This is to certify that the

dissertation entitled

Bootstrap approximation to the

distributions of M-estimators

presented by

Soumendra Nath Lahiri

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

6-29-89

Major professor

Date ____________________

BOOTSTRAP APPROXIMATIONS TO THE DISTRIBUTIONS OF M – ESTIMATORS.

By

Soumendra Nath Lahiri

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and probability

1989

ABSTRACT

BOOTSTRAP APPROXIMATIONS TO THE DISTRIBUTIONS OF M – ESTIMATORS.

By

Soumendra Nath Lahiri

Consider the linear regression model $Y_i = x_i\beta + \epsilon_i$, where ϵ_i's are random variables with common distribution F and x_i's are known constants. Let $\bar{\beta}_n$ be the M - estimator of β corresponding to a nondecreasing, bounded score function ψ. This thesis analyzes the asymptotic behaviors of certain bootstrap approximations to the distribution of normalized $\bar{\beta}_n$. It is shown that the ordinary bootstrap procedure as such does not work in the present set up. As remedies, several modifications of this procedure have been rmulated. For studying the asymptotic behaviors of these procedures, Edgeworth expansions of the distributions of $\bar{\beta}_n$ and the modified bootstrap estimators are obtained. It is proved that all the proposed modifications lead to a faster rate of approximation, viz. $o\left(\text{Max} \left\{ |x_j| / \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} : 1 \le j \le n \right\} \right)$ than the usual normal approximation. For the special case, when the score function ψ is odd and the underlying error distribution F is smooth and symmetric, it is observed that by taking the resampling distribution to be a suitable symmetrized kernel estimator of F, one can have even a higher rate of approximation, namely $o\left(\text{Max} \left\{ |x_j|^2 / \sum_{i=1}^{n} x_i^2 : 1 \le j \le n \right\} \right)$.

Second part of the thesis considers the bootstrap approximations to the distributions of M - estimators in a multivariate setting under a different model.

Let $X_1, \ldots, X_n$ be independent and identically distributed k – dimensional random vectors with common distribution F_θ, $\theta \in \Theta \subset \mathbb{R}^p$ for some $p \geq 1$. Let ψ be a function from $\mathbb{R}^k \times \mathbb{R}^p$ into $\mathbb{R}^p$ and $\hat{\theta}_n$ be the M - estimator of θ corresponding to ψ. Under some regularity conditions on ψ, an Edgeworth expansion of the bootstrapped M - estimator is proved. Using this and the Edgeworth expansion for $\hat{\theta}_n$ (obtained by Bhattacharya and Ghosh (1978) : 'On The Validity of Formal Edgeworth expansion.', Ann. Statist. 6, 434 – 445), the rate of bootstrap approximation is shown to be $o(n^{-1/2})$. This extends a result of Singh (1981) ('On the Accuracy Of Efron's Bootstrap.', Ann. Statist. 9, 1187 – 1195) about the sample mean to the M–estimators.

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis adviser, Professor Hira Lal Koul for all the helpful advice, guidance and encouragement I have received over last two years. I am thankful to Professor H. L. Koul and Professor R. Lepage for their constructive suggestions that significantly improved my unprofessional presentation of this work and brought it up to the present form.

I would also like to thank Professors R. V. Erickson, R. Lepage and H. Salehi for serving on my committee. Finally, I want to thank Ms. Loretta Ferguson for helping me with the typing of this manuscript.

TABLE OF CONTENTS

INTRODUCTION

Ever since introduction, the bootstrap method has found its applications in a variety of statistical problems and in most of the cases with overwhelming success. The superiority of the bootstrap approximation in certain estimation problems has been reported in the introductory paper, Efron (1979), on the basis of some numerical studies. Soon these empirical results were substantiated from the theoretical standpoint by Singh (1981), Bickel and Freedman (1981), Beran (1982), Babu and Singh (1984) and Hall (1988) among others. In fact, it was Singh (1981) who showed for the first time that the rate of bootstrap approximation to the distribution of the normalized sample mean is faster than the usual large sample normal approximation. He derived an almost sure Edgeworth Expansion for the distribution of the bootstrapped statistic and compared it with the standard Edgeworth expansion for the distribution of normalized sample mean to arrive at the conclusion. It became clear from this work that the distribution of the bootstrapped statistic corrects itself for the possible skewness of the underlying distribution and thus povides a better approximation than the normal law. Subsequently similar results on the rate of bootstrap approximation have been established in a number of cases when the statistic of interest is a smooth functional of the underlying distribution. See Babu and Singh (1983) for results on studentized k–sample means, Helmers (1988) for results on U–statistics, Bose (1988) for bootstrapping an autoregression model.

In this thesis we shall consider the behaviour of bootsrtrap approximation to the distributions of M–estimators in two different problems. The first problem concerns a simple linear regression model

$$Y_i = x_i \beta + \epsilon_i, \quad i = 1,\dots,n$$

where ϵ_i's are independent with a common distribution F and x_i's are known, nonrandom constants. Here the model differs from the others mentioned earlier (except Bose (1988)) at the point that the observed values $Y_1,....,Y_n$ are not identically distributed. Bootstrap approximation in similar nonidentical set up has been considered in Freedman (1981), Bickel and Freedman (1983) and Liu (1988). The first two papers prove the bootstrap central limit theorem for the least square estimators of the multiple regression parameters and Liu (1988) establishes the second order correctness of the bootstrap method for the sample mean of independent but not necessarily identical observations. Here we consider bootsrapping the M–estimator $\bar{\beta}_n$ of β corresponding to a nondecreasing, bounded score function ψ (see Section 1.1 of Chapter 1 for definition). Under certain smoothness conditions on ψ and/or F, an Edgeworth expansion for the distribution of normalized $\bar{\beta}_n$ has been obtained. This result is of independent interest for two reasons. First, such expansions for the M– estimators in the general regression context were not known earlier ; second, the method of proof is somewhat different from the conventional approach (cf. Ringland (1983)) based on Bhattacharya and Ghosh (1978).

Bootstrapping $\bar{\beta}_n$ under the present model leads to some intriguing phenomena. In Section 1.3 of Chapter 1, we give an example which shows that the usual bootstrap procedure does not work in the present set up. The bootstrapped statistic in the example does not even converge to the limiting distribution of the unbootstrapped statistic. To overcome this drawback of the usual bootstrap procedure, we propose different modifications and show that each of these modifications actually attains a faster rate than the normal approximation.

In the second problem, we consider the M–estimators of a higher dimensional parameter in a multivariate setting. The Edgeworth expansion of the normalized M–estimator was obtained by Bhattacharya and Ghosh (1978) and Bhattacharya (1985) under some smoothness conditions on the score function ψ. Here we follow the usual bootstrap procedure and select the bootstrap samples from the empirical distribution of the observations. Using the smoothness of ψ and a result of Babu and Singh (1984), we obtain an almost sure expansion of the distribution of the bootstrapped statistic along the line of Bhattacharya and Ghosh (1978). Comparison of these two expansions establishes the superiority of the bootstrap approximation to the normal approximation. This extends a result (part (d) of Theorem 1) of Singh (1981) about the sample mean to the M–estimators.

CHAPTER 1

1.1. Introduction.

Consider a simple linear regression model

$$(1.1) \qquad Y_i = x_i \beta + \epsilon_i , \qquad i=1,\dots,n,$$

where $\epsilon_1,\dots,\epsilon_n$ are independent and identically distributed (i.i.d.) random variables (r.v.'s) with common distribution function (d.f.) F and where $x_1,\dots,x_n$ are known nonrandom constants. Let ψ be a nondecreasing and bounded function from $\mathbb{R}$ into $\mathbb{R}$. Define an estimator $\overline{\beta}_n$ of β to be a solution of the equation (in t)

$$(1.2) \qquad \sum_{i=1}^{n} x_i \, \psi (Y_i - x_i t) = 0.$$

Estimators $\{\overline{\beta}_n\}$ are known as M–estimators of β (Huber : 1973 , 1981). Assume that

$$(1.3) \qquad E \, \psi (\epsilon_1) = 0.$$

The condition (1.3) ensures the asymptotic unbiasedness of $\overline{\beta}_n$. For easy reference later on, let

$$(1.4) \qquad a_n^2 = \sum_{i=1}^{n} x_i^2 \text{ and } M_n = \text{Max} \{ \, |x_i| : 1 \le i \le n \, \}.$$

The asymptotic normality of $a_n(\overline{\beta}_n - \beta)$ has been studied extensively in the literature under much more general settings : see Huber (1973, 1981) and the references therein. Relatively very little is known about the Edgeworth expansions for the distributions of these estimators , specially when the score function ψ is not smooth. Ringland (1983) considered the one–way layout model with p populations ($p^3/n \to \infty$) and obtained a two–term Edgeworth expansion for the distribution of studentized M–estimators. His method of proof was along the line of Bhattacharya and Ghosh (1978). In particular, he required the score function ψ to be smooth and the design matrix elements to be 0's and 1's only. When $p = 1$, this forces $x_i = 1$ for all i which is too restrictive in the regression context .

For the one parameter case this paper gives an Edgeworth expansion of the distribution of $a_n(\bar{\beta}_n - \beta)$ when ψ is not necessarily smooth and the constants x_i's satisfy only some mild growth conditions. The method of proof is completely different from that of Ringland (1983). Monotonicity of ψ enables one to obtain bounds on the probabilities involving $\bar{\beta}_n$ in terms of the probabilities relating to the sums of independent random variables. Thus one can apply the classical Edgeworth expansion techniques to these bounds for obtaining an approximate expansion of the distribution of $\bar{\beta}_n$. Then, the smoothness of ψ and/or F is used to simplify these expressions into the stated forms.

BOOTSTRAPPING $\bar{\beta}_n$: In order to describe the bootstrapping of $\bar{\beta}_n$, let F_n be an estimator of the underlying error d.f. F based on the estimated residuals $\bar{\epsilon}_i = Y_i - x_i\bar{\beta}_n$, $i = 1,...,n$. Also let $\epsilon_1^*,..., \epsilon_n^*$ be a bootstrap sample from F_n and define $Y_i^* = x_i \bar{\beta}_n + \epsilon_i^*$ for $i = 1 ,..., n$. In accordance with (1.2), the bootstrap estimator β_n^* of β is defined as a solution of the equation (in t)

$$\sum_{i=1}^{n} x_i \psi (Y_i^* - x_i t) = 0. \tag{1.5}$$

The role played by β in the original problem is to be replaced by $\bar{\beta}_n$ in the bootstrap set up. Accordingly one should have

$$E_n \psi(\epsilon_1^*) = E_n \psi(Y_1^* - x_1\bar{\beta}_n) = 0. \tag{1.6}$$

where E_n denotes the expectation under F_n. In general, the choice of F_n that will satisfy condition (1.6) and at the same time be a good estimator of F seems to depend heavily on the forms of F and ψ. In the case of bootstrapping the least square estimator $\hat{\beta}_n$ of β, the corresponding requirement is $E_n(\epsilon_1^*) = 0$. Freedman (1981) considered the problem of bootstrapping $\hat{\beta}_n$ and ensured this condition by centering the estimated residuals $\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n$ and then taking the bootstrap samples from the empirical distribution of these centered values. In fact, he has pointed out that if one does not center the estimated residuals, the distribution of $a_n(\beta_n^* - \hat{\beta}_n)$ does not converge to that of $a_n(\hat{\beta}_n - \beta)$. Similar remark applies to the present context as well. We give an example at the beginning of Section 1.3 where (1.6) does not hold and $a_n(\beta_n^* - \bar{\beta}_n)$ does not have the same limiting distribution as the unbootstrapped statistic $a_n(\bar{\beta}_n - \beta)$. Therefore, one should consider only those F_n's for which condition (1.6) is satisfied.

Clearly, (1.6) is not satisfied for general design points if $\bar{\beta}_n$ is defined by (1.5) and F_n is taken to be the empirical distribution function (e.d.f.) of the estimated residuals $\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n$. Therefore, one has to look for appropriate modifications, if any, of the usual bootstrap procedure. In fact, there are at least two ways of attaining this. One is to change the resampling distribution and the other is to change the defining equation (1.5). As an example of the first possibility, F_n is taken to be a suitable weighted empirical distribution and β_n^* is defined as a solution of (1.5) (see Section 1.3 for details). As an example of the other case, equation (1.5) is modified according to Shorack (1982) and β_n^* is defined as a solution of the resulting equation (cf. equation (3.8) in Section 1.3). In the second case, it is shown that one can take F_n to be either the e.d.f. of the estimated residuals $\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n$ or some smoother estimator of F depending on the degree of smoothness of ψ and F. With either modification the resulting bootstrap procedure corrects one term in the Edgeworth expansion of the distribution of the normalised $\bar{\beta}_n$ and the rate of bootstrap approximation becomes $o(M_n/a_n)$.

Finally, in the case when the error d.f. F is smooth and symmetric and the score function ψ is odd, the rate of bootstrap approximation corresponding to a symmetrized kernel density estimator of F is shown to be $o((M_n/a_n)^2)$. This result is similar to a result of Babu and Singh (1984) about the sample mean where the the resampling distribution is taken to be the symmetrized e.d.f. of the observations centered about the sample mean. In a nut shell, for all the cases considered here bootstrap approximation is shown to have a better rate than the normal approximation .

The layout of this chapter is as follows. Section 1.2 contains theorems giving the Edgeworth expansions for $\bar{\beta}_n$. Section 1.3 deals with the bootstrap approximations to the distribution of $\bar{\beta}_n$ and Section 1.4 contains the proofs of the results stated in Sections 1.2 and 1.3.

1.2. **Edgeworth expansions for $\bar{\beta}_n$.**

This section gives the Edgeworth expansion for the distribution of normalized $\bar{\beta}_n$ under some assumptions on ψ, F and x_i's. Parts of these assumptions are on the underlying model (1.1) and will be assumed throughout the paper without explicit reference. The rest of the assumptions are required for the validity of some results in this section. Whenever used, one or more of these will always be mentioned in the statement of the corresponding assertion. Before stating the assumptions, we need to fix some notation. For x real, write

$$\mu(x) = E\,\psi(\epsilon_1 - x), \qquad V(x) = \sigma^2(x) = \mathrm{Var}\,\psi(\epsilon_1 - x),$$

$$\mu_3(x) = E\,(\psi(\epsilon_1 - x) - \mu(x))^3 \quad \text{and} \quad \mu_4(x) = E\,(\psi(\epsilon_1 - x) - \mu(x))^4.$$

Since ψ is bounded all these quantities are well defined.

