1. Introduction and main results. Let $F_1, F_2, \ldots, F_k$ be the distributions of $k$ populations with means $\mu_1, \mu_2, \ldots, \mu_k$. Let $\theta = \sum_i^k l_i \mu_i$ where $l_1, l_2, \ldots, l_k$ are non-zero constants. Let $(X_{i1}, X_{i2}, \ldots, X_{in_i})$, $i = 1, 2, \ldots, k$, be independent random samples of sizes $n_1, n_2, \ldots, n_k$ from $F_1, F_2, \ldots, F_k$. Let $n$ denote the vector $(n_1, n_2, \ldots, n_k)$ and $N = \sum_i^k n_i$. A natural estimator for $\theta$ is $\hat{\theta}_n = \sum_i^k l_i \bar{X}_i$ and a consistent estimator for its variance is $v_n^2 = \sum_i^k l_i^2 s_i^2 / n_i$ where $\bar{X}_i = n_i^{-1} \sum_j^{n_i} X_{ij}$ and $s_i^2 = (1/n_i) \sum_j^{n_i} (X_{ij} - \bar{X}_i)^2$. Here we study the accuracy of the bootstrap approximation to the distribution of the standardized random variable $t_n = (\hat{\theta}_n - \theta) / v_n$. This approximation is discussed in the next paragraph. Although one could base an inference about $\theta$ on the difference $\hat{\theta}_n - \theta$ itself, it turns out that the bootstrap approximation is asymptotically more accurate for $t_n$ than for $(\hat{\theta}_n - \theta)$.

Let $G_i$ denote the empirical distribution function based on $(X_{i1}, X_{i2}, \ldots, X_{i(n_i)})$, $i = 1, 2, \ldots, k$. The dependence of $G_i$'s on the sample sizes is suppressed in the notation. Now let $(Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$, $i = 1, 2, \ldots, k$, denote independent random samples from the populations $G_1, G_2, \ldots, G_k$; let $\bar{Y}_i = n_i^{-1} \sum_j^{n_i} Y_{ij}$ and $\gamma_i^2 = n_i^{-1} \sum_j^{n_i} (Y_{ij} - \bar{Y}_i)^2$. Then, by definition, the distribution of $t_n^* = \sum_i^k l_i(\bar{Y}_i - \bar{X}_i)/\sum_i^k l_i^2 \gamma_i^2$ under $G_1, G_2, \ldots, G_k$ is the bootstrap distribution of $t_n$. Under the conditions given below, the bootstrap distribution of $t_n^*$ is shown to be asymptotically close to the actual distribution of $t_n$ up to $o(N^{-1/2})$. In applications the bootstrap distribution is approximated by drawing samples of sizes $n_1, n_2, \ldots, n_k$ from $G_1, G_2, \ldots, G_k$ a large number of times, say $M$ times, calculating $t_n^*$ each time and finally forming an empirical histogram. It is shown here that this second stage approximation is good up to $o(N^{-1/2})$ provided $M/(N \log N) \to \infty$.

We now state the main results proved in this note. Throughout, we make the following assumptions, to be referred to as A in the sequel.

A. $F_i$ has finite 6th moment for all $1 \leq i \leq k$. For at least one $i$, $F_i$ is continuous. Without loss of generality we shall assume that $F_i$ is continuous. The $n_i$'s tend to infinity at the same rate. In other words, the $N/n_i \leq \lambda < \infty$ for all $i = 1, 2, \ldots, k$. In practice this last condition means that the $n_i$'s are of comparable size.

In what follows, for any distribution $F_i$ let $F^{-1}(t) = \inf\{x: F(x) : \geq t\}$, where $0 < t < 1$.

**Theorem.** If $H_n$ denotes the d.f. of $t_n$ and $H_n^*$ denotes the d.f. of $t_n^*$ then, under A, as $N \to \infty$

$$N^{1/2} \sup_{x \in R} |H_n(x) - H_n^*(x)| \to 0$$

and

$$N^{1/2} |H_n^{-1}(t) - H_n^*{-1}(t)| \to 0$$

a.s. for all $t \in (0, 1)$. Further let $H_{n,M}$ denote the approximation to $H_n^*$ described in the second paragraph above with $M$ samples from $G_i$'s. If $M/(N \log N) \to \infty$ as $N \to \infty$, then

---

Received November 1981; revised August 1982.

