

Mixtures of Global and Local Edgeworth Expansions and Their Applications

Gutti Jogesh Babu*

Pennsylvania State University

and

Z. D. Bai

National Sun Yat-Sen University, Kaohsiung, Taiwan

Edgeworth expansions which are local in one coordinate and global in the rest of the coordinates are obtained for sums of independent but not identically distributed random vectors. Expansions for conditional probabilities are deduced from these. Both lattice and continuous conditioning variables are considered. The results are then applied to derive Edgeworth expansions for bootstrap distributions, for Bayesian bootstrap distribution, and for the distributions of statistics based on samples from finite populations. This results in a unified theory of Edgeworth expansions for resampling procedures. The Bayesian bootstrap is shown to be second order correct for smooth positive “priors,” whenever the third cumulant of the “prior” is equal to the third power of its standard deviation. Similar results are established for weighted bootstrap when the weights are constructed from random variables with a lattice distribution. © 1996 Academic Press, Inc.

1. INTRODUCTION

Let $\{a_{1,N}, \dots, a_{N,N}\}$ be a sequence of row vectors in \mathbb{R}^k . Let $\{Y_j\}$ be a sequence of i.i.d. random variables. We obtain Edgeworth expansions for

$$P\left(\sum_{j=1}^N a_{j,N}(Y_j - E(Y_j)) \in H, \sum_{j=1}^N Y_j = n\right),$$

Received June 24, 1994; revised July 1996.

AMS 1991 subject classifications: 62E20, 60F05.

Key words and phrases: Asymptotic expansions, Bayesian bootstrap, bootstrap, Chebyshev–Hermite polynomial, Dirchlet distribution, expansions for conditional distributions, gamma distribution, lattice distribution, local limit theorems, sampling without replacement, weighted bootstrap.

* Research supported in part by NSF Grants DMS-9208066 and DMS-9626189.

when Y_1 has a lattice distribution. Such a result is a combination of global and local expansions. The result does not follow from any of the known expansions on lattice or strongly non-lattice structures, as the vector considered is neither. From this, expansions for conditional probabilities

$$P\left(\sum_{j=1}^N a_{j,N}(Y_j - E(Y_j)) \in H \mid \sum_{j=1}^N Y_j = n\right), \quad (1)$$

are derived using local expansions for $P(\sum_{j=1}^N Y_j = n)$.

In a fundamental paper Bhattacharya and Ghosh (1978) have demonstrated that Edgeworth expansions for a wide class of statistics can be derived from Edgeworth expansions for multivariate sample means. This technique has been used by Babu and Singh (1983, 1984) to show the superiority of the bootstrap method and by Babu and Singh (1985) to obtain Edgeworth expansions for the ratio statistic and similar statistics based on samples from finite populations. The method is also used by Babu and Singh (1989) to obtain global Edgeworth expansions for functions of means of random vectors, when one of the coordinates has a lattice distribution and the remaining part of the vector has a strongly non-lattice distribution. In this paper we concentrate on sample means of k -variate random vectors. The Edgeworth expansions for smooth functions of multivariate sample means follow from similar expansions for multivariate means as in Bhattacharya and Ghosh (1978).

We shall also consider the case where Y_1 is absolutely continuous and obtain Edgeworth expansion for $(\sum_{j=1}^N a_{j,N} Y_j) / (\sum_{j=1}^N Y_j)$. Whenever the density h_N of a sum of non-identically distributed random vectors $(\sum_{j=1}^N a_{j,N}(Y_j - E(Y_j)), \sum_{j=1}^N Y_j)$ exists, we use the notation

$$P(H, z; N) = \int_H h_N(x, z) dx. \quad (2)$$

In order to get Edgeworth expansions for (2), we require a result similar to Theorem 19.3 of Bhattacharya and Ranga Rao (1986) with a better error estimate, under assumptions weaker than their condition (19.29). Condition (19.29) is too stringent for the applications considered in this paper. The main results are stated in Section 2. Applications of these results to the classical bootstrap, Bayesian bootstrap, weighted bootstrap, and to the estimators based on samples drawn without replacement from a finite population, are given in Section 3. Brief sketches of the proofs of the main results are presented in Section 4. Technical lemmas required in the proofs of the theorems are given in the Appendix. Lemma 5 in the Appendix describes a sieve method, which is used to obtain bounds for the integral of a characteristic function.

2. MAIN RESULTS

We introduce some notation before stating the results. Let \hat{f} denote the characteristic function of a non-degenerate random variable Y_1 . To simplify the notation we drop the subscript N from $a_{j,N}$ and denote it by a_j instead. Let $x = (x_1, \dots, x_k)$ denote a row vector in \mathbb{R}^k . For any integer $r \geq 1$, let $\phi_r(z_1, \dots, z_r) = (2\pi)^{-r/2} \exp(-2^{-1} \sum_{i=1}^r z_i^2)$ and $\phi = \phi_1$. For a real valued measurable function h on \mathbb{R}^k , $\delta > 0$ and $S \geq 3$, let

$$M_{h,s} = \sup_x |h(x)| (1 + \|x\|)^{-s},$$

$$\omega(h, \delta, x) = \sup_{\|x - z\| < \delta} |h(x) - h(z)|, \quad \text{and} \quad \omega(h, \delta) = \int \omega(h, \delta, x) \phi_k(x) dx.$$

Furthermore let

$$\mu = E(Y_1), \quad \sigma^2 = \text{Var}(Y_1) > 0, \quad \gamma_3 = E(Y_1 - \mu)^3 \sigma^{-3}, \quad (3)$$

$$V_N^2 = N^{-1} \sum_{j=1}^N a_j' a_j, \quad y_n = (n - N\mu)/\sigma \sqrt{N}, \quad (4)$$

and for any row vector t let

$$d_N(t) = N^{-1} \sum_{j=1}^N e^{it a_j'}. \quad (5)$$

Note that a_j need not be centered in defining the matrix V_N^2 . Later in the statements of the theorems, $\sum_{j=1}^N a_j$ is assumed to be $= 0$.

Under the assumptions of the theorems stated below, V_N^2 is positive definite for all large N ; see (35) of the Appendix. Hence there exists a non-singular matrix V_N^{-1} such that $(V_N^{-1}) V_N^2 V_N^{-1}$ is the identity matrix. Define

$$U_N = \sum_{j=1}^N a_j V_j^{-1} Y_j / (\sigma \sqrt{N}) \quad \text{and} \quad W_N = \sum_{j=1}^N (Y_j - \mu) / (\sigma \sqrt{N}). \quad (6)$$

Motivation for considering U_N comes from the bootstrap methodology. Suppose $\{Y_j\}$ denotes a sequence of i.i.d. Poisson random variables with mean 1, and $a_j = X_j - \bar{X}_N$, where X_j are univariate random variables, and $\bar{X}_N = N^{-1} \sum_{j=1}^N X_j$. Then $U_N = \sqrt{N} (\bar{X}_N^* - \bar{X}_N) / S_N$ denotes the bootstrapped version of the standardized quantity $\sqrt{N} (\bar{X}_N - \mu_x) / \sigma_x$, where μ_x and σ_x denote the mean and standard deviation of X_1 , and S_N denotes the sample standard deviation. More details on this special case will be provided in Section 3.

We now describe the polynomials $Q_{r,N}$ in $(k+1)$ variables, that appear in Edgeworth expansions,

$$\psi_{s,N} = \phi_{k+1} + \sum_{r=1}^{s-2} N^{-r/2} Q_{r,N} \phi_{k+1}, \quad (7)$$

which in turn occur in the theorems. Let

$$A_m = (-1)^m N^{-1} \sum_{j=1}^N \left(D_{k+1} + \sum_{i=1}^k c_{ji} D_i \right)^m$$

denote a differential operator, where $D_i = \partial/\partial x_i$, x_i is the i th coordinate of x , $D_{k+1} = \partial/\partial y$ and c_{ji} denotes the i th coordinate of the vector $c_j = a_j V_N^{-1}$. Now define $Q_{r,N}(x, y)$ by

$$Q_{r,N}(x, y) = (\phi_{k+1}(x, y))^{-1} \left(\sum^* \prod_{m=1}^r \frac{1}{r_m!} \left(\frac{\gamma_{m+2} A_{m+2}}{(m+2)!} \right)^{r_m} \phi_{k+1}(x, y) \right), \quad (8)$$

where \sum^* denotes the sum over all non-negative integers r_m satisfying $\sum_{1 \leq m \leq r} m r_m = r$.

