ON THE BOOTSTRAP AND CONFIDENCE INTERVALS[1]

by

Peter Hall
University of North Carolina, Chapel Hill[2]

ABSTRACT. We derive an explicit formula for the first term in an *unconditional* Edgeworth-type expansion of coverage probability for the nonparametric bootstrap technique applied to a very broad class of "studentized" statistics. The class includes sample mean, k-sample mean, sample correlation coefficient, maximum likelihood estimators expressible as functions of vector means, etc. We suggest that the bootstrap is really an empiric one-term Edgeworth inversion, with the bootstrap simulations implicitly estimating the first term in an Edgeworth expansion. This view of the bootstrap is reinforced by our discussion of the iterated bootstrap, which inverts an Edgeworth expansion to arbitrary order by simulating simulations.

AMS (1980) SUBJECT CLASSIFICATION: Primary 62E20, Secondary 62G05.

KEY WORDS AND PHRASES: Bootstrap, central limit theorem, Edgeworth expansion, rate of convergence, Studentized statistic.

SHORT TITLE: Bootstrap.

---

[1]Research partially supported by AFOSR Contract No. F49620 82 C 0009.

[2]On leave from Australian National University

1. Introduction and summary. In this paper it is argued that Efron's bootstrap [8,9] is in effect an empiric one-term Edgeworth inversion, at least from the point of view of its role in constructing confidence intervals and hypothesis tests. The bootstrap argument uses the empiric distribution to estimate the first term in an Edgeworth expansion, and correct for it. That task may be performed directly, as described in [10], although the bootstrap provides a correction which is smoother (i.e. uniform) and automatic (i.e. does not require a preliminary theoretical calculation of the expansion's first term). There exist other smooth inversions [11], although we know of none other which is automatic.

We shall show that the bootstap may be iterated to yield an approximation with an error of arbitrarily small order, in the same way that direct Edgeworth inversions were iterated in [10]. Bootstrap iteration involves simulations of simulations, and although it is almost prohibitively laborious to implement in practice, its theoretical properties provide new insight into the nature of the bootstrap algorithm.

We shall obtain an explicit formula for the first error term after the bootstrap correction, in a very wide class of "Studentized" statistics including the mean, variance, correlation coefficient, maximum likelihood estimators expressible as explicit or implicit functions of vector means, etc. The formula for the error term may be used to predict the performance of the bootstrap; see the comments following display (3.2) below. Our expansion is very different from those described by Singh [13] and Babu and Singh [2] in the care of k-sample Studentized means.

The expansions studied by Babu and Singh hold with probability one along a sample path, and are of interest if a statistician wishes to know the order of approximation for his particular sample. In contrast, our expansions are not conditional on a sample path, and so are useful in describing coverage probabilities or significance levels in the classical frequentist sense. Our general formulation of the problem includes k-sample means as a particular case.

The basic result is described in Section 2 and proved in Section 5. Its implications, and issues arising from its proof, are discussed in Section 3. These matters lead naturally to bootstrap iteration, described in Section 4.

Several ideas in our proofs are borrowed from [5,6,10], and in those places we omit many details. In addition to the works of Babu and Singh cited above, the reader is referred to Beran [3,4] and Bickel and Freedman [7] for an asymptotic account of the bootstrap.

2. Notation and basic result.

Let $\underset{\sim}{Y}, \underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots$ be independent and identically distributed d-dimensional random vectors with mean $\underset{\sim}{\mu} = E(\underset{\sim}{Y})$. Define $\mathcal{Y} \equiv \{\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n\}$ and $\overline{\underset{\sim}{Y}} \equiv n^{-1} \sum_{r=1}^{n} \underset{\sim}{Y}_r$. Denote the i'th elements of $\underset{\sim}{Y}, \underset{\sim}{Y}_r, \overline{\underset{\sim}{Y}}$ and $\underset{\sim}{\mu}$ by $Y_i, Y_{ri}, \overline{Y}_i$ and μ_i, respectively. Let f be a real-valued function on $\mathbb{R}^d$ with at least one continuous derivative. We shall consider the problem of interval inference based on $\gamma_1(\mathcal{Y}) \equiv f(\overline{\underset{\sim}{Y}}) - f(\underset{\sim}{\mu})$. Typically $f(\overline{\underset{\sim}{Y}}) \equiv \hat{\theta}$ is an estimate of a parameter $f(\underset{\sim}{\mu}) \equiv \theta$.

Under appropriate regularity conditions, $n^{\frac{1}{2}}\gamma_1(\mathcal{Y})$ is approximately normally distributed with zero mean and variance given by

(2.1) $\sigma^2 \equiv \sum_i \sum_j \sigma_{ij} f_i(\underset{\sim}{\mu}) f_j(\underset{\sim}{\mu})$,

where $\sigma_{ij} \equiv \text{cov}(Y_i, Y_j)$ and $f_i(\underset{\sim}{\mu}) \equiv (\partial/\partial\mu_i) f(\underset{\sim}{\mu})$. Usually the value of σ^2 would not be known, and would be replaced by an estimate such as

(2.2) $\hat{\sigma}^2 \equiv \sum_i \sum_j \hat{\sigma}_{ij} f_i(\underset{\sim}{\bar{Y}}) f_j(\underset{\sim}{\bar{Y}})$,

where $\hat{\sigma}_{ij} \equiv n^{-1} \sum_{r=1}^{n} (Y_{ri} - \bar{Y}_i)(Y_{rj} - \bar{Y}_j)$. We would focus attention on $\gamma_2(\mathcal{Y}) \equiv \hat{\sigma}^{-1} \gamma_1(\mathcal{Y})$, rather than $\gamma_1(\mathcal{Y})$.

Let us adjoin to the vector $\underset{\sim}{Y}_r$ the set of those products $Y_{ri}Y_{rj}$, $1 \le i \le j \le d$, which do not already appear in $\underset{\sim}{Y}_r$. Thus, $\underset{\sim}{Y}_r$ is expanded to $\underset{\sim}{Y}_r^o$, say, of length $d_o \le d(d+3)/2$. Let $\underset{\sim}{Y}^o$ and $\underset{\sim}{\mu}^o = E(\underset{\sim}{Y}^o)$ denote the corresponding lengthenings of $\underset{\sim}{Y}$ and $\underset{\sim}{\mu}$, and let $\underset{\sim}{\bar{Y}}^o \equiv n^{-1} \sum_{r=1}^{n} \underset{\sim}{Y}_r^o$. We may write $\gamma_2(\mathcal{Y})$ as a function of $\underset{\sim}{\bar{Y}}^o$ alone: $g(\underset{\sim}{\bar{Y}}^o | \underset{\sim}{\mu}) \equiv \{f(\underset{\sim}{\bar{Y}}) - f(\underset{\sim}{\mu})\}/\hat{\sigma}$. Clearly $g(\underset{\sim}{\mu}^o | \underset{\sim}{\mu}) = 0$. The asymptotic variance of $n^{\frac{1}{2}} g(\underset{\sim}{\bar{Y}}^o | \underset{\sim}{\mu})$ is unity, which is reflected in the fact that if we construct the quantity σ^2 at (2.1) for $g(\cdot | \underset{\sim}{\mu})$ instead of $f(\cdot)$, we obtain precisely 1.

The bootstrap argument runs as follows. Condition on the sample $\mathcal{Y}$, and let $\underset{\sim}{Z}, \underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n$ be independent and identically distributed with the n-point distribution $P(\underset{\sim}{Z} = \underset{\sim}{Y}_r | \mathcal{Y}) = n^{-1}$, $1 \le r \le n$. Set $\underset{\sim}{\bar{Z}} \equiv n^{-1} \sum_{r=1}^{n} \underset{\sim}{Z}_r$. We may work out the distribution of $g(\underset{\sim}{\bar{Z}}^o | \underset{\sim}{\bar{Y}})$, conditional on $\mathcal{Y}$, to arbitrary accuracy using simulation. The distribution is discrete. Define

$$t_\alpha = t_\alpha(\mathcal{Y}) \equiv \inf\{t: P[n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}}^o | \underset{\sim}{\bar{Y}}) \le t | \mathcal{Y}] \ge \alpha\},$$

for $0 < \alpha < 1$. Then t_α is the (nonparametric) bootstrap approximation to the upper $(1-\alpha)$-level critical point of $n^{\frac{1}{2}} g(\underset{\sim}{\bar{Y}}^o | \underset{\sim}{\mu})$, and may be used in the construction of confidence intervals or hypothesis tests. Theorem 2.1 below describes the accuracy of the approximation.

Before stating the theorem we list notation and technical conditions. Given any vector $\underset{\sim}{u}^{o} = (u_1,\ldots,u_d,u_{11},u_{12},\ldots,u_{dd})^T = (u_1,\ldots,u_{d^o})^T$ of length d^o, define $\underset{\sim}{u} \equiv (u_1,\ldots,u_d)^T$ (the first d elements of $\underset{\sim}{u}^o$),

$$\sigma^2(\underset{\sim}{u}^o) \equiv \sum_i \sum_j (u_{ij} - u_i u_j)\, f_i(\underset{\sim}{u}) f_j(\underset{\sim}{u}) \ ,$$

$$g(\underset{\sim}{u}^o|\underset{\sim}{v}) \equiv \{f(\underset{\sim}{u}) - f(\underset{\sim}{v})\} \ / \ \{\sigma^2(\underset{\sim}{u}^o)\}^{\frac{1}{2}}$$

and $g_{i_1\ldots i_p}(\underset{\sim}{u}^o|\underset{\sim}{v}) \equiv (\partial^p/\partial u_{i_1}\ldots\partial u_{i_p})\, g(\underset{\sim}{u}^o|\underset{\sim}{v})$. (Recall from the definition of $\underset{\sim}{Y}^o$ that only those products Y_iY_j not already occurring in $\underset{\sim}{Y}$ appear in the extension $\underset{\sim}{Y}^o$. If $Y_iY_j = Y_k \in \underset{\sim}{Y}$, we define $u_{ij} = u_k$.) Let $B(\underset{\sim}{u},\rho)$ denote the open ball in $\mathbb{R}^d$ centered at $\underset{\sim}{u}$ and of radius ρ, and define $B(\underset{\sim}{u}^o,\rho)$ analogously. Given an extended mean vector $\underset{\sim}{\mu}^o$ with $\sigma^2(\underset{\sim}{\mu}^o) > 0$, assume that on some set $B(\underset{\sim}{\mu}^o,\rho) \times B(\underset{\sim}{\mu},\rho)$ of vectors $(\underset{\sim}{u}^o,\underset{\sim}{v})$, $\sigma^2(\underset{\sim}{u}^o) > 0$ and $g(\underset{\sim}{u}^o|\underset{\sim}{v})$ has four continuous derivatives with respect to $\underset{\sim}{u}^o$. Let $h(\underset{\sim}{u}^o|\underset{\sim}{v})$ denote any of these derivatives of order one, two or three. Assume that $h(\underset{\sim}{u}^o|\underset{\sim}{u})$ has four continuous derivative with respect to $\underset{\sim}{u}^o$, for each choice of h and all $\underset{\sim}{u}^o \in B(\underset{\sim}{\mu}^o,\rho)$. Suppose the distribution of $\underset{\sim}{Y}$ satisfies $E(\underset{\sim}{Y}^o) = \underset{\sim}{\mu}^o$ and $E(\|\underset{\sim}{Y}^o\|^{8+\epsilon}) < \infty$, for some $\epsilon > 0$. Further assume that $\underset{\sim}{Y}^o$ satisfies Cramér's continuity condition:

$$\limsup_{|\underset{\sim}{t}|\to\infty} |E\{\exp(i < \underset{\sim}{t}, \underset{\sim}{Y}^o >)\}| < 1 \ . \tag{2.3}$$

(One consequence of (2.3) is that $\underset{\sim}{Y}^o$ has nonsingular variance matrix.) Note that by definition of g,

$$\text{(2.4)} \qquad g(\underset{\sim}{u}^o|\underset{\sim}{u}) = 0 \text{ and } \sum_{i=1}^{d} \sum_{j=1}^{d} (u_{ij} - u_i u_j) g_i(\underset{\sim}{u}^o|\underset{\sim}{u}) g_j(\underset{\sim}{u}^o|\underset{\sim}{u}) = 1$$

for all $\underset{\sim}{u}^o \in B(\underset{\sim}{\mu}^o, \rho)$. Define

$$\ell_{i_1 \ldots i_p} \equiv g_{i_1 \ldots i_p}(\underset{\sim}{\mu}^o|\underset{\sim}{\mu}), \quad \ell_{i_1 \ldots i_p}^{(j)} \equiv (\partial/\partial u_j) g_{i_1 \ldots i_p}(\underset{\sim}{u}^o|\underset{\sim}{u})\Big|_{\underset{\sim}{u}^o = \underset{\sim}{\mu}^o}$$

and $\alpha_{i_1 \ldots i_p} \equiv E\{\prod_{j=1}^{p} (Y_{i_j} - \mu_{i_j})\}$. Define the odd, quintic polynomial

$$\text{(2.5)} \qquad \psi(z) \equiv (z/4) \sum_i \sum_j \sum_p \{2 + (z^2-3)(\ell_p^2 \alpha_{pp} - 1)\}$$

$$\times \{z^2(\ell_p \alpha_{ij} \sum_k \ell_{ij}^{(k)} \alpha_{kp} + \ell_{ij}\, \ell_p\, \alpha_{ijp})$$

$$+ (1/3)(z^2-1)(3\ell_i \ell_j \ell_p \sum_k \sum_m \ell_k^{(m)} \alpha_{ijk}\, \alpha_{mp} + \ell_i \ell_j \ell_p \sum_k \ell_k\, \alpha_{ijkp}$$

$$- 3\ell_i \ell_j \ell_p\, \alpha_{ij} \sum_k \ell_k\, \alpha_{kp})\} .$$

Let t_α be as in the previous paragraph, and let Φ and ϕ be the standard normal distribution and density functions, respectively.

