = \log\left[\frac{P(y=1)}{1-P(y=1)}\right] = \boldsymbol{x}'\boldsymbol{\beta}, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the p-dimensional vector of regression parameters. Throughout this paper we assume that all the covariates are non-random. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is commonly used for inference purposes. For a given sample $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, the likelihood is given by

$$L(\boldsymbol{\beta}|y_1, \ldots, y_n, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{i=1}^{n} (p(\boldsymbol{\beta}|\boldsymbol{x}_i))^{y_i} (1 - p(\boldsymbol{\beta}|\boldsymbol{x}_i))^{1-y_i},$$

where $p(\boldsymbol{\beta}|\boldsymbol{x}_i) = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}})^{-1}$. The MLE $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is defined as the maximizer of the likelihood, which is obtained by solving

$$\sum_{i=1}^{n}(y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))\boldsymbol{x}_i = 0. \tag{1.2}$$

To construct the confidence intervals for some function of the regression parameter $\boldsymbol{\beta}$ or to perform test on it, it is essential to find a good approximation of the distribution of $\hat{\boldsymbol{\beta}}_n$. $\hat{\boldsymbol{\beta}}_n$ being the MLE, the distribution of $\hat{\boldsymbol{\beta}}_n$ is approximately Gaussian under certain regularity conditions. Asymptotic normality as well as other large sample properties of $\hat{\boldsymbol{\beta}}_n$ have been studied extensively in the literature (Amemiya, 1976; Fahrmeir and Kaufmann, 1985; Gourieroux and Monfort, 1981; Haberman, 1974; McFadden, 1974). For the sake of completeness, we establish the Berry-Esseen bound corresponding to the normal approximation of $\hat{\boldsymbol{\beta}}_n$ and compare it with the rate of approximation by the Bootstrap method developed in this paper.

In the last few decades, several variants of Bootstrap have been developed in linear regression. As an alternative to asymptotic normality, the approach proposed in Efron (1979) has been shown to work in a wide class of models including multiple linear regression. Depending on whether the covariates are non-random or random in linear regression setup, Freedman (1981) proposed the residual Bootstrap or the paired Bootstrap. A few other variants of Bootstrap methods in linear regression setup are as follows: the wild Bootstrap (Liu, 1988; Mammen, 1993), the weighted Bootstrap (Barbe and Bertail, 1995; Lahiri, 1992) and the perturbation Bootstrap (Das and Lahiri, 2019). Using similar mechanism of the residual and the paired Bootstrap, Moulton and Zeger (1989) and Moulton and Zeger (1991) developed the standardized Pearson residual resampling and the observation vector resampling Bootstrap methods in generalized linear models (GLM). Lee (1990) considered the logistic regression model with random covariates and showed that the conditional distribution of these resample-based Bootstrap estimators for the given data are close to the distribution of the original estimator in almost sure sense. Diciccio and Efron (1992) and Diciccio and Efron (1996) developed different Bootstrap confidence intervals for parameters in exponential family models. They also established the second order accuracy of the proposed Bootstrap confidence intervals via assuming certain smoothness conditions implicitly on the underlying distributions. However the responses in the logistic regression setup being binary, such smoothness conditions may not hold. One such scenario is explored in Theorem 3.2. As an alternative to Bootstrap, Sun et al. (2000) developed simultaneous confidence regions for the mean function in GLM using inverse Edgeworth expansions. Claeskens et al. (2003) proposed a couple of Bootstrap methods for logistic regression in univariate case, namely 'linear one-step Bootstrap' and 'quadratic one-step Bootstrap'. 'Linear one-step Bootstrap' was developed following the linearization principle proposed in Davison et al. (1986), whereas, 'Quadratic one-step Bootstrap' was constructed based on the quadratic approximation of the estimators as discussed in Ghosh (1994). The validity of these two Bootstrap methods for approximating the underlying distribution in almost sure sense was established in Claeskens et al. (2003). Further a finite sample bias correction of logistic regression estimator was also proposed in Claeskens et al. (2003) which they did using quadratic one-step Bootstrap method.

In order to have an explicit understanding about the sample size requirement for practical implementations of any asymptotically valid method, it is essential to study the error rate of the approximation. For example in the simplest case of the mean of iid random variables, Berry-Esseen theorem tells us that the error rate for the normal approximation is of order $O(n^{-1/2})$ (cf. Theorem 12.4 of Bhattacharya and Rao (1986)), i.e. the normal approximation is first order correct. Whereas the error rate for the corresponding Bootstrap approximation is of order $o(n^{-1/2})$ in almost sure or probability sense (cf. Singh (1981)), i.e. the Bootstrap approximation is second order correct. Moreover, the error rate of normal approximation cannot be improved to second order in general, except in specific situations e.g., when the underlying random variables have symmetric absolutely continuous distributions. Second order correctness (SOC) implies that the difference between the cdf of the sample mean and the corresponding Bootstrap cdf is close to 0 for large enough n even after multiplying with $n^{1/2}$. The first order correctness does not imply the same in case of the normal approximation; the difference between the cdf of the sample mean and the normal cdf may not necessarily be small after multiplication by $n^{1/2}$ even when n is substantially large. Therefore SOC has the potential of drawing more accurate inference results based on Bootstrap compared to its normal approximation-based alternatives. In other words, to achieve a desired level of accuracy, much smaller sample size is required in case of second-order correct Bootstrap approximation than that based on normal approximation. The results on SOC in linear regression have been studied extensively for different Bootstrap methods, specially for the residual, the weighted and the perturbation Bootstrap methods (Barbe and Bertail, 1995; Das and Lahiri, 2019; Lahiri, 1992). However, to the best of our knowledge, SOC has not been explored in the context of Bootstrap methods for logistic regression. In this paper, we propose Perturbation Bootstrap in Logistic Regression (PEBBLE) as a second order correct approximation of the distribution of $\hat{\boldsymbol{\beta}}_n$. Whenever the underlying estimator is a minimizer of certain objective function, perturbation Bootstrap simply produces a Bootstrap version of the estimator by finding the minimizer of a random objective function, suitably developed by perturbing the original objective function using some non-negative random variables. For the sake of comparison with PEBBLE, we also establish the error rate for the normal approximation of the studentized version of the distribution of $\hat{\boldsymbol{\beta}}_n$ which comes out to be of $O(n^{-1/2}\log n)$. The extra "$\log n$" term in the error rate appears due to the underlying lattice structure. Therefore, the inference based on PEBBLE can be expected to be more accurate than that based on the asymptotic normality.

In order to establish SOC for the proposed method, we start with studentization of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and its perturbation Bootstrap version. We show that unlike in the case of multiple linear regression, here SOC may not be achieved, in general, only by studentization of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ due to the possible lattice nature of the distribution of the logistic regression estimator $\hat{\boldsymbol{\beta}}_n$. The lattice nature of the distribution is induced by the binary nature of the response variables. The lattice structure can be avoided if the covariates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ do not cluster around only a few points. For $p = 1$, this is same as having the covariates not clustering closely around a collection of equally spaced numbers (see for example assumption 2 in Kong and Levin (1996)). Such an assumption is clearly restrictive and is hard to check in practical applications. However without such smoothness assumption, extra correction terms generally appear in the Edgeworth expansion and hence usual Edgeworth expansion does not remain valid. For example, one can compare Theorem 20.8 and Corollary 23.2 in Bhattacharya and Rao (1986) to get an idea about the correction terms. Generally these correction terms cannot be approximated with an error of order $o(n^{-1/2})$, which makes SOC unachievable even with studentization. A detailed discussion in this direction can be found in Remark 3.4 based on the existing Bootstrap methods in logistic regression. In this paper, we relax such restrictive assumptions. Instead, we adopt a smoothing technique developed in Lahiri (1993) and obtain usual Edgeworth expansion without any correction factor only under some moment conditions. The same smoothing technique is used for the Bootstrap version as well and SOC is established for PEBBLE by comparing the Edgeworth expansions across the original and the Bootstrap cases. An interesting property of the smoothing is that it has negligible effect on the asymptotic variance of $\hat{\boldsymbol{\beta}}_n$ and therefore it is not required to incorporate the effect of the smoothing in the form of the studentization. In order to prove the results, we establish the Edgeworth expansion of a smoothed version of a sequence of sample means of independent (but not necessarily identically distributed) random vectors in Lemma 2.3 of the supplementary material file Das and Das (2024). This lemma may be of independent interest for establishing SOC of Bootstrap in other related problems.

The rest of the paper is organized as follows. The perturbation Bootstrap version of the logistic regression estimator is described in Section 2. Main results including theoretical properties of the

Bootstrap along with normal approximation are discussed in Section 3. SOC of PEBBLE for one-dimensional smooth functions of $\boldsymbol{\beta}$ is presented in Section 4. In Section 5, finite-sample performance of PEBBLE is evaluated comparing with other related existing methods by simulation experiments. Section 6 gives an illustration of PEBBLE in healthcare operations decision data set. Section 7 contains the proofs of Theorem 3.1 and Theorem 3.3. Concluding remarks are made in the Section 8. Proofs of Theorem 3.2 and Theorem 4.1, statements of the auxiliary lemmas, and additional simulation results are relegated to the supplementary material file Das and Das (2024).

2. Description of PEBBLE

In this section, we define PEBBLE estimator for logistic regression. Let $G_1^*, \ldots, G_n^*$ be n independent copies of a non-negative and non-degenerate random variable G^* with mean μ_{G^*}, $Var(G^*) = \mu_{G^*}^2$ (i.e. the variance is square of the mean) and $\mathbf{E}(G^* - \mu_{G^*})^3 = \mu_{G^*}^3$ (i.e. the third central moment is the cube of the mean). These quantities serve as perturbing random quantities in the construction of PEBBLE. It is the minimizer of a carefully constructed objective function which involves the observed values $y_1, \ldots, y_n$ as well as the estimated probability of successes $\hat{p}(\boldsymbol{x}_i) = e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n}(1 + e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n})^{-1}$, $i = 1, \ldots, n$. Formally, the PEBBLE estimator $\hat{\boldsymbol{\beta}}_n^*$ is defined as

$$\hat{\boldsymbol{\beta}}_n^* = \arg\max_{\boldsymbol{t}} \left[\sum_{i=1}^n \left\{ (y_i - \hat{p}(\boldsymbol{x}_i))\boldsymbol{x}_i'\boldsymbol{t} \right\} (G_i^* - \mu_{G^*}) + \mu_{G^*} \sum_{i=1}^n \left\{ \hat{p}(\boldsymbol{x}_i)(\boldsymbol{x}_i'\boldsymbol{t}) - \log(1 + e^{\boldsymbol{x}_i'\boldsymbol{t}}) \right\} \right].$$

In other words, $\hat{\boldsymbol{\beta}}_n^*$ is the solution of the equation

$$\sum_{i=1}^n (y_i - \hat{p}(\boldsymbol{x}_i))\boldsymbol{x}_i(G_i^* - \mu_{G^*})\mu_G^{*-1} + \sum_{i=1}^n (\hat{p}(\boldsymbol{x}_i) - p(\boldsymbol{t}|\boldsymbol{x}_i))\boldsymbol{x}_i = 0, \tag{2.1}$$

since the derivative of the LHS of (2.1) with respect to $\boldsymbol{t}$ is negative definite. Now let us briefly describe the motivation behind defining the PEBBLE estimator $\hat{\boldsymbol{\beta}}_n^*$ as a solution of (2.1). The first term in the LHS of (2.1) is the ideal perturbation Bootstrap version of the LHS of the defining equation (1.2), and hence it is the main contributing factor to the Bootstrap randomization. Now, if we examine the second term on the LHS of (2.1) closely, then the leading term in the Taylor's expansion of $p(\hat{\boldsymbol{\beta}}_n^*|\boldsymbol{x}_i)$ around $p(\hat{\boldsymbol{\beta}}_n|\boldsymbol{x}_i)$ for each $\boldsymbol{x}_i$ gives rise to the matrix $n^{-1}\sum_{i=1}^n p(\hat{\boldsymbol{\beta}}_n|\boldsymbol{x}_i)(1 - p(\hat{\boldsymbol{\beta}}_n|\boldsymbol{x}_i))\boldsymbol{x}_i\boldsymbol{x}_i'$. This matrix is clearly an estimator of the variance of the original pivotal quantity $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$, and the reason behind incorporating the second term in the LHS of (2.1) is essentially to introduce such a matrix in the Bootstrapped setup. For details, we refer to the Taylor's expansion considered in the proof of Theorem 3.3, presented in Section 7. One immediate choice for the distribution of G^* is Beta$(1/2, 3/2)$ since the required conditions of G^* are satisfied for this distribution. Other choices can be found in Das et al. (2019). The moment characteristics of G^* are assumed to be true for the rest of this paper. Any additional assumption on G^* will be stated in respective theorems.

3. Main results

In this section, we describe the theoretical results of Bootstrap as well as the normal approximation. In the Subsection 3.1, we state a Berry-Esseen type theorem for a studentized version of the logistic regression estimator $\hat{\boldsymbol{\beta}}_n$. In the Subsection 3.2, we explore the effectiveness of Bootstrap in approximating the distribution of the studentized version. Theorem 3.2 shows that SOC is not achievable solely by studentization even when $p = 1$. As a remedy, we introduce a smoothing in the studentization pivots and show that the PEBBLE achieves SOC.

Before moving to our first theorem on normal approximation, let us define the class of sets that we will consider in the following theorems. Let $\partial\boldsymbol{B}$ denote the boundary of $\boldsymbol{B}$. For any set $\boldsymbol{A} \subseteq \mathcal{R}^m$ and $\eta > 0$, $\boldsymbol{A}^\eta = \{\boldsymbol{x} \in \mathcal{R}^m : d(\boldsymbol{x}, \boldsymbol{A}) < \eta\}$ where $d(\boldsymbol{x}, \boldsymbol{A}) = \inf\{\|\boldsymbol{x} - \boldsymbol{y}\| : \boldsymbol{y} \in \boldsymbol{A}\}$ and $\|\cdot\|$ is the Euclidean norm. For any natural number m, the class of sets $\mathcal{A}_m$ is the collection of Borel subsets of $\mathcal{R}^m$ satisfying

$$\sup_{B \in \mathcal{A}_m} \Phi((\partial B)^\epsilon) = O(\epsilon) \quad \text{as } \epsilon \downarrow 0.$$

Here Φ denotes the normal distribution with mean $\mathbf{0}$ and dispersion matrix being the identity matrix. We use the class $\mathcal{A}_m$ for the uniform asymptotic results on normal and Bootstrap approximations. A detailed explanation of the class $\mathcal{A}_m$ is presented below in Remark 3.1. $\mathbf{P}_*$ denotes the conditional Bootstrap probability of G^* given data $\{y_1, \ldots, y_n\}$.

