The Annals of Statistics
1988, Vol. 16, No. 4, 1696–1708

BOOTSTRAP PROCEDURES UNDER SOME NON-I.I.D. MODELS[1]

BY REGINA Y. LIU

Rutgers University

It is shown in this article that the classical i.i.d. bootstrap remains a valid procedure for estimating the sampling distributions of certain symmetric estimators of location, as long as the random observations are independently drawn from distributions with (essentially) a common location. This may be viewed as a robust property of the classical i.i.d. bootstrap. Also included is a study of the second order properties of a different bootstrap procedure proposed by Wu in the context of heteroscedasticity in regression.

1. Introduction. The bootstrap resampling procedure is known to be a good general procedure for estimating a sampling distribution under i.i.d. models. [See, e.g., Efron (1979), Bickel and Freedman (1981) and Singh (1981).] In practical situations the i.i.d. setup is often violated, and so it is natural to wonder how the bootstrap performs under non-i.i.d. models. In this article we focus particularly on the bootstrap procedure under (essentially) two models: (i) $X_1, \dots, X_n$ are independent observations drawn from distributions $G_1, \dots, G_n$ with a common mean or a common center of symmetry but not necessarily identical. (ii) The simple regression model $Y_i = \beta x_i + e_i$, $i = 1, \dots, n$, where $\mathrm{Var}(e_i)$'s are possibly different.

For motivation, consider the following two practical examples for which model (i) arises naturally.

EXAMPLE 1. Suppose n measurements of a certain object are taken using a number of different instruments by personnel with varying skills. However, the collected data do not take into account the different instruments and personnel.

EXAMPLE 2. Suppose in a given produce line the quality of the product varies over time due to some deterministic or random factors. As a result the variance of a certain measurement of the output does not necessarily remain the same.

With regard to model (i), we are asking the specific question: Suppose we choose one of the standard estimates for location, e.g., the mean, the median, a trimmed mean or the average of a few quantiles, and then apply the classical bootstrap based on i.i.d. samples from the empirical population to estimate the sampling distribution of the chosen estimate. Are the results still asymptotically

---

Received November 1986; revised December 1987.
[1]Research supported by NSF Grant DMS-85-02945.
AMS 1980 *subject classification*. 62G05.
Key words and phrases. Bootstrap, non-i.i.d. models, sampling distributions, second order asymptotics, regression, *L*-statistics, Edgeworth expansions.

correct? Interestingly enough, the answer is yes, even though the classical bootstrap does not seem intuitively appropriate here for the simple reason that the original data are not i.i.d. while the bootstrap still draws i.i.d. samples. Moreover, this "incorrect" bootstrap not only captures the first order limit, but also retains the second order asymptotic properties in the case of the sample mean. In the next paragraph we provide a brief explanation for this phenomenon, which represents a robust property of the bootstrap.

Throughout the paper we use the notation $\bar{\cdot}_n$ to indicate the average of the elements $\cdot_1, \dots, \cdot_n$, for example, $\overline{G}_n = (1/n)\sum_{i=1}^{n} G_i$. The result that the asymptotic properties of the classical bootstrap still hold under model (i) is primarily a consequence of the simple probabilistic identity $\mathrm{Var}(\overline{X}_n) = \mathrm{Var}(\bar{\xi}_n)$, where $\{\xi_i\}$ are taken to be i.i.d. random variables having c.d.f. $\overline{G}_n$ while assuming that G_i's have the same mean. Since the empirical distribution function F_n based on the X_i's well approximates $\overline{G}_n$, typically at the rate $O_p(n^{-1/2})$, the classical bootstrap which draws samples for F_n should correctly estimate the standard error of $\bar{\xi}_n$ which equals the standard error of $\overline{X}_n$. Thus, the problem basically does not reflect the difference between the stationary model $(\overline{G}_n \times \cdots \times \overline{G}_n)$ and the nonstationary model $(G_1 \times \cdots \times G_n)$. This observation also extends to symmetric L-statistics (linear combinations of order statistics). Using a representation similar to the ones derived in Babu and Singh (1984) and Liu, Singh and Lo (1986), we note that bootstrapping symmetric L-statistics is mathematically equivalent to bootstrapping a sample mean whose component variables are of mean 0, provided that the underlying distributions are all symmetric about a common center.

The claim made previously regarding the second order asymptotic properties is established by examining the formal one-term Edgeworth expansions for both the original sampling distribution and the bootstrap distribution. Details are given in later sections.

As for the regression problem with $\mathrm{Var}(e_i) = \sigma_i^2$ mentioned in (ii), we search for modifications of the classical bootstrap procedure which provide consistent results regardless of the nonidentical error variances. Partial success has been achieved in this search. In particular, we have found the bootstrap procedure proposed by Wu (1986) useful. In this procedure i.i.d. observations are drawn from an external population (having mean 0 and variance 1) which is totally unrelated to the original data set. We devote Section 4B to the issue of second order asymptotic properties of this procedure. The main discovery is that if the external population, in addition to having mean 0 and variance 1, also has its third central moment equal to 1, then Wu's bootstrap shares the usual second order asymptotic properties of the classical bootstrap.

The rest of the article is organized as follows: In Section 2, we study the bootstrap distributions of the studentized as well as the nonstudentized statistics based on $\overline{X}_n$ under model (i). In Section 3, we extend the first order results obtained for the classical bootstrap in Section 2 to a class of symmetric L-statistics. In the last section, besides studying Wu's bootstrap procedure in the regression problems, we also use this bootstrap to provide a competing alternative for bootstrapping the mean (Section 4A).