THEOREM 2.1. *Under the above conditions,*

$$\text{(2.6)} \qquad P\{n^{\frac{1}{2}} g(\underset{\sim}{\bar{Y}}^o|\underset{\sim}{\mu}) \leq t_\alpha(\mathcal{Y})\} = \alpha + n^{-1}\psi\{z(\alpha)\}\, \phi\{z(\alpha)\} + o(n^{-1})$$

uniformly in α, $0 < \alpha < 1$, *where* $z(\alpha)$ *is the solution of* $\Phi(z) = \alpha$.

An alternative approach is to bootstrap the statistic $g^*(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}^o)$, where

$$g^*(\underset{\sim}{u}|\underset{\sim}{v}^o) \equiv \{f(\underset{\sim}{u}) - f(\underset{\sim}{v})\} \,/\, \{\sigma^2(\underset{\sim}{v}^o)\}^{\frac{1}{2}} ,$$

instead of $g(\underset{\sim}{\overline{Y}}^o|\underset{\sim}{\mu})$. In that case we use the critical point estimate

$$t^*_\alpha = t^*_\alpha(\mathcal{Y}) \equiv \inf\{t: P[n^{\frac{1}{2}} g^*(\underset{\sim}{\overline{Z}}|\underset{\sim}{\overline{Y}}^o) \leq t|\mathcal{Y}] \geq \alpha\} ,$$

instead of t_α. Construct a new polynomial ψ^* whose coefficients are obtained by replacing $\ell_{i_1 \ldots i_p}$ and $\ell^{(j)}_{i_1 \ldots i_p}$ by $\ell^*_{i_1 \ldots i_p}$ and $\ell^{(j)*}_{i_1 \ldots i_p}$, respectively, in formula (2.5), where $g^*_{i_1 \ldots i_p}(\underset{\sim}{u}|\underset{\sim}{v}^o) \equiv (\partial^p/\partial u_{i_1} \ldots \partial u_{i_p}) g^*(\underset{\sim}{u}|\underset{\sim}{v}^o)$,

$$\ell^*_{i_1 \ldots i_p} \equiv g^*_{i_1 \ldots i_p}(\underset{\sim}{\mu}|\underset{\sim}{\mu}^o), \quad \ell^{(j)*}_{i_1 \ldots i_p} \equiv (\partial/\partial u_j) g^*_{i_1 \ldots i_p}(\underset{\sim}{u}|\underset{\sim}{u}^o)|_{\underset{\sim}{u}^o = \underset{\sim}{\mu}^o} .$$

Instead of Theorem 2.1, we obtain:

THEOREM 2.2. *Under the above conditions, with the smoothness constraints applied to* $g^*(\underset{\sim}{u}|\underset{\sim}{v}^o)$ *instead of* $g(\underset{\sim}{u}^o|\underset{\sim}{v})$,

$$(2.7) \qquad P\{n^{\frac{1}{2}} g^*(\underset{\sim}{\overline{Y}}|\underset{\sim}{\mu}^o) \leq t^*_\alpha(\mathcal{Y})\} = \alpha + n^{-1} \psi^*\{z(\alpha)\}\phi\{z(\alpha)\} + o(n^{-1})$$

uniformly in α, $0 < \alpha < 1$.

Theorem 2.2 may be derived almost identically to Theorem 2.1, and so the proof will not be given here. The simulations required to calculate t^*_α are usually less time-consuming than those needed to derive t_α, since only the numerator in the formula $g^*(\underset{\sim}{\overline{Z}}|\underset{\sim}{\overline{Y}}^o) \equiv \{f(\underset{\sim}{\overline{Z}}) - f(\underset{\sim}{\overline{Y}})\} / \sigma(\underset{\sim}{\overline{Y}})$ is being simulated. However, result (2.7) can only be used to construct confidence intervals for $f(\underset{\sim}{\mu})$ when $\sigma^2(\underset{\sim}{\mu}^o)$ is known, whereas (2.6) can be used even when $\sigma^2(\underset{\sim}{\mu}^o)$ is unknown.

3. Discussion. Our discussion will be confined to Theorem 2.1.

3.1. *A Cornish-Fisher view of the bootstrap.* Let $s \equiv n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma}$ be a general "Studentized" statistic. In a great many cases, s admits an Edgeworth expansion:

$$P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq x\} = \Phi(x) + \sum_{j=1}^{k} n^{-j/2}\pi_{1j}(x)\phi(x) + o(n^{-k/2})$$

uniformly in x, where π_{1j} is a polynomial of degree 3j-1. Whenever it exists, this expansion may be inverted to yield an expansion of (inverse) Cornish-Fisher type:

$$P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq x + \sum_{j=1}^{k} n^{-j/2}\pi_{2j}(x)\} = \Phi(x) + o(n^{-k/2})$$

uniformly on compact intervals, where $\{\pi_{2j}\}$ is a new sequence of polynomials. An alternative, equivalent form is

$$P[n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq z(\alpha) + \sum_{j=1}^{k} n^{-j/2}\,\pi_{2j}\{z(\alpha)\}] = \alpha + o(n^{-k/2}) \tag{3.1}$$

uniformly in $\alpha \in (\varepsilon, 1-\varepsilon)$, any $\varepsilon > 0$, where z is the solution of $\Phi(z) = \alpha$.

The coefficients of the π_{2j}'s depend on the sampling distribution through its moments. Let Π_{2j} denote the version of π_{2j} with each population moment replaced by the corresponding sample moment. It is not difficult to show that for each k (modulo regularity conditions), and for each $\alpha \in (0,1)$,

$$t_\alpha = z(\alpha) + \sum_{j=1}^{k} n^{-j/2}\,\Pi_{2j}\{z(\alpha)\} + o_p(n^{-k/2})$$

as $n \to \infty$. In this sense, the bootstrap critical point estimate t_α is asymptotically equivalent to the critical point estimate obtained by empiric inverse Cornish-Fisher expansion.

3.2. *The bootstrap and Edgeworth inversion.* If we replace each π_{2j} in (3.1) by Π_{2j}, then that expansion fails for $k \geq 2$. This follows from the fact that the cumulants of $\Theta_1(z) \equiv n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma}^{-1} - \sum_{j=1}^{k} n^{-j/2}\,\Pi_{2j}(z)$ coincide with those of $\Theta_2(z) \equiv n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma}^{-1} - \sum_{j=1}^{k} n^{-j/2}\,\pi_{2j}(z)$ only to order $n^{-\frac{1}{2}}$. Even-order moments of $\Theta_1(z)$ contain terms of order n^{-1} which do not appear in the corresponding moments of $\Theta_2(z)$. As a result, with $z=z(\alpha)$:

$$
\begin{aligned}
&P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq t_\alpha\} \\
&= P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq z + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(z)\} + o(n^{-1}) \\
&= P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq z + n^{-\frac{1}{2}}\,\Pi_{21}(z) + n^{-1}\,\pi_{22}(z)\} + o(n^{-1}) \\
&= P\{n^{\frac{1}{2}}(\hat{\theta}-\theta)/\hat{\sigma} \leq z + n^{-\frac{1}{2}}\,\pi_{21}(z) + n^{-1}\,\pi_{22}(z)\} + n^{-1}\,\psi(z)\,\phi(z) + o(n^{-1}) \\
&= \alpha + n^{-1}\,\psi(z)\phi(z) + o(n^{-1}),
\end{aligned}
$$

where ψ has the meaning it did in Theorem 2.1. This is essentially the argument used in Section 5 to prove Theorem 2.1.

Note that t_α and $z + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(z)$ produce confidence intervals with the same coverage probability, up to terms $o(n^{-1})$. The latter critical point estimate results from a direct Edgeworth inversion of the type studied in [10].

3.3. *Tightening the bootstrap.* As an illustration, we shall consider use of the bootstrap to set confidence intervals for a population mean. There, $f(u_1) \equiv u_1$, representing the mean, and $\underset{\sim}{u}^{o} = (u_1,u_{11})^T = (u_1,u_2)^T$, with

u_2 representing mean square. The "Studentized" mean is given by

$$g(u_1,u_2|v_1) = (u_1-v_1)/(u_2-u_1^2)^{\frac{1}{2}} .$$

Assume without loss of generality that $E(Y)=0$ and $E(Y^2)=1$. Then $\ell_1=1$, $\ell_{12}=-\frac{1}{2}$, $\ell_1^{(2)}=-\frac{1}{2}$, $\ell_{11}^{(1)}=2$, $\ell_{12}^{(2)}=\frac{3}{4}$, and other terms are zero. Let $\mu_3=E(Y^3)$ and $\mu_4=E(Y^4)$. It follows from the definition of ψ at (2.5) that

$$\psi(z) = -(z/6)(1+2z^2)\{\mu_4-3-(3/2)\mu_3^2\} , \quad -\infty < z < \infty \quad . \tag{3.2}$$

Therefore $\psi(z)$ vanishes if kurtosis and skewness are both zero. Formula (3.2) suggests that the bootstrap approximation may be noticeably in error for skew, leptokurtic distributions. On the other hand, contributions of skewness and kurtosis have some tendency to cancel in the case of skew, platykurtic distributions.

It was shown in [10] that statistics which are representable as functions of vector means admit directly invertible Edgeworth expansions. See [1,12] for alternative approaches. A simple way of tightening the bootstrap approximation is to directly invert the expansion in Theorem 2.1. To this end, define

$$\hat{\psi}(z) \equiv -(z/6)(1+2z^2) \{\hat{\mu}_4-3-(3/2)\hat{\mu}_3^2\} ,$$

where $\hat{\mu}_4 \equiv \hat{\sigma}^{-4} n^{-1} \sum_{r=1}^{n} (Y_r-\bar{Y})^4$, $\hat{\mu}_3 = \hat{\sigma}^{-3} n^{-1} \sum_{r=1}^{n} (Y_r-\bar{Y})^3$, $\hat{\sigma}^2 = n^{-1} \sum_{r=1}^{n} (Y_r-\bar{Y})^2$.

Under the conditions of Theorem 2.1,

$$P[n^{\frac{1}{2}} g(\underset{\sim}{\bar{Y}}^o|\underset{\sim}{\mu}) \le t_\alpha(\mathcal{Y}) - n^{-1} \hat{\psi}\{z(\alpha)\}] = \alpha + o(n^{-1}) \tag{3.3}$$

uniformly in $\alpha \in (\varepsilon,1-\varepsilon)$, any $\varepsilon > 0$. In a sense, this procedure is the *reverse* of one suggested by Abramovitch and Singh [1]: they first-order corrected and then bootstrapped, while we bootstrap and then first-order correct. The end result is very similar.

