Statistica Sinica **35** (2025), 91-110
doi:https://doi.org/10.5705/ss.202022.0296

AN ADAPTIVELY RESIZED PARAMETRIC BOOTSTRAP FOR INFERENCE IN HIGH-DIMENSIONAL GENERALIZED LINEAR MODELS

Qian Zhao* and Emmanuel J. Candès

University of Massachusetts Amherst and Stanford University

Abstract: Accurate statistical inference in logistic regression models remains a critical challenge when the ratio between the number of parameters and sample size is not negligible. This is because approximations based on either classical asymptotic theory or bootstrap calculations are grossly off the mark. This paper introduces a resized bootstrap method to infer model parameters in arbitrary dimensions. As in the parametric bootstrap, we resample observations from a distribution, which depends on an estimated regression coefficient sequence. The novelty is that this estimate is actually far from the maximum likelihood estimate (MLE). This estimate is informed by recent theory studying properties of the MLE in high dimensions, and is obtained by appropriately shrinking the MLE towards the origin. We demonstrate that the resized bootstrap method yields valid confidence intervals in both simulated and real data examples. Our methods extend to other high-dimensional generalized linear models.

Key words and phrases: Bootstrap, confidence interval, generalized linear models, high-dimensional statistics.

1. Introduction

The bootstrap is a well-known resampling procedure introduced in Efron's seminal paper (Efron, 1979) for approximating the distribution of a statistic of interest. Its popularity stems from a combination of several elements: it is conceptually rather straightforward; it is flexible and can be deployed in a whole suite of delicate inference problems (Efron, 1981, 1985; Efron, Halloran and Holmes, 1996); and finally, whenever theoretical calculations are impossible, the bootstrap often provides an excellent approximation to the distribution under study. As a result, researchers from a spectacular array of disciplines have used the bootstrap for hypothesis testing (Politis, Romano and Wolf, 1999, Ch. 1), model selection (Shao, 1996), density estimation (Franke and Härdle, 1992), and many other important statistical inference problems.

The bootstrap can usually be understood via the plug-in principle (Efron and Tibshirani, 1994, Ch. 4). Suppose we observe $X_i \in \mathbb{R}^p$, $i = 1, \ldots, n$, sampled independently and identically from a distribution F. We wish to infer

*Corresponding author. E-mail: qzhao1@stanford.edu

the distribution of a statistic $t_F(X_1, X_2, \ldots, X_n)$, which can be a complicated functional of the data aimed at estimating the number of modes F has. For instance, we may be interested in the 90% quantile of $t_F(X_1, \ldots, X_n)$. The subscript F indicates which distribution X_i are sampled from; here, the X_i's are i. i. d. samples from F. The plug-in principle estimates the distribution of $t_F(X_1, \ldots, X_n)$ by that of $t_{\hat{F}}(X_1^*, \ldots, X_n^*)$, wherein $\hat{F}$ is an estimate of F, and $(X_1^*, \ldots, X_n^*)$ is a draw from $\hat{F}$. In other words, by resampling observations from $\hat{F}$, we obtain a distribution we hope closely resembles that of $t_F(X_1, \ldots, X_p)$.

Naturally, statisticians have since the beginning studied the accuracy of the bootstrap. Broadly speaking, the bootstrap is known to be consistent, i.e., $t_{\hat{F}}(X_1^*, \ldots, X_n^*) \longrightarrow t_F(X_1, \ldots, X_n)$ in distribution, under the conditions that (1) the distribution of $t_F(X_1, X_2, \ldots, X_n)$ varies smoothly near F, and (2) $\hat{F}$ converges to F (Bickel and Freedman, 1981; Diciccio and Romano, 1988; Politis, Romano and Wolf, 1999, Ch. 1). The second condition is typically satisfied for appropriately chosen estimates $\hat{F}$ whenever the data dimension p is fixed. In addition to general theory, statisticians have carried out detailed studies for specific statistics including the sample mean (Bickel and Freedman, 1981; Hall, 1992), regression coefficients (Shorack, 1982; Bickel and Freedman, 1982, 1981; Mammen, 1993), and continuous functions of the empirical measure (Gine and Zinn, 1990), and so forth.

Motivated by the abundance of high-dimensional data, researchers are increasingly studying statistical methods in the *high-dimensional setting* in which the number of variables p grows with the number of observations n. Specifically, this article concerns the accuracy of bootstrap methods when p and n are both very large and perhaps grow with a fixed ratio. In linear regression for example, while the residual bootstrap is weakly consistent if p is fixed and $n \to \infty$, it is inconsistent when $n, p \to \infty$ in such a way that $p/n \to \kappa > 0$; to be sure, Bickel and Freedman (1982) displays a data-dependent contrast, i.e., a linear combination of coefficients, for which the estimated contrast distribution is asymptotically incorrect. Motivated by results from high-dimensional maximum likelihood theory (El Karoui et al., 2013; El Karoui, 2013, 2018), El Karoui and Purdom (2018) proposed to use corrected residuals to achieve correct inference. Another example is this: although the nonparametric bootstrap can be used to construct a valid confidence region for the spectrum of a covariance matrix when the problem dimension is fixed (Beran and Srivastava, 1985; Eaton and Tyler, 1991), it yields incorrect estimates of the distribution of the largest eigenvalue if $p/n \to \kappa > 0$ (El Karouis and Purdom, 2016). With the exception of these two studies, the accuracy of the bootstrap in other high-dimensional problems has not been much researched.

In this paper, we study the bootstrap for inferring the distribution of the maximum likelihood estimator (MLE) in high-dimensional logistic regression models. We find that the standard parametric bootstrap and the pairs bootstrap

are both incorrect (Sec. 1.2), a finding which echoes with El Karoui and Purdom (2018). We also show that recent high-dimensional maximum likelihood theory (HDT) developed for multivariate Gaussian covariates does not correctly predict the distribution of the MLE when the covariates are heavy tailed; this is analogous to findings in El Karoui (2018). Both these failures call for solutions and in this paper, we design a novel resized bootstrap by combining the bootstrap method with insights from HDT. We demonstrate that the resized bootstrap yields confidence intervals attaining nominal coverage regardless of the covariate distribution. Finally, we extend our methods to other generalized linear models.

1.1. High-dimensional maximum likelihood theory

We begin by briefly reviewing recent theory about M-estimators in the high-dimensional setting in which both the number of observations n and the number of variables p go to ∞ while the ratio p/n approaches a constant $\kappa > 0$. This high-dimensional theory (HDT) generalizes the classical asymptotic setting, and offers a more accurate characterization of the distribution of M-estimators when both n and p are large. In particular, a considerable amount of research has studied the behavior of M-estimators in high-dimensional regression and penalized regression (El Karoui et al., 2013; El Karoui, 2013; Zhang and Zhang, 2014; van de Geer et al., 2014; Donoho and Montanari, 2016; Celentano, Montanari and Wei, 2023; Bellec, Shen and Zhang, 2022).