1 On leave from the Indian Statistical Institute.

2 Research supported by NSF Grant MCS-81-02341.

AMS 1970 subject classifications. 62G05; 62G15.

Key words and phrases. Bootstrap, Edgeworth expansions, empirical distribution function, confidence interval.
for almost all sample sequences \( \{X_i\} \)

\[
N^{1/2} \sup_{x \in R} |H_{n,M}(x) - H_n^*(x)| \to 0
\]

and

\[
N^{1/2} |H_{n,M}^{-1}(t) - H_n^{-1}(t)| \to 0
\]

as.s. for all \( t \in (0, 1) \) as \( N \to \infty \). The a.s. here refers to the random mechanism generating the samples from \( G_i's \). (We assume that all the second stage sample sequences are defined on the same space.)

It may be mentioned here that (1) in the above theorem is an extension of (1.5) in [8] which is a result involving \( (X - \mu)/\sigma \). For constructing a confidence interval for \( \theta \), one may replace an actual quantile \( H_n^{-1}(a) \) of \( t_n \) by its bootstrap approximation \( H_{n,M}^{-1}(a) \). This approximation in the one sample case has been investigated by Efron [6] on simulated data from an asymmetric population. The procedure performed quite well (see Table 5 of [6]).

2. Proof of the theorem. We first develop some notation. Let \( \phi_x, \Phi_x \) denote the density and the d.f. of a normal variable with mean zero and dispersion matrix \( \Sigma \); let \( \phi, \Phi \) denote the density and the d.f. of a standard normal variable in \( R \); let \( c \) denote a constant, the later may denote different constants at different places. For non-negative integral vectors \( \beta = (\beta_1, \cdots, \beta_r) \) and \( x \in R^r \) let \( x^\beta = \prod_{i=1}^r x_i^{\beta_i}, \beta! = \beta_1! \beta_2! \cdots \beta_r! \), \( |\beta| = \beta_1 + \cdots + \beta_r \) and \( D^\beta = D_1^{\beta_1} \cdots D_r^{\beta_r} \), where \( D_i^{\beta_i} \) denotes the \( \beta_i \)th order derivative with respect to the \( i \)th variable. Finally let \( \|x\|^2 = x_1^2 + \cdots + x_r^2 \) and \( x \cdot y = x_1 y_1 + \cdots + x_r y_r \), where \( x = (x_1, \cdots, x_r) \) and \( y = (y_1, \cdots, y_r) \).

We shall show that

\[
P(t_n^* \leq x) = \Phi(x) + N^{-1/2} \int_{-\infty}^{x} d(y) \phi(y) \, dy + o(N^{-1/2}) \quad \text{a.s.}
\]

where \( d \) is a polynomial whose coefficients depend upon \( F_i's \). The same steps will also yield

\[
P(t_n \leq x) = \Phi(x) + N^{-1/2} \int_{-\infty}^{x} d(y) \phi(y) \, dy + o(N^{-1/2}).
\]

Clearly (5) and (6) imply (1). Before proving (5) we shall deduce (2), (3) and (4) from (5) and (6).

To prove (3), first note that in distribution (given the original sample) \( H_{n,M} \) is the same as the empirical d.f. of \( H_n^{-1}(U_i) \) where \( U_1, U_2, \cdots, U_M \) are i.i.d. \( U[0, 1] \) random variables. If \( E_M \) denotes the empirical d.f. of \( U_1, \cdots, U_M \) then \( \#(H_n^{-1}(U_i) \leq x) = ME_M(H_n(x)) \). Hence, using a well known bound on \( E_m \), we have

\[
P(\sup_{x \in R} |H_{n,M}(x) - H_n(x)| \geq 4M^{-1/2} (\log M)^{1/2})
\]
\[
\leq P(\sup_{x \in [0,1]} |E_M(t) - t| \geq 4M^{-1/2} (\log M)^{1/2}) = O(M^{-2}).
\]

Consequently, in view of Borel-Cantelli lemma,

\[
\lim_{M \to \infty} M^{-1/2} (\log M)^{-1/2} \sup_{x \in R} |H_{n,M}(x) - H_n(x)| \leq 4 \quad \text{a.s.}
\]

Here dependence of \( M \) on \( N \) is suppressed. The claim (3) in the theorem follows from this, since \( M/(N \log N) \to \infty \).