In particular $Q_{r,N}$ in (8) for $r=1$ is given by

$$\begin{aligned} Q_{1,N}(x, y) &= \frac{\gamma_3}{6} \left(6 \sum_{j \leq j < m < r \leq k} A_{j,m,r} x_j x_m x_r + 3 \sum_{j \leq j \neq r \leq k} A_{j,j,r} (x_j^2 - 1) x_r \right. \\ &\quad \left. + \sum_{j=1}^k A_{j,j,j} (x_j^3 - 3x_j) + 3(\|x\|^2 - k) y + (y^3 - 3y) \right), \\ &= \frac{\gamma_3}{6} \left(\sum_{i=1}^N ((c_i x')^3 - 3(c_i \mathbf{1}')^2 (c_i x')) + 3(\|x\|^2 - k) y + (y^3 - 3y) \right), \end{aligned} \quad (9)$$

where $\mathbf{1} = (1, \dots, 1)$, and $A_{i,m,r} = N^{-1} \sum_{i=1}^N c_{ij} c_{im} c_{ir}$ and γ_3 is defined in (3). If $k=1$, then the first two sums in the definition of $Q_{1,N}$ do not appear. If $k=2$, then the first sum in the definition of $Q_{1,N}$ does not appear.

If Y_1 has a lattice distribution, then without loss of generality, we assume that its span is 1. In this case, let $F_N(H, n) = P(U_N \in H, W_N = y_n)$. We now state the main results.

THEOREM 1. *Suppose Y_1 has a lattice distribution with span 1 and $\sum_{j=1}^N a_j = 0$. For some $M > 0$ and an integer $s \geq 3$, let $E|Y_1|^s < \infty$, and*

$\sum_{j=1}^N \|a_j\|^s < MN$ for all N . Suppose that for any $0 < K < L < \infty$, there exists a $\gamma = \gamma(K, L) < 1$ such that

$$\limsup_{N \rightarrow \infty} \sup_{k \leq \|t\| \leq LN^{(s-3)}} |D_N(t)| < \gamma. \quad (10)$$

Then for any real valued measurable function h on \mathbb{R}^k satisfying $M_{h,s} < \infty$, we have

$$\begin{aligned} & \left| \int_{\mathbb{R}^k} h(x) (\sigma \sqrt{N} F_N(dx, n) - \psi_{s,N}(x, y_n) dx) \right| \\ &= o(M_{h,s} N^{-(s-2)/2}) + O(\omega(h, \delta_N)), \end{aligned}$$

uniformly in y_n for some $\delta_N = o(N^{-(s-2)/2})$.

Remark 1. If Cramér's condition holds for the common distribution of a sequence $\{X_j\}$ of i.i.d. random vectors, and if $a_i = X_i - N^{-1} \sum_{j=1}^N X_j$, then by Lemma 2 of Babu and Singh (1984), condition (10) holds with probability 1 for all $s \geq 3$. If the distribution, which assigns mass $1/N$ to each of the points a_1, \dots, a_N , converges weakly to a strongly non-lattice distribution, then condition (10) holds with $s = 3$. It can be shown, under condition (10), that the eigenvalues of V_N are bounded and bounded away from 0. So if (10) holds, then it also holds when a_j is replaced by $a_j V_N^{-1}$.

If Y_1 has lattice distribution with span 1, then let

$$F_N(H | y_n) = F_N(H, n) / P_N(n),$$

where

$$P_N(n) = P(W_N = y_n) = P\left(\sum_{j=1}^N Y_j = n\right).$$

By Theorem 13 on local Edgeworth expansions on pages 205–206 of Petrov (1975), we have

$$\sigma \sqrt{N} P_N(n) - \zeta_{s,N}(y_n) = o(N^{-(s-2)/2}), \quad (11)$$

uniformly in y_n , where

$$\zeta_{s,N}(y) = \phi(y) + \sum_{j=1}^{s-2} N^{-j/2} q_j(y) \phi(y), \quad (12)$$

and q_j are certain linear combinations of Chebyshev–Hermite polynomials. See equation (1.14) on page 139 of Petrov (1975) for details. In particular

$$q_1(y) = \frac{1}{6} \gamma_3 (y^3 - 3y),$$

where γ_3 is the 3rd cumulant of Y_1/σ defined in (3). By Taylor series method we can expand the ratio $\psi_{s,N}(x, y)/\zeta_{s,N}(y)$ in terms of powers of $N^{-1/2}$. Let $\tilde{\psi}_{s,N}(x|y)$ denote the sum of the terms of the ratio involving $N^{-r/2}$ for non-negative integers $r \leq s-2$. Theorem 2 of Babu and Singh (1984) is a consequence of Remark 1 and Theorem 2 below.

THEOREM 2. *Under the conditions of Theorem 1, we have for any $G > 0$,*

$$\sup_{|y_n| \leq G} \left| \int_{\mathbb{R}^k} h(x)(F_N(dx|y_n) - \tilde{\psi}_{s,N}(x|y_n) dx) \right| \\ = o(M_{h,s} N^{-(s-2)/2}) + O(\omega(h, \delta_N)),$$

for some $\delta_N = o(N^{-(s-2)/2})$.

We now consider the continuous case. If the characteristic function \hat{f} of the distribution of Y_1 satisfies

$$\int |\hat{f}(t)|^r dt < \infty \quad (13)$$

for some $r > 0$, then by Theorem 4.1 of Bhattacharya and Ranga Rao (1986), W_N has a bounded continuous density p_N for all large N . By Theorem 15 on local Edgeworth expansions on pages 206-207 of Petrov (1975), we have

$$p_N(y) - \zeta_{s,N}(y) = o(N^{-(s-2)/2}), \quad (14)$$

uniformly in y , where $\zeta_{s,N}$ is given in (12). If (13) holds, then the density f_N of (U_N, W_N) exists for all large N . Let $\tilde{f}_N(x|y) = f_N(x, y)/p_N(y)$.

For the continuous case, we do not need the full force of condition (10). Instead a weaker version (15) stated in Theorem 3 is sufficient.

THEOREM 3. *Suppose (13) holds for some $r > 0$, and $\sum_{j=1}^N a_j = 0$. For some $M > 0$ and an integer $s \geq 3$, let $E|Y_1|^s < \infty$, and $\sum_{j=1}^N \|a_j\|^s < NM$ for all N . Suppose for any $0 < K < L < \infty$, there exists a $\gamma = \gamma(K, L) < 1$ satisfying*

$$\limsup_{N \rightarrow \infty} \sup_{K \leq \|t\| \leq L} |d_N(t)| < \gamma. \quad (15)$$

Then we have

$$\int_{\mathbb{R}^{k+1}} (1 + \|(x, y)\|)^s |f_N(x, y) - \psi_{s,N}(x, y)| dx dy = o(N^{-(s-2)/2}), \quad (16)$$

and uniformly in y ,

$$\int_{\mathbb{R}^k} (1 + \|x\|)^s |f_N(x, y) - \psi_{s, N}(x, y)| dx = o(N^{-(s-2)/2}). \quad (17)$$

THEOREM 4. *Under the conditions of Theorem 3, we have for any $G > 0$,*

$$\sup_{|y| \leq G} \int_{\mathbb{R}^k} (1 + \|x\|)^s |\tilde{f}_N(x|y) - \tilde{\psi}_{s, N}(x|y)| dx = o(N^{-(s-2)/2}).$$

If h is an indicator function of a set, then the normal measure of the δ_N neighborhood of the boundary of the set will enter into the error term. Thus local variations of h influence the error term, in Theorems 1 and 2. On the other hand, when the density f_N of the distribution of (U_N, W_N) exists, these variations will not have much influence on the error term, only an upper bound of h matters. Thus $h(x, y)$ is replaced by terms such as $1 + \|(x, y)\|^s$ and $1 + \|x\|^s$ in Theorems 3 and 4.

We now study the expansions of the distribution F_N^o of $N\mu U_N/(\sum_{j=1}^N Y_j)$ given $\sum_{j=1}^N Y_j > 0$. We also study the expansions of its density f_N^o , if it exists. Note that if the density exists, then

$$f_N^o(x) = \int f_N(x(1 + (\sigma/\mu \sqrt{N}) y), y)(1 + (\sigma/\mu \sqrt{N}) y)^k dy, \quad (18)$$

and if Y_1 has a lattice distribution, then

$$F_N^o(x) = \sum_{n=1}^{\infty} F_N(x(1 + (\sigma/\mu \sqrt{N}) y_n), n) \Big/ P\left(\sum_{j=1}^N Y_j > 0\right),$$

where y_n is defined in (4). To describe the formal expansion $\psi_{s, N}^c$ of f_N^o we first use Taylor series method to expand $\int \psi_{s, N}(x(1 + (\sigma/\mu \sqrt{N}) y), y)(1 + (\sigma/\mu \sqrt{N}) y)^k dy$ in terms of powers of $1/\sqrt{N}$. Let $\psi_{s, N}^c$ denote the sum of the terms involving $N^{-r/2}$ for nonnegative integers $r \leq s-2$. Similarly, the density $\psi_{s, N}^d$ of the formal expansion of F_N^o is the $(s-1)$ term Taylor expansion, in terms of $1/\sqrt{N}$, of the function $(1/\sqrt{N} \sigma) \sum_{n=1}^{\infty} \psi_{s, N}(x(1 + (\sigma/\mu \sqrt{N}) y_n), y_n)(1 + (\sigma/\mu \sqrt{N}) y_n)^k$. It can be verified easily that $\psi_{3, N}^c = \psi_{3, N}^d$. The next two theorems give expansions f_N^o and F_N^o .