A more detailed argument shows that under appropriate moment conditions, the right-hand side of (3.3) may be written as $\alpha + n^{-3/2}\,\psi_1(z)\phi(z) + O(n^{-2})$, where ψ_1 is an *even* polynomial. This observation is particularly relevant when using the tightening procedure to construct a *symmetric, two-sided* confidence interval. Recalling that $g(\tilde{\bar{Y}}^0|\tilde{\mu}) = \{f(\tilde{\bar{Y}}) - f(\tilde{\mu})\}/\hat{\sigma}$, with $\hat{\sigma}$ defined at (2.2), we see that the interval

$$(f(\tilde{\bar{Y}}) - n^{-\frac{1}{2}}\hat{\sigma}[t_{1-\alpha/2}(\mathcal{Y}) - n^{-1}\,\hat{\psi}\{z(1-\alpha/2)\}]\,,$$

$$f(\tilde{\bar{Y}}) - n^{-\frac{1}{2}}\,\hat{\sigma}[t_{\alpha/2}(\mathcal{Y}) - n^{-1}\,\hat{\psi}\{z(\alpha/2)\}])$$

covers $f(\tilde{\mu})$ with probability $1-\alpha + O(n^{-2})$, not just $1-\alpha + O(n^{-3/2})$. This property motivates bootstrap tightening: a single correction which improves the bootstrap approximation to order $n^{-3/2}$, actually gives an error as small as order n^{-2} in the case of two-sided confidence intervals.

A disadvantage of the approach described by (3.3) is that it does not apply uniformly in α. In this sense, it is not "smooth"; the bootstrap approximation *is* smooth. Unless $\mu_4-3-(3/2)\mu_3^2 = 0$, (3.3) will fail either as $\alpha \to 0$ or as $\alpha \to 1$. The problems created by this discontinuity are illustrated by simulations summarized in [10]. An alternative, smooth bootstrap tightening may be constructed as follows. Calculate the value $\hat{\lambda} \equiv \hat{\mu}_4-3-(3/2)\hat{\mu}_3^2$. Conditional on $\mathcal{Y}$, construct a known, "comparison distribution" with zero mean and unit variance, having its 3rd and 4th moments μ_{03} and μ_{04} satisfying $\hat{\lambda} = \mu_{04} - 3 - (3/2)\mu_{03}^2$. By simulation or otherwise, calculate that value $\beta = \beta(\alpha)$ such that the bootstrap confidence interval $(-\infty, t_\beta)$ for the comparison distribution covers the comparison version of $n^{\frac{1}{2}} g(\underset{\sim}{\overline{Y}}|\underset{\sim}{\mu})$ precisely 100α % of the time. Then

$$(3.4) \qquad P\{n^{\frac{1}{2}} g(\underset{\sim}{\overline{Y}}|\underset{\sim}{\mu}) \le t_{\beta(\alpha)}(\mathcal{Y})\} = \alpha + o(n^{-1})$$

uniformly in $\alpha \in (0,1)$. The right-hand side of (3.4) may be expanded as $\alpha + n^{-3/2}\psi_2(z)\phi(z) + O(n^{-2})$ for an *even* polynomial ψ_2. Therefore the two-sided interval

$$(f(\underset{\sim}{\overline{Y}}) - n^{-\frac{1}{2}}\hat{\sigma} t_{\beta(1-\alpha/2)}(\mathcal{Y}),\ f(\underset{\sim}{\overline{Y}}) - n^{-\frac{1}{2}}\hat{\sigma} t_{\beta(\alpha/2)}(\mathcal{Y}))$$

covers $f(\underset{\sim}{\mu})$ with probability $1-\alpha + O(n^{-2})$, not just $1-\alpha + O(n^{-3/2})$.

3.4. *Bootstrap brethren.* The arguments described above may be refined in many ways, for example to yield approximations with errors of order $n^{-k/2}$ for arbitrary k. The idea of a "comparison distribution" in Subsection 3.3 is based on a tabular method for correcting normal approximations [11]; in the present case, simulation rather than tabulation is used to calculate the correction. Of course, if simulation can be conducted very rapidly and cheaply then even traditional statistical tables become obsolete. We suggest that the "parametric simulation" used to generate statistical tables, and the "nonparametric simulation" used to construct nonparametric bootstrap critical points, are just two extremes of a vast array of techniques which use simulation to solve problems that are essentially numerical. The "comparison distribution" method is only one example of a simulation approach intermediate between nonparametric and completely parametric. It is possible to construct a smooth (i.e. uniform in $\alpha \in (0,1)$) simulation-based approximation, based on the sample $\mathcal{Y}$ *only through its first four moments,* which corrects two-sided confidence intervals to order n^{-2}. (The tabular analogue of this method was mentioned in [11].) Among these bootstrap brethren, the bootstrap itself stands impressively tall because of its great flexibility and its "automatic" correction of normal approximations to order n^{-1}. Other, intermediate approaches may prove well suited to specific problems.

3.6 *Bias correction.* The bootstrap technique discussed in this paper has been termed the "percentile method" by Efron (1982). Efron has proposed a bias correction for the percentile method. That correction is intended for use only with two-sided confidence intervals, and so has been omitted from our work so far. Its aim is to "centre" a two-sided interval; it does not appear to have a substantial effect on coverage probability.

Versions of Theorems 2.1 and 2.2 may be proved for the bias-corrected percentile method, using arguments in Section 5. We shall illustrate in the case of Theorem 2.1. First, we introduce necessary notation. Notice that the definition of $\psi(z)$ at (2.5) includes a term whose coefficient is $(1/3)(z^2-1)$. Let $\eta_1(z)$ be the version of $\psi(z)$ in which that coefficient is replaced by $(1/3)(z^2+1)$. Define the constants

$$\xi \equiv -(1/3) \sum_i \sum_j \sum_k \ell_i \ell_j \ell_k \, \alpha_{ijk}, \quad \eta \equiv (\xi^2/2) - \xi \sum_i \sum_j \ell_{ij} \, \alpha_{ij} \, .$$

Let $z_0 \equiv \Phi^{-1}[P\{g(\underset{\sim}{Z}^0|\underset{\sim}{\bar{Y}}) \le 0|\mathcal{Y}\}]$, $\alpha' = \Phi\{2z_0 + z(\alpha)\}$ and $t'_\alpha = t_{\alpha'}$.

THEOREM 3.1 *Under the conditions of Theorem 2.1,*

$$P\{n^{\frac{1}{2}}g(\underset{\sim}{\bar{Y}}^0|\underset{\sim}{\mu}) \le t'_\alpha(\mathcal{Y})\} = \alpha + n^{-\frac{1}{2}}\,\xi\,\phi(z) + n^{-1}\{\eta_1(z) + \eta z\}\,\phi(z) + o(n^{-1})$$

uniformly in $\alpha \in (0,1)$.

As expected, the bias correction introduces a skewness term of order $n^{-\frac{1}{2}}$. The effect of that term vanishes when the bias correction is used to construct a two-sided interval.

4. The iterated bootstrap. In Subsections3.1 and 3.2 we demonstrated that the bootstrap may be viewed as a first-order inversion of an Edgeworth expansion. In Subsections 3.3 and 3.4 we described relatively simple corrections of second order. We shall show now that an arbitrarily high degree of correction may be obtained by iterating the bootstrap argument. Once again, we shall tailor our discussion to the more general context of Theorem 2.1, where $\sigma^2(\underset{\sim}{\mu}^0)$ is not assumed known.

So as to clearly explain the procedure, we shall abbreviate our earlier notation. Distributions, empiric or otherwise, will be represented by their distribution functions. Let G_1 and G_2 be any two distributions on $\mathbb{R}^d$, let $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ be their means, and define

$$h(G_2|G_1) \equiv n^{\frac{1}{2}}\{f(\underset{\sim}{\mu}_2) - f(\underset{\sim}{\mu}_1)\} \,/\, \{\sigma^2(\underset{\sim}{\mu}_2^0)\}^{\frac{1}{2}}\,.$$

Given a distribution F_{i-1}, draw a random n-sample from it and take F_i to be the (empiric) distribution function of that sample. Repeat this for all $i \ge 1$, with F_0 being the (non-random) distribution of $\underset{\sim}{Y}$.

In this notation the random sample $\mathcal{Y}$ introduced in Section 2 has distribution function F_1, and our aim is to determine approximate critical points for the statistic $h(F_1|F_0)$. Those points are calculated as follows.

Given an $(r-1)$'th order approximation $t_\alpha^{(r-1)}(F_1)$, define $t_\alpha^{(r)}(F_1)$ by $t_\alpha^{(r)}(F_1) \equiv t_\alpha^{(r-1)}(F_1) + t_{\alpha,r}(F_1)$, where for any $i \geq 1$,

$$t_{\alpha,r}(F_i) \equiv \inf\{t\colon P[h(F_{i+1}|F_i) \leq t_\alpha^{(r-1)}(F_{i+1}) + t|F_i] \geq \alpha\} .$$

Assuming Cramér's condition (2.3) and appropriate moment conditions on F_0, we have

$$(4.1) \qquad P\{h(F_1|F_0) \leq t_\alpha^{(r)}(F_1)\} = \alpha + n^{-(r+1)/2}\, \psi^{(r)}\{z(\alpha)\}\phi\{z(\alpha)\} + o(n^{-(r+1)/2})$$

uniformly in $\alpha \in (\varepsilon, 1-\varepsilon)$, any $\varepsilon > 0$, where $\psi^{(r)}$ is a polynomial. In the notation of Theorem 2.1, $t_\alpha^{(1)} \equiv t_\alpha$ and $\psi^{(1)} \equiv \psi$.

Calculation of $t_\alpha^{(r)}(F_1)$ requires simulation up to and including the level of F_{r+1}. We shall illustrate methodology in the case $r=3$. Note that $t_\alpha^{(r)}(F_1) = \sum_{j=1}^{r} t_{\alpha,j}(F_1)$; we shall show how to construct $t_{\alpha,1}(F_1)$, $t_{\alpha,2}(F_1)$ and $t_{\alpha,3}(F_1)$. Define $t_{\alpha,1}(F_1) = t_\alpha^{(1)}(F_1)$ using the usual bootstrap argument, i.e. by simulating conditional on F_1. To calculate $t_{\alpha,2}(F_1)$, first compute $t_{\alpha,1}(F_2)$ for each F_2 derivable by sampling F_1, by simulating conditional on F_2:

$$t_{\alpha,1}(F_2) \equiv \inf\{t\colon P[h(F_3|F_2) \leq t|F_2] \geq \alpha\} .$$

Next calculate $t_{\alpha,2}(F_1)$ by simulating conditional on F_1:

$$t_{\alpha,2}(F_1) \equiv \inf\{t\colon P[h(F_2|F_1) \leq t_{\alpha,1}(F_2) + t|F_1] \geq \alpha\} .$$

To derive $t_{\alpha,3}(F_1)$, first calculate $t_{\alpha,1}(F_3)$, then $t_{\alpha,2}(F_2)$ (using $t_{\alpha,1}(F_3)$), and finally $t_{\alpha,3}(F_1)$ (using $t_{\alpha,1}(F_2)$ and $t_{\alpha,2}(F_2)$).

The time taken to construct successive approximations $t_\alpha^{(r)}(F_1)$ increases rapidly and exponentially with r. We have introduced the iterated bootstrap to clarify the role of the ordinary bootstrap as an empiric one-term Edgeworth inversion. It could not reasonably be regarded as a general practical tool for the continuous case which is the subject of this paper.