2. The classical bootstrap on $\bar{X}_n$. Throughout Sections 2 and 3, $\{Y_1, \ldots, Y_n\}$ indicates a bootstrap sample which is formed by drawing an i.i.d. sample of size n from F_n, where F_n is the empirical d.f. based on $X_1, \ldots, X_n$. Recall that G_i is the c.d.f. of X_i, $i = 1, \ldots, n$.

THEOREM 1. *Let $X_1, \ldots, X_n$ be a set of independent random observations and let μ_i and σ_i^2 denote, respectively, the mean and the variance of X_i, $i = 1, \ldots, n$. Also let $\nu_n^2 = (1/n)\sum_{i=1}^n \sigma_i^2$ and $V_n^2 = (1/n)\sum_{i=1}^n (X_i - \bar{X}_n)^2$. If* (i) $\lim_{n \to \infty}(1/n)\sum_{i=1}^n(\mu_i - \bar{\mu}_n)^2 = 0$, (ii) $\liminf_{n \to \infty} \nu_n^2 > 0$ *and* (iii) $E|X_i|^{2+\delta} \le K < \infty$, *for some $\delta > 0$ and for all i, then*

$$\lim_{n \to \infty} \|P^*\big(\sqrt{n}\,(\bar{Y}_n - \bar{X}_n) \le x\big) - P\big(\sqrt{n}\,(\bar{X}_n - \bar{\mu}_n) \le x\big)\|_\infty = 0 \quad a.s.$$

(P^* stands for the bootstrap probability and $\|\ \|_\infty$ stands for the sup-norm over x.)

The proof of Theorem 1 uses a modified version of Marcinkiewicz–Zygmund SLLN (strong law of large numbers) which we state in

LEMMA 1. *Let $W_1, W_2, \ldots$ be independent random variables with $E|W_i|^{p+\varepsilon} \le K$ for some p, $0 < p < 2$, and $\varepsilon > 0$. Then*

$$n^{-1/p} \sum_{i=1}^{n} (W_i - a_i) \to 0 \quad a.s., \tag{2.1}$$

where $a_i = EW_i$ if $p \ge 1$ and $a_i = 0$ otherwise.

PROOF. The proof is sketched here only for $p < 1$. Arguments for the other case should be similar. Note that by Kronecker's lemma, (2.1) follows from

$$\sum_{n=1}^{\infty} n^{-1/p} W_n < \infty \quad \text{a.s.} \tag{2.2}$$

To obtain (2.2) we first concentrate on the truncated random variables $T_n = n^{-1/p} W_n I_{[|W_n| \le n^{1/p}]}$ and show that $\sum_{n=1}^\infty E|T_n|^2 < \infty$, which in turn implies that $\sum_{n=1}^\infty (T_n - ET_n) < \infty$ a.s. Since $\sum_{n=1}^\infty P(n^{-1/p}W_n \ne T_n) < \infty$, by Markov's inequality it suffices to show that $\sum_{n=1}^\infty |ET_n|$ converges. This is clearly implied by the assumption $E|W_n|^{p+\varepsilon} \le K$, for all n. □

PROOF OF THEOREM 1. Writing

$$V_n^2 - \nu_n^2 = \frac{1}{n} \sum_{i=1}^{n} \Big[\big(X_i - \mu_i + \mu_i - \bar{\mu}_n + \bar{\mu}_n - \bar{X}_n\big)^2 - \sigma_i^2\Big],$$

and using Lemma 1 as well as condition (i), we immediately see that $V_n^2 - \nu_n^2 \to 0$ a.s. Therefore, the result follows if we prove both

$$\|P\big(\sqrt{n}\,(\bar{X}_n - \bar{\mu}_n)/\nu_n \le x\big) - \Phi(x)\|_\infty \to 0$$

and

$$\|P^*\big(\sqrt{n}\,(\bar{Y}_n - \bar{X}_n)/V_n \le x\big) - \Phi(x)\|_\infty \to 0 \quad \text{a.s.},$$

as $n \to \infty$, where $\Phi(\cdot)$ indicates the c.d.f. of a standard normal distribution. The former part is an immediate consequence of the central limit theorem. To show the latter, we first apply the Berry–Esseen theorem to obtain

$$\|P^*\big(\sqrt{n}\,(\overline{Y}_n - \overline{X}_n)/V_n \le x\big) - \Phi(x)\|_\infty \le cV_n^{-3}n^{-1/2}E^*|Y_1 - \overline{X}_n|^3,$$

for some constant $c > 0$. Note that the right-hand side does not exceed $4cV_n^{-3}n^{-1/2}(E^*|Y_1|^3 + |\overline{X}_n|^3)$. Since $E^*|Y_1|^3 = (1/n)\sum_{i=1}^n|X_i|^3$, we only need to show that $n^{-3/2}\sum_{i=1}^n|X_i|^3 \to 0$ a.s. This indeed follows by Lemma 1 if we let $p = 2/3$. □

REMARK. It is apparent from the proof that the condition (i) $(1/n)\sum_{i=1}^n(\mu_i - \bar{\mu}_n)^2 \to 0$ is necessary for the consistency result in Theorem 1. It also seems unlikely that without this condition one can place an asymptotically correct error bound on the estimate $\overline{X}_n$ for $\bar{\mu}_n$.