Consider a logistic model in which the covariates $X \in \mathbb{R}^p$ are mutivariate Gaussian and $\mathbb{P}(Y = 1|X) = \sigma(X^\top \beta)$, where $\sigma(t) = 1/(1 + e^{-t})$ is the usual sigmoid function. Zhao, Sur and Candès (2020) showed that if $\hat{\beta}$ denotes the MLE, then

$$\frac{\sqrt{n}(\hat{\beta}_j - \alpha_\star \beta_j)}{\sigma_\star/\tau_j} \xrightarrow{d} \mathcal{N}(0, 1), \tag{1.1}$$

where β_j (resp. $\hat{\beta}_j$) is the jth (resp. estimated) model coefficient. In contrast to classical asymptotic theory, which states that the MLE is unbiased, the MLE is centered at $\alpha_\star \beta_j$, for some $\alpha_\star > 1$ whenever κ is positive. The standard deviation is $\sigma_\star/\tau_j$; here, τ_j is the conditional standard deviation of the jth variable given all the other variables whereas the parameters $\alpha_\star$ and $\sigma_\star$ are determined by κ and the signal strength γ defined as $\gamma^2 = \mathrm{Var}(X^\top \beta)$. The parameters $\alpha_\star$ and $\sigma_\star$ both increase as either the dimensionality κ increases or the signal-to-noise ratio γ increases (Sur and Candès, 2019, Fig. 7). To be complete, we stress that Eqn. (1.1) holds with the proviso that the magnitude of β_j is not extremely large. Zhao, Sur and Candès (2020) hypothesized that HDT holds when $\tau_j \beta_j = o(1)$. Empirically, they observed that the theoretical inflation and std. dev. are reasonably correct when $\tau_j \beta_j/\gamma \leq 0.15$; however, when $\tau_j \beta_j/\gamma = 0.5$, the empirical std.dev. is 36% larger than the theoretical std.dev.

The approximation (1.1) happens to be very accurate for moderately large sample sizes, e.g., when $n = 4000$ and $p = 400$ (Zhao, Sur and Candès, 2020), and is accurate for relatively small sample sizes, i.e., $n = 200$ and $p = 20$ (Sur and Candès, 2019, Appendix G). Further, (1.1) is expected to hold for sub-Gaussian covariates, see Zhao, Sur and Candès (2020) for empirical studies supporting this claim.

Having said all of this, (1.1) does not hold when the covariates follow a general distribution. For instance, El Karoui (2018) studied ridge regression in linear models where the covariates follow a multivariate t-distribution, and proved that the variance of the ridge estimate does depend on the geometry of the covariates. Similarly, for a logistic regression we expect that $\alpha_\star$ and $\sigma_\star$ would also depend on the degrees of freedom of the t-distribution. In Section 1 of the supplementary material (SM), we give a conjecture about the distribution of the MLE, and compare it with empirical observations. Aside from these two scenarios, we know very little about the distribution of M-estimators, or the inflation and std.dev. of the MLE when the covariates follow an arbitrary distribution.

1.2. An example with non-Gaussian covariates

Having succinctly described the high-dimensional theory, we simulate a high-dimensional logistic regression model with 4,000 observations and 400 covariates ($n = 4000$ and $p = 400$). We sample covariates from a multivariate t-distribution and standardize each variable so that $\mathrm{Var}(X_j) = 1/p$. We pick 50 non-null variables and sample their coefficients from a mixture of Gaussians $\mathcal{N}(5, 1)$ and $\mathcal{N}(-5, 1)$ with equal weights.

Figure 1 presents a histogram of a coordinate of the MLE from repeated experiments. From the bell-shaped curve, we conclude that the MLE is approximately Gaussian. Although the value of the true coefficient under study is 4.78, the average MLE is 5.56, which shows that the MLE is biased upward and the inflation factor is roughly equal to $\alpha_j = \bar{\beta}_j/\beta_j = 1.16$. The empirical standard deviation (std.dev.) of the MLE is equal to 1.34; however, the classical theory estimates that the std.dev. equals 1.15. We thus see that because of both a poor centering and a poor assessment of variability, the classical Wald confidence interval would significantly undercover β_j. Now HDT from Section 1.1 estimates the bias to be $\alpha_\star = 1.14$ and the standard deviation to be $\sigma_\star/\tau_j = 1.25$. This implies that while capturing the bias, HDT slightly underestimates the std.dev. of the MLE.

Next, we apply the parametric bootstrap and pairs bootstrap and display in Figure 1 the density curves of the bootstrap MLEs from one experiment.

- For the parametric bootstrap, we generate samples by fixing the covariates at the observed values and sample responses from a logistic model whose coefficients equal the MLE; put another way, we choose $\hat{F} = F_{\hat{\beta}}$. The para-

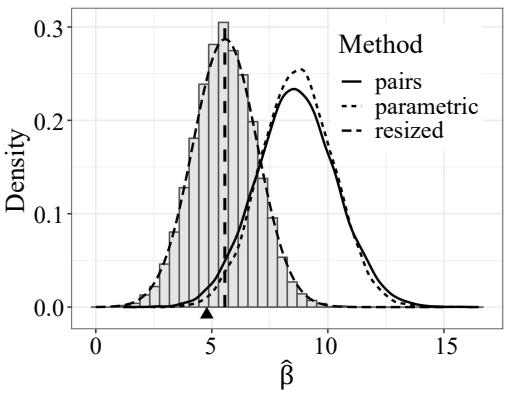


Figure 1. Histogram of the logistic MLE of a randomly chosen coefficient in 10,000 repeated experiments. Here, the covariates are sampled from a multivariate t-distribution with 8 degrees of freedom. The bootstrap MLE densities are displayed for the parametric bootstrap (dashed), the pairs bootstrap (solid) and the proposed resized bootstrap (longdash). The triangle indicates the true coefficient and the dashed line indicates the average MLE.

metric bootstrap (dashed) does not begin to describe the MLE distribution since the average value is 8.68, about twice that of the true coefficient, and the std.dev. is 1.55.

- The pairs bootstrap generates bootstrap samples by sampling with replacement from the observed data, i.e., we choose $\hat{F}$ to be the empirical distribution. The pairs bootstrap also fails to approximate the MLE distribution since the solid curve shifts to the right and is much wider than the histogram (mean is 8.63 and std.dev. is 1.71).

Finally, the longdashed curve in Figure 1 shows the accuracy of the proposed resized bootstrap. We can see that this best describes the MLE distribution; for instance, both the mean (5.54) and standard deviation (1.39) are close to the true values.