Result (4.1) may be proved along the lines of Theorem 2.1, and so will not be derived here. Briefly, the argument runs as follows - we illustrate with the case r=2. Let $z=z(\alpha)$ be the solution of $\Phi(z)=\alpha$. A version of Theorem 2.1 in which the pair (F_o,F_1) is replaced by (F_1,F_2), reveals that

$$P\{h(F_2|F_1) \le t_{\alpha,1}(F_2)|F_1\} = \Phi(z) + n^{-1}\,\psi_1(z|F_1)\phi(z) + n^{-3/2}\,\psi_2(z|F_1)\phi(z) + O_p(n^{-2}),$$

where $\psi_i(\cdot|G)$ denotes a polynomial whose coefficients are smooth functions of means with respect to G, and where $\psi_1(\cdot|F_o) \equiv \psi(\cdot)$, defined at (2.5). Consequently, inverting as in steps (ii) and (iii) from Section 5,

$$t_{\alpha,2}(F_1) = -n^{-1}\,\psi_1(z|F_1) + n^{-3/2}\,\psi_3(z|F_1) + O_p(n^{-2}),$$

and arguing as in steps (iv) and (vi) of Section 5,

$$P\{h(F_1|F_o) \le t_{\alpha,1}(F_1) + t_{\alpha,2}(F_1)\}$$

$$= P\{h(F_1|F_o) \le t_{\alpha,1}(F_1) - n^{-1}\,\psi_1(z|F_1)\} + n^{-3/2}\,\psi_4(z|F_o)\phi(z) + O(n^{-2})$$

$$= P\{h(F_1|F_o) \leq t_{\alpha,1}(F_1) - n^{-1}\,\psi_1(z|F_o)\} + n^{-3/2}\,\psi_5(z|F_o)\phi(z) + O(n^{-2})$$

$$= P\{h(F_1|F_o) \leq t_{\alpha,1}(F_1) - n^{-1}\,\psi_1(z|F_o)\} + n^{-3/2}\,\psi_6(z|F_o)\phi(z) + O(n^{-2})$$

$$= \{\alpha + n^{-1}\,\psi_1(z|F_o) + n^{-3/2}\,\psi_7(z|F_o)\} + n^{-3/2}\,\psi_6(z|F_o)\phi(z) + O(n^{-2})$$

$$= \alpha + n^{-3/2}\{\psi_6(z|F_o) + \psi_7(z|F_o)\}\phi(z) + O(n^{-2})\ .$$

The general proof uses induction over r.

<u>5. Proof of Theorem 2.1</u>. It is inconvenient to continue using the superscript on vectors $\overline{\underset{\sim}{Y}}^{o}$, $\underset{\sim}{\mu}^{o}$, etc. We shall drop it below, writing $g(\overline{\underset{\sim}{Y}}|\underset{\sim}{\mu})$ instead of $g(\overline{\underset{\sim}{Y}}^{o}|\underset{\sim}{\mu})$, and so on. This amounts to assuming that the vectors of interest are already sufficiently long, so that further extension is unnecessary. Write $\underset{\sim}{V}_o = (v_{ij}) \equiv \text{var}(\underset{\sim}{Y})$ $(\text{var}(\underset{\sim}{Y}^{o})$ in the old notation).

We begin by introducing polynomials π_{ij} and Π_{ij}, for i,j=1,2. Bhattacharya and Ghosh [5, Theorem 2] provide explicit conditions under which

$$P\{n^{\frac{1}{2}}g(\overline{\underset{\sim}{Y}}|\underset{\sim}{\mu}) \leq x\} = \Phi(x) + \sum_{j=1}^{2} n^{-j/2}\,\pi_{1j}(x)\phi(x) + o(n^{-1})$$

uniformly in x, where π_{1j} is a polynomial of degree 3j-1 whose coefficients depend on the first j+2 moments of $\underset{\sim}{Y}$ and the first j+1 derivatives of $g(\underset{\sim}{u}|\underset{\sim}{\mu})$ at $\underset{\sim}{u}=\underset{\sim}{\mu}$. Applied to the bootstrap distribution, this would give us formally

$$P\{n^{\frac{1}{2}}g(\overline{\underset{\sim}{Z}}|\overline{\underset{\sim}{Y}}) \leq x|\mathcal{Y}\} = \Phi(x) + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{1j}(x)\phi(x) + o(n^{-1})\ .$$

Notice that only derivatives of order three or less are involved; hence the condition on derivatives of order three or less in the paragraph preceding Theorem 2.1.

The polynomials π_{2j} are defined to be those polynomials which are such that, for each $y \in \mathbb{R}$, the quantity $x = x(y) = y + \sum_{j=1}^{2} n^{-j/2} \pi_{2j}(y)$ satisfies $\Phi(x) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(x)\phi(x) = \Phi(y) + O(n^{-3/2})$ as $n \to \infty$. The coefficients of the π_{2j}'s are of course simple functions of the coefficients of the π_{1j}'s; in fact, $\pi_{21} = -\pi_{11}$. The random-coefficient polynomials Π_{21} and Π_{22} are defined analogously. Fortunately we do not yet require explicit expressions for any of these polynomaials.

By way of notation, define $V_{ij} \equiv n^{-1} \sum_{r=1}^{n} (Y_{ri} - \bar{Y}_i)(Y_{rj} - \bar{Y}_j)$ and let $\underset{\sim}{V} = (V_{ij})$ be the usual estimate of $\underset{\sim}{V}_0$. As $n \to \infty$, $\underset{\sim}{V} \to \underset{\sim}{V}_0$ in probability. Given a nonnegative integer-valued vector $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_d)^T$, and a smooth function f of d variables, let

$$D^{\underset{\sim}{\alpha}} f(\underset{\sim}{x}) \equiv (\partial/\partial x_1)^{\alpha_1} \ldots (\partial/\partial x_d)^{\alpha_d} f(\underset{\sim}{x}) .$$

Our proof of Theorem 2.1 contains six steps. Let $C_1, C_2, \ldots$ denote positive constants not depending on n, and $\mathcal{C}$ the class of all samples $\mathcal{Y}$. Write R for the probability measure on $\mathbb{R}$ generated by $n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}} | \underset{\sim}{\bar{Y}})$, conditional on $\mathcal{Y}$.

Step (i). Here we expand the conditional distribution of $n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}} | \underset{\sim}{\bar{Y}})$, and prove:

PROPOSITION 5.1. *Under the conditions of Theorem 2.1, there exist* $\mathcal{C}_n^{(1)} \subseteq \mathcal{C}$ *with* $P(\tilde{\mathcal{C}}_n^{(1)}) = o(n^{-1})$, *and constants* $C_1 > 0$ *and* $\varepsilon_1 \in (0, \frac{1}{2})$, *such that the random variable*

$$\Delta_n(x) \equiv \int_{(-\infty, x]} d(R - \Phi - \sum_{j=1}^{2} n^{-j/2} \Pi_{ij}\, \phi)$$

satisfies

$$\sup_{\mathcal{Y} \in \mathcal{C}_n^{(1)}} \sup_{-\infty < x < \infty} |\Delta_n(x)| \leq C_1 n^{-1-\varepsilon_1} , \tag{5.1}$$

and such that

$$\sup_{\mathcal{Y}\in C_n^{(1)}} |\Pi_{ij}(x)| \le C_1(1+|x|^{3j-1}) \tag{5.2}$$

for $-\infty<x<\infty$ *and* $j=1,2$.

PROOF. We begin by defining several classes C_{nj} of samples $\mathcal{Y}$.

The random coefficients of polynomials Π_{11} and Π_{12} are continuous functions of moments of $\mathcal{Y}$ of order 4 or less. Under the condition $E(\|\underset{\sim}{Y}\|^8)<\infty$, each such sample moment satisfies $P(|M-EM|>\eta)=o(n^{-1})$ for each $\eta>0$. Therefore we may choose $C_2>0$ and $C_{n1}\subseteq C$ such that $P(\tilde{C}_{n1})=o(n^{-1})$ and the absolute value of each coefficient of Π_{11} and Π_{12} is dominated by C_2 whenever $\mathcal{Y}\in C_{n1}$. By choosing $C_n^{(1)}\subseteq C_{n1}$ we may ensure that (5.2) holds. For any $p>0$,

$$E(\|\underset{\sim}{Z}-\underset{\sim}{\bar{Y}}\|^p|\mathcal{Y}) = n^{-1}\sum_{r=1}^{n}\{\sum_{j=1}^{d}(Y_{rj}-\bar{Y}_j)^2\}^{p/2} \le (2d)^p\, n^{-1}\sum_{r=1}^{n}\sum_{j=1}^{d}(|Y_{rj}|^p+|\bar{Y}_j|^p),$$

and so

$$P\{E(\|\underset{\sim}{Z}-\underset{\sim}{\bar{Y}}\|^p|\mathcal{Y}) \ge (2d)^p\sum_{j=1}^{d}(E|Y_{1j}|^p+|E\bar{Y}_j|^p)+(2d)^{p+1}\}$$

$$\le P\{\sum_{j=1}^{d}|\sum_{r=1}^{n}(|Y_{rj}|^p-E|Y_{rj}|^p)| + n\sum_{j=1}^{d}||\bar{Y}_j|^p-|E\bar{Y}_j|^p| \ge 2dn\}$$

$$\le \sum_{j=1}^{d} n^{-(8+\varepsilon)/p}\, E\{|\sum_{r=1}^{n}(|Y_{rj}|^p-E|Y_{rj}|^p)|^{(8+\varepsilon)/p}\} + \sum_{j=1}^{d}E|\bar{Y}_j-E\bar{Y}_j|^{8+\varepsilon}$$

$$= O(n^{-(8+\varepsilon)/2p}),$$

provided $1<p<(8+\varepsilon)/2$. Therefore we may choose $\varepsilon_2\in(0,\tfrac{1}{2})$ and $C_3>0$ such that the set $C_{n2}\equiv\{E(\|\underset{\sim}{Z}-\underset{\sim}{\bar{Y}}\|^{4+2\varepsilon_2}|\mathcal{Y})\le C_3\}$ has $P(\tilde{C}_{n2})=o(n^{-1})$.

Note that $\underset{\sim}{V}=(V_{ij})\to\underset{\sim}{V}_o=(v_{ij})$ in probability as $n\to\infty$, and that $\underset{\sim}{V}_o$ is positive definite. Suppose all eigenvectors of $\underset{\sim}{V}_o$ lie within [2a,1/2a], where $a > 0$. Choose $C > 0$ so small that any symmetric $d \times d$ matrix (u_{ij}) satisfying $|u_{ij}-v_{ij}| \leq b$ for all i and j has all its eigenvalues within [a,1/a]. Let C_{n3} be the class of all samples $\mathcal{Y}$ such that $|V_{ij}-v_{ij}| \leq b$ for all i and j. If $\mathcal{Y} \in C_{n3}$ then $C_4^{-1}\|\underset{\sim}{t}\| \leq \|\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{t}\| \leq C_4\|\underset{\sim}{t}\|$ for all $\underset{\sim}{t} \in \mathbb{R}^d$, where C_4 depends only on $\underset{\sim}{V}_o$ and b. The inequality

$$|V_{ij}-v_{ij}| \leq |n^{-1} \sum_{r=1}^{n} \{Y_{ri}Y_{rj} - E(Y_{ri}Y_{rj})\}| + |\overline{Y}_i||\overline{Y}_j - E(\overline{Y}_j)|$$

$$+ |E(\overline{Y}_j)||\overline{Y}_i - E(\overline{Y}_i)|$$

may be used to prove that $P(\tilde{C}_{n3}) = o(n^{-1})$.

Let

$$\underset{\sim}{Z}_r^* \equiv \begin{cases} \underset{\sim}{V}^{-\frac{1}{2}}(\underset{\sim}{Z}_r - \underset{\sim}{\overline{Y}}) & \text{if } \|\underset{\sim}{V}^{-\frac{1}{2}}(\underset{\sim}{Z}_r - \underset{\sim}{\overline{Y}})\| \leq n^{1/2} \\ \underset{\sim}{0} & \text{otherwise} \end{cases}$$

and $\underset{\sim}{Z}_r^{\dagger} \equiv \underset{\sim}{Z}_r^* - E(\underset{\sim}{Z}_r^*|\mathcal{Y})$. Define Q and $Q^{\dagger}$ to be the probability measures on $\mathbb{R}^d$ generated by $n^{-\frac{1}{2}} \sum_{r=1}^{n} \underset{\sim}{Z}_r$ and $n^{-\frac{1}{2}} \sum_{r=1}^{n} \underset{\sim}{Z}_r^{\dagger}$, respectively, both conditional on $\mathcal{Y}$. Let $\underset{\sim}{V}^{\dagger} \equiv (V_{ij}^{\dagger})$ be the variance matrix and $\{\chi_\nu^{\dagger}\}$ the cumulant sequence of $\underset{\sim}{Z}_1^{\dagger}$, again conditional on $\mathcal{Y}$. Observe that $\underset{\sim}{V}^{\dagger} \to \underset{\sim}{I} \equiv (\delta_{ij})$ in probability as $n \to \infty$. Choose $c > 0$ so small that any symmetric $d \times d$ matrix (u_{ij}) satisfying $|u_{ij} - \delta_{ij}| \leq c$ for all i and j has all its eigenvalues within [1/2,2]. Let C_{n4} be the class of samples $\mathcal{Y}$ such that $|V_{ij}^{\dagger} - \delta_{ij}| \leq c$ for all i and j, and define $\underset{\sim}{W} \equiv \underset{\sim}{V}^{-\frac{1}{2}}(\underset{\sim}{Z}-\underset{\sim}{\overline{Y}})$. The identity

$$V_{ij}^{\dagger} = \delta_{ij} - E\{(\underset{\sim}{W})_i(\underset{\sim}{W})_j \ I(\|\underset{\sim}{W}\| > n^{\frac{1}{2}})|\mathcal{Y}\}$$

$$- E\{(\underset{\sim}{W})_i \ I(\|\underset{\sim}{W}\| > n^{\frac{1}{2}})|\mathcal{Y}\} \ E\{(\underset{\sim}{W})_j \ I(\|\underset{\sim}{W}\| > n^{\frac{1}{2}})|\mathcal{Y}\},$$

together with Markov's inequality, may be used to prove that $P(C_{n3} \cap \tilde{C}_{n4}) = o(n^{-1})$. For example, for any $\eta > 0$,

$$P[C_{n3};\ |E\{(\underset{\sim}{W})_i (\underset{\sim}{W})_j\ I(\|\underset{\sim}{W}\| > n^{\frac{1}{2}}) | \mathcal{Y}\}| > \eta]$$

$$\leq P[C_4^2\ E\{\|\underset{\sim}{Z} - \underset{\sim}{\bar{Y}}\|^2\ I(C_4 \|\underset{\sim}{Z} - \underset{\sim}{\bar{Y}}\| > n^{\frac{1}{2}}) | \mathcal{Y}\} > \eta]$$

$$\leq C_4^2\ \eta^{-1} (C_4^{-1}\ n^{\frac{1}{2}})^{-3}\ E(\|\underset{\sim}{Z} - \underset{\sim}{\bar{Y}}\|^5) = O(n^{-3/2}).$$