We now proceed to examine whether the second order asymptotic properties of the classical bootstrap are retained under model (i). Let us set aside the validity of the one-term Edgeworth expansion for a later discussion. If we assume that the expansion does exist, we will see that the bootstrap still corrects the skewness term. To elaborate this point, we consider the nonstudentized and the studentized cases separately. For the nonstudentized case, the formal Edgeworth expansions, if they exist, should be

$$\begin{aligned} &P\big(\sqrt{n}\,(\overline{X}_n - \bar{\mu}_n) \le x\big) \\ &\quad = \Phi\left(\frac{x}{\nu_n}\right) - \frac{\bar{\mu}_{3,n}}{6\nu_n^3\sqrt{n}}\left(\frac{x^2}{\nu_n^2} - 1\right)\phi\left(\frac{x}{\nu_n}\right) + o(n^{-1/2}), \end{aligned} \tag{2.3}$$

$$\begin{aligned} &P^*\big(\sqrt{n}\,(\overline{Y}_n - \overline{X}_n) \le x\big) \\ &\quad = \Phi\left(\frac{x}{V_n}\right) - \frac{\hat{K}_3}{6V_n^3\sqrt{n}}\left(\frac{x^2}{V_n^2} - 1\right)\phi\left(\frac{x}{V_n}\right) + o(n^{-1/2}) \quad \text{a.s.}, \end{aligned} \tag{2.4}$$

where $\bar{\mu}_{3,n} = (1/n)\sum_{i=1}^n E_{G_i}(X_i - \mu_i)^3$, $\hat{K}_3 = (1/n)\sum_{i=1}^n(X_i - \overline{X}_n)^3$ and $\phi(\cdot) = \Phi'(\cdot)$. It is easy to see that the skewness terms are matched if

$$\hat{K}_3 - \bar{\mu}_{3,n} \to 0 \quad \text{a.s. as } n \to \infty. \tag{2.5}$$

This holds if we assume that $E|X_i|^{3+\delta} \le K < \infty$ for some $\delta > 0$ and $n^{-1}\sum_{i=1}^n|\mu_i - \bar{\mu}_n|^3 \to 0$. To prove this, one only needs to incorporate the preceding assumptions into the proof of Theorem 1. (The homogeneity conditions on μ_i in terms of the second and the third moments may appear artificial. In practice they ought to be interpreted as saying that the population means are essentially the same.) Note that in this nonstudentized case the leading terms themselves do not match. This phenomenon of partial $n^{-1/2}$-term correction of the bootstrap has been studied in Liu and Singh (1987).

As far as the studentized case is concerned, the formal Edgeworth expansions, if they exist, should be

$$(2.6)\quad P\left(\sqrt{n}\,\frac{\overline{X}_n - \bar{\mu}_n}{V_n} \le x\right) = \Phi(x) + \frac{\bar{\mu}_{3,n}}{6\nu_n^3\sqrt{n}}(2x^2+1)\phi(x) + o(n^{-1/2})$$

and

$$(2.7)\quad \begin{aligned} &P^*\left(\sqrt{n}\,\frac{\overline{Y}_n - \overline{X}_n}{V_n^*} \le x\right) \\ &\quad = \Phi(x) + \frac{\hat{K}_3}{6V_n^3\sqrt{n}}(2x^2+1)\phi(x) + o(n^{-1/2}) \quad \text{a.s.,} \end{aligned}$$

where $V_n^{*2} = n^{-1}\Sigma_{i=1}^n (Y_i - \overline{Y}_n)^2$, the variance of the bootstrap sample. In analogy with the nonstudentized case, we need condition (2.5) to obtain the correction of skewness term by the bootstrap. Note also that both leading terms in (2.6) and (2.7) are identical. Thus, there is a total $n^{-1/2}$-term correction by the bootstrap in the studentized case. [See Liu and Singh (1987) for "total correction."]

We now turn to the issue of validity of the expansions (2.3), (2.4), (2.6) and (2.7).

Theorem 2. (i) *The expansions* (2.3) *and* (2.4) *hold under conditions* (a), (b) *and* (c):

(a) *There exists a nonlattice distribution H with mean* 0 *and variance* 1, *and a sequence k_n with $k_n/\log n \to \infty$, such that k_n of the population G_i's are of the form $G_i(x) = H((x-\mu_i)/\sigma_i)$ with the σ_i's bounded away from* 0.
(b) $E|X_i|^{3+\delta_0} \le M_1 < \infty$ *for some* $\delta_0 > 0$.
(c) $\liminf_{n\to\infty} \nu_n^2 > 0$ *and* $(1/n)\Sigma_{i=1}^n(\mu_i - \bar{\mu}_n)^2 = o(n^{-1/2})$.

(ii) *The expansions* (2.6) *and* (2.7) *will hold if, in addition to* (a), (b) *and* (c), *we also assume that H is continuous and* $E|X_i|^{6+\delta} \le M_2 < \infty$ *for some* $\delta > 0$.

Proof. The proof for the nonstudentized case (2.3) and (2.4) is along the same lines as the proof for the case of i.i.d. mean given in Feller (1971). In extending the one-term Edgeworth expansion to the non-i.i.d. case, the main difficulty lies in showing that

$$\int_{\varepsilon\sqrt{n}\le|\eta|\le M\sqrt{n}} \frac{1}{\eta}\left|\prod_{j=1}^n \psi_j\left(\frac{\eta}{\sqrt{n}}\right)\right| d\eta = o(n^{-1/2})$$

for any $0 < \varepsilon < M$, where $\psi_j(\eta) = E\exp[i\eta(X_j - \mu_j)]$. Hence it suffices to show that

$$\sup_{\varepsilon\le\eta\le M}\left|\prod_{j=1}^n \psi_j(\eta)\right| = o(n^{-1/2}),$$

which is clearly implied by condition (a).

