The Annals of Statistics
1986, Vol. 14, No. 4, 1431–1452

ON THE BOOTSTRAP AND CONFIDENCE INTERVALS[1]

BY PETER HALL[2]

University of North Carolina

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.

1. Introduction. Several authors [2, 3, 4, 15] have used sample path properties to argue that Efron's [8, 9] bootstrap is an empiric one-term Edgeworth inversion. In this paper we shall tackle the problem from a different viewpoint—that of coverage probability of confidence intervals. Arguing in that vein we shall show the bootstrap to have two important advantages over simpler inversions—it is smooth (i.e., uniform) and automatic (i.e., does not require preliminary theoretical calculation of the expansion's first term).

We shall show that the bootstrap 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 [12]. 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. Bootstrap iteration is very different from the type of "Edgeworth correction" advocated by Abramovitch and Singh [1], Hall [12] or Withers [16, 17]. All those techniques require explicit calculation of terms in Edgeworth expansions, and explicit correction for them. In comparison, bootstrap iteration involves no calculation of Edgeworth expansions and no explicit correction. All "corrections" are implicit, via a nested sequence of resampling operations.

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 Eq. (3.2). Our expansion is entirely different from earlier expansions

Received February 1985; revised March 1986.
[1] Research partially supported by AFOSR Contract No. F49620-82-C-0009.
[2] On leave from Australian National University.
AMS 1980 *subject classifications.* Primary 62E20; secondary 62G05.
Key words and phrases. Bootstrap, central limit theorem, Edgeworth expansion, rate of convergence, Studentized statistic.

connected with the bootstrap, such as those described by Babu and Singh [13] and Singh [15] in the case of k-sample Studentized means. The expansions studied by Babu and Singh hold with probability 1 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. They help to explain the encouraging performance of the bootstrap method for constructing confidence intervals, described empirically in other work (e.g., [10, 11]). 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, 12], and in those places we omit many details. In addition to the works of Babu and Singh already cited, 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 $\mathbf{Y}, \mathbf{Y}_1, \mathbf{Y}_2, \dots$ be independent and identically distributed d-dimensional random vectors with mean $\mu = E(\mathbf{Y})$. Define $\mathscr{Y} \equiv \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ and $\overline{\mathbf{Y}} \equiv n^{-1}\Sigma_{r=1}^{n}\mathbf{Y}_r$. Denote the ith elements of $\mathbf{Y}$, $\mathbf{Y}_r$, $\overline{\mathbf{Y}}$ and μ 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 $\gamma_1(\mathscr{Y}) \equiv f(\overline{\mathbf{Y}}) - f(\mu)$. Typically $f(\overline{\mathbf{Y}}) \equiv \hat{\theta}$ is an estimate of a parameter $f(\mu) \equiv \theta$.

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

$$\sigma^2 \equiv \Sigma_i \Sigma_j \sigma_{ij} f_i(\mu) f_j(\mu), \tag{2.1}$$

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

$$\hat{\sigma}^2 \equiv \Sigma_i \Sigma_j \hat{\sigma}_{ij} f_i(\overline{\mathbf{Y}}) f_j(\overline{\mathbf{Y}}), \tag{2.2}$$

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

Let us adjoin to the vector $\mathbf{Y}_r$ the set of those products $Y_{ri}Y_{rj}$, $1 \le i \le j \le d$, which do not already appear in $\mathbf{Y}_r$. Thus, $\mathbf{Y}_r$ is expanded to $\mathbf{Y}_r^0$, say, of length $d_0 \le d(d+3)/2$. Let $\mathbf{Y}^0$ and $\mu^0 = E(\mathbf{Y}^0)$ denote the corresponding lengthenings of $\mathbf{Y}$ and μ, and let $\overline{\mathbf{Y}}^0 \equiv n^{-1}\Sigma_{r=1}^{n}\mathbf{Y}_r^0$. We may write $\gamma_2(\mathscr{Y})$ as a function of $\overline{\mathbf{Y}}^0$ alone: $g(\overline{\mathbf{Y}}^0|\mu) \equiv \{f(\overline{\mathbf{Y}}) - f(\mu)\}/\hat{\sigma}$. Clearly $g(\mu^0|\mu) = 0$. The asymptotic variance of $n^{1/2}g(\overline{\mathbf{Y}}^0|\mu)$ is unity, which is reflected in the fact that if we construct the quantity σ^2 at (2.1) for $g(\cdot|\mu)$ instead of $f(\cdot)$, we obtain precisely 1.

The bootstrap argument runs as follows. Condition on the sample $\mathscr{Y}$, and let $\mathbf{Z}, \mathbf{Z}_1, \dots, \mathbf{Z}_n$ be independent and identically distributed with the n-point distribution $P(\mathbf{Z} = \mathbf{Y}_r|\mathscr{Y}) = n^{-1}$, $1 \le r \le n$. Set $\overline{\mathbf{Z}} \equiv n^{-1}\Sigma_{r=1}^{n}\mathbf{Z}_r$. We may work out the distribution of $g(\bar{\mathbf{Z}}^0|\bar{\mathbf{Y}})$, conditional on $\mathscr{Y}$, to arbitrary accuracy using simulation. The distribution is discrete. Define

$$t_\alpha = t_\alpha(\mathscr{Y}) \equiv \inf\left\{t\colon P\left[n^{1/2}g(\bar{\mathbf{Z}}^0|\bar{\mathbf{Y}}) \le t|\mathscr{Y}\right] \ge \alpha\right\},$$

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

Before stating the theorem we list notation and technical conditions. Given any vector $\mathbf{u}^0 = (u_1, \dots, u_d, u_{11}, u_{12}, \dots, u_{dd})^T = (u_1, \dots, u_{d^0})^T$ of length d^0, define $\mathbf{u} \equiv (u_1, \dots, u_d)^T$ (the first d elements of u^0),

$$\sigma^2(\mathbf{u}^0) \equiv \Sigma_i \Sigma_j (u_{ij} - u_i u_j) f_i(\mathbf{u}) f_j(\mathbf{u}),$$

$$g(\mathbf{u}^0|\mathbf{v}) \equiv \{f(\mathbf{u}) - f(\mathbf{v})\}/\{\sigma^2(\mathbf{u}^0)\}^{1/2},$$

$$g_{i_1 \cdots i_p}(\mathbf{u}^0|\mathbf{v}) \equiv \left(\partial^p/\partial u_{i_1} \cdots \partial u_{i_p}\right) g(\mathbf{u}^0|\mathbf{v}).$$

