The Annals of Statistics
1983, Vol. 11, No. 2, 569–576

INVERTING AN EDGEWORTH EXPANSION

BY PETER HALL

Australian National University

We provide a method for inverting a general Edgeworth expansion, so as to correct a statistic for the effects of non-normality. This technique is applied to the special case of the "Studentized" mean. Explicit formulae are given for the correction terms.

1. Introduction. The general theory of Edgeworth and Edgeworth-type expansions has been described by many authors; see for example Wallace (1958), Chambers (1967), Chibishov (1972, 1973), Bhattacharya and Rao (1976), Bhattacharya and Ghosh (1978), Barndorff-Nielsen and Cox (1979), Ibragimov and Khas'minskii (1979) and Pedersen (1979). The work of Bhattacharya and Ghosh (1978) provided a rigorous and very general development of the somewhat heuristic results of earlier authors. Suppose the statistic $\hat{\theta} = \hat{\theta}_n$ admits the Edgeworth expansion

$$P\{n^{1/2}(\hat{\theta}_n - \theta) \leq x\} = \Phi(x) + n^{1/2}\xi_{11}(x)\phi(x) + n^{-1}\xi_{12}(x)\phi(x) + \cdots, \tag{1.1}$$

where the functions ξ_{11}, ξ_{12}, $\cdots$ depend on characteristics of the underlying distribution of the sample values, but do not depend on n. Here Φ and ϕ denote the standard normal distribution and density functions, respectively. We have assumed that $n^{1/2}(\hat{\theta} - \theta)$ has unit asymptotic variance, but we shall show in the next section that this restriction may be removed. The expansion (1.1) leads to the "corrected" expansion

$$P\{n^{1/2}(\hat{\theta}_n - \theta) \leq x - n^{-1/2}\xi_{11}(x)\} = \Phi(x) + n^{-1}\xi_{22}(x)\phi(x) + \cdots, \tag{1.2}$$

at least in some local sense. (Condition (1.2) may not be valid uniformly in all x, even if (1.1) is valid uniformly in x.) Suppose we can estimate the function ξ_{11}, and that its estimate $\hat{\xi}_{11}$ behaves "reasonably," in that it admits an Edgeworth expansion:

$$P[n^{1/2}\{\hat{\xi}_{11}(x) - \xi_{11}(x)\} \leq y] = \Phi\{y/\sigma_1(x)\} + n^{-1/2}\eta_1(x, y) + \cdots.$$

Then it is at least plausible that (1.2) implies

$$P\{n^{1/2}(\hat{\theta}_n - \theta) \leq x - n^{-1/2}\hat{\xi}_{11}(x)\} = \Phi(x) + n^{-1}\xi_{32}(x)\phi(x) + n^{-3/2}\xi_{33}(x)\phi(x) + \cdots, \tag{1.3}$$

where (in general) the function ξ_{32} in (1.3) is different from ξ_{22} appearing in (1.2). Of course, (1.3) leads to another "corrected" expansion, an analogue of (1.2):

$$P\{n^{1/2}(\hat{\theta}_n - \theta) \leq x - n^{-1/2}\hat{\xi}_{11}(x) - n^{-1}\xi_{32}(x)\} = \Phi(x) + n^{-3/2}\xi_{43}(x) + \cdots.$$

(This expansion requires smoothness assumptions about the function $\hat{\xi}_{11}$.) The function ξ_{32} may be replaced by an estimate, and this procedure iterated *ad infinitum*, to produce an approximation of the form

$$P\{n^{1/2}(\hat{\theta}_n - \theta) \leq x - n^{-1/2}\hat{\eta}_1(x) \cdots - n^{-(k-1)/2}\hat{\eta}_{k-1}(x)\} = \Phi(x) + O(n^{-k/2}). \tag{1.4}$$

This result might be interpreted as an "inversion" of the Edgeworth expansion (1.1).

In practice the number of correction terms which can be used depends on at least two factors. First of all, if too many corrections are incorporated and the sample size, n, is not sufficiently large, then overcorrection may occur, resulting in a *worse* normal approximation than would have been obtained without the correction. See Section 3 for an example

Received December 1981; revised November 1982.
AMS 1980 *subject classification.* Primary 62E20, 62H10; secondary 60F05, 60G05.
Key words and phrases. Confidence interval, Edgeworth expansion, normal approximation.

of this phenomenon. Secondly, there are considerable algebraic difficulties in deriving an explicit form for the function $\hat{\eta}_j$ for large j, at least in a general context. For these reasons we shall not proceed beyond $k = 3$ in deriving the formula (1.4). In the case of the "Studentized" mean, taking $k = 3$ permits correction for the primary and secondary effects of skewness, and the primary effect of kurtosis.