Similar arguments apply to the bootstrap part (2.4). We only need to show that

$$\sup_{\varepsilon \le \eta \le M} [\psi_n^*(\eta)]^n = o(n^{-1/2}) \quad \text{a.s.,} \tag{2.8}$$

where $\psi_n^*(\eta) = E_{F_n}\{\exp[i\eta(Y_1 - \overline{X}_n)]\}$. Note that by definition

$$\psi_n^*(\eta) = \frac{1}{n} \sum_{j=1}^{n} e^{i\eta(X_j - \overline{X}_n)}.$$

If we assume, wlog, that the first k_n of G_i are of the form $G_i(x) = H((x - \mu_i)/\sigma_i)$, then

$$|\psi_n^*(\eta)| \le \frac{k_n}{n} \left| \frac{1}{k_n} \sum_{j=1}^{k_n} e^{i\eta X_j} \right| + \frac{n - k_n}{n}.$$

Since the first k_n of the σ_i's are bounded away from 0, we have

$$\sup_{\varepsilon \le \eta \le M} \left| \frac{1}{k_n} \sum_{j=1}^{k_n} E_{G_j} e^{i\eta X_j} \right| \le \rho < 1$$

for some $\rho > 0$. Now (2.8) is a consequence of

$$\lim_{n \to \infty} \sup_{\varepsilon \le \eta \le M} \left| \frac{1}{k_n} \sum_{j=1}^{k_n} \left(e^{i\eta X_j} - E_{G_j} e^{i\eta X_j} \right) \right| \le \frac{1 - \rho}{2} \quad \text{a.s.}$$

To establish this, we begin by bounding the supremum by the maximum over n^2 equidistant points with a remainder term of $O(n^{-1})$. We then use an exponential bound together with Bonferroni's inequality to show that the probability for the maximum to exceed $(1 - \rho)/2$ is $n^2 O(e^{-\delta k_n})$ for some $\delta > 0$. Since $k_n/\log n \to \infty$, the result follows from the Borel–Cantelli lemma.

For the studentized case, we follow the procedure in Babu and Singh (1983) for proving the validity of the one-term Edgeworth expansion. The statistic $\sqrt{n}(\overline{X}_n - \bar{\mu}_n)/V_n$ can be written as a function of two normalized means $\sqrt{n}(\overline{X}_n - \bar{\mu}_n)$ and $\sqrt{n}[n^{-1}\sum_{i=1}^{n}(X_i - \mu_i)^2 - \nu_n^2]$, with a remainder term which is negligible for the expansion. The remainder term can be handled using condition (c). The moment condition $E|X_i|^{6+\delta} \le M_2$ is used to guarantee that the $(3 + \delta/2)$th moment of the vector (X_i, X_i^2) be bounded. The continuity of H implies that the vector (X_i, X_i^2) is nonlattice. The extra work needed due to the non-i.i.d. structure is similar to that for the nonstudentized case, and is thus omitted. □

It may be worth mentioning that condition (a) is weaker than the following simple condition: There exists a δ, $0 < \delta < 1$, such that $[\delta n]$ of the underlying distributions $\{G_i\}$ are the same and the common distribution is nonlattice. This is the case of Example 1, where the observations are not identically distributed simply because k different instruments are used for measuring.

We remark here that Section 4A provides an alternative bootstrap procedure in estimating a common mean in contrast to the classical bootstrap procedure discussed in this section.

3. The classical bootstrap on symmetric L-statistics. Our main objective in this section is to assert the following fact without giving complete mathematical details: As far as the first order limit is concerned, bootstrapping symmetric L-statistics when the underlying distributions are symmetric around a common center is equivalent to bootstrapping a sample mean when the underlying populations have a common mean. Besides the examples given in the Introduction, we mention here a natural setup under which the theory for symmetric L-statistics presented in this section is applicable.

EXAMPLE 3. Let

$$T_i = \beta x_i + e_i$$

be a simple regression model, where x_i's are known nonzero numbers and e_i's have a common symmetric distribution centered at 0. It is natural to consider the class of estimators of β consisting of symmetric L-statistics based on the ratios $X_i \equiv T_i/x_i$, $i = 1, \dots, n$. The distributions of X_i's clearly have β as their common center of symmetry but otherwise they are different. The results of this section will establish that the bootstrap which draws i.i.d. samples from the empirical d.f. formed by X_i's is consistent for this class of estimators of β.

The subsequent discussion hinges on Theorem 3, which is a variation of the Bahadur–Kiefer representation of quantiles. We omit the proof of the theorem because it follows the standard techniques used for proving the representation. See, for example, Babu and Singh (1984) for the key steps involved in the proof.

Define, for a c.d.f. H, $H^{-1}(t) = \inf\{x: H(x) \geq t\}$. Let $F_{n,*}$ denote the empirical d.f. based on a bootstrap sample $Y_1, \dots, Y_n$ and recall that F_n is the empirical d.f. based on $X_1, \dots, X_n$.

THEOREM 3. *For some $0 < \alpha < \beta < 1$ and $\varepsilon > 0$, assume that $\overline{G}_n$ is twice differentiable on $[\overline{G}_n^{-1}(\alpha) - \varepsilon, \overline{G}_n^{-1}(\beta) + \varepsilon]$, where both the derivatives are bounded above and the first derivative is bounded away from 0. Let $\bar{g}_{n,t} = \overline{G}_n'(\overline{G}_n^{-1}(t))$. Then*

$$\sup_{\alpha \leq t \leq \beta} \left| F_n^{-1}(t) - \overline{G}_n^{-1}(t) + \left[F_n\left(\overline{G}_n^{-1}(t)\right) - t\right]/\bar{g}_{n,t}\right| = O\left(n^{-3/4}(\log n)^{3/4}\right) \quad a.s.$$

and

$$\sup_{\alpha \leq t \leq \beta} \left| F_{n,*}^{-1}(t) - F_n^{-1}(t) + \left[F_{n,*}\left(\overline{G}_n^{-1}(t)\right) - F_n\left(\overline{G}_n^{-1}(t)\right)\right]/\bar{g}_{n,t}\right|$$

$$= O_p^*\left(n^{-3/4}(\log n)^{3/4}\right) \quad a.s.,$$

where O_p^* stands for O_p under the bootstrap probability.