3.1. Rate of normal approximation

In this subsection we explore the rate of normal approximation of suitable studentized version of the logistic regression estimator $\hat{\boldsymbol{\beta}}_n$, uniformly over the class of sets $\mathcal{A}_p$. From the definition (1.2) of $\hat{\boldsymbol{\beta}}_n$, we have that $\sum_{i=1}^n (y_i - \hat{p}(\boldsymbol{x}_i))\boldsymbol{x}_i = 0$. Now using Taylor's expansion of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$, it is easy to see that the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is $\boldsymbol{L}_n^{-1}$ where $\boldsymbol{L}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' e^{\boldsymbol{x}_i'\boldsymbol{\beta}}(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}})^{-2}$. An estimator of $\boldsymbol{L}_n$ can be obtained by replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_n$ in the form of $\boldsymbol{L}_n$. Hence we can define the studentized version of $\hat{\boldsymbol{\beta}}_n$ as

$$\tilde{\mathbf{H}}_n = \sqrt{n}\hat{\boldsymbol{L}}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}),$$

where $\hat{\boldsymbol{L}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n}\left(1 + e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n}\right)^{-2}$. Other studentized versions can be constructed by considering other estimators of $\boldsymbol{L}_n$. For details of the construction of different studentized versions, one can look into Lahiri (1994). The result on normal approximation will hold for other studentized versions also as long as it involves the estimator of $\boldsymbol{L}_n$ which is $\sqrt{n}$–consistent.

Classical Berry-Esseen theorem states that the error in normal approximation for the distribution of the mean of a sequence of independent random variables is $O(n^{-1/2})$, provided the average third absolute moment is bounded (cf. Theorem 12.4 in Bhattacharya and Rao (1986)). In the same spirit, we establish a Berry-Esseen theorem for $\tilde{\mathbf{H}}_n$ in Theorem 3.1 where one can point out that there is an extra multiplicative "$\log n$" factor besides the usual $n^{-1/2}$ term. This extra factor is due to the error incurred in Taylor's approximation of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. Since the underlying setup in logistic regression has lattice nature, this error cannot in general be corrected by higher order approximations, like Edgeworth expansions. Further, one important tool in deriving the error rate in normal approximation, and later for deriving the higher order result for the Bootstrap is to find the rate of convergence of $\hat{\boldsymbol{\beta}}_n$ to $\boldsymbol{\beta}$. To this end, we state our first theorem as follows.

Theorem 3.1. *Suppose that $n^{-1}\sum_{i=1}^n \|\boldsymbol{x}_i\|^3 = O(1)$ and $\boldsymbol{L}_n \to \boldsymbol{L}$ as $n \to \infty$ where $\boldsymbol{L}$ is a positive definite matrix.*

(a) *Then there exists a positive constant C_0 such that*

$$\mathbf{P}\Big(\hat{\boldsymbol{\beta}}_n \text{ solves } (1.2) \text{ and } \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| \leq C_0 n^{-1/2}(\log n)^{1/2}\Big) = 1 - o\big(n^{-1/2}\big).$$

(b) *Then we have*

$$\sup_{\boldsymbol{B}\in\mathcal{A}_p}\left|\mathbf{P}\big(\tilde{\mathbf{H}}_n\in\boldsymbol{B}\big)-\Phi(\boldsymbol{B})\right|=O\big(n^{-1/2}\log n\big).$$

The proof of Theorem 3.1 is presented in Section 7. Theorem 3.1 shows that the normal approximation of the distribution of $\tilde{\mathbf{H}}_n$, the studentized logistic regression estimator, has near optimal Berry-Esseen rate. However, an extra '$\log n$' term appears due to the possible lattice structure of the distribution of $\hat{\boldsymbol{\beta}}_n$. The rate can be significantly improved by Bootstrap, as described in the next subsection. The convergence of $\boldsymbol{L}_n$, the variance-covariance matrix of $\sqrt{n}\hat{\boldsymbol{\beta}}_n$, to a positive definite matrix $\boldsymbol{L}$ is required so that $\boldsymbol{L}_n^{-1}$ can be considered for the asymptotic analysis of $\hat{\boldsymbol{\beta}}_n$. Moreover, the condition $n^{-1}\sum_{i=1}^n\|\boldsymbol{x}_i\|^3=O(1)$ is required to utilize Lemma 2.5 (mentioned in the supplementary material file Das and Das (2024)) for obtaining Berry-Esseen inequality in case of normal approximation of $\hat{\boldsymbol{\beta}}_n$, presented in Theorem 3.1. Similar or stronger regularity conditions are generally used in the literature, for example see Amemiya (1976), Fahrmeir and Kaufmann (1985).

3.2. Rate of Bootstrap approximation

In this subsection, we extensively study the rate of Bootstrap approximation for the distribution of the logistic regression estimator. To that end, before exploring the rate of convergence of Bootstrap, it is essential to define the suitable studentized versions in both the original and the Bootstrap settings. Similar to the original case, the asymptotic variance of the PEBBLE estimator $\hat{\boldsymbol{\beta}}_n^*$ is needed to be found to define the studentized version in the Bootstrap setting. Using Taylor's expansion, from (2.1) one can notice that the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^*-\hat{\boldsymbol{\beta}}_n)$ is $\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1}$ where $\hat{\boldsymbol{L}}_n=n^{-1}\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n}(1+e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n})^{-2}$ and $\hat{\boldsymbol{M}}_n=n^{-1}\sum_{i=1}^n\big(y_i-\hat{p}(\boldsymbol{x}_i)\big)^2\boldsymbol{x}_i\boldsymbol{x}_i'$. Therefore the studentized version in Bootstrap setting can be defined as

$$\mathbf{H}_n^*=\sqrt{n}\hat{\boldsymbol{M}}_n^{*-1/2}\boldsymbol{L}_n^*\big(\hat{\boldsymbol{\beta}}_n^*-\hat{\boldsymbol{\beta}}_n\big),$$

where $\boldsymbol{L}_n^*=n^{-1}\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n^*}\big(1+e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n^*}\big)^{-2}$ and $\hat{\boldsymbol{M}}_n^*=n^{-1}\sum_{i=1}^n\big(y_i-\hat{p}(\boldsymbol{x}_i)\big)^2\boldsymbol{x}_i\boldsymbol{x}_i'\mu_{G^*}^{-2}(G_i^*-\mu_{G^*})^2$. Analogously, we define the original studentized version as

$$\mathbf{H}_n=\sqrt{n}\hat{\boldsymbol{M}}_n^{-1/2}\hat{\boldsymbol{L}}_n\big(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}\big),$$

which is used for investigating SOC of Bootstrap for rest of this section. In the next theorem we show that $\mathbf{H}_n^*$ fails to be SOC in approximating the distribution of $\mathbf{H}_n$ even when $p=1$.

Theorem 3.2. *Suppose that $p=1$ and x denotes the only covariate in the model (1.1). Let $x_1,\dots,x_n$ be the observed values of x and β be the true value of the regression parameter. Define $\mu_n=n^{-1}\sum_{i=1}^n x_i p(\beta|x_i)$. Assume the following conditions hold:*

(C.1) *$x_1,\dots,x_n$ are non random and are all integers.*
(C.2) *$x_{i_1},\dots,x_{i_m}=1$ where $\{i_1,\dots,i_m\}\subseteq\{1,\dots,n\}$ with $m\geq(\log n)^2$.*
(C.3) $\max\{|x_i|:i=1,\dots,n\}=O(1)$ *and* $\liminf_{n\to\infty}\big[n^{-1}\sum_{i=1}^n|x_i|^6\big]>0$.
(C.4) *$\sqrt{n}|\mu_n|<M_1$ for $n\geq M_1$ where M_1 is a positive constant.*
(C.5) *The distribution of G^* has an absolutely continuous component with respect to Lebesgue measure and $\mathbf{E}G^{*4}<\infty$.*

Then there exists an interval $\boldsymbol{B}_n$ such that

$$\lim_{n\to\infty} \mathbf{P}\Big(\sqrt{n}\big|\mathbf{P}_*(\mathbf{H}_n^* \in \boldsymbol{B}_n) - \mathbf{P}(\mathbf{H}_n \in \boldsymbol{B}_n)\big| > 0\Big) = 1.$$

The proof of Theorem 3.2 is presented in the supplementary material file Das and Das (2024). Theorem 3.2 shows that unlike in the case of multiple linear regression, in general, the Bootstrap cannot achieve SOC in logistic regression even with studentization. This is due to the discrete nature of the responses $y_1, \ldots, y_n$. Now let us delve into the form of the set $\boldsymbol{B}_n$. $\boldsymbol{B}_n$ is of the form $f_n(\boldsymbol{E}_n \times \mathcal{R})$ with $\boldsymbol{E}_n = (-\infty, z_n]$ and $z_n = \big(3/(4n) - \mu_n\big)$. $f_n(\cdot)$ is a continuous function which is obtained from the Taylor expansion of $\mathbf{H}_n$. Since $\boldsymbol{E}_n \times \mathcal{R}$ is a convex subset of $\mathcal{R}^2$, it is also a connected set. Since $f_n(\cdot)$ is a continuous function, $\boldsymbol{B}_n$ is a connected subset of $\mathcal{R}$ and hence is an interval. Let us briefly describe the utility of the assumptions for establishing the negative result presented in Theorem 3.2. Assumption (C.1) is needed to have the lattice structure of $\hat{\boldsymbol{\beta}}_n$ with (C.2) ensuring that a 'substantial' number of the covariates are clustered. (C.1) and (C.2) along with assumption (C.3) are essential to employ Lemma 2.10 (see supplementary material file Das and Das (2024)) for getting the desired expansion of $n^{-1/2}\sum_{i=1}^{n}(y_i - p(\beta|x_i)x_i$. (C.4) is required to essentially show that the same expansion is valid for the original pivotal quantity $\mathbf{H}_n$. (C.3) along-with (C.6) are required to establish Edgeworth expansion for the Bootstrapped pivot $\mathbf{H}_n^*$ using Lemma 2.9 (see supplementary material file Das and Das (2024)).

Now we define the smoothed versions of $\mathbf{H}_n$ and $\mathbf{H}_n^*$ which are necessary in achieving SOC by PEBBLE for general p. Note that the primary reason behind Bootstrap's failure is the lattice nature of the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. Hence if one can somehow smooth the distribution $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$, or more generally the distribution of $\mathbf{H}_n$, so that the smoothed version has density with respect to Lebesgue measure, then the Bootstrap may be shown to achieve SOC by employing theory of Edgeworth expansions. To that end, following the prescription of Lahiri (1993), consider a $p-$dimensional standard normal random vector Z, independent of $y_1, \ldots, y_n$. Define the smoothed version of $\mathbf{H}_n$ as

$$\check{\mathbf{H}}_n = \mathbf{H}_n + \hat{\boldsymbol{M}}_n^{-1/2} b_n Z, \tag{3.1}$$

where $\{b_n\}_{n\geq 1}$ is a suitable sequence such that it has negligible effect on the variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and hence on the studentization factor. The required conditions on $\{b_n\}_{n\geq 1}$ are stated in Theorem 3.3. To define the smoothed studentized version in Bootstrap setting, consider another $p-$dimensional standard normal vector Z^* which is independent of $y_1, \ldots, y_n, G_1^*, \ldots, G_n^*$ and Z. Define the smoothed version of $\mathbf{H}_n^*$ as

$$\check{\mathbf{H}}_n^* = \mathbf{H}_n^* + \hat{\boldsymbol{M}}_n^{*-1/2} b_n Z^*. \tag{3.2}$$

The following theorem can be distinguished as the main theorem of this section as it shows that the smoothing does the trick for PEBBLE to achieve SOC. Thus the inference on $\boldsymbol{\beta}$ based on the Bootstrap after smoothing is much more accurate than the normal approximation. To state the main theorem, define $W_i = \big(\tilde{y}_i \boldsymbol{x}_i', \big[\tilde{y}_i^2 - \mathbf{E}\tilde{y}_i^2\big]\tilde{z}_i'\big)'$ where $\tilde{y}_i = (y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))$ and $\tilde{z}_i = (x_{i1}^2, x_{i1}x_{i2}, \ldots, x_{i1}x_{ip}, x_{i2}^2, x_{i2}x_{i3}, \ldots, x_{i2}x_{ip}, \ldots, x_{ip}^2)'$ with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$, $i \in \{1, \ldots, n\}$.

Theorem 3.3. *Suppose that $n^{-1}\sum_{i=1}^{n}\|\mathbf{x}_i\|^6 = O(1)$ and the matrix $n^{-1}\sum_{i=1}^{n} Var(W_i)$ converges to some positive definite matrix as $n \to \infty$. Further, the sequence $\{b_n\}_{n\geq 1}$ is chosen such that $b_n = O(n^{-d})$ and $n^{-1/p_1}\log n = o(b_n^2)$, where $d > 0$ is a constant and $p_1 = \max\{p+1, 4\}$.*

(a) *Then there exists a positive constant C such that*

$$\mathbf{P}_*\left(\hat{\boldsymbol{\beta}}_n^* \textit{ solves } (2.1) \textit{ and } \|\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\| \le Cn^{-1/2}(\log n)^{1/2}\right) = 1 - o_p\left(n^{-1/2}\right).$$

(b) *Then the following holds true:*

$$\sup_{\boldsymbol{B}\in\mathcal{A}_p} \left|\mathbf{P}_*\left(\check{\mathbf{H}}_n^* \in \boldsymbol{B}\right) - \mathbf{P}\left(\check{\mathbf{H}}_n \in \boldsymbol{B}\right)\right| = o_p\left(n^{-1/2}\right).$$

The proof of Theorem 3.3 is presented in Section 7. Here we briefly explain why the assumptions are essential for proving Theorem 3.3. The two conditions, such as $n^{-1}\sum_{i=1}^n \|\mathbf{x}_i\|^6 = O(1)$ and the convergence of the matrix $n^{-1}\sum_{i=1}^n Var(W_i)$ to a positive definite matrix are required to employ Theorem 20.6 of Bhattacharya and Rao (1986) in order to establish Lemma 2.4 (see supplementary material file Das and Das (2024)), which is again essential for establishing the above theorem. These types of conditions are quite common for carrying out asymptotic analysis of maximum likelihood type estimators. The conditions on b_n are essential to ensure that the effect of the smoothing is negligible on the variance of the unsmoothed pivotal quantities and hence no modification is required for the studentization considered in defining the smoothed pivots $\check{\mathbf{H}}_n$ and $\check{\mathbf{H}}_n^*$. Theorem 3.3 shows that SOC of PEBBLE can be achieved by a simple smoothing in the studentized pivotal quantities. As a result, confidence regions for $\boldsymbol{\beta}$ can be constructed based on Euclidean norms of the pivotal quantities $\check{\mathbf{H}}_n$ and $\check{\mathbf{H}}_n^*$ which will be more accurate than that based on normal approximation. For some $\alpha \in (0, 1)$, let $\left(\|\check{\mathbf{H}}_n^*\|\right)_\alpha$ be the αth quantile of the Bootstrap distribution of $\|\check{\mathbf{H}}_n^*\|$. Then the $100(1-\alpha)\%$ confidence region of $\boldsymbol{\beta}$ is given by the set $C_{1-\alpha} \subset \mathcal{R}^p$ where

$$C_{1-\alpha} = \left\{\boldsymbol{\beta} : \|\check{\mathbf{H}}_n\| \le \left(\|\check{\mathbf{H}}_n^*\|\right)_{(1-\alpha)}\right\}.$$

This confidence region can also be used to perform tests on $\boldsymbol{\beta}$. SOC result for one-dimensional smooth functions of $\boldsymbol{\beta}$ and the construction of the confidence intervals based on it are presented in the next section.