Define $C_{n5} \equiv C_{n1} \cap C_{n2} \cap C_{n3} \cap C_{n4}$. Then $P(\tilde{C}_{n5}) = o(n^{-1})$. Throughout the remainder of Step (i) we shall consider only samples $\mathcal{Y} \in C_{n5}$.

Let $A \subseteq \mathbb{R}^d$, random but measurable in the σ-field generated by $\mathcal{Y}$. Define $A^\dagger \equiv \underset{\sim}{V}^{-\frac{1}{2}} A + n^{\frac{1}{2}} E(\underset{\sim}{Z}_1^* | \underset{\sim}{Y})$ and, in notation of [6, pp. 53-54],

$$H \equiv Q - \sum_{r=1}^{d+2} n^{-r/2} P_r(-\Phi_{\underset{\sim}{0},\underset{\sim}{V}^\dagger} : \{\chi_\nu^\dagger\}).$$

Write $\{\chi_\nu\}$ for the cumulant sequence of $\underset{\sim}{Z}$, conditional on $\mathcal{Y}$. The argument preceding (20.15) of [6, p. 209] yields:

$$(5.3)\qquad |\int_A d[Q - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{\underset{\sim}{0},\underset{\sim}{V}} : \{\chi_\nu\})]$$

$$- \int_{A^\dagger} d[Q^\dagger - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{\underset{\sim}{0},\underset{\sim}{V}^\dagger} : \{\chi_\nu^\dagger\})]|$$

$$\leq C_5\ n^{-1}\ E\{\|\underset{\sim}{W}\|^4\ I(\|\underset{\sim}{W}\| > n^{\frac{1}{2}}) | \mathcal{Y}\} \leq C_5\ n^{-1}\ C_4^4 (C_4^{-1} n^{\frac{1}{2}})^{-2\varepsilon_2}\ C_3,$$

the last inequality following since $\mathcal{Y} \in C_{n2} \cap C_{n3}$. Using (9.12) and (14.74) of [6, pp. 72, 133], we obtain:

$$(5.4)\qquad \left|\int_{A^\dagger} d[Q^\dagger - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{0,V^\dagger}: \{\chi_\nu^\dagger\}) - H]\right|$$

$$\leq C_6 \sum_{r=3}^{d+2} n^{-r/2} E(\|Z_1^*\|^{r+2}|\mathcal{Y}) \leq C_7\, n^{-1-\varepsilon_2},$$

again since $\mathcal{Y} \in C_{n2} \cap C_{n3}$. Combining (5.3) and (5.4), we obtain on C_{n5}:

$$(5.5)\qquad \left|\int_A d[Q - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{0,V}: \{\chi_\nu\})] - \int_{A^\dagger} dH\right| \leq C_8\, n^{-1-\varepsilon_2}.$$

Let K be a probability measure on $\mathbb{R}^d$ with support confined to $B(0,1)$ and satisfying $|(D^\alpha \hat{K})(t)| \leq C_9 \exp(-\|t\|^{\frac{1}{2}})$, all $|\alpha| \leq d+5$ and $t \in \mathbb{R}^d$. (The hat denotes Fourier-Stieltjes transform.) Set $K_\delta(E) \equiv K(\delta^{-1}E)$ for $0 < \delta < 1$. Arguing as in [6, p. 210] we obtain:

$$(5.6)\qquad \left|\int_{A^\dagger} dH\right| \leq C_{10} \sup_{0\leq\alpha\leq\beta, |\beta|\leq d+5} \int |(D^{\beta-\alpha}\hat{H})(t)(D^\alpha \hat{K}_\delta)(t)|\,dt$$

$$+ \int I\{A^\dagger \Delta B(x,2\delta) \neq \emptyset\} \left|\sum_{r=0}^{d+2} n^{-r/2} P_r(-\Phi_{0,V^\dagger} : \{\chi_\nu^\dagger\})\right| dx,$$

where Δ denotes symmetric difference. The constants C_j here and below do not depend on δ. Following the argument of [6, pp. 210-211], with a little modification, we derive for a certain random variable A_n:

$$(5.7)\qquad \sup_{0\leq\alpha\leq\beta, |\beta|\leq d+5} \left|\int_{\{\|t\| \leq A_n\}} |(D^{\beta-\alpha}\hat{H})(t)(D^\alpha \hat{K}_\delta)(t)|\,dt \leq C_{11}\, n^{-1-\varepsilon_2},$$

$$(5.8)\qquad \sup_{0\leq\alpha\leq\beta, |\beta|\leq d+5} \left|\int_{\{\|t\| > A_n\}} |(D^{\beta-\alpha}\hat{H})(t)(D^\alpha \hat{K}_\delta)(t)|\,dt \leq \kappa + C_{11}\, n^{-1-\varepsilon_2},$$

where $\kappa \equiv \sup_{0\leq\alpha\leq\beta, |\beta|\leq d+5} \int_{\{\|t\| > c_n\}} |(D^{\beta-\alpha}\hat{Q}^\dagger)(t)(D^\alpha \hat{K}_\delta)(t)|\,dt$ and $c_n = n^{\frac{1}{2}}/\{16\, E(\|W\|^3|\mathcal{Y})\}$.

Let $\psi_n(\underset{\sim}{t}) \equiv n^{-1} \sum_{r=1}^{n} \exp(i< \underset{\sim}{t}, \underset{\sim}{Y}_r >)$ denote the (empiric) characteristic function of $\mathcal{Y}$, and $\psi(\underset{\sim}{t}) \equiv E\{\psi_n(\underset{\sim}{t})\}$ the characteristic function of $\underset{\sim}{Y}$. We shall need:

LEMMA 5.1. *There exist constants* $C_{12} \in (0,1)$, C_{13} *and* C_{14} *such that whenever* $n > d+5$, $m \in [1,n-d-5]$, $0 < \delta < 1$ *and* $\mathcal{Y} \in \mathcal{C}_{n5}$,

$$\kappa \leq C_{13}\, n^{d+5} \int [C_{12}^{n} + \{C_{14} |\psi_n(n^{-\frac{1}{2}}\underset{\sim}{t}) - \psi(n^{-\frac{1}{2}}\underset{\sim}{t})|\}^{m}] \exp(-C_4^{-\frac{1}{2}} \|\delta \underset{\sim}{t}\|^{\frac{1}{2}}) d\underset{\sim}{t}\,.$$

PROOF. Set $q(\underset{\sim}{t}) \equiv E\{\exp(n^{-\frac{1}{2}}\, i < \underset{\sim}{t}, \underset{\sim}{Z}_1^{\dagger} >) | \mathcal{Y}\}$. Then $\hat{Q}(\underset{\sim}{t}) = q^{n}(\underset{\sim}{t})$, from which it may be deduced that $|(D^{\underset{\sim}{\gamma}}\hat{Q}^{\dagger})(\underset{\sim}{t})| \leq (2n)^{|\gamma|} |q(\underset{\sim}{t})|^{n-|\underset{\sim}{\gamma}|}$. (Note that $|\underset{\sim}{Z}_1^{\dagger}| \leq 2n^{\frac{1}{2}}$.) Therefore if $n > d+5$,

$$(5.9)\qquad \kappa \leq C_9\, (2n)^{d+5} \int_{\{\|\underset{\sim}{t}\| > c_n\}} |q(\underset{\sim}{t})|^{n-d-5} \exp(-\|\delta \underset{\sim}{t}\|^{\frac{1}{2}}) d\underset{\sim}{t}\,.$$

If $\mathcal{Y} \in \mathcal{C}_{n5}$ then $\|\underset{\sim}{t}\| > c_n$ implies that $16\, C_4 E(\|\underset{\sim}{W}\|^{3} | \mathcal{Y}) \|\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{t}\| > n^{\frac{1}{2}}$, and hence that $\|\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{t}\| > C_{15}\, n^{\frac{1}{2}}$. Now,

$$(5.10)\qquad |q(\underset{\sim}{t})| \leq |E\{\exp(n^{-\frac{1}{2}}\, i < \underset{\sim}{t}, \underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{Z} >) | \mathcal{Y}\}| + P(\|\underset{\sim}{W}\| > n^{\frac{1}{2}} | \mathcal{Y})$$
$$\leq |\psi_n(n^{-\frac{1}{2}}\, \underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{t})| + C_{16}\, n^{-1}\,.$$

Results (5.9) and (5.10) together give:

$$(5.11)\qquad \kappa \leq C_{17}\, n^{d+5} \int_{\{\|n^{-\frac{1}{2}}\underset{\sim}{t}\| > C_{15}\}} \{|\psi_n(n^{-\frac{1}{2}}\underset{\sim}{t})| + C_{16}\, n^{-1}\}^{n-d-5}$$
$$\times \exp(-C_4^{-\frac{1}{2}} \|\delta \underset{\sim}{t}\|^{\frac{1}{2}}) d\underset{\sim}{t}\,.$$

Choose $\eta_1 \in [\frac{1}{2},1)$ and $\eta_2 > 0$ such that $\sup_{\|\underset{\sim}{u}\| > \xi} |\psi(\underset{\sim}{u})| \leq \max(\eta_1, 1-\eta_2\, \xi^2)$ for all $\xi > 0$. Let $\eta_3 = \eta_3(\xi) \in [\sup_{\|\underset{\sim}{u}\| > \xi} |\psi(\underset{\sim}{u})|, 1)$. If

$|\psi_n(\underset{\sim}{u})| > \eta_3$ then $C_{16}\, n^{-1} \le C_{16}\, n^{-1}\, \eta_3^{-1}|\psi_n(\underset{\sim}{u})|$, and

$$\{|\psi_n(\underset{\sim}{u})| + C_{16}\, n^{-1}\}^{n-d-5} \le \exp(C_{16}/\eta_3)|\psi_n(\underset{\sim}{u})|^{n-d-5} .$$

Choose η_4 so small that $\eta_3\,(1+\eta_4) < 1$. If $|\psi_n(\underset{\sim}{u}) - \psi(\underset{\sim}{u})| \le \eta_4|\psi(\underset{\sim}{u})|$ then $|\psi_n(\underset{\sim}{u})| \le \eta_3(1+\eta_4)$ for $\|\underset{\sim}{u}\| > \xi$. If $|\psi_n(\underset{\sim}{u}) - \psi(\underset{\sim}{u})| > \eta_4|\psi(\underset{\sim}{u})|$ then $|\psi_n(\underset{\sim}{u})|^{n-d-5} \le \{(1+\eta_4^{-1})|\psi_n(\underset{\sim}{u}) - \psi(\underset{\sim}{u})|\}^m$. Finally, if $|\psi(\underset{\sim}{u})| \le \eta_3$ then $\{|\psi_n(\underset{\sim}{u})| + C_{16}\, n^{-1}\}^{n-d-5} \le \exp(C_{16}/\eta_3)\, \eta_3^{n-d-5}$. Combining these results we conclude that for any $\|\underset{\sim}{u}\| > \xi$,

$$\{|\psi_n(\underset{\sim}{u})| + C_{16}\, n^{-1}\}^{n-d-5} \le \exp(C_{16}/\eta_3)\; [\{\eta_3(1+\eta_4)\}^{n-d-5} + \{(1+\eta_4^{-1})|\psi_n(\underset{\sim}{u}) - \psi(\underset{\sim}{u})|\}^m] .$$