We would like to point out that unlike in the i.i.d. case, one cannot prove the result for the $U[0,1]$ population and then extend it to a general c.d.f. by the standard quantile transformation since G_i's are allowed to be different here. Therefore, one should directly work with the general G_i's.

Let $\beta = 1 - \alpha$, $\alpha < \frac{1}{2}$ and W be a c.d.f. on $[\alpha, 1-\alpha]$ which is symmetric around $\frac{1}{2}$. Suppose G_i's are all symmetric about a common center θ. If we define

$$L_n = \int_\alpha^{1-\alpha} F_n^{-1}(t)\, dW(t)$$

and

$$L_n^* = \int_\alpha^{1-\alpha} F_{n,*}^{-1}(t)\, dW(t),$$

then Theorem 3 immediately implies

Corollary 1. *Under the conditions given in Theorem* 3,

$$L_n - \theta = \frac{1}{n}\sum_{i=1}^{n} \xi_{\overline{G}_n}(X_i) + O\big(n^{-3/4}(\log n)^{3/4}\big) \quad a.s.$$

and

$$L_n^* - L_n = \frac{1}{n}\sum_{i=1}^{n} \xi_{\overline{G}_n}(Y_i) - \frac{1}{n}\sum_{i=1}^{n} \xi_{\overline{G}_n}(X_i) + O_p^*\big(n^{-3/4}(\log n)^{3/4}\big) \quad a.s.,$$

where

$$\xi_{\overline{G}_n}(X_i) = -\int_\alpha^{1-\alpha} \frac{I_{(X_i \le \overline{G}_n^{-1}(t))} - t}{\bar{g}_{n,t}}\, dW(t).$$

This demonstrates that bootstrapping $L_n - \theta$ is mathematically equivalent to bootstrapping the sample mean $\bar{\xi}_{\overline{G}_n}$, treating $\{\xi_{\overline{G}_n}(X_i)\}$ as the original data instead of $\{X_i\}$ themselves. The crucial point that remains to be checked is

$$E_{G_i}\xi_{\overline{G}_n}(X_i) = 0 \quad \text{for each } i = 1, \dots, n.$$

Note that

$$E_{G_i}\xi_{\overline{G}_n}(X_i) = -\int_{-\infty}^{\infty}\int_\alpha^{1-\alpha} \frac{I_{(x \le \overline{G}_n^{-1}(t))} - t}{\bar{g}_{n,t}}\, dW(t)\, dG_i(x).$$

Interchanging two integrals and then integrating the inner part, we are left with

$$E_{G_i}\xi_{\overline{G}_n}(X_i) = -\int_\alpha^{1-\alpha} \frac{G_i\big(\overline{G}_n^{-1}(t)\big) - t}{\bar{g}_{n,t}}\, dW(t).$$

Now, we break the integral into two parts: $\int_\alpha^{1/2} * + \int_{1/2}^{1-\alpha} *$, and write this as (I) + (II). Note that the assumed symmetry of G_i's around θ implies that $\overline{G}_n$ is also symmetric around θ. Applying the symmetry properties of G_i's, $\overline{G}_n$, $\bar{g}_{n,\cdot}$ and W together, we obtain at once that (I) $= -$(II).

As a direct consequence of Corollary 1, the consistency of the bootstrap for estimating the distribution of $L_n - \theta$ can be stated immediately. The proof is based on the representations in Corollary 1 which clearly establish the equivalence of bootstrapping $L_n - \theta$ and bootstrapping a sample mean.

THEOREM 4. *Under the conditions given in Theorem* 3 *and the condition that* $\liminf_{n\to\infty}(1/n)\sum_{i=1}^n \operatorname{Var}\xi_{\overline{G}_n}(X_i) > 0$, *we have*

$$\lim_{n\to\infty} \|P^*\big(\sqrt{n}\,(L_n^* - L_n) \le x\big) - P\big(\sqrt{n}\,(L_n - \theta) \le x\big)\|_\infty = 0 \quad a.s.$$

There are many commonly used statistics which belong to the class of L-statistics we considered. As examples, we mention the sample median, the hinged mean (the average of the first and the third quartiles), the (symmetrically) trimmed mean $[dW(t) = dt/(1-2\alpha)$ for $\alpha \le t \le 1-\alpha]$ and the Winsorized means $[W(t)$ puts mass α at $t = \alpha$ and $t = 1 - \alpha$ and $dW(t) = dt$ for $\alpha < t < 1 - \alpha]$. It is also worth noting that our theory can be applied in obtaining an adaptive trimmed mean, which selects from finitely many trimming proportions on the basis of estimated s.e.'s (standard errors). In other words, we can use bootstrap to estimate the s.e.'s of several trimmed means with different trimming proportions and select the one with the smallest estimated s.e.

We conclude this section with the remark that the classical bootstrap remains valid under the nonexchangeable models considered here for all location estimators which, after being centered, can be represented as a sample mean, provided that each summand in the sample mean has expectation 0. This includes, for example, M-estimators with symmetric weight functions.