(Recall from the definition of $\mathbf{Y}^0$ that only those products $Y_i Y_j$ not already occurring in $\mathbf{Y}$ appear in the extension $\mathbf{Y}^0$. If $Y_i Y_j = Y_k \in \mathbf{Y}$, we define $u_{ij} = u_k$.) Let $B(\mathbf{u}, \rho)$ denote the open ball in $\mathbb{R}^d$ centered at $\mathbf{u}$ and of radius ρ, and define $B(\mathbf{u}^0, \rho)$ analogously. Given an extended mean vector $\mathbf{u}^0$ with $\sigma^2(\boldsymbol{\mu}^0) > 0$, assume that on some set $B(\boldsymbol{\mu}^0, \rho) \times B(\boldsymbol{\mu}, \rho)$ of vectors $(\mathbf{u}^0, \mathbf{v})$, $\sigma^2(\mathbf{u}^0) > 0$ and $g(\mathbf{u}^0|\mathbf{v})$ has four continuous derivatives with respect to $\mathbf{u}^0$. Let $h(\mathbf{u}^0|\mathbf{v})$ denote any of these derivatives of order 1, 2 or 3. Assume that $h(\mathbf{u}^0|\mathbf{u})$ has four continuous derivatives with respect to $\mathbf{u}^0$, for each choice of h and all $\mathbf{u}^0 \in B(\boldsymbol{\mu}^0, \rho)$. Suppose the distribution of $\mathbf{Y}$ satisfies $E(\mathbf{Y}^0) = \boldsymbol{\mu}^0$ and $E(\|\mathbf{Y}^0\|^{8+\varepsilon}) < \infty$, for some $\varepsilon > 0$. Lengthen $\mathbf{Y}^0$ to $\mathbf{Y}^1$ (of length d_1) by adjoining those triples $Y_i Y_j Y_k$ that do not already appear, let $\mathbf{t}$ be of length d_1, and assume Cramér's condition,

$$\limsup_{|\mathbf{t}| \to \infty} |E\{\exp(i\langle \mathbf{t}, \mathbf{Y}^1\rangle)\}| < 1. \tag{2.3}$$

(One consequence of (2.3) is that $\mathbf{Y}^1$ has nonsingular variance matrix.) Note that by definition of g,

$$g(\mathbf{u}^0|\mathbf{u}) = 0 \quad \text{and} \quad \sum_{i=1}^{d} \sum_{j=1}^{d} (u_{ij} - u_i u_j) g_i(\mathbf{u}^0|\mathbf{u}) g_j(\mathbf{u}^0|\mathbf{u}) = 1 \tag{2.4}$$

for all $\mathbf{u}^0 \in B(\boldsymbol{\mu}^0, \rho)$. Define

$$l_{i_1 \cdots i_p} \equiv g_{i_1 \cdots i_p}(\boldsymbol{\mu}^0|\boldsymbol{\mu}),\ l^{(j)}_{i_1 \cdots i_p} \equiv (\partial/\partial u_j) g_{i_1 \cdots i_p}(\mathbf{u}^0|\mathbf{u})|_{\mathbf{u}^0 = \boldsymbol{\mu}^0}$$

and $\alpha_{i_1 \cdots i_p} \equiv E\{\prod_{j=1}^p (Y_{i_j} - \mu_{i_j})\}$. In terms of these functions, let ψ be the odd, quintic polynomial defined by (5.28) in Section 5. 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 previous conditions,*

$$P\{n^{1/2}g(\bar{\mathbf{Y}}^0|\boldsymbol{\mu}) \le t_\alpha(\mathscr{Y})\} = \alpha + n^{-1}\psi\{z(\alpha)\}\phi\{z(\alpha)\} + o(n^{-1}), \tag{2.5}$$

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

An alternative approach would be to bootstrap the statistic $g^*(\bar{\mathbf{Y}}|\boldsymbol{\mu}^0)$, where

$$g^*(\mathbf{u}|\mathbf{v}^0) \equiv \{f(\mathbf{u}) - f(\mathbf{v})\}/\{\sigma^2(\mathbf{v}^0)\}^{1/2},$$

instead of $g(\bar{\mathbf{Y}}^0|\boldsymbol{\mu})$. In that case we would use the critical point estimate

$$t_\alpha^* = t_\alpha^*(\mathscr{Y}) \equiv \inf\{t\colon P[n^{1/2}g^*(\bar{\mathbf{Z}}|\bar{\mathbf{Y}}^0) \le t|\mathscr{Y}] \ge \alpha\},$$

instead of t. If $t(\mathscr{Y})$ on the left-hand side of (2.5) were to be replaced by $t^*(\mathscr{Y})$, then a term of order $n^{-1/2}$ would be introduced into the right-hand side, due to the fact that the statistic being bootstrapped is not pivotal. However, if the true sampling variance of the estimator were known then the statistic would be pivotal; a version of Theorem 2.1 is available for this case. In fact, under conditions similar to those in Theorem 2.1,

$$P\{n^{1/2}g^*(\bar{\mathbf{Y}}|\boldsymbol{\mu}^0) \le t_\alpha^*(\mathscr{Y})\} = \alpha + n^{-1}\psi^*\{z(\alpha)\}\psi\{z(\alpha)\} + o(n^{-1}),$$

where ψ^* is essentially the same as ψ but with g^* replacing g in the definition.

3. Discussion.

Our discussion will be confined to Theorem 2.1.

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

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

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,

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

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

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

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

The coefficients of π_{2j} 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 $\Theta_1(z) \equiv n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} - \sum_{j=1}^{k} n^{-j/2}\Pi_{2j}(z)$ coincide with those of $\Theta_2(z) \equiv n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} - \sum_{j=1}^{k} n^{-j/2}\pi_{2j}(z)$ only to order $n^{-1/2}$. Even-order moments of $\Theta_1(z)$ contain terms of order n^{-1} that do not appear in the corresponding moments of $\Theta_2(z)$. As a result, with $z = z(\alpha)$,

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

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 [12, 16, 17]. Edgeworth expansions are generally not monotone functions, and so problems are bound to arise if we attempt to invert them. In particular, the statement

$$P\left\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq z + \sum_{j=1}^{2} n^{-j/2}\Pi_{2j}(z)\right\} = \alpha + n^{-1}\psi(z)\phi(z) + o(n^{-1})$$

does not hold uniformly in α. On the other hand, the bootstrap involves inversion of a *monotone* function (the distribution function of the resampled bootstrap statistic), and as a result, the statement

$$P\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq t_\alpha\} = \alpha + n^{-1}\psi(z)\phi(z) + o(n^{-1})$$

is available uniformly in α. Simulations summarised in [12] indicate the disadvantages of the former approximation; in particular, it can overcorrect for small n. Neither coverage statement holds uniformly in F.

3.3. *An example.* 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 $\mathbf{u}^0 = (u_1, u_{11})^T = (u_1, u_2)^T$, with u_2 representing mean square. The Studentized mean is given by

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

Assume without loss of generality that $E(Y) = 0$ and $E(Y^2) = 1$. Then $l_1 = 1$, $l_{12} = -\frac{1}{2}$, $l_1^{(2)} = -\frac{1}{2}$, $l_{11}^{(1)} = 2$, $l_{12}^{(2)} = \frac{3}{4}$, and other terms are zero. Let $\mu_3 = E(Y^3)$ and $\mu_4 = E(Y^4)$. It follows from the definition of ψ that

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

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, platykurtic distributions. On the other hand, contributions of skewness and kurtosis have some tendency to cancel in the case of skew, leptokurtic distributions.