Remark 3.1. *The class of sets $\mathcal{A}_m$ used to state the uniform asymptotic results is somewhat abstract and hence some explanation is necessary. The class $\mathcal{A}_m$ identifies a collection of Borel subsets of $\mathcal{R}^m$ for which $\left[\epsilon^{-1}\Phi((\partial B)^\epsilon)\right]$ is bounded uniformly by a constant for sufficiently small ϵ. Or equivalently, $\mathcal{A}_m$ is a collection of Borel subsets of $\mathcal{R}^m$ with boundaries having upper Standard Gaussian-Minkowski content $\left(i.e.\ \limsup_{\epsilon\to 0+}\left[\epsilon^{-1}\Phi((\partial B)^\epsilon)\right]\right)$ upper bounded by a uniform constant. Such a class may be a generic one like the class of Borel measurable convex sets, the class of Euclidean balls, the class of rectangles, the class of half spaces etc., or it may be constructed based on the definition. For example if $\mathcal{A}_m$ is the collection of rectangles, then $\Phi\left((\partial B)^\epsilon\right) \le M\epsilon$ with $M = 4^m(2\pi)^{-m/2}$, uniformly for any $B \in \mathcal{A}_m$. Whereas if $\mathcal{A}_m$ is the collection of Borel measurable convex sets, then $\Phi\left((\partial B)^\epsilon\right) \le M\epsilon$ with $M = \sqrt{2}(m-1)(\Gamma((m-1)/2))(\Gamma(m/2))^{-1}$, uniformly for any $B \in \mathcal{A}_m$, due to Corollary 3.2 of Bhattacharya and Rao (1986). The two major reasons behind considering such a class in our asymptotic analysis go as follows; firstly, to obtain asymptotic normality or to obtain valid Edgeworth expansions for the normalized part of the underlying pivots, and secondly, to bound the remainder term by required small magnitude with sufficiently large probability (or Bootstrap probability).*

Remark 3.2. *Note that the random quantities Z and Z^* introduced in the pivots (3.1) and (3.2) respectively, are essential in achieving SOC of PEBBLE. The assumption of Z and Z^* being normally* *distributed is very essential, since a normal random vector has exponentially decaying tail with respect to its norm. One application of it is to use Mill's ratio inequality in the proof of Lemma 2.3 (see supplementary material file Das and Das (2024)). Moreover, the characteristic function of the normal random vector has an exponential form which is essential to blend it with the polynomials arising in the Edgeworth expansions. However, it is not essential to have $\mathbb{I}_p$ as the varaiance of Z and Z^*. Theorem 3.3 still remains true even if we replace $\mathbb{I}_p$ by any diagonal matrix with diagonal elements being independent of n.*

Remark 3.3. *The results on Bootstrap approximation presented in Theorem 3.3, can be established in almost sure sense as well. In that case, the only additional requirement is to have $n^{-1}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^{12} = O(1)$, since $y_1,\dots,y_n$ can take either 0 or 1. An almost sure version of part (a) of Theorem 3.3 is necessary to establish Theorem 3.2. Note that the requirement for almost sure version is met under the assumptions of Theorem 3.2.*

Remark 3.4. *In this remark, we comment on the incompetency of the existing Bootstrap methods to achieve SOC in logistic regression. Without proper smoothing, any Bootstrap method in logistic regression should fail to be second order correct, similar to what is observed in Theorem 3.2. This is primarily because the distribution of $\hat{\boldsymbol{\beta}}_n$ may possibly be lattice which introduces correction factors in the corresponding Edgeworth expansion and these correction factors cannot usually be approximated with an error $o(n^{-1/2})$ using Bootstrap even after adequate studentization. Before introducing necessary smoothing, the pivotal quantities must also be precisely centered and properly studentized. Moreover, the Bootstrap pivotal quantity should be a good quadratic approximation of the original pivotal quantity. All of these conditions are essential in order to achieve SOC by any Bootstrap method in logistic regression or more generally in any inference problem where the distribution of the underlying estimator may have a lattice structure. The pivotal quantities $(\check{\mathbf{H}}_n, \check{\mathbf{H}}_n^*)$ based on PEBBLE, considered above in this section, are constructed with these considerations.*

For the existing Bootstrap methods in the literature, at least one of these necessary steps seems to be not followed. For example in case of the Pearson residual resampling considered in Moulton and Zeger (1991), the resampling is carried out from non-centered standardized residuals which essentially leads to the corresponding Bootstrapped estimator which is not centered precisely. On the other hand, the Bootstrap methodologies explored in Lee (1990) and Claeskens et al. (2003) are properly centered and studentized. However, the one step Bootstrap defined in equation (3) of Claeskens et al. (2003) is based on a linear approximation (i.e. it can only be first order correct) whereas the quadratic Bootstrap defined in equation (5) of Claeskens et al. (2003) and the Bootstrap method of Lee (1990) are defined based on quadratic approximations. Let us focus on the quadratic Bootstrap estimator θ_n^* of Claeskens et al. (2003) and consider the logistic regression setup explored in Theorem 3.2. It can be shown that there exists an interval $\boldsymbol{B}_{1n}$ such that

$$\limsup_{n\to\infty}\mathbf{P}\Big(\sqrt{n}\Big|\mathbf{P}_*\big(\sqrt{n}(\theta_n^* - \hat{\beta}_n) \in \boldsymbol{B}_{1n}\big) - \mathbf{P}\big(\sqrt{n}(\hat{\beta}_n - \beta) \in \boldsymbol{B}_{1n}\big)\Big| > 0\Big) > 0. \tag{3.3}$$

To that end, define the sets $\boldsymbol{B}_{1n} = (-\infty, z_{1n})$ where $z_{1n} = [1/(2\sqrt{n}) - \sqrt{n}\mu_n]$ and $\boldsymbol{A}_n = \{n|\hat{\mu}_n - \mu_n| \in \cup_{j=0}^{\infty}(j+1/8, j+7/8)\}$, where $\mu_n = n^{-1}\sum_{i=1}^{n} x_i p(\beta|x_i)$ (as defined earlier) and $\hat{\mu}_n = n^{-1}\sum_{i=1}^{n} x_i \hat{p}(x_i)$. Now note that, unlike in the case of Theorem 3.2, here both $n\bar{\mathrm{W}}_{n1}$ and $n\bar{\mathrm{W}}_{n1}^*$ are sum of centered lattice random variables. Hence we need to follow the analysis of $\mathbf{H}_n$ in case of both $\sqrt{n}(\theta_n^* - \hat{\beta}_n)$ and $\sqrt{n}(\hat{\beta}_n - \beta)$. Clearly on the set $\boldsymbol{A}_n$, all the steps up to equation (3.6) in the proof of Theorem 3.2 (see supplementary material file Das and Das (2024)) remain valid. Therefore it is enough to show that

$\mathbf{P}(\boldsymbol{A}_n) \geq 2^{-1}\Big[\Phi(2/\sqrt{\boldsymbol{L}}) - \Phi(1/\sqrt{\boldsymbol{L}})\Big]$ and for sufficiently large n, on the set $\boldsymbol{A}_n$ we have

$$\sqrt{n}\Big|\mathbf{P}_*\Big(\sqrt{n}\bar{W}^*_{n1} \in \boldsymbol{B}_{1n}\Big) - \mathbf{P}\Big(\sqrt{n}\bar{W}_{n1} \in \boldsymbol{B}_{1n}\Big)\Big| > 1/16,$$

when $\{n\}$ is a sequence of perfect squares. Here $\bar{W}_{n1} = n^{-1}\sum_{i=1}^{n}(y_i - p(\beta|x_i))x_i$, $\bar{W}^*_{n1} = n^{-1}\sum_{i=1}^{n}(y^*_i - \hat{p}(x_i))x_i$ and $\boldsymbol{L}$ (> 0) is the limit of $\boldsymbol{L}_n = n^{-1}\sum_{i=1}^{n} x_i^2 e^{x_i\beta}(1+e^{x_i\beta})^{-2}$. Now by Lemma 2.10 (see supplementary material file Das and Das (2024)), we have

$$\sup_{x\in\mathcal{R}}\Big|\mathbf{P}\Big(\sqrt{n}\bar{W}_{n1} \leq x\Big) - \Phi_{\sigma_n^2}(x) - n^{-1/2}P_1\big(-\Phi_{\sigma_n^2} : \{\bar{\chi}_{\nu,n}\}\big)(x)$$
$$+ n^{-1/2}\big(n\mu_n + \sqrt{n}x - [n\mu_n + \sqrt{n}x] - 1/2\big)\frac{d}{dx}\Phi_{\sigma_n^2}(x)\Big| = o\big(n^{-1/2}\big)$$
$$\textit{and} \quad \sup_{x\in\mathcal{R}}\Big|\mathbf{P}\Big(\sqrt{n}\bar{W}^*_{n1} \leq x\Big) - \Phi_{\sigma_n^2}(x) - n^{-1/2}P_1\big(-\Phi_{\sigma_n^2} : \{\bar{\chi}_{\nu,n}\}\big)(x)$$
$$+ n^{-1/2}\big(n\hat{\mu}_n + \sqrt{n}x - [n\hat{\mu}_n + \sqrt{n}x] - 1/2\big)\frac{d}{dx}\Phi_{\sigma_n^2}(x)\Big| = o\big(n^{-1/2}\big),$$

where σ_n^2, $P_1\big(-\Phi_{\sigma_n^2} : \{\bar{\chi}_{\nu,n}\}\big)(\cdot)$ are all defined in Lemma 2.10 (see supplementary material file Das and Das (2024)) and $[x]$ is the greatest integer $\leq x$. Again, note that

$$\big(n\mu_n + \sqrt{n}z_{1n} - [n\mu_n + \sqrt{n}z_{1n}] - 1/2\big) = 0,$$

and on the set $\boldsymbol{A}_{1n}$ we have

$$\big|n\hat{\mu}_n + \sqrt{n}z_{1n} - [n\hat{\mu}_n + \sqrt{n}z_{1n}] - 1/2\big| > 1/8.$$

Therefore, clearly we have

$$\sqrt{n}\Big|\mathbf{P}_*\Big(\sqrt{n}\bar{W}^*_{n1} \in \boldsymbol{B}_{1n}\Big) - \mathbf{P}\Big(\sqrt{n}\bar{W}_{n1} \in \boldsymbol{B}_{1n}\Big)\Big| \geq \big|n\hat{\mu}_n + \sqrt{n}z_{1n} - [n\hat{\mu}_n + \sqrt{n}z_{1n}] - 1/2\big| - o(n^{-1/2}) > 1/16,$$

for sufficiently large n. Now we have to claim that $\mathbf{P}(\boldsymbol{A}_n) \geq 2^{-1}\Big[\Phi(2/\sqrt{\boldsymbol{L}}) - \Phi(1/\sqrt{\boldsymbol{L}})\Big]$ for large enough n. To establish this claim, note that

$$\mathbf{P}(\boldsymbol{A_n}) \geq \mathbf{P}\Big(n|\hat{\mu}_n - \mu_n| \in \cup_{j=0}^{2\sqrt{n}}(j + 1/8, j + 7/8)\Big)$$

and one can consider the fact that $\sqrt{n}(\hat{\mu}_n - \mu_n)$ is asymptotically distributed as $N(0, \boldsymbol{L})$ which is due to Theorem 3.1 and the delta method. Hence the quadratic Bootstrap of Claeskens et al. (2003) also cannot achieve SOC. Similar argument can also be made for the Bootstrap method defined in Lee (1990) as well. Therefore as mentioned before, in general, a Bootstrap method should not achieve SOC in logistic regression without necessary smoothing, even if it is based on a quadratic approximation and the pivotal quantities are properly centered and studentized.

4. Validity of PEBBLE for smooth functions

In logistic regression, often we are interested in drawing inference about some function of $\boldsymbol{\beta}$, like odds ratio, success probability etc., instead of $\boldsymbol{\beta}$ itself. In this section we explore the second order inference on a real-valued function of $\boldsymbol{\beta}$, say $f(\boldsymbol{\beta})$ where $f : \mathcal{R}^p \to \mathcal{R}$. To be precise, here our goal is to approximate the distribution of $\sqrt{n}\Big(f(\hat{\boldsymbol{\beta}}_n) - f(\boldsymbol{\beta})\Big)$ using PEBBLE. Similar to $\hat{\boldsymbol{\beta}}_n$, we need to invoke necessary smoothing beside performing studentization here as well. Since $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ is asymptotically normally distributed, the delta method implies that the asymptotic variance of $\sqrt{n}\Big(f(\hat{\boldsymbol{\beta}}_n) - f(\boldsymbol{\beta})\Big)$ is $\big(f'(\boldsymbol{\beta})\big)'\boldsymbol{L}_n^{-1}\big(f'(\boldsymbol{\beta})\big)$ where $f'(\boldsymbol{\beta})$ is the column vector of the partial derivatives of f with respect to $\boldsymbol{\beta}$ and the matrix $\boldsymbol{L}_n$ is as defined in the Subsection 3.1. Now upon observing that $\hat{\boldsymbol{L}}_n, \hat{\boldsymbol{M}}_n$ (defined in the previous section) are estimators of $\boldsymbol{L}_n$, an estimator of the asymptotic variance is $s_n^2 = \big(f'(\hat{\boldsymbol{\beta}}_n)\big)'\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1}\big(f'(\hat{\boldsymbol{\beta}}_n)\big)$. Then let us define the smoothed studentized version of $\sqrt{n}\big(f(\hat{\boldsymbol{\beta}}_n) - f(\boldsymbol{\beta})\big)$ as

$$T_n = s_n^{-1}\Big[\sqrt{n}\Big(f(\hat{\boldsymbol{\beta}}_n) - f(\boldsymbol{\beta})\Big) + b_n\big(f'(\hat{\boldsymbol{\beta}}_n)\big)'\hat{\boldsymbol{L}}_n^{-1}Z\Big], \tag{4.1}$$

where b_n is as defined in the statement of Theorem 3.3. Z is a standard normal random vector in $\mathcal{R}^p$ independent of responses $y_1, \ldots, y_n$. Similarly, let us define the Bootstrap version of $\boldsymbol{T}_n$ as

$$T_n^* = s_n^{*-1}\Big[\sqrt{n}\Big(f(\hat{\boldsymbol{\beta}}_n^*) - f(\hat{\boldsymbol{\beta}}_n)\Big) + b_n\big(f'(\hat{\boldsymbol{\beta}}_n^*)\big)'\boldsymbol{L}_n^{*-1}Z^*\Big], \tag{4.2}$$

where $s_n^{*2} = \big(f'(\hat{\boldsymbol{\beta}}_n^*)\big)'\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\boldsymbol{L}_n^{*-1}\big(f'(\hat{\boldsymbol{\beta}}_n^*)\big)$. $\boldsymbol{L}_n^*$ and $\hat{\boldsymbol{M}}_n^*$ are as defined in the previous section and Z^* is a standard normal random vector in $\mathcal{R}^p$ independent of responses $y_1, \ldots, y_n$, and $G_1^*, \ldots, G_n^*$, and also independent of Z.