4. Wu's bootstrap and regression problems.

A. *Estimating a common mean.* We begin by using the idea of the weighted bootstrap in Wu's (1986) regression to construct a bootstrap procedure in the context of estimating a common mean (or almost so) from possibly different distributions. Let

$$(+) \qquad Y_i = \overline{X}_n + (X_i - \overline{X}_n)t_i,$$

where $t_1, \dots, t_n$ are i.i.d. random variables with mean 0 and variance 1 and are chosen completely independent of data X_i's. The distribution of X_i has mean μ_i and variance σ_i^2. The μ_i's are "roughly" the same while the σ_i^2's are possibly different. The random variables $Y_1, \dots, Y_n$ form a bootstrap sample. For each i, $i = 1, \dots, n$, Y_i has the same mean $\overline{X}_n$, and the variance $\operatorname{Var}(Y_i) = (X_i - \overline{X}_n)^2$, which in its own right is an estimator of σ_i^2. Thus, the bootstrap sample constructed from $(+)$ in essence reflects the heteroscedasticity of the original data. It is therefore natural to expect that this bootstrap will provide a consistent procedure for estimating the sampling distribution of $\overline{X}_n$ and it will also possess the second order property (one-term Edgeworth correction) of the classical bootstrap, under possibly some further conditions on t_i. More precisely, we state Theorem 5 to establish the consistency of this bootstrap procedure and we examine the $n^{-1/2}$-term correction phenomenon afterward.

THEOREM 5. *If* $E|t_1|^3 < \infty$, $E|X_i|^{2+\delta} \le K < \infty$, *for some* $\delta > 0$ *and* $i = 1, \dots, n$, $\liminf_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_i^2 > 0$ *and* $\lim_{n\to\infty}(1/n)\sum_{i=1}^n(\mu_i - \bar{\mu}_n)^2 = 0$, *then*

$$\lim_{n\to\infty} \|P^*\big(\sqrt{n}\,(\bar{Y}_n - \bar{X}_n) \le x\big) - P\big(\sqrt{n}\,(\bar{X}_n - \bar{\mu}_n) \le x\big)\|_\infty = 0 \quad a.s.$$

The proof is very similar to that of Theorem 1. The condition $E|t_i|^3 < \infty$ can be relaxed to $E|t_i|^{2+\delta} < \infty$, for some $\delta > 0$, by using a more general version of the Berry–Esseen bound.

To study the second order asymptotics of this particular bootstrap procedure, we need to consider $E_t(\sqrt{n}\,(\bar{Y}_n - \bar{X}_n))^3$, where E_t stands for the expectation w.r.t. t_i's treating X_i's as fixed numbers (i.e., the conditional expectation given $X_1, \dots, X_n$). Note that

$$E_t\left[\sum_{i=1}^n (Y_i - \bar{X}_n)\right]^3 = \sum_{i=1}^n (X_i - \bar{X}_n)^3 E t_i^3.$$

Thus, if $Et_i^3 = 1$,

$$E_t\big(\sqrt{n}\,(\bar{Y}_n - \bar{X}_n)\big)^3 = n^{-1/2}\left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^3\right],$$

which is $n^{-1/2}n^{-1}\sum_{i=1}^n \mu_{3,i} + o(n^{-1/2})$ assuming that $E|X_i|^{3+\delta} \le K < \infty$ for some $\delta > 0$ and for all i. This last expression is equal to $E[n^{1/2}\,(\bar{X}_n - \bar{\mu}_n)]^3 + o(n^{-1/2})$ if $n^{-1}\sum_{i=1}^n|\mu_i - \bar{\mu}_n|^3 \to 0$. Therefore the skewness term in the formal Edgeworth expansions will match if the t_i's have third central moment equal to 1, besides having mean 0 and variance 1. Moreover, similar calculations show that if $Et_i^3 = 1$, then the first three cumulants of the studentized statistics $\sqrt{n}\,(X_n - \bar{\mu}_n)/V_n$ and $\sqrt{n}\,(\bar{Y}_n - \bar{X}_n)/V_n^*$ (conditional on X_i), where $V_n^* = [n^{-1}\sum_{i=1}^n(Y_i - \bar{Y}_n)^2]^{1/2}$ match up to $o(n^{-1/2})$. Consequently, there is a total $n^{-1/2}$-term corrections by the bootstrap in this case. We keep the discussion brief here because similar phenomena have been discussed in detail in Section 2.

Regarding the $n^{-1/2}$-term correction by the bootstrap, the $n^{-1/2}$-term is automatically missing in the expansion if the original population is assumed to be symmetric. Thus we would like to correct the next term, namely the n^{-1}-term, in the expansion of the studentized mean. Similar calculations reveal that the condition for this is to require the first four moments of t_i to be 0, 1, 0 and 1, respectively. The only distribution that satisfies this requirement is the two points distribution $P(t_1 = 1) = P(t_1 = -1) = \frac{1}{2}$, which is a lattice distribution. This choice of t_i makes even the existence of one-term Edgeworth expansion for the statistics based on Y_i's unlikely.

The issue of the validity of Edgeworth expansions for the studentized or nonstudentized statistics based on $\bar{Y}_n$ itself poses interesting probabilistic questions. We focus on the nonstudentized case, i.e., $\sqrt{n}\,(\bar{Y}_n - \bar{X}_n)$. In view of the arguments in Feller [(1971), Chapter 16, Section 4], we know the main difficulty is to establish that

$$\sup_{\varepsilon \le \eta \le M}\left|\prod_{j=1}^n E_t \exp\big(i\eta\big(X_j - \bar{X}_n\big)t_j\big)\right| = o(n^{-1/2}) \quad \text{a.s.} \tag{4.1}$$