It was shown in [10] that statistics that 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,

$$(3.3) \qquad P\left[n^{1/2} g(\bar{\mathbf{Y}}^0 | \boldsymbol{\mu}) \le t_\alpha(\mathscr{Y}) - n^{-1}\hat{\psi}\{z(\alpha)\}\right] = \alpha + o(n^{-1})$$

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 $g(\bar{\mathbf{Y}}^0 | \boldsymbol{\mu}) = \{f(\bar{\mathbf{Y}}) - f(\boldsymbol{\mu})\}/\hat{\sigma}$, with $\hat{\sigma}$ defined at (2.2), we see that the interval

$$\Big(f(\bar{\mathbf{Y}}) - n^{-1/2}\hat{\sigma}\left[t_{1-\alpha/2}(\mathscr{Y}) - n^{-1}\hat{\psi}\{z(1 - \alpha/2)\}\right],$$
$$f(\bar{\mathbf{Y}}) - n^{-1/2}\hat{\sigma}\left[t_{\alpha/2}(\mathscr{Y}) - n^{-1}\hat{\psi}\{z(\alpha/2)\}\right]\Big)$$

covers $f(\boldsymbol{\mu})$ 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. Unless $\mu_4 - 3 - \frac{3}{2}\mu_3^2 = 0$, (3.3) will fail either as $\alpha \to 0$ or as $\alpha \to 1$. An alternative, smooth bootstrap tightening may be constructed as follows. Calculate the value $\hat{\lambda} \equiv \hat{\mu}_4 - 3 - \frac{3}{2}\hat{\mu}_3^2$. Conditional on $\mathscr{Y}$, construct a known, "comparison distribution" with zero mean and unit variance, having its third and fourth moments μ_{03} and μ_{04} satisfying $\hat{\lambda} = \mu_{04} - 3 - \frac{3}{2}\mu_{03}^2$. 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 $n^{1/2}g(\overline{\mathbf{Y}}|\boldsymbol{\mu})$ precisely $100\alpha\%$ of the time. Then

$$P\left\{n^{1/2}g(\overline{\mathbf{Y}}|\boldsymbol{\mu}) \le t_{\beta(\alpha)}(\mathscr{Y})\right\} = \alpha + o(n^{-1}) \tag{3.4}$$

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

$$\left(f(\overline{\mathbf{Y}}) - n^{-1/2}\hat{\sigma}t_{\beta(1-\alpha/2)}(\mathscr{Y}),\ f(\overline{\mathbf{Y}}) - n^{-1/2}\hat{\sigma}t_{\beta(\alpha/2)}(\mathscr{Y})\right)$$

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

3.4. *Methods related to the bootstrap.* If resampling can be conducted rapidly and inexpensively then even traditional statistical methods based on tabulated critical points may become obsolete. The "parametric resampling" used to generate statistical tables, and the "nonparametric resampling" used to construct nonparametric bootstrap critical points, are just two extremes of a vast array of techniques that use resampling to solve problems that are essentially numerical.

For example, it is possible to construct a smooth (i.e., uniform in $\alpha \in (0,1)$) resampling-based approximation, based on the sample $\mathscr{Y}$ *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 [13].) Among these methods related to the bootstrap, 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.

The bootstrap technique discussed in this paper has been termed the "percentile-t method". Efron [9] 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. Expansions of coverage probability similar to those given here may be derived for the bias-corrected interval, and terms of order n^{-1} persist in the expansion.

4. The iterated bootstrap. In Sections 3.1 and 3.2 we demonstrated that the bootstrap may be viewed as a first-order inversion of an Edgeworth expansion; in Sections 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 $\sigma^2(\mu^0)$ 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 μ_1 and μ_2 be their means, and define

$$h(G_2|G_1) \equiv n^{1/2}\{f(\mu_2) - f(\mu_1)\}/\{\sigma^2(\mu_2^0)\}^{1/2}.$$

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 $i \geq 1$, with F_0 being the (nonrandom) distribution of $\mathbf{Y}$. In this notation the random sample $\mathscr{Y}$ 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. (Of course, mention of a *single* random sample drawn from F_{i-1} serves only to *define* F_i. Calculation of F_i, or more importantly of functionals of F_i, requires repeated resampling from F_{i-1}. Calculation of F_{i+1} requires repeated resampling from every one of the many samples drawn from F_{i-1}, and so on. Therein lies the exponential computational tedium of the iterated bootstrap.)

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

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

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 be regarded as a general practical tool for the continuous case that is the subject of this paper. If the iterated bootstrap were to be applied, one would have to decide on the order of the iteration, r. Since the support of F_r decreases with increasing r, one cannot iterate indefinitely. (A decision on order must also be made for techniques based on Edgeworth inversion or Cornish–Fisher inversion.)

Result (4.1) may be proved along the lines of Theorem 2.1, and so will not be derived here.

5. Proof of Theorem 2.1. It is inconvenient to continue using the superscript on vectors $\overline{\mathbf{Y}}^0$, $\boldsymbol{\mu}^0$, etc., so we shall drop it, writing $g(\overline{\mathbf{Y}}|\boldsymbol{\mu})$ instead of $g(\overline{\mathbf{Y}}^0|\boldsymbol{\mu})$, and so on. This amounts to assuming that the vectors of interest are already sufficiently long, so that further extension is unnecessary. Write $\mathbf{V}_0 = (v_{ij}) \equiv \operatorname{var}(\mathbf{Y})$ ($\operatorname{var}(\mathbf{Y}^0)$ 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

$$P\{n^{1/2}g(\overline{\mathbf{Y}}|\boldsymbol{\mu}) \le x\} = \Phi(x) + \sum_{j=1}^{2} n^{-j/2}\pi_{1j}(x)\phi(x) + o(n^{-1})$$

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

$$P\{n^{1/2}g(\overline{\mathbf{Z}}|\overline{\mathbf{Y}}) \le x|\mathscr{Y}\} = \Phi(x) + \sum_{j=1}^{2} n^{-j/2}\Pi_{1j}(x)\phi(x) + o(n^{-1}).$$

Notice that only derivatives of order 3 or less are involved; hence the condition on derivatives of order 3 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 π_{2j} are of course simple functions of the coefficients of π_{1j}; 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 polynomials.

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

$$D^{\boldsymbol{\alpha}}f(\mathbf{x}) \equiv (\partial/\partial x_1)^{\alpha_1} \cdots (\partial/\partial x_d)^{\alpha_d} f(\mathbf{x}).$$

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

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

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

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

satisfies

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

and such that

$$\sup_{\mathscr{Y} \in \mathscr{C}_n^{(1)}} |\Pi_{1j}(x)| \le C_1(1 + |x|^{3j-1}) \tag{5.2}$$

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

Proof. We begin by defining several classes $\mathscr{C}_{nj}$ of samples $\mathscr{Y}$.