Lemma 5.1 follows from this inequality and (5.11), on taking $\xi = C_{15}$, $\eta_3 = \max\{\tfrac{1}{2}, \sup_{\|\underset{\sim}{u}\|>\xi} |\psi(\underset{\sim}{u})|\}$, $\eta_4 = \tfrac{1}{4}$ if $\eta_3 = \tfrac{1}{2}$,

$$\eta_4 = (1/2)[\{\max(\eta_1, 1-\eta_2\, \xi^2)\}^{-1}-1]$$

if $\eta_3 \ne \tfrac{1}{2}$, and $C_{12} = \max\{(1/2)(1+\eta_1),\ 1-(1/2)\eta_2\, \xi^2\}$. □

Choose $\delta = n^{-\lambda}$, where $\lambda > 0$ will be selected shortly. Since

$$E\{|\psi_n(\underset{\sim}{u}) - \psi(\underset{\sim}{u})|^m\} \le C_{17}(m)\; n^{-m/2}$$

uniformly in $\underset{\sim}{u}$, then by Markov's inequality, the set

$$C_{n6} \equiv \{\mathcal{Y}: \int|\psi_n(n^{-\frac{1}{2}+\lambda}\underset{\sim}{t}) - \psi(n^{-\frac{1}{2}+\lambda}\underset{\sim}{t})|^m \exp(-C_4^{-\frac{1}{2}}\, \|\underset{\sim}{t}\|^{\frac{1}{2}})d\underset{\sim}{t} \le n^{-d(\lambda+1)-7}\}$$

satisfies $P(\tilde{C}_{n6}) = O(n^{d(\lambda+1)+7} n^{-m/2}) = o(n^{-1})$, provided m is chosen sufficiently large. Henceforth we shall work only with samples $\mathcal{Y} \in C_{n7} \equiv C_{n5} \cap C_{n6}$. In that case, Lemma 5.1 implies

$$\kappa \leq C_{18}\, n^{d(\lambda+1)+5}\, C_{12}^{n} + C_{13}\, C_{14}^{m}\, n^{-2} \leq C_{19}\, n^{-2},$$

and so by (5.6), (5.7) and (5.8),

$$(5.12) \qquad \left|\int_{A^{\dagger}} dH\right| \leq C_{20}\, n^{-1-\varepsilon_2} + \xi,$$

where $\xi \equiv \sum_{r=0}^{\ell+2} \xi_r$,

$$\xi_r \equiv n^{-r/2} \int I(\underset{\sim}{x}) \left|P_r(-\Phi_{\underset{\sim}{0},\underset{\sim}{V}^{\dagger}} : \{\chi_{\nu}^{\dagger}\})\right| d\underset{\sim}{x}$$

and $I(\underset{\sim}{x}) \equiv I\{A^{\dagger} \equiv B(\underset{\sim}{x},\, 2n^{-\lambda}) \neq \emptyset\}$.

An argument using (9.12) and (14.74) of [6, pp. 72 and 133] gives $\xi_r \leq C_{21}\, \zeta_r$, where

$$\zeta_r \equiv \int I(\underset{\sim}{x})\, (1+\|\underset{\sim}{x}\|^{3r})\, \exp\{-(1/2)\, \underset{\sim}{x}^{T}\, \underset{\sim}{V}^{\dagger-1}\, \underset{\sim}{x}\} d\underset{\sim}{x}.$$

Let $\underset{\sim}{a} \equiv n^{\frac{1}{2}}\, \underset{\sim}{V}^{\frac{1}{2}}\, E(\underset{\sim}{Z}_1^{*}|\mathcal{Y})$, and note that for $\mathcal{Y} \in C_{n7}$,

$$B(\underset{\sim}{x}) \equiv \underset{\sim}{V}^{\frac{1}{2}}\, B(\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{x},\, 2n^{-\lambda}) \subseteq B(\underset{\sim}{x},\, C_{22}\, n^{-\lambda}).$$

Thus,

$$\zeta_r = (\det \underset{\sim}{V}^{-\frac{1}{2}}) \int I(\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{x})(1+\|\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{x}\|^{3r}) \quad \exp\{-(1/2)\underset{\sim}{x}^{T}\, \underset{\sim}{V}^{-\frac{1}{2}}\, \underset{\sim}{V}^{\dagger-1}\underset{\sim}{V}^{-\frac{1}{2}}\underset{\sim}{x}\} d\underset{\sim}{x}$$

$$\leq C_{23} \int I\{(\underset{\sim}{A}+\underset{\sim}{a}) \Delta B(\underset{\sim}{x}) \neq \emptyset\}\, \exp(-D_{24}\|\underset{\sim}{x}\|^2) d\underset{\sim}{x}$$

$$\leq C_{23} \int_{(\delta A)^{\eta}} \exp(-C_{24}\|\underset{\sim}{x}+\underset{\sim}{a}\|^2) d\underset{\sim}{x},$$

where $\eta \equiv C_{22}\, n^{-\lambda}$. An elementary argument shows that

$$\exp(-\|\underset{\sim}{x}+\underset{\sim}{y}\|^2) \leq \exp(-\|\underset{\sim}{x}\|^2) + C_{25}\|\underset{\sim}{y}\| \exp(-\|\underset{\sim}{x}+\underset{\sim}{y}\|^2/2)$$

for all $\underset{\sim}{x}$, $\underset{\sim}{y} \in \mathbb{R}^d$. If $\mathcal{Y} \in C_{n7}$ then $\|\underset{\sim}{a}\| \leq C_{26}\, n^{-3/2}$, using Markov's inequality. Combining these estimates we conclude that

$$\zeta_r \leq C_{23} \int_{(\delta A)^\eta} \exp(-C_{24}\|\underset{\sim}{x}\|^2)d\underset{\sim}{x} + C_{27}\, n^{-3/2},$$

and so by (4.12),

$$\text{(5.13)} \qquad |\int_{A^\dagger} dH| \leq C_{28}\, n^{-1-\varepsilon_1} + C_{23} \int_{(\delta A)^\eta} \exp(-C_{24}\|\underset{\sim}{x}\|^2)d\underset{\sim}{x}\,.$$

We now restrict attention to sets $A \in \{A(t)\colon -\infty<t<\infty\}$, where

$$A(t) \equiv \{\underset{\sim}{x} \in \mathbb{R}^d\colon\ n^{\frac{1}{2}} g(\underset{\sim}{\bar{Y}}+n^{-\frac{1}{2}}\underset{\sim}{x}\,|\,\underset{\sim}{\bar{Y}}) \leq t\}\,.$$

LEMMA 5.2. *Take* $\lambda = 1+\varepsilon_2$ *and* $\eta = C_{22}\, n^{-\lambda}$. *There exists* $C_{n8} \subseteq C_{n7}$, *with* $P(\tilde{C}_{n8}) = o(n^{-1})$, *such that*

$$\sup_{\mathcal{Y} \in C_{n8}}\ \sup_{-\infty<t<\infty} \int_{\{\delta A(t)\}^\eta} \exp(-C_{24}\|\underset{\sim}{x}\|^2)d\underset{\sim}{x} \leq C_{29}\, n^{-1-\varepsilon_2}\,.$$

PROOF. Suppose $g(\underset{\sim}{u}|\underset{\sim}{v})$ and its first four derivatives with respect to $\underset{\sim}{u}$ are continuous in $(\underset{\sim}{u},\underset{\sim}{v}) \in B(\underset{\sim}{\mu},\rho_1) \times B(\underset{\sim}{\mu},\rho_1)$. In view of property (2.4), not all terms $g_i(\underset{\sim}{\mu}|\underset{\sim}{\mu})$ can vanish. Let $\rho_2 \in (0,\rho_1/2]$ be such that for some i, $g_i(\underset{\sim}{u}|\underset{\sim}{v})$ is bounded away from zero in $B(\underset{\sim}{\mu},\rho_2) \times B(\underset{\sim}{\mu},\rho_2)$. Let C_{n9} be the class of all samples $\mathcal{Y}$ for which $\|\underset{\sim}{\bar{Y}}-\underset{\sim}{\mu}\| \leq \rho_2/2$. The set $C_{n8} = C_{n7} \cap C_{n9}$ satisfies $P(\tilde{C}_{n8}) = o(n^{-1})$, and has the property that for some $C_{29} > 0$ and some $i_o \in \{1,\ldots,d\}$,

$$\sup_{\mathcal{Y} \in C_{n8}}\ \sup_{i,j,k} \{|g_i(\underset{\sim}{\bar{Y}}|\underset{\sim}{\bar{Y}})|,\ |g_{ij}(\underset{\sim}{\bar{Y}}|\underset{\sim}{\bar{Y}})|,\ g_{ijk}(\underset{\sim}{\bar{Y}}|\underset{\sim}{\bar{Y}})|\} \leq C_{29}$$

and $\inf_{\mathcal{Y} \in C_{n8}} |g_{i_o}(\underset{\sim}{\bar{Y}}|\underset{\sim}{\bar{Y}})| > 1/C_{29}$. We assume throughout the argument below that $\mathcal{Y} \in C_{n8}$.

For any vector $\underset{\sim}{x} = (x_1,\dots,x_d)^T \in \mathbb{R}^d$, define the random-coefficient cubic polynomial

$$p_n(\underset{\sim}{x}) \equiv \sum_{i=1}^{d} x_i g_i(\underset{\sim}{\overline{Y}}|\underset{\sim}{\overline{Y}}) + (1/2)\, n^{-\frac{1}{2}} \sum_{i=1}^{d}\sum_{j=1}^{d} x_i x_j g_{ij}(\underset{\sim}{\overline{Y}}|\underset{\sim}{\overline{Y}})$$

$$+ (1/6)\, n^{-1} \sum_{i=1}^{d}\sum_{i=1}^{d}\sum_{i=1}^{d} x_i x_j x_k g_{ijk}(\underset{\sim}{\overline{Y}}|\underset{\sim}{\overline{Y}}) .$$

Then $n^{\frac{1}{2}}\, g(\underset{\sim}{\overline{Y}} + n^{-\frac{1}{2}}\underset{\sim}{x}|\underset{\sim}{\overline{Y}}) = p_n(\underset{\sim}{x}) + \Delta_{n1}(\underset{\sim}{x})$, where

$$\sup_{\underset{\sim}{x} \in B(\underset{\sim}{0},\log n)} |\Delta_{n1}(\underset{\sim}{x})| \le C_{30}\, n^{-3/2} (\log n)^4 .$$

Therefore, remembering that $\eta = C_{22} n^{-\lambda}$ and $\lambda = 1+\varepsilon_2$:

$$\{\delta A(t)\}^{\eta} \subseteq \{\underset{\sim}{x}_1 \in \mathbb{R}^d\colon \text{ for some } \underset{\sim}{x}_2 \in \mathbb{R}^d,\ |\underset{\sim}{x}_1 - \underset{\sim}{x}_2| \le \eta \text{ and}$$

$$n^{\frac{1}{2}}\, g(\underset{\sim}{\overline{Y}} + n^{-\frac{1}{2}}\, \underset{\sim}{x}_2|\underset{\sim}{\overline{Y}}) = t\}$$

$$\subseteq \{A_1(t)\} \cup B(0,\ \log n)^{\sim} ,$$

where $A_1(t) \equiv \{\underset{\sim}{x}\colon \|\underset{\sim}{x}\| \le 2 \log n \text{ and } p_n(\underset{\sim}{x}) \in [t - C_{31}\, n^{-1-\varepsilon_2}, t + C_{31}\, n^{-1-\hat{\varepsilon}_2}]\}$.