2. Why Does the Bootstrap Fail?

The pairs bootstrap fails in the high-dimensional setting because it effectively inflates the dimensionality ratio $\kappa = p/n$. In particular, when n is large, the number of unique pairs (X_i^*, Y_i^*) in a bootstrap sample is approximately $(1-1/e)n$ on average (Mendelson et al., 2016). Consequently, the effective dimensionality ratio $\kappa e/(e-1)$ in the bootstrap sample is larger than κ. Because the bias and variance of the MLE increase as κ increases (Sur and Candès, 2019, Fig. 7), the pairs bootstrap tends to over-estimate both the bias and standard error.

While the pairs bootstrap over-estimates κ, the parametric bootstrap fails because the signal strength γ is inflated in the bootstrap samples. Suppose for

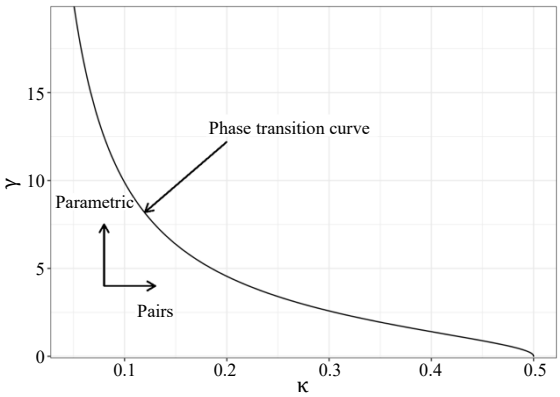


Figure 2. According to the high-dimensional theory (Sec. 1.1), the asymptotic distribution of the MLE depends on the problem dimension κ and the signal strength γ. The pairs bootstrap over-estimates κ whereas the parametric bootstrap over-estimates γ. Therefore, both methods lead to incorrect estimates of the MLE distribution. The blue region shows pairs of values of (κ, γ) where the MLE exists when $\beta_0 = 0$.

simplicity that the covariates are independent $\mathcal{N}(0, 1)$. Then (Sur and Candès, 2019, Thm. 2) shows that

$$\lim_{n,p\to\infty} \mathrm{Var}(X_{\mathrm{new}}^\top \hat{\beta}) \overset{a.s.}{=} \alpha_\star^2 \gamma^2 + \kappa \sigma_\star^2 > \gamma^2, \tag{2.1}$$

whereas $\mathrm{Var}(X^\top \beta) = \gamma^2$. Here, X_{new} is a new random sample independent from the training set. Because a higher γ leads to higher bias and variance (Sur and Candès, 2019, Fig. 7), the parametric bootstrap also tends to over-estimate the bias and standard error of the MLE.

In addition to over-estimating the bias and standard error, another problem of using the bootstrap is that when working with bootstrap samples, the MLE may cease to exist. We can explain this issue via the phase transition: for every ratio κ and intercept β_0, there exists an asymptotic threshold $\gamma(\kappa, \beta_0)$ such that the MLE does not exist once the signal strength $\gamma > \gamma(\kappa, \beta_0)$. Similarly, for every γ and β_0, there exists a threshold $\kappa(\gamma, \beta_0)$ such that the MLE does not exist once $\kappa > \kappa(\gamma, \beta_0)$. Because the pairs bootstrap over-estimates κ while the parametric bootstrap over-estimates γ, the bootstrap MLE may not exist if either κ or γ exceeds the phase transition threshold. Figure 2 provides a visual illustration of these points.

3. A Resized Bootstrap Method

We propose to construct parametric bootstrap samples from $F_{\beta_\star}$, where $\beta_\star$ is obtained by shrinking the MLE towards zero. We would like $\beta_\star$ to obey $\mathrm{Var}(X_{\mathrm{new}}^\top \beta_\star) = \gamma^2 = \mathrm{Var}(X^\top \beta)$ as to preserve the signal-to-noise ratio. We

set out to estimate γ in Section 3.1 since γ is unobserved. Upon obtaining $\beta_\star$, we follow the standard parametric bootstrap procedure to generate bootstrap samples. That is to say, the bth bootstrap sample consists of (x_i, Y_i^b), $i = 1, \ldots, n$, where x_i is the vector of features for the ith sample and Y_i^b is sampled from our GLM with features x_i and coefficients $\beta_\star$. We then compute the bootstrap MLE $\hat{\beta}^b \in \mathbb{R}^p$ by fitting the GLM using pairs (x_i, Y_i^b). Repeating this process B times yields B bootstrap MLEs. We then infer the inflation and std.dev. of the MLE from the bootstrap MLE.

We summarize the procedure in Algorithm 1 and discuss how to compute confidence intervals using the bootstrap MLE in Section 3.2. We evaluate our method through simulated examples in Section 4.

Algorithm 1: Resized bootstrap procedure.

Input: Observed data (x_i, y_i), $1 \le i \le n$, and a GLM formula.

Compute resized coefficients $\beta_\star$;

for $b = 1, \ldots, B$ **do**

  Simulate Y_i^b given x_i using $\beta_\star$ as model coefficients;

  Fit a GLM for (x_i, Y_i^b) to obtain the bootstrap MLE $\hat{\beta}^b$;

end

Estimate the standard deviation of the MLE $\hat{\sigma}_j$ (See Eqn. (3.7));

Estimate a common factor $\hat{\alpha}$ by regressing $\bar{\beta}$ onto $\beta_\star$ with weights proportional to $1/\hat{\sigma}_j^2$;

Output: $\hat{\alpha}$ and $\hat{\sigma}_j$

3.1. Estimating the signal strength

Since we would like to have $\mathrm{Var}({X_{\mathrm{new}}}^\top \beta_\star) = \gamma^2$, we discuss how to estimate γ from observed data (see Alg. 2 for a summary). We begin by reviewing the existing ProbeFrontier method, which applies to Gaussian covariates, and then introduce a new approach applicable to general covariate distributions.

The ProbeFrontier method (Sur and Candès, 2019) estimates γ by using the phase transition curve $\kappa(\beta_0, \gamma)$: if the intercept equals β_0 and the signal strength equals γ, then the MLE does not exist almost surely (asymptotically) if $\kappa > \kappa(\beta_0, \gamma)$; that is, the cases and controls can be perfectly separated by a hyperplane (see Sec. 2). The ProbeFrontier method identifies the threshold $\hat{\kappa}_s$ at which the MLE ceases to exist by subsampling observations. It then estimates $\hat{\gamma}$ in such a way that $\kappa(\beta_0, \hat{\gamma}) = \hat{\kappa}_s$ holds. While the ProbeFrontier method accurately estimates γ when the covariates are Gaussian, it does not apply here because the phase transition curve actually depends on the covariate distribution. For example, if the covariates are from a multivariate t-distribution, then the phase transition curve depends on the degrees of freedom of the t-distribution (Tang and Ye, 2020).