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

$$\begin{aligned} E(\|\mathbf{Z} - \bar{\mathbf{Y}}\|^p | \mathscr{Y}) &= n^{-1} \sum_{r=1}^{n} \left\{ \sum_{j=1}^{d} (Y_{rj} - \bar{Y}_j)^2 \right\}^{p/2} \\ &\le (2d)^p n^{-1} \sum_{r=1}^{n} \sum_{j=1}^{d} (|Y_{rj}|^p + |\bar{Y}_j|^p), \end{aligned}$$

and so

$$\begin{aligned} &P\left\{ E(\|\mathbf{Z} - \bar{\mathbf{Y}}\|^p | \mathscr{Y}) \ge (2d)^p \sum_{j=1}^{d} (E|Y_{1j}|^p + |E\bar{Y}_j|^p) + (2d)^{p+1} \right\} \\ &\quad \le P\left\{ \sum_{j=1}^{d} \left| \sum_{r=1}^{n} (|Y_{rj}|^p - E|Y_{rj}|^p) \right| + n \sum_{j=1}^{d} \||\bar{Y}_j|^p - |E\bar{Y}_j|^p| \ge 2dn \right\} \\ &\quad \le \sum_{j=1}^{d} n^{-(8+\varepsilon)/p} E\left\{ \left| \sum_{r=1}^{n} (|Y_{rj}|^p - E|Y_{rj}|^p) \right|^{(8+\varepsilon)/p} \right\} + \sum_{j=1}^{d} E|\bar{Y}_j - E\bar{Y}_j|^{8+\varepsilon} \\ &\quad = O(n^{-(8+\varepsilon)/2p}), \end{aligned}$$

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

Note that $\mathbf{V} = (V_{ij}) \to \mathbf{V}_0 = (v_{ij})$ in probability as $n \to \infty$, and that $\mathbf{V}_0$ is positive definite. Suppose all eigenvalues of $\mathbf{V}_0$ lie within $[2a, 1/2a]$, where $a > 0$. Choose $a > 0$ so small that any symmetric $d \times d$ matrix (u_{ij}) satisfying $|u_{ij} - v_{ij}| \le b$ for all i and j has all its eigenvalues within $[a, 1/a]$. Let $\mathscr{C}_{n3}$ be the class of all samples $\mathscr{Y}$ such that $|V_{ij} - v_{ij}| \le b$ for all i and j. If $\mathscr{Y} \in \mathscr{C}_{n3}$ then $C_4^{-1}\|\mathbf{t}\| \le \|\mathbf{V}^{-1/2}\mathbf{t}\| \le C_4\|\mathbf{t}\|$ for all $\mathbf{t} \in \mathbb{R}^d$, where C_4 depends only on $\mathbf{V}_0$ and b. The inequality

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

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

Let

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

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

$$V_{ij}^\dagger = \delta_{ij} - E\{(\mathbf{W})_i(\mathbf{W})_j I(\|\mathbf{W}\| > n^{1/2})|\mathscr{Y}\}$$
$$- E\{(\mathbf{W})_i I(\|\mathbf{W}\| > n^{1/2})|\mathscr{Y}\} E\{(\mathbf{W})_j I(\|\mathbf{W}\| > n^{1/2})|\mathscr{Y}\},$$

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

$$P\big[\mathscr{C}_{n3}; |E\{(\mathbf{W})_i(\mathbf{W})_j I(\|\mathbf{W}\| > n^{1/2})|\mathscr{Y}\}| > \eta\big]$$
$$\le P\big[C_4^2 E\{\|\mathbf{Z} - \overline{\mathbf{Y}}\|^2 I(C_4\|\mathbf{Z} - \overline{\mathbf{Y}}\| > n^{1/2})|\mathscr{Y}\} > \eta\big]$$
$$\le C_4^2\eta^{-1}\big(C_4^{-1}n^{1/2}\big)^{-3} E\big(\|\mathbf{Z} - \overline{\mathbf{Y}}\|^5\big) = O(n^{-3/2}).$$

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

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

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

Write $\{\chi_\nu\}$ for the cumulant sequence of $\mathbf{Z}$, conditional on $\mathscr{Y}$. The argument preceding (20.15) of [6, page 209] yields

$$
\begin{aligned}
(5.3)\qquad & \left| \int_A d\left[Q - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{0,\mathbf{V}} : \{\chi_\nu\}) \right] \right. \\
& \left. - \int_{A^\dagger} d\left[Q^\dagger - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{0,\mathbf{V}^\dagger} : \{\chi_\nu^\dagger\}) \right] \right| \\
& \le C_5 n^{-1} E\{\|\mathbf{W}\|^4 I(\|\mathbf{W}\| > n^{1/2}) | \mathscr{Y}\} \le C_5 n^{-1} C_4^4 (C_4^{-1} n^{1/2})^{-2\varepsilon_2} C_3,
\end{aligned}
$$

the last inequality following since $\mathscr{Y} \in \mathscr{C}_{n2} \cap \mathscr{C}_{n3}$. Using (9.12) and (14.74) of [6, pages 72 and 133], we obtain

$$
\begin{aligned}
(5.4)\qquad & \left| \int_{A^\dagger} d\left[Q^\dagger - \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{0,\mathbf{V}^\dagger} : \{\chi_\nu^\dagger\}) - H \right] \right| \\
& \le C_6 \sum_{r=3}^{d+2} n^{-r/2} E(\|\mathbf{Z}_1^*\|^{r+2} | \mathscr{Y}) \le C_7 n^{-1-\varepsilon_2},
\end{aligned}
$$

again since $\mathscr{Y} \in \mathscr{C}_{n2} \cap \mathscr{C}_{n3}$. Combining (5.3) and (5.4), we obtain on $\mathscr{C}_{n5}$,

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

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

$$
\begin{aligned}
(5.6)\qquad \left| \int_{A^\dagger} dH \right| & \le C_{10} \sup_{0 \le \alpha \le \beta,\, |\beta| \le d+5} \int |(D^{\beta-\alpha} \hat{H})(\mathbf{t})(D^\alpha \hat{K}_\delta)(\mathbf{t})| \, d\mathbf{t} \\
& \quad + \int I(A, \delta)(\mathbf{x}) \left| \sum_{r=0}^{d+2} n^{-r/2} P_r(-\phi_{0,\mathbf{V}^\dagger} : \{\chi_\nu^\dagger\}) \right| d\mathbf{x},
\end{aligned}
$$

where $I(A, \delta)(\mathbf{x}) = 1$ if either $\mathbf{x} \in A$ and $\tilde{A} \cap B(\mathbf{x}, \delta) \ne \varnothing$, or $\mathbf{x} \in \tilde{A}$ and $A \cap B$ $(\mathbf{x}, \delta) \ne \varnothing$; and $I(A, \delta)(\mathbf{x}) = 0$ otherwise. The constants C_j here and following do not depend on δ. Following the argument of [6, pages 210–211], with a little modification, we derive for a certain random variable A_n,