Theorem 4.1. *Consider all the regularity conditions of Theorem 3.3. Additionally, consider a function $f : \mathcal{R}^p \to \mathcal{R}$ having continuous partial derivatives of order ≤ 2 and define T_n and T_n^* as in (4.1) and (4.2) respectively. Then we have*

$$\sup_{x \in \mathcal{R}}\big|\mathbf{P}_*(T_n^* \leq x) - \mathbf{P}(T_n \leq x)\big| = o_p\big(n^{-1/2}\big).$$

The above theorem establishes SOC of PEBBLE for any real valued smooth function of the regression parameter $\boldsymbol{\beta}$. This essentially implies that more accurate inference can be drawn for different smooth functions of $\boldsymbol{\beta}$, than that based on the normal approximation obtained by applying the delta method. Now we describe the forms of confidence intervals of $f(\boldsymbol{\beta})$ which can be constructed based on Theorem 4.1. Define $T_{\alpha,n}^*$ to be the αth quantile of the Bootstrap distribution of T_n^* for some $\alpha \in (0, 1)$. Then $100(1 - \alpha)\%$ two-sided confidence interval of $f(\boldsymbol{\beta})$ based on PEBBLE is the set $\big\{f(\boldsymbol{\beta}) : T_{\alpha/2,n}^* \leq T_n \leq T_{1-\alpha/2,n}^*\big\}$ which is essentially given by the interval

$$\left[\left\{f(\hat{\boldsymbol{\beta}}_n) - \frac{s_n u_{1j}^*(\alpha/2)}{\sqrt{n}}\right\}, \left\{f(\hat{\boldsymbol{\beta}}_n) - \frac{s_n l_{1j}^*(\alpha/2)}{\sqrt{n}}\right\}\right],$$

where $l_{1j}^*(\alpha) = \big[T_{\alpha,n}^* - s_n^{-1}b_n\big(f'(\hat{\boldsymbol{\beta}}_n)\big)'\hat{\boldsymbol{L}}_n^{-1}Z\big]$ and $u_{1j}^*(\alpha) = \big[T_{1-\alpha,n}^* - s_n^{-1}b_n\big(f'(\hat{\boldsymbol{\beta}}_n)\big)'\hat{\boldsymbol{L}}_n^{-1}Z\big]$. Similarly $100(1 - \alpha)\%$ lower and upper confidence intervals of $f(\boldsymbol{\beta})$ using PEBBLE are respectively given by

$$\left(-\infty, \left\{f(\hat{\boldsymbol{\beta}}_n) - \frac{s_n l_{1j}^*(\alpha)}{\sqrt{n}}\right\}\right] \quad \text{and} \quad \left[\left\{f(\hat{\boldsymbol{\beta}}_n) - \frac{s_n u_{1j}^*(\alpha)}{\sqrt{n}}\right\}, \infty\right).$$

4.1. Examples

Theorem 4.1 concludes that perturbation Bootstrap achieves SOC in drawing inference about any real valued smooth function of $\boldsymbol{\beta}$. The forms of the perturbation Bootstrap confidence intervals are also mentioned above. Explicit forms of the confidence intervals for three important special cases are discussed below.

Example 4.1: A linear combination of regression parameters

Sometimes it is of interest to assess whether a particular covariate is relevant or not. Or in general, one may be interested in comparing the contribution of two or more covariates in explaining the response variable. In such cases, it is essential to be able to draw inference about a linear combination of the regression parameters. Taking that into account, let us consider $f(\boldsymbol{\beta}) = \boldsymbol{d}'\boldsymbol{\beta}$ for some $\boldsymbol{d} \in \mathcal{R}^p$, and interest remains in drawing inference about $f(\boldsymbol{\beta})$. Then, s_n^2 and s_n^{*2} respectively become $s_n^2 = \boldsymbol{d}'(\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1})\boldsymbol{d}$ and $s_n^{*2} = \boldsymbol{d}'(\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\boldsymbol{L}_n^{*-1})\boldsymbol{d}$. Hence the form of T_n^* becomes

$$T_n^* = \Big[\boldsymbol{d}'(\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\boldsymbol{L}_n^{*-1})\boldsymbol{d}\Big]^{-1/2}\Big[\sqrt{n}\boldsymbol{d}'(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) + b_n\boldsymbol{d}'\boldsymbol{L}_n^{*-1}Z^*\Big]. \tag{4.3}$$

Then $100(1-\alpha)\%$ two-sided, lower one-sided and upper one-sided confidence intervals for $\boldsymbol{d}'\boldsymbol{\beta}$ are given by

$$\left[\left\{\boldsymbol{d}'\hat{\boldsymbol{\beta}}_n + \frac{b_n\boldsymbol{d}'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_nT^*_{1-\alpha/2,n}}{\sqrt{n}}\right\}, \left\{\boldsymbol{d}'\hat{\boldsymbol{\beta}}_n + \frac{b_n\boldsymbol{d}'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_nT^*_{\alpha/2,n}}{\sqrt{n}}\right\}\right],$$

$$\left(-\infty, \left\{\boldsymbol{d}'\hat{\boldsymbol{\beta}}_n + \frac{b_n\boldsymbol{d}'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_nT^*_{\alpha,n}}{\sqrt{n}}\right\}\right] \text{ and } \left[\left\{\boldsymbol{d}'\hat{\boldsymbol{\beta}}_n + \frac{b_n\boldsymbol{d}'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_nT^*_{1-\alpha,n}}{\sqrt{n}}\right\}, \infty\right),$$

respectively, where $s_n^2 = \boldsymbol{d}'(\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1})\boldsymbol{d}$ and $T^*_{\alpha,n}$ is the αth quantile of the Bootstrap distribution of T_n^*, defined at (4.3). In the special case of $f(\boldsymbol{\beta}) = \beta_j$, the jth component of $\boldsymbol{\beta}$, $100(1-\alpha)\%$ confidence intervals become

$$\left[\left\{\hat{\beta}_{jn} + \frac{b_n\left(\hat{\boldsymbol{L}}_n^{-1}\right)'_{j\cdot}Z}{\sqrt{n}} - \frac{\hat{\sigma}_{jn}T^*_{1-\alpha/2,n}}{\sqrt{n}}\right\}, \left\{\hat{\beta}_{jn} + \frac{b_n\left(\hat{\boldsymbol{L}}_n^{-1}\right)'_{j\cdot}Z}{\sqrt{n}} - \frac{\hat{\sigma}_{jn}T^*_{\alpha/2,n}}{\sqrt{n}}\right\}\right],$$

$$\left(-\infty, \left\{\hat{\beta}_{jn} + \frac{b_n\left(\hat{\boldsymbol{L}}_n^{-1}\right)'_{j\cdot}Z}{\sqrt{n}} - \frac{\hat{\sigma}_{jn}T^*_{\alpha,n}}{\sqrt{n}}\right\}\right] \text{ and } \left[\left\{\hat{\beta}_{jn} + \frac{b_n\left(\hat{\boldsymbol{L}}_n^{-1}\right)'_{j\cdot}Z}{\sqrt{n}} - \frac{\hat{\sigma}_{jn}T^*_{1-\alpha,n}}{\sqrt{n}}\right\}, \infty\right),$$

where $(\hat{\boldsymbol{L}}_n^{-1})'_{j\cdot}$ is the jth row of $\hat{\boldsymbol{L}}_n^{-1}$, $\hat{\sigma}^2_{jn}$ is the (j,j)th element of $\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1}$ and $T^*_{\alpha,n}$ is the αth quantile of the Bootstrap distribution of $T_n^* = \sigma_{jn}^{*-1}\big[\sqrt{n}(\hat{\beta}^*_{jn} - \hat{\beta}_{jn}) + b_n(\boldsymbol{L}_n^{*-1})'_{j\cdot}Z^*\big]$. Here $(\boldsymbol{L}_n^{*-1})'_{j\cdot}$ is the jth row of $\boldsymbol{L}_n^{*-1}$ and σ^{*2}_{jn} is the (j,j)th element of $\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\boldsymbol{L}_n^{*-1}$.

Example 4.2: Odds ratio

Odds ratio in the context of logistic regression is given by the ratio of the probabilities of observing a success in the response variable at the presence and absence of certain covariates. In other words, it measures the odds of a success in the response variable based on one or more covariates. The odds ratio at the covariate value $\boldsymbol{x} = \boldsymbol{x}_0$ is defined as $OR(x_0) = p(\boldsymbol{\beta}|\boldsymbol{x}_0)\big(1 - p(\boldsymbol{\beta}|\boldsymbol{x}_0)\big)^{-1} = e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$. If it is of interest to learn whether a particular covariate is relevant in explaining the response, then one can test whether the odds ratio, corresponding to one unit increase in that covariate keeping others fixed, is equal to 1 or not. This can be done by looking into two-sided confidence intervals for $OR(\boldsymbol{x}_0)$. For example, if we are interested to know whether jth covariate variable is relevant then we should consider $\boldsymbol{x}_0 = (0, \dots, 0, 1, 0, \dots, 0)'$ where 1 is at jth position.

Let us consider $f(\boldsymbol{\beta}) = OR(\boldsymbol{x}_0) = e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$. Then the $100(1-\alpha)\%$ two-sided, lower one-sided and upper one-sided confidence intervals for $e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$ are given by

$$\Bigg[\bigg\{e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n} + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_n T^*_{1-\alpha/2,n}}{\sqrt{n}}\bigg\}, \bigg\{e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n} + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_n T^*_{\alpha/2,n}}{\sqrt{n}}\bigg\}\Bigg],$$

$$\Bigg(-\infty, \bigg\{e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n} + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_n T^*_{\alpha,n}}{\sqrt{n}}\bigg\}\Bigg] \text{ and } \Bigg[\bigg\{e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n} + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}} - \frac{s_n T^*_{1-\alpha,n}}{\sqrt{n}}\bigg\}, \infty\Bigg),$$

respectively, where $s_n^2 = e^{2\boldsymbol{x}_0'\boldsymbol{\beta}}\big[\boldsymbol{x}_0'(\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1})\boldsymbol{x}_0\big]$ and $T^*_{\alpha,n}$ is the αth quantile of the Bootstrap distribution of $T_n^* = e^{-\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*}\big[\boldsymbol{x}_0'(\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\boldsymbol{L}_n^{*-1})\boldsymbol{x}_0\big]^{-1/2}\big[\sqrt{n}(e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*} - e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}) + b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*}\boldsymbol{x}_0'\boldsymbol{L}_n^{*-1}Z^*\big]$.

Example 4.3: Success probability

The success probability of the response y at the covariate value $\boldsymbol{x} = \boldsymbol{x}_0$ is defined as $p(\boldsymbol{\beta}|\boldsymbol{x}_0) = e^{\boldsymbol{x}_0'\boldsymbol{\beta}}(1 + e^{\boldsymbol{x}_0'\boldsymbol{\beta}})^{-1}$. We generally estimate this success probability by the predicted probability $\hat{p}(\boldsymbol{x}_0) = e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}})^{-1}$. Using our proposed Bootstrap method, we can draw second order accurate inference on $p(\boldsymbol{\beta}|\boldsymbol{x}_0)$ based on $\hat{p}(\boldsymbol{x}_0)$. Now we present the forms of the Bootstrap confidence intervals for $p(\boldsymbol{\beta}|\boldsymbol{x}_0)$ which can also be used to perform hypothesis testing. In the setup of Theorem 4.1, $f(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\boldsymbol{x}_0)$ and hence the $100(1-\alpha)\%$ two-sided, lower one-sided and upper one-sided Bootstrap confidence intervals for $p(\boldsymbol{\beta}|\boldsymbol{x}_0)$ are given by

$$\Bigg[\bigg\{\hat{p}(\boldsymbol{x}_0) + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n})^2} - \frac{s_n T^*_{1-\alpha/2,n}}{\sqrt{n}}\bigg\}, \bigg\{\hat{p}(\boldsymbol{x}_0) + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n})^2} - \frac{s_n T^*_{\alpha/2,n}}{\sqrt{n}}\bigg\}\Bigg],$$

$$\Bigg(-\infty, \bigg\{\hat{p}(\boldsymbol{x}_0) + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n})^2} - \frac{s_n T^*_{\alpha,n}}{\sqrt{n}}\bigg\}\Bigg] \text{ and } \Bigg[\bigg\{\hat{p}(\boldsymbol{x}_0) + \frac{b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n}\boldsymbol{x}_0'\hat{\boldsymbol{L}}_n^{-1}Z}{\sqrt{n}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n})^2} - \frac{s_n T^*_{1-\alpha,n}}{\sqrt{n}}\bigg\}, \infty\Bigg),$$

respectively, where $s_n^2 = e^{2\boldsymbol{x}_0'\boldsymbol{\beta}}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n})^{-4}\big[\boldsymbol{x}_0'(\hat{\boldsymbol{L}}_n^{-1}\hat{\boldsymbol{M}}_n\hat{\boldsymbol{L}}_n^{-1})\boldsymbol{x}_0\big]$ and $T^*_{\alpha,n}$ is the αth quantile of the Bootstrap distribution of $T_n^* = e^{-\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*})^2\big[\boldsymbol{x}_0'(\boldsymbol{L}_n^{*-1}\hat{\boldsymbol{M}}_n^*\ \boldsymbol{L}_n^{*-1})\boldsymbol{x}_0\big]^{-1/2}\big[\sqrt{n}\big(p^*(\boldsymbol{x}_0) - \hat{p}(\boldsymbol{x}_0)\big) + b_n e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*})^{-2}\boldsymbol{x}_0'\boldsymbol{L}_n^{*-1}Z^*\big]$, where $p^*(\boldsymbol{x}_0) = e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*}(1 + e^{\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_n^*})^{-1}$.