Let us briefly discuss the way in which the formula (1.4) might be used. Suppose we wish to construct a one-sided confidence interval for θ, of the form $\mathscr{I} = (\hat{\lambda}, \infty)$, which covers θ with (unconditional) probability very nearly equal to 0.95. If $x_0 = \Phi^{-1}(0.95) = 1.64485 \cdots$ is the 95% point of the standard normal distribution, and if we take $\hat{\lambda} = \hat{\lambda}_1 = \hat{\theta}_n - n^{-1/2}x_0$, then $\mathscr{I}$ will cover θ with probability equal to $0.95 + O(n^{-1/2})$. But if we define instead

$$\hat{\lambda} = \hat{\lambda}_2 = \hat{\theta}_n - n^{-1/2}\{x_0 - n^{-1/2}\hat{\eta}_1(x_0) - n^{-1}\hat{\eta}_2(x_0)\},$$

the coverage probability will equal $0.95 + O(n^{-3/2})$, and so for large samples will be closer to 0.95. In many practical situations the sign of $\hat{\eta}_1(x_0)$ will equal the sign of the skewness of the underlying distribution.

In the case of a two-sided interval, a simple correction of the type leading to (1.3) yields a coverage probability with an error of only $O(n^{-3/2})$. Better approximations may be obtained by iteration.

In studying the asymptotic theory of inverse expansions in Section 2 we are guided by a detailed and rigorous account given by Bhattacharya and Ghosh (1978) of the theory of direct expansions. We examine the case of the "Studentized" mean in greatest detail, and prove other results which demonstrate the broad scope of the method of inverse expansion. In Section 3 we report on Monte Carlo trials of our procedure.

2. Asymptotic theory. Let $Y, Y_1, Y_2, \cdots$ be independent and identically distributed m-vectors, and $f_1, \cdots, f_k$ be real-valued Borel-measurable functions on $\mathbb{R}^m$. Define

$$Z = (f_1(Y), \cdots, f_k(Y)), \quad Z_i = (f_1(Y_i), \cdots, f_k(Y_i)), \qquad 1 \le i \le n,$$

$\bar{Z} = n^{-1}\sum_1^n Z_i$ and $\mu = E(Z) = E(\bar{Z})$. Let H be a real-valued Borel-measurable function on $\mathbb{R}^k$. Bhattacharya and Ghosh (1978) derived an Edgeworth expansion for the quantity $n^{1/2}\{H(\bar{Z}) - H(\mu)\}$. We shall begin by following their development.

Define $\alpha_{u_1,\ldots,u_p} = E[\prod_{i=1}^p \{f_{u_i}(Y) - \mu^{(u_i)}\}]$ for any vector $(u_1, \cdots, u_p) \subseteq \{1, \cdots, k\}^p$, provided the expectation exists. Note that $\alpha_u = 0$. Set

$$\ell_{u_1,\ldots,u_p} = (\partial^p/\partial x^{(u_1)} \cdots \partial x^{(u_p)})H(x)|_{x=\mu},$$

and let $\sigma^2 = \ell_u\ell_v\alpha_{uv}$, $a_1 = \frac{1}{2}\ell_{uv}\alpha_{uv}$ and $a_3 = \ell_u\ell_v\ell_w\alpha_{uvw} + 3\ell_u\ell_v\ell_{wx}\alpha_{uw}\alpha_{vx}$. (Here and below we use the summation notation, so that $\ell_u\ell_v\alpha_{uv}$ stands for $\Sigma_u \Sigma_v \ell_u\ell_v\alpha_{uv}$, etc.) The first term in an Edgeworth expansion may be described as follows.