For any real valued function h defined on $\mathbb{R}$, let $h^{(i)}$ denote the i–th derivative of h whenever it exists and $\|h\|$ denote the supremum norm of h. For convenience, h', h'', h''' will replace $h^{(1)}$, $h^{(2)}$ and $h^{(3)}$ respectively. Define

$$A = -\mu'(0)/\sigma(0), \quad d_{1n} = \sum_{i=1}^{n} x_i^3 / a_n^3, \quad d_{2n} = \sum_{i=1}^{n} x_i^4 / a_n^4,$$

$$\text{(2.1)} \qquad d_{3n} = \sum_{i=1}^{n} |x_i^3| / a_n^3 \quad \text{and} \quad d_{4n} = \mathrm{Max}\,\{ d_{3n}^2, d_{2n} \}.$$

Next recall the definition of M_n and a_n from (1.4). Note that $d_{3n} = O\,(M_n/a_n)$ and $d_{4n} = O\,(M_n^2/a_n^2)$. Let $b_n = \log a_n$ (whenever it is defined). For $c > 0$, define the set $A_n(c) = \{ i : 1 \leq i \leq n, \ |x_i| > c.M_n\}$ and let $k_n(c)$ denote the number of elements in $A_n(c)$.

In addition to (1.3), assume that conditions (A.1) – (A.3) below are satisfied by the underlying model (1.1) throughout this chapter.

(A.1) : $a_n \to \infty$ as $n \to \infty$.

(A.2) : $A = -\mu'(0)/\ \sigma(0) > 0$ (Whenever it exists).

(A.3) : There exists a constant c , $0 < c < 1$ such that $b_n = o\ (k_n(c))$ as $n \to \infty$.

Next, we list down the remaining assumptions used in this section.

(A.4) : $b_n^4\, M_n = o\ (a_n)$ as $n \to \infty$.

(A.5) : $b_n^6\, M_n = o\ (a_n)$ as $n \to \infty$.

(A.6) : There exist constants $M > 0$, $\delta > 0$ and $0 < q < 1$ such that

$$\sup \{ \ |E \exp(it\ \psi(\epsilon_1 - x))| : |x| < \delta \text{ and } |t| > M \} < q.$$

REMARK 2.1 : First two assumptions are typical for proving the asymptotic normality of $\overline{\beta}_n$ and occur frequently in the literature (see, for example, Huber : 1973,1981). Assumption (A.3) is rather uncommon and deserves some clarifications. For obtaining the Edgeworth expansions of normalized sums of independent r.v.'s, one usually assumes that the absolute values of the characteristic functions of all the summands are uniformly bounded away from 1 outside every neighbourhood of zero. But in the present context, this will require $\min \{ \ |x_i| : i \geq 1 \ \} > c$ for some constant $c > 0$ which will rule out many frequently used designs. Condition (A.3) relaxes this requirement on x_i's. Another typical assumption made for proving the asymptotic normality of $\overline{\beta}_n$ is that $M_n/a_n = o\ (1)$ as $n \to \infty$. Condition (A.4) and (A.5) are somewhat stronger versions of this and are required for obtaining the Edgeworth expansions upto the desired orders. Note that both the conditions are trivially satisfied for bounded x_i's as well as for $x_i \equiv i$. Condition (A.6) is actually a modified Cramer's condition (see Bhattacharya and Rao (1976), page 207 for the statement of Cramer's condition) and will be used for obtaining higher order expansions. See Remark 2.4 and the Proposition following it for a sufficient condition.

Before stating the theorems, we put down the explicit form of the Edgeworth expansions. To that effect, write H_i for the Hermite polynomial of degree i, $i \geq 1$ (see Feller (1966), page 514). Let ϕ and Φ respectively denote the density and the d.f. of a standard normal r.v.. For Theorems 2.1 and 2.2, define

$$H_{1n}(x) = \Phi(x) - d_{1n}\,[(\mu''(0)/\sigma(0) \;-\; \mu'(0)\,V'(0)/\sigma^3(0))\,x^2/\,2A^2$$

$$+ (\,\mu_3(0)\,/\,6\sigma^3(0)\,)\,H_2(x)\,]\,\phi(x)$$

$$H_{2n}(x) = H_{1n}(x) - \phi(x)\,[d_{2n}\{\,(\mu'''(0)/\sigma(0) + 3\,A\,V''(0)/\,2\,V(0)\,)\,x^3/6A^3$$

$$+ (\,\mu_3'(0)\,/\,6A\sigma^3(0)\,)\,x\,H_2(x) + ((\,\mu_4(0) - 3\sigma^4(0))\,/\,24\sigma^4(0))\,H_3(x)\,\}$$

$$+ d_{1n}^{\;2}\,\{(\mu''(0)/\sigma(0) + A\,V'(0)/\,V(0)\,)\,(\,\mu_3(0)/12A^2\,\sigma^3(0)\,)\,x^2H_3(x)$$

$$+ (\,\mu_3^{\,2}(0)/\,72\sigma^6(0)\,)\,H_5(x) + (\mu''(0)/\sigma(0) + A\,V'(0)/\,V(0))^2\,x^5/\,8A^4$$

$$- (\mu''(0)\,V'(0)/\sigma^3(0) + 3AV'^2(0)/2V^2(0))x^3/4A^2$$

$$- (\mu_3(0)\,V'(0)/4A\sigma^5(0))\,xH_2(x)\}].$$

Under the hypotheses of the following Theorems , the functions μ, V, μ_3, μ_4 have sufficiently many derivatives so that H_{1n} and H_{2n} are well defined. Now we state the theorems of this section.

THEOREM 2.1 : *Suppose that* ψ *has a uniformly continuous, bounded second derivative.* (a) *If* $\psi(\epsilon_1)$ *is nonlattice and condition* (A.4) *holds, then*

$$\sup_x | P(a_n(\beta_n - \beta) \leq x) - H_{1n}(Ax) | = o (M_n/a_n).$$

(b) *Suppose that* ψ *has a uniformly continuous, bounded third derivative. If, in addition, conditions* (A.5) *and* (A.6) *hold, then*

$$\sup_x | P(a_n(\beta_n - \beta) \leq x) - H_{2n}(Ax) | = o (d_{4n}) = o (M_n^2 / a_n^2)$$

where d_{4n} *is as defined in* (2.1).

Next we state a version of Theorem 1 under the corresponding regularity conditions on F without assuming the differentiability of ψ.

THEOREM 2.2 : *Assume that* F *has a uniformly continuous, density* f.

(a) *If* $\psi(\epsilon_1)$ *is nonlattice and condition* (A.4) *holds, then*

$$\sup_x | P(a_n(\beta_n - \beta) \leq x) - H_{1n}(Ax) | = o (M_n/a_n).$$

(b) *Suppose that* f *has a uniformly continuous, bounded second derivative. If, in addition, conditions* (A.5) *and* (A.6) *hold, then*

$$\sup_x | P(a_n(\beta_n - \beta) \leq x) - H_{2n}(Ax) | = o (d_{4n}) = o (M_n^2 / a_n^2).$$

REMARK 2.2 : The same technique can be used to obtain higher order expansions under stronger smoothness conditions on ψ and/or F. The corresponding expressions become more and more messy as the order of expansion increases.

REMARK 2.3 : In Theorems 2.1 and 2.2, the smoothness conditions on ψ and/or F has been used to guarantee that the functions μ, V, m_3, m_4 etc. have sufficiently many derivatives. In fact it is possible to achieve the same results by varying the degree of smoothness on ψ and F. In practice one often encounters score functions ψ which are sufficiently smooth except possibly at a finite number of points. It can be shown that if F is well behaving in some neighbourhoods of these points, then also the above expansions hold.

REMARK 2.4 : Direct verification of assumption (A.6) may pose some difficulty in some cases. But (A.6) is true quite generally if ψ and F satisfies some mild regularity conditions as is evidenced by the following result.

PROPOSITION : *Suppose that* F *has a nonzero absolutely continuous component* Q *with density* q *with respect to the Lebesgue measure on* $\mathbb{R}$ *and* ψ *has a continuous nonvanishing derivative on some interval* (a , b) *for which* Q { (a , b) } > 0. *Then* (A.6) *holds.*

1.3. Bootstrap Approximations.

We start this section with the following example. It shows that if condition (1.6) does not hold for some choice of the resampling distribution F_n, then the corresponding bootstrap procedure cannot be even first order correct.

EXAMPLE : In addition to ψ being a nondecreasing, bounded, real valued function, assume that ψ has a bounded uniformly continuous second derivative (e.g. one may take $\psi(x) = \tan^{-1}(x)$). Also suppose that F and ψ jointly satisfy the hypotheses of Theorem 2.1 (a) and $E\,\psi(\epsilon_1) = 0$. For the sake of clarity in the resulting expressions, we take $x_i = 0$ or 1 according as i is even or odd. Note that for this choice of x_i's, $a_n^2 = 0\,(n)$ and $b_n = 0\,(\log n)$. By Theorem 2.1, it follows that

(3.1) $$a_n(\overline{\beta}_n \ \beta\,) \text{ converges in distribution to } N\,(\,0,\, A^{-2}).$$

where $A = E\psi'(\epsilon_1)/\,[\,E\,\psi^2(\epsilon_1)]^{1/2}$ as in (2.1).

Next consider bootstrapping $\overline{\beta}_n$. Let F_n denote the empirical distribution of the estimated residuals $\overline{\epsilon}_i = Y_i - x_i\overline{\beta}_n$, $i = 1,\dots, n$. Take independent sample $\epsilon_1^*, \dots, \epsilon_n^*$ from F_n. Note that in this case $E_n(\psi(\epsilon_1^*))$ is not necessarily zero. Hence, condition (1.6) does not hold. For $t \in \mathbb{R}$, write

$$S_n^*(t) = \sum_{i=1}^{n} x_i\, \psi(\,\epsilon_i^* - x_i t\,)$$

$$\tau_n^*(t) = \text{Standard deviation of } S_n^*(t) \text{ under } F_n.$$

By the monotonicity of ψ and the definition of β_n^*, it follows that for all $t \in \mathbb{R}$,

(3.2) $$P_n(S_n^*(t) < 0) \le P_n(\,(\beta_n^* - \beta) \le t) \le P_n(S_n^*(t) \le 0).$$

where P_n denotes the bootstrap probability under F_n.

Now, using the Berry – Esseen Theorem for independent random varibles (cf. Theorem 12.4 of Bhattacharya and Rao (1976)), one can conclude that almost surely, for all $t \in \mathbb{R}$,

$$\sup_y \left| P_n\left((S_n^*(t) - E_n S_n^*(t)) \le y\, \tau_n^*(t) \right) - \Phi(y) \right|$$

(3.3)

$$\le 2.75 \left\{ \sum_{i=1}^{n} |x_i|^3 E_n \left| \psi(\epsilon_i^* - x_i t) - E_n \psi(\epsilon_i^* - x_i t) \right|^3 \right\} / [\tau_n^*(t)]^3$$

Here, as before, E_n denotes the expectation under P_n and Φ denotes the distribution function of $N(0, 1)$.

Next we state the following results without proofs. Result 3.1 has been derived in the proof of Theorem 2.1 below (see equation (4.4)) and Result 3.2 is a consequence of Lemma 4.2 of Section 1.4 below.

RESULT 3.1 : *Let* ψ *have a bounded second derivative and* $\bar{\beta}_n$ *be defined by equation* (1.2). *Then there exists* $N > 1$ *such that for all* $n > N$,

$$P\left(a_n(\bar{\beta}_n - \beta) > b_n \right) < a_n^{-3}.$$

RESULT 3.2 : *Let* F_n *denote the empirical distribution of the estimated residuals* $\bar{\epsilon}_i$, $i = 1,\ldots, n$. *Then for every* $M > 0$ *and every* $k \ge 1$,

$$\sup \left\{ \left| E_n[\psi(\epsilon_1^* - x)]^k - E\,[\psi(\epsilon_1 - x)]^k \right| : |x| \le M \right\} = o(1), \quad \text{a.s..}$$

By Result 3.2 and the uniform continuity of ψ, it follows that for all x with $|x| \le \log n$,

(3.4) $$\left| [\tau_n^*(x/a_n)]^2 - [\tau_n^*(0)]^2 \right| = o(n), \quad \text{a.s..}$$

Hence, from (3.2), (3.3) and (3.4), it follows that for all x with $|x| \le \log n$,

$$(3.5)\qquad \sup_{|x|\le \log n} |\, P_n (a_n(\beta_n^* - \bar\beta_n) \le x) - \Phi(-[E_n S_n^*(x/a_n)]/\tau_n^*(x/a_n)) \,|$$

$$= o(n^{-1/2}), \qquad \text{a.s..}$$

Next we simplify $(-[E_n S_n^*(x/a_n)]/\tau_n^*(x/a_n))$. Since $\bar\beta_n$ satisfies (1.2), taking Taylor's expansion , we get

$$0 = \sum_{i=1}^n x_i\, \psi(\epsilon_i - x_i(\bar\beta_n - \beta))$$

$$= \sum_{i=1}^n x_i\, \psi(\epsilon_i) - (\bar\beta_n - \beta) \sum_{i=1}^n x_i^2\, \psi'(\epsilon_i) + (\bar\beta_n - \beta)^2 \sum_{i=1}^n x_i^3\, \psi''(\xi_i)/2$$

where ξ_i is a point between ϵ_i and $\bar\epsilon_i = Y_i - x_i\bar\beta_n$, $1 \le i \le n$. Now use Result 3.1 and the Law of Iterated Logarithm (LIL) to conclude that

$$(3.6)\qquad a_n(\bar\beta_n - \beta)\, E\, \psi'(\epsilon_1) = \sum_{i=1}^n x_i\, \psi(\epsilon_i) + o\,(1) \qquad \text{a.s..}$$

For j = 1, 2, ..., n, a two term Taylor's expansion together with the LIL and the fact that $\sum_{i=1}^n x_i = \sum_{i=1}^n x_i^2 = a_n^2$, yields,

$$E_n \psi(\epsilon_j^* - x_j x/a_n) = n^{-1}\sum_{i=1}^n \psi(\epsilon_i) - [x x_j/a_n + a_n^2(\bar\beta_n - \beta)/n] E\psi'(\epsilon_1)$$

$$+ R_{jn}(x)$$

where $\sup \{ |R_{jn}(x)| : 1 \le j \le n,\ |x| \le \log n\} = O\,(n^{-1}(\log n)^2)$ a.s..

Therefore, by (3.6) and the Result 3.2, one has

$$-[E_n S_n^*(x/a_n)]/\tau_n^*(x/a_n) = Ax + \sum_{i=1}^n (x_i - 1)\psi(\epsilon_i)/a_n\sigma(0) + R_n(x)$$

where $\sup\{|R_n(x)| : |x| \leq \log n\} = o(1)$ a.s.. Recall that $\sigma^2(0) = E\psi^2(\epsilon_1)$. Define $B_n = \sum_{i=1}^{n} (x_i - 1)\psi(\epsilon_i)/a_n\sigma(0)$. Then B_n has a limiting nondegenarate normal distribution (viz. N(0, 1)). Also, from (3.5), it follows that

$$\sup_{|x|\leq \log n} | P_n (a_n(\beta_n^* - \bar{\beta}_n) \leq x) - \Phi (Ax + B_n) | = o(1) \quad \text{a.s..}$$

Comparison of this with (3.1) shows that the usual bootstrap procedure fails to capture the limiting distribution of the unbootstrapped statistic and as a result, is not even first order correct.