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

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

where

$$
\kappa \equiv \sup_{0 \le \alpha \le \beta,\, |\beta| \le d+5} \int_{\{\|\mathbf{t}\| > c_n\}} |(D^{\beta-\alpha} \hat{Q}^\dagger)(\mathbf{t})(D^\alpha \hat{K}_\delta)(\mathbf{t})| \, d\mathbf{t}
$$

$(\mathbf{x}, \delta) \ne$

and

$$c_n = n^{1/2}/\{16E(\|\mathbf{W}\|^3|\mathscr{Y})\}.$$

Let $\psi_n(\mathbf{t}) \equiv n^{-1}\Sigma_{r=1}^n \exp(i\langle \mathbf{t}, \mathbf{Y}_r\rangle)$ denote the (empiric) characteristic function of $\mathscr{Y}$, and $\psi(\mathbf{t}) \equiv E\{\psi_n(\mathbf{t})\}$ the characteristic function of $\mathbf{Y}$. 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 $\mathscr{Y} \in \mathscr{C}_{n5}$,*

$$\kappa \le C_{13} n^{d+5} \int \Big[C_{12}^n + \big\{ C_{14}|\psi_n(n^{-1/2}\mathbf{t}) - \psi(n^{-1/2}\mathbf{t})| \big\}^m \Big] \exp\big(-C_4^{-1/2}\|\delta\mathbf{t}\|^{1/2}\big)\, d\mathbf{t}.$$

PROOF. Set $q(\mathbf{t}) \equiv E\{\exp(n^{-1/2} i\langle \mathbf{t}, \mathbf{Z}_1^\dagger\rangle)|\mathscr{Y}\}$. Then $\hat{Q}^\dagger(\mathbf{t}) = q^n(\mathbf{t})$, from which it may be deduced that $|(D^\gamma \hat{Q}^\dagger)(\mathbf{t})| \le (2n)^{|\gamma|}|q(\mathbf{t})|^{n-|\gamma|}$. (Note that $|\mathbf{Z}_1^\dagger| \le 2n^{1/2}$.) Therefore, if $n > d+5$,

$$\kappa \le C_9 (2n)^{d+5} \int_{\{\|\mathbf{t}\| > c_n\}} |q(\mathbf{t})|^{n-d-5} \exp\big(-\|\delta \mathbf{t}\|^{1/2}\big)\, d\mathbf{t}. \tag{5.9}$$

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

$$\begin{aligned} |q(\mathbf{t})| &\le |E\{\exp(n^{-1/2} i\langle \mathbf{t}, \mathbf{V}^{-1/2}\mathbf{Z}\rangle)|\mathscr{Y}\}| + P(\|\mathbf{W}\| > n^{1/2}|\mathscr{Y}) \\ &\le |\psi_n(n^{-1/2}\mathbf{V}^{-1/2}\mathbf{t})| + C_{16} n^{-1}. \end{aligned} \tag{5.10}$$

Results (5.9) and (5.10) together give

$$\begin{aligned} \kappa \le C_{17} n^{d+5} \int_{\{\|n^{-1/2}\mathbf{t}\| > C_{15}\}} & \big\{ |\psi_n(n^{-1/2}\mathbf{t})| + C_{16} n^{-1} \big\}^{n-d-5} \\ & \times \exp\big(-C_4^{-1/2}\|\delta\mathbf{t}\|^{1/2}\big)\, d\mathbf{t}. \end{aligned} \tag{5.11}$$

Choose $\eta_1 \in [\frac{1}{2}, 1)$ and $\eta_2 > 0$ such that $\sup_{\|\mathbf{u}\| > \xi}|\psi(\mathbf{u})| \le \max(\eta_1, 1 - \eta_2\xi^2)$ for all $\xi > 0$. Let $\eta_3 = \eta_3(\xi) \in [\sup_{\|\mathbf{u}\| > \xi}|\psi(\mathbf{u})|, 1)$. If $|\psi_n(\mathbf{u})| > \eta_3$ then $C_{16} n^{-1} \le C_{16} n^{-1}\eta_3^{-1}|\psi_n(\mathbf{u})|$, and

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

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

$$\begin{aligned} \big\{ |\psi_n(\mathbf{u})| + C_{16} n^{-1} \big\}^{n-d-5} \le \exp(C_{16}/\eta_3) \Big[& \{\eta_3(1+\eta_4)\}^{n-d-5} \\ & + \big\{ (1 + \eta_4^{-1})|\psi_n(\mathbf{u}) - \psi(\mathbf{u})| \big\}^m \Big]. \end{aligned}$$

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

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

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

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

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

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

$$C_{n6} \equiv \Big\{\mathscr{Y}: \int|\psi_n(n^{-1/2+\lambda}\mathbf{t}) - \psi(n^{-1/2+\lambda}\mathbf{t})|^m \times\exp\big(-C_4^{-1/2}\|\mathbf{t}\|^{1/2}\big)\,d\mathbf{t} \le n^{-d(\lambda+1)-7}\Big\}$$

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

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

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

$$\left|\int_{A^\dagger} dH\right| \le C_{20}n^{-1-\varepsilon_2} + \xi, \tag{5.12}$$

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

$$\xi_r \equiv n^{-r/2}\int I(\mathbf{x})|P_r\big(-\phi_{\mathbf{0},\mathbf{V}^\dagger}: \{\chi_\nu^\dagger\}\big)|\,d\mathbf{x}$$

and $I(\mathbf{x}) \equiv I(A^\dagger, 2n^{-\lambda})$.

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

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

Let $\mathbf{a} \equiv n^{1/2}\mathbf{V}^{1/2}E(\mathbf{Z}_1^*|\mathscr{Y})$, and note that for $\mathscr{Y} \in \mathscr{C}_{n7}$,

$$B(\mathbf{x}) \equiv \mathbf{V}^{1/2}B(\mathbf{V}^{-1/2}\mathbf{x}, 2n^{-\lambda}) \subseteq B\big(\mathbf{x}, C_{22}n^{-\lambda}\big).$$

Thus,

$$\begin{aligned}\zeta_r &= (\det\mathbf{V}^{-1/2})\textstyle\int I(\mathbf{V}^{-1/2}\mathbf{x})\big(1 + \|\mathbf{V}^{-1/2}\mathbf{x}\|^{3r}\big)\\ &\qquad\times\exp\big\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{V}^{-1/2}\mathbf{V}^{\dagger-1}\mathbf{V}^{-1/2}\mathbf{x}\big\}\,d\mathbf{x}\\ &\le C_{23}\int I\big(A + \mathbf{a}, C_{22}n^{-\lambda}\big)\exp\big(-C_{24}\|\mathbf{x}\|^2\big)\,d\mathbf{x}\\ &\le C_{23}\int_{(\partial A)^\eta}\exp\big(-C_{24}\|\mathbf{x} + \mathbf{a}\|^2\big)\,d\mathbf{x},\end{aligned}$$

