

## ASYMPTOTIC PROPERTIES OF THE RESIDUAL BOOTSTRAP FOR LASSO ESTIMATORS

A. CHATTERJEE AND S. N. LAHIRI

(Communicated by Edward C. Waymire)

ABSTRACT. In this article, we derive the asymptotic distribution of the bootstrapped Lasso estimator of the regression parameter in a multiple linear regression model. It is shown that under some mild regularity conditions on the design vectors and the regularization parameter, the bootstrap approximation converges weakly to a random measure. The convergence result rigorously establishes a previously known heuristic formula for the limit distribution of the bootstrapped Lasso estimator. It is also shown that when one or more components of the regression parameter vector are zero, the bootstrap may fail to be consistent.

### 1. INTRODUCTION

Consider the following regression model:

$$(1.1) \quad y_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n,$$

where  $y_i$  is the response,  $\mathbf{x}'_i = (x_{i,1}, \dots, x_{i,p})$  is a  $p \times 1$  covariate vector,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is the regression parameter and  $\{\epsilon_i\}$  are *iid* errors. The Lasso estimator of  $\boldsymbol{\beta}$  is defined as the minimizer of the  $l_1$ -penalized least square criterion function,

$$(1.2) \quad \hat{\boldsymbol{\beta}}_n := \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^p} \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{u})^2 + \lambda_n \sum_{j=1}^p |u_j|,$$

where  $\lambda_n$  is a regularization parameter. The Lasso estimate was introduced by [Tib96] as an estimation and variable selection method. The Lasso is preferred over the ordinary least squares estimation method (which corresponds to  $\lambda_n = 0$  in (1.2)) as a model selection tool because, with a suitable choice of  $\lambda_n$ , it leads to a sparse solution. Many authors have studied the model-consistency properties of the Lasso and investigated conditions under which the Lasso can recover the true sparsity pattern; see [ZY06], [Wai06], [Zou06] and the references therein.

---

Received by the editors January 22, 2009 and, in revised form, December 23, 2009 and March 2, 2010.

2010 *Mathematics Subject Classification*. Primary 62J07; Secondary 62G09, 62E20.

*Key words and phrases*. Consistency, bootstrap, penalized regression, random measure.

This research was partially supported by NSF grant DMS-0707139.

In an important paper, [KF00] derived the asymptotic distribution of the Lasso estimator. They showed that under mild regularity conditions on the error variables, the design vectors  $\mathbf{x}_i$ , and the regularization parameter  $\lambda_n$ ,

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \underset{\mathbf{u} \in \mathbb{R}^p}{\operatorname{argmin}} V(\mathbf{u}) \quad \text{as } n \rightarrow \infty,$$

where  $V(\cdot)$  is a random process over  $\mathbb{R}^p$ . Except for some very special cases, no closed form formula for either the limiting random vector or the limit distribution is available. As a result, the use of the asymptotic distribution of the Lasso estimator for constructing confidence intervals or for conducting large sample tests is not very appealing in practice. As an alternative, [KF00] considered using the bootstrap method to approximate the distribution of the Lasso estimator. In the setup of (1.1), the standard bootstrap method is the residual bootstrap of [Fre81]. [KF00] considered the residual bootstrap for the Lasso estimator and sketched out its asymptotic behavior using some heuristic arguments. In this paper, we derive the asymptotic distribution of the bootstrapped Lasso estimator and show that it is given by a random probability measure on  $\mathbb{R}^p$ . We provide a precise description of the random limit distribution. From this, it follows that under the regularity conditions of this paper, the heuristic formula for the limit distribution of the bootstrapped Lasso estimator, given by [KF00], is indeed valid, thereby proving the conjecture by [KF00]. See Section 3 for details.

As a corollary to the main result (cf. Theorem 3.1), we show that under mild regularity conditions, the bootstrap approximation is inconsistent whenever there are one or more zero-components of the parameter vector  $\boldsymbol{\beta}$ , thereby corroborating the claim of [KF00]. We also use the asymptotic results to identify some important special cases where the limiting random measure is degenerate and the bootstrap approximation is consistent. In the course of the proof of the main result, we also establish strong consistency of the Lasso estimator (cf. Lemma 4.2), which may be of some independent interest.

The rest of the paper is organized as follows. In Section 2, we describe the bootstrap method. In Section 3, we state the main result and the corollaries. The proofs of all the results are given in Section 4.

## 2. DESCRIPTION OF THE BOOTSTRAP METHOD

In a regression setup such as (1.1), the standard approach to bootstrapping is the *residual bootstrap* method (cf. [Fre81]), which we briefly describe now. Letting  $\hat{\boldsymbol{\beta}}_n$  denote the Lasso estimator of  $\boldsymbol{\beta}$  given by (1.2), our goal is to approximate the distribution of a centered and scaled version of  $\hat{\boldsymbol{\beta}}_n$ ,

$$\mathbf{T}_n = \sqrt{n} \left( \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right), \quad n \geq 1,$$

using the residual bootstrap. To that end, define the residuals

$$e_i = y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_n, \quad i = 1, \dots, n.$$

Consider the set of centered residuals  $\{e_1 - \bar{e}_n, \dots, e_n - \bar{e}_n\}$ , where  $\bar{e}_n = n^{-1} \sum_i e_i$ . For the residual bootstrap, one selects a random sample  $\{e_i^*\}_{i=1}^n$ , of size  $n$  from  $\{e_i - \bar{e}_n : i = 1, \dots, n\}$  with replacement. The bootstrap version of (1.1) is now given by

$$y_i^* = \mathbf{x}'_i \hat{\boldsymbol{\beta}}_n + e_i^*, \quad i = 1, \dots, n.$$

Next, based on the bootstrap dataset  $\{y_i^*, \mathbf{x}_i\}_{i=1}^n$ , as in [KF00], we define the bootstrap version of the Lasso estimator (cf. (1.2)) as

$$(2.1) \quad \beta_n^* := \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^p} \sum_{i=1}^n (y_i^* - \mathbf{x}_i' \mathbf{u})^2 + \lambda_n \sum_{j=1}^p |u_j|.$$

Then, the bootstrap version of  $\mathbf{T}_n$  is given by

$$\mathbf{T}_n^* = \sqrt{n} \left( \beta_n^* - \hat{\beta}_n \right), \quad n \geq 1.$$