As indicated in the introduction and implied by the above example, we shall confine ourselves only to those cases in which condition (1.6) holds. First we consider a situation where (1.6) is ensured by changing the resampling distribution.

Weighted Empirical Bootstrap

Assume that x_i's *are either all nonnegative or all nonpositive.* For $n \geq 1$, write $p_n = \sum_{i=1}^{n} |x_i|$. Let, F_{1n} be the d.f. putting mass $|x_i|/p_n$ at $\bar{\epsilon}_i$, $i = 1, \ldots, n$. Take the resampling distribution F_n to be F_{1n} and draw the bootstrap samples $\epsilon_1^*, \ldots, \epsilon_n^*$ from F_n. With $Y_i^* = x_i \bar{\beta}_n + \epsilon_i^*$, $i = 1, \ldots, n$, define the bootstrap estimator β_n^* of β as a solution of (1.5). Note that for this choice of F_n, $E_n\psi(\epsilon_1^*) = p_n^{-1} \sum_{i=1}^{n} |x_i| \psi(Y_i - x_i\bar{\beta}_n)$. Hence (1.2) and the fact that all x_i's are of the same sign jointly imply that

$$E_n\psi(\epsilon_1^*) = (\text{Sign of } x_1)\, p_n^{-1} \sum_{i=1}^{n} x_i\, \psi(Y_i - x_i\bar{\beta}_n) = 0.$$

Hence, in this case (1.6) holds.

Before stating the theorems we introduce some more notation. For any resampling distribution F_n, write

$$m_n(x) = E_n\psi(\epsilon_1^* - x), \qquad W_n(x) = s_n^2(x) = E_n\,\psi^2(\epsilon_1^* - x) - m_n^2(x)$$

(3.7)

$$m_{i,n}(x) = E_n(\psi(\epsilon_1^* - x) - m_n(x))^i, \; i = 3, 4 \text{ and } A_n = -m_n'(0)/s_n(0).$$

Next define

$$H_{1n}^*(x) = \Phi(x) - d_{1n}\,[(m_n''(0)/s_n(0) - m_n'(0)\,W_n'(0)/s_n^3(0))\,x^2/2A_n^2$$

$$+ (\, m_{3,n}(0) \,/\, 6s_n^3(0)\,)\,H_2(x)\,]\,\phi(x)$$

$$H_{2n}^*(x) = H_{1n}^*(x) - \phi(x)\,[d_{2n}\{\,(m_n'''(0)/s_n(0) + 3\,A_nW_n''(0)/\,2\,W_n(0)\,)\,x^3/6A_n^3$$

$$+ (\, m_{3,n}'(0)/6A_ns_n^3(0))\,xH_2(x) + ((\, m_{4,n}(0) - 3s_n^4(0))/24s_n^4(0))\,H_3(x)\}$$

$$+ d_{1n}^2\{(m_n''(0)/s_n(0) + A_nW_n'(0)/W_n(0))\,(m_{3,n}(0)/12A_n^2\,s_n^3(0))\,x^2H_3(x)$$

$$+ (m_{3,n}^2(0)/72s_n^6(0))H_5(x) + (m_n''(0)/s_n(0) + A_nW_n'(0)/W_n(0))^2\,x^5/8A_n^4$$

$$- (m_n''(0)\,W'(0)/s^3(0) + 3A_nW_n'^2(0)/2W_n^2(0))x^3/4A_n^2$$

$$- (m_{3,n}(0)\,W_n'(0)/4A_ns_n^5(0))\,xH_2(x)\}].$$

REMARK 3.1 : In the statements of Theorems 3.1 – 3.4, H_{1n}^* and H_{2n}^* are defined by the same expressions but in each case the functions m_n, W_n, $m_{3,n}$ and $m_{4,n}$ are to be defined using the corresponding resampling distribution F_n.

In the following, let P_n denote the bootstrap probability under F_n. We are now ready to state

THEOREM 3.1 : *Assume that the hypotheses of Theorem* 2.1 (*a*) *hold and that for every* $c > 0$, $\sum_{n=1}^{\infty} \exp(-cp_n^2/a_n^2) < \infty$. *If* β_n^* *is defined as a solution of* (1.5) *with* $F_n = F_{1n}$ *then,*

(a) $$\sup_x \mid P_n(a_n(\beta_n^* - \bar{\beta}_n) \le x) - H_{1n}^*(A_n x) \mid = o(M_n/a_n) \quad \text{a.s..}$$

(b) $$\sup_x \mid P_n(A_n a_n(\beta_n^* - \bar{\beta}_n) \le x) - P(A a_n(\bar{\beta}_n - \beta) \le x) \mid$$
$$= o(M_n/a_n) \text{ a.s..}$$

where A and A_n *are as defined in* (2.1) *and* (3.7) *respectively.*

Modified Scores Bootstrap :

Now we consider bootstrapping $\bar{\beta}_n$ using Shorack's modification. For any resampling distribution F_n, define β_n^* as a solution t of

(3.8) $$\sum_{i=1}^{n} x_i \{ \psi(Y_i^* - x_i t) - E_n \psi(\epsilon_1^*) \} = 0.$$

Clearly with this modification, $E_n\{ \psi(Y_1^* - x_1\bar{\beta}_n) - E_n\psi(\epsilon_1^*) \} = 0$ for any resampling distribution F_n and any x_i's. Let G_n denote the empirical distribution of the estimated residuals $\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n$. If ψ is smooth, one can take $F_n = G_n$ and still have the Edgeworth expansion for β_n^*. More precisely, the following analog of Theorem 3.1 is true.

THEOREM 3.2: *Suppose that the hypotheses of Theorem* 2.1(a) *hold and* β_n^* *is defined as a solution of* (3.8) *with* $F_n = G_n$. *Then,*

(a) $$\sup_x | P_n(a_n(\beta_n^* - \bar{\beta}_n) \leq x) - H_{1n}^*(A_n x) | = o(M_n/a_n) \quad \text{a.s..}$$

(b) $$\sup_x | P_n(A_n a_n(\beta_n^* - \bar{\beta}_n) \leq x) - P(A a_n(\bar{\beta}_n - \beta) \leq x)|$$
$$= o(M_n/a_n) \quad \text{a.s..}$$

Now consider the case when ψ is not necessarily smooth and the differentiability conditions are imposed solely on F. Here, instead of taking the samples from G_n, one should take the bootstrap samples from some smoother estimator of F to guarantee the validity of Edgeworth expansion for the bootstrapped estimator β_n^*. Let k be a known probability density on the real line and $\{e_n\}$ be a sequence of positive real numbers, $e_n \to 0$ as $n \to \infty$. Define

(3.9) $$g_n(x) = e_n^{-1} [\int k((x-y)/e_n) \, dG_n(y)].$$

Now take F_n to be the d.f. corresponding to g_n. In this case properties of F_n depends largely on the assumptions made about k and $\{e_n\}$. For $r = 1, 2$, let C(r) refer to the following conditions on k and $\{e_n\}$:

(i) For every $c>0$, $\sum_{n=1}^{\infty} \exp(- c n e_n^{2(r+2)}) < \infty$,

(ii) $\int |u|\, k(u)\, du < \infty$ and

(iii) For $s = 0, 1, \ldots, (r+1)$, $k^{(s)}$ is of bounded variation.

THEOREM 3.3 : *Assume that the hypotheses of Theorem* 2.2(a) *hold and that* β_n^* *is defined by* (3.8) *taking* F_n *to be the d.f. corresponding to the density* g_n. *If* k *and* $\{e_n\}$ *satisfy condition* C(1), *then*

(a)
$$\sup_x |\, P_n(\, a_n(\, \beta_n^* - \bar{\beta}_n\,) \leq x\,) - H_{1n}^*(\, A_n x)\,| = o\,(\, M_n/a_n\,) \quad \text{a.s..}$$

(b)
$$\sup_x |\, P_n(\, A_n a_n(\, \beta_n^* - \bar{\beta}_n\,) \leq x\,) - P(\, A a_n(\, \bar{\beta}_n - \beta) \leq x\,)|$$
$$= o\,(\, M_n/a_n\,) \quad \text{a.s..}$$

Theorems 3.1 – 3.3 show that appropriate bootstrap estimators correct the terms of order $O(d_{1n})$ (see equation (2.1) of Section 1.2 for the definition of d_{jn}, $1 \leq j \leq 4$) in the Edgeworth expansion for the distribution of normalized $\bar{\beta}_n$ and thus attain a higher rate than the normal approximation. In fact, under some symmetry assumptions on the model, the accuracy of bootstrap procedure can be increased considerably with a minor modification. Assume that the score function ψ is odd and the underlying d.f. F is symmetric about zero (i.e. $F(-x) + F(x) = 1$). Under these conditions all the terms of order $O\,(d_{3n})$ in the Edgeworth expansion of $\bar{\beta}_n$ vanish. As a result, the rate of normal approximation is typically of the order of $O\,(d_{4n})$. In such situations if one draws the bootstrap samples from an asymmetric resampling distribution F_n, the terms of order $O(d_{3n})$ do not necessarily vanish from the corresponding expansion for β_n^*. Therefore, the rate of bootstrap approximation can at the best be $o(\, d_{3n}\,)$ which is much worse than the normal approximation. In particular this implies that the ordinary bootstrap procedure fails in such situations. However, one can overcome this by changing the resampling distribution to a suitable symmetric distribution. Only the case with smooth F is considered below.

Let g_n be as in (3.8). Since $g_n(x)$ may not be symmetric, we symmetrize g_n and take the estimating density at x to be $f_n(x) = [g_n(x) + g_n(-x)]/2$. Now choose F_n to be the d.f. corresponding to f_n. Note that for this choice of F_n, (3.8) reduces to (1.5) and the corresponding bootstrap estimators are the same.

THEOREM 3.4 : *Assume that the hypotheses of Theorem* 2.2(b) *hold and* k *and* $\{e_n\}$ *satisfy condition* C(2). *Then for odd* ψ, *symmetric* F *and* F_n *equal to the d.f. corresponding to* f_n,

(a) $$\sup_x \mid P_n(a_n(\beta_n^* - \bar{\beta}_n) \leq x) - H_{2n}^*(A_n x) \mid = o\,(d_{4n}) = o\,(M_n^2/a_n^2) \quad \text{a.s..}$$

(b) $$\sup_x \mid P_n(A_n a_n(\beta_n^* - \bar{\beta}_n) \leq x) - P(A a_n(\bar{\beta}_n - \beta) \leq x)\mid = o\,(d_{4n})$$
$$= o\,(M_n^2/a_n^2) \quad \text{a.s..}$$

In (a) and (b), β_n^* is defined as a solution of (1.5) or (3.8).

1.4. **Proofs.**

We start by stating Esseen's lemma (Lemma 2 of Feller (1966), page 512).

LEMMA 4.1 (Esseen) : *Let* F *be a probability distribution with vanishing expectation and characteristic function* φ. *Suppose that* G *is a function on the real line such that* $F - G$ *vanishes at* $\pm\infty$ *and* G *has a derivative* g *with* $|g| \le m$. *Finally, suppose that* g *has a continuously differentiable Fourier transform* γ *such that* $\gamma(0) = 1$ *and* $\gamma'(0) = 0$. *hen for all real* x *and* $a > 0$,

$$| F(x) - G(x) | \le \int_{-a}^{a} \{ | \varphi(t) - \gamma(t) | / (\pi |t|)\} \, dt + 24m/ a\pi.$$

Repeated use of this lemma with proper choice of a and G will give the expansions upto the desired order. For the sake of completeness, we include here an inequality due to Hoeffding (Theorem 2 of Hoeffding (1963)).

HOEFFDING'S INEQUALITY : *If* $X_1, X_2, \ldots, X_n$ *are independent r.v.'s with* $a_i \le X_i \le b_i$ $(1 \le i \le n)$, *then for any* $t > 0$,

$$P(\bar{X} - \mu > t) \le \exp(-2n^2 t^2 / \sum_{i=1}^{n} (b_i - a_i)^2)$$

where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ *and* $\mu = E(\bar{X})$.

Before proving Theorem 2.1 we need to have some more notation. For $t \in \mathbb{R}$, let

$$S_n(t) = \sum_{i=1}^{n} x_i \psi(Y_i - x_i t), \qquad \mu_n(x) = E\, S_n(\beta + x/a_n),$$

$$V_n(x) = \tau_n^2(x) = \text{Var } S_n(\beta + x/a_n), \quad \bar{\mu}_{4,n}(x) = \sum_{i=1}^{n} x_i^4 \, \sigma^4(x\, x_i/a_n),$$

$$\mu_{i,n}(x) = i^{\text{th}} \text{ central moment of } S_n(\beta + x/a_n), \ i = 3, 4.$$

$$\varphi_n(x,t) = E \exp (it (S_n(\beta + x/a_n) - \mu_n(x))),$$

$$v_n(x,t) = [\log \varphi_n(x,t)] + V_n(x)\, t^2/2, \quad w(x,t) = E \exp (it\, \psi\, (\epsilon_1 - x)).$$

Next for real numbers x and y, define

$$K_{1n}(x,y) = \Phi(y) - (\mu_{3,n}(x)/ 6\tau_n^3(x))\, H_2(y)\, \phi(y),$$

$$K_{2n}(x,y) = K_{1n}(x,y) - \phi(y)\, [((\mu_{4,n}(x) - 3\bar{\mu}_{4,n}(x))/24\tau_n^4(x))\, H_3(y) + (\mu_{3,n}^2(x) / 72\tau_n^6(x))\, H_5(y)],$$

$$\gamma_{1n}(x,t) = [1 + (\mu_{3,n}(x)/ 6\tau_n^3(x))\, (it)^3] \exp (-t^2/ 2)$$

$$\gamma_{2n}(x,t) = \gamma_{1n}(x,t) + [((\mu_{4,n}(x) - 3\bar{\mu}_{4,n}(x))/24\tau_n^4(x))\, (it)^4 + (\mu_{3,n}^2(x) / 72\tau_n^6(x))\, (it)^6] \exp (-t^2/2).$$

In the proofs that follow, we shall use $D > 0$ as a generic constant, independent of n, x, y etc.

PROOF OF THEOREM 2.1 : Proofs of parts (a) and (b) follow more or less the same route . First we outline the arguments common to both the parts and then complete the remaining steps in the proof of each part separately.