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

$$\exp\big(-\|\mathbf{x} + \mathbf{y}\|^2\big) \le \exp\big(-\|\mathbf{x}\|^2\big) + C_{25}\|\mathbf{y}\|\exp\big(-\|\mathbf{x} + \mathbf{y}\|^2/2\big)$$

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

$$\zeta_r \le C_{23}\int_{(\partial A)^\eta}\exp\big(-C_{24}\|\mathbf{x}\|^2\big)\,d\mathbf{x} + C_{27}n^{-3/2},$$

and so by (5.12),

$$\left|\int_{A^\dagger} dH\right| \le C_{28} n^{-1-\varepsilon_1} + C_{23} \int_{(\partial A)^\eta} \exp\left(-C_{24}\|\mathbf{x}\|^2\right) d\mathbf{x}. \tag{5.13}$$

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

$$A(t) \equiv \left\{\mathbf{x} \in \mathbb{R}^d\colon n^{1/2} g\left(\overline{\mathbf{Y}} + n^{-1/2}\mathbf{x} | \overline{\mathbf{Y}}\right) \le t\right\}.$$

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

$$\sup_{\mathscr{Y} \in \mathscr{C}_{n8}} \sup_{-\infty < t < \infty} \int_{\{\partial A(t)\}^\eta} \exp\left(-C_{24}\|\mathbf{x}\|^2\right) d\mathbf{x} \le C_{29} n^{-1-\varepsilon_2}.$$

PROOF. Suppose $g(\mathbf{u}|\mathbf{v})$ and its first four derivatives with respect to $\mathbf{u}$ are continuous in $(\mathbf{u}, \mathbf{v}) \in B(\boldsymbol{\mu}, \rho_1) \times B(\boldsymbol{\mu}, \rho_1)$. In view of property (2.4), not all terms $g_i(\boldsymbol{\mu}|\boldsymbol{\mu})$ can vanish. Let $\rho_2 \in (0, \rho_1/2]$ be such that for some i, $g_i(\mathbf{u}|\mathbf{v})$ is bounded away from zero in $B(\boldsymbol{\mu}, \rho_2) \times B(\boldsymbol{\mu}, \rho_2)$. Let $\mathscr{C}_{n9}$ be the class of all samples $\mathscr{Y}$ for which $\|\overline{\mathbf{Y}} - \boldsymbol{\mu}\| \le \rho_2/2$. The set $\mathscr{C}_{n8} = \mathscr{C}_{n7} \cap \mathscr{C}_{n9}$ satisfies $P(\tilde{\mathscr{C}}_{n8}) = o(n^{-1})$, and has the property that for some $C_{29} > 0$ and some $i_0 \in \{1, \dots, d\}$,

$$\sup_{\mathscr{Y} \in \mathscr{C}_{n8}} \sup_{i,j,k} \left\{|g_i(\overline{\mathbf{Y}}|\overline{\mathbf{Y}})|, |g_{ij}(\overline{\mathbf{Y}}|\overline{\mathbf{Y}})|, g_{ijk}(\overline{\mathbf{Y}}|\overline{\mathbf{Y}})|\right\} \le C_{29}$$

and $\inf_{\mathscr{Y} \in \mathscr{C}_{n8}} |g_{i_0}(\overline{\mathbf{Y}}|\overline{\mathbf{Y}})| > 1/C_{29}$. We assume throughout the argument below that $\mathscr{Y} \in \mathscr{C}_{n8}$.

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

$$\begin{aligned} p_n(\mathbf{x}) \equiv & \sum_{i=1}^d x_i g_i(\overline{\mathbf{Y}}|\overline{\mathbf{Y}}) + \tfrac{1}{2} n^{-1/2} \sum_{i=1}^d \sum_{j=1}^d x_i x_j g_{ij}(\overline{\mathbf{Y}}|\overline{\mathbf{Y}}) \\ & + \tfrac{1}{6} n^{-1} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d x_i x_j x_k g_{ijk}(\overline{\mathbf{Y}}|\overline{\mathbf{Y}}). \end{aligned}$$

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

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

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

$$\begin{aligned} \{\partial A(t)\}^\eta \subseteq & \left\{\mathbf{x}_1 \in \mathbb{R}^d\colon \text{for some } \mathbf{x}_2 \in \mathbb{R}^d, |\mathbf{x}_1 - \mathbf{x}_2| \le \eta \text{ and} \right. \\ & \left. \qquad n^{1/2} g\left(\overline{\mathbf{Y}} + n^{-1/2}\mathbf{x}_2 | \overline{\mathbf{Y}}\right) = t\right\} \\ \subseteq & \left\{A_1(t)\right\} \cup B(\mathbf{0}, \log n)^{\sim}, \end{aligned}$$

where $A_1(t) \equiv \{\mathbf{x}\colon \|\mathbf{x}\| \le 2\log n \text{ and } p_n(\mathbf{x}) \in [t - C_{31} n^{-1-\varepsilon_2}, t + C_{31} n^{-1-\varepsilon_2}]\}$. The lemma follows easily from this result and the statement:

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

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

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

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

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

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

for all $n \ge 1$."

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 $d \ge 2$. Without loss of generality, $|c_d|$ is the *smallest* $|c_i|$. Let $\mathbf{X}^* = (X_1, \dots, X_{d-1})^T$, $\mathbf{x}^* = (x_1, \dots, x_{d-1})^T$ and

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

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

$$\begin{aligned} &\sup_{-\infty < t < \infty} P\big\{\|\mathbf{X}\| \le b \log n \text{ and } r_n(\mathbf{X}) \in \big[t - n^{-\beta}, t + n^{-\beta}\big]\big\} \\ &\quad \le \sup_{-\infty < t < \infty} \sup_{|x_d| \le b \log n} P\big\{\|\mathbf{X}^*\| \le b \log n \text{ and} \\ &\qquad\qquad r_n^*(\mathbf{X}^* | x_d) \in \big[t - n^{-\beta}, t + n^{-\beta}\big]\big\} \\ &\quad \le C(\beta, 2b, d-1) n^{-\beta}. \end{aligned}$$

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 $\mathscr{Y} \in \mathscr{C}_{n8}$,

$$\begin{aligned} \sup_{-\infty < t < \infty} &\big|P\{n^{1/2} g(\overline{\mathbf{Z}} | \overline{\mathbf{Y}}) \le t | \mathscr{Y}\} \\ &- \sum_{r=0}^{2} n^{-r/2} P_r(-\Phi_{\mathbf{0},\mathbf{V}} : \{\chi_\nu\})\{A(t)\}\big| \le C_{31} n^{-1-\varepsilon_2}. \end{aligned} \tag{5.14}$$

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, pages 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 $\mathscr{C}_n^{(1)}$ is a suitably chosen subset of $\mathscr{C}_{n8}$. □

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

$$t_{\alpha,\pm} \equiv \inf\left\{t\colon \Phi(t) + \sum_{j=1}^{2} n^{-j/2}\Pi_{1j}(t)\phi(t) \mp C_1 n^{-1-\varepsilon_1} \geq \alpha\right\}, \tag{5.15}$$

where C_1 and ε_1 are as in Proposition 5.1 and the + and − 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 $C_{32} > 0$, $\sup_{\alpha \in T}|t_{\alpha,\pm}| \leq \log n$ provided $\mathscr{Y} \in \mathscr{C}_n^{(1)}$ and $n \geq C_{32}$.