Let  $\hat{G}_n(\cdot)$  denote the (conditional) distribution of  $\mathbf{T}_n^*$  given the observations  $\{y_i\}_{i=1}^n$ ; i.e.,

$$\hat{G}_n(B) = \mathbf{P}_* (\mathbf{T}_n^* \in B), \quad B \in \mathcal{B}(\mathbb{R}^p),$$

where  $\mathbf{P}_*$  denotes the bootstrap probability measure conditional on the observations  $\{y_i\}_{i=1}^n$ ,  $n \geq 1$  or, equivalently, conditional on the error variables  $\{\epsilon_i : i \geq 1\}$  and where  $\mathcal{B}(\mathbb{R}^p)$  denotes the Borel  $\sigma$ -field on  $\mathbb{R}^p$ . Then,  $\hat{G}_n(\cdot)$  is the residual bootstrap estimator of the unknown distribution  $G_n$  (say) of  $\mathbf{T}_n$ . The residual bootstrap estimator of a population parameter  $\theta_n = \varphi(G_n)$ , defined through a functional  $\varphi(\cdot)$  of  $G_n(\cdot)$ , is given by  $\varphi(\hat{G}_n)$ . For example, the bootstrap estimator of the  $\alpha$ -quantile ( $0 < \alpha < 1$ ) of the unknown distribution of  $\|\mathbf{T}_n\|$  is given by the  $\alpha$ -quantile of the conditional distribution of  $\|\mathbf{T}_n^*\|$  under  $\mathbf{P}_*$ ; The latter can be computed using the data and may be used for constructing a confidence region for  $\beta$ .

Next, let

$$(2.2) \quad V(\mathbf{u}) = -2\mathbf{u}'\mathbf{W} + \mathbf{u}'\mathbf{C}\mathbf{u} + \lambda_0 \sum_{j=1}^p \left[ u_j \operatorname{sgn}(\beta_j) \mathbf{1}(\beta_j \neq 0) + |u_j| \mathbf{1}(\beta_j = 0) \right],$$

where  $\mathbf{W} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{C})$ ,  $\mathbf{C} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'$  and where  $\lambda_0 \in [0, \infty)$ . Here and in the following, we write  $\operatorname{sgn}(x)$  to denote the sign of  $x \in \mathbb{R}$  (cf. (4.1)) and  $\mathbf{1}(\cdot)$  to denote the indicator function, where  $\mathbf{1}(S) = 0$  or  $1$ , according to whether the statement  $S$  is false or true. [KF00] show that under some regularity conditions,

$$\mathbf{T}_n \xrightarrow{d} \mathbf{T}_\infty,$$

where  $\mathbf{T}_\infty = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^p} V(\mathbf{u})$ . Write  $G_\infty(\cdot)$  to denote the distribution of the limiting random vector  $\mathbf{T}_\infty$ . For the bootstrap approximation to be useful, one would expect  $\hat{G}_n(\cdot)$  to be close to  $G_n(\cdot)$  or, equivalently, to be close to the limit distribution  $G_\infty(\cdot)$ , for large  $n$ . However, in the next section we show that this is *not* always the case. The bootstrap estimator  $\hat{G}_n(\cdot)$  may be inconsistent for  $G_\infty(\cdot)$ , and thus it may fail to provide a valid approximation to  $G_n(\cdot)$ .

### 3. MAIN RESULTS

As mentioned in the Introduction, [KF00] considered the residual based bootstrap for the Lasso estimator and sketched out some of its asymptotic properties to indicate that the residual bootstrap approximation may not work properly in this problem. In this section, we obtain a complete description of the limit behavior of the bootstrapped Lasso estimator  $\mathbf{T}_n^*$ , which, in particular, pinpoints the situations where the residual bootstrap fails. The weak limit of the conditional distribution of the bootstrapped Lasso estimator  $\mathbf{T}_n^*$  is a *random probability measure*; i.e.,  $\hat{G}_n(\cdot)$  converges weakly to a random probability measure on  $\mathbb{R}^p$ . To describe it, suppose that the error variables  $\{\epsilon_i : i \geq 1\}$  are defined on a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  and

let  $\mathbf{W}_1$  and  $\mathbf{W}_2$  be two independent copies of  $\mathbf{W} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{C})$ , also defined on  $(\Omega, \mathcal{F}, \mathbf{P})$ . Also, without loss of generality, suppose that  $\beta_j \neq 0$  for  $j = 1, \dots, p_0$  and  $\beta_j = 0$  for  $j = p_0 + 1, \dots, p$  for some  $0 \leq p_0 \leq p$ . Let  $p_1 = p - p_0$  denote the number of components of  $\boldsymbol{\beta}$  that are zero. For  $\mathbf{u} = (u_1, \dots, u_p)'$ ,  $\mathbf{t} = (t_1, \dots, t_p)' \in \mathbb{R}^p$ , define

$$(3.1) \quad V_\infty(\mathbf{t}; \mathbf{u}) = -2\mathbf{u}'\mathbf{W}_1 + \mathbf{u}'\mathbf{C}\mathbf{u} + \lambda_0 \sum_{j=1}^{p_0} u_j \operatorname{sgn}(\beta_j) \\ + \lambda_0 \sum_{j=p_0+1}^p \left\{ \operatorname{sgn}(t_j) \left[ u_j - 2(u_j + t_j) \mathbf{1}(\operatorname{sgn}(t_j)[u_j + t_j] < 0) \right] \right. \\ \left. + |u_j| \mathbf{1}(t_j = 0) \right\},$$

where summation over the empty set is assumed to be zero. (Thus, the second sum is zero if  $p_0 = p$ .) For each  $\mathbf{t} \in \mathbb{R}^p$ , let  $\mu_\infty(\mathbf{t}, \cdot)$  denote the probability distribution of  $\operatorname{argmin}(V_\infty(\mathbf{t}, \cdot))$ . Thus, for each  $\mathbf{t}$ ,  $\mu_\infty(\mathbf{t}, \cdot)$  is a probability measure on  $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ . Let  $\mathbb{M}$  denote the set of all probability measures on  $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$  and let  $\varrho$  denote the Prohorov metric (or any other metric metricizing the topology of weak convergence) on  $\mathbb{M}$ . Let  $\mathcal{M}$  denote the Borel  $\sigma$ -field on  $(\mathbb{M}, \varrho)$ . Then, it follows (cf. [Kal86]) that  $\mu_\infty(\mathbf{W}_2, \cdot)$  is an  $(\mathcal{F}, \mathcal{M})$ -measurable random measure. Similarly, the bootstrap estimator  $\hat{G}_n(\cdot)$  is also an  $(\mathcal{F}, \mathcal{M})$ -measurable random measure. The following result gives the limit distribution of the random measure  $\hat{G}_n(\cdot)$ .