Note that boundedness of ψ, ψ'' and continuity of ψ'' guarantee that $\int \psi'(y)dy < \infty$ and ψ' is uniformly continuous. This in turn implies that ψ' is bounded. Therefore, the function μ is twice continuously differentiable with a bounded second derivative. Hence there exist constants $\eta_1 > 0$ and $c_1 > 0$ such that for $|x| < \eta_1$,

$$(4.1) \qquad |\mu(x)| > c|x|.$$

This inequality will be used to obtain a bound on the probability of the deviation of $\overline{\beta}_n$ from β. Next observe that monotonicity of ψ implies $S_n(t)$ is nonincreasing in t for every $n \geq 1$. This and the definition of $\overline{\beta}_n$ gives,

$$(4.2) \qquad \begin{aligned} P(S_n(\beta + x/a_n) < 0) \leq P(a_n (\overline{\beta}_n - \beta) \leq x) \leq P(S_n(\beta + x/a_n) \leq 0) \\ P(S_n(\beta + x/a_n) > 0) \leq P(a_n (\overline{\beta}_n - \beta) \geq x) \leq P(S_n(\beta + x/a_n) \geq 0) \end{aligned}$$

By Hoeffding's inequality, (4.1) and (4.2), there exists a constant $C > 0$ such that for all $0 < u < \eta_1 M_n$,

$$(4.3) \qquad P(| \overline{\beta}_n - \beta | > u) \leq 2 \exp (-C u^2 a_n^2)$$

Now take $u = b_n/a_n$ (recall that $b_n = \log a_n$) in (4.3) to get an $N > 1$ such that for all $n > N$,

$$(4.4) \qquad P(a_n | \overline{\beta}_n - \beta | > b_n) \leq a_n^{-3}.$$

Therefore, it is enough to consider the expansion of $P(a_n | \overline{\beta}_n - \beta | \leq x)$ for $|x| \leq b_n$. In view of (4.2), (4.3) and the form of $H_{1n}(x)$, it is enough to find an expansion of $P(S_n(\beta + x/a_n) \leq 0)$, that holds uniformly for $|x| \leq b_n$, and to appraise $\sup \{ P(S_n(\beta + x/a_n) = 0), |x| \leq b_n \}$.

PROOF OF (a): Given an $\eta > 0$, choose an integer N and a constant $b > 0$ large enough such that for all y and $|x| \leq b_n$, $24 |K_{1n}(x,y)| < b\eta$. This is possible since both ψ and its derivative are bounded. Take $a = b\, M_n/a_n$ in Lemma 4.1. Then for all y in $\mathbb{R}$ and for all x with $|x| \le b_n$,

$$(4.5)\qquad \begin{aligned} &| P([S_n(\beta + x/a_n) - \mu_n(x)]/\tau_n(x) \le y) - K_{1n}(x , y) | \\ &\le \int_{-a}^{a} | \varphi_n(x,t/\tau_n(x)) - \gamma_{1n}(x,t) | / |t| \, dt + \eta\, M_n/a_n \end{aligned}$$

As is customary, the integral on the R.H.S. is broken into two parts ; one ranging over $|t| \le \delta\, a_n/ M_n$ (call it I) and the other over $\delta\, a_n/ M_n < |t| < a$ (call it II) for some $\delta > 0$ which will be chosen later. Since ψ is bounded, continuous and nondecreasing, ψ is uniformly continuous. Therefore for any $D > 0$,

$$\begin{aligned} &\sup\{ | w(x ,t) - w(0 ,t) | : |t| < D \} \\ &\le D \int | \psi(y-x) - \psi(y) | \, dF(y) \to 0 \text{ as } x \to 0. \end{aligned}$$

Hence there exists a $\delta > 0$ such that for $|x| < 2\delta$ and $|t| < 2\delta$,

$$(4.6)\qquad |w(x ,t)| > .5.$$

This guarantees that $v_n(x , t/\tau_n(x))$ is well defined for large n when $|x| \le b_n$ and $|t| \le \delta\, a_n/M_n$. Since ψ is bounded, $v_n(x , t/\tau_n(x))$ is infinitely differentiable in t over $|t| \le \delta\, a_n/M_n$. Next note that for any complex number u with $|u|<1$

$$\log (1+u) = u - u^2 r(u) \quad \text{where } |r(u)| < 1/(1-\delta_1) \text{ for all } |u| < \delta_1 < 1.$$

Therefore, Taylor's expansion of $v_n(x , t/\tau_n(x))$ around $t = 0$, continuity of the functions $V(x)$, $\mu_3(x)$, $\mu_4(x)$ and the above result together yield (possibly with a smaller $\delta > 0$)

$$(4.7) \qquad | v_n(x, t/\tau_n(x)) - (it)^3 \mu_{3,n}(x)/6\tau_n^3(x) | < D |t|^4 (\Sigma x_j^4 / \tau_n^4(x))$$

for all $|x| \leq b_n$, $|t| < \delta a_n/M_n$ and large n. Without loss of generality we may suppose that for the same set of values of x, t and n,

$$(4.8) \qquad | v_n(x,t/\tau_n(x)) | \leq t^2/4, \qquad | (it)^3 \mu_{3,n}(x)/6\tau_n^3(x) | \leq t^2/4.$$

Note that for all complex u and z,

$$(4.9) \qquad \begin{aligned} &|\exp(u) - 1 - z| \leq (|u-z| + |z|^2) \exp(\gamma) \\ &|\exp(u) - 1 - z - z^2/2| \leq (|u-z| + |z|^3) \exp(\gamma), \ \gamma > \max(|u|,|z|). \end{aligned}$$

Now choose $\delta > 0$ such that (4.6) – (4.8) hold simultaneously. For this choice of δ, one may use bounds (4.7) – (4.9) to conclude that uniformly in $|x| \leq b_n$ and for large n,

$$I = \int_{|t| \leq \delta a_n / M_n} \{ | \varphi_n(x,t/\tau_n(x)) - \gamma_{1n}(x,t) | / |t| \} dt$$

$$= \int_{|t| \leq \delta a_n / M_n} |t|^{-1} |\exp(v_n(x,t/\tau_n(x))) - 1 - (it)^3 \mu_{3,n}(x)/6\tau_n^3(x)| \exp(-t^2/2) dt$$

$$\leq D \int \{ |t|^3 (\Sigma x_j^4) / \tau_n^4(x) + |t|^5 (\Sigma x_j^3)^2 / \tau_n^6(x) \} \exp(-t^2/4) dt$$

$$(4.10) \qquad \leq D (M_n/a_n)^2$$

This takes care of the first part of the integral. Now we estimate II . Note that for real numbers x and t, the differentiability of ψ gives

$$|w(x,t) - w(0,t)| \leq D\,|tx|.$$

By the nonlatticeness of $\psi(\epsilon_1)$ and the above inequality, it follows that there exist $0 < q < 1$ and $N > 1$ (both depending on η through 'a' of (4.5)) such that for all $n > N$,

$$\sup \{|w(xx_j/a_n,\ tx_j/\tau_n(x))| : j \in A_n(c),\ |x| \leq b_n,\ \delta a_n/M_n \leq |t| \leq ba_n/M_n\} \tag{4.11}$$
$$< q.$$

Hence by condition (2.3) it follows that for all $n > N$ and $|x| \leq b_n$,

$$II = \int_{\delta a_n/M_n \leq |t| \leq b\, a_n/M_n} \{|\varphi_n(x,t/\tau_n(x)) - \gamma_{1n}(x,t)|/|t|\}\, dt$$

$$\leq D\,[\,q^{k_n(c)} + \int_{\delta a_n / M_n \leq |t|} \{\,|\,\gamma_{1n}(x,t)\,|\,/\,|t|\,\}\, dt\,] \tag{4.12}$$

$$\leq D\,(\,M_n/\,a_n\,)^2.$$

By (4.5), (4.10) and (4.12) it follows that given an $\eta > 0$, there exists $N > 1$ and a $D>0$ (both depending on F only through the nonlatticeness of $\psi(\epsilon_1)$ and the values of the function μ, σ, μ_3 and their derivatives at zero) such that for all $n > N$,

$$\sup_{|x| \leq b_n} \sup_{y} |P(\,(S_n(\beta + x/a_n) - \mu_n(x))/\tau_n(x) \leq y\,) - K_{1n}(\,x\,,y\,)\,|$$

$$\leq D\,(\,M_n/a_n\,)^2 + \eta\, M_n/a_n.$$

Since $\eta > 0$ is arbitrary, this gives the Edgeworth expansion of normalised $S_n(\beta + x/a_n)$ with a remainder term of the order of $o\,(d_{3n})$ uniformly in $|x| \le b_n$. The smoothness conditions on ψ ensures that the functions μ, V and μ_3 have a second derivative and μ'' is uniformly continuous. Taking Taylor's expansions (of the terms involving x) around $x = 0$, one gets,

$$\mu_n(x)/\tau_n(x) = x\mu'(0)/\sigma(0) + d_{1n}(\mu''(0)/\sigma(0) - \mu'(0)V'(0)/\sigma^3(0))x^2/2$$
$$+ Q_{1n}(x),$$

$$\mu_{3,n}(x)/\tau_n^3(x) = (\,\mu_{3,n}(0)/\sigma^3(0)\,)\,(d_{1n}) + Q_{2n}(x),$$

where the remainder terms satisfy

$$| Q_{1n}(x) | \le D\, x^2\, (d_{3n}) \sup \{\, | \mu''(y) - \mu''(0) | : |y| \le b_n M_n/a_n \,\},$$

(4.13)

$$| Q_{2n}(x) | \le D\, b_n\, (M_n/a_n)^2$$

for all x with $|x| \le b_n$. Here the constant D depends only on the values of functions μ, V, m_3 and their derivatives at 0. Using the above expansions, uniform continuity of μ'' and (4.2) one can conclude that,

$$\sup_{|x| \le b_n} | P(\, a_n(\,\overline{\beta}_n - \beta\,) \le x) - H_{1n}(Ax) | = o\,(\,(M_n/\,a_n)\,).$$

This together with (4.4) completes the proof of part (a).

PROOF OF PART (b) : The steps in the proof are similar to those in part (a). We

will mention only the major differences here. Given $\eta > 0$, choose $b > 0$ large enough such that for all y in $\mathbb{R}$ and $|x| \leq b_n$, $| 24 K'_{2n}(x , y) | < b\eta$. Take $a = b d_{4n}$ in the Esseen's lemma and break up the integral into two parts as before. Note that for any complex u with $|u| < 1$,

$$\log (1+u) = u - u^2/2 + u^3 r(u) \text{ where } |r(u)| < 1/(1-\delta_1) \text{ for } |u| < \delta_1 < 1.$$

Using the differentiability of $v(x, t)$ in t and the above result, choose $\delta > 0$ such that for $|x| \leq b_n$, $t| \leq \delta a_n/M_n$ and large n,

$$| v_n(x,t/\tau_n(x)) - (it)^3 \mu_{3,n}(x)/6\tau_n^3(x) - (it)^4 [\mu_{4,n}(x) - \bar{\mu}_{4,n}(x)]/24\tau_n^4(x) |$$
$$< D |t|^5 (\sum |x_j|^5)/\tau_n^5(x),$$

$$|v_n(x,t/\tau_n(x))| \leq t^2/4, \quad |(it)^3 \mu_{3,n}(x)/6\tau_n^3(x) + (it)^4 \mu_{4,n}(x)/24\tau_n^4(x)| \leq t^2/4.$$

Now use the second part of (4.8) to conclude that for $|t| \leq \delta M_n/a_n$,

$$|\varphi_n(x,t/\tau_n(x)) - \gamma_{2n}(x,t)| \leq D (d_{4n}M_n/a_n) \{|t|^5 + |t|^9\} \exp(-t^2/4).$$

Hence, it follows that for large n, uniformly in $|x| \leq b_n$,

$$\text{(4.14)} \quad I = \int_{|t| \leq \delta a_n / M_n} \{|\varphi_n(x,t/\tau_n(x)) - \gamma_{2n}(x,t)|/|t|\} \, dt \leq D d_{4n}M_n/a_n.$$

For estimating II, one has to use condition (A.6) instead of the nonlatticeness of $\psi(\epsilon_1)$. In fact condition (A.6) guarantees that

(4.15) $\sup \{ |\varphi_n(x, t/\tau_n(x))| : |x| \le b_n,\ \delta a_n/M_n \le |t| \le b\, d_{4n} \} < q^{k_n(c)}$.

Using (4.14) and (4.15) one can conclude (as in part a) that

(4.16) $$\sup_{|x| \le b_n} \sup_{y} |P((S_n(\beta + x/a_n) - \mu_n(x))/\tau_n(x) \le y) - K_{2n}(x , y)| = o(d_{4n}) = o((M_n/a_n)^2).$$

Now observe that the differentiability conditions on ψ implies that the functions μ, V, μ_3 and μ_4 are three times differentiable and μ''' is uniformly continuous. A tedious computation of Taylor's expansion gives

$$\mu_n(x)/\tau_n(x) = (\mu'(0)/\sigma(0))\, x + (d_{1n}) \{\mu''(0)/\sigma(0) - \mu'(0)V'(0)/\sigma^3(0)\}\, x^2/2$$

$$+ [(3d_{1n}^2/2) \{ 3\, \mu'(0)V'^2(0)/2\sigma^5(0) - \mu''(0)V'(0)/\sigma^3(0) \}$$

$$+ d_{2n} \{\mu'''(0)/\sigma(0) - 3\mu'(0)V''(0)/2\sigma^3(0)\}]\, x^3/6 + Q_{3n}(x),$$

$$\mu_{3,n}(x)/\tau_n^3(x) = d_{1n}\, \mu_3(0)/\sigma^3(0) + d_{2n}\, \mu_3'(0)\, x/\sigma^3(0))$$

$$- 3\, d_{1n}^2\, \mu_3(0)\, V'(0)\, x\, /2\sigma^5(0) + Q_{4n}(x),$$

$$[\, \mu_{4,n}(x) - \bar{\mu}_{4,n}(x)\,]/\tau_n^4(x) = d_{2n}\, [\, \mu_4(0) - \bar{\mu}_4(0)\,]/\sigma^4(0) + Q_{5n}(x),$$

$$\mu_{3,n}^2(x)/\tau_n^6(x) = d_{1n}^2\, \mu_3^2(0)/\sigma^6(0) + Q_{6n}(x)$$

where for all x with $|x| \le b_n$, the remainder terms satisfy

$$(4.17)\quad \begin{aligned} &|Q_{3n}(x)| \le D\,|x|^3\, d_{4n} \sup\{\,|\mu'''(y) - \mu'''(0)| : |y| < M_n b_n/a_n\,\}, \\ &\mathrm{Max}\,\{\,|Q_{in}(x)| \;: i = 4, 5, 6\,\} < D\, b_n^2\,(d_{4n}\, M_n/a_n). \end{aligned}$$

The constant D depends only on the values of the functions μ , V , μ_3 , μ_4 and their derivatives at zero. As in the previous case it now follows from (4.2), (4.16) and (4.17) that

$$\sup_{|x| \le b_n} |\, P(\,a_n(\,\hat{\beta}_n - \beta\,) \le x) - H_{2n}(x)\,| = o\,(d_{4n}) = o\,(\,(M_n/\,a_n)^2).$$

By (4.4) the proof of part (b) is now complete.