5. Simulation study

In this section, we compare the performance of PEBBLE with other existing methods via simulation experiments. For comparative study, we consider the Normal approximation (Normal), Pearson

Residual Resampling Bootstrap (PRRB) of Moulton and Zeger (1991), One-Step Bootstrap (OSB) and Quadratic Bootstrap (QB) of Claeskens et al. (2003). To explore the comparative performance in both scenarios-where the true coefficients are either entirely positive or consist of both positive and negative signs, we consider two possible values of $\boldsymbol{b}$ given by $\boldsymbol{b} = (1, .5, -2, -0.75, 1.5, -1, 1.85, -1.6)$ and $(1, .5, 2, 0.75, 1.5, 1, 1.85, 1.6)$. Note that $\boldsymbol{b}$ is of length 8. For the scenarios where $p \leq 8$, we take the true parameter vector $\boldsymbol{\beta}$ to be the first p-many elements of $\boldsymbol{b}$. The rows of the design matrix $\boldsymbol{X}$ is generated from multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma} = \{\sigma_{ij}\}_{p\times p}$ where $\sigma_{ij} = 0.5^{|i-j|}$.

In PEBBLE, we take three alternative values of b_n^2 given by

(a) $b_n^2 = n^{-\frac{1}{p_1+1}}$
(b) $b_n^2 = n^{-\frac{1}{p_1}}(\log n)^2$
(c) $b_n^2 = n^{-\frac{1}{2p_1}}$,

where $p_1 = \max\{p+1, 4\}$. Both Z and Z^* are drawn from independent multivariate normal distribution with mean $\mathbf{0}$ and variance $\frac{1}{4}\mathbb{I}_p$, $\mathbb{I}_p$ being $p \times p$ identity matrix. G_i^*'s are generated from $Beta(\frac{1}{2}, \frac{3}{2})$. Details on the forms of the confidence region for $\boldsymbol{\beta}$ based on PEBBLE are provided immediately after the statement of Theorem 3.3; and the form of the confidence intervals for one-dimensional smooth functions of $\boldsymbol{\beta}$ based on PEBBLE are presented in Section 4. PEBBLE as well as other methods namely Normal, PRRB, OSB and QB are all implemented in **R**. For the experiment, we consider 500 Bootstrap iterations. In order to find coverage, such experiment is repeated 500 times for each (n, p) scenario.

Thus, in total, we consider six scenarios under different combinations of values of $\boldsymbol{b}$ and b_n^2, given by:

- Scenario 1: $\boldsymbol{b} = (1, .5, -2, -0.75, 1.5, -1, 1.85, -1.6)$, $b_n^2 = n^{-\frac{1}{p_1+1}}$.
- Scenario 2: $\boldsymbol{b} = (1, .5, 2, 0.75, 1.5, 1, 1.85, 1.6)$, $b_n^2 = n^{-\frac{1}{p_1+1}}$.
- Scenario 3: $\boldsymbol{b} = (1, .5, -2, -0.75, 1.5, -1, 1.85, -1.6)$, $b_n^2 = n^{-\frac{1}{p_1}}(\log n)^2$.
- Scenario 4: $\boldsymbol{b} = (1, .5, 2, 0.75, 1.5, 1, 1.85, 1.6)$, $b_n^2 = n^{-\frac{1}{p_1}}(\log n)^2$.
- Scenario 5: $\boldsymbol{b} = (1, .5, -2, -0.75, 1.5, -1, 1.85, -1.6)$, $b_n^2 = n^{-\frac{1}{2p_1}}$.
- Scenario 6: $\boldsymbol{b} = (1, .5, 2, 0.75, 1.5, 1, 1.85, 1.6)$, $b_n^2 = n^{-\frac{1}{2p_1}}$.

In order to access the performance of all the methods for various dimensional coefficient vectors and sample sizes, we consider the following cases $(n, p) = (30, 3), (50, 3), (50, 4), (100, 3), (100, 4), (100, 6), (200, 3), (200, 4), (200, 6)$ and $(200, 8)$ for all six scenarios. We observe that PEBBLE yields better results for taking $b_n^2 = n^{-\frac{1}{p_1+1}}$ (i.e., scenario 1 and 2) compared to other two alternatives, in general. Therefore, our recommended choice for b_n^2 is $n^{-\frac{1}{p_1+1}}$. In the main paper we only demonstrate the results for lower $n : p$ cases (i.e. small sample size, high dimension) $(n, p) = (30, 3), (50, 4), (100, 6)$ and $(200, 8)$ for scenario 1 and 2. The full simulation study results for all six scenarios are provided in the supplementary material file Das and Das (2024).

In Table 1, we note down the empirical coverage of 90% confidence region of $\boldsymbol{\beta}$, upper, middle and lower 90% Confidence intervals (CIs) corresponding to the minimum and maximum components of $\boldsymbol{\beta}$ for scenario 1. We also note down the average empirical coverages of upper, middle and lower 90% CI over all components of $\boldsymbol{\beta}$. Average widths of 90% CI corresponding to all applicable cases are also noted in parenthesis. In Table 2, we list the values of these attributes for scenario 2. It is observed that in most of the cases PEBBLE performs better than other methods. Specifically for $(n, p) = (200, 8)$, it is noted that PEBBLE outperforms other methods by a wide margin. In this case, except for PEBBLE,

Table 1. Scenario 1: Comparative performance study of PEBBLE with Normal approximation (Normal), Pearson Residual Resampling Bootstrap (PRRB), One-Step Bootstrap (OSB) and Quadratic Bootstrap (QB) for components of $\boldsymbol{\beta}$. Empirical coverages of 90% confidence region of $||\boldsymbol{\beta}||$ (column 1), upper, lower and middle confidence intervals (CIs) of the minimum absolute value of $\boldsymbol{\beta}$ (columns 2,3,4), upper, lower and middle CIs of the maximum absolute value of the $\boldsymbol{\beta}$ (columns 5,6,7), upper, lower and middle CIs of the all components of $\boldsymbol{\beta}$, on average (columns 8,9,10) are presented, computed over 500 experiments. Average widths of the middle CIs are provided in parenthesis.

(n, p)	Methods	$\|\|\beta\|\|$	β_{min} middle (width)	β_{min} upper	β_{min} lower	β_{max} middle (width)	β_{max} upper	β_{max} lower	β avg. middle (width)	β avg. upper	β avg. lower
(30, 3)	PEBBLE	0.924	0.902 (2.98)	0.872	0.928	0.922 (4.12)	0.922	0.904	0.905 (3.25)	0.893	0.920
	Normal	0.948	0.940 (2.31)	0.956	0.896	0.966 (2.85)	0.916	0.998	0.957 (2.43)	0.937	0.940
	PRSB	0.938	0.918 (2.16)	0.926	0.862	0.944 (2.66)	0.918	0.944	0.929 (2.27)	0.913	0.910
	OSB	0.946	0.942 (2.35)	0.934	0.902	0.940 (2.66)	0.920	0.954	0.932 (2.38)	0.917	0.931
	QB	0.968	0.956 (2.50)	0.942	0.922	0.964 (3.06)	0.934	0.978	0.937 (2.53)	0.911	0.947
(50, 4)	PEBBLE	0.894	0.886 (3.05)	0.872	0.924	0.898 (4.12)	0.922	0.886	0.896 (2.85)	0.893	0.906
	Normal	0.926	0.922 (2.14)	0.954	0.892	0.946 (2.65)	0.896	0.982	0.935 (2.04)	0.919	0.923
	PRSB	0.916	0.904 (1.98)	0.928	0.872	0.924 (2.42)	0.890	0.934	0.899 (1.88)	0.901	0.892
	OSB	0.946	0.912 (2.19)	0.938	0.916	0.922 (2.44)	0.908	0.936	0.915 (2.03)	0.920	0.913
	QB	0.940	0.902 (2.09)	0.930	0.916	0.910 (2.40)	0.892	0.936	0.907 (1.98)	0.918	0.909
(100, 6)	PEBBLE	0.930	0.888 (1.82)	0.868	0.906	0.910 (2.88)	0.936	0.894	0.905 (2.13)	0.908	0.911
	Normal	0.870	0.870 (1.24)	0.868	0.870	0.910 (1.69)	0.884	0.940	0.873 (1.35)	0.876	0.890
	PRSB	0.858	0.842 (1.21)	0.858	0.864	0.884 (1.65)	0.872	0.906	0.846 (1.32)	0.865	0.874
	OSB	0.932	0.784 (1.29)	0.828	0.824	0.836 (1.66)	0.850	0.868	0.792 (1.37)	0.837	0.845
	QB	0.954	0.796 (1.37)	0.850	0.828	0.868 (1.84)	0.856	0.894	0.805 (1.45)	0.848	0.854
(200, 8)	PEBBLE	0.842	0.864 (1.76)	0.848	0.944	0.856 (2.30)	0.962	0.774	0.846 (1.94)	0.866	0.876
	Normal	0.406	0.664 (0.94)	0.878	0.668	0.740 (1.19)	0.708	0.958	0.685 (1.00)	0.782	0.794
	PRSB	0.492	0.650 (0.97)	0.876	0.670	0.734 (1.22)	0.702	0.944	0.683 (1.02)	0.780	0.799
	OSB	0.854	0.472 (0.97)	0.798	0.570	0.564 (1.16)	0.632	0.838	0.490 (1.00)	0.680	0.715
	QB	0.848	0.478 (0.98)	0.800	0.574	0.544 (1.14)	0.638	0.842	0.484 (0.98)	0.683	0.714

coverages of 90% confidence intervals for rest of the methods are noted to be far away from 0.9, in general. An overall picture of the superior performance of PEBBLE over other methods can be easily observed by examining the last three columns, where the average coverage accuracy for all components of $\boldsymbol{\beta}$ is presented. It is noted that for all the simulation scenarios, the average coverage over all coordinates is much closer to 0.90 for PEBBLE compared to other methods. Based on Tables 1 and 2, it can be said that for relatively smaller $n : p$ scenarios, the PEBBLE CIs are a little wider compared to other methods, but, as n increases (for fixed p), PEBBLE CI widths become closer to those observed for other methods. Lastly, we do not observe any notable differences in result when all coordinates of true $\boldsymbol{b}$ are of the same sign (scenario 2) compared to when $\boldsymbol{b}$ contains both positive and negative values (scenario 1).

In Tables 3 and 4, we consider the odds ratio $f(\boldsymbol{\beta}) = e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$ with all the components of $\boldsymbol{x}_0$ being 1 for scenario 1 and 2 respectively. We compare the empirical coverages of 90% lower, middle and upper CIs of $f(\boldsymbol{\beta})$ based on PEBBLE with Normal, PRRB, OSB and QB methods over 500 iterations. Here as well, PEBBLE is observed to outperform other methods. In particular, when $(n, p) = (100, 6)$ or $(200, 8)$, the overall performance of PEBBLE is notably better than the other methods. Here again, it is noted that the width of the PEBBLE CIs are little longer than that based on other methods. However, the widths of PEBBLE CIs tend to become similar to other methods as n increases keeping p fixed. The findings in Tables 1, 2, 3 and 4 demonstrate the validity of PEBBLE as a second order correct method for drawing statistical inference in the logistic regression model.

6. Application to healthcare operations decision

Vaginal delivery is known to be the most common type of birth. However,with advancement of medical procedures, caesarian delivery is often considered as an alternative way for delivery in presence

Table 2. Scenario 2: Comparative performance study of PEBBLE with Normal approximation (Normal), Pearson Residual Resampling Bootstrap (PRRB), One-Step Bootstrap (OSB) and Quadratic Bootstrap (QB) for components of $\boldsymbol{\beta}$. Empirical coverages of 90% confidence region of $||\boldsymbol{\beta}||$ (column 1), upper, lower and middle confidence intervals (CIs) of the minimum absolute value of $\boldsymbol{\beta}$ (columns 2,3,4), upper, lower and middle CIs of the maximum absolute value of the $\boldsymbol{\beta}$ (columns 5,6,7), upper, lower and middle CIs of the all components of $\boldsymbol{\beta}$, on average (columns 8,9,10) are presented, computed over 500 experiments. Average widths of the middle CIs are provided in parenthesis.

(n, p)	Methods	$\|\|\beta\|\|$	β_{min} middle (width)	β_{min} upper	β_{min} lower	β_{max} middle (width)	β_{max} upper	β_{max} lower	β avg. middle (width)	β avg. upper	β avg. lower
(30, 3)	PEBBLE	0.910	0.906 (3.50)	0.886	0.904	0.890 (4.29)	0.836	0.928	0.899 (4.03)	0.856	0.923
	Normal	0.904	0.958 (2.67)	0.958	0.910	0.932 (2.92)	0.998	0.870	0.944 (2.89)	0.981	0.890
	PRSB	0.954	0.860 (2.28)	0.868	0.862	0.880 (2.46)	0.922	0.856	0.881 (2.45)	0.908	0.854
	OSB	0.978	0.904 (2.66)	0.914	0.916	0.926 (2.70)	0.956	0.878	0.910 (2.75)	0.945	0.887
	QB	0.940	0.878 (2.55)	0.912	0.906	0.890 (2.35)	0.954	0.848	0.883 (2.55)	0.940	0.867
(50, 4)	PEBBLE	0.904	0.904 (3.06)	0.882	0.910	0.908 (3.82)	0.866	0.924	0.904 (2.97)	0.873	0.916
	Normal	0.932	0.932 (2.24)	0.928	0.902	0.948 (2.54)	0.982	0.902	0.936 (2.13)	0.946	0.897
	PRSB	0.900	0.870 (1.98)	0.890	0.870	0.916 (2.24)	0.922	0.892	0.880 (1.87)	0.890	0.875
	OSB	0.968	0.912 (2.30)	0.924	0.88	0.928 (2.46)	0.946	0.922	0.916 (2.14)	0.932	0.903
	QB	0.982	0.918 (2.36)	0.922	0.894	0.952 (2.61)	0.958	0.926	0.929 (2.24)	0.941	0.912
(100, 6)	PEBBLE	0.926	0.926 (2.52)	0.916	0.934	0.868 (3.26)	0.812	0.956	0.899 (2.69)	0.862	0.935
	Normal	0.864	0.934 (1.67)	0.932	0.912	0.908 (2.00)	0.998	0.830	0.920 (1.76)	0.965	0.876
	PRSB	0.840	0.846 (1.35)	0.854	0.864	0.846 (1.63)	0.970	0.760	0.842 (1.43)	0.909	0.818
	OSB	0.956	0.938 (1.72)	0.930	0.924	0.920 (1.94)	0.990	0.832	0.922 (1.77)	0.954	0.881
	QB	0.980	0.950 (1.82)	0.934	0.930	0.918 (1.93)	0.990	0.826	0.933 (1.87)	0.962	0.888
(200, 8)	PEBBLE	0.932	0.890 (2.26)	0.884	0.924	0.892 (2.95)	0.856	0.946	0.895 (2.44)	0.855	0.934
	Normal	0.828	0.930 (1.39)	0.952	0.872	0.904 (1.57)	0.990	0.826	0.906 (1.41)	0.968	0.839
	PRSB	0.622	0.772 (0.97)	0.874	0.768	0.764 (1.10)	0.946	0.682	0.748 (0.99)	0.905	0.719
	OSB	0.960	0.882 (1.32)	0.950	0.824	0.760 (1.39)	0.998	0.650	0.811 (1.31)	0.975	0.732
	QB	0.968	0.884 (1.35)	0.960	0.820	0.778 (1.45)	0.998	0.660	0.822 (1.34)	0.978	0.735

Table 3. Scenario 1: Comparative performance study of PEBBLE with Normal approximation (Normal), Pearson Residual Resampling Bootstrap (PRRB), One-Step Bootstrap (OSB) and Quadratic Bootstrap (QB) for the odds ratio $e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$ with components of $\boldsymbol{x}_0$ being 1. Empirical coverages of 90% middle, upper and lower CIs of the four methods are presented, computed over 500 experiments. Average width of the middle CIs are provided in parenthesis.