By Proposition 5.1, if $\mathscr{Y} \in \mathscr{C}_n^{(1)}$ and $P\{n^{1/2}g(\bar{\mathbf{Z}}|\bar{\mathbf{Y}}) \leq t|\mathscr{Y}\} \geq \alpha$, then

$$P\{n^{1/2}g(\bar{\mathbf{Z}}|\bar{\mathbf{Y}}) \leq t|\mathscr{Y}\} + C_1 n^{-1-\varepsilon_1} - \Delta_n(t) \geq \alpha.$$

Therefore, $t_{\alpha,-} \leq t_\alpha$, and likewise, $t_{\alpha,+} \geq t_\alpha$, whence for $\mathscr{Y} \in \mathscr{C}_n^{(1)}$,

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

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

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

Then

$$\Phi(x) + \sum_{j=1}^{2} n^{-j/2}\pi_{1j}(x)\phi(x) = \Phi(y) + \delta_1(y),$$

where $|\delta_1(y)| \leq C_{33}n^{-3/2}$ uniformly in $|y| \leq \log n$. Thus,

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

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

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

Let $\mathscr{T}$ 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 $t \in \mathscr{T}$, 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 $\mathscr{T}$ for all $\alpha \in T$. By (5.17) and (5.18),

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

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

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

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

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

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

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 $\mathscr{Y} \in \mathscr{C}_n^{(2)}$ for a suitable $\mathscr{C}_n^{(2)} \subseteq \mathscr{C}_n^{(1)}$ with $P(\mathscr{C}_n^{(2)}) = o(n^{-1})$. 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

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

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

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

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

provided $\alpha \in T$, $n \geq C_{35}$ and $\mathscr{Y} \in \mathscr{C}_n^{(2)}$. Therefore,

$$(5.22)\qquad \begin{aligned} &P\left\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq y_- + \sum_{j=1}^{2} n^{-j/2}\Pi_{2j}(y_-)\right\} - P(\tilde{\mathscr{C}}_n^{(2)}) \\ &\leq P\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq t_\alpha\} \\ &\leq P\left\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq y_+ + \sum_{j=1}^{2} n^{-j/2}\Pi_{2j}(y_+)\right\} + P(\tilde{\mathscr{C}}_n^{(2)}), \end{aligned}$$

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

$$(5.23)\qquad P\left\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq z + \sum_{j=1}^{2} n^{-j/2}\Pi_{2j}(z)\right\} = \alpha + n^{-1}\psi(z)\phi(z) + o(n^{-1})$$

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). 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

$$\begin{aligned} \sup_{0<\alpha<\alpha_1} P\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq t_\alpha\} &\leq P\{n^{1/2}g(\overline{\mathbf{Y}}|\mu) \leq t_{\alpha_1}\} \\ &= \alpha_1 + n^{-1/2}\psi\{z(\alpha_1)\} + o(n^{-1}) = o(n^{-1}). \end{aligned}$$

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 [12], and may be proved in the same way. As in [12], 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.$(\log n)^{1/2}$ for $\alpha \in T_1$, ensure that this extra generality is easily achieved. For example, the polynomial Π_{22} depends on $\mathscr{Y}$ 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 $\mathscr{C}_{n,10}$ be the set of samples $\mathscr{Y}$ 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), $P(\tilde{\mathscr{C}}_{n,10}) = o(n^{-1})$, and also for $n \geq 2$,

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

since Π_{22} is of degree 5.

For these reasons we shall give only an identification of expansion (5.23). Following [5, 12] 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 [12], $\pi_{21}(z) = -\pi_{11}(z) = a_1 + (a_3/6)(z^2 - 1)$, where $a_1 \equiv \frac{1}{2}l_{ij}\alpha_{ij}$ and $a_3 = l_i l_j l_k \alpha_{ijk} + 3 l_i l_j l_{km}\alpha_{ik}\alpha_{jm}$. The sample versions of $l_{i_1 \cdots i_p}$ and $\alpha_{i_1 \cdots i_p}$ used to construct Π_{21} are

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

respectively. To order $n^{-1/2}$,

$$L_{i_1 \cdots i_p} \simeq l_{i_1 \cdots i_p} + l_{i_1 \cdots i_p}^{(j)} \bar{\Delta}_j$$

and

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

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

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

Therefore, to order $n^{-1/2}$, and since each $\alpha_i = 0$,

$$\begin{aligned} \Pi_{21}(z) \simeq \pi_{21}(z) &+ \tfrac{1}{2}\left(l_{ij}^{(k)} \alpha_{ij} \bar{\Delta}_k + l_{ij} W_{ij} \right) \\ &+ \tfrac{1}{6}(z^2 - 1)\left\{ l_i l_j l_k^{(m)} \alpha_{ijk} \bar{\Delta}_m + l_i l_j l_k W_{ijk} - 3 l_i l_j l_k \alpha_{ij} \bar{\Delta}_k \right. \\ &\qquad + 6 l_i l_j^{(p)} l_{km} \alpha_{ik} \alpha_{jm} \bar{\Delta}_p + 3 l_i l_j l_{km}^{(p)} \alpha_{ik} \alpha_{jm} \bar{\Delta}_p \\ &\qquad \left. + 3 l_i l_j l_{km} \left(\alpha_{ik} W_{jm} + \alpha_{jm} W_{ik} \right) \right\}. \end{aligned}$$

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

$$\begin{aligned} U^* \equiv U - n^{-1/2}\Big[&\tfrac{1}{2}\left(l_{ij}^{(k)} \alpha_{ij} \bar{\Delta}_k + l_{ij} W_{ij} \right) + \tfrac{1}{6}(z^2 - 1) \\ &\times \left\{ 3 l_i l_j l_k^{(m)} \alpha_{ijk} \bar{\Delta}_m + l_i l_j l_k W_{ijk} - 3 l_i l_j l_k \alpha_{ij} \bar{\Delta}_k + 6 l_i l_j^{(p)} l_{km} \alpha_{ik} \alpha_{jm} \bar{\Delta}_p \right. \\ &\left. + 3 l_i l_j l_{km}^{(p)} \alpha_{ik} \alpha_{jm} \bar{\Delta}_p + 3 l_i l_j l_{km} \left(\alpha_{ik} W_{jm} + \alpha_{jm} W_{ik} \right) \right\} \Big], \end{aligned}$$

where

$$U \equiv n^{1/2}\left\{ l_i \bar{\Delta}_i + \tfrac{1}{2} l_{ij} \bar{\Delta}_i \bar{\Delta}_j + \tfrac{1}{6} l_{ijk} \bar{\Delta}_i \bar{\Delta}_j \bar{\Delta}_k \right\} - \sum_{r=1}^{2} n^{-r/2} \pi_{2r}(z).$$