PROOF OF THEOREM 2.2 : Note that the hypotheses of Theorem 2.2 differ from those of Theorem 2.1 only in the differentiability conditions on the functions ψ and F. From the proof of Theorem 2.1 it is evident that the differentiability of the function ψ has been used to guarantee that the functions μ, V, μ_3 and μ_4 have sufficiently many derivatives. Since ψ is bounbed and nondecreasing, therefore for every $k \ge 1$, ψ^k is of bounded variation. An application of integration by parts gives,

$$\int \psi^k\,(y - x)\, dF(y) = \psi^k\,(\,\infty\,) - \int F\,(y + x)\, d\psi^k(y).$$

As a consequence of this relation, the function μ, V, μ_3 and μ_4 will have sufficient smoothness as required in the proof of Theorem 2.1. The only cases where the differentiability of ψ has been used for different reasons are (4.6) and (4.11). But under the hypotheses of both the parts, F has a density and hence this follows easily by Scheffe's Theorem.

PROOF OF THE PROPOSITION : Let $p = Q\{(a, b)\}$. Then $0 < p \le 1$. For any set B of $\mathbb{R}$, let 1_B denote the indicator of the set B. Note that by the Riemann Lebesgue Lemma,

$$\int \exp(it\,\psi(y))\,1_{(a,b)}(y)\,dQ(y)$$
$$= \int \exp(ity)\,[\,1_{(\psi(a),\,\psi(b))}(y)\,q(\psi^{-1}(y))/\psi^{(1)}(y)\,]\,dy$$
$$\to 0 \text{ as } |t| \to \infty.$$

Hence, there exists a constant $M > 0$ such that for $|t| > M$,

$$\left|\int \exp(it\,\psi(y))\,1_{(a,b)}(y)\,dQ(y)\right| < p/4. \tag{4.18}$$

Therefore, for any x in $\mathbb{R}$ and $|t| > M$,

$$|\,E\exp(it\,\psi(\epsilon_1 - x))\,|$$
$$\le (1-p) + \left|\int \exp(it\,\psi(y-x))\,1_{(a,b)}(y)\,dQ(y)\right|$$
$$\le (4-3p)/4 + \left|\int [\exp(it\,\psi(y-x)) - \exp(it\,\psi(y))]\,1_{(a,b)}(y)\,dQ(y)\right|$$
$$\le (4-3p)/4 + \int |\,q(y+x)\,1_{(a-x,\,b-x)}(y) - q(y)\,1_{(a,b)}(y)\,|\,dy.$$

Note that the continuity of q on (a, b) implies,

$$q(y+x)\,1_{(a-x,\,b-x)}(y) \to q(y)\,1_{(a,b)}(y) \quad \text{as } x \to 0.$$

Therefore, the above integral goes to zero because

$$\int q(y+x)\,1_{(a-x,\,b-x)}(y)\,dy = \int q(y)\,1_{(a,b)}(y)\,dy \quad \text{for all x in } \mathbb{R}.$$

Hence, there exists $\delta > 0$ such that whenever $|x| < \delta$ and $|t| > M$,

$$| E \exp (it \psi(\epsilon_1 - x)) | < (4 - 3p)/4 + p/4 = (2 - p)/2 < 1.$$

This completes the proof of the proposition.

For the proofs of Theorems 3.1 – 3.4, define $w_n(x,t) = E_n (\exp (it\psi(\epsilon_1^* - x)))$, $w_{1n}(x,t) = p_n^{-1} \sum_{j=1}^{n} x_j \exp (it\psi(\epsilon_j - x))$ and $w_{2n}(x,t) = n^{-1} \sum_{j=1}^{n} \exp (it\psi(\epsilon_j - x))$. The basic facts required for proving Theorem 3.1 and 3.2 are given in Lemma 4.2 below.

LEMMA 4.2 : *Let* F_n *be either of the resampling distributions of Theorem* 3.1 *and* 3.2. *Then, for any* $M > 0$,

(4.20) $$\sup \{ | w_n(x,t) - w (x,t) | : |t| \leq M, |x| \leq M \} = o (1) \quad \text{a.s..}$$

Let h *be a function with a bounded first derivative. Then for every* $M > 0$,

(4.21) $$\sup \{|E_n h(\epsilon_1^* - x) - E h(\epsilon_1 - x) | : |x| \leq M \} = o (1) \quad \text{a.s..}$$

PROOF OF LEMMA 4.2 : First we prove (4.20). For $|t| \leq M$, $|x| \leq M$ and $F_n = F_{1n}$,

$$| w_n(x,t) - w_{1n}(x,t) |$$

$$\leq \| \psi^{(1)} \| \, |t| \, [\sum_{j=1}^{n} |x_j| \, | \bar{\epsilon}_j - \epsilon_j |]/p_n,$$

$$\leq D a_n^2 \, |\bar{\beta}_n - \beta| \, / \, p_n.$$

By the assumption on x_i's and (4.3), the R.H.S. tends to zero a.s. as n tends to infinity. Similarly, for $F_n = G_n$, $|t| \leq M$, $|x| \leq M$,

$$| w_n(x,t) - w_{2n}(x,t) | \leq D a_n \, | \bar{\beta}_n - \beta | \, / \, n.$$

By (4.3), this tends to zero a.s. as n goes to infinity. Therefore, it is enough to show that for i = 1, 2,

$$\sup \{ \mid w_{in}(x,t) - w(x,t) \mid : |t| \leq M, \ |x| \leq M \} = o(1) \qquad \text{a.s..}$$

This is proved by adapting the idea of the proof of Lemma 2 in Babu and Singh (1984). Fix $\eta > 0$. Then there exists a constant $C > 0$ (independent of η) such that for all $n \geq 1$ and for all u with $|u| < C\eta$,

$$\sup \{ \mid w_{in}(x+u, \ t+u) - w_{in}(x,t) \mid : |t| \leq M, \ |x| \leq M, \ i = 1, \ 2 \} < \eta$$

and

$$\sup \{ \mid w(x+u, \ t+u) \ - \ w(x,t) \mid : |t| \leq M, \ |x| \leq M \} < \eta.$$

Define $B(M,\eta) = \{ j : j$ is an integer between $-M/(C\eta)$ and $M/(C\eta) \}$. Then, for i = 1, 2,

$$\sup \{ \mid w_{in}(x,t) - w(x,t) \mid : |t| \leq M, \ |x| \leq M \}$$

$$\leq 2\eta + \max \{ \mid w_{in}(1C\eta , jC\eta) - w(1C\eta , jC\eta) \mid : i, \ j \in B(M,\eta) \}.$$

Therefore by Hoeffding's inequality it follows that

$$P(\sup \{ \mid w_{1n}(x,t) - w(x,t) \mid : |t| \leq M, \ |x| \leq M \} > 4\eta)$$

$$\leq P(\max \{ \mid w_{1n}(iC\eta , jC\eta) - w(iC\eta , jC\eta) \mid : i, j \in B(M,\eta) \} > 2\eta)$$

$$\leq D \, \eta^{-2} \exp (- (\eta p_n)^2 / 2a_n^2).$$

Similarly,

$$P(\sup \{ \mid w_{2n}(x,t) - w(x,t) \mid : |t| \leq M , \ |x| \leq M \} > 4\eta)$$

$$\leq D \, \eta^{-2} \exp (- \eta^2 n / 2).$$

By Borel Cantelli lemma, first part of the lemma follows. The other part can be proved similarly.

PROOF OF THEOREM 3.1 : Now we sketch the proof of Theorem 3.1. Since ψ is bounded, by Lemma 4.2 all (central) moments of $\psi(\epsilon_1 - x)$ under F_n converges a.s. to the corresponding (central) moments of $\psi(\epsilon_1 - x)$ uniformly over $|x| \leq M$. Let **N** denote the set of all positive integers. Fix a sample point for which (4.20) holds for every M in **N** and $m_n(x)$, $s_n(x)$, $m_{3,n}(x)$, $m_{4,n}(x)$ and their derivatives respectively converge to $\mu(x)$, $\sigma(x)$, $\mu_3(x)$, $\mu_4(x)$ and the corresponding derivatives uniformly over $|x| \leq 1$. For this sample point, using Lemma 4.2 one can get bounds in the inequalities (in the present set up) corresponding to (4.1), (4.3), (4.4), (4.5), (4.10), (4.12) and (4.13) uniformly over all $n \geq N$ for some $N > 1$. Hence one can retrace the proof of Theorem 2.1(a) to obtain Theorem 3.1 (a). Part (b) follows easily from Lemma 4.2.

PROOF OF THEOREM 3.2 : Similar to the proof of Theorem 3.1.

PROOF OF THEOREM 3.3 : Let G_{1n} denote the empirical distribution function of $\epsilon_1,\ldots,\epsilon_n$. Define $g_{1n}(x) = \int [\, k((x-y)/e_n)\; dG_{1n}(y)\,]/e_n$. First we show that the estimators $g_n^{(r)}(x)$ converge to $f^{(r)}(x)$ uniformly in x for r = 0, 1, a.s.. Under the hypothesis of Theorem 3.3, Lemma 2.2 of Schuster (1969) and a simple modification of Lemma 1 of Bhattacharya (1967) guarantee that

(4.22) $$\max\{\, \| g_{1n}^{(r)} - f^{(r)} \| \;:\; r = 0, 1 \,\} = o(1) \text{ as } n \to \infty \quad \text{a.s..}$$

Therefore, it is enough to show that

(4.23) $$\max\{\, \| g_{1n}^{(r)} - g_n^{(r)} \| \;:\; r = 0, 1 \,\} = o(1) \text{ as } n \to \infty \quad \text{a.s..}$$

Now,
$$\| g_{1n}^{(r)} - g_n^{(r)} \|$$
$$\leq \sum_{i=1}^{n} [\sup_x \ | k^{(r)}((x-\epsilon_i)/e_n) - k^{(r)}((x-\bar{\epsilon}_i)/e_n)|]/n\, e_n^{r+1}$$
$$\leq D \sum_{i=1}^{n} | \bar{\epsilon}_i - \epsilon_i | / (n\, e_n^{r+2})$$
$$\leq D\, a_n \, | \bar{\beta}_n - \beta | / (n^{1/2} e_n^{r+2}).$$

The last step follows by an application of Cauchy Schwartz inequality. By (4.3) and the assumption on $\{ e_n \}$, (4.23) follows. Hence (4.22) and (4.23) jointly imply that

(4.24) $$\max \{ \| g_{1n}^{(r)} - f^{(r)} \| : r = 0, 1 \} = o(1) \text{ as } n \to \infty \quad \text{a.s..}$$

For proving Theorem 3.3 we need the following Lemma.

LEMMA 4.3 : *Let* F_n *be the distribution corresponding to the density* g_n. *Then* ,

$$\sup \{ \ | w_n(x,t) - w\,(x,t) \ | : t \in \mathbb{R},\ x \in \mathbb{R} \} = o(1) \text{ as } n \to \infty \qquad \text{a.s..}$$

For any bounded function h,

$$\sup \{ \ | \ddot{E}_n h(\epsilon_1^* - x) - E\, h(\epsilon_1 - x) \ | : x \in \mathbb{R} \} = o(1) \text{ as } n \to \infty \text{ a.s..}$$

PROOF OF LEMMA 4.3 : It is to see that for all x and for all t

$$| w_n(x,t) - w\,(x,t) |$$
$$\leq \int | g_n(y) - f(y) | \, dy$$

By (4.24) and Scheffe's theorem it follows that

$$\int | g_n(y) - f(y) | \, dy \to 0 \text{ as } n \to \infty \text{ a.s.}$$

This proves the first part of the Lemma. Proof of the other part is similar.

Now we give an outline of the proof of Theorem 3.3. Fix a sample point for which $\max \{ \| g_n^{(r)} - f^{(r)} \| : r = 0, 1 \} \to 0$ as $n \to \infty$. It is enough to show that for this sample point the inequalities in the proof of part (a) of Theorem 2.1 holds uniformly in all sufficiently large n when F is replaced by F_n. Note that for any real number x,

$$\sup \{ | w(x,t) - w(0,t) | : t \in \mathbb{R} \} \le \int | f(y+x) - f(y) | \, dy$$

which tends to zero as x tends to zero. Hence, by the nonlatticeness of $\psi(\epsilon_1)$, Lemma 4.3 and the above observation, it follows that there exist $N > 1$, $\delta > 0$ and $0 < q < 1$ such that for all $n \ge N$,

$$\begin{aligned} &\operatorname{Inf} \{ | w_n(x,t) | : |t| > M, \ |x| \le \delta \} > .5, \\ (4.25) \quad & \\ &\sup \{ | w_n(x,t) | : |t| > M, \ |x| \le \delta \} < (1+q)/2 < 1. \end{aligned}$$

Also by Lemma 4.3,

$$\begin{aligned} &\operatorname{Max} \{ \| m_n^{(r)} - \mu^{(r)} \|, \ \| s_n^{(r)} - \sigma^{(r)} \| : r = 0, 1, 2 \} \to 0 \text{ as } n \to \infty \\ (4.26) \quad & \\ &\operatorname{Max} \{ \| m_{i,n}^{(r)} - \mu_i^{(r)} \| : i = 3, 4; \ r = 0, 1, 2 \} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Using (4.25) and (4.26), one can get bounds in the inequalities corresponding to (4.4), (4.5), (4.10), (4.11) and (4.12) uniformly over all sufficiently large n. As for the counter part of (4.13) in this case, note that,

$$\begin{aligned} &\sup \{ | m_n^{(2)}(y) - m_n^{(2)}(0) | : |y| \le M_n b_n / a_n \} \\ &\le \sup \{ | \mu^{(2)}(y) - \mu^{(2)}(0) | : |y| \le M_n b_n / a_n \} + 2 \| m_n^{(2)} - \mu^{(2)} \| \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence part (a) of the Theorem 3.3 follows. Part (b) is trivial in view of Lemma4.3.

PROOF OF THEOREM 3.4 : Using the conditions on $\{e_n\}$, k and the symmetry of the underlying density f, one can show (as in the proof of Theorem 3.3) that

$$\max \{ \; \| f_n^{(r)} - f^{(r)} \| : \; r = 0, \; 1, \; 2 \} \to 0 \text{ as } n \to \infty \qquad \text{a.s.}$$

Therefore the conclusions of Lemma 4.3 hold in this case as well. Hence, one can complete the proof along the line of proofs of Theorem 3.3 and Theorem 2.2(b) with a similar observation on Q_{3n}.

CHAPTER 2

2.1. Introduction.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d) p–dimensional random vectors with distribution function (d.f.) F_θ where θ lies in an open subset Θ of $\mathbb{R}^m$. Let $\psi: \mathbb{R}^p \times \Theta \to \mathbb{R}^m$ be a measurable function with respect to (w.r.t) the Borel σ–algebras on $\mathbb{R}^p \times \Theta$ and $\mathbb{R}^m$ such that

$$(1.1) \qquad \int \psi(x, \theta)\, dF_\theta(x) = 0 \quad \text{for all } \theta \in \Theta.$$

Let $\psi_1, \ldots, \psi_m$ denote the components of ψ. Then M–estimator $\hat{\theta}_n$ of θ corresponding to ψ is defined as a solution of the m–equations (in t)

$$(1.2) \qquad \sum_{j=1}^{n} \psi_i(X_j, t) = 0, \quad i = 1, 2, \ldots, m.$$

For $n \geq 1$, denote the empirical distribution function of $X_1, X_2, \ldots, X_n$ by F_n. Let $X_1^*, \ldots, X_n^*$ be a random sample of size n from F_n. Define the bootstrapped M–estimator θ_n^* as a solution of the system of equations (in t)

$$(1.3) \qquad \sum_{j=1}^{n} \psi_i(X_j^*, t) = 0, \quad i = 1, 2, \ldots, m.$$

In sections 2.2 and 2.3 below, under some regularity conditions on ψ and F_θ, it is shown that $\hat{\theta}_n$ exists for sufficiently large values of n and tends to θ as $n \to \infty$ with probability 1 under θ. It is also shown that with high (conditional) probability under F_n, θ_n^* exists and tends to θ at the rate $0(n^{-1/2}(\log n)^{1/2})$.