The lemma follows easily from this result and the statement within quotation marks below:

"Let $\underset{\sim}{X}$ have the $N(\underset{\sim}{0},\underset{\sim}{I})$ distribution on $\mathbb{R}^d$, let β, $b > 0$, and let

$$r_n(\underset{\sim}{x}) = \sum_{i=1}^{d} c_i x_i + n^{-\frac{1}{2}} \sum_{i=1}^{d}\sum_{j=1}^{d} c_{ij} x_i x_j + n^{-1} \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} c_{ijk} x_i x_j x_k$$

be a cubic polynomial whose symmetric coefficients c_i, c_{ij} and c_{ijk} satisfy

$$\sup_{i,j,k} (|c_i|, |c_{ij}|, |c_{ijk}|) \le b, \text{ and for some } i,\ |c_i| > 1/b .$$

There exists $C > 0$, depending only on β, b and d, such that

$$\sup_{-\infty<t<\infty} P\{\|\underset{\sim}{X}\| \leq b \log n \text{ and } r_n(\underset{\sim}{X}) \in [t-n^{-\beta},\ t+n^{-\beta}]\} \leq C(\beta,b,d)n^{-\beta}$$

for all $n \geq 1$."

Our proof of this statement is by induction over d. The statement is obviously true if d=1. Suppose it is true for d-1, some $d \geq 2$. Without loss of generality, $|c_d|$ is the *smallest* $|c_i|$. Let $\underset{\sim}{X}^* = (X_1,\ldots,X_{d-1})^T$, $\underset{\sim}{x}^* = (x_1,\ldots,x_{d-1})^T$ and

$$\begin{aligned} r_n^*(\underset{\sim}{x}^*|x_d) \equiv & \sum_{i=1}^{d-1} (c_i + 2\, n^{-\frac{1}{2}}\, c_{id}\, x_d + 3\, n^{-1}\, c_{idd}\, x_d^2)\, x_i \\ & + n^{-\frac{1}{2}} \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} (c_{ij} + 3\, n^{-\frac{1}{2}}\, c_{ijd}\, x_d) x_i x_j \\ & + n^{-\frac{1}{2}} \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} \sum_{k=1}^{d-1} c_{ijk}\, x_i x_j x_k \\ = & \sum_{i=1}^{d-1} c_i^*\, x_i + n^{-\frac{1}{2}} \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} c_{ij}^*\, x_i x_j + n^{-1} \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} \sum_{k=1}^{d-1} c_{ijk}^*\, x_i x_j x_k, \end{aligned}$$

say. For all $n \geq n_0(b)$, we have $|c_i^*|$, $|c_{ij}^*|$ and $|c_{ijk}^*| \leq 2b$ for all i,j,k, and $|c_i^*| > 1/2b$ for some i, no matter what the value of $|x_d| \leq b \log n$. Consequently,

$$\sup_{-\infty<t<\infty} P\{\|\underset{\sim}{X}\| \leq b \log n \text{ and } r_n(\underset{\sim}{X}) \in [t-n^{-\beta},\ t+n^{-\beta}]\}$$

$$\leq \sup_{-\infty<t<\infty} \sup_{|x_d| \leq b \log n} P\{\|\underset{\sim}{X}^*\| \leq b \log n \text{ and } r_n^*(\underset{\sim}{X}^*|x_d) \in [t-n^{-\beta}, t+n^{-\beta}]\}$$

$$\leq C(\beta,2b,d-1)n^{-\beta}.$$

This proves the statement, and so completes the proof of Lemma 5.2. □

Combining (5.5), (5.13) and Lemma 5.2, we conclude that if $\mathcal{Y} \in C_{n8}$,

$$(5.14) \qquad \sup_{-\infty<t<\infty} |P\{n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}}|\underset{\sim}{\bar{Y}}) \le t|\mathcal{Y}\} - \sum_{t=0}^{2} n^{-r/2} P_r(-\Phi_{\underset{\sim}{0},\underset{\sim}{V}}: \{\chi_\nu\})\{A(t)\}| \le C_{31} n^{-1-\varepsilon_2}.$$

The remainder of the proof, which only involves unravelling terms in the polynomial expansion (5.14) to obtain the terms in (5.1), may be conducted by modifying arguments used to establish Lemma 2.1 and Theorem 2(b) of [5, pp. 443-444 and 445-446]. The cubic polynomial technique used to prove our Lemma 5.2 may be employed to overcome minor technical difficulties. The set $C_n^{(1)}$ is a suitably chosen subset of C_{n8}.

Step (ii). Here we develop an approximation to t_α. Define

$$(5.15) \qquad t_{\alpha,\pm} \equiv \inf\{t: \Phi(t) + \sum_{j=1}^{2} n^{-j/2} \Pi_{1j}(t)\phi(t) \mp C_1 n^{-1-\varepsilon_1} \ge \alpha\},$$

where C_1 and ε_1 are as in Proposition 5.1 and the +,- signs are taken respectively. The left-hand side of the inequality within braces in (5.15) converges to $1 \mp C_1 n^{-1-\varepsilon_1}$ as $t \to \infty$. Therefore the set in (5.15) is guaranteed nonempty, and $t_{\alpha,\pm}$ well-defined, if we confine attention to $\alpha \in T \equiv [3C_1 n^{-1-\varepsilon_1}, 1-3C_1 n^{-1-\varepsilon_1}]$.

Notice that $\phi(\log n)$ and $1-\Phi(\log n)$ are both $O(n^{-\lambda})$ for all $\lambda > 0$. This observation and inequality (5.2) lead us to conclude that for some $C_{32} > 0$, $\sup_{\alpha \in T} |t_{\alpha,\pm}| \le \log n$ provided $\mathcal{Y} \in C_n^{(1)}$ and $n \ge C_{32}$.

By Proposition 5.1, if $\mathcal{Y} \in C_n^{(1)}$ and $P\{n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}}|\underset{\sim}{\bar{Y}}) \le t|\mathcal{Y}\} \ge \alpha$ then

$$P\{n^{\frac{1}{2}} g(\underset{\sim}{\bar{Z}}|\underset{\sim}{\bar{Y}}) \le t|\mathcal{Y}\} + C_1 n^{-1-\varepsilon_1} - \Delta_n(t) \ge \alpha.$$

Therefore, $t_{\alpha,-} \leq t_\alpha$, and likewise, $t_{\alpha,+} \geq t_\alpha$, whence for $y \in C_n^{(1)}$:

(5.16) $$t_{\alpha,-} \leq t_\alpha \leq t_{\alpha,+} .$$

Step (iii). Here we invert the expansion defining $t_{\alpha,\pm}$. We begin by discussing the deterministic polynomials π_{1j} and π_{2j}, which are related to one another as follows. Let

(5.17) $$x = x(y) = y + \sum_{j=1}^{2} n^{-j/2} \pi_{2j}(y) .$$

Then $\Phi(x) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(x)\phi(x) = \Phi(y) + \delta_1(y)$, where $|\delta_1| \leq C_{33} n^{-3/2}$ uniformly in $|y| \leq \log n$. Thus,

(5.18) $$\Phi(x) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(x)\phi(x) \geq \Phi(y) - C_{33} n^{-3/2}$$

uniformly in $|y| \leq \log n$. Let $z=z(\alpha)$ be the solution of $\Phi(z)=\alpha$, and write $t=t(\alpha)$ for the solution of

(5.19) $$\Phi(t) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(t)\phi(t) = \alpha.$$

Let T be the set of values of t such that the left-hand side of (5.19) lies within $T_1 \equiv [C_1 n^{-1-\varepsilon_1}, 1-C_1 n^{-1-\varepsilon_1}]$. If n is sufficiently large then the left-hand side is a strictly increasing function of t for $t \in T$, and so if $\alpha \in T_1$ then $t(\alpha)$ is uniquely defined. Assume $\alpha \in T(\subseteq T_1)$, and take $y \equiv z(\alpha + C_1 n^{-1-\varepsilon_1} + C_{33} n^{-3/2})$ in (5.17) and (5.18). Let $t' = t(\alpha + C_1 n^{-1-\varepsilon_1})$. For sufficiently large n, each of $x(y)$, y and t' is in T for all $\alpha \in T$. By (5.17) and (5.18),

$$\Phi(x) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(x)\phi(x) \geq (\alpha + C_1 n^{-1-\varepsilon_1} + C_{33} n^{-3/2}) - C_{33} n^{-3/2}$$

$$= \Phi(t') + \sum_{j=1}^{n} n^{-j/2} \pi_{ij}(t')\phi(t') .$$

Consequently, $x \geq t'$; that is,

$$y + \sum_{j=1}^{2} n^{-j/2} \pi_{2j}(y) \geq \inf\{t: \Phi(t) + \sum_{j=1}^{2} n^{-j/2} \pi_{1j}(t)\phi(t) - C_1 n^{-1-\varepsilon_1} \geq \alpha\}.$$

Translated to the random polynomials Π_{1j} and Π_{2j}, this argument suggests the following. There exist positive constants C_{34} and C_{35} such that, with $y_+ \equiv z(\alpha + C_1 n^{-1-\varepsilon_1} + C_{34} n^{-3/2})$, we have

$$(5.20) \qquad y_+ + \sum_{j=1}^{2} n^{-j/2} \Pi_{2j}(y_+) \geq t_{\alpha,+}$$

whenever $\alpha \in T$ and $n \geq C_{35}$. Techniques used early in the proof of Proposition 5.1 show that this translation is correct, provided we make the restriction $\mathcal{Y} \in C_n^{(2)}$ for a suitable $C_n^{(2)} \subseteq C_n^{(1)}$ with $P(C_n^{(2)}) = o(n^{-1})$. The proof becomes quite straightforward when it is noted that the constant C_{33} cited earlier may be taken to be a continuous function of the first four population moments. To obtain the constant C_{34}, we should ensure that the sample moments are sufficiently close to the population moments with probability $1-o(n^{-1})$ - see the second sentence of the proof of Proposition 5.1.

Of course, inequality (5.20) has a counterpart providing a lower bound to $t_{\alpha,-}$. Let $y_- \equiv z(\alpha - C_1 n^{-1-\varepsilon_1} - C_{34} n^{-3/2})$. Then

$$(5.21) \qquad y_- + \sum_{j=1}^{2} n^{-j/2} \Pi_{2j}(y_-) \leq t_{\alpha,-}$$

provided $\alpha \in T$, $n \geq C_{35}$ and $\mathcal{Y} \in C_n^{(2)}$.

Step (iv). Here we combine Steps (i) - (iii). Note from (5.16), (5.20) and (5.21) that

$$y_- + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(y_-) \le t_\alpha \le y_+ + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(y_+),$$

provided $\alpha \in T$, $n \ge C_{35}$ and $\mathcal{Y} \in C_n^{(2)}$. Therefore

$$(5.22)\qquad P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le y_- + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(y_-)\} - P(\tilde{C}_n^{(2)})$$

$$\le P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le t_\alpha\}$$

$$\le P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le y_+ + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(y_+)\} + P(\tilde{C}_n^{(2)}),$$

provided $\alpha \in T$ and $n \ge C_{35}$. If we prove that, with $z = z(\alpha)$,

$$(5.23)\qquad P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le z + \sum_{j=1}^{2} n^{-j/2}\,\Pi_{2j}(z)\} = \alpha + n^{-1}\psi(z)\phi(z) + o(n^{-1})$$

uniformly in $\alpha \in T_1$, then it is immediate from (5.22) that expansion (2.6) holds uniformly in $\alpha \in T$. We shall establish (5.23) in Step (vi) below. In the meantime, Step (v) extends (2.6) from $\alpha \in T$ to $\alpha \in (0,1)$.

Step (v). Let $\alpha_1 = 3C_1\, n^{-1-\varepsilon_1}$ and $\alpha_2 = 1 - 3C_1\, n^{-1-\varepsilon_1}$. Then $T = [\alpha_1, \alpha_2]$. Since t_α is nondecreasing in α, and since (2.6) holds for $\alpha \in T$, then

$$\sup_{0<\alpha<\alpha_1} P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le t_\alpha\} \le P\{n^{\frac{1}{2}}\, g(\underset{\sim}{\bar{Y}}|\underset{\sim}{\mu}) \le t_{\alpha_1}\}$$

$$= \alpha_1 + n^{-\frac{1}{2}}\psi\{z(\alpha_1)\} + o(n^{-1}) = o(n^{-1}).$$

The case $\alpha_2 < \alpha < 1$ may be treated similarly.