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

$$\begin{aligned}\kappa_2^* \simeq \kappa_2 - \Big[&\big\{l_{ij}^{(k)}l_p\alpha_{ij}E(\bar\Delta_k\bar\Delta_p) + l_{ij}l_pE(W_{ij}\bar\Delta_p)\big\}\\ &+\tfrac{1}{3}(z^2-1)\big\{3l_il_jl_k^{(m)}l_p\alpha_{ijk}E(\bar\Delta_m\bar\Delta_p) + l_il_jl_kl_pE(W_{ijk}\bar\Delta_p)\\ &\quad -3l_il_jl_kl_p\alpha_{ij}E(\bar\Delta_k\bar\Delta_p) + 6l_il_j^{(p)}l_{km}l_q\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_q)\\ &\quad +3l_il_jl_{km}^{(p)}l_q\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_q)\\ &\quad +3l_il_jl_{km}l_qE(\alpha_{ik}W_{jm}\bar\Delta_q + \alpha_{jm}W_{ik}\bar\Delta_q)\big\}\Big],\end{aligned} \tag{5.25}$$

$$\begin{aligned}\kappa_4^* \simeq \kappa_4 - 2n\Big[&\big\{l_{ij}^{(k)}l_{p_1}l_{p_2}l_{p_3}\alpha_{ij}E(\bar\Delta_k\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad + l_{ij}l_{p_1}l_{p_2}l_{p_3}E(W_{ij}\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\big\}\\ &+\tfrac{1}{3}(z^2-1)\big\{3l_il_jl_k^{(m)}l_{p_1}l_{p_2}l_{p_3}\alpha_{ijk}E(\bar\Delta_m\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad + l_il_jl_kl_{p_1}l_{p_2}l_{p_3}E(W_{ijk}\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad -3l_il_jl_kl_{p_1}l_{p_2}l_{p_3}\alpha_{ij}E(\bar\Delta_k\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad +6l_il_j^{(p)}l_{km}l_{p_1}l_{p_2}l_{p_3}\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad +3l_il_jl_{km}^{(p)}l_{p_1}l_{p_2}l_{p_3}\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\\ &\quad +6l_il_jl_{km}l_{p_1}l_{p_2}l_{p_3}\alpha_{ik}E(W_{jm}\bar\Delta_{p_1}\bar\Delta_{p_2}\bar\Delta_{p_3})\big\}\Big]\\ + 6\Big[&\big\{l_{ij}^{(k)}l_p\alpha_{ij}E(\bar\Delta_k\bar\Delta_p) + l_{ij}l_pE(W_{ij}\bar\Delta_p)\big\}\\ &+\tfrac{1}{3}(z^2-1)\big\{3l_il_jl_k^{(m)}l_p\alpha_{ijk}E(\bar\Delta_m\bar\Delta_p) + l_il_jl_kl_pE(W_{ijk}\bar\Delta_p)\\ &\quad -3l_il_jl_kl_p\alpha_{ij}E(\bar\Delta_k\bar\Delta_p) + 6l_il_j^{(p)}l_{km}l_q\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_q)\\ &\quad +3l_il_jl_{km}^{(p)}l_q\alpha_{ik}\alpha_{jm}E(\bar\Delta_p\bar\Delta_q)\\ &\quad +3l_il_jl_{km}l_qE(\alpha_{ik}W_{jm}\bar\Delta_q + \alpha_{jm}W_{ik}\bar\Delta_q)\big\}\Big].\end{aligned} \tag{5.26}$$

Notice that $E(W_{i_1\cdots i_p}\Delta_j) = n^{-1}\alpha_{i_1\cdots i_pj}$,

$$\begin{aligned}n^2E(\bar\Delta_i\bar\Delta_j\bar\Delta_k\bar\Delta_l) &= \alpha_{ij}\alpha_{kl} + \alpha_{ik}\alpha_{jl} + \alpha_{il}\alpha_{jk} + O(n^{-1}),\\ n^2E(W_{i_1\cdots i_p}\bar\Delta_j\bar\Delta_k\bar\Delta_l) &= \alpha_{i_1\cdots i_pj}\alpha_{kl} + \alpha_{i_1\cdots i_pk}\alpha_{jl} + \alpha_{i_1\cdots i_pl}\alpha_{jk}\\ &\quad -\alpha_{i_1\cdots i_p}\alpha_{jkl} + O(n^{-1}).\end{aligned}$$

Therefore, the right-hand sides of (5.25) and (5.26) may be written as $\kappa_2 - n^{-1}\delta_2(z)$ and $\kappa_4 - n^{-1}\delta_4(z) + O(n^{-2})$, respectively, where δ_2 and δ_4 are even, quadratic polynomials. Consequently

$$P(U^* \le x) = P(U \le x) - n^{-1}\left\{\tfrac{1}{2}\delta_2(z)(\partial/\partial x) + \tfrac{1}{24}\delta_4(z)(\partial/\partial x)^3\right\}\phi(x) + o(n^{-1}) \tag{5.27}$$
$$= P(U \le x) + (x/4n)\left\{2\delta_2(z) + \tfrac{1}{6}\delta_4(z)(x^2 - 3)\right\}\phi(x) + o(n^{-1})$$

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 $n^{1/2}g(\overline{\mathbf{Y}}|\mu) - \sum_{r=1}^{2} n^{-r/2}\Pi_{2r}(z)$. Therefore, (5.23) follows on taking $x = z$ in (5.27), and noting that

$$\psi(z) = \tfrac{1}{4}z\left\{2\delta_2(z) + \tfrac{1}{6}\delta_4(z)(z^2 - 3)\right\}. \tag{5.28}$$ □

REFERENCES

[1] Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.
[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] Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72** 45–58.
[11] Efron, B. (1986). Better bootstrap confidence intervals (with discussion). To appear in *J. Amer. Statist. Assoc.*
[12] Hall, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* **11** 569–576.
[13] Hall, P. (1985). A tabular method for correcting skewness. *Math. Proc. Cambridge Phil. Soc.* **97** 525–540.
[14] Pfanzagl, J. (1979). Nonparametric minimum contrast estimators. In *Selecta Statistica Canadiana* **5** 105–140. Dept. of Mathematics, McMaster Univ., Hamilton, Ontario.
[15] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.
[16] Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11** 577–587.
[17] Withers, C. S. (1984). Asymptotic expansions for distributions and quantiles with power series cumulants. *J. Roy. Statist. Soc. Ser. B* **46** 389–396.

Department of Statistics
Faculty of Economics and Commerce
Australian National University
GPO Box 4
Canberra, ACT 2601
Australia