For such sequences of estimators, an almost sure Edgeworth expansion of the distribution of $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$ is given.

The method of the proof is similar to that of Bhattacharya and Ghosh (1978). Using the assumptions, an almost sure representation for $\hat{\theta}_n$ is obtained. In fact, it is shown that there exists a sufficiently smooth function H and a Borel measurable function $f: \mathbb{R}^p \to \mathbb{R}^k$, for some integer $k \geq 1$, such that

$$\hat{\theta}_n = H(\bar{Z} + R_n)$$

where $\bar{Z} = 1/n \sum_{j=1}^{n} Z_j$, $Z_j = f(X_j)$, $j = 1, 2, ..., n$ and with probability 1, $\|R_n\| = o(n^{-(s-2)/2})$ for some integer $s \geq 3$.

Next, for almost all sample sequences $(X_1, X_2, ...)$, outside a set of conditional probability $o(n^{-(s-2)/2})$, θ_n^* is expressed as

$$\theta_n^* = H(\bar{Y} + R_n^*)$$

where $\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j$, $Y_j = f(X_j^*)$, $j = 1,...,n$ and

$P_n(\|R_n^*\| > 0(n^{-s/2}(\log n)^{s/2}) = o(n^{-(s-2)/2})$ almost surely (a.s.). Here P_n refers to the conditional probability given $(X_1, X_2,...,X_n)$. It should be pointed out that for almost all sample sequences $(X_1, X_2,...)$, θ_n^* is expressed in terms of the same function H.

The arguments in the proof following this point can be divided into two steps. In step 1, $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$ is closely approximated by $\sqrt{n}(H(\bar{Y}) - H(\bar{Z}))$.

Properties of H, R_n and R_n^* guarantee that for almost all sample sequence $(X_1, X_2, \ldots)$, the error of approximation, say D_n, is small with high conditional probability. More precisely, D_n satisfies

$$P_n(\|D_n\| > 0(n^{-(s-1)/2}(\log n)^{s/2}) = o(n^{-(s-2)/2}) \text{ a.s..}$$

Representation of θ_n^* in terms of the same function H is crucial for carrying out this step.

In step 2, an almost sure asymptotic expansion for the conditional distribution of $\sqrt{n}\,(H(\bar{Y}) - H(\bar{Z}))$ is obtained. This, together with step 1, gives the almost sure asymptotic expansion for the distribution of $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$. Corresponding expansion for the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ was obtained by Bhattacharya and Ghosh (1978). Comparison of these two expansion shows that the bootstrap distribution of $\sqrt{n}(\theta_n^* - \hat{\theta}_n)$ approximates the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ under θ, at the rate of $o(n^{-1/2})$.

2.2. Assumptions and main results.

Before proceeding further, we collect here the notations to be used in the rest of chapter 2. Let Z^+ denote the set of all non–negative integers. Also, let ℓ be a positive integer. For $\nu = (\nu_1 \ldots \nu_\ell)' \in (Z^+)^\ell$ and $x = (x_1, \ldots, x_\ell)'$ in $\mathbb{R}^\ell$, write $x^\nu = \prod_{i=1}^{\ell} x_i^{\nu_i}$, $\nu! = \prod_{i=1}^{\ell} (\nu_i!)$ and $|\nu| = \nu_1 + \ldots + \nu_\ell$. For a function $f : \mathbb{R}^\ell \to \mathbb{R}$ having sufficiently many partial derivatives, denote by $D_j f$ the partial derivative of f w.r.t. its j–th co–ordinate, j=1,..., ℓ and set $D^\nu f = D_1^{\nu_1} \ldots D_\ell^{\nu_\ell} f$. Let Φ_A and ϕ_A respectively denote the distribution function and the density of normal distribution with mean zero and covariance matrix A for some positive definite

matrix A. For any matrix A, write A' = transpose of A. By $\| \ \|$ and $< , >$ denote, respectively, the norm and the inner product on appropriate Euclidean spaces. For a Borel set $B \subseteq \mathbb{R}^{\ell}$, let $B^{\epsilon} = \{x: \|x-y\| < \epsilon$ for some $y \in B\}$, $\epsilon > 0$ and ∂B = boundary of B.

Next, denote the underlying parameter value by θ_o. For $i = 1,\ldots, m$, $0 \le |\nu| \le s-1$ and $j \ge 1$, define the variables $\{Z_{\nu,i,j}\}$ and $\{Y_{\nu,i,j}\}$ by $Z_{\nu,i,j} = D^{\nu} \psi_i(X_j, \theta_o)$, $Y_{\nu,i,j} = D^{\nu}\psi_i(X_j^*, \theta_o)$. Write $Y_j^{(\nu)}$, $Z_j^{(\nu)}$ for the m–dimensional random vectors $(Y_{\nu,i,j})_{i = 1,\ldots,m}$ and $(Z_{\nu,i,j})_{i = 1,\ldots,m}$ respectively. Set $Z_j = (Z_j^{(\nu)})_{0 \le |\nu| \le s-1}$, $Y_j = (Y_j^{(\nu)})_{0 \le |\nu| \le s-1}$. Then $(Z_1, Z_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are i.i.d k–dimensional random vectors with $k = m\sum_{r=0}^{s-1} \binom{m+r-1}{r}$. Write $\bar{Z}^{(\nu)} = \sum_{j=1}^{n} Z_j^{(\nu)}/n$ and $\bar{Z} = \sum_{j=1}^{n} Z_j/n$ and define $\bar{Y}^{(\nu)}$ and $\bar{Y}$ similarly. In the following we shall write P to denote the product probability measure $\otimes_1^{\infty} F_{\theta_o}$ on the space of all infinite sequences in $\mathbb{R}^p$ and E to denote the expectation under P. $X_1, X_2, \ldots$ are then considered as co–ordinate variables. Also write E_n to denote the expectation under P_n.

$$\text{Let } \mu_{\nu,i} = E\,(Z_{\nu,i,1}), \quad i = 1,\ldots, m, \quad 0 \le |\nu| \le s-1,$$

(2.1)
$$\mu_{\nu} = (\mu_{\nu,i})_{i = 1, \ldots, m} \text{ and } \mu = (\mu_{\nu})_{0 \le |\nu| \le s-1}.$$

Also, let $\Sigma = E\,(Z_1 - \mu)(Z_1 - \mu)'$ and $S_n = E_n(Y_1 - E_nY_1)(Y_1 - E_nY_1)'$.
Finally, define

$$M = (\text{Grad } H(\mu))\,(\Sigma)\,(\text{Grad } H(\mu))'$$

$$M_n = (\text{Gard } H(\bar{Z}))\,(S_n)\,(\text{Grad } H(\bar{Z})'.$$

Now we state the assumptions.

(A1) There exists a Borel set $C \subseteq \mathbb{R}^p$ such that $F_\theta(C) = 1 \ \forall \ \theta \in \Theta$ and the components of ψ have continuous νth order partial derivatives in θ for $1 \leq |\nu| \leq s$ at each $(x, \theta) \in C \times \Theta$ for some integer $s \geq 3$.

(A2) $E \, \|D^\nu \psi(X_1, \theta_o)\|^s < \infty$ for $0 \leq |\nu| \leq s-1$, and there exists an $\epsilon > 0$ such that

$$\underset{|\nu|=s}{\text{Max}} \; E\big(\sup_{\|\theta-\theta_o\|<\epsilon} \|D^\nu \psi(X_1, \theta)\|^s \big) < \infty.$$

(A3) $D = ((\, E\, D_j \psi_i(X_1, \theta_o) \,))$ is non–singular.

We are now ready to state the main result.

THEOREM: *Let assumptions* (A1) – (A3) *hold. Then,*

(a) for almost all sample sequences $(X_1, X_2, \ldots)$, *there exists a sequence of statistics* $\{\theta_n^*\}$ *and a constant* $d_1 > 0$ *such that*

$$P_n\left(\|\theta_n^* - \theta_o\| < d_1 n^{-1/2} (\log n)^{1/2}, \ \theta_n^* \text{ solves (1.3)} \right) = 1 - o(n^{-(s-2)/2}).$$

(b) There exists a sequence $\{\hat{\theta}_n\}$ *of statistics such that*

$$P\left(\hat{\theta}_n \text{ solves (1.2) and } \|\hat{\theta}_n - \theta_o\| < d_1 \cdot n^{-1/2} (\log n)^{1/2} \text{ eventually} \right) = 1.$$

(c) Let $\{\hat{\theta}_n\}$ *and* $\{\theta_n^*\}$ *be two sequences of statistics which respectively satisfy (b) and (a). Suppose that the characteristic function of* Z_1 *under* θ_o *satisfies the Cramer's condition*

(A.4) $$\limsup_{\|t\| \to \infty} |E(e^{i\langle t, Z_1 \rangle})| < 1.$$

Then, there exist polynomials $a_1(F_n, \cdot), \ldots, a_{s-2}(F_n, \cdot)$ *such that for almost all sample sequence* $(X_1, X_2, \ldots)$,

$$\sup_{B\in\mathscr{B}} |P_n(\sqrt{n}\,(\theta_n^* - \hat{\theta}_n) \in B) - \int_B (1 + \sum_{r=1}^{s-2} n^{-r/2} a_r(F_n, x)) d\Phi_{M_n}(x)|$$

$$= o\,(n^{-(s-2)/2})$$

where $\mathscr{B}$ *is a class of Borel subsets of* $\mathbb{R}^m$ *satisfying*

$$\text{(2.2)} \qquad \sup_{B\in\mathscr{B}} \Phi_M((\partial B)^\epsilon) = 0(\epsilon) \text{ as } \epsilon \downarrow 0$$

and $a_1(F_n, \cdot), \ldots, a_{s-2}(F_n, \cdot)$ *are polynomials whose co–efficients are continuous functions of moments of* F_n *of order* s *or less.*

(d) If conditions (A1) – (A4) *are satisfied with* s = 3, *then for almost all sample sequence* $(X_1, X_2, \ldots)$

$$\sup_{B\in\mathscr{B}_1} |P_n(\sqrt{n}\, M_n^{1/2}(\theta_n^* - \hat{\theta}_n) \in B) - P\,(\sqrt{n}\, M^{1/2}(\hat{\theta}_n - \theta_o) \in B)| = o\,(n^{-1/2})$$

where $\mathscr{B}_1$ *is a class of Borel subset of* $\mathbb{R}^m$ *satisfying*

$$\text{(2.3)} \qquad \sup_{B\in\mathscr{B}_1} \Phi_I((\partial B)^\epsilon) = 0(\epsilon) \quad \text{as} \quad \epsilon \downarrow 0.$$

REMARK 2.1. Conditions (A1) – (A3) are similar to those of Bhattacharya (1985) and are somewhat weaker than the conditions in Bhattacharya and Ghosh (1978). Under some additional conditions, e.g. the contnuity of the maps $\theta \to F_\theta$ and $\theta \to D(\theta) = E_\theta((D_j\ \psi_i\ (X_1, \theta)))$, Bhattacharya and Ghosh (1978) have obtained results similar to (b) and (c) of the Theorem uniformly in θ lying in compact subsets of Θ. But, in our case, such a uniformity does not seem to be necessary. Given the data $X_1,\ldots,X_n$, if we can find θ_n^* and $\hat{\theta}_n$ satisfying (a) and (b), we can use the approximations in part (c) and (d) without any knowledge about θ. One such situation is of course that (1.2) and (1.3) have unique solutions. In the case of multiple solutions there is no rule which definitely specifies $\hat{\theta}_n$ satisfying (b) (or θ_n^* satisfying (a)) even in the presence of such uniformity.

REMARK 2.2 : Part (d) of the Theorem extends the pioneering result of Singh(1981) concerning the improvement of the rate of approximation by bootstrap in the case of sample mean. Taking $\psi(x,t) = x - t$ for the sample mean, it is easy to see that assumptions (A1)–(A4) reduce exactly to the set of conditions required for the validity of the corresponding result (part D of Theorem 1) of Singh(1981).

REMARK 2.3 : Though conditions (2.2) and (2.3) look similar, they are not equivalent in general. If the largest eigenvalue Λ of M is less than or equal to 1, then every class of Borel sets satisfying (2.3) also satisfies (2.2). But for $\Lambda>1$, a class of Borel sets satisfying (2.3) need not satisfy (2.2) as shown by the following example.

EXAMPLE : We consider the case $m = 1$. Let $M = (\Lambda)$ with $\Lambda > 1$ and $c = (4\Lambda)^{.5}$. Also , let $a_n = (c \log n)^{1/2}$, $n > 2$. Define the set B by $B = \{ a_n : n > 2 \}$. Then $\partial B = B$ and $(\partial B)^\epsilon = \bigcup_{n>2} (a_n - \epsilon , a_n + \epsilon)$. Therefore, for $0 < \epsilon < a_3$,

$$\int_{(\partial B)^\epsilon} \exp(-x^2/2)\, dx$$

$$\leq \sum_{n>2} \int_{(a_n-\epsilon,\, a_n+\epsilon)} \exp(-x^2/2)\, dx$$

$$\leq 2\epsilon \sum_{n>2} \exp(-(a_n-\epsilon)^2/2)$$

$$\leq 2\epsilon \sum_{n>2} \exp(-(a_n^2 - 2\epsilon\, a_n)/2).$$

Now choose $\sigma = (16\Lambda)^{-1/4}$ and let N be an integer such that $(1-2\sigma)a_N > 2a_3$. Then,

$$\int_{(\partial B)^\epsilon} \exp(-x^2/2)\,dx \le 2\epsilon\,(N + \sum_{n>N} \exp(-\sigma a_n^2)) = 0(\epsilon).$$

Hence $\mathscr{B} = \{B\}$ satisfies condition (2.3). Now for sufficiently small $\epsilon > 0$ and for all integer n satisfying $(n+1)\epsilon \ge 1$, $a_{n+1} - a_n < \epsilon$. Write $a = (-c \log \epsilon)^{1/2}$. Then,

$$\begin{aligned} &\int_{(\partial B)^\epsilon} \exp(-x^2/2\Lambda)\,dx \\ &\ge \int_{(a,\infty)} \exp(-x^2/2\Lambda)\,dx \\ &\ge (a^{-1} - a^{-3}\Lambda)\exp(-a^2/2\Lambda) \\ &= a^{-1}(1-a^{-2}\Lambda)\,\epsilon^{c/2\Lambda}. \end{aligned}$$

Hence, it follows that

$$\begin{aligned} &\lim_{\epsilon\downarrow 0} \epsilon^{-1} \int_{(\partial B)^\epsilon} \exp(-x^2/2\Lambda)\,dx \\ &\ge \lim_{\epsilon\downarrow 0} a^{-1}(1-a^{-2}\Lambda)\,\epsilon^{(c-2\Lambda)/2\Lambda} = +\infty. \end{aligned}$$

So, $\mathscr{B}$ does not satisfy condition (2.2).