The claims (2) and (4) on quantiles follow using Lemma 1 given below which is an easy consequence of Taylor's expansion.

**Lemma 1.** Let \( L_N \) be a sequence of d.f.'s on the real line such that, for a polynomial
with its coefficients bounded in $N$,
\[ L_N(x) = \int_\mathbb{R} \left[ 1 + N^{-1/2} a_N(y) \right] \phi(y) \, dy + o(N^{-1/2}) \]
uniformly in $x$. Then for each $\alpha \in (0, 1)$,
\[ L_N^\alpha(\alpha) = z - (\phi(z) \sqrt{N})^{-1} \int_{-\infty}^z a_N(y) \phi(y) \, dy + o(N^{-1/2}) \]
where $z = \phi^{-1}(\alpha)$.

The proof of (5) is based on Lemmas 2–5 that follow. In the proofs below we assume w.l.o.g. that $l_1 = l_2 = \cdots = l_k = 1$.

All proofs are given for a single sequence of realizations of the $X_i$ for which $G_i$ converges weakly to $F_i$ and $\int x^6 \, dG_i \rightarrow \int x^6 \, dF_i$ for $j = 1, 2, \cdots, k$. Thus in view of A the results hold a.s.

**Lemma 2.** Let $Y$ be a random vector in $\mathbb{R}^2$ with mean zero and dispersion matrix $V = ((\mu_{ij}))$. Suppose for some $b > 1$, $\max(|v_{11}|, |v_{12}|, |v_{22}|, E \| Y \|^3) < b$. Let $a > 2$ be such that $\Delta(a) < 1/10$, where $\Delta(a) = (1/a) + E(\| Y \|^2(\| Y \| > a))$. Then for all $\| t \| \leq a^{-3} \sqrt{N}$ and all non-negative integral vectors $a$ with $|a| \leq 3$,
\[ |D^a(g^N(t/\sqrt{N}) - (1 - (i/6 \sqrt{N}) E(t \cdot Y)^3) \exp(-t' V t/2))| \leq c b^a \Delta(a) N^{-1/2} (\| t \|^a + 1) \exp(-t' V t/2 + c a^{-1} b^3 \| t \|^2) \]
for $t \in \mathbb{R}^2$, $g(t) = E(\exp(it \cdot Y))$.

The proof is similar to the proof of Theorem 9.9 of [3].

**Lemma 3.** Suppose $A$ holds. Let $\lambda_j = (N/n_j)^{1/2}$, $Z_j = [\lambda_j(Y_j - \bar{X}_j), \lambda_j^3((Y_j - \bar{X}_j)^2 - s_i^2)]$, $g_i$ denote the characteristic function of $Z_i/\sqrt{n_j}$, $B_j$ denote the dispersion matrix of $Z_i$ and $B = \sum_{j=1}^N B_j$. Then for any $\eta > 0$,
\[ \max_{|\beta| = 3} \int_{|t| \leq \eta} \| D^\beta \left[ \prod_{j=1}^k g_j(t) - e^{-t B t/2} \left( 1 - \frac{i}{6 \sqrt{N}} \sum_{j=1}^k \lambda_j E(t \cdot Z_j)^3 \right) \right] \| dt = o(N^{-1/2}). \]