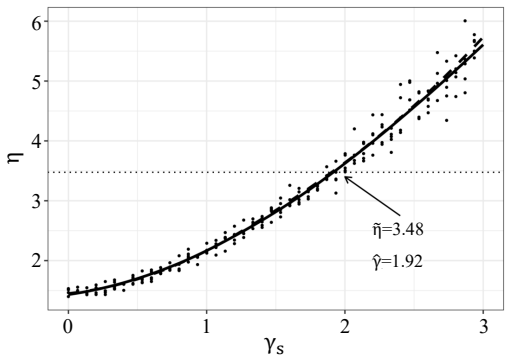


Figure 3. An illustration of using $\eta = \mathrm{sd}({X_{\mathrm{new}}}^\top \hat{\beta})$ to estimate the signal strength $\gamma = \mathrm{sd}({X_{\mathrm{new}}}^\top \beta)$. The dashed curve shows η versus γ. This is obtained by generating 100 random samples for each γ, the dashed curve being the smoothed LOESS fit. The solid curve shows an estimated curve using one dataset only; it is a smoothed version of $\hat{\eta}(\gamma)$ (black points). The dotted line shows $\tilde{\eta}$, and the estimated $\hat{\gamma} = 1.92$ approximates $\gamma = 2$ well. Here, we sample covariates from a multivariate t-distribution and responses from a logistic model. The coefficients β are sampled once and then re-scaled to achieve a value of γ shown on the x-axis.

As an alternative, we estimate γ by using a one-to-one relation between

$$\gamma^2 = \mathrm{Var}(X_{\mathrm{new}}^\top \beta) \quad \text{and} \quad \eta^2 = \mathrm{Var}(X_{\mathrm{new}}^\top \hat{\beta}). \tag{3.1}$$

The dashed curve in Figure 3 plots η as γ varies, and we observe that $\eta(\gamma)$ increases monotonically when γ increases. (Once again, this is because both the bias and the variance of the MLE increase as γ increases (Sur and Candès, 2019, Fig. 7).) Since the MLE does not exist when γ exceeds the phase transition threshold γ_s, which satisfies $\kappa(\beta_0, \gamma_s) = \kappa$, we expect that η would increase to infinity as γ approaches the threshold.

The one-to-one relation between γ and $\eta(\gamma)$ suggests that, if $\mathrm{Var}(X^\top \beta_\star) \cong \mathrm{Var}(X^\top \beta)$, then $\mathrm{Var}(X^\top \hat{\beta}_\star) \cong \mathrm{Var}(X^\top \hat{\beta})$, where $\hat{\beta}_\star$ denotes the MLE when the true coefficient is $\beta_\star$. Thus, we estimate γ^2 by $\mathrm{Var}(X^\top \beta_\star)$, where $\beta_\star$ obeys

$$\mathrm{Var}(X_{\mathrm{new}}^\top \hat{\beta}_\star) = \eta^2. \tag{3.2}$$

In this paper, we set $\beta_\star$ to be a rescaled version of the MLE, i.e., $\beta_\star = s \times \hat{\beta}$. Because the MLE is biased upwards in absolute magnitude, the rescaling factor s is less than one and shrinks the MLE towards zero.

Although we cannot compute η directly because it is evaluated at a new observation X_{new}, we estimate η by using the SLOE estimator introduced in Yadlowsky et al. (2021). We briefly describe SLOE here, and defer detailed formulae to SM Section 2. The SLOE estimator proceeds in two steps. First, it approximates $\mathrm{Var}(X_{\mathrm{new}}^\top \hat{\beta})$ by the variance of $x_i^\top \hat{\beta}_{(i)}$ where $\hat{\beta}_{(i)}$ is the leave-ith-

observation-out MLE. Second, instead of re-evaluating $\hat{\beta}_{(i)}$ for each observation, SLOE uses the first-order approximation of the score equation to approximate $\hat{\beta}_{(i)}$ from the MLE. The theory is this: Yadlowsky et al. (2021) proves that the SLOE estimator is consistent in logistic regression models with Gaussian covariates. Furthermore, we expect that SLOE yields reliable estimates for a broad class of covariates, for which the Euclidean norm $\|X\|$ is concentrated and the Hessian at the MLE is positive definite.

Now that we are able to approximate $\eta(\gamma)$ at a given γ, we estimate $\eta = \mathrm{Var}(X^\top \hat{\beta})^{1/2}$ and denote it as $\tilde{\eta}$. Next, we estimate the curve $\eta(t)$ at a sequence of signal strengths t, from which we estimate γ by setting $\hat{\gamma}$ such that $\tilde{\eta} = \hat{\eta}(\hat{\gamma})$. To implement this, we pick a sequence of scaling factors $\{0 = s_1, \ldots, s_L = 1\}$. At each s_l, we set the coefficients to be $\beta^{s_l} = s_l \times \hat{\beta}$ and the signal strength corresponding to s_i as $\gamma(s_l) = \mathrm{sd}(\boldsymbol{X}\beta^{s_l})$, where $\boldsymbol{X}$ refers to the observed covariate matrix. We use β^{s_l} as the true coefficient to generate new responses (as in a parametric bootstrap) and then use this sample to obtain one estimate of $\hat{\eta}(\gamma(s_l))$. Repeating the process J times yields J estimates $\hat{\eta}_j(\gamma(s_l))$ for every s_l. We next fit a smoothed curve $\hat{\eta}(\gamma(s_l))$ through the points $\hat{\eta}_j(\gamma(s_l))$, $l = 1, \ldots, L$, $j = 1, \ldots, J$. Finally, we set $\hat{\gamma}$ such that $\hat{\eta}(\hat{\gamma}) = \tilde{\eta}$.

We demonstrate our method in Figure 3, which shows $\hat{\eta}(t)$ estimated from a single dataset. The estimated curve offers an excellent fit across all values of γ. In this example, the estimated $\tilde{\eta} = 3.48$ (dotted horizontal line), and this corresponds to $\hat{\gamma} = 1.92$ on the solid curve. This estimate is close to the actual signal strength set to $\gamma = 2$.

Algorithm 2: Estimating signal strength

Input: Observed data (x_i, y_i), $1 \le i \le n$, and a GLM formula.
Estimate $\tilde{\eta} = \mathrm{Var}(X_{\mathrm{new}}^\top \hat{\beta})$ via leave-one-out techniques;
Pick a sequence $\{0 = s_1, \ldots, s_L = 1\}$;
for $l = 1, \ldots L$ **do**
    Set $\beta^{s_l} = s_l \times \hat{\beta}$ and $\gamma_l = \mathrm{sd}(\boldsymbol{X}\beta^{s_l})$;
    for $j = 1, \ldots, J$ **do**
        Simulate $Y_{l,i}^j$ given x_i using β^{s_l} as model coefficients for each observation $i = 1, \ldots, n$;
        Fit a GLM for $(x_i, Y_{l,i}^j)$ to estimate $\hat{\eta}_j(\gamma(s_l))$;
    end
end
Fit a smooth curve $\hat{\eta}(\gamma)$;
Estimate $\hat{\gamma}$ by solving $\hat{\eta}(\hat{\gamma}) = \tilde{\eta}$;
Output: Estimated $\hat{\gamma}$.