REMARK 2.4 : Condition (A.4) may be difficult to verify in some situations. A sufficient condition for (A.4) is given in Bhattacharya and Ghosh (1978) as assumption (A6) on page 439. In our set up, this can be stated as : (A6) of Bhattacharya and Ghosh (1978) together with the assumptions that C in (A1) is open and the matrix $((E\,\psi_i(X_1, \theta_0) \cdot \psi_j(X_1, \theta_0)))$ is nonsingular.

2.3. Proofs.

First we state and prove some lemmas.

LEMMA 3.1. *If* E $\|Z_1\|^s < \infty$ *for some* $s \geq 3$ *and* Z_1 *satisfies* (A.4), *then for almost all sample sequences and for sufficiently large* n,

$$|P_n(\sqrt{n}\,(\dot{\bar{Y}} - \bar{Z}) \in B) - \int_B (1 + \sum_{r=1}^{s-2} n^{-r/2}\, b_r(F_n, x))\, d\Phi_{S_n}(x)|$$

$$\leq o(n^{-(s-2)/2}) + c_1\, \Phi_{S_n}((\partial B)^{e^{-dn}})$$

for every Borel set B *in* $\mathbb{R}^k$. *Here* $d > 0$, c_1 *are constants (independent of the sample sequence) and* $b_r(F_n, \cdot)$, $r = 1, \ldots, s-2$ *are polynomials whose co-efficients are continuous functions of moments of* F_n *of order* s *or less.*

Lemma 3.1 is an easy consequence of Theorem 2 in Babu and Singh (1984). So we omit the proof. The next lemma gives an almost sure asymptotic expansion for the distribution of $\sqrt{n}\,(H(\bar{Y}) - H(\bar{Z}))$.

LEMMA 3.2 : *Let* $Q = \{x \in \mathbb{R}^k : \|x - \mu\| < \delta_1\}$ *for some* $\delta_1 > 0$ *and let* $H: \mathbb{R}^k \to \mathbb{R}^m$ *have continuous partial derivatves of all orders on* Q. *If Grad* $H(\mu)$ *is of full rank then, for almost all sample sequences,*

$$\sup_{B \in \mathcal{B}} |P_n(\sqrt{n}\,(H(\bar{Y}) - (H(\bar{Z})) \in B) - \int_B (1 + \sum_{r=1}^{s-2} n^{-r/2}\, a_r(F_n, x))\, d\,\Phi_{M_n}(x)|$$

$$= o\,(n^{-(s-2)/2})$$

where $a_1(.\,,.), \ldots, a_{s-2}(.\,,.)$ *and* $\mathcal{B}$ *are as in the statement of part (c) of the Theorem.*

PROOF OF LEMMA 3.2: without loss of generality, we may assume that the first m–columns of Grad $H(\mu)$ are linearly independent. Write,

$$\gamma_{s,n}(x) = (1 + \sum_{r=1}^{s-2} n^{-r/2} b_r(F_n, x))\, \phi_{S_n}(x), \quad x \in \mathbb{R}^k \text{ and}$$

$$g_n(x) = \sqrt{n}\,(H(\bar{Z} + x/\sqrt{n}) - H(\bar{Z})), \quad x \in \mathbb{R}^k \quad \text{so that}$$

$$\sqrt{n}\,(H(\bar{Y}) - H(\bar{Z})) = g_n(\sqrt{n}\,(\bar{Y} - \bar{Z})$$

First, we show that

$$\text{(3.1)} \qquad \int_{(g_n^{-1}B)} \gamma_{s,n}(x)\,dx$$

$$= \int_B (1 + \sum_{r=1}^{s-2} n^{-r/2} a_r(F_n, x))d\,\Phi_{M_n}(x) + o(n^{-(s-2)/2})$$

holds uniformly over all Borel sets B in $\mathbb{R}^m$. To that effect let V_n denote the set $\{x \in \mathbb{R}^k : \|x\| \leq \log n\}$ and define the function $k_n : V_n \to \mathbb{R}^k$ by

$$k_n(x) = \begin{bmatrix} g_n(x) \\ (x)_{m+1}^{k} \end{bmatrix}$$

where $(x)_{m+1}^{k}$ denotes the vector of last (k–m) elements of $x \in \mathbb{R}^k$. Then,

$$\text{Grad } k_n(x) = \begin{bmatrix} \text{Grad} & g_n(x) \\ 0 & I_{k\text{-}m} \end{bmatrix}$$

By SLLN, $\bar{Z} \to \mu$ almost surely (P). Therefore k_n has continuous partial derivatives of all orders and a non–singular gradient on V_n eventually, a.s.(P). For sufficiently large n;

(3.2)
$$\int_{(g_n^{-1}B)} \gamma_{s,n}(x)dx$$

$$= \int_{(g_n^{-1}B)\cap V_n} \gamma_{s,n}(x)dx + o(n^{-(s-2)/2})$$

$$= \int_{\{(\omega)_1^m \in B\}\cap k_n(V_n)} \gamma_{s,n}(k_n^{-1}(\omega)).|\det \text{Grad } k_n(k_n^{-1}(\omega))|^{-1} d\omega$$

$$+o(n^{-(s-2)/2})$$

where $(\omega)_1^m$ is the vector of first m elements of $\omega \in \mathbb{R}^k$.

Next, we approximate $\det \text{Grad } k_n(x)$ by taking co–ordinatewise Taylor's expansion.

$$\det \text{Grad } k_n(x)$$

$$= \det \begin{bmatrix} \text{Grad } H(\bar{Z} + x/\sqrt{n}) \\ 0 \quad I_{k-m} \end{bmatrix}$$

$$= \det \begin{bmatrix} \text{Grad } H(\bar{Z}) + \sum_{r=1}^{s-2} n^{-r/2} A_{r,n}(x) + n^{-(s-1)/2} R_n(x) \\ 0 \qquad\qquad I_{k-m} \end{bmatrix}$$

Here $A_{r,n}(x)$ are $m \times k$ matrices of polynomials in x and $R_n(x)$ is a $m \times k$ matrix which satisfies $\|R_n(x)\| \le c_2 \cdot \|x\|^{s-1}$, $x \in V_n$ eventually, a.s. for some nonrandom constant c_2.

With $B_n = \begin{bmatrix} \text{Grad } H(\bar{Z}) \\ 0 \quad I_{k-m} \end{bmatrix}$, we have

$$\text{(3.3)} \qquad \det \text{Grad } k_n(x)$$

$$= (\det B_n)(1 + q_{1,n}(n^{-1/2}x) + n^{-(s-1)/2}R_{1n}(x))$$

where $q_{1,n}$ is a polynomial of degree $\le (s-2)$ and the remainder term R_{1n} is $o(\sqrt{n})$ uniformly on V_n. Therefore , for all large n, we can write

$$\text{(3.4)} \qquad (\det \text{Grad } k_n(x))^{-1}$$

$$= (\det B_n)^{-1}(1 + q_{2,n}(n^{-1/2}x) + n^{-(s-1)/2} R_{2,n}(x))$$

where $q_{2,n}$ and $R_{2,n}$ respectively have properties similar to $q_{1,n}$ and $R_{1,n}$ in (3.3). Next observe that for almost all sample sequences, there exists a $\delta > 0$ such that $\{\bar{Z} + x: \|x\| \le \delta\} \subseteq Q$ for sufficiently large values of n. Define the function Γ_n on $E = \{x: \|x\| \le \delta\}$ by

$$\Gamma_n(x) = \begin{bmatrix} H(\bar{Z}+x) \quad -H(\bar{Z}) \\ (x)_{m+1}^{k} \end{bmatrix}, x \in E.$$

Then, $k_n(x) = \sqrt{n}\,\ddot{\Gamma}_n(n^{-1/2}x)$, $x \in V_n$ holds for all n such that $\log n \le \delta \cdot \sqrt{n}$. Notice that Γ_n is a diffeomorphism (cf. Milnor (1965), page 4) onto its image. Hence Γ_n^{-1} has continuous partial derivatives of all orders. In particular we can express Γ_n^{-1} as the sum of a vector of polynomials $q_{3,n}$ and a remainder term $0(\|\omega\|^s)$. As a consequence, for all $\omega \in k_n(V_n)$,

$$(3.5) \qquad k_n^{-1}(\omega) = n^{1/2}(\Gamma_n^{-1}(n^{-1/2}\omega))$$

$$= n^{1/2}\, q_{3,n}(n^{-1/2}\,\omega) + \sqrt{n}\, 0(\|n^{-1/2}\omega\|^s)$$

$$= B_n^{-1}\omega + \sum_{r=1}^{s-2} n^{-r/2}\, q_{4,r,n}(\omega) + n^{-(s-1)/2}\cdot 0(\|\omega\|^s)$$

where for r = 1, ..., (s–2), $q_{4,r,n}$ is a vector of polynomials. Now, using (3.4) and (3.5) in (3.2), we have

$$\int_{g_n^{-1}B} \gamma_{s,n}(x)\, dx$$

$$= \int_{\{(\omega)_1^m \epsilon B\} \cap k_n(V_n)} \gamma_{s,n}(k_n^{-1}(\omega)).\ |(\det\ B_n)^{-1}(1 + q_{2,n}(n^{-1/2}k_n^{-1}(\omega))|\, d\omega$$

$$+ o(n^{-(s-2)/2}).$$

$$= |\det\ B_n|^{-1}.$$

$$\int_{\{(\omega)_1^m \epsilon B\} \cap k_n(V_n)} \gamma_{s,n}(n^{1/2}q_{3,n}(n^{-1/2}\omega))(1 + q_{2,n}(q_{3,n}(n^{-1/2}\omega)))\, d\omega$$

$$+ o(n^{-(s-2)/2})$$

$$= \int_{\{(\omega)_1^m \epsilon B\} \cap k_n(V_n)} (1 + \sum_{r=1}^{s-2} n^{-r/2} a_{1,r}(F_n, \omega))\, d\Phi_{M_n}(\omega) + o(n^{-(s-2)/2})$$

where $a_{1,r}(F_n,\cdot)$, r = 1,...,(s–2) are polynomials whose co–efficients are continuous functions of moments of F_n of order s or less. Now, integrate out the variables $(\omega_{m+1},...\omega_k)$ to get (3.1).

Next, write $\xi_{s,n}(x) = (1 + \sum_{r=1}^{s-2} n^{-r/2} a_r(F_n, x))\, \phi_{M_n}(x),\quad x \in \mathbb{R}^m$. From Lemma 3.1, it follows that for almost all sample sequences and for large n,

$$(3.6)\qquad |P_n(\sqrt{n}(H(\bar{Y}) - H(\bar{Z})) \in B) - \int_B \xi_{s,n}(x)dx|$$

$$\leq o(n^{-(s-2)/2}) + c_1\, \Phi_{S_n}((\partial g_n^{-1}B)^{e^{-d.n}})$$

for every Borel subset B of $\mathbb{R}^m$. Following the arguments given in Bhattacharya and Ghosh (1978) (page 444–445) it can be shown that there exists a constant $a > 0$ such that for large n

$$(3.7)\qquad \Phi_{S_n}((\partial g_n^{-1}B)^{e^{-dn}})) \leq \Phi_{M_n}((\partial B)^{e^{-an}} \cap V_n) + o(n^{-(s-2)/2})$$

holds for every Borel set $B \subseteq \mathbb{R}^m$. Next use condition (2.2) to conclude that

$$\sup_{B \in \mathscr{B}} \Phi_{M_n}((\partial B)^{e^{-an}} \cap V_n) = o(n^{-(s-2)/2})$$

This completes the proof of lemma 3.2.

LEMMA 3.3. *Let,* $U_1,\ldots,U_n$ *be i.i.d. random vectors with common mean* β. *Let* Λ *denote the largest eigen value of the dispersion matrix of* U_1. *Suppose that* $E\|U_1\|^s < \infty$ *for some integer* $s \geq 3$. *Then,*

$$P(\sqrt{n}\, \|\bar{U} - \beta\| > ((s-1)\, \Lambda \log n)^{1/2}) \leq J.n^{-(s-2)/2} (\log n)^{-s/2}$$

where $\bar{U}_n = n^{-1} \sum_{i=1}^{n} U_i$ *and* J *is a function of* Λ *which is bounded on bounded set of values of* Λ.

PROOF OF LEMMA 3.3 : See Von Bahr (1967).

We are now ready to prove the theorem.

PROOF OF (a): By assumption (A1), $\psi_1,\dots,\psi_m$ have continuous partial derivatives of order s on $C \times \Theta$. Taking Taylor's expansion of $\psi_i(x,\cdot)$ around θ_o for i = 1,...,m, we have

$$\text{(3.8)} \qquad \psi_i(x,t) = \psi_i(x,\theta_o) + \sum_{1\le|\nu|\le s-1} (t-\theta_o)^\nu D^\nu \psi_i(x,\theta_o)/\nu! + R_{n,i}(x,t)$$

where the remainder term $R_{n,i}(x, t)$ satisfies

$$|R_{n,i}(x, t)| \le c\,\|t - \theta_o\|^s \max_{|\nu|=s} \sup_{\|\theta-\theta_o\|<\|t-\theta_o\|} |D^\nu \psi_i(x, \theta)|$$

for some constant c. Using (3.8), rewrite equation (1.3) as

$$\text{(3.9)} \qquad 0 = \Upsilon_{0,i} + \sum_{1\le|\nu|\le s-1} (t-\theta_o)^\nu . \Upsilon_{\nu,i}/\nu! + R^*_{n,i}(t)$$

where $R^*_{n,i}(t) = \frac{1}{n}\sum_{j=1}^{n} R_{n,i}(X^*_j, t)$.

By SLLN, the dispersion matrix of $Y_1^{(\nu)}$ under P_n

$$= \frac{1}{n}\sum_{j=1}^{n} Z_j^{(\nu)} \cdot Z_j^{(\nu)\prime} - \bar{Z}^{(\nu)} \cdot \bar{Z}^{(\nu)\prime}$$

$$\xrightarrow{a.s.} E_{\theta_o} Z_1{}^{(\nu)} \cdot Z_1^{(\nu)\prime} - \mu^{(\nu)} \cdot \mu^{(\nu)\prime}$$

for all ν, $0 \le |\nu| \le s-1$ and hence are bounded in norm.