**Theorem 3.1** (Limit distribution of usual bootstrap). *Suppose that*

$$(C.1) \quad \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \rightarrow \mathbf{C}, \text{ where } \mathbf{C} \text{ is positive definite. Further,}$$

$$n^{-1} \sum_{i=1}^n \|\mathbf{x}_i\|^3 = O(1) \quad \text{as } n \rightarrow \infty.$$

$$(C.2) \quad \lambda_n/\sqrt{n} \rightarrow \lambda_0 \geq 0 \text{ for some } \lambda_0 \in [0, \infty).$$

$$(C.3) \quad \text{The errors } \{\epsilon_i\}_{i=1}^n \text{ are iid with } \mathbf{E}(\epsilon_1) = 0, \text{ and } \mathbf{Var}(\epsilon_1) = \sigma^2 \in (0, \infty).$$

Then,  $\hat{G}_n(\cdot)$  converges weakly to  $\mu_\infty(\mathbf{T}_\infty, \cdot)$  as random elements of the metric space  $(\mathbb{M}, \varrho)$ , where  $\mathbf{T}_\infty = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^p} V(\mathbf{u})$  (cf. (2.2)) is the distributional limit of  $\mathbf{T}_n$ .

Theorem 3.1 shows that under conditions (C.1) - (C.3), the weak limit of the conditional distribution of the bootstrapped Lasso estimator is given by the random measure  $\mu_\infty(\mathbf{T}_\infty, \cdot)$ , where the randomness is driven by the random vector  $\mathbf{T}_\infty$ . Based on their heuristic arguments, [KF00] claimed that  $\hat{G}_n(\cdot)$  converged weakly to the distribution of the  $\operatorname{argmin}$  of the random process  $V^*(\cdot)$ , which in our notation is given by

$$(3.2) \quad V^*(\mathbf{u}) \equiv -2\mathbf{u}'\mathbf{W} + \mathbf{u}'\mathbf{C}\mathbf{u} + \lambda_0 \left[ \sum_{j=1}^{p_0} u_j \operatorname{sgn}(\beta_j) + \sum_{j=p_0+1}^p |u_j + T_{\infty,j}| \right].$$

It can be shown that  $V_\infty(\mathbf{t}; \mathbf{u})$  can also be written as

$$(3.3) \quad V_\infty(\mathbf{t}; \mathbf{u}) = -2\mathbf{u}'\mathbf{W}_1 + \mathbf{u}'\mathbf{C}\mathbf{u} + \lambda_0 \left[ \sum_{j=1}^{p_0} u_j \operatorname{sgn}(\beta_j) + \sum_{j=p_0+1}^p |u_j + t_j| - |t_j| \right].$$

A comparison of (3.2) and (3.3) shows that the heuristic formula of [KF00] is essentially correct if  $\mathbf{W}$  is replaced by a copy  $\mathbf{W}_1$  that is independent of the limiting random variable  $\mathbf{T}_\infty$  of the original Lasso estimator. In addition to providing a correct interpretation to the heuristic formula, Theorem 3.1 also gives regularity conditions that ensure the validity of the conjecture.

From Theorem 3.1, we also note that when  $\beta$  has at least one zero component and  $\lambda_0 \neq 0$ , the limit  $\mu_\infty(\mathbf{T}_\infty, \cdot)$  is a *nondegenerate random measure*. Thus, in this case, the bootstrap estimator  $\widehat{G}_n(B)$  of the probability  $G_n(B) = \mathbf{P}(\mathbf{T}_n \in B)$ , instead of converging to the *deterministic target*  $G_\infty(B) = \lim_{n \rightarrow \infty} G_n(B)$ , indeed converges (in distribution) to a random variable  $\mu_\infty(B, \cdot)$  for all  $B$  satisfying  $G_\infty(\partial B) = 0$ , where  $\partial B$  denotes the boundary of the set  $B$ . This, in particular, implies that the residual bootstrap approximation to the distribution of the Lasso estimator is inconsistent, which we state formally in the following corollary.

**Corollary 3.2** (Inconsistency of the bootstrap). *Suppose that the conditions of Theorem 3.1 hold. If  $\{j : \beta_j = 0\}$  is nonempty and  $\lambda_0 > 0$ , then*

$$\varrho(\widehat{G}_n(\cdot), G_\infty(\cdot)) \rightarrow_p 0 \quad \text{as } n \rightarrow \infty.$$

Corollary 3.2 implies that when  $\beta_j = 0$  for some  $j$ , bootstrap confidence regions for  $\beta$  cannot attain the nominal level even when the sample size  $n$  is large. Similarly, the bootstrap test for the hypothesis  $H_0 : \beta_j = 0$  for all  $j \in J$  (for a given nonempty set  $J$ ) fail to achieve the desired level of significance under the null hypothesis. The latter has some important negative implications in the context of model selection, where one needs to identify the zero components for selecting the true model from an initial larger model with superfluous covariates.

Next we consider some important special cases, where the bootstrap approximation is consistent. Note that if all components of  $\beta$  are nonzero or the regularization factor  $\lambda_n = o(n^{1/2})$  (i.e.  $\lambda_0 = 0$  in (C.2)), the function  $V_\infty(\mathbf{t}; \mathbf{u})$  reduces to  $V(\mathbf{u})$  (cf. (2.2)) for all  $\mathbf{t}, \mathbf{u} \in \mathbb{R}^p$ , and hence  $\mu_\infty(\mathbf{t}, \cdot)$  equals  $G_\infty(\cdot)$  for all  $\mathbf{t}$  and the residual bootstrap is consistent. We state this formally in the following corollary.

**Corollary 3.3** (Consistency of the bootstrap). *Suppose the conditions of Theorem 3.1 hold. If either  $p_0 = p$  or  $\lambda_0 = 0$ , then*

$$\sup_{A \in \mathcal{C}} \left| \widehat{G}_n(A) - G_n(A) \right| \rightarrow_p 0 \quad \text{as } n \rightarrow \infty,$$

where  $\mathcal{C}$  is the collection of all convex measurable sets in  $\mathbb{R}^p$ .