3.2. Constructing confidence intervals

We consider two ways of computing confidence intervals (CI) from bootstrapped MLEs: first, assuming that the MLE is approximately Gaussian, i.e.,

$$\frac{\hat{\beta}_j - \alpha_j \beta_j}{\sigma_j} \approx \mathcal{N}(0,1), \tag{3.3}$$

where α_j and σ_j denote the bias and standard deviation, inverting Eqn. (3.3) yields the following $(1-q)$ CI for β_j:

$$\left[\frac{1}{\hat{\alpha}_j}\left(\hat{\beta}_j - z_{1-q/2}\,\hat{\sigma}_j\right), \frac{1}{\hat{\alpha}_j}\left(\hat{\beta}_j - z_{q/2}\,\hat{\sigma}_j\right)\right]. \tag{3.4}$$

Here, z_q is the quantile of a standard Gaussian, while $\hat{\alpha}_j$ and $\hat{\sigma}_j$ refer to estimates of α_j and σ_j.

When the normal approximation is inadequate, we use the approximation

$$\frac{\hat{\beta}_j - \alpha_j \beta_j}{\sigma_j} \overset{d}{\approx} \frac{\hat{\beta}_j^b - \hat{\alpha}_j \beta_{\star,j}}{\hat{\sigma}_j}, \tag{3.5}$$

where the right-hand side refers to the distribution of $\hat{\beta}_j^b$ conditional on the observed covariates. We obtain a $(1-q)$ CI as

$$\left[\frac{1}{\hat{\alpha}_j}\left\{\hat{\beta}_j - t_j^b\left(1-\frac{q}{2}\right)\hat{\sigma}_j\right\}, \frac{1}{\hat{\alpha}_j}\left\{\hat{\beta}_j - t_j^b\left(\frac{q}{2}\right)\hat{\sigma}_j\right\}\right], \tag{3.6}$$

where $t_j^b[q]$ denotes the quantile of the right-hand side of (3.5). We refer to the confidence interval in (3.6) as the "bootstrap-t" confidence interval, and examine the approximation (3.5) in Section 4.2.

Finally, we describe how to estimate the bias α_j and the standard deviation σ_j. To estimate σ_j, we use the standard deviation of the bootstrap MLE, i.e.,

$$\hat{\sigma}_j^2 = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_j^b - \bar{\beta}_j)^2, \quad \text{where} \quad \bar{\beta}_j = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_j^b. \tag{3.7}$$

We estimate α_j by weighted regression: that is, we regress $\bar{\beta}^b$ onto $\beta_\star$ by assigning to each MLE coordinate a weight inversely proportional to its estimated variance $\hat{\sigma}_j^2$. We assume a common bias factor because all the α_j's are equal when the covariates are multivariate Gaussian. In practice, we can plot $\bar{\beta}_j^b$ versus $\beta_{\star,j}$: if bias factors are all equal, then the points should align on a line, which we observe in all our simulations (Fig. 4).

3.3. When is the resized bootstrap adequate?

When the covariates are multivariate Gaussian, Zhao, Sur and Candès (2020) observed that while Eqn. (1.1) is accurate when β_j is moderately large (assuming the covariates X_j are standardized to have zero mean and unit variance), the std.dev. of $\hat{\beta}_j$ increases as the absolute magnitude of β_j increases. This result implies that the resized coefficient $\beta_{\star,j}$ should be close to β_j in order to correctly estimate the MLE distribution. However, the resized coefficients only satisfy $\mathrm{Var}(X^\top \beta_\star) = \gamma^2$, and yet $\beta_{\star_j} \neq \beta_j$ in general. Therefore, we expect that the CIs to be approximately correct when β_j is moderately large, but inaccurate when β_j is large. We explore the performance of our method when the model coefficients are large in SM Section 5. While we expect that correct inference can be obtained by shrinking the large and small coefficients separately, we leave this study for future research.

4. Numerical Studies

We now study the accuracy of the proposed resized bootstrap method by simulating GLMs with non-Gaussian covariates. In this section, we consider an example of logistic regressions. Results in other settings (with various levels of signal strength, problem dimensions and class imbalance) and with other types of GLM (including Probit and Poisson regressions) are reported in Sections 3.2–3.4 of the Supplementary Material (SM). We also consider an example where the sample size is small ($n = 400$) in SM Section 4. Lastly, we study the situation when the M-estimator is obtained by minimizing a general loss function that may not be the negative log-likelihood in SM Section 6. `R` code used for these simulations is publicly available at `https://github.com/zq00/glmboot`. The `R` package `glmhd` (`https://github.com/zq00/glmhd`) implements the resized bootstrap method and provides tutorials.

4.1. Simulation design

First, we set $n = 4000$ and $p = 400$ ($\kappa = p/n = 0.1$). Without specifying, we sample covariates from a multivariate t-distribution (MVT) with $\nu = 8$ degrees of freedom whose covariance matrix Σ is a circulant matrix equal to $\Sigma_{ij} = 0.5^{\min(|i-j|, p-|i-j|)}$. This structure implies that the Σ_{ii}^{-1}'s are all equal. (If the covariates were Gaussian, then the variance of a predictor conditional on the others is the same regardless of the predictor. In turn, HDT then predicts that in this case all the MLE coefficients have equal standard deviation.)

After sampling the covariates, we sample responses from a logistic model. We sample model coefficients by first picking 50 non-null variables; then, we sample the magnitude of the non-null coefficients from an equal mixture of $\mathcal{N}(5, 1)$ and $\mathcal{N}(-5, 1)$. This signal strength ensures that the MLE exists. At the same time, the magnitude of the coefficient is sufficiently large so that we can tell a large

proportion the the non-null variables apart from the nulls. For instance, when $\beta_j = 4.78$ as in the example in Section 1.2 , over 90% of the 95% CI excludes 0, and approximately 90% of the non-null coefficients from the mixture distribution satisfy this property.

4.2. Results

We report below the estimated inflation and standard deviation of the MLE as well as the coverage proportions. We also examine the MLE distribution and the assumption that the bias factors α_j are all equal.