THEOREM 5. *Suppose $P(Y_1 > 0) = 1$ and the conditions of Theorem 3 are satisfied. Then we have*

$$\int_{\mathbb{R}^k} (1 + \|x\|)^s |f_N^o(x) - \psi_{s, N}^c(x)| dx = o(N^{-(s-2)/2}).$$

THEOREM 6. Suppose $P(Y_1 \geq 0) = 1$ and the conditions of Theorem 1 are satisfied. Then we have

$$\int_{\mathbb{R}^k} h(x)(F_N^o(dx) - \psi_{s,N}^d dx) = o(M_{h,s} N^{-(s-3)/2}) + O(\sqrt{N} \omega(h, \delta_N)),$$

where $\delta_N = o(N^{-(s-2)/2})$.

Remark 2. Note that the error term in Theorem 6 is not as sharp as the one given in Theorem 1. For many applications the given error estimates are adequate as Y_1 will have enough moments. However, with some tedious analysis it may be possible to improve the error term.

Remark 3. It can be verified that the one-term Edgeworth expansions are given by

$$\psi_{3,N}(x, y) = (1 + N^{-1/2} Q_{1,N}(x, y)) \phi_{k+1}(x, y), \quad (19)$$

$$\tilde{\psi}_{3,N}(x|y) = (1 + N^{-1/2} (Q_{1,N}(x, y) - (\gamma_3/6)(y^3 - 3y))) \phi_k(x), \quad (20)$$

$$\psi_{3,N}^c(x) = (1 + N^{-1/2} Q_{1,N}(x, 0)) \phi_k(x) = \tilde{\psi}_{3,N}(x|0). \quad (21)$$

Remark 4. From the proofs of the results, it follows that the error bounds, appearing in the conclusions of the theorems as δ_N , $o(\cdot)$ and $O(\cdot)$ -term, depend only on σ_1 and M_1 , whenever

$$E|Y_1|^s \leq M_1 < \infty \quad \text{and} \quad \sigma \geq \sigma_1 > 0. \quad (22)$$

So in particular, the bounds in Theorems 1–2 hold uniformly for all lattice random variables Y_1 with span 1, as long as (22) holds.

3. APPLICATIONS

The following examples illustrate the applicability of Edgeworth expansions for conditional probabilities. For the applications considered in this section we use the notation, $\bar{X}_N = N^{-1} \sum_{j=1}^N X_j$ and $a_j = X_j - \bar{X}_N$, where X_j is a row vector in \mathbb{R}^k . If X_j are univariate random variables, then in addition, we use the notation

$$\mu_x = E(X_1), \quad \sigma_x = \sqrt{\text{Var}(X_1)}, \quad \text{and} \quad s_N^2 = \frac{1}{N} \sum_{j=1}^N (X_j - \bar{X}_N)^2.$$

I. Sampling without Replacement from a Finite Population

Suppose $\{Y_1, \dots, Y_N\}$ are i.i.d. Bernoulli random variables with $P(Y_1 = 1) = n/N$. If $\sum_{j=1}^N Y_j = n$, then $\sum_{j=1}^N X_j Y_j$ represents the sum of n items sampled

without replacement from a finite population X_1, \dots, X_N , and $y_n = ((n - NP(Y_1 = 1))/(\sigma \sqrt{N})) = 0$. Following a direct approach, Babu and Singh (1985) derived a one-term Edgeworth expansion for the mean of a sample from a finite population. The present methods give us an s -term Edgeworth expansion for any integer $s \geq 1$. Suppose \bar{x}_n denotes the mean of the n sampled units and $\delta \leq n/N \leq 1 - \delta$ for some $0 < \delta < 1/2$. If condition (10) holds, and if $\sum_{j=1}^N \|X_j\|^s < MN$ for some $M > 0$ and an integer $s \geq 3$, then from Theorem 2 and Remark 4, we can get an $(s-2)$ -term Edgeworth expansion for $\mathcal{G}_n = (nN/(N-n))^{1/2} (\bar{x}_n - \bar{X}_N) V_N^{-1}$. The distribution of \mathcal{G}_n is the conditional distribution of U_N given $W_N = y_n = 0$. By (9) and (20), the one-term Edgeworth expansion is given by

$$\begin{aligned} \tilde{\psi}_{3,N}(x|0) = & \left(1 + \frac{1}{6}(2n-N)(nN(N-n))^{-1/2} \left(\sum_{j=1}^k A_{j,j,j}(x_j^3 - 3x_j) \right. \right. \\ & + 3 \sum_{1 \leq j \neq r \leq k} A_{j,j,r}(x_j^2 - 1)x_r \\ & \left. \left. + 6 \sum_{1 \leq j < m < r \leq k} A_{j,m,r}x_jx_mx_r \right) \right) \phi_k(x). \end{aligned}$$

In the univariate case (when $k=1$), the last two sums above vanish and we are left with

$$\tilde{\psi}_{3,N}(x|0) = 1 + \frac{2p-1}{6\sqrt{Np(1-p)}} \frac{(1/N) \sum_{j=1}^N a_j^3}{((1/N) \sum_{j=1}^N a_j^2)^{3/2}} (x^3 - 3x), \quad (23)$$

where $p = n/N$. For related results on the sub-sample method see Babu (1992). It is interesting to note that (23) agrees with the one-term empirical Edgeworth expansion in the i.i.d. situation, provided $p = n/N \approx \frac{1}{2}(1 - 1/\sqrt{5})$.

II. Bootstrap

Let $N=n$. If $\{X_1, \dots, X_N\}$ represents an i.i.d. sample of size N from a k -variate population F , then the distribution of the bootstrap mean can be identified with $N^{-1} \sum_{j=1}^N n_j X_j$, where (n_1, \dots, n_N) is a realization from the multinomial distribution $M(N; 1/N, \dots, 1/N)$. For an alternative approach to the bootstrap, consider i.i.d. random variables Y_1, \dots, Y_N with a common Poisson distribution with mean 1. Then the bootstrap distribution of the sample sum centered at its mean, is given by (1). Clearly in this case, the moment condition of Theorem 2 on Y_1 holds. By Remark 1, this leads to Theorem 2 of Babu and Singh (1984). In the special case of $s=3$, and $k=1$, this yields under strong non-lattice condition on F , that

$$\sqrt{N} \sup_z |P(\sqrt{N}(\bar{X}_N - \mu_x) \leq z\sigma_x) - P^*(\sqrt{N}(\bar{X}_N^* - \bar{X}_N) \leq z\sigma_N)| \rightarrow 0,$$

for almost all sample sequences $\{X_i\}$, where \bar{X}_N^* denotes the bootstrapped mean, and P^* denotes the probability induced by the bootstrap sampling scheme. Theorem 2 yields similar results for statistics which are smooth functions of multivariate means. See Corollary 2 of Babu and Singh (1984) for the details. The class of statistics for which these results are applicable include, sample means, sample variances, central and non-central t -statistics (with possibly non-normal populations), sample coefficient of variation, maximum likelihood estimators, least squares estimators, correlation coefficients, regression coefficients, and smooth transforms of these statistics. It may be noted that if Cramér's condition holds for F , then by Theorem 2,

$$N \sup_z |P(\sqrt{N}(\bar{X}_N - \mu_x) \leq z\sigma_x) - P^*(\sqrt{N}(\bar{X}_N^* - \bar{X}_N) \leq z\sigma_N)|$$

converges weakly to a random variable.

It is interesting to note that even when $\{X_1, \dots, X_N\}$ represents a realization of N random variables, not necessarily independent or identically distributed, Theorem 2 gives an Edgeworth expansion for the bootstrap distribution of a smooth function of the mean of X_j , as long as the conditions of Theorem 1 on a_j hold. The Bootstrap procedures for the sample mean of independent but not identically distributed random variables X_i , along with some examples to motivate such a study, were considered by Liu (1988). Suppose μ_i and σ_i denote the mean and the standard deviation of X_i . Theorem 2(ii) of Liu (1988) follows from Theorems 1 and 2, as Student's t can be expressed as a smooth function of the multivariate mean

$$\left(\frac{1}{N} \sum_{i=1}^N (X_i - \mu_i), \frac{1}{N} \sum_{i=1}^N ((X_i - \mu_i)^2 - \sigma_i^2) \right).$$

III. Bayesian Bootstrap

Let $N = n$. The random weighting scheme using multinomial distribution can be generalized to obtain, what is known as the Bayesian bootstrap. Rubin (1981) suggested using the spacings of a sample of size $(N-1)$ from the uniform distribution as the random weights, instead of n_i/N . This can be arrived at by starting with a standard exponential random variable Y_1 , and considering the posterior mean given the data $\{X_1, \dots, X_N\}$. In general if Y_1 has the standard gamma distribution with mean $r(r > 0)$, and if $\{X_1, \dots, X_N\}$ represents an i.i.d. sample from a k -variate population, then the conditional distribution of $\sum_{j=1}^N a_j(Y_j/rN)$ given $\sum_{j=1}^N Y_j = rN$ is the same as the distribution of $\sum_{j=1}^N a_j Z_j$, where (Z_1, \dots, Z_N) has the N -variate Dirichlet distribution $D(N; r, \dots, r)$. Note that in this case, $(Y_1(\sum_{j=1}^N Y_j)^{-1}, \dots, Y_N(\sum_{j=1}^N Y_j)^{-1})$ also has the same Dirichlet distribution $D(N; r, \dots, r)$.