Suppose that the t_j's satisfy Cramér's condition, so that

$$\limsup_{|\eta|\to\infty} Ee^{i\eta t_1} < 1. \tag{4.2}$$

(Note that this condition makes it unnecessary to worry about the cases when $|X_i - \overline{X}_n|$ gets too large.) Then it suffices to have k_n of the $|X_i - \overline{X}_n|$'s be greater than some positive number δ_0 where $k_n/\log n \to \infty$. Assume $E|X_i|^{4+\delta} \le K$ for all i. If $\liminf_{n\to\infty}(1/n)\sum_{i=1}^n \sigma_i^2 > 0$, then $\liminf_{n\to\infty}(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = m$ a.s. for some $m > 0$. Using the Markov and Bonferroni inequalities, we conclude that

$$P\Big(\max_{1\le i\le n}|X_i| > \sqrt{n}/(\log n)\Big) = O(n^{-1-\delta/2}).$$

Thus we may assume that for a given sequence of X_i's, $\max_{1\le i\le n}|X_i - \overline{X}_n| < 2\sqrt{n}/(\log n)$ for all large n. Suppose l_n of the $|X_i - \overline{X}_n|$'s are less than δ_0. Then

$$l_n\delta_0^2 + 4(n - l_n)n/(\log n)^2 > n(m/2)$$

for all large n a.s. This inequality implies

$$(n - l_n) > \text{constant}\cdot(\log n)^2,$$

which guarantees that $k_n/\log n \to \infty$ (a.s.). This observation together with condition (4.2) amounts to condition (4.1).

B. *Regression.* Consider now the simple linear model $Y_i = \beta x_i + e_i$, where x_i's are nonzero real numbers, $Ee_i = 0$, $\mathrm{Var}(e_i) = \sigma_i^2$ and e_i's are independent. The least square estimate of β is $\hat\beta = (\sum_{i=1}^n x_iY_i)/\sum_{j=1}^n x_j^2$. Clearly $\mathrm{Var}(\hat\beta) = \sum_{i=1}^n x_i^2\sigma_i^2/(\sum_{j=1}^n x_j^2)^2$. Let $r_i = Y_i - x_i\hat\beta$ be the residuals. The classical bootstrap sample is $Y_i^* = x_i\hat\beta + r_i^*$ for $i = 1,\dots,n$, where $r_1^*,\dots,r_n^*$ is a random sample from the empirical d.f. based on $(r_1 - \bar r_n),\dots,(r_n - \bar r_n)$. If $\hat\beta_b$ denotes the least square estimate based on Y_i^*'s, then the bootstrap variance of $\hat\beta_b$, $\mathrm{Var}^*(\hat\beta_b)$, is $n^{-1}\sum_{i=1}^n(r_i - \bar r_n)^2/\sum_{j=1}^n x_j^2$ which is equivalent to $(n^{-1}\sum_{i=1}^n\sigma_i^2)/(\sum_{j=1}^n x_j^2)$ asymptotically. Thus the classical bootstrap does not provide a consistent estimator for the s.e. of $\hat\beta$ if the error variances are allowed to be different. However, it is not difficult to modify this bootstrap to achieve the consistency. Let us form the empirical d.f. on $\{(x_i/(\overline{x_n^2})^{1/2})(r_i - \bar r_n)\}$ instead of just $\{r_i - \bar r_n\}$ [recall that $\overline{x_n^2} = (1/n)\sum_{i=1}^n x_i^2$] and let $\tilde\beta_b$ denote the resulting bootstrap least square estimator of $\hat\beta$. Then, under the conditions that x_i's are bounded and $\liminf_{n\to\infty} n^{-1}\sum_{i=1}^n x_i^2 > 0$ (i.e., $\sum_{i=1}^n x_i^2$ grows roughly like n),

$$\mathrm{Var}^{**}(\tilde\beta_b) = \frac{\sum_{i=1}^n x_i^2 r_i^2}{\left(\sum_{j=1}^n x_j^2\right)^2} - \frac{n^{-1}(\sum_{i=1}^n x_i r_i)^2}{\left(\sum_{j=1}^n x_j^2\right)^2}$$

(** stands for the bootstrap probability under the weighted empirical d.f.), which is equal to

$$\frac{\sum_{i=1}^n x_i^2\sigma_i^2}{\left(\sum_{j=1}^n x_j^2\right)^2} + O_p(n^{-3/2}).$$

In fact, one can center the weighted empirical before bootstrapping in order to make the bootstrap variance exactly equal to $\Sigma_{i=1}^n x_i^2 r_i^2/(\Sigma_{j=1}^n x_j^2)^2$ [= (I), say]. A natural pivot related to $\hat{\beta}$ is $T_n = (\hat{\beta} - \beta)/(\mathrm{I})^{1/2}$. Denote the bootstrap counterpart of T_n to be $T_n^* = (\tilde{\beta}_b - \hat{\beta})/(\mathrm{II})^{1/2}$, where $(\mathrm{II}) = \Sigma_{i=1}^n x_i^2 r_i^{*2}/(\Sigma_{i=1}^n x_i^2)^2$ and $\{r_i^*\}_{i=1}^n$ are i.i.d. samples from $\{(x_i/(\overline{x_n^2})^{1/2})(r_i - \bar{r}_n)\}$'s. By computing the first three cumulants of T_n and T_n^*, we can show that the bootstrap will correct the skewness term (i.e., the $n^{-1/2}$-term) in the Edgeworth expansion of the sampling distribution of T_n.

As for Wu's bootstrap on this simple regression, we assert that under this weighting scheme, the bootstrap sample is

$$Y_i^*, \dots, Y_n^*, \qquad Y_i^* = x_i\hat{\beta} + r_i t_i, \tag{4.3}$$

where $\{t_i\}_1^n$ are i.i.d. with $Et_i = 0$ and $\operatorname{Var} t_i = 1$, for all i. Note that in Wu (1986), r_i is multiplied with an adjusting factor which is equal to 1 asymptotically. Let $\breve{\beta}_b$ indicate the least square estimator of $\hat{\beta}$ under the bootstrap (4.3). Then

$$\operatorname{Var}_t(\breve{\beta}_b) = \frac{\Sigma_{i=1}^n x_i^2 r_i^2 E t_i^2}{\left(\Sigma_{j=1}^n x_j^2\right)^2} = \frac{\Sigma_{i=1}^n x_i^2 \sigma_i^2}{\left(\Sigma_{j=1}^n x_j^2\right)^2} + O_p(n^{-3/2})$$

since $\operatorname{Var} t_i = 1$. Hence $Et_i = 0$ and $\operatorname{Var}(t_i) = 1$ for all i are sufficient for proving the consistency of the bootstrap. If we also assume that $Et_i^3 = 1$, then the third moments of $\sqrt{n}(\hat{\beta} - \beta)$ and all first three moments of T_n will be estimated correctly up to $O(n^{-1})$ by this bootstrap.