Step (vi). Here we prove (5.23). That result is of the type established as Theorem 3 of [10], and may be proved in the same way. As in [10], the argument rests heavily on ideas from [5]. The only real difference here is that in (5.23), the expansion is to hold over slightly more than just bounded z. However, the terms $n^{-j/2}$ multiplying the polynomials, and the fact that $|z(\alpha)|$ is not larger than const. $(\log n)^{\frac{1}{2}}$ for $\alpha \in T_1$, ensure that this extra generality is easily achieved. For example, the polynomial Π_{22} depends on $\mathcal{Y}$ only through the first four sample moments. If M is any one of these moments and if $0 < \rho < \varepsilon/2(8+\varepsilon)$, then for any $c > 0$,

$$(5.24) \qquad P(|M-EM| > cn^{-\rho}) \leq (c^{-1}n^{\rho})^{2+(\varepsilon/4)} \, E|M-E(M)|^{2+(\varepsilon/4)} = o(n^{-1}) .$$

Let $C_{n,10}$ be the set of samples $\mathcal{Y}$ for which the difference between any coefficient Γ of Π_{22} and its limit $\gamma \equiv p \lim \Gamma$ satisfies $|\Gamma-\gamma| \leq n^{-\rho}$. In view of (5.24), $P(\tilde{C}_{n,10}) = o(n^{-1})$, and also for $n \geq 2$,

$$\sup_{\mathcal{Y} \in C_{n,10}} \; \sup_{z(\alpha):\alpha \in T_1} n^{-1}|\Pi_{22}(z) - \pi_{22}(z)| \leq C_{36} \, n^{-1} \, n^{-\rho}(\log n)^{5/2} = o(n^{-1}),$$

since Π_{22} is of degree 5.

For these reasons we shall give only an identification of expansion (5.23). Following [5,10] we work with Taylor expansions to order n^{-1} in probability. ("To order $n^{-j/2}$" means that terms of order $n^{-j/2}$ are included but smaller terms are excluded.) We shall use the summation notation and define $\Delta_{ri} \equiv Y_{ri} - \mu_i$, $\bar{\Delta}_i \equiv n^{-1} \sum_{r=1}^{n} \Delta_{ri}$.

According to Lemma 1 of [10], $\pi_{21}(z) = -\pi_{11}(z) = a_1 + (a_3/6)(z^2-1)$, where $a_i \equiv (1/2)\ \ell_{ij}\ \alpha_{ij}$ and $a_3 = \ell_i \ell_j \ell_k\ \alpha_{ijk} + 3\ell_i\ell_j\ell_{km}\ \alpha_{ij}\alpha_{km}$. But by (2.4), $\ell_i\ell_j\alpha_{ij} = 1$, and so

$$\pi_{21}(z) = (1/2)z^2\ \ell_{ij}\alpha_{ij} + (1/6)(z^2-1)\ell_i\ell_j\ell_k\ \alpha_{ijk}\,.$$

The sample versions of $\ell_{i_1 \ldots i_p}$ and $\alpha_{i_1 \ldots i_p}$ used to construct Π_{21} are

$$L_{i_1\ldots i_p} \equiv g_{i_1\ldots i_p}(\bar{\underset{\sim}{Y}}|\bar{\underset{\sim}{Y}}) \quad \text{and} \quad A_{i_1\ldots i_p} \equiv n^{-1} \sum_{r=1}^{n} \prod_{j=1}^{p} (\Delta_{ri_j} - \bar{\Delta}_{i_j}),$$

respectively. To order $n^{-\frac{1}{2}}$, $L_{i_1\ldots i_p} \approx \ell_{i_1\ldots i_p} + \ell^{(j)}_{i_1\ldots i_p}\ \bar{\Delta}_j$ and

$$A_{i_1\ldots i_p} = \alpha_{i_1\ldots i_p} + W_{i_1\ldots i_p} - \alpha_{i_1\ldots i_p}(i_j)\bar{\Delta}_{i_j}$$

(in summation notation), where $\alpha_{i_1\ldots i_p}(i_j) \equiv \alpha_{i_1\ldots i_{j-1}i_{j+1}\ldots i_p}$ and

$$W_{i_1\ldots i_p} \equiv n^{-1} \sum_{r=1}^{n} (\Delta_{ri_1}\ldots\Delta_{ri_p} - \alpha_{i_1\ldots i_p})\,.$$

Therefore to order $n^{-\frac{1}{2}}$,

$$\Pi_{21}(z) \approx \pi_{21}(z) + (1/2)z^2\ (\ell^{(k)}_{ij}\alpha_{ij}\ \bar{\Delta}_k + \ell_{ij}\ W_{ij} - 2\ell_{ij}\alpha_i\bar{\Delta}_j)$$

$$+ (1/6)(z^2-1)(3\ell_i\ell_j\ell^{(m)}_k\alpha_{ijk}\ \ \bar{\Delta}_m + \ell_i\ell_j\ell_k\ W_{ijk} - 3\ell_i\ell_j\ell_k\ \alpha_{ij}\ \bar{\Delta}_k)\,.$$

By definition, each α_i=0. To order 1, $\Pi_{22}(z) \approx \pi_{22}(z)$. Therefore to order n^{-1}, $n^{\frac{1}{2}}\ g(\bar{\underset{\sim}{Y}}|\underset{\sim}{\mu}) - \sum_{r=1}^{2} n^{-r/2}\ \Pi_{2r}(z)$ equals

$$U^* \equiv U - n^{-\frac{1}{2}}\{(1/2)z^2(\ell_{ij}^{(k)}\alpha_{ij}\,\bar{\Delta}_k + \ell_{ij}\,W_{ij})$$

$$+ (1/6)(z^2-1)(3\ell_i\ell_j\ell_k^{(m)}\alpha_{ijk}\,\bar{\Delta}_m + \ell_i\ell_j\ell_k\,W_{ijk} - 3\ell_i\ell_j\ell_k\alpha_{ij}\,\bar{\Delta}_k)\}\,,$$

where $U \equiv n^{\frac{1}{2}}\{\ell_i\,\bar{\Delta}_i + (1/2)\ell_{ij}\,\bar{\Delta}_i\,\bar{\Delta}_j + (1/6)\,\ell_{ijk}\,\bar{\Delta}_i\,\bar{\Delta}_j\,\bar{\Delta}_k\} - \sum_{r=1}^{2} n^{-r/2}\pi_{2r}(z)$.

Let κ_i and κ_i^* denote the i'th cumulants of U and U*, respectively. With approximations holding to order n^{-1}, we have $\kappa_1^* = \kappa_1$, $\kappa_3^* \cong \kappa_3$,

$$\kappa_2^* \cong \kappa_2 - [z^2\{\ell_{ij}^{(k)}\ell_p\alpha_{ij}E(\bar{\Delta}_k\bar{\Delta}_p) + \ell_{ij}\ell_p\,E(W_{ij}\bar{\Delta}_p)\}$$

$$+ (1/3)(z^2-1)\{3\ell_i\ell_j\ell_k^{(m)}\ell_p\,\alpha_{ijk}\,E(\bar{\Delta}_m\bar{\Delta}_p) + \ell_i\ell_j\ell_k\ell_p\,E(W_{ijk}\,\bar{\Delta}_p)$$

$$- 3\ell_i\ell_j\ell_k\ell_p\,\alpha_{ij}\,E(\bar{\Delta}_k\bar{\Delta}_p)\}]\,,$$

$$\kappa_4^* \cong \kappa_4 - 2n[z^2\{\ell_{ij}^{(k)}\ell_p^3\,\alpha_{ij}\,E(\bar{\Delta}_k\bar{\Delta}_p^3) + \ell_{ij}\ell_p^3\,E(W_{ij}\,\bar{\Delta}_p^3)\}$$

$$+ (1/3)(z^2-1)\{3\ell_i\ell_j\ell_k^{(m)}\ell_p^3\,\alpha_{ijk}\,E(\bar{\Delta}_m\bar{\Delta}_p^3) + \ell_i\ell_j\ell_k\ell_p^3\,E(W_{ijk}\,\bar{\Delta}_p^3)$$

$$- 3\ell_i\ell_j\ell_k\ell_p^3\,\alpha_{ij}\,E(\bar{\Delta}_k\bar{\Delta}_p^3)\}]$$

$$+ 6[z^2\{\ell_{ij}^{(k)}\ell_p\,\alpha_{ij}\,E(\bar{\Delta}_k\bar{\Delta}_p) + \ell_{ij}\ell_p\,E(W_{ij}\,\bar{\Delta}_p)\}$$

$$+ (1/3)(z^2-1)\{3\ell_i\ell_j\ell_k^{(m)}\ell_p\,\alpha_{ijk}\,E(\bar{\Delta}_m\bar{\Delta}_p) + \ell_i\ell_j\ell_k\ell_p\,E(W_{ijk}\,\bar{\Delta}_p)$$

$$- 3\ell_i\ell_j\ell_k\ell_p\,\alpha_{ij}\,E(\bar{\Delta}_k\bar{\Delta}_p)\}]\,.$$

Notice that $E(W_{i_1\cdots i_p}\,\bar{\Delta}_j) = n^{-1}\,\alpha_{i_1\cdots i_p j}$, $E(W_{i_1\cdots i_p}\,\bar{\Delta}_j^3) \sim 3n^{-2}\,\alpha_{i_1\cdots i_p j}\,\alpha_{jj}$ as $n\to\infty$, and $\bar{\Delta}_i = W_i$. Therefore to order n^{-1}, $\kappa_2^* \cong \kappa_2 - n^{-1}\,\delta_2(z)$ and $\kappa_4^* \cong \kappa_4 - 6n^{-1}\,\delta_4(z)$, where $\delta_2(z) \equiv \sum_i \sum_j \sum_p \gamma_{ijp}(z)$,

$$\delta_4(z) \equiv \sum_i \sum_j \sum_p (\ell_p^2 \alpha_{pp} - 1)\, \gamma_{ijp}(z)\,,$$

$$\gamma_{ijp}(z) \equiv z^2(\ell_p\, \alpha_{ij} \sum_k \ell_{ij}^{(k)}\, \alpha_{kp} + \ell_{ij}\ell_p\, \alpha_{ijp})$$

$$+ (1/3)(z^2-1)(3\ell_i\ell_j\ell_p \sum_k \sum_m \ell_k^{(m)}\, \alpha_{ijk}\, \alpha_{mp} + \ell_i\ell_j\ell_p \sum_k \ell_k\, \alpha_{ijkp}$$

$$- 3\ell_i\ell_j\ell_p\, \alpha_{ij} \sum_k \ell_k\, \alpha_{kp}).$$

(These expressions are not be interpreted in the convention of summation notation.) Consequently,

$$\text{(5.25)} \quad P(U^* \le x) = P(U \le x) - n^{-1}\{(1/2!)\delta_2(z)(\partial/\partial x) + (1/4!)\; 6\delta_4(z)(\partial/\partial x)^3\}\phi(x) + o(n^{-1})$$

$$= P(U \le x) + (x/4n)\{2\delta_2(z) + \delta_4(z)(x^2-3)\}\phi(x) + o(n^{-1})$$

uniformly in $-\infty<x<\infty$ and $z=z(\alpha)$ with $\alpha \in T_1$. The definition of π_{21} and π_{22} ensures that $P\{U \le z(\alpha)\} = \alpha + o(n^{-1})$. Finally, U^* has the same expansion, to order n^{-1}, as $n^{\frac{1}{2}}\, g(\underset{\sim}{\overline{Y}}|\underset{\sim}{\mu}) - \sum_{r=1}^{2} n^{-r/2}\, \Pi_{2r}(z)$. Therefore (5.23) follows on taking $x=z$ in (5.25), and noting that

$$\psi(z) = (z/4)\{2\delta_2(z) + \delta_4(z)(z^2-3)\}\;. \qquad \square$$

REFERENCES

1. ABRAMOVITCH, L. and SINGH, K. (1984). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.*

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

3. BERAN, R.J. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* 10, 212-225.

4. BERAN, R.J. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.* 12, 101-118.

5. BHATTACHARYA, R.N. and GHOSH, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* 6, 435-451.

6. BHATTACHARYA, R.N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

7. BICKEL, P.J. and FREEDMAN, D. (1981). Some asymptotics on the bootstrap. *Ann. Statist.* 9, 1196-1217.

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

9. EFRON, B. (1982). *The Jackknife, the Bootstrap and other Resampling Plans.* SIAM, Philadelphia.

10. HALL, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* 11, 569-576.

11. HALL, P. (1985). A tabular method for correcting skewness. *Math. Proc. Cambridge Phil. Soc.*

12. PFANZAGL, J. (1979). Nonparametric minimum contrast estimators. In *Selecta Statistica Canadiana, V.*

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