4.2.1. Estimated inflation and variance

From Section 1.1 we know that the MLE is just too sure in the sense that the estimated magnitude is biased upwards. As an illustration, Figure 4 plots the average MLE versus the model coefficients when the covariates are from (modified) ARCH model (see SM Sec. 3.2). Since the scatterplot lies near a line, we can see that the α_j's do not seem to much depend on the magnitude of the coefficients; additionally, the plot confirms the bias of the MLE since the line has a slope greater than 1. For information, we get a very similar plot for the multivariate t-covariates.

We now examine the accuracy of the estimated inflation using existing high-dimensional theory and the resized bootstrap (recall that both estimate a common bias factor). Table 1 reports the estimated inflation and variance of a single null and a single non-null variable. As observed in Section 1.2, HDT captures the bias, and Table 1 shows that the resized bootstrap estimate is also reasonably accurate. As to the standard deviation, while both methods slightly underestimate the std.dev., the resized bootstrap is more accurate and its relative error is less than 1%. In particular, the resized bootstrap captures the increased std.dev. of the MLE of non-null variables in comparison to null variables. In contrast, classical calculations based on the Fisher information significantly underestimate the std.dev.. Since the resized bootstrap yields a more accurate std.dev., we would expect enhanced CIs.

4.2.2. Coverage proportion

Section 3.2 introduced two types of CIs, based on the assumptions that the MLE is approximately Gaussian (3.3) or that the standardized bootstrap MLE approximates the distribution of the standardized MLE (3.5). Before evaluating accuracy, we examine these assumptions by showing a normal Q-Q plot of the MLE (Fig. 5(a)) and a Q-Q plot of the standardized bootstrap MLE versus the standardized MLE (Fig. 5(b)). Here, we standardize the bootstrap MLE by the estimated inflation and estimated std.dev. and the MLE by the correct bias and std.dev. Along the points align on the 45 degree line in both plots, we conclude that both assumptions are reasonable and, therefore, expect that both CIs would

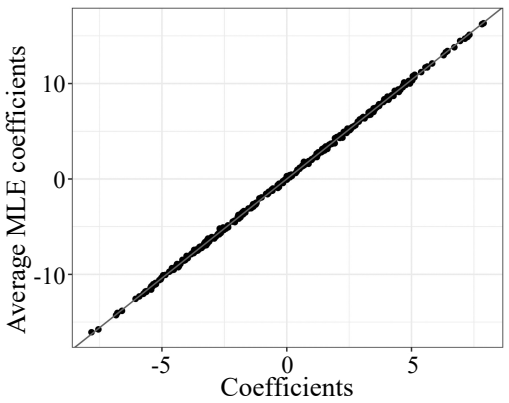


Figure 4. Average MLE versus model coefficients for the non-null variables. The x-axis shows the magnitude of each non-null coefficient and the y-axis shows the average MLE over 832 repetitions. The gray line has zero intercept and slope equal to 2.08. In this example, the covariates are sampled from the modified ARCH model described in SM Section 3.2.

Table 1. Estimated inflation and std.dev. of the MLE. The correct values (empirical bias and std.dev.) have been obtained from 10,000 repetitions. The std.dev. from classical theory is calculated by the `glm` function in `R` and averaged over 10,000 repetitions. The resized bootstrap estimates are computed by taking an average over 1,000 repetitions and uses an estimated signal strength γ. We highlight the number closest to the empirical observation in bold.

	Inflation			Standard Deviation			
	High-dim Theory	Resized Bootstrap	Empirical Bias	Classical Theory	High-dim Theory	Resized Bootstrap	Empirical Std.dev.
$\beta = 0$	-	-	-	1.232	1.259	**1.316**	1.327
$\beta = 5.519$	1.151	**1.159**	1.160	1.244	1.259	**1.327**	1.337

perform well.

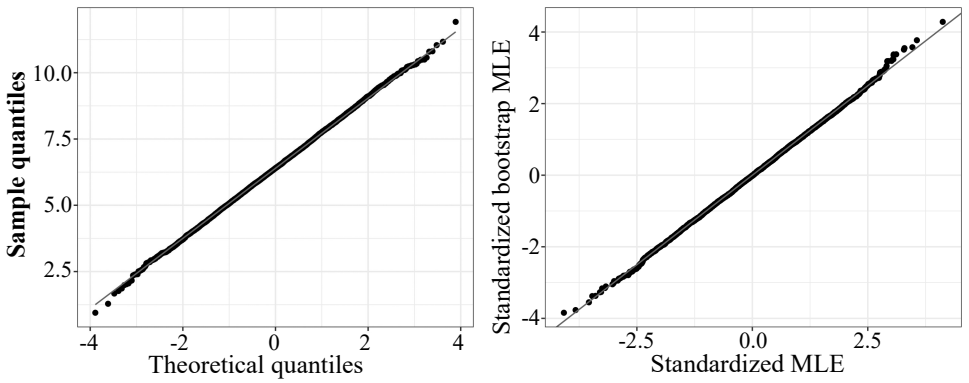
Denote the confidence interval for β_j in the ith simulation as $\mathrm{CI}_{i,j}$, and define the proportion of times a single variable β_j is covered as

$$q_j := \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\{\beta_j \in \mathrm{CI}_{i,j}\}. \tag{4.1}$$

Define the coverage proportion of *all of the variables in the ith experiment only* as

$$\bar{q}_i = \frac{1}{p}\sum_{j=1}^{p} \mathbb{I}\{\beta_j \in \mathrm{CI}_{i,j}\} \tag{4.2}$$

We report both coverage of a single non-null coefficient q_j and the proportion of variables covered in a *single-shot experiment* $\bar{q} = \sum_{i=1}^{N} \bar{q}_i/N$ in Tables 2 and 3 respectively (we report the coverage proportion q_j for a single null variable in



(a) (b)

Figure 5. (a) Normal Q-Q plot of the MLE. (b) Q-Q plot of the standardized bootstrap MLE (in one simulated example) versus the standardized MLE. In this example, the covariates are sampled from a multivariate t-distribution.

Table 2. Coverage proportion of a single *non-null* variable (q_j in Eqn. (4.1)) with standard deviation between parentheses. This example uses multivariate-t covariates. We highlight the number closest to the empirical observation in bold.

Nominal coverage	Theoretical CI		Standard Bootstrap		Resized Bootstrap			
					Known γ		Estimated γ	
	Classical	High-Dim	Parametric	Pairs	Boot-g	Boot-t	Boot-g	Boot-t
95	87.3	93.5	71.1	76.3	93.6	93.9	**94.2**	**94.4**
	(0.3)	(0.3)	(1.6)	(1.3)	(0.7)	(0.7)	(0.8)	(0.8)
90	79.4	87.9	61.2	66.6	88.5	88.7	88.6	**89.1**
	(0.3)	(0.3)	(1.7)	(1.4)	(1.0)	(1.0)	(1.1)	(1.1)
80	67.4	77.2	46.8	52.7	79.5	79.6	80.8	**80.0**
	(0.5)	(0.4)	(1.7)	(1.5)	(1.2)	(1.2)	(1.3)	(1.4)

SM Sec. 3.1). Both the Gaussian approximation (Boot-g) and bootstrap MLE distribution (Boot-t) are used to compute the CIs. The two CIs not only differ in their formulae, but also in the number of bootstrap samples: we use $B = 10000$ bootstrap samples to compute the boot-t CI, but only $B = 100$ bootstrap samples to compute the boot-g CI. This is because boot-g CI requires only estimates of the bias and variance, while boot-t CI requires an estimate of the entire distribution.