When $r=4$, Tu and Zheng (1987) have shown the second order accuracy of the distribution of $N^{-1/2} \sum_{j=1}^N a_j V_N^{-1} Z_j / \sqrt{\text{Var}(Z_1)}$ in approximating the distribution of the mean of a random sample $\{X_j\}$ from a univariate population.

Note that if

$$(Z_1, \dots, Z_N) = \left(Y_1 \left(\sum_{j=1}^N Y_j \right)^{-1}, \dots, Y_N \left(\sum_{j=1}^N Y_j \right)^{-1} \right),$$

then the dispersion of $N^{-1/2} \sum_{j=1}^N a_j V_N^{-1} Z_j$ is $(N/(N-1)) \text{Var}(Z_1) I$, where I is the identity matrix. Now

$$\begin{aligned} \text{Var}(Z_1) &= \text{Var} \left(\frac{Y_1}{\sum_{j=1}^N Y_j} \right) \\ &= N^{-2} \text{Var} \left(\frac{Y_1}{N^{-1} \sum_{j=1}^N Y_j} \right) \\ &\approx (N\mu)^{-2} \text{Var} Y_1 = (N\mu)^{-2} \sigma^2. \end{aligned}$$

Consequently, we define the Bayesian bootstrap distribution of $\sqrt{N}(\bar{X}_N - \mu_x)/\sigma_x$ to be the distribution of

$$N^{-1/2} \left(\sum_{j=1}^N a_j V_N^{-1} Z_j \right) \Big/ (\sigma/N\mu) = N\mu U_N \Big/ \left(\sum_{j=1}^N Y_j \right).$$

In practice, it is easier to generate a sequence of i.i.d. random variables than a sequence subject to a restriction on the sum, which may be a reason to consider the distribution of $(\mu \sqrt{N} \sum_{j=1}^N a_j Y_j)(\sigma \sum_{j=1}^N Y_j)^{-1}$ instead of the conditional distribution, in generalizing the Bayesian bootstrap. If a gamma “prior” is used, then by Theorem 4 and equations (9) and (20), or by Theorem 5 and equations (9) and (21), it follows that the Bayesian bootstrap is second order correct in approximating the distribution of the sample mean of k -variate sequence $\{X_1, \dots, X_N\}$ only if $E(Y_1 - \mu)^3/\sigma^3 = 1$. This holds if and only if $r = \mu = 4$, in which case, the distribution of the vector of weights (Z_1, \dots, Z_N) is $D(N; 4, \dots, 4)$. This can be seen in the univariate case ($k=1$) from the following arguments. Note that in the univariate case, if κ_3 is defined by

$$\kappa_3 = \left(\frac{1}{N} \sum_{j=1}^N (X_j - \bar{X}_N)^3 \right) \left(\frac{1}{N} \sum_{j=1}^N (X_j - \bar{X}_N)^2 \right)^{-3/2},$$

then

$$P\left(\left(\mu \sqrt{N} \sum_{j=1}^N (X_j - \bar{X}_N) Y_j\right) / \left(\sigma \sum_{j=1}^N Y_j\right) \leq z\right) \\ = \int_{-\infty}^z \left(1 + \frac{\gamma_3 \kappa_3}{6 \sqrt{N}} (w^3 - 3w)\right) \phi(w) dw + o\left(\frac{1}{\sqrt{N}}\right). \quad (24)$$

On the other hand the classical theory of Edgeworth expansions yields

$$P(\sqrt{N} (\bar{X}_N - \mu_x) / \sigma_x \leq z) \\ = \int_{-\infty}^z \left(1 + \frac{k_3^o}{6 \sqrt{N}} (w^3 - 3w)\right) \phi(w) dw + o\left(\frac{1}{\sqrt{N}}\right), \quad (25)$$

under non-lattice condition on the distribution of X_1 , where k_3^o denotes the third cumulant of X_1 / σ_x . By the strong law of large numbers, the two expansions (24) and (25) above together yield that, the difference between the sampling distribution of $\sqrt{N} (\bar{X}_N - \mu_x) / \sigma_x$ and the corresponding Bayesian bootstrap distribution is $o(N^{-1/2})$ uniformly in z , if and only if $\gamma_3 = 1$.

Lo (1991) considered $(\mu \sqrt{N} \sum_{j=1}^N a_j Y_j) (\sigma \sum_{j=1}^N Y_j)^{-1}$ for univariate X_j , when Y_1 is not necessarily a gamma variable, and obtained first order asymptotic results. Suppose Y_1 is a positive random variable and $\{X_1, \dots, X_N\}$ represents an i.i.d. sample from a k -variate population. The distribution of $\mu \sqrt{N} (\sum_{j=1}^N a_j V_N^{-1} Y_j) / (\sigma \sum_{j=1}^N Y_j)$, may still be called the Bayesian bootstrap distribution of $N^{-1/2} \sum_{j=1}^N (X_1 - E(X_1)) \Sigma^{-1}$, where Σ^2 denotes the dispersion matrix of X_1 . Theorem 5 gives Edgeworth expansions for this generalized Bayesian bootstrap distribution. If $\gamma_3 = E((Y_1 - \mu)^3) \sigma^{-3} = 1$, then the one-term Edgeworth expansion $\psi_{3,N}^c$ in (21) leads to the second order correctness of the generalized Bayesian bootstrap approximation. So Theorem 5 allows us to choose Y_1 from a wide class of distributions, not just the gamma family of distributions. Hence we can choose “priors” from a variety of smooth distributions.

By considering the two-term Edgeworth expansions, it can be shown that the difference between the bootstrap distribution P_B^N and the Bayesian bootstrap distribution P_{BB}^N is $o(N^{-1})$, whenever γ_3 corresponding to the Bayesian “prior” is 1. That is, if $M_{h,4} < \infty$ and $E(Y_1^4) < \infty$, then for some $\delta_N = o(N^{-1})$,

$$\left| \int h(x) (P_B^N - P_{BB}^N)(dx) \right| = o(M_{h,4} N^{-1} + \omega(h, \delta_N)).$$

In summary, the results show that the Bayesian bootstrap is second order correct for any smooth positive “prior” as long as the third cumulant of the “prior” is equal to the third power of its standard deviation. As a consequence, among the standard gamma “priors”, the only one that leads to second order correctness is the one with mean 4.

IV. Weighted Bootstrap

For the Bayesian bootstrap, the resampling distribution is defined by assigning a set of random weights to the original sample points with weights continuously distributed. By Theorem 6, results similar to those mentioned in the previous section on Bayesian bootstrap will also hold for the distribution of $(\mu/\sigma) \sqrt{N} (\sum_{j=1}^N a_j V_N^{-1} Z_j)$, if the random variables Y_j are generated from a lattice distribution and the weights $Z_i = Y_i / \sum_{j=1}^N Y_j$ are used. As in the case of the Bayesian bootstrap, the one-term Edgeworth expansion matches with that of the standardized sample mean up to an error term of the order $o(N^{1/2})$, whenever $\gamma_3 = E((Y_1 - \mu)^3) \sigma^{-3} = 1$. So one achieves second order correctness of the weighted bootstrap whenever $\gamma_3 = 1$, if the weights are constructed using random variables from a lattice distribution. In particular, the second order accuracy is achieved if Y_1 has the negative binomial distribution with the parameters $r=4$ and $p \in (0, 1)$. The general case of first order approximations for weighted bootstrap is considered by Præstgaard and Wellner (1993).

4. PROOFS OF THE THEOREMS

In this section, we briefly sketch the proofs of the Theorems. The technical details required in the proofs are separated as lemmas and presented in the Appendix.

We start with some notation. For non-negative integral vectors $\alpha = (\alpha, \dots, \alpha_k)$ and $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, we write

$$|\alpha| = \alpha_1 + \dots + \alpha_k, \quad x^\alpha = x_1^{\alpha_1} \dots x_k^{\alpha_k}, \quad \text{and} \quad D^\alpha = D_1^{\alpha_1} \dots D_k^{\alpha_k},$$

where D_j denotes the partial derivative with respect to the j th coordinate. For any integrable function τ , its Fourier transform is denoted by $\hat{\tau}$. Let $\hat{f}_N(t, v)$ be the characteristic function of $(\sum_{j=1}^N a_j Y_j, \sum_{j=1}^N Y_j)$, given by

$$\hat{f}_N(t, v) = \prod_{j=1}^N \hat{f}(ta'_j + v),$$

and define

$$\hat{f}_{N, C}(t, v) = \prod_{j \in C} \hat{f}(ta'_j + v),$$

where C is a subset of $\{1, \dots, N\}$. Finally for any integer $\xi \geq 0$, let

$$f_{N, \xi}(t, v) = \max\{|\hat{f}_{N, C}(t, v)|\},$$

where the maximum is taken over all subsets of $\{1, \dots, N\}$ of size $N - \xi$.