In situations covered by Corollary 3.3, the bootstrap approximation can be used to construct valid large sample confidence regions for  $\beta$ . This is of much practical importance, as the bootstrap automatically yields a valid confidence region without any analytical work on the part of the user.

*Remark.* Note that the regression model (1.1) assumes that the underlying regression parameter  $\beta$  is a *fixed* quantity and does not depend the sample size  $n$ . We have limited our study of the bootstrap in this fixed parameter framework. But there can be situations where such a fixed parameter framework is unsuitable. For example, if the underlying parameters are *small*, it is common to consider a contiguous framework where the underlying  $\beta_j$ 's are assumed to change (or decay) with the sample size  $n$ . In a series of papers, Pötscher and his co-authors (see [PL09], [LP08] and the references therein) have studied model selection properties

and asymptotics of various penalized regression methods in such contiguous settings and developed interesting *impossibility* and uniform consistency results, with uniformity over the underlying parameter space. The study of the bootstrap in such contiguous settings is an interesting issue requiring careful analysis and is beyond the scope of the current paper.

#### 4. PROOFS

Let  $C, C(\cdot)$  denote generic positive constants that depend on their arguments, but not on  $n$ . Also, recall that we write  $\mathbf{1}(\cdot)$  to denote the indicator function and

$$(4.1) \quad \text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $(\Omega, \mathcal{F}, \mathbf{P})$  denote the underlying probability space and let  $\mathcal{E} = \sigma\langle \epsilon_i : i \geq 1 \rangle$  denote the sub- $\sigma$ -field of  $\mathcal{F}$  generated by  $\{\epsilon_i : i \geq 1\}$ . For a random vector  $\mathbf{Z}$  and a  $\sigma$ -field  $\mathcal{C}$ , write  $\mathcal{L}(\mathbf{Z})$  and  $\mathcal{L}(\mathbf{Z}|\mathcal{C})$  to denote the probability distribution of  $\mathbf{Z}$  and the conditional distribution of  $\mathbf{Z}$  given  $\mathcal{C}$ , respectively. For any random vector  $\mathbf{Y}$ , set  $\mathcal{L}(\mathbf{Z}|\sigma\langle \mathbf{Y} \rangle) = \mathcal{L}(\mathbf{Z}|\mathbf{Y})$ , for notational simplicity. Write  $\mathbf{X}_n$  for the  $n \times p$  matrix with rows  $\mathbf{x}'_i$ ,  $i = 1, \dots, n$ , and let  $\mathbf{C}_n = n^{-1} \mathbf{X}'_n \mathbf{X}_n$ . Unless otherwise indicated, limits in the order symbols are taken by letting  $n \rightarrow \infty$ . Recall that  $\mathbf{P}_*$  denotes conditional probability given  $\mathcal{E}$  and  $\mathbf{E}_* = \mathbf{E}(\cdot|\mathcal{E})$ .

Our first result proves strong consistency of  $\hat{\beta}_n$ , which may be of some independent interest.

**Lemma 4.1.** *Suppose that conditions (C.1) and (C.2) hold. Then*

$$(4.2) \quad \left\| \sum_{i=1}^n \epsilon_i \mathbf{x}_i \right\| = o\left(n^{1/2} \log n\right), \quad \text{a.s. } (\mathbf{P}).$$

*Proof of Lemma 4.1.* It is enough to show that each component of the vector

$$\sum_{i=1}^n \epsilon_i \mathbf{x}_i = o\left(n^{1/2} \log n\right)$$

almost surely. Hence, for notational simplicity, suppose that the  $\mathbf{x}_i$ 's are scalar. As a first step, we show that with probability one,

$$(4.3) \quad \sum_{i=2}^{\infty} \frac{x_i \epsilon_i}{i^{1/2} (\log i)} \quad \text{converges.}$$

By Khinchine-Kolmogorov's 1-series theorem (cf. Theorem 8.34 in [AL06]), it is enough to show that

$$(4.4) \quad \sum_{i=2}^{\infty} \frac{x_i^2}{i(\log i)^2} < \infty.$$

Let  $s_n^2 = \sum_{i=2}^n x_i^2$ ,  $n \geq 2$ . Then by (C.3),  $s_n^2 = O(n)$  as  $n \rightarrow \infty$ . Now, using Abel's summation formula (cf. Lemma 8.4.1 in [AL06]), we have

$$\begin{aligned} \sum_{i=2}^n \frac{x_i^2}{i(\log i)^2} &= \frac{s_n^2}{n(\log n)^2} - \sum_{j=2}^n s_j^2 \left[ (j+1)^{-1} (\log(j+1))^{-2} - j^{-1} (\log j)^{-2} \right] \\ &= O([\log n]^{-2}) + O\left( \sum_{j=2}^n \frac{j \left| (j+1) [\log(j+1)]^2 - j (\log j)^2 \right|}{j(j+1)(\log j)^2 [\log(j+1)]^2} \right) \\ &= O([\log n]^{-2}) + O\left( \sum_{j=2}^n \frac{j [\log(j+1)]^2}{j(j+1)(\log j)^2 [\log(j+1)]^2} \right) \\ &= O([\log n]^{-2}) + O\left( \sum_{j=2}^n j^{-1} (\log j)^{-2} \right) = O(1), \end{aligned}$$

as

$$\begin{aligned} \left| (j+1) [\log(j+1)]^2 - j (\log j)^2 \right| &\leq \left| j \left\{ [\log(j+1)]^2 - (\log j)^2 \right\} \right| + [\log(j+1)]^2 \\ &\leq 2j \log(j+1) [\log(1+j^{-1})] + [\log(j+1)]^2 \\ &= O([\log(j+1)]^2). \end{aligned}$$

Thus (4.3) and (4.4) hold. Now using Kronecker's Lemma (cf. Lemma 8.4.2 in [AL06]), one can conclude (4.2) from (4.3).  $\square$