While the resized bootstrap slightly undercovers a single coefficient (Tbl. 2), the relative error is within 2% in all of the levels we examined. Similarly, the proportion of variables covered in a single-shot experiment (Tbl. 3) is also close to the nominal coverage and the relative error is within 1%. In addition, boot-g and boot-t CI achieve similar accuracy at every level we examined. Since boot-g CI uses a smaller sample size, we prefer boot-g CI when the Gaussian assumption holds. We can verify the normality assumption by comparing the quantiles of bootstrap MLEs with normal quantiles. Table 2 shows the coverage of a non-null

Table 3. The proportion of covered variables in a single-shot experiment ($\bar{q}$ in Eqn. (4.2)). The standard deviation is given between parentheses.

Nominal coverage	Theoretical CI		Standard Bootstrap		Resized Bootstrap			
					Known γ		Estimated γ	
	Classical	High-Dim	Parametric	Pairs	Boot-g	Boot-t	Boot-g	Boot-t
95	92.5	93.7	90.8	93.3	94.6	94.9	94.7	**95.0**
	(0.02)	(0.02)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
90	86.6	88.2	84.5	87.8	89.5	89.7	89.7	**89.9**
	(0.02)	(0.02)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
80	75.7	77.7	73.6	77.5	79.4	79.5	79.6	**79.7**
	(0.03)	(0.03)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)

variable, and we report coverage of a null variable in the supplement. Comparing the coverage probability using the estimated signal strength $\hat{\gamma}$ versus its true value γ shows that the method with estimated parameters perform as well as if we had an oracle.

As to the other methods, the HDT CIs slightly undercover since variability is underestimated as seen earlier. Classical CIs significantly undercover. Neither the parametric nor the pairs bootstrap provide the correct coverage, and this is consistent with observations from Figure 1.

5. Application to a Real Data Set

Having observed that the resized bootstrap procedure provides more accurate inference compared to classical and high-dimensional theory, we now analyze a real data set. In this study by Lim, Jun and Lee (2019), researchers aim to understand which factors are associated with restrictive spirometry pattern (RSP), which is a lung condition. In particular, they hypothesize that glomerular hyperfiltration (GHF), which assesses the kidney function, may be associated with the risk of RSP. To evaluate their hypothesis, they collected participants data from from the Korea National Health and Nutrition Examination Survey (KNHANES) from 2009-2015. They performed a logistic regression, where the response variable is RSP (defined as FVC $< 80\%$ AND FEV1/FVC ≥ 0.7) and the covariates include demographic variables, medical history, medications used, and a variety of health-related variables.

For the purpose of illustrating our approach, we fit a logistic regression using subsamples of sample size $n = 200$ and include $p = 18$ covariates including the intercept ($\kappa = 18/200 = 0.09$). We only include binary variables such that both positive and negative classes occur in at least 5% of all the samples. We examine whether the confidence intervals of the model coefficients, i.e., the log odds ratios, cover the "true" coefficients, which we estimate by the logistic MLE using the full data that contains about 22,000 observations. Figure 6 shows the CI for each covariate using classical theory (solid line), resized bootstrap (dashed line),

Figure 6. Confidence interval for each variable using classical theory (solid line) and the resized bootstrap (dashed line). The black points indicate true model coefficients, estimated using the full data set. While we include demographic variables in the logistic model, we do not present their fitted coefficients as in Table 2 of the paper.

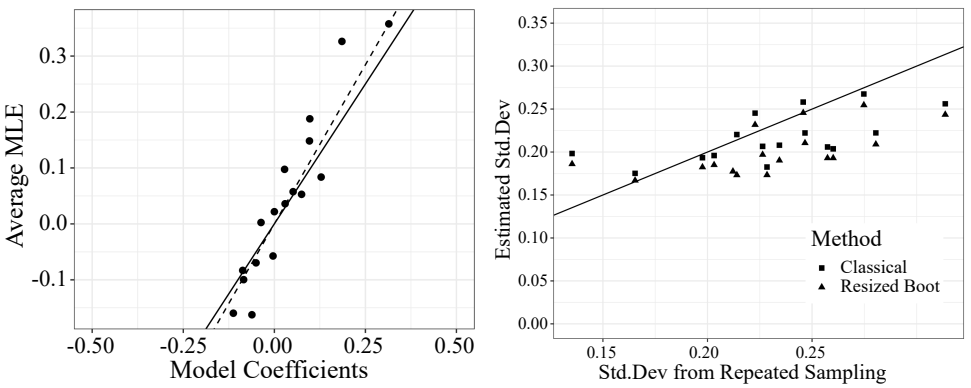
and the estimated coefficient using the full data (black points). Because the estimated γ is random, we repeat 10 times and use the average as the estimated signal strength. The resized bootstrap CI is closer to zero compared to CI using the classical theory, and is slightly more accurate. For instance, the coefficient for waist circumference is covered by the dashed line segment, but is not covered by the solid line segment.

Then, we generate $B = 24$ disjoint subsamples of sample size $n = 200$ and compare classical theory and the resized bootstrap based on the estimated inflation, std.dev., and the coverage proportion of CIs. First, we examine the bias of the MLE by plotting the average of the logistic MLE estimated using each subsample versus the true coefficients (Fig. 7(a)). While the average MLEs are scattered across, their absolute magnitude is slightly larger than the true coefficients. The resized bootstrap yields an estimate $\hat{\alpha}_b = 1.14$ (dashed). Though this is a small adjustment, it allows the resized bootstrap to produce more accurate CI as observed in Figure 6.

Next, we plot the average estimated std.dev. versus the empirical std.dev. in Figure 7(b) calculated across batches. The resized bootstrap and the classical estimates are similar, and both methods tend to underestimate the true standard deviation. In Table 4, we evaluate the proportion of variables covered in each batch as well as the coverage probability of the variable "systolic blood pressure". Since both methods under-estimates the std.dev., we expect that the bootstrap provides some improvement in coverage, but does not yield correct coverage either, and this is indeed what we observe in the table. In this example, we use the large sample coefficient as a proxy for the true model coefficients, and our

Table 4. Coverage probability of confidence intervals (the coverage standard deviation is between parentheses). The first columns report the coverage proportion for the variable "systolic blood pressure". The next two columns compute the proportion of variables covered in each batch and report the average over 24 batches.