Proof of Theorem 1. Note that the Fourier transform of $\psi_{s, N}$ is given by

$$\hat{\psi}_{s, N}(t, v) = \left(1 + \sum_{r=1}^{s-2} N^{-j/2} P_{j, N}(it, iv)\right) \exp(-\frac{1}{2}(\|t\|^2 + v^2)), \quad (26)$$

where $\hat{\Delta}_r = N^{-1} \sum_{j=1}^N (ia_j V_N^{-1} t' + iv)^r$, and

$$P_{j, N}(it, iv) = \sum^* \prod_{m=1}^j \frac{1}{r_m!} \left(\frac{\gamma_{m+2} \hat{\Delta}_{m+2}}{(m+2)!} \right)^{r_m} \exp\left(-\frac{1}{2}(\|t\|^2 + v^2)\right).$$

Here \sum^* denotes the sum over all non-negative integers r_m satisfying $\sum_{1 \leq m \leq j} m r_m = j$. We use Lemma 1, which is similar to Lemma 11.6 of Bhattacharya and Ranga Rao (1986), that gives bounds for a signed measure in terms of derivatives of its Fourier transform. We also use Lemma 4 that gives bounds on the error, when a distribution is convoluted with a smooth distribution. Fix a small $\eta > 0$ and let ε in Lemma 4 be $= \eta N^{-(s-2)/2}$. Now the theorem can be established using Lemma 1 and the estimates of the Fourier transforms in Lemmas 2 and 3, along the lines of the proof of Theorem 20.1 of Bhattacharya and Ranga Rao (1986). See also the proof of Theorem 3.

Proof of Theorem 3. Essential part of any proof of the validity of Edgeworth expansions consists of an inversion formula and the estimation of the derivatives of the Fourier transform in several regions. By Lemma 11.6 of Bhattacharya and Ranga Rao (1986), the integral in (16) is dominated by a constant multiple of

$$\sup_{|\alpha| \leq 2 + k + s} \int_{\mathbb{R}^{k+1}} |D^\alpha(\hat{f}_N(t, v) - \hat{\psi}_{s, N}(t, v))| dt dv. \quad (27)$$

We estimate (27) by dividing the range of the integration into several, possibly overlapping, regions:

- (i) $\|t\| \leq N^{-1/2} \log N$, $|v| \leq MN^{-1/2} \log N$, for some $M > 0$
- (ii) $\|t\| \leq N^{-1/2} \log N$, $H \geq |v| > MN^{-1/2} \log N$, for some $M > 0$, $H > 0$
- (iii) $N^{-1/2} \log N \leq \|t\| \leq \delta$, for some $\delta > 0$
- (iv) $\delta \leq \|t\| \leq L$, for some $\delta > 0$, $L > 0$
- (v) $\|(t, v)\| > T$, for some $T > 0$.

We expand $D^\alpha(\hat{f}_N(t, v))$ in the region (i), and estimate the integrand in (27). Lemma 2 is used for the regions (iii) and (iv), and Lemma 3 is used for the region (ii) to estimate the integrand in (27). Finally Lemma 6 is used in region (v), as in the proof of Theorem 19.5 of Bhattacharya and Ranga Rao (1986), to arrive at (16). To prove (17), we use inequality (32) of Lemma 1 instead of Lemma 11.6 of Bhattacharya and Ranga Rao (1986). In addition, Lemmas 2, 3 and 6 are used as above, together with the arguments similar to the proof of Theorem 19.3 of Bhattacharya and Ranga Rao (1986) to get (17). The details are omitted.

Proofs of Theorems 2 and 4. Under the conditions of Theorem 1 or Theorem 3, the functions $\int h(x) \psi_{s,N}(x, y) dx$, and $\zeta_{s,N}$ defined in (12), are bounded. Furthermore, for any $G \geq 0$, $\liminf_{N \rightarrow \infty} \inf_{|y| \leq G} |\zeta_{s,N}(y)| > 0$. Theorems 2 follows from (11) and Theorem 1. Theorem 4 follows from (14) and Theorem 3.

Proofs of Theorems 5 and 6. Let $Z_i = Y_i(\sum_{j=1}^N Y_j)^{-1}$ whenever $\sum_{j=1}^N Y_j > 0$, and let H_N denote the indicator function of $(|\sum_{j=1}^N (Y_j - \mu)| > N\mu/2)$. Note that $\mu > 0$, and $E(Z_j H_N) = E(Z_1 H_N)$ for all j . By Theorem 2 of Michel (1976), there exists a constant k_1 such that

$$\begin{aligned} E(H_N) &= O(N^{2-3s/2}) + NP(|Y_1 - \mu| > k_1 \mu N) \\ &= O(N^{(2-3s)/2} + N^{1-s} E(|Y_1|^s I(|Y_1| > 2^{-1/2} k_1 \mu N))) \\ &= o(N^{1-s}). \end{aligned}$$

By (35) there exists a $\lambda > 0$ such that $V_N^2 - \lambda I$ is non-negative definite and hence $\|a_j V_N^{-1}\| \leq \|a_j\| \lambda^{-1/2}$ for all large N . Furthermore, as $\sum_{j=1}^N Z_j = 1$, it follows that $E(Z_1 H_N) = N^{-1} E(H_N)$. So by (28) we have for some constant $M_2 > 0$,

$$\begin{aligned} M_2 E \left(h \left(\sigma N U_N \left(\sum_{j=1}^N Y_j \right)^{-1} \right) H_N \right) \\ \leq E(H_N) + N^{s/2} E \left(\sum_{j=1}^N \|a_j V_N^{-1}\|^s Z_j H_N \right) \\ \leq E(H_N) + N^{s/2} \lambda^{-s/2} \left(\sum_{j=1}^N \|a_j\|^s \right) E(Z_1 H_N) \\ \leq E(H_N) + N^{s/2} \lambda^{-s/2} \left(\sum_{j=1}^N \|a_j\|^s \right) N^{-1} E(H_N) \\ = o(N^{-(s-2)/2}). \end{aligned} \tag{29}$$

To complete the proof of Theorem 5, we have by (16), (18), and (29) that

$$\begin{aligned}
 & \int_{\mathbb{R}^k} (1 + \|x\|)^s |f_N^o(x) - \psi_{s,N}^c(x)| \, dx \\
 & \leq \int_{\mathbb{R}^k} (1 + \|x\|)^s \int_{|y| \leq \mu \sqrt{N}/2\sigma} |(f_N(x(1 + (\sigma/\mu \sqrt{N}) y), y) \\
 & \quad - \psi_{s,N}(x(1 + (\sigma/\mu \sqrt{N}) y), y))(1 + (\sigma/\mu \sqrt{N}) y)^k| \, dy \, dx \\
 & \quad + E \left(\left(1 + \left\| N\mu U_N \left(\sum_{j=1}^N Y_j \right)^{-1} \right\|^s \right) H_N \right) + o(N^{-(s-2)/2}) \\
 & \leq \int_{\mathbb{R}^k} \int_{|y| \leq \mu \sqrt{N}/2\sigma} (1 + \|x\|/|1 + (\sigma/\mu \sqrt{N}) y|)^s \\
 & \quad \times |f_N(x, y) - \psi_{s,N}(x, y)| \, dy \, dx + o(N^{-(s-2)/2}) \\
 & = o(N^{-(s-2)/2}).
 \end{aligned}$$

This completes the proof of Theorem 5.

Now we turn to the remaining part of the proof of Theorem 6. Note that the factor $P(\sum_{j=1}^N Y_j > 0) = 1 - E(Y_1 = 0)^N$, which appears in F_N^o does not contribute to the main terms of the expansion. From the definition of $\psi_{s,N}^d(x)$, it is not difficult to show that

$$\begin{aligned}
 & \left| \int h(x) \left(\frac{1}{\sqrt{N} \sigma} \sum_{n=1}^{\infty} \psi_{s,N}(x(1 + (\sigma/\mu \sqrt{N}) y_n), y_n) \right. \right. \\
 & \quad \left. \left. \times (1 + (\sigma/\mu \sqrt{N}) y_n)^k - \psi_{s,N}^d(x) \right) dx \right| \\
 & = o(N^{-(s-2)/2}),
 \end{aligned}$$

and

$$\begin{aligned}
 & \left| \int h(x) \frac{1}{\sqrt{N} \sigma} \sum_{n: 2|\sigma y_n| \geq \sqrt{N} \mu} \psi_{s,N}(x(1 + (\sigma/\mu \sqrt{N}) y_n), y_n) \right. \\
 & \quad \left. \times (1 + (\sigma/\mu \sqrt{N}) y_n)^k \, dx \right| \\
 & = o(N^{-(s-2)/2}).
 \end{aligned}$$

Finally, by applying Theorem 1, we obtain

$$\begin{aligned} & \left| \frac{1}{\sqrt{N} \sigma} \sum_{n: 2|\sigma y_n| < \sqrt{N} \mu} \int h(x/(1 - (\sigma y_n/(\mu \sqrt{N}))) \right. \\ & \quad \left. \times (\sigma \sqrt{N} F_N(dx, n) - \psi_{s, N}(x, y_n) dx) \right| \\ & = o(M_{h, s} N^{-(s-3)/2}) + O(\sqrt{N} \omega(h, \delta_N)). \end{aligned}$$

The above three inequalities, combined with (29), yield Theorem 6.