**Proof.** Define
\[ f_j(t) = (1 - (i/6 \sqrt{n_j}) E(t \cdot Z_j)^3) \exp(-t' B_j t/2). \]
First note that
\[ \max_{|\beta| = 3} |D^\beta(e^{-t B t/2}(1 - (i/6 \sqrt{N}) \sum_{j=1}^k \lambda_j E(t \cdot Z_j)^3) - \prod_{j=1}^k f_j(t))| = O(N^{-1}(\| t \|^{3k+1} + 1)) \exp(-t' B t/2), \]
and for any non-empty subset $J$ of $\{1, 2, \cdots, k\}$,
\[ \max_{|\beta| = 3} |D^\beta \prod_{j \in J} g_j^\alpha(t)| \leq \max_{|\beta| = 3} E \| \sum_{j \in J} n_j^{1/2} \sum_{j=1}^k Z_j \|^3 |\beta| \]
\[ \leq 1 + k^3 \max_{1 \leq j \leq k} n_j^{-3/2} E \| \sum_{j=1}^k Z_j \|^3 = O(1), \]
where $Z_j$ are independent copies of $Z_j$. The last inequality above follows from the proof of Lemma 14.7 of [3] as sup $E \| Z_j \|^b \leq b < \infty$ from some $b > 0$. Also for $1 \leq j \leq k$
\[ \max_{|\beta| = 3} |D^\beta \prod_{j \in J} f_j(t)| = O((1 + \| t \|^{3k}) \exp(-t' B t/2)). \]
By (8), (9) and Lemma 2, we have for any \( a \geq 2, \beta \leq 3 \) and \( \| t \| \leq \lambda^{-1} a^{-\sqrt{N}} \),
\[
\| D^\beta (\prod_{j=1}^\beta g_j^\beta(t) - \prod_{j=1}^\beta f_j(t)) \| 
\leq \sum_{j=1}^\beta \| D^\beta (\prod_{j=1}^\beta f_j(t)) \| \leq O(N^{-1/2} + N^{-1/2}(1 + \| t \|^d)) \exp(-t'B_1 t/2 + c a^{-1}\| t \|^2),
\]
where \( r(a) = 1/a + \sup E(\| Z \|^2 I(\| Z \| > a)) \). It now follows from (7) and (10) that, for a \( \| \beta \| \leq 3 \), the integral in Lemma 3 over \( \| t \| \leq \lambda^{-1} a^{-\sqrt{N}} \) is \( O(r(a) N^{-1/2} + O(N^{-1})) \).

Since \( F_1 \) is continuous, the dispersion matrix of \( X = (X_{11}, (X_{11} - \mu_1)^2) \) is positive definite and the c.f. \( h \) of \( X \) satisfies the condition \( |h(t)| < 1 \) for all \( t \neq 0 \). As a result of this and the fact that weak convergence implies convergence of c.f.'s. uniformly over compact sets, it follows that
\[
\sup\{ |g_1(t)| ; \| t \| \leq \lambda^{-1} a^{-\sqrt{N}}, \eta \sqrt{N} \} \leq \delta < 1
\]
for all large \( N \). Also,
\[
\inf(\| t'B_1 t \| / \| t \|^2; t \neq 0) \geq b > 0
\]
for all large \( N \) under A. Finally for any \( \| \beta \| \leq 3 \)
\[
\| D^\beta (g_1^\beta(t)) \| \leq n_1^\beta E(\| Z \|^2 I(\| Z \| > a)) \leq c N^{3} \| g_1(t) \|^n - 3.
\]
Thus for \( \| \beta \| \leq 3 \), the integral in the lemma over \( \lambda^{-1} a^{-\sqrt{N}} \leq \| t \| \leq \eta \sqrt{N} \) is \( O(N^{-1}) \). The claim now follows by letting \( a \to \infty \).

Next, an inversion theorem is obtained by combining a modification of Lemma 5 in [9] with Lemma 11.6 of [3]. The proof is deleted.

**Lemma 4.** Let \( P \) be a probability on \( R^k \) and \( Q \) denote a measure with density \( (1 + N^{-1/2} p(y)) \phi_2(y) \) where \( p(y) \) is a polynomial and \( \Sigma \) is a positive definite matrix of order \( k \times k \). Let the coefficients of \( p(y) \), \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) be bounded by \( M > 0 \) where \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) denote the maximum and minimum eigenvalues of \( \Sigma \). Then for any \( \epsilon > 0 \)
\[
|P(C) - Q(C)| \leq c(k) \max_{|\beta| \leq k+1} \int_{t < \epsilon \sqrt{N}} |D^\beta (\hat{P}(t) - \hat{Q}(t))| \, dt 
+ c(M)\Phi_2((\partial C)\epsilon/\sqrt{N}) + O(N^{-1}).
\]
Here \( \hat{P} \) and \( \hat{Q} \) stand for c.f.'s of \( P \) and \( Q \); \( (\partial C)\epsilon/\sqrt{N} \) is the \( \epsilon/\sqrt{N} \) neighborhood of the boundary of \( C \).