(n, p)	Methods	middle (width)	upper	lower
(30, 3)	PEBBLE	0.906 (1.13)	0.926	0.890
	Normal	0.902 (0.78)	0.958	0.854
	PRSB	0.892 (0.81)	0.978	0.844
	OSB	0.902 (0.84)	0.986	0.856
	QB	0.922 (0.94)	0.994	0.870
(50, 4)	PEBBLE	0.896 (0.53)	0.856	0.930
	Normal	0.894 (0.33)	0.912	0.904
	PRSB	0.880 (0.32)	0.912	0.898
	OSB	0.918 (0.36)	0.932	0.886
	QB	0.916 (0.34)	0.928	0.868
(100, 6)	PEBBLE	0.910 (1.20)	0.924	0.886
	Normal	0.868 (0.71)	0.932	0.838
	PRSB	0.834 (0.70)	0.910	0.804
	OSB	0.804 (0.72)	0.906	0.806
	QB	0.820 (0.81)	0.930	0.820
(200, 8)	PEBBLE	0.892 (0.25)	0.844	0.936
	Normal	0.766 (0.16)	0.800	0.812
	PRSB	0.726 (0.17)	0.804	0.794
	OSB	0.544 (0.16)	0.742	0.718
	QB	0.528 (0.15)	0.740	0.698

Table 4. Scenario 2: Comparative performance study of PEBBLE with Normal approximation (Normal), Pearson Residual Resampling Bootstrap (PRRB), One-Step Bootstrap (OSB) and Quadratic Bootstrap (QB) for the odds ratio $e^{\boldsymbol{x}_0'\boldsymbol{\beta}}$ with components of $\boldsymbol{x}_0$ being 1. Empirical coverages of 90% middle, upper and lower CIs of the four methods are presented, computed over 500 experiments. Average width of the middle CIs are provided in parenthesis.

(n, p)	Methods	middle (width)	upper	lower
(30, 3)	PEBBLE	0.912 (0.53)	0.852	0.986
	Normal	0.898 (0.40)	0.900	0.928
	PRSB	0.880 (0.43)	0.918	0.854
	OSB	0.896 (0.47)	0.934	0.892
	QB	0.852 (0.38)	0.906	0.854
(50, 4)	PEBBLE	0.928 (0.89)	0.882	0.934
	Normal	0.902 (0.36)	0.910	0.904
	PRSB	0.882 (0.34)	0.930	0.868
	OSB	0.912 (0.40)	0.926	0.900
	QB	0.914 (0.41)	0.936	0.898
(100, 6)	PEBBLE	0.906 (0.73)	0.862	0.934
	Normal	0.906 (0.48)	0.902	0.902
	PRSB	0.848 (0.41)	0.872	0.840
	OSB	0.920 (0.54)	0.958	0.900
	QB	0.934 (0.55)	0.962	0.902
(200, 8)	PEBBLE	0.896 (0.31)	0.866	0.914
	Normal	0.878 (0.21)	0.874	0.904
	PRSB	0.772 (0.16)	0.768	0.844
	OSB	0.908 (0.22)	0.862	0.930
	QB	0.910 (0.23)	0.882	0.930

Table 5. Real Data Analysis : The estimated coefficients and corresponding middle, upper and lower 90% CIs are noted for all the covariates; the type of delivery is the dependent variable, which takes values 1 or 0 based on if the delivery was caesarian or not.

Variables	$\hat{\beta}$	90% CI	90% CI	90% CI
		(mid)	(upper)	(lower)
Age	-0.010	(-0.151, 0.300)	>-0.100	<0.237
Delivery number	0.263	(-0.544, 0.740)	>-0.398	<0.601
Delivery time	-0.427	(-0.643, 0.466)	>-0.521	<0.348
Blood pressure	-0.251	(-0.709, 0.680)	>-0.548	<0.531
Heart problem	1.702	(-0.139, 2.327)	>0.145	<2.105

of several existing medical complications and preferences. Recently a few studies showed how the recommended type of delivery may depend on various clinical aspects of the mother including age, blood pressure and heart problem (Amorim et al., 2017; Pieper, 2012; Rydahl et al., 2019). We consider a dataset[1] concerning delivery results of 80 pregnant women along with several important related clinical covariates of the subjects. We regress the type of delivery (caesarian or not) on several related covariates namely age, delivery number, delivery time, blood pressure and presence of heart problem. Delivery time can take three values 0 (timely), 1 (premature) and 2 (latecomer). Blood pressure is denoted by 0, 1, 2 for the cases low, normal and high respectively. The covariate denoting the presence of heart problem is also binary; where 0 and 1 denote apt behaviour and inept condition, respectively. We perform a logistic regression and CIs of the regression parameters are computed using PEBBLE. Results are provided in Table 5.

[1] https://archive.ics.uci.edu/ml/datasets/Caesarian+Section+Classification+Dataset

It is noted that although 90% CIs for all the covariates contain zero, however, the 90% CI for heart problem belong to the positive quadrant mostly; also the upper 90% CI completely belongs to the positive quadrant, which implies women with heart problems tend to have caesarian procedure, coinciding with the findings in Yap et al. (2008) and Balci et al. (2011).

7. Proof of the results

7.1. Notations

In this section, we present the proofs of Theorems 3.1 and Theorem 3.3. Proofs of Theorem 3.2 and Theorem 4.1 are relegated to the supplementary material file Das and Das (2024). The statements as well as the proofs of a set of auxiliary lemmas, which are necessary to prove the theorems, are also provided in the supplementary material file Das and Das (2024). Before proceeding to the proofs, let us define a few notations. Suppose that Φ_V and ϕ_V respectively denote the normal distribution and its density with mean 0 and covariance matrix V. We write $\Phi_V = \Phi$ and $\phi_V = \phi$ when the dispersion matrix V is the identity matrix. $\mathcal{N}$ denotes the set of natural numbers. $C(\cdot), C_1(\cdot), \ldots$ denote generic constants which depend on only their arguments. For a non-negative integral vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)'$ and a function $f = (f_1, f_2, \ldots, f_l) : \mathcal{R}^l \rightarrow \mathcal{R}^l$, $l \geq 1$, let $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_l$, $\boldsymbol{\alpha}! = \alpha_1! \ldots \alpha_l!$, $f^{\boldsymbol{\alpha}} = (f_1^{\alpha_1}) \ldots (f_l^{\alpha_l})$, $D^{\boldsymbol{\alpha}} f_1 = D_1^{\alpha_1} \cdots D_l^{\alpha_l} f_1$, where $D_j f_1$ denotes the partial derivative of f_1 with respect to the jth component of $\boldsymbol{\alpha}$, $1 \leq j \leq l$. We write $D^{\boldsymbol{\alpha}} = D$ if $\boldsymbol{\alpha}$ has all the component equal to 1. For $\mathbf{t} = (t_1, t_2, \cdots t_l)' \in \mathcal{R}^l$ and α as above, define $\mathbf{t}^{\boldsymbol{\alpha}} = t_1^{\alpha_1} \cdots t_l^{\alpha_l}$. For any two vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{R}^k$, $\boldsymbol{\alpha} \leq \boldsymbol{\beta}$ implies that each of the component of $\boldsymbol{\alpha}$ is smaller than that of $\boldsymbol{\beta}$. For a set $\boldsymbol{A}$ and real constants a_1, a_2, $a_1\boldsymbol{A} + a_2 = \{a_1 y + a_2 : y \in \boldsymbol{A}\}$, ∂A is the boundary of A and A^ϵ denotes the $\epsilon-$neighbourhood of A for any $\epsilon > 0$. For any natural number m, the class of sets $\mathcal{A}_m$ is the collection of Borel subsets of $\mathcal{R}^m$ satisfying

$$\sup_{B \in \mathcal{A}_m} \Phi((\partial B)^\epsilon) = O(\epsilon) \quad \text{as } \epsilon \downarrow 0. \tag{7.1}$$

7.2. Proofs of Theorem 3.1 and Theorem 3.3

Proof of Theorem 3.1: Recall that the studentized pivot is given by

$$\tilde{\mathbf{H}}_n = \sqrt{n}\hat{\boldsymbol{L}}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}),$$

where $\hat{\boldsymbol{L}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' e^{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n}\big(1 + e^{\boldsymbol{x}_i'}\hat{\boldsymbol{\beta}}_n\big)^{-2}$. $\hat{\boldsymbol{\beta}}_n$ is the solution of (1.2). By Taylor's theorem, from (1.2) we have

$$\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n (y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))\boldsymbol{x}_i - (2n)^{-1}\sum_{i=1}^n \boldsymbol{x}_i e^{z_i}(1 - e^{z_i})(1 + e^{z_i})^{-3}\big[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\big]^2, \tag{7.2}$$

where $|z_i - \boldsymbol{x}_i'\boldsymbol{\beta}| \leq |\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})|$ for all $i \in \{1, \ldots, n\}$. Now due to the assumption $n^{-1}\sum_{i=1}^n \|\boldsymbol{x}_i\|^3 = O(1)$, by Lemma 2.1 (see supplementary material file Das and Das (2024)), taking $t = 3$, we have

$$\mathbf{P}\Big(\Big|n^{-1}\sum_{i=1}^n (y - p(\boldsymbol{\beta}|\boldsymbol{x}_i))x_{ij}\Big| \leq C_{40}(p)n^{-1/2}(\log n)^{1/2}\Big) = o\big(n^{-1/2}\big), \tag{7.3}$$

for any $j \in \{1, \ldots, p\}$. Again by assumption, $\boldsymbol{L}_n$ converges to some positive definite matrix $\boldsymbol{L}$. Moreover,

$$\Big\|(2n)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i e^{z_i}(1-e^{z_i})(1+e^{z_i})^{-3}\big[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big]^2\Big\| \le \Big(n^{-1}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^3\Big)\|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}\|^2.$$

Hence (7.2) can be rewritten as

$$(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}) = f_n(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}),$$

where f_n is a continuous function from $\mathcal{R}^p$ to $\mathcal{R}^p$ satisfying $\mathbf{P}\big(\|f_n(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\| \le C_{40}n^{-1/2}(\log n)^{1/2}\big) = 1-o\big(n^{-1/2}\big)$ whenever $\|(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\| \le C_{40}n^{-1/2}(logn)^{1/2}$. Therefore, part (a) of Theorem 3.1 follows by Brouwer's fixed point theorem (See also equation (2.33) at page 448 of Bhattacharya and Ghosh (1978)). Now we prove part (b). Note that from (7.2) and the fact that $\boldsymbol{L}_n$ converges to some positive definite matrix $\boldsymbol{L}$, for sufficiently large n we have,

$$\tilde{\mathbf{H}}_n = \hat{\boldsymbol{L}}_n^{1/2}\big[\boldsymbol{L}_n^{-1}\Lambda_n + R_{1n}\big]. \tag{7.4}$$

Here $\Lambda_n = n^{-1/2}\sum_{i=1}^{n}(y-p(\boldsymbol{\beta}|\boldsymbol{x}_i))\boldsymbol{x}_i$ and $2R_{1n} = -\boldsymbol{L}_n^{-1}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{x}_i e^{z_i}(1-e^{z_i})(1+e^{z_i})^{-3}\big[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\big]^2$ with $|z_i-\boldsymbol{x}_i'\boldsymbol{\beta}| \le |\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})|$ for all $i \in \{1, \ldots, n\}$. $\boldsymbol{L}_n$ and $\hat{\boldsymbol{L}}_n$ are as defined earlier. Now applying part (a) we have $\mathbf{P}\Big(\|R_{1n}\| = O\big(n^{-1/2}(\log n)\big)\Big) = 1-o\Big(n^{-1/2}\Big)$. Again by Taylor's theorem we have

$$\hat{\boldsymbol{L}}_n - \boldsymbol{L}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big(1-e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)\big(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)^{-3}\big[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\big] + \boldsymbol{L}_{1n}, \tag{7.5}$$

where by part (a), we have $\mathbf{P}\Big(\|\boldsymbol{L}_{1n}\| = O\Big(n^{-1}(\log n)\Big)\Big) = 1-o\Big(n^{-1/2}\Big)$. Hence using Lemma 2.6 (see supplementary material file Das and Das (2024)), part (a) and Taylor's theorem, one can show that $\mathbf{P}\Big(\|\hat{\boldsymbol{L}}_n^{1/2}-\boldsymbol{L}_n^{1/2}\| = O\Big(n^{-1/2}(\log n)^{1/2}\Big)\Big) = 1-o\Big(n^{-1/2}\Big)$. Therefore (7.3) and (7.5) will imply that

$$\tilde{\mathbf{H}}_n = \boldsymbol{L}_n^{-1/2}\Lambda_n + R_{2n},$$

where $\mathbf{P}\big(\|R_{2n}\| = O\big(n^{-1/2}(\log n)\big)\big) = 1-o\big(n^{-1/2}\big)$. Hence for any set $B \in \mathcal{A}_p$, there exists a constant $C_{41}(p) > 0$ such that

$$\begin{aligned}
&\Big|\mathbf{P}\Big(\tilde{\mathbf{H}}_n \in B\Big) - \Phi(B)\Big| \le \Big|\mathbf{P}\Big(\tilde{\mathbf{H}}_n \in B\Big) - \mathbf{P}\Big(\boldsymbol{L}_n^{-1/2}\Lambda_n \in B\Big)\Big| + \Big|\mathbf{P}\Big(\boldsymbol{L}_n^{-1/2}\Lambda_n \in B\Big) - \Phi(B)\Big| \\
&\le \mathbf{P}\big(\|R_{2n}\| > C_{41}(p)n^{-1/2}(\log n)\big) + 2\mathbf{P}\big(\boldsymbol{L}_n^{-1/2}\Lambda_n \in (\partial B)^{C_{41}(p)n^{-1/2}(\log n)}\big) \\
&\quad + \big|\mathbf{P}\big(\boldsymbol{L}_n^{-1/2}\Lambda_n \in B\big) - \Phi(B)\big| \\
&= O\Big(n^{-1/2}(\log n)\Big).
\end{aligned}$$

The last equality is a consequence of Lemma 2.5 (see supplementary material file Das and Das (2024)) and the bound on $\|R_{2n}\|$. Therefore part (b) is proved.