APPENDIX

For the first lemma, we consider a real valued measurable function g on $\mathbb{R}^k \times \mathbb{Z}$, endowed with the product of Lebesgue measure on \mathbb{R}^k and the counting measure on \mathbb{Z} . Let

$$\hat{g}(t, v) = \sum_m e^{ivm} \int_{\mathbb{R}^k} e^{itx'} g(x, m) dx$$

denote the Fourier transform of the function g on the product space $\mathbb{R}^k \times \mathbb{Z}$. We analyze its marginal transform in the proof of Lemma 1.

LEMMA 1. *Let g be a real valued function on $\mathbb{R}^k \times \mathbb{Z}$ satisfying*

$$\sum_m \int_{\mathbb{R}^k} (1 + \|x\|)^{s+k+1} |g(x, m)| dx < \infty \quad (30)$$

for some non-negative integer s . Then there exists a constant $c(k)$ depending only on k such that, for all integers m ,

$$\begin{aligned} & \int_{\mathbb{R}^k} (1 + \|x\|^s) |g(x, m)| dx \\ & \leq c(k) \max_{|\alpha| \leq 1+k+s} \int_{-\pi}^{\pi} \left(\int_{\mathbb{R}^k} |D^\alpha \hat{g}(t, v)| dt \right) dv. \end{aligned} \quad (31)$$

Suppose g_1 is a real valued function on \mathbb{R}^{k+1} satisfying

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^k} (1 + \|x\|)^{s+k+1} |g_1(x, y)| dx \right) dy < \infty,$$

for some non-negative integer s and suppose

$$\hat{g}_1(t, v) = \int_{\mathbb{R}^{k+1}} e^{itx' + ivy} g_1(x, y) dx dy$$

denotes the Fourier transform of g_1 on \mathbb{R}^{k+1} . Then there exists a constant $c'(k)$ depending only on k such that, for all y ,

$$\int_{\mathbb{R}^k} (1 + \|x\|^s) |g_1(x, y)| dx \leq c'(k) \max_{|\alpha| \leq 1+k+s} \int_{-\infty}^{\infty} \left(\int_{\mathbb{R}^k} |D^\alpha \hat{g}_1(t, v)| dt \right) dv. \quad (32)$$

Proof. We assume that

$$\int_{-\pi}^{\pi} \left(\int_{\mathbb{R}^k} |D^\alpha \hat{g}(t, v)| dt \right) dv < \infty,$$

for vectors α of non-negative integers satisfying $|\alpha| \leq 1+k+s$. Otherwise the result is trivial. For fixed m , let $g_m(x) = g(x, m)$ and $h_{m, \alpha}(x) = x^\alpha g_m(x)$. Note that the Fourier transform \hat{g}_m of g_m is given by

$$\hat{g}_m(t) = \int_{\mathbb{R}^k} e^{itx} g(x, m) dx,$$

and that the Fourier transform $\hat{h}_{m, \alpha}$ of $h_{m, \alpha}$ is given by $(-1)^{|\alpha|} D^\alpha \hat{g}_m$. For each t , the Fourier transform of the function $u(m) = \hat{h}_{m, \alpha}(t)$ on \mathbb{Z} , is given by $\hat{u}(v) = (-1)^{|\alpha|} D^\alpha \hat{g}(t, v)$. So by (30), Fubini's theorem, and the Fourier inversion

$$u(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ivm} \hat{u}(v) dv,$$

we have

$$D^\alpha \hat{g}_m(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ivm} D^\alpha \hat{g}(t, v) dv.$$

The result now follows from an application of Lemma 11.6 of Bhattacharya and Ranga Rao (1986) to g_m . This completes the proof of (31). A similar proof yields (32).

The basic ideas of the proofs of the next two lemmas are inspired by Erdős and Rényi (1959).

LEMMA 2. Let \tilde{Y} denote the difference of two independent copies of Y . Let $m_2 > m_1 > 0$ be such that $P(m_1 < |\tilde{Y}| < m_2) = p > 0$, and let $\xi \leq N$ be a non-negative integer. Then we have:

(a) If for some $\delta > 0$, and $\Delta > 0$, $|d_N(t)| \leq 1 - \Delta \|t\|^2$ for $\|t\| \leq \delta$, then for all v , and $\|t\| \leq \delta/m_2$,

$$f_{N,\xi}(t, v)^2 \leq \exp(\xi - N \Delta p m_1^2 \|t\|^2).$$

(b) For any $K, L > 0$, and for all v ,

$$\sup_{K \leq \|t\| \leq L} f_{N,\xi}(t, v)^2 \leq \exp\{\xi - Np(1 - \sup_{Km_1 \leq \|t\| \leq Lm_2} |d_N(t)|)\}.$$

Proof. Note that the characteristic function of \tilde{Y} is given by $|\hat{f}|^2$. Consequently,

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N |\hat{f}(ta'_j + v)|^2 &= E \left[\left(\frac{1}{N} \sum_{j=1}^N e^{ita'_j \tilde{Y}} \right) e^{iv \tilde{Y}} \right] \\ &\leq E |d_N(t \tilde{Y})| \\ &\leq 1 - p + E |d_N(t \tilde{Y}) I(m_1 \leq |\tilde{Y}| \leq m_2)| \end{aligned} \quad (33)$$

$$\leq 1 - p + p \sup_{m_1 \|t\| \leq \|u\| \leq m_2 \|t\|} |d_N(u)|. \quad (34)$$

Since $x \leq e^{x-1}$ for any non-negative x , we have for any integer $\xi \geq 0$ and for any subset C of $\{1, \dots, N\}$ of $N - \xi$ integers, that

$$\begin{aligned} \prod_{j \in C} |\hat{f}(ta'_j + v)|^2 &\leq \exp \left(\xi - N + \sum_{j \in C} |\hat{f}(ta'_j + v)|^2 \right) \\ &\leq \exp \left(\xi - N + \sum_{j=1}^N |\hat{f}(ta'_j + v)|^2 \right). \end{aligned}$$

The results (a) and (b) now follow from (33) and (34) respectively.

LEMMA 3. Let $E(|Y_1|^3) \leq M_1$, $\sigma \geq \sigma_1$, $\sum_{j=1}^N a_j = 0$, and $\sum_{j=1}^N \|a_j\|^3 \leq NM$, for some $M > 1$, $M_1 < \infty$, $\sigma_1 > 0$. Let $H = \pi$ if Y_1 has a lattice distribution with span 1, and let H be any positive number if Y_1 has a non-lattice distribution. Then for $\|t\| \leq N^{-1/2} \log N$ and $H \geq |v| > MN^{-1/2} \log N$, we have for any integer $0 \leq \xi < N$.

$$f_{N,\xi}(t, v) \leq k_2 \exp(\xi - k_3(\log N^2)),$$

where $k_2 > 0$ and $k_3 > 0$ are constants, and they may depend on M , H , ξ , σ_1 and on M_1 .

Proof. Since for all real ω ,

$$|e^{-i\omega\mu}\hat{f}(\omega) - 1 + \frac{1}{2}\sigma^2\omega^2| \leq \frac{1}{6}|\omega|^3 M_1,$$

there exist $0 < \delta < \pi/4$, and $\Delta > 0$ depending only on σ_1 and M_1 such that $\frac{1}{2} \leq |\hat{f}(\omega)| \leq 1 - \Delta\omega^2$, whenever $|\omega| < 4\delta$. Suppose $\|t\| \leq N^{-1/2} \log N$, and $MN^{-1/2} \log N \leq |v| \leq 2\delta$. Then for all large N , $|ta'_j + v| < 4\delta$ and hence we have $\frac{1}{2} \leq |\hat{f}(ta'_j + v)|$ and

$$|\hat{f}_N(t, v)| \leq \exp\left(-\Delta \sum_{j=1}^N (ta'_j + v)^2\right).$$

As $\sum_{j=1}^N a_j = 0$, it follows that

$$|\hat{f}_N(t, v)| \leq \exp(-N \Delta v^2) \leq \exp(-M \Delta (\log N)^2),$$

and hence $f_{N, \xi}(t, v) \leq 2^\xi \exp(-M \Delta (\log N)^2)$. Now suppose $2\delta \leq |v|$ and $\|t\| \leq N^{-1/2} \log N$. Note that for any $H > 4\delta$ (if Y_1 has lattice distribution with span 1, then $H = \pi$), we have

$$\sup_{\delta \leq |\omega| \leq H + \delta} |\hat{f}(\omega)| = \rho(\delta, H) < 1.$$

If the characteristic function of Y_1 satisfies (13), then $\sup_{\delta \leq |\omega|} |\hat{f}(\omega)| = \rho(\delta, \infty) < 1$. Since $\sup_{1 \leq j \leq N} \|a_j\| \leq MN^{1/3}$, if $\|t\| \leq N^{-1/2} \log N$, then $\|ta'_j\| \leq MN^{-1/6} \log N$. Hence for large N , $\delta \leq \|ta'_j + v\| \leq H + \delta$, whenever $\|t\| \leq N^{-1/2} \log N$, and $2\delta \leq |v| \leq H$. Consequently, in this case, it follows that $f_{N, \xi}(t, v) \leq \rho(\delta, H)^{N-\xi}$. This completes the proof.