Finally Lemma 5 justifies converting a multivariate one-term Edgeworth expansion into an univariate one. This result is a modification of Lemma 2.1 of [2]. A proof for the present version is contained in [1].

**Lemma 5.** Let \( t = (t_1, t_2, \ldots, t_r) \) be a vector, \( L = (L_{ij}) \) be a \( r \times r \) matrix and \( q \) be a polynomial in \( r \) variables. Let \( M \geq \max\{ |v_i|, |u_i|, |t_i|, |L_{ij}|, |c_i| \} \), where \( V = ((v_i)) \) is a positive definite matrix \( ((v_i)) = V^{-1} \) and \( c_i \) are the coefficients of \( q \). Let \( |t_i| > t_0 > 0 \). Then there exists a polynomial \( p \) in one variable, whose coefficients are continuous functions of \( t_i, L_{ij}, v_i, u_i \), and \( c_i \), such that
\[
\int_{-\pi \leq y \leq \pi} \left( 1 + N^{-1/2} q(z) \phi_V(z) \right) \, dz = \int_{-\pi}^\pi \left( 1 + N^{-1/2} p(y) \phi(y) \right) \, dy + o(N^{-1/2})
\]
where the \( o(\cdot) \) term depends on \( M \) and \( t_0 \) only.

We now briefly sketch the proof of (5) using the lemmas. Define, \( \xi_j = n_j^{-1} \sum_{i=1}^n (Y_j - \bar{X}_j)^2 - s_j^2 \) and \( s^2 = \sum_i s_i^2/n_j \). From Lemmas 3 and 4 it follows that for a measurable \( C \) and
\[ \varepsilon > 0, \]
\[ P[(\sqrt{N} \sum_j (\bar{Y}_j - \bar{X}_j), N^{3/2} \sum_j (\xi_j/n_j)) \in C] \]
\[ = \int_C \phi_B(x)[1 + N^{-1/2}a_N(x)] dx + o(N^{-1/2}) + O(\Phi_B(aB^{1/\sqrt{N}})) \]

where \( a_N \) is a polynomial whose coefficients are polynomials in \( \{\lambda_j\} \), and the moments of \( G_j \) of order 6 or less. Note that \( B = (b_{ij})_{2 \times 2} \) with \( b_{11} = Ns^2 \), the variance of \( \sqrt{N} \sum_j \bar{Y}_j \). Now (11) combined with Lemma 5 entails

\[ P(s^{-1} \sum_j (\bar{Y}_j - \bar{X}_j)(1 - (\%)/s^2 \sum_j (\xi_j/n_j)) \leq x) \]
\[ = \Phi(x) + N^{-1/2} \int_{-\infty}^X b(y)\phi(y) dy + o(N^{-1/2}), \]

where \( b \) is a polynomial whose coefficients are continuous functions of \( B^{-1} \), \( \lambda_j \) and the moments of \( G_j \) of order 6 or less.

Define \( C_N = (\sqrt{N} \sum_j |\bar{Y}_j - \bar{X}_j| < \log N \) and \( D_N = (N^{3/2} \sum_j (\xi_j/n_j) | < \log N \). On \( C_N \cap D_N \) one has

\[ t^*_N = s^{-1} \sum_j (\bar{Y}_j - \bar{X}_j)(1 - (\%)/s^2 \sum_j (\xi_j/n_j)) + O(N^{-1}(\log N)^2) \]

(taking \( l_1 = \cdots = l_k = 1 \)). Since the 6th moments of \( \{Y_{ij}\} \) are bounded, it follows from the proof of Theorem 2 of [7] that

\[ [1 - P(C_N)] + [1 - P(D_N)] = o(N^{-1/2}). \]

Thus (12), (13) and (14) yield (5)

REFERENCES