**Lemma 4.2** (Strong consistency). *Suppose that conditions (C.1) - (C.3) are satisfied. Then*

$$(4.5) \quad \|\mathbf{T}_n\| = O(\log n), \quad \text{with probability 1.}$$

*Proof of Lemma 4.2.* Suppose  $\beta_j \neq 0$  for all  $1 \leq j \leq p_0$  and  $\beta_j = 0$  for all  $(p_0 + 1) \leq j \leq p$ . Let  $\eta_{1,n}$  and  $\eta_{p,n}$  denote the smallest and largest eigenvalues of  $\mathbf{C}_n$ . Note that  $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \operatorname{argmin}_{\mathbf{u}} V_n(\mathbf{u})$ , where

$$\begin{aligned} V_n(\mathbf{u}) &= \mathbf{u}' \mathbf{C}_n \mathbf{u} - 2\mathbf{u}' \mathbf{W}_n + \lambda_n \left[ \sum_{j=1}^{p_0} \left( |\beta_j + \frac{u_j}{\sqrt{n}}| - |\beta_j| \right) + \sum_{j=p_0+1}^p \left| \frac{u_j}{\sqrt{n}} \right| \right] \\ &\geq \eta_{1,n} \|\mathbf{u}\|^2 - 2\|\mathbf{u}\| \|\mathbf{W}_n\| + \lambda_n \left[ \sum_{j=1}^{p_0} \left( |\beta_j| - \left| \frac{u_j}{\sqrt{n}} \right| \right) + \sum_{j=p_0+1}^p \left| \frac{u_j}{\sqrt{n}} \right| \right] \\ (4.6) \quad &\geq \|\mathbf{u}\| \left[ \eta_{1,n} \|\mathbf{u}\| - 2\|\mathbf{W}_n\| - \frac{\lambda_n}{\sqrt{n}} \sqrt{p} \right] = V_{1,n}(\mathbf{u}) \quad (\text{say}), \end{aligned}$$

where  $\mathbf{W}_n = n^{-1/2} \sum_{i=1}^n \epsilon_i \mathbf{x}_i$ . Note that  $\mathbf{W}_n$  is a scaled sum of independent random vectors, with mean  $\mathbf{0}$  and finite variance. By Lemma 4.1, we have

$$\|\mathbf{W}_n\| = O(\log n), \quad \text{with probability 1.}$$

This implies that for  $\|\mathbf{u}\| \geq C \log n$ ,

$$\begin{aligned} \frac{\eta_{1,n} \|\mathbf{u}\|}{2} &\geq 2\|\mathbf{W}_n\| + \frac{\lambda_n}{\sqrt{n}} \sqrt{p} \\ \Rightarrow V_n(\mathbf{u}) &\geq C \frac{\eta_{1,n}}{2} \log n (> 0), \end{aligned}$$

with probability 1. Since  $V_n(\mathbf{0}) = 0$ , this implies that the minimizer of  $V_n(\mathbf{u})$  can not lie in the set  $\{\|\mathbf{u}\| \geq C \log n\}$ . Thus, the minimizer of  $V_n(\mathbf{u})$  lies in the set  $\{\|\mathbf{u}\| < C \log n\}$  with probability 1, proving the lemma.  $\square$

**Lemma 4.3.** *Let  $s_n^2 = n^{-1} \sum_{j=1}^n (e_j - \bar{e}_n)^2$  and  $\hat{\mu}_{3,n} = n^{-1} \sum_{j=1}^n |e_j - \bar{e}_n|^3$ . Assume that  $\frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_j\|^2 = O(1)$  and condition (C.3) holds. Then*

$$|s_n^2 - \sigma^2| + n^{-1/2} \hat{\mu}_{3,n} \xrightarrow{a.s.} 0,$$

where we recall that  $\sigma^2 = \mathbf{Var}(\epsilon_1)$ .

*Proof of Lemma 4.3.* First consider the first term :  $|s_n^2 - \sigma^2|$ . Define

$$\sigma_n^2 = n^{-1} \sum_{j=1}^n (\epsilon_j - \bar{\epsilon}_n)^2,$$

where  $\bar{\epsilon}_n = n^{-1} \sum_{j=1}^n \epsilon_j$ . Then, we can write, using Lemma 4.2,

$$\begin{aligned} (s_n - \sigma_n)^2 &\leq \frac{1}{n} \sum_{j=1}^n \left( [e_j - \bar{e}_n] - [\epsilon_j - \bar{\epsilon}_n] \right)^2 \\ &\leq \frac{1}{n} \sum_{j=1}^n (e_j - \epsilon_j)^2 \\ &\leq \left( \frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_j\|^2 \right) \|\hat{\beta}_n - \beta\|^2 \\ &= o(1), \quad \text{with probability 1.} \end{aligned}$$

Since  $\sigma_n^2 \rightarrow \sigma^2$  almost surely, it follows that  $|s_n^2 - \sigma^2| = o(1)$  with probability 1.

Next consider  $\hat{\mu}_{3,n}$ . Using the condition on the  $\mathbf{x}_i$ 's, we get

$$(4.7) \quad \max_{1 \leq i \leq n} \|\mathbf{x}_i\| \leq \left( \sum_{i=1}^n \|\mathbf{x}_i\|^2 \right)^{1/2} = O(n^{1/2}).$$

Hence, by the Marcinkiewicz-Zygmund Strong Law of Large Numbers, Lemma 4.2, and (4.7), we have

$$\begin{aligned} n^{-1/2} \hat{\mu}_{3,n} &\leq 8n^{-3/2} \sum_{i=1}^n \left| \epsilon_i - \mathbf{x}_i(\hat{\beta}_n - \beta) \right|^3 \\ &\leq 32 \left[ n^{-3/2} \sum_{i=1}^n |\epsilon_i|^3 + \left( n^{-3/2} \sum_{i=1}^n \|\mathbf{x}_i\|^3 \right) \|\hat{\beta}_n - \beta\|^3 \right] \\ &= o(1), \quad \text{with probability 1.} \end{aligned}$$

This completes the proof of the lemma.  $\square$

**Lemma 4.4.** *For each  $n \geq 1$ , let  $\{e_{i,n}^\dagger : 1 \leq i \leq n\}$  be a collection of random variables on  $(\Omega, \mathcal{F}, \mathbf{P})$  such that given  $\mathcal{E}$ ,  $\{e_{i,n}^\dagger : 1 \leq i \leq n\}$  are iid with  $\mathbf{E}_* e_{i,n}^\dagger = 0$ , and  $|\mathbf{E}_*(e_{i,n}^\dagger)^2 - \tau^2| + n^{-1/2} \mathbf{E}_* |e_{i,n}^\dagger|^3 \rightarrow 0$  with probability 1 for some  $\tau \in (0, \infty)$ . Also, suppose condition (C.1) holds. Then*

$$\mathcal{L}\left(n^{-1/2} \sum_{i=1}^n \mathbf{x}_i e_{i,n}^\dagger \mid \mathcal{E}\right) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \tau^2 C) \quad \text{with probability 1.}$$