By Theorem 10.1 of Bhattacharya and Ranga Rao (1986), for any positive integer s , there exists a probability measure J on \mathbb{R}^k and $k_4, k_5 > 0$, such that $k_4 = \int \|x\|^{s+k+2} dJ(x) < \infty$, $J(\{x: \|x\| < 1\}) \geq \frac{3}{4}$ and $\hat{J}(t) = 0$ if $\|t\| \geq k_5$ ($t \in \mathbb{R}^k$). For any $\varepsilon > 0$, let J_ε denote the probability measure given by $J_\varepsilon(E) = J(\varepsilon^{-1}E)$ for all borel subsets E of \mathbb{R}^k . The next lemma follows easily from Lemma 24.1 of Bhattacharya and Ranga Rao (1986), which is a strengthened version of Lemma 11.1 of Bhattacharya and Ranga Rao (1986). The inequality is similar to the inequality (4.1) of Babu and Singh (1984), and it is stated here for ready reference.

LEMMA 4. *Let P be a finite measure and Q a signed measure on \mathbb{R}^k . Let h be a real valued measurable function on \mathbb{R}^k satisfying $M_{h,s} < \infty$. Then*

there exists a constant k_6 depending only on s and k such that for any $0 < \varepsilon < 1$,

$$\left| \int h d(P - Q) \right| \leq k_6 \left(\int (1 + \|x\|^s) d|J_\varepsilon * (P - Q)|(x) \right. \\ \left. + \beta k_4 \varepsilon^2 + \beta 3^{-\varepsilon^{-1/4}} + \sup_{\|x\| \leq 2\varepsilon^{1/4}} \int \omega(h, 2\varepsilon, x - z) d|Q|(z) \right),$$

where $\beta = M_{h,s} \int (1 + \|x\|^s) d(P + |Q|)(x)$.

Further for any $0 < \|x\| < 1$ and $0 < \delta < 1$, we have for some constant k_7 ,

$$\int \omega(h, \delta, x - z) \phi_k(z) dz \\ \leq 3 \int \omega(h, \delta, z) \phi_k(z) dz + k_7 M_{h,s} \|x\|^{1-k-s} \exp(-\tfrac{1}{8} \|x\|^{-2}).$$

The next two lemmas are required for the continuous case treated in Theorems 3, 4 and 5. Lemma 5 describes a sieve method needed to estimate the characteristic function in Lemma 6.

LEMMA 5. Let a_1, \dots, a_N be vectors in \mathbb{R}^k satisfying

- (a) $\sum_{j=1}^N a_j = 0$
- (b) $\sum_{j=1}^N \|a_j\|^s \leq NM$, for some $M > 0$ and $s > 2$
- (c) for any $0 < K < L < \infty$, (15) holds for some $\gamma = \gamma(K, L) < 1$.

Then for some $\theta > 0$ and $N_0 \geq 1$, we can select $m \geq \theta N$ groups of distinct vectors $\{b_{ij}, i = 1, \dots, k+1, j = 1, \dots, m\}$ from a_1, \dots, a_N , such that $|\det A_j| \geq \theta$ and the eigenvalues of $A_j A_j'$ are all bounded below by θ^2 for all $N \geq N_0$ where

$$A_j = \begin{pmatrix} b'_{1j} & \cdots & b'_{(k+1)j} \\ 1 & \cdots & 1 \end{pmatrix}.$$

Proof. By assumptions (a) and (c) there exists $N_1 \geq 1$ such that for any row vector l of unit length,

$$\frac{1}{2} l V_N^2 l' = \frac{1}{2N} \sum_{j=1}^N l a'_j a_j l' \\ \geq \left| \frac{1}{N} \sum_{j=1}^N (e^{i l a'_j} - 1 - i l a'_j) \right| \\ = |d_N(l) - 1| \geq 1 - |d_N(l)| \geq 1 - \gamma(2^{-1}, 2) > 0,$$

for all $N \geq N_1$. This implies, for some $\lambda > 0$ and for all $N \geq N_1$, that

$$\frac{1}{N} \sum_{j=1}^N a'_j a_j - \lambda I \quad \text{is non-negative definite.} \quad (35)$$

Let l be a row vector of unit length in \mathbb{R}^k , $0 < K < L < \infty$ and let

$$H(l) = H(l, K, L) = \{1 \leq i \leq N : K \leq la'_i, \|a\| \leq L\}.$$

Let $v_i = la'_i$. By (a), (b) and (35), we have for all $N \geq N_1$, that

$$\begin{aligned} 2L \# \{i \in H(l)\} &\geq 2 \sum_{i \in H(l)} v_i \\ &= 2 \left(\sum_{v_i > 0} v_i - \sum_{0 < v_i < K} v_i - \sum_{v_i \geq K, \|a_i\| > L} v_i \right) \\ &\geq \sum_{i=1}^N |v_i| - 2KN - 2 \sum_{\|a_i\| > L} \|a_i\| \\ &\geq \sum_{|v_i| \geq L} |v_i| - 2KN - 2 \sum_{\|a_i\| > L} \|a_i\| \\ &\geq \frac{1}{L} \sum_{|v_i| \leq L} v_i^2 - 2KN - 2 \sum_{\|a_i\| > L} \|a_i\| \\ &\geq \frac{1}{L} \sum_{i=1}^N v_i^2 - 2KN - \frac{1}{L} \sum_{|v_i| > L} v_i^2 - 2 \sum_{\|a_i\| > L} \|a_i\| \\ &\geq \frac{\lambda N}{L} - 2KN - L^{1-s} \sum_{i=1}^N v_i^s - 2L^{1-s} \sum_{i=1}^N \|a_i\|^s \\ &\geq N \left(\frac{\lambda}{L} - 2K - 3ML^{1-s} \right). \end{aligned}$$

Hence, $\# \{i \in H(l)\} \geq N\theta(K, L)$, where

$$\theta(K, L) = (2L^2)^{-1} (\lambda - 2KL - 4ML^{2-s}) > 0,$$

provided $L > (8M\lambda^{-1})^{1/(s-2)}$ and $K \leq (\lambda/4L)$. Note that $\theta(K, L)$ is independent of l . Hence there exist $\theta > 0$, $0 < K < L < \infty$ such that for all vectors l of unit length,

$$\# \{i \in H(l, K, L)\} \geq \theta N.$$

Next let $0 < K' < L' < \infty$. For $b \in \mathbb{R}^k$, $K' \leq \|b\| \leq L'$ and $v > 0$, let

$$M(b, v) = \{1 \leq j \leq N: |ba'_j - 1| > v\} \quad \text{and} \quad K(v, v) = \#\{j \in M(b, v)\}.$$

By (c) there exist $N_0 = N_0(K', L') \geq N_1$ and $\gamma(K', L') < 1$ such that for all $N \geq N_0$,

$$\begin{aligned} N\gamma(K', L') &\geq \left| \sum_{j=1}^N e^{i(ba'_j - 1)} \right| \\ &\geq \left| \sum_{j \notin M(b, v)} (1 + (e^{i(ba'_j - 1)} - 1)) \right| - \sum_{j \in M(b, v)} |e^{i(ba'_j - 1)}| \\ &\geq (N - K(b, v))(1 - v) - K(b, v). \end{aligned}$$

This implies that for all $N \geq N_0$,

$$K(b, v) \geq N(1 - v - \gamma(K', L'))/(2 - v) > vN,$$

provided $v = v(K', L')$ is chosen small enough. Note that v is independent of b .