Proof of Theorem 3.3: By applying Taylor's theorem, it follows from (2.1) that

$$\hat{\boldsymbol{L}}_n(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{p}(\boldsymbol{x}_i))\boldsymbol{x}_i\mu_{G^*}^{-1}(G_i^* - \mu_{G^*})$$
$$- (2n)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i e^{z_i^*}(1 - e^{z_i^*})(1 + e^{z_i^*})^{-3}\big[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)\big]^2, \tag{7.6}$$

where $|z_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}| \leq |\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)|$ for all $i \in \{1,\ldots,n\}$. Now rest of part (a) of Theorem 3.3 follows exactly in the same line as the proof of part (a) of Theorem 3.1. To establish part (b), assume that $W_i = \big(Y_i\boldsymbol{x}_i', \big[Y_i^2 - \mathbf{E}Y_i^2\big]\boldsymbol{z}_i'\big)'$ and $W_i^* = \big(\hat{Y}_i\big[(G_i^* - \mu_{G^*})\mu_{G^*}^{-1}\big]\boldsymbol{x}_i', \hat{Y}_i^2\big[\mu_{G^*}^{-2}(G_i^* - \mu_{G^*})^2 - 1\big]\boldsymbol{z}_i'\big)'$. Here $Y_i = (y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))$ and $\hat{Y}_i = (y_i - \hat{p}(\boldsymbol{x}_i))$. First we show that

$$\check{\mathbf{H}}_n = \sqrt{n}\Big(H\big(\bar{W}_n + n^{-1/2}b_nZ\big)\Big) + R_n \quad\text{and}\quad \check{\mathbf{H}}_n^* = \sqrt{n}\Big(\hat{H}\big(\bar{W}_n^* + n^{-1/2}b_nZ\big)\Big) + R_n^*,$$

for some functions $H, \hat{H} : \mathcal{R}^k \rightarrow \mathcal{R}^p$ where $k = p + q$ with $2q = p(p+1)$. $H(\cdot), \hat{H}(\cdot)$ have continuous partial derivatives of all orders, $H(\mathbf{0}) = \hat{H}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{P}\big(\|R_n\| = o\big(n^{-1/2}\big)\big) = 1 - o\big(n^{-1/2}\big)$ & $\mathbf{P}_*\big(\|R_n^*\| = o\big(n^{-1/2}\big)\big) = 1 - o_p\big(n^{-1/2}\big)$. Next step is to apply Lemma 2.3, Lemma 2.4 and Lemma 2.7 (see supplementary material file Das and Das (2024)) to claim that suitable Edgeworth expansions exist for both $\check{\mathbf{H}}_n$ and $\check{\mathbf{H}}_n^*$. The last step is to conclude SOC of Bootstrap by comparing the Edgeworth expansions. Now (7.2) and part (a) of Theorem 3.1 imply that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \boldsymbol{L}_n^{-1}\Big[\Lambda_n - \xi_n/2\Big] + R_{3n}, \tag{7.7}$$

where $\mathbf{P}\big(\|R_{3n}\| \leq C_{42}(p)n^{-1}(\log n)^{3/2}\big)\big) = 1 - o\big(n^{-1/2}\big)$. Here $\Lambda_n = n^{-1/2}\sum_{i=1}^{n} Y_i\boldsymbol{x}_i$ and $\xi_n = n^{-3/2}\sum_{i=1}^{n}\boldsymbol{x}_i e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big(1 - e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)\big(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)^{-3}\big[\boldsymbol{x}_i'\big(\boldsymbol{L}_n^{-1}\Lambda_n\big)\big]^2$. Clearly, $\mathbf{P}\big(\|\xi_n\| \leq C_{43}(p)n^{-1/2}(\log n)\big)\big) = 1 - o\big(n^{-1/2}\big)$. Therefore, by Taylor's theorem we have

$$\sqrt{n}\big(\hat{\boldsymbol{L}}_n - \boldsymbol{L}_n\big)\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\big) = \xi_n + R_{4n}, \tag{7.8}$$

where $\mathbf{P}\big(\|R_{4n}\| \leq C_{44}(p)n^{-1}(\log n)^2\big)\big) = 1 - o\big(n^{-1/2}\big)$. Again noting (7.8), by equation (5) at page 52 of Turnbull (1930) we have

$$\hat{\boldsymbol{M}}_n^{-1/2} - \boldsymbol{L}_n^{-1/2} = -\boldsymbol{L}_n^{-1/2}\boldsymbol{Z}_{1n}\boldsymbol{L}_n^{-1/2} + \boldsymbol{Z}_{2n}, \tag{7.9}$$

where $\big(\hat{\boldsymbol{M}}_n - \boldsymbol{L}_n\big) = \boldsymbol{L}_n^{1/2}\boldsymbol{Z}_{1n} + \boldsymbol{Z}_{1n}\boldsymbol{L}_n^{1/2}$ and $\|Z_{2n}\| \leq \|\hat{\boldsymbol{M}}_n - \boldsymbol{L}_n\|^2$. It is also easy to show that

$$\mathbf{P}\Big(\|\hat{\boldsymbol{M}}_n - \boldsymbol{M}_n\| \leq C_{45}(p)n^{-1}(\log n)\Big) + \mathbf{P}\Big(\|\boldsymbol{M}_n - \boldsymbol{L}_n\| \leq C_{45}(p)n^{-1/2}(\log n)\Big) = 1 - o\big(n^{-1/2}\big),$$

where $\boldsymbol{M}_n = n^{-1}\sum_{i=1}^{n} Y_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$. Hence we have $\mathbf{P}\big(\|\boldsymbol{Z}_{2n}\| \leq C_{46}(p)n^{-1}(\log n)^2\big)\big) = 1 - o\big(n^{-1/2}\big)$. Therefore from (7.7)-(7.9), Lemma 2.8 (see supplementary material file Das and Das (2024)) and the fact that $b_n = O(n^{-d})$ (for some $d > 0$) will imply that

$$\check{\mathbf{H}}_n = \boldsymbol{L}_n^{-1/2}\Big[\Lambda_n + b_nZ + \xi_n/2\Big] - \boldsymbol{L}_n^{-1/2}\Big[\int_0^\infty e^{-t\boldsymbol{L}_n^{1/2}}(\boldsymbol{M}_n - \boldsymbol{L}_n)e^{-t\boldsymbol{L}_n^{1/2}}dt\Big]\boldsymbol{L}_n^{-1/2}\Lambda_n + R_{5n}, \tag{7.10}$$

where $\mathbf{P}\big(\|R_{5n}\| \leq C_{47}(p)n^{-1/2}(\log n)^{-1}\big) = 1 - o\big(n^{-1/2}\big)$. Now writing $W_i = (W_{i1}', W_{i2}')'$ and $\bar{W}_n = n^{-1}\sum_{i=1}^n W_i = (\bar{W}_{n,1}', \bar{W}_{n2}')'$ with W_{i1} has first p components of W_i for all $i \in \{1, \ldots, n\}$, we have

$$\begin{aligned}
\Lambda_n + b_n Z &= \sqrt{n}\big(\bar{W}_{n1} + n^{-1/2} b_n Z\big) \\
\xi_n &= n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big(1 - e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)\big(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)^{-3}\Big[\bar{W}_{n1}'\boldsymbol{L}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{L}_n^{-1}\bar{W}_{n1}\Big]^2 \\
&= \sqrt{n}\Big(\bar{W}_{n1}'\tilde{\boldsymbol{M}}_1\bar{W}_{n1}, \ldots, \bar{W}_{n1}'\tilde{\boldsymbol{M}}_p\bar{W}_{n1}\Big)',
\end{aligned}$$

where $\tilde{\boldsymbol{M}}_k = n^{-1}\sum_{i=1}^n x_{ik} e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big(1 - e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)\big(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\big)^{-3}\Big(\boldsymbol{L}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{L}_n^{-1}\Big)$ for $k \in \{1, \ldots, p\}$. Hence writing $\tilde{W}_{n1} = \bar{W}_{n1} + n^{-1/2} b_n Z$ we have

$$\boldsymbol{L}_n^{-1/2}\Big[\Lambda_n + b_n Z + \xi_n/2\Big] = \sqrt{n}\Big[\boldsymbol{L}_n^{-1/2}\tilde{W}_{n1} + \Big(\tilde{W}_{n1}'\breve{\boldsymbol{M}}_1\tilde{W}_{n1}, \ldots, \tilde{W}_{n1}'\breve{\boldsymbol{M}}_p\tilde{W}_{n1}\Big)'\Big] + R_{51n}, \tag{7.11}$$

where $\mathbf{P}\big(\|R_{51n}\| = o\big(n^{-1/2}\big)\big) = 1 - o\big(n^{-1/2}\big)$, since $b_n = O(n^{-d})$ and $\|\tilde{\boldsymbol{M}}_k\| = O(1)$ for any $k \in \{1, \ldots, p\}$. Here $\breve{\boldsymbol{M}}_k = \sum_{j=1}^p L_{kjn}^{-1/2}\tilde{\boldsymbol{M}}_k$, $k \in \{1, \ldots, p\}$, with $L_{kjn}^{-1/2}$ being the (k, j)th element of $\boldsymbol{L}_n^{-1/2}$. Again the jth row of $\big(\boldsymbol{M}_n - \boldsymbol{L}_n\big)$ is $\bar{W}_{n2}'\boldsymbol{E}_{jn}$ where $\boldsymbol{E}_{jn}$ is a matrix of order $q \times p$ with $\|\boldsymbol{E}_{jn}\| \leq q$, $j \in \{1, \ldots, p\}$. Therefore from (7.2) and (7.5) we have

$$\boldsymbol{L}_n^{-1/2}\Big[\int_0^\infty e^{-t\boldsymbol{L}_n^{1/2}}\big(\boldsymbol{M}_n - \boldsymbol{L}_n\big)e^{-t\boldsymbol{L}_n^{1/2}}dt\Big]\boldsymbol{L}_n^{-1/2}\Lambda_n = \sqrt{n}\Big(\tilde{W}_{n2}'\bar{\boldsymbol{M}}_1\tilde{W}_{n1}, \ldots, \tilde{W}_{n2}'\bar{\boldsymbol{M}}_p\tilde{W}_{n1}\Big)', \tag{7.12}$$

where $\tilde{W}_{n2} = \bar{W}_{n2} + n^{-1/2} b_n Z_1$ with $Z_1 \sim N_q\big(\mathbf{0}, \boldsymbol{I}_q\big)$, independent of Z & $\{y_1, \ldots, y_n\}$. $\bar{\boldsymbol{M}}_k = \int_0^\infty \big[\sum_{j=1}^p m_{kjn}(t)\check{\boldsymbol{M}}_j(t)\big]dt$ where $m_{kjn}(t)$ is the (k, j)th element of the matrix $\boldsymbol{L}_n^{-1/2}e^{-t\boldsymbol{L}_n^{1/2}}$ and $\check{\boldsymbol{M}}_j(t) = \boldsymbol{E}_{jn}e^{-t\boldsymbol{L}_n^{1/2}}\boldsymbol{L}_n^{-1/2}$, $k, j \in \{1, \ldots, p\}$. Now define the $(p+q) \times (p+q)$ matrices $\{\boldsymbol{M}_1^\dagger, \ldots, \boldsymbol{M}_p^\dagger\}$ where $\boldsymbol{M}_k^\dagger = \begin{bmatrix}\breve{\boldsymbol{M}}_k & \mathbf{0} \\ \bar{\boldsymbol{M}}_k & \mathbf{0}\end{bmatrix}$. Therefore from (7.10)-(7.12) we have

$$\check{\mathbf{H}}_n = \sqrt{n}\Big[\Big(\boldsymbol{L}_n^{-1/2}\ \ \mathbf{0}\Big)\tilde{W}_n + \Big(\tilde{W}_n'\boldsymbol{M}_1^\dagger\tilde{W}_n, \ldots, \tilde{W}_n'\boldsymbol{M}_p^\dagger\tilde{W}_n\Big)'\Big] + R_n = \sqrt{n}H\big(\tilde{W}_n\big) + R_n, \tag{7.13}$$

where the function $H(\cdot)$ has continuous partial derivatives of all orders, $\tilde{W}_n = \big(\tilde{W}_{n1}', \tilde{W}_{n2}'\big)'$ and $R_n = R_{5n} + R_{51n} + R_{6n}$.