Suppose one is interested in the sampling distribution of the least square estimator of a certain linear combination of $\boldsymbol{\beta}$, say $\mathbf{l}'\boldsymbol{\beta}$, of a general linear model $\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e}$. Here X is a $n \times p$ matrix, $\boldsymbol{\beta}$ and $\mathbf{l}$ are $p \times 1$ vectors and $\mathbf{e}$ is a $n \times 1$ vector. The idea of bootstrapping from the empirical d.f. based on weighted residuals extends readily. It is so because the least square estimator $\mathbf{l}'\hat{\boldsymbol{\beta}}$ can be expressed in a form of $\Sigma_{i=1}^n w_i Y_i$, where the vector $\mathbf{w} = (w_1, \dots, w_n) = \mathbf{l}'(X'X)^{-1}X'$. Using the same idea of weighted empirical d.f. previously mentioned, we only need to multiply the residuals $(r_i - \bar{r}_n)$ with $w_i/(\overline{w_n^2})^{1/2}$ before forming the empirical d.f. However, the main drawback is that the weights and thus the whole procedure depend on $\mathbf{l}$. It is not known to us at this point if any single weight vector can be so chosen as to work for all $\mathbf{l}$. On the other hand, Wu's bootstrap, different from the classical one [see, e.g., Freedman (1981) on bootstrapping regression models], does provide a consistent estimator for the standard error of $\mathbf{l}'\hat{\boldsymbol{\beta}}$ for all $\mathbf{l}$. In this bootstrap, the bootstrap sample is

$$\mathbf{Y}^* = X\hat{\boldsymbol{\beta}} + \mathbf{e}^*. \tag{4.4}$$

Here $\mathbf{e}^* = (e_1^*, \dots, e_n^*)$ and $e_i^* = r_i t_i$, where the t_i's are i.i.d. random variables with mean 0 and variance 1. Let $\tilde{\boldsymbol{\beta}}_b$ represent the least square estimate of $\hat{\boldsymbol{\beta}}$ under the bootstrap of (4.4). Assume that the elements of X are bounded and the column of X have length $O(n)$. Then we have

$$\operatorname{Var}_t(\mathbf{l}'\tilde{\boldsymbol{\beta}}_b - \mathbf{l}'\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n w_i^2 r_i^2 E t_i^2,$$

which is $\sum_{i=1}^{n} w_i^2 \sigma_i^2 + o(1)$ since $Er_i^2 = \sigma_i^2 + o(1)$ and $Et_i^2 = 1$, for all i. Furthermore,

$$E_t(\mathbf{l}'\tilde{\boldsymbol{\beta}}_b - \mathbf{l}'\hat{\boldsymbol{\beta}})^3 = \sum_{i=1}^{n} w_i^3 r_i^3$$

provided $Et_i^3 = 1$ and

$$\sum_{i=1}^{n} w_i^3 r_i^3 = \sum_{i=1}^{n} w_i^3 \mu_{3,i} + o(1),$$

where $\mu_{3,i} = Ee_i^3$. Therefore, the condition $Et_i^3 = 1$ will suffice to correct the skewness term in the Edgeworth expansion of the sampling distribution of $\mathbf{l}'\hat{\boldsymbol{\beta}}$.

In concluding this section we reemphasize that, in addition to the conditions $Et_i = 0$ and $Et_i^2 = 1$, the key condition $Et_i^3 = 1$ is needed in order to obtain the second order properties of Wu's bootstrap. Two examples satisfying these conditions will now be given.

EXAMPLE 4. $t_i = D_i - ED_i$, $i = 1, \ldots, n$, and $D_1, \ldots, D_n$ are i.i.d. with gamma distribution having density $g_D(x) = [\alpha^\beta/(\beta - 1)!]x^{\beta-1}e^{-\alpha x}I_{(x>0)}$, where $\alpha = 2$ and $\beta = 4$.

EXAMPLE 5. $t_i = W_i Z_i - (EW_i)(EZ_i)$, $i = 1, \ldots, n$, where $W_1, \ldots, W_n$ are i.i.d. normally distributed with mean $\frac{1}{2}(\sqrt{17/6} + \sqrt{1/6})$ and variance $\frac{1}{2}$, $Z_1, \ldots, Z_n$ are i.i.d. normally distributed with mean $\frac{1}{2}(\sqrt{17/6} - \sqrt{1/6})$ and variance $\frac{1}{2}$, and W_i's and Z_i's are independent.

Acknowledgments. I am very grateful to Kesar Singh for his encouragement and helpful discussions on this work. I would also like to thank the referee and the Associate Editor for their valuable comments.

REFERENCES

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BABU, G. J. and SINGH, K. (1984). Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhyā Ser. A* **46** 195–206.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

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

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

LIU, R. and SINGH, K. (1987). On a partial correction by the bootstrap. *Ann. Statist.* **15** 1713–1718.

LIU, R., SINGH, K. and LO, S. H. (1986). On a representation related to the bootstrap. *Sankhyā Ser. A*. To appear.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1295.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903