We shall now describe the selection of b_{ij} . Choose b_1 to be one of the a_j with $K \leq \|a_j\| \leq L$. Let l_1 be a unit length vector orthogonal to b_1 . Then choose b_2 one of the a_i with $i \in H(l_1)$. Suppose b_1, \dots, b_j are selected for $j \leq k-1$, then choose a unit length vector l_j orthogonal to (b_1, \dots, b_j) . Now select b_{j+1} to be one of the a_i with $i \in H(l_j)$. Having chosen b_1, \dots, b_k , define $d_1 = b_1, \dots, d_j = b_j P_j$, where P_j denotes the projection to the orthogonal complement of the space generated by $\{b_1, \dots, b_{j-1}\}$. Note that $K \leq l_j b'_{j+1}$ implies $K \leq l_j P'_{j+1} b'_{j+1} = l_j d'_{j+1}$. This implies that $K \leq \|d_j\| \leq L$ for all j . Now $B = (b'_1, \dots, b'_k) = A\Gamma$, where Γ is an upper triangular matrix with all the diagonal elements equal to 1 and $A = (d'_1, \dots, d'_k)$. Let $e_0 \in \mathbb{R}^k$ be the row vector with all the entries equal to 1. Since

$$\det \Gamma = 1 \quad \text{and} \quad A' A = \text{diag}(\|d_1\|^2, \dots, \|d_k\|^2), \quad (36)$$

we have $|\det B| = |\det A| \geq K^k$. If ρ_{\max} and ρ_{\min} respectively denote the maximum and minimum eigenvalues of BB' , then

$$\rho_{\max} \leq \text{tr}(BB') \leq (1 + L^2)(k + 1)$$

and

$$\rho_{\min} \geq \det(BB')/\rho_{\max}^k \geq [(1 + L^2)(k + 1)]^{-k} K^{2k}.$$

Consequently $k/\rho_{\max} \leq \|e_0 B^{-1}\|^2 \leq k/\rho_{\min}$. So $K' \leq \|e_0 B^{-1}\| \leq L'$, for some $0 < K' < L' < \infty$. Clearly K', L' , and hence N_0 depend only on K, L chosen

earlier. Now choose b_{k+1} to be one of the a_i with $|1 - e_0 B^{-1} a_i| > \theta_1 = \min\{\theta(K, L), v(K', L')\}$. Observe that

$$A = \begin{pmatrix} b'_1 & \cdots & b'_{k+1} \\ 1 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} d'_1 & \cdots & d'_k & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \Gamma & A^{-1} b'_{k+1} \\ e_0 & 1 \end{pmatrix}.$$

By (36) and

$$\left| \det \begin{pmatrix} \Gamma & A^{-1} b'_{k+1} \\ e_0 & 1 \end{pmatrix} \right| = |1 - e_0 \Gamma^{-1} A^{-1} b'_{k+1}| \geq \theta_1,$$

we have $|\det A| \geq K^k \theta_1$. If λ_{\min} and λ_{\max} are the minimum and maximum eigenvalues of AA' then

$$\lambda_{\max} \leq \text{tr}(AA') \leq (1 + L^2)(k + 1)$$

and

$$\lambda_{\min} \geq \det(AA') / \lambda_{\max}^k \geq \theta_1^2 [(1 + L^2)(k + 1)]^{-2} K^{2k} = \theta_2 > 0.$$

So the eigenvalues of AA' are bounded below by θ_2 .

Since for each l and b , both $H(l)$ and $M(b, \theta_1)$ contain at least $N\theta_1$ many a_i , it is possible to repeat the above procedure m times, with $m \geq \theta N$ for some $\theta > 0$. This completes the proof.

LEMMA 6. Suppose $\sum_{j=1}^N \|a_j\|^s \leq MN$ for some $s > 2$, $M > 0$ and for any $0 < K < L < \infty$, there exists a $\gamma = \gamma(K, L) < 1$, satisfying (15). If the characteristic function \hat{f} of Y_1 satisfies (13) for some positive integer r , then for any $T > 0$ and for any integer $\xi \geq 0$, there exist $\eta = \eta(T, \xi) < 1$ and N_0 such that

$$\int_{\|(t, v)\| > T} |\hat{f}_{N, C}(t, v)| dt dv \leq \eta^N, \quad (37)$$

for all $N \geq N_0$ and for all subsets C of $\{1, \dots, N\}$ of size $N - \xi$.

Proof. Without loss of generality, by subtracting $N^{-1} \sum_{j=1}^N a_j$ from each a_i if necessary, we assume that $\sum_{j=1}^N a_j = 0$. Note that $|\hat{f}_{N, C}(t, v)|$ is not affected by this. By Lemma 5, it is possible to choose a $\theta > 0$, $N_0 \geq \xi/\theta$, and $2m$ groups of distinct vectors $\{b_{ij}, i = 1, \dots, k + 1, j = 1, \dots, 2m\}$ from a_1, \dots, a_N , such that $m \geq \theta N$, $|\det A_j| \geq \theta$ and the eigenvalues of $A_j A'_j$ are all bounded below by θ^2 for all $N \geq N_0$, where

$$A_j = \begin{pmatrix} b'_{1j} & \cdots & b'_{(k+1)j} \\ 1 & \cdots & 1 \end{pmatrix}.$$

Note that $2m - \xi \geq m$, for $N \geq N_0 \geq \xi/\theta$. By dropping the groups that contain a_i for $i \notin C$, we are left with at least m groups of $(k+1)$ distinct vectors $\{b_{ij}, i = 1, \dots, k+1, j = 1, \dots, m\}$ from $\{a_i; i \in C\}$. This leads to

$$\begin{aligned} & \int_{\|(t, v)\| > T} |\hat{f}_{N, C}(t, v)| \, dt \, dv \\ & \leq \int_{\|(t, v)\| > T} \prod_{j=1}^m \prod_{i=1}^{k+1} |\hat{f}(tb'_{ij} + v)| \, dt \, dv \\ & \leq \frac{1}{m} \sum_{j=1}^m \int_{\|(t, v)\| > T} \prod_{i=1}^{k+1} |\hat{f}(tb'_{ij} + v)|^m \, dt \, dv. \end{aligned} \quad (38)$$

For the j th summand we make the transformation $w = (w_1, \dots, w_{k+1}) = (t, v) A_j$. Note that $\|(t, v)\| > T$ implies $\|w\| > \theta T$. It follows that the j th summand of the last term of inequality (38) is

$$\begin{aligned} & \leq \frac{1}{\theta} \int_{\|w\| > \theta T} \prod_{i=1}^{k+1} |\hat{f}(w_i)|^m \, dw \\ & \leq \frac{1}{\theta} \sum_{q=1}^{k+1} \int_{|w_q| > \delta T} \prod_{i=1}^{k+1} |\hat{f}(w_i)|^m \, dw \\ & \leq \frac{1}{\theta} \sum_{q=1}^{k+1} \int_{|w_q| > \delta T} |\hat{f}(w_q)|^{m-r} \prod_{i=1}^{k+1} |\hat{f}(w_i)|^r \, dw \\ & \leq \frac{k+1}{\theta} \rho(\delta T)^{m-r} \left(\int |\hat{f}(\omega)|^r \, d\omega \right)^{k+1}, \end{aligned}$$

where $\delta = \theta/(k+1)$ and for any $v > 0$.

$$\rho(v) = \sup_{|\omega| > v} |\hat{f}(\omega)|.$$

By (13), the distribution corresponding to the characteristic function \hat{f}^r has a continuous density. So by Riemann-Lebesgue lemma $|\hat{f}(w)| \rightarrow 0$ as $|w| \rightarrow \infty$. Consequently $\rho(v) < 1$ for any $v > 0$. This leads to inequality (37) completing the proof.

ACKNOWLEDGMENT

We thank the Editor and the referees for their helpful comments and suggestions.

REFERENCES

- [1] Babu, G. J. (1992). Subsample and half-sample methods. *Ann. Inst. Statist. Math.* **44** 703–720.
- [2] Babu, G. J., and Singh, K. (1983). Inference on means using the bootstrap. *Statist.* **11** 999–1003.
- [3] Babu, G. J., and Singh, K. (1984). On one term Edgeworth correction by Efron's Bootstrap. *Sankhyā Ser. A* **46** 219–232.
- [4] Babu, G. J., and Singh, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17** 261–278.
- [5] Babu, G. J., and Singh, K. (1989). On Edgeworth expansions in the mixture cases. *Ann. Statist.* **17** 443–447.
- [6] Bhattacharya, R. N., and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.
- [7] Bhattacharya, R. N., and Ranga Rao, R. (1986). *Normal Approximation and Asymptotic Expansion*, 2nd ed. Wiley, New York.
- [8] Erdős, P., and Rényi, A. (1959). On the central limit theorem for samples from finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **4** 49–61.
- [9] Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **6** 1696–1708.
- [10] Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā Ser. A* **53** 320–333.
- [11] Michel, R. (1976). Nonuniform central limit bounds with applications to probabilities of deviations. *Ann. Probab.* **4** 102–106.
- [12] Petrov, V. V. (1975). *Sums of Independent Random Variables*, English Translation. Springer-Verlag, New York.
- [13] Præstgaard, J., and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086.
- [14] Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134.
- [15] Tu, D. S., and Zheng, Z. G. (1987). On the Edgeworth expansions of random weighing methods. *Chinese J. Appl. Probab. Statist.* **3** 340–347.