Following the same line of arguments, writing $\bar{W}_{n1}^* = n^{-1}\sum_{i=1}^n W_{i1}^* = n^{-1}\sum_{i=1}^n \hat{Y}_i\mu_{G^*}^{-1}(G_i^* - \mu_{G^*})\boldsymbol{x}_i$ and $\bar{W}_{n2}^* = n^{-1}\sum_{i=1}^n W_{i2}^* = n^{-1}\sum_{i=1}^n \hat{Y}_i^2\big[\mu_{G^*}^{-2}(G_i^* - \mu_{G^*})^2 - 1\big]\boldsymbol{z}_i$, it can be shown that

$$\check{\mathbf{H}}_n^* = \sqrt{n}\Big[\Big(\hat{\boldsymbol{M}}_n^{-1/2}\ \ \mathbf{0}\Big)\tilde{W}_n^* + \Big(\tilde{W}_n^{*\prime}\boldsymbol{M}_1^{*\dagger}\tilde{W}_n^*, \ldots, \tilde{W}_n^{*\prime}\boldsymbol{M}_p^{*\dagger}\tilde{W}_n^*\Big)'\Big] + R_n^* = \sqrt{n}\hat{H}\big(\tilde{W}_n^*\big) + R_n^*, \tag{7.14}$$

where $\tilde{W}_n^* = \big(\tilde{W}_{n1}^{*\prime}, \tilde{W}_{n2}^{*\prime}\big)'$ with $\tilde{W}_{n1}^* = \bar{W}_{n1}^* + n^{-1/2} b_n Z^*$ and $\tilde{W}_{n2}^* = \bar{W}_n^* + n^{-1/2} b_n Z_1^*$; Z_1^* being a $N_q\big(\mathbf{0}, \boldsymbol{I}_q\big)$ distributed random vector independent of $\{G_1^*, \ldots, G_n^*\}$ and Z^*. $\boldsymbol{M}_k^{*\dagger} = \begin{bmatrix}\breve{\boldsymbol{M}}_k^* & \mathbf{0} \\ \bar{\boldsymbol{M}}_k^* & \mathbf{0}\end{bmatrix}$ where

$\check{\boldsymbol{M}}^*_k = \sum_{j=1}^p \hat{M}^{-1/2}_{kjn} \tilde{\boldsymbol{M}}^*_j$ with $\hat{M}^{-1/2}_{kjn}$ being the (k,j)th element of $\hat{\boldsymbol{M}}_n^{-1/2}$, $\tilde{\boldsymbol{M}}^*_j$ being same as $\tilde{\boldsymbol{M}}_j$ after replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_n$. $\bar{\boldsymbol{M}}^*_j = \int_0^\infty \left[\sum_{j=1}^p m^*_{kjn} \check{\boldsymbol{M}}^*_j(t)\right] dt$ where $m^*_{kjn}(t)$ is the (k,j)th element of the matrix $\hat{\boldsymbol{M}}_n^{-1/2} e^{-t\hat{\boldsymbol{M}}_n^{1/2}}$ and $\check{\boldsymbol{M}}^*_j(t) = \boldsymbol{E}_{jn} e^{-t\hat{\boldsymbol{M}}_n^{1/2}} \hat{\boldsymbol{M}}_n^{-1/2}$. Also $\mathbf{P}_*\left(\|R^*_n\| \le C_{49} n^{-1/2} (\log n)^{-1}\right) = 1 - o_p\left(n^{-1/2}\right)$. Now by applying Lemma 2.3, Lemma 2.4 and Lemma 2.7 (see supplementary material file Das and Das (2024)) with $s = 3$, Edgeworth expansions of the densities of $\check{\mathbf{H}}_n$ and $\check{\mathbf{H}}^*_n$ can be found uniformly over the class $\mathcal{A}_p$ upto an error $o\left(n^{-1/2}\right)$ and $o_p\left(n^{-1/2}\right)$ respectively. Call those Edgeworth expansions $\tilde{\psi}_{n,3}(\cdot)$ and $\tilde{\psi}^*_{n,3}(\cdot)$ respectively. Now if $\tilde{\psi}_{n,3}(\cdot)$ is compared with $\check{\psi}_{n,3}(\cdot)$ of Lemma 2.4 (see supplementary material file Das and Das (2024)), then $\check{\boldsymbol{M}}^\dagger_n = I_p$. Similarly for $\tilde{\psi}^*_{n,3}(\cdot)$ also $\check{\boldsymbol{M}}^\dagger_n = I_p$. Therefore, $\tilde{\psi}_{n,3}(\cdot)$ and $\tilde{\psi}^*_{n,3}(\cdot)$ have the forms

$$\tilde{\psi}_{n,3}(\boldsymbol{x}) = \left[1 + n^{-1/2} q_1(\boldsymbol{\beta}, \mu_W, \boldsymbol{x}) + \sum_{j=1}^{m_2-1} b_n^{2j} q_{2j}(\boldsymbol{\beta}, \boldsymbol{L}_n, \boldsymbol{x})\right] \phi(\boldsymbol{x})$$

$$\tilde{\psi}^*_{n,3}(\boldsymbol{x}) = \left[1 + n^{-1/2} q_1(\hat{\boldsymbol{\beta}}_n, \hat{\mu}_W, \boldsymbol{x}) + \sum_{j=1}^{m_2-1} b_n^{2j} q_{2j}(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{M}}_n, \boldsymbol{x})\right] \phi(\boldsymbol{x}),$$

where $m_2 = \inf\{j : b_n^{2j} = o(n^{-1/2})\}$, μ_W is the vector containing $\{n^{-1} \sum_{i=}^n \mathbf{E}(y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))^2 x_{ij}^{l_1} x_{ij'}^{l_2} : j, j' \in \{1, \ldots, p\}, l_1, l_2 \in \{0,1,2\}, l_1 + l_2 = 2\}$ and $\{n^{-1} \sum_{i=}^n \mathbf{E}(y_i - p(\boldsymbol{\beta}|\boldsymbol{x}_i))^3 x_{ij}^{l_1} x_{ij'}^{l_2} x_{ij''}^{l_3} : j, j', j'' \in \{1, \ldots, p\}\ , l_1, l_2, l_3 \in \{0,1,2,3\}, l_1 + l_2 + l_3 = 3\}$. $\hat{\mu}_W$ is the vector of $\{n^{-1} \sum_{i=}^n (y_i - \hat{p}(\boldsymbol{x}_i))^2 x_{ij}^{l_1} x_{ij'}^{l_2} : j, j' \in \{1, \ldots, p\}, l_1, l_2 \in \{0,1,2\}, l_1 + l_2 = 2\}$ and $\{n^{-1} \sum_{i=}^n (y_i - \hat{p}(\boldsymbol{x}_i))^3 x_{ij}^{l_1} x_{ij'}^{l_2} x_{ij''}^{l_3} : j, j', j'' \in \{1, \ldots, p\}, l_1, l_2 \in \{0,1,2,3\}, l_1 + l_2 + l_3 = 3\}$. $q_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is a polynomial in $\boldsymbol{c}$ whose coefficients are continuous functions of $(\boldsymbol{a}, \boldsymbol{b})'$. $q_{2j}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ are polynomials in $\boldsymbol{c}$ whose coefficients are continuous functions of $\boldsymbol{a}$ and $\boldsymbol{b}$. Now Theorem 3.3 follows by comparing $\tilde{\psi}_{n,3}(\cdot)$ and $\tilde{\psi}^*_{n,3}(\cdot)$ and due to part (a) of Theorem 3.1.

8. Conclusion

In this paper we investigate the second order correct approximation of the distribution of the logistic regression estimator. Second order correctness (SOC) is necessary for achieving higher level of accuracy in drawing inferences even when the sample size is relatively small. SOC in logistic regression is quite different from that in linear regression because the distribution of the logistic regression estimator is discrete, which may result in the distribution taking a lattice form. The possible lattice nature of the distribution reduces the accuracy of normal approximation by introducing an extra '$\log n$' term in the usual Berry-Essen rate of $O(n^{-1/2})$. Due to the same reason, it is observed that the proposed perturbation Bootstrapped estimator or PEBBLE, does not achieve SOC even after usual studentization, unlike in case of linear regression. The smoothing developed in Lahiri (1993) has been utilized to define smoothed versions of both logistic regression estimator and its Bootstrap version. The advantage of this smoothing is two-fold. Firstly, it makes both the original and Bootstrap distributions to be absolutely continuous with respect to Lebesgue measure, i.e. the smoothing removes the lattice structure which might gives rise to complications. On the other hand, it has negligible effect on the underlying variances, resulting the same studentization to work. Further SOC is established for PEBBLE by utilizing the Edgeworth expansion theory. Reproducible codes used for Tables in Sections 5 and 6 are made available on github[2].

The discrete nature of the underlying distribution prevails not only in logistic regression, but in many sub-models of Generalized linear models (or GLM). Therefore, it would be interesting to explore whether similar smoothing works in achieving SOC for GLM. Moreover, if the underlying dimension is high compared to the sample size, then the problem becomes more interesting since the smoothing actually depends on the dimension. We leave the theoretical investigation in that direction as a potential future work.

Supplementary Material

Proofs of Theorem 3.2 and Theorem 4.1 from the main paper are presented. All auxiliary lemmas are provided along with their proofs. Additional simulation results are also reported.

Software: Reproducible R codes used for Tables in Sections 5 and 6 are made available on github, at the following link: https://github.com/priyamdas2/PEBBLE.

References

Amemiya, T. (1976). The maximum likelihood, the minimum chi-square and the nonlinear weighted least-squares estimator in the general qualitative response model. *J. Amer. Statist. Assoc.* **71** 347–351. DOI: 10.1080/01621459.1976.10480346.

Amorim, M. M., Souza, A. S., and Katz, L. (2017). Planned caesarean section versus planned vaginal birth for severe pre-eclampsia. *Cochrane Database Syst. Rev.* **10** CD009430. DOI: 10.1002/14651858.CD009430.pub2.

Balci, A., Drenthen, W., Mulder, B., Roos-Hesselink, J., Voors, A., Vliegen, H., Moons, P., Sollie, K., Dijk, A., Veldhuisen, D., and Pieper, P. (2011). Pregnancy in women with corrected tetralogy of fallot: occurrence and predictors of adverse events. *Am. Heart J.* **161** 307–313. DOI: 10.1016/j.ahj.2010.10.027.

Barbe, P. and Bertail, P. (1995). *The Weighted Bootstrap - Lecture Notes in Statistics*. Springer. DOI: 10.1007/978-1-4612-2532-4.

Berkson, J. (1944). Application of the logistic function to bio-assay. *J. Amer. Statist. Assoc.* **39** 357–365. DOI: 10.2307/2280041.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal edgeworth expansion. *Annals of Statistics* **6** 434–451.

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions*. John Wiley & Sons. DOI: 10.1137/1.9780898719895.fm.

Claeskens, G., Aerts, M., and Molenberghs, G. (2003). A quadratic bootstrap method and improved estimation in logistic regression. *Stat. Probab. Lett.* **61** 383–394. DOI: 10.1016/S0167-7152(02)00397-8.

Cox, D. R. (1958). The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B Stat. Method.* **20** 215–232. DOI: 10.1111/j.2517-6161.1958.tb00292.x.

Das, D. and Das, P. (2024). Supplementary material for “pebble: a second order correct bootstrap method in logistic regression.” .

Das, D., Gregory, K., and Lahiri, S. N. (2019). Perturbation bootstrap in adaptive lasso. *Ann. Statist.* **47** 2080–2116. DOI: 10.1214/18-AOS1741.full.

Das, D. and Lahiri, S. N. (2019). Second order correctness of perturbation bootstrap m-estimator of multiple linear regression parameter. *Bernoulli* **25** 654–682. DOI: 10.3150/17-BEJ1001.full.

[2] https://github.com/priyamdas2/PEBBLE

Davison, A. C., Hinkley, D. V., and Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika* **73** 555–566. DOI: 10.2307/2336519.

Diciccio, T. and Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika* **79** 231–245. DOI: 10.2307/2336835.

Diciccio, T. and Efron, B. (1996). Bootstrap confidence intervals. *Statist. Sci.* **11** 189–212. DOI: 10.1214/ss/1032280214.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26. DOI: 10.1214/aos/1176344552.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368. DOI: 10.1214/aos/1176346597.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228. DOI: 10.1214/aos/1176345638.

Ghosh, J. K. (1994). *Higher order asymptotics*, volume 4. NSF-CBMS Regional Conf. Ser. Probab. Statist. DOI: 10.1214/cbms/1462297300.

Gourieroux, C. and Monfort, A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *J. Econometrics* **17** 83–97. DOI: 10.1016/0304-4076(81)90060-9.

Haberman, S. J. (1974). Log-linear models for frequency tables derived by indirect observation: maximum likelihood equations. *Ann. Statist.* **2** 911–924. DOI: 10.1214/aos/1176342813.

Hosmer, D. W., Lameshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.

Kong, F. and Levin, B. (1996). Edgeworth expansions for the conditional distributions in logistic regression models. *J. Statist. Plann. Inference* **52** 109–129. DOI: 10.1016/0378-3758(95)00106-9.

Lahiri, S. N. (1992). Bootstrapping m-estimators of a multiple linear regression parameter. *Ann. Statist.* **20** 1548–1570. DOI: 10.1016/0047-259X(92)90112-S.

Lahiri, S. N. (1993). Bootstrapping the studentized sample mean of lattice variables. *J. Multivariate Anal.* **45** 247–256. DOI: 10.1006/jmva.1993.1037.

Lahiri, S. N. (1994). On two-term edgeworth expansions and bootstrap approximations for studentized multivariate m-estimators. *Sankhya A* **56** 201–226. http://www.jstor.org/stable/25050986. Accessed 14 Oct. 2024.

Lee, K. W. (1990). Bootstrapping logistic regression models with random regressors. *Commun. Stat. Theory Methods* **19** 2527–2539. DOI: 10.1080/03610929008830332.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. DOI: 10.1214/aos/1176351062.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21** 255–285. DOI: 10.1214/aos/1176349025.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics, New York: Academic Press* pages 105–142.

Moulton, L. H. and Zeger, S. L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics* **45** 381–394. DOI: 10.2307/2531484.

Moulton, L. H. and Zeger, S. L. (1991). Bootstrapping generalized linear models. *Comput. Stat. Data Anal.* **11** 53–63. DOI: 10.1016/0167-9473(91)90052-4.

Pieper, P. G. (2012). The pregnant woman with heart disease: management of pregnancy and delivery. *Neth. Heart J.* **20** 33–37. 10.1007/s12471-011-0209-y.

Rydahl, E., Declercq, E., Juhl, M., and Maimburg, R. (2019). Cesarean section on a rise - does advanced maternal age explain the increase? a population register-based study. *PLoS One* **14** e0210655. DOI: 10.1371/journal.pone.0210655 PMCID: PMC6345458.

Singh, K. (1981). On the asymptotic accuracy of efron's bootstrap. *Ann. Statist.* **9** 1187–1195. DOI: 10.1214/aos/1176345636.

Sun, J., Loader, C., and McCormick, W. (2000). Confidence bands in generalized linear models. *Ann. Statist.* **28** 429–460. DOI: 10.1214/aos/1016218225.

Turnbull, H. W. (1930). A matrix form of taylor's theorem. *Proc. Edinburgh Math. Soc.* **2** 33–54. DOI: 10.1017/S0013091500007537.

Yap, S., Drenthen, W., Pieper, P., Moons, P., Mulder, B., Mostert, B., Vliegen, H., Dijk, A., Meijboom, F., Steegers, E., and Roos-Hesselink, J. (2008). Risk of complications during pregnancy in women with congenital aortic stenosis. *Int. J. Cardiol.* **126** 240–246. DOI: 10.1016/j.ijcard.2007.03.134.