*Proof of Lemma 4.4.* Let  $A \in \mathcal{F}$  be such that  $\mathbf{P}(A) = 1$  and for every  $\omega \in A$ ,

$$|\mathbf{E}_*(e_{i,n}^\dagger)^2 - \tau^2| + n^{-1/2}\mathbf{E}_*|e_{i,n}^\dagger|^3 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Fix  $\omega \in A$ . For this  $\omega$ , we will use the Cramer-Wold device to prove the result. Accordingly, consider a  $\mathbf{t} = (t_1, \dots, t_p)' \in \mathbb{R}^p$ ,  $\mathbf{t} \neq \mathbf{0}$ . Let

$$s_n^2(\mathbf{t}) = \mathbf{Var}_*\left(n^{-1/2} \sum_{i=1}^n \mathbf{t}'\mathbf{x}_i e_{i,n}^\dagger\right).$$

Note that  $s_n^2(\mathbf{t}) \rightarrow \mathbf{t}'\mathbf{C}\mathbf{t}\tau^2 \in (0, \infty)$ . Then, by the Berry-Esseen Theorem (cf. Chapter 12, [BRR86]),

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \mathbf{P}_*\left(n^{-1/2} \sum_{i=1}^n \mathbf{t}'\mathbf{x}_i e_{i,n}^\dagger \leq x\right) - \Phi(x/s_n(\mathbf{t})) \right| &\leq C \frac{\sum_{i=1}^n \mathbf{E}_* \left| n^{-1/2} \mathbf{t}'\mathbf{x}_i e_{i,n}^\dagger \right|^3}{\left(\sum_{i=1}^n \mathbf{E}_* \left| n^{-1/2} \mathbf{t}'\mathbf{x}_i e_{i,n}^\dagger \right|^2\right)^{3/2}} \\ &\leq C \frac{\|\mathbf{t}\|^3 n^{-3/2} \sum_{i=1}^n \|\mathbf{x}_i\|^3 \mathbf{E}_* |e_{i,n}^\dagger|^3}{s_n^3(\mathbf{t})} \\ &= o(1). \end{aligned}$$

This completes the proof of the lemma. □

*Proof of Theorem 3.1.* Using Skorohod’s theorem (cf. Theorem 1.10.3, [vdVW96]), without loss of generality, we may suppose that there exists a version  $\mathbf{T}_\infty$  (also denoted as  $\mathbf{T}_\infty$  for notational simplicity) on  $(\Omega, \mathcal{F}, \mathbf{P})$  such that  $\mathbf{T}_n \rightarrow \mathbf{T}_\infty$  *a.s.* ( $\mathbf{P}$ ). Let  $A \in \mathcal{F}$  be such that  $\mathbf{P}(A) = 1$  and

$$\lim_{n \rightarrow \infty} \varrho\left(\mathcal{L}\left(\mathbf{W}_n^* \Big| \mathcal{E}\right)(\omega), \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{C})\right) + \|\mathbf{T}_n(\omega) - \mathbf{T}_\infty(\omega)\| = 0 \quad \text{for all } \omega \in A,$$

where  $\mathbf{W}_n^* = n^{-1/2} \sum_{i=1}^n \mathbf{x}_i e_i^*$ . Recall that  $\mathbb{M}$  denotes the set of all probability measures on  $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ , equipped with the topology of weak convergence, and  $\varrho(\cdot, \cdot)$  is a metric metricizing the topology of weak convergence on  $\mathbb{M}$ . Then,  $(\mathbb{M}, \varrho)$  is a complete separable metric space. Then,  $\hat{G}_n$  and  $\hat{G}_\infty$  can be considered  $(\mathcal{F}, \mathcal{M})$  measurable,  $\mathbb{M}$ -valued random elements. By the definition of weak convergence for  $\mathbb{M}$ -valued random elements and the Bounded Convergence Theorem, Theorem 3.1 will follow if we show that

$$(4.8) \quad \lim_{n \rightarrow \infty} \varrho\left(\hat{G}_n(\omega, \cdot), \hat{G}_\infty(\omega, \cdot)\right) = 0 \quad \text{for all } \omega \in A.$$

To that end, for  $\mathbf{a} = (a_j : p_0 + 1 \leq j \leq p) \in \{0, 1\}^{p_1}$ , define the sets

$$A_{\mathbf{a}} = \{\omega \in A : \mathbf{T}_{\infty, j}(\omega) \in \mathbf{I}(a_j), \quad j = p_0 + 1, \dots, p\},$$

with  $\mathbf{I}(0) = \{0\}$  and  $\mathbf{I}(1) = \mathbb{R} \setminus \{0\}$ . Then,

$$\begin{aligned} V_n^*(\mathbf{u}) &= \mathbf{u}'\mathbf{C}_n\mathbf{u} - 2\mathbf{u}'\mathbf{W}_n^* + \lambda_n \sum_{j=1}^p \left\{ \left| \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right| - |\hat{\beta}_{n,j}| \right\} \\ &= \mathbf{u}'\mathbf{C}_n\mathbf{u} - 2\mathbf{u}'\mathbf{W}_n^* + \lambda_n \sum_{j=1}^{p_0} \left\{ \left| \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right| - |\hat{\beta}_{n,j}| \right\} \\ &\quad + \lambda_n \sum_{j=p_0+1}^p \left\{ \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right) \mathbf{1} \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \geq 0 \right) \right. \\ &\quad \left. - \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right) \mathbf{1} \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} < 0 \right) - |\hat{\beta}_{n,j}| \right\}. \end{aligned}$$