By Lemma 3.3, there exist constants d_2 , d_3 such that almost surely (P),

$$P_n(\|Y^{(\nu)} - E_n Y_1^{(\nu)}\| > d_2 n^{-1/2}(\log n)^{1/2}) < d_3 n^{-(s-2)/2}(\log n)^{-s/2}$$

for $0 \leq |\nu| \leq s-1$, when n is sufficiently large. Also note that by the LIL, $\|E_n Y_1^{(\nu)} - \mu_\nu\| = 0(n^{-1/2}(\log\log n)^{1/2})$ almost surely (P). Therefore it follows that for almost all sample sequence, there exist an integer $n_o \geq 1$ such that for all $n \geq n_o$,

$$\text{(3.10)} \qquad P_n(\|Y^{(\nu)} - \mu_\nu\| > d_4\, n^{-1/2}(\log n)^{1/2}))$$

$$< d_3\, n^{-(s-2)/2}(\log n)^{-s/2} \quad \text{for some constant } d_4 > 0.$$

Set $R_n^*(t) = (R_{n,1}^*(t), \ldots, R_{n,m}^*(t))'$. By similar arguments, it can be shown that for almost all sample sequences, there exists a constant d_5 such that (without loss of generality)

$$\text{(3.11)} \qquad P_n(\|R_n^*(t)\| > \|t-\theta_o\|^s\,(d_5 + d_4 n^{-1/2}(\log n)^{1/2}))$$

$$< d_3\, n^{-(s-2)/2}\,(\log n)^{-s/2} \quad \text{for all} \quad n \geq n_o.$$

Hence, for almost all sample sequences, there exists $n_o \geq 1$ such that for all $n \geq n_o$, outside a set of P_n–probability $d_6\, n^{-(s-2)/2}(\log n)^{-s/2}$, we can write (3.9) as

$$\text{(3.12)} \qquad (t-\theta_o) = (D+\eta_n^*)^{-1}(\delta_n^* + \sum_{2\leq|\nu|\leq s-1}(t-\theta_o)^\nu\,\mu_\nu/\nu! + d_7\|t-\theta_o\|^s\,\epsilon_n^*)$$

where d_6 and d_7 are cönstants and η_n^*, δ_n^* and ϵ_n^* are random elements depending on $(X_1^*,\ldots,X_n^*)$ and norms of η_n^* and δ_n^* are $0(n^{-1/2}(\log n)^{1/2}$ while $\|\epsilon_n^*\| \leq 1$.

Hence, there exist an integer $n_1 \geq n_o$(depending on the sample sequence) and a constant d_8 such that for all $n \geq n_1$, r.h.s. of (3.12) is less than $d_8\, n^{-1/2}(\log n)^{1/2}$ whenever $\|t - \theta_o\|$ is less than $d_8\, n^{-1/2}(\log n)^{1/2}$. By Brower's fixed point theorem (Milnor (1965), page 14) it follows that there exist statistics $\{\theta_n^*\}$ such that for all $n \geq n_1$,

$$\text{(3.13)} \qquad P_n(\|\theta_n^* - \theta_o\| < d_8\, n^{-1/2}(\log n)^{1/2},\ \theta_n^* \text{ solves (1.3)})$$

$$> 1 - d_6\, n^{-(s-2)/2}\, (\log n)^{-s/2}.$$

This completes the proof of part (a).

PROOF OF (b): The proof is essentially the same as that of (a). Only exception is that we use LIL instead of Lemma 3.3 to get bounds on the deviations $\|Z^{(\nu)} - \mu^{(\nu)}\|$ for $0 \leq |\nu| \leq s-1$. This is also pointed out in Remark 1.8 of Bhattacharya and Ghosh (1978).

PROOF 0F (c): Using (3.8) we can write

$$0 = \frac{1}{n}\sum_{j=1}^{n} \psi_i(X_j\,, \hat{\theta}_n)$$

$$= Z_{0,i} + \sum_{1\leq|\nu|\leq s-1} (\hat{\theta}_n - \theta_o)^{\nu}\ Z_{\nu,i}\,/\nu! + \bar{R}_{n,i}$$

where $\bar{R}_{n,i} = n^{-1}\sum_{j=1}^{n} R_{n,i}(X_j, \hat{\theta}_n)$. Set $\bar{R}_n = (\bar{R}_{n1},\ldots, \bar{R}_{nm})'$. Then by assumption (A2), it follows that there exists a constant $d_9 > 0$ such that

$$\text{(3.14)} \qquad P\,(\|\bar{R}_n\| < d_9\, n^{-s/2}(\log n)^{s/2} \text{ eventually}) = 1$$

For $i = 1,\dots,m$ define the function f_i: $\mathbb{R}^{k+m} \to \mathbb{R}$ by

$$f_i(\omega, \theta) = \omega_{0,i} + \sum_{1\le|\nu|\le s-1} \omega_{\nu,i} (\theta - \theta_o)^\nu/\nu!$$

where $\omega = (\omega_{\nu,i})_{0\le|\nu|\le s-1;\ 1\le i\le m} \in \mathbb{R}^k$, $\theta \in \mathbb{R}^m$. Then, $f = (f_1,\dots,f_m)$ has continuous partial derivatives of all orders, $f(\mu, \theta_o) = 0$ and by assumption $D = ((D_{k+j} f_i(\mu,\theta_o)))_{1\le i,j\le m}$ is non–singular. Hence by the implicit function theorem, there exists a unique function $H : \mathbb{R}^k \to \mathbb{R}^m$ and a neighborhood Q of μ such that

$$f(\omega, H(\omega)) = 0$$

for all $\omega \in Q$ and H has continuous partial derivatives of all orders. Now, by LIL and (3.14),

$$\|\bar{Z}^{(\nu)} - \mu_\nu\| < d_o\, n^{-1/2}(\log n)^{1/2} \quad \text{for } 1 \le |\nu| \le s-1$$

$$\|\bar{Z}^{(0)} + \bar{R}_n\| < d_o\, n^{-1/2} (\log n)^{1/2}$$

hold eventually, almost surely under P for some constant $d_o > 0$. Hence, by the uniqueness of H, we have

$$\hat{\theta}_n = H(\tilde{Z}) \tag{3.15}$$

where $\tilde{Z} = (\tilde{Z}_{\nu,i})$ is given by

$$\tilde{Z}_{\nu,i} = \bar{Z}_{\nu,i} \quad \text{for } 1 \le |\nu| \le s-1 \text{ and } i = 1,\dots,m$$

$$\tilde{Z}_{0,i} = \bar{Z}_{0,i} + \bar{R}_{n,i} \quad \text{for } i = 1,\dots,m.$$

This gives the almost sure representation for $\hat{\theta}_n$.

Next expand the r.h.s of the equation $0 = n^{-1} \sum_{j=1}^{n} \psi_i(X_j^*, \theta_n^*)$ into Taylor's series around θ_0 as in (3.8). Using (3.10) and (3.11) it can be shown (exactly in the same way as in Bhattacharya and Ghosh (1978)) that for almost all sample sequence, outside a set of P_n-probability $O(n^{-(s-2)/2}(\log n)^{-s/2})$, θ_n^* has the representation

$$\theta_n^* = H(\tilde{Y}).$$

Here, $\tilde{Y} = (\tilde{Y}_{\nu,i})$ is defined as

$$\tilde{Y}_{\nu,i} = Y_{\nu,i} \text{ for } 1 \le |\nu| \le s-1 \text{ and } i=1,\dots,m$$

$$\tilde{Y}_{0,i} = Y_{0,i} + R_{ni}^*(\theta_n^*) \text{ for } i = 1,\dots,m$$

By (3.11) and (3.13), it follows that for almost all sample sequences, there exists a constant $d_{10} > 0$ such that

$$\text{(3.17)} \qquad P_n(\| R_n^*(\theta_n^*) \| > d_{10}\, n^{-s/2} (\log n)^{s/2}) = 0(n^{-(s-2)/2} (\log n)^{-s/2})$$

Fix $0 < \delta_2 < \delta_1$. Since H has continuous partial derivatives on Q, by the mean value theorem, there is a constant $d_{11} > 0$ such that whenever ω_1, ω_2 lies in $\{\omega : \|\mu - \omega\| \le \delta_2\}$,

$$\text{(3.18)} \qquad \|H(\omega_1) - H(\omega_2)\| < d_{11} \|\omega_1 - \omega_2\|.$$

Write $D_n = \sqrt{n}\,(\hat{\theta}_n - \theta_n^*) - \sqrt{n}\,(H(\tilde{Y}) - H(\tilde{Z}))$. Then (3.14), (3.17) and (3.18) jointly imply that for almost all sample sequences, there exists a constant $d_{12} > 0$ such that

(3.19) $$P_n(\|D_n\| > d_{12}\, n^{-(s-1)/2}\, (\log n)^{s/2})$$

$$= P_n\,(\|(H(\tilde{Y}) - H(Y)) - (H(\tilde{Z}) - H(Y)\| > d_{12}\, n^{-s/2}\, (\log n)^{s/2})$$

$$\leq P_n(\|R_n^*(\theta_n^*)\| > d_{10}\, n^{-s/2}\, (\log n)^{s/2})$$

$$= 0(n^{-(s-2)/2}(\log n)^{-s/2}).$$

Let $\epsilon_n = d_{10}\, n^{-(s-1)/2}\, (\log n)^{s/2}$. Then, it follows from Lemma 3.2 and (3.6) that

(3.20) $$\sup_{B\in\mathcal{B}} |\, P_n(\sqrt{n}\,(\theta_n^* - \hat{\theta}_n) \in B) - \int_B \xi_{s,n}(x)\, dx|$$

$$\leq \sup_{B\in\mathcal{B}} |\, P_n(\sqrt{n}\,(\theta_n^* - \hat{\theta}_n) \in B)\; - P_n(\sqrt{n}\,(H(Y) - H(Z)) \in B)| + o(n^{-(s-2)/2})$$

$$\leq P_n(\|D_n\| > \epsilon_n) + \sup_{B\in\mathcal{B}} P_n(\sqrt{n}(H(Y) - H(Z) \in (\partial B)^{\epsilon_n}) + o(n^{-(s-2)/2})$$

$$= 0\,(\sup_{B\in\mathcal{B}} \Phi_{M_n}((\partial B)^{\alpha\epsilon_n} \cap V_n)) + o(n^{-(s-2)/2}).$$

for some constant $\alpha > 0$. Now use the smoothness of H at μ and the LIL to get

$$\|\, M_n^{-1} - M^{-1}\, \| = 0\,(n^{-1/2}(\log\log n)^{1/2}))\quad \text{a.s.}(P).$$

Hence, it follows that

$$0\,(\sup_{B\in\mathscr{B}} \Phi_{M_n}(\partial B)^{\epsilon_n \alpha} \cap V_n))$$

$$= 0\,(\sup_{B\in\mathscr{B}} \Phi_M((\partial B)^{\epsilon_n \alpha}))$$

$$= 0(\epsilon_n) = o(n^{-(s-2)/2}).$$

This completes this proof of (c).

PROOF OF (d): Write $\beta_n = \epsilon_n \|M_n^{-1/2}\|$, $n \geq 1$. Then, as in the derivation (3.20), one can show that for almost all sample sequences,

$$\sup_{B\in\mathscr{B}_1} |P_n(\sqrt{n}\, M_n^{-1/2}(\theta_n^* - \hat{\theta}_n) \in B) - \int_{M_n^{1/2}B} \xi_{s,n}(x)\, dx|$$

$$\leq o(n^{-(s-2)/2}) + 0\,(\sup_{B\in\mathscr{B}_1} \Phi_{M_n}(\partial M_n^{1/2}B)^{\epsilon_n}))$$

$$= o(n^{-(s-2)/2}) + 0(\sup_{B\in\mathscr{B}_1} \Phi_{M_n}(M_n^{1/2}(\partial B)^{\beta_n}))$$

$$= o(n^{-(s-2)/2})$$

The last step follows by the condition (2.3). By exactly similar arguments as in the bootstrap case, it follows that

$$\sup_{B\in\mathscr{B}_1} |P\,(\sqrt{n}\, M^{-1/2}(\hat{\theta}_n - \theta_o) \in B) - \int_{M^{1/2}B} \overline{\xi}_{s,n}(x)\, dx| = o(n^{-(s-2)/2})$$

where $\xi_{s,n}(\cdot) = (1 + \sum_{r=1}^{s-2} n^{-r/2} a_r(\theta_o, \cdot))\, \phi_M(\cdot)$ and $a_r(\theta_o, \cdot)$, are polynomials obtained by replacing the moments of F_n by the corresponding moments of F_{θ_o}. Hence, the result follows from the SLLN and the continuity of the co–efficients of the polynomials $a_r(\cdot, \cdot)$ in the moments of the corresponding distributions.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Babu G.J. and Singh K (1983). Inference on means using bootstrap. *Ann. statist.* 11, 999 – 1003.

2. Babu G. J. and Singh K (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhya Ser A*, *46*, 219 – 232.

3. Bahr, B. Von (1967). On the central limit theorem in $\mathbb{R}^k$, *Ark. Mat* *7*, 61 – 69.

4. Beran R. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* 10, 212 – 225.

5. Bhattacharya, P.K. (1967). Estimation of a probability density function and its derivatives. *Sankhya Ser.* A 29, 373 – 382.

6. Bhattacharya R.N. (1985). Some recent results on Cramer–Edgeworth expansions with applications. *Multivariate Analysis* – VI, P.R. Krishnaiah (editor) 57 – 77.

7. Bhattacharya R. N. and Ghosh J. K. (1978). On the Validity of the formal Edgeworth expansion. *Ann. Statist.* *6*, 434 – 451.

8. Bickel P. and Freedman (1983). Bootstrapping regression models with many parameters. In *A Festschrift For Elrich L. Lehmann* (P.J.Bickel, K.A.Docksum, and J.L.Hodges Jr. eds.) 28 – 48, Wadsworth, Belmont, Calif.

9. Bose A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* 16, 1709 – 1722.

10. Feller, W. (1966). *An Introduction to probability theory and its applications.* Vol. 2, Weily, New York.

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

12. Hall P. (1988). Rate of convergence in bootstrap approximation. *Ann.* Probab. 16, 1665 – 1684.

13. Helmers (1988). Bootstrap approximations for studentized U–statistics. Preprint.

14. Hoeffding, W. (1963). Probability inequalities for sums of independent bounded random variables. *J. Amer. Statist. Assoc.* 58, 13 – 30.

15. Huber, P.J. (1973). Robust regression : Asymptotics, Conjectures and Monte Carlo. *Ann. Statist.* 1, 799 – 821.

16. Huber, P.J. (1981). ***Robust Statistics.*** Weily, New York.

17. Liu R. (1988). Bootstrap procedures under some non i.i.d. models. *Ann. Statist.* 16, 1696 – 1708.

18. Milnor J. W. (1965). ***Topology from the differentiable viewpoint.*** Univ. Press of Virginia, Charlottesville.

19. Ringland, J.T. (1983). Robust multiple comparisons. *J. Amer. Statist. Assoc.* 78, 145 – 151.

20. Schuster, E.F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* 40, 1187 – 1195.

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

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