Nominal	I. Single variable		II. Single experiment	
Coverage	Classical	Resized Bootstrap	Classical	Resized Bootstrap
95	87.5 (6.9)	91.7 (6.0)	92.2 (1.3)	94.9 (1.1)
90	87.5 (6.9)	87.5 (7.2)	85.8 (1.6)	88.2 (1.4)
80	83.3 (7.8)	83.3 (8.1)	72.3 (1.9)	75.3 (1.7)



(a) (b)

Figure 7. Bias and std.dev. of the MLE. (a) Average MLE for the variables versus true coefficients. The black points show the average MLE averaged over $B = 24$ batches. The dashed line shows the resized bootstrap estimate of the bias factor ($\hat{\alpha}_b = 1.14$). (b) Average estimated standard deviation of the MLE for each variable versus standard deviation across batches. The square and triangular points respectively use classical theory and the resized bootstrap. In both plots, the solid line is the 45 degree line.

results suggest that when the sample size is small, while the resized bootstrap may not yield accurate coverage, it may perform better than the classical theory.

6. Discussion

In this paper, we demonstrated that the distribution of the MLE in large logistic regression models depends on the distribution of the covariates and that bootrstrap methods fail to approximate this distribution. This is in line with previous findings concerned with linear regression (El Karoui, 2018; El Karoui and Purdom, 2018). To fix this problem, we introduced a resized bootstrap, which correctly adjusts inference. The key is to resample from a parametric distribution obtained by shrinking the MLE towards zero in a data-dependent fashion, where the amount of shrinkage is informed by insights from HDT. Resized bootstrap CIs yield correct coverage proportions for different types of covariate distributions

and types of GLMs. Our findings echo previous results in El Karoui and Purdom (2018) and Lopes, Blandino and Aue (2019); combining HDT with bootstrap resampling methods can provide improved estimates.

We conclude with several future research questions. First, while the resized bootstrap procedure provides a high-quality approximation to the MLE distribution, it slightly underestimates the standard deviation. Therefore, future research on the theoretical accuracy of the procedure might lead to improvements in the design of the resized MLE, for example, by adjusting the coefficients to not only match the standard deviation of the linear predictor, but also a few higher moments. Second, one drawback of the resized bootstrap is its relatively high computational cost: we need to compute the MLE many times to estimate γ and the MLE distribution. Although a few hundred bootstrap samples suffice to yield accurate CIs when the MLE is approximately Gaussian, being able to reduce the computational cost would make it even more suitable for larger datasets. Third, as mentioned in Section 3.3, the resized bootstrap is expected to accurately estimate the distribution of the MLE for coefficients with moderate magnitudes. While the resized bootstrap is reasonably accurate for relatively large β_j (see Supplementary Material), novel insights might further enhance it.

Supplementary Material

Additional materials contain the following: (1) a conjecture about the MLE distribution when the covariates follow a multivariate t-distribution; (2) a description of the SLOE estimator; (3) additional logistic regression examples; (4) simulated examples for Probit and Poisson regressions; (5) a simulated example when the sample size is small; (6) simulations when the coefficients are sparse; (7) application of the resized bootstrap method to the case when the M-estimator minimizes a general function.

Acknowkedgments

E. C. was supported by the Office of Naval Research grant N00014-20-1-2157, the National Science Foundation grant DMS-2032014, the Simons Foundation under award 814641, and the ARO grant 2003514594.

References

Bellec, P., Shen, Y. and Zhang, C. (2022). Asymptotic normality of robust M-estimators with convex penalty. *Electron. J. Stat.* **16**, 5591–5622.

Beran, R. and Srivastava, M. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13**, 95–115.

Bickel, P. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.

Bickel, P. and Freedman, D. (1982). Bootstrapping regression models with many parameters. Technical report no. 7. University of California, Berkley.

Celentano, M., Montanari, A. and Wei, Y. (2023). The Lasso with general Gaussian designs with applications to hypothesis testing. *Ann. Statist.* **51**, 2194–2220.

Diciccio, T. and Romano, J. (1988). A review of bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50**, 338–354.

Donoho, D. and Montanari, A. (2016). High-dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166**, 935–969.

Eaton, M. and Tyler, D. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19**, 260–271.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–26.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76**, 312–319.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**, 45–58.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Efron, B., Halloran, E. and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93**, 13429–13434.

El Karoui, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. *arXiv: 1311.2445*.

El Karoui, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170**, 95–175.

El Karoui, N., Bean, D., Bickel, P., Chinghway Limb and Yu, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110**, 14557–14562.

El Karoui, N. and Purdom, E. (2016). The bootstrap, covariance matrices and PCA in moderate and high-dimensions. *arXiv: 1608.00948*.

El Karoui, N. and Purdom, E. (2018). Can we trust the bootstrap in high-dimensions? The case of linear models. *J. Mach. Learn. Res.* **19**, 1–66.

Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20**, 121–145.

Gine, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18**, 851–869.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

Lim, H. I., Jun, S. J. and Lee, S. W. (2019). Glomerular hyperfiltration may be a novel risk factor of restrictive spirometry pattern: Analysis of the Korea National Health and Nutrition Examination Survey (KNHANES) 2009-2015. *PLoS One* **9**, 2009–2015.

Lopes, M., Blandino, A. and Aue, A. (2019). Bootstrapping spectral statistics in high dimensions. *Biometrika* **106**, 781–801.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255–285.

Mendelson, A., Zuluaga, M., Hutton, B. and Ourselin, S. (2016). What is the distribution of the number of unique original items in a bootstrap sample? *arXiv: 1602.05822*.

Politis, D., Romano, J. and Wolf, M. (1999). *Subsampling*. Springer.

Shao, J. (1996). Bootstrap model selection. *J. Amer. Statist. Assoc.* **91**, 655–665.

Shorack, G. (1982). Bootstrapping robust regression. *Comm. Statist. Theory Methods* **11**, 961–972.

Sur, P., and Candès and J., E. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116**, 14516–14525.

Tang, W. and Ye, Y. (2020). The existence of maximum likelihood estimate in high-dimensional binary response generalized linear models. *Electron. J. Stat.* **14**, 4028 – 4053.

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42**, 1166–1202.

Yadlowsky, S., Yun, T., McLean, C. and D'Amour, A. (2021). SLOE: A faster method for statistical inference in high-dimensional logistic regression. *Advances in Neural Information Processing Systems* **34 (NeurIPS 2021)**.

Zhang, C. and Zhang, S. (2014). Confidence intervals for low-dimensional parameters in high-dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**, 217–242.

Zhao, Q., Sur, P. and Candès, E. (2020). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28**, 1835–1861.

(Received August 2022; accepted April 2023)