Now consider the  $j$ -th summand of the last term on the right side. For  $\hat{\beta}_{n,j} \neq 0$ ,

$$\begin{aligned} &\lambda_n \left\{ \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right) \mathbf{1} \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \geq 0 \right) - \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right) \mathbf{1} \left( \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} < 0 \right) - |\hat{\beta}_{n,j}| \right\} \\ &= \begin{cases} \lambda_n \left\{ \frac{u_j}{\sqrt{n}} \mathbf{1} \left( u_j \geq -\sqrt{n}\hat{\beta}_{n,j} \right) - \left( 2\hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right) \mathbf{1} \left( u_j < -\sqrt{n}\hat{\beta}_{n,j} \right) \right\} & \text{if } \hat{\beta}_{n,j} > 0, \\ \lambda_n \left\{ 2\hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \mathbf{1} \left( u_j \geq -\sqrt{n}\hat{\beta}_{n,j} \right) - \frac{u_j}{\sqrt{n}} \mathbf{1} \left( u_j < -\sqrt{n}\hat{\beta}_{n,j} \right) \right\} & \text{if } \hat{\beta}_{n,j} < 0, \end{cases} \\ &= \begin{cases} \frac{\lambda_n}{\sqrt{n}} \left\{ u_j - 2 \left( u_j + \sqrt{n}\hat{\beta}_{n,j} \right) \mathbf{1} \left( u_j < -\sqrt{n}\hat{\beta}_{n,j} \right) \right\} & \text{if } \hat{\beta}_{n,j} > 0, \\ \frac{\lambda_n}{\sqrt{n}} \left\{ 2 \left( \sqrt{n}\hat{\beta}_{n,j} + u_j \right) \mathbf{1} \left( u_j \geq -\sqrt{n}\hat{\beta}_{n,j} \right) - u_j \right\} & \text{if } \hat{\beta}_{n,j} < 0, \end{cases} \\ &= \operatorname{sgn} \left( \hat{\beta}_{n,j} \right) \frac{\lambda_n}{\sqrt{n}} \left[ u_j - 2 \left( u_j + \sqrt{n}\hat{\beta}_{n,j} \right) \mathbf{1} \left\{ \operatorname{sgn} \left( \hat{\beta}_{n,j} \right) \left( u_j + \sqrt{n}\hat{\beta}_{n,j} \right) < 0 \right\} \right]. \end{aligned}$$

For  $\hat{\beta}_{n,j} = 0$ ,

$$\left| \hat{\beta}_{n,j} + \frac{u_j}{\sqrt{n}} \right| - |\hat{\beta}_{n,j}| = \frac{|u_j|}{\sqrt{n}}.$$

Hence on the set  $\{\mathbf{u} : |u_j| < \sqrt{n}|\hat{\beta}_{n,j}|, j = 1, \dots, p_0\}$ ,

$$\begin{aligned} V_n^*(\mathbf{u}) &= \mathbf{u}'\mathbf{C}_n\mathbf{u} - 2\mathbf{u}'\mathbf{W}_n^* + \frac{\lambda_n}{\sqrt{n}} \sum_{j=1}^{p_0} \operatorname{sgn} \left( \hat{\beta}_{n,j} \right) u_j \\ &\quad + \frac{\lambda_n}{\sqrt{n}} \sum_{j=p_0+1}^p \left\{ \left[ u_j - 2 \left( u_j + \sqrt{n}\hat{\beta}_{n,j} \right) \mathbf{1} \left( \operatorname{sgn} \left( \hat{\beta}_{n,j} \right) \left[ u_j + \sqrt{n}\hat{\beta}_{n,j} \right] < 0 \right) \right] \right. \\ &\quad \left. \times \operatorname{sgn} \left( \hat{\beta}_{n,j} \right) + |u_j| \mathbf{1} \left( \hat{\beta}_{n,j} = 0 \right) \right\} \quad (\text{because } \operatorname{sgn}(\hat{\beta}_{n,j}) = 0 \text{ if } \hat{\beta}_{n,j} = 0). \end{aligned}$$

Next fix  $\mathbf{a} \in \{0, 1\}^{p_1}$  and  $\omega \in A_{\mathbf{a}}$ . Let  $J(\mathbf{a}) = \{j : p_0 + 1 \leq j \leq p, a_j = 1\}$ . We shall show that

$$(4.9) \quad \varrho \left( \hat{G}_n(\omega, \cdot), \mathcal{L} \left( \operatorname{argmin}_{\mathbf{u}} V_{\infty}(\mathbf{u}, \mathbf{T}_{\infty}(\omega)) \right) \right) \xrightarrow{a.s.} 0,$$

as  $n \rightarrow \infty$ . By our definition of  $A$ , as  $n \rightarrow \infty$ ,

$$(4.10) \quad \mathcal{L}(\mathbf{W}_n^* | \mathcal{E})(\omega) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{C})$$

and

$$\widehat{\beta}_{n,j}(\omega) = \beta_j + \frac{1}{\sqrt{n}} T_{n,j}(\omega) \rightarrow \beta_j, \quad \text{for all } 1 \leq j \leq p_0.$$

In particular

$$(4.11) \quad \text{sgn}(\widehat{\beta}_{n,j}(\omega)) \rightarrow \text{sgn}(\beta_j), \quad \text{as } n \rightarrow \infty \text{ for all } 1 \leq j \leq p_0.$$

For  $j \in J(\mathbf{a})$ ,  $\sqrt{n}\widehat{\beta}_{n,j}(\omega) \rightarrow T_{\infty,j}(\omega) \neq 0$ , and hence

$$(4.12) \quad \text{sgn}(\widehat{\beta}_{n,j}(\omega)) \rightarrow \text{sgn}(T_{\infty,j}(\omega)).$$

Next, consider  $j \in \{p_0 + 1, \dots, p\}$  with  $a_j = 0$ . In this case, by the definition of  $A_{\mathbf{a}}$ ,

$$\sqrt{n}\widehat{\beta}_{n,j}(\omega) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and hence if  $\widehat{\beta}_{n,j}(\omega) > 0$ , then

$$\begin{aligned} & \left[ u_j - 2(u_j + \sqrt{n}\widehat{\beta}_{n,j}(\omega)) \right. \\ & \quad \left. \times \mathbf{1}(\text{sgn}(\widehat{\beta}_{n,j}(\omega)) [u_j + \sqrt{n}\widehat{\beta}_{n,j}(\omega)] < 0) \right] \text{sgn}(\widehat{\beta}_{n,j}(\omega)) \\ &= - (u_j - 2\sqrt{n}\widehat{\beta}_{n,j}(\omega)) \mathbf{1}(u_j < -\sqrt{n}\widehat{\beta}_{n,j}(\omega)) \\ & \quad + u_j \mathbf{1}(-\sqrt{n}\widehat{\beta}_{n,j}(\omega) \leq u_j \leq 0) + u_j \mathbf{1}(u_j > 0) \\ &= |u_j| - 2\sqrt{n}\widehat{\beta}_{n,j}(\omega) \mathbf{1}(u_j < -\sqrt{n}\widehat{\beta}_{n,j}(\omega)) \\ & \quad - 2u_j \mathbf{1}(-\sqrt{n}\widehat{\beta}_{n,j}(\omega) \leq u_j \leq 0) \\ (4.13) \quad &= |u_j| + \theta \sqrt{n}\widehat{\beta}_{n,j}(\omega), \end{aligned}$$

where  $\theta = \theta(u_j) \in [-2, 2]$  for all  $u_j \in \mathbb{R}$ . Using similar considerations for  $\widehat{\beta}_{n,j}(\omega) < 0$  and using (4.11), (4.12) and (4.13), it now follows that for any compact set  $K \in \mathbb{R}^p$ ,

$$\begin{aligned} & \sup_{\mathbf{u} \in K} \left| V_n^*(\mathbf{u}) - \left[ \mathbf{u}' \mathbf{C}_n \mathbf{u} - 2\mathbf{u}' \mathbf{W}_n^* + \lambda_0 \sum_{j=1}^{p_0} \text{sgn}(\beta_j) u_j \right. \right. \\ & \quad \left. \left. + \lambda_0 \sum_{j \in J(\mathbf{a})} \text{sgn}(T_{\infty,j}(\omega)) \left\{ \left( u_j - 2(u_j + T_{\infty,j}(\omega)) \right) \right. \right. \right. \\ & \quad \left. \left. \times \mathbf{1}(\text{sgn}(T_{\infty,j}(\omega)) [u_j + T_{\infty,j}(\omega)] < 0) \right\} \right. \\ & \quad \left. \left. + \lambda_0 \sum_{j \notin J(\mathbf{a})} |u_j| \right] \right| \\ (4.14) \quad &= o(1) \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Now using standard results on weak convergence (cf. [KF00], [KP90]), from (4.10) and (4.14) we get

$$\mathcal{L}(V_n^*(\mathbf{u}) \mid \mathcal{E})(\omega) \xrightarrow{d} \mathcal{L}(V_\infty(\mathbf{T}_\infty(\omega), \mathbf{u})), \quad \text{uniformly on compact sets,}$$

and

$$\mathcal{L}\left(\operatorname{argmin}_{\mathbf{u}} V_n^*(\mathbf{u}) \mid \mathcal{E}\right)(\omega) \xrightarrow{d} \mathcal{L}\left(\operatorname{argmin}_{\mathbf{u}} V_\infty(\mathbf{T}_\infty(\omega), \mathbf{u})\right).$$

Hence (4.9) follows. This completes the proof of the theorem.  $\square$

*Proof of Corollary 3.2.* Since  $\mu_\infty(\mathbf{T}_\infty, \cdot)$  is a *nondegenerate* random measure, it follows that there exists an  $\eta \in (0, 1)$  such that

$$P\left(\varrho(\mu_\infty(\mathbf{T}_\infty, \cdot), G_\infty(\cdot)) > \eta\right) > \eta.$$

The result now follows from this and Theorem 3.1.  $\square$

*Proof of Corollary 3.3.* Under the assumptions of Corollary 3.3,  $\mu_\infty(\mathbf{T}_\infty, \cdot)$  does not depend on  $\mathbf{T}_\infty$  and is equal to  $G_\infty$ , which is now given by a normal distribution on  $\mathbb{R}^p$ . Hence, the result follows by combining Theorem 3.1, Theorem 2 of [KF00], Theorem 2.11 (on page 22) and Corollary 3.2 of [BRR86] on the uniformity classes for weak convergence.  $\square$

#### ACKNOWLEDGEMENT

The authors thank an anonymous referee for some insightful and constructive comments that improved the presentation of the paper.

#### REFERENCES

- [AL06] Krishna B. Athreya and Soumendra N. Lahiri. *Measure theory and probability theory*. Springer Texts in Statistics. Springer, New York, 2006. MR2247694
- [BRR86] R. N. Bhattacharya and R. Ranga Rao. *Normal approximation and asymptotic expansions*. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1986. Reprint of the 1976 original. MR0855460
- [Fre81] D. A. Freedman. Bootstrapping regression models. *Ann. Statist.*, 9(6):1218–1228, 1981. MR0630104
- [Kal86] Olav Kallenberg. *Random measures*. Akademie-Verlag, Berlin, fourth edition, 1986. MR0854102
- [KP90] JeanKyung Kim and David Pollard. Cube root asymptotics. *Ann. Statist.*, 18(1):191–219, 1990. MR1041391
- [KF00] Keith Knight and Wenjiang Fu. Asymptotics for lasso-type estimators. *Ann. Statist.*, 28(5):1356–1378, 2000. MR1805787
- [LP08] Hannes Leeb and Benedikt M. Pötscher. Sparse estimators and the oracle property, or the return of Hodges’ estimator. *J. Econometrics*, 142(1):201–211, 2008. MR2394290
- [PL09] Benedikt M. Pötscher and Hannes Leeb. On the distribution of penalized maximum likelihood estimators: the LASSO, SCAD, and thresholding. *J. Multivariate Anal.*, 100(9):2065–2082, 2009. MR2543087
- [Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996. MR1379242
- [vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York, 1996. MR1385671
- [Wai06] M. J. Wainwright, Sharp thresholds for high-dimensional and noisy recovery of sparsity. [arXiv:math/0605740v1](https://arxiv.org/abs/math/0605740v1) (2006).

- [ZY06] Peng Zhao and Bin Yu. On model selection consistency of Lasso. *J. Mach. Learn. Res.*, 7:2541–2563, 2006. MR2274449
- [Zou06] Hui Zou. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476):1418–1429, 2006. MR2279469

DEPARTMENT OF STATISTICS, TEXAS A&M UNIVERSITY, COLLEGE STATION, TEXAS 77843-3143  
*E-mail address:* `cha@stat.tamu.edu`

DEPARTMENT OF STATISTICS, TEXAS A&M UNIVERSITY, COLLEGE STATION, TEXAS 77843-3143  
*E-mail address:* `snlahiri@stat.tamu.edu`