LEMMA 1. *If H has three continuous derivatives in a neighbourhood of μ, if $E|f_u(Y)|^3 < \infty$ for each u and if the characteristic function ψ of Z satisfies*

$$\limsup_{\|t\|\to\infty} |\psi(t)| < 1,$$

then

$$P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le \sigma x] = \Phi(x) + n^{-1/2}\xi_{11}(x)\phi(x) + o(n^{-1/2}) \tag{2.1}$$

uniformly in x as $n \to \infty$, where $\xi_{11}(x) = -\{a_1/\sigma + \frac{1}{6}(a_3/\sigma^3)(x^2 - 1)\}$.

In many practical situations a statistic can be Studentized by dividing by an estimate of σ^2. For example, suppose the values of the ℓ_u's are known, and define

$$\hat{\alpha}_{uv} = n^{-1}\sum_{i=1}^n f_u(Y_i)f_v(Y_i) - \{n^{-1}\sum_{i=1}^n f_u(Y_i)\}\{n^{-1}\sum_{i=1}^n f_v(Y_i)\}.$$

Extend the vector Z by adjoining those products $f_u(Y)f_v(Y)$ for pairs (u, v) with $\ell_u\ell_v \neq 0$, giving rise to a new vector Z^*. If Z_n^* and $\bar{Z}^*$ are obtained from Z_n and $\bar{Z}$ by the same method of extension, we may write $\hat{\alpha}_{uv} = a_{uv}(\bar{Z}^*)$, where the function a_{uv} does not depend on n. The Studentized statistic $\{H(\bar{Z}) - H(\mu)\}/(\sum\sum \ell_u\ell_v\hat{\alpha}_{uv})^{1/2}$ may therefore be written as $K(\bar{Z}^*)$, where K does not depend on n. A version of Lemma 1 in which $\sigma^2 = 1$ and Z is replaced by Z^* can be stated for this statistic.

If $\sigma^2 = 1$ then (2.1) implies that

$$P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x - n^{-1/2}\xi_{11}(x)] = \Phi(x) + o(n^{-1/2}) \tag{2.2}$$

uniformly in x on compact intervals. The result (2.2) cannot be assumed to hold uniformly in all x, for if $a_3 < 0$ then the left hand side tends to zero as $x \to +\infty$, while the right hand side tends to $1 + o(n^{-1/2})$.

The parameters a_1 and a_3 can usually be estimated very easily, the precise form of the estimates depending on the nature of the problem under consideration. Our next result shows that a_1 and a_3 in (2.2) can often be replaced by these estimates.

THEOREM 1. *If the conditions of Lemma* 1 *hold, if* $\sigma^2 = 1$ *and*

$$P(|\hat{a}_1 - a_1| > \varepsilon) + P(|\hat{a}_3 - a_3| > \varepsilon) = o(n^{-1/2}) \tag{2.3}$$

for each $\varepsilon > 0$, *then with* $\hat{\xi}_{11}(x) = -\{\hat{a}_1 + \tfrac{1}{6}\hat{a}_3(x^2 - 1)\}$ *we have*

$$P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x - n^{-1/2}\hat{\xi}_{11}(x)] = \Phi(x) + o(n^{-1/2}) \tag{2.4}$$

uniformly in x *on compact intervals.*

Further refinement of the term $o(n^{-1/2})$ on the right hand side of (2.4) depends intimately on the form of the estimates $\hat{a}_1$ and $\hat{a}_2$. Let us consider initially the important special case of the Studentized mean. The problem of deriving asymptotic expansions or approximations for Student's t-statistic is quite old, dating back to the work of Pearson and Adyanthaya (1929) and Bartlett (1935). See also Geary (1936, 1947) and Gayen (1949). Some of this work has been reviewed by Bowman, Beauchamp and Shenton (1977) and Cressie (1980).

Suppose $Z = (Y, Y^2)$ for a scalar random variable Y, and define

$$H(z) = H(z^{(1)}, z^{(2)}) = (z^{(1)} - m)/\{z^{(2)} - (z^{(1)})^2\}^{1/2},$$

where $m = E(Y)$. Then $\mu = (m, s^2 + m^2)$ where $s^2 = \operatorname{var}(Y)$, and $\sigma^2 = \ell_1^2\alpha_{11} = 1$. It is common to normalize using the unbiased estimate of s^2, rather than the biased estimate implicit in the use of $H(z)$. Thus, $H(z)$ should really be replaced by $(1 - 1/n)^{1/2}H(z)$. However, this change does not alter the Edgeworth expansion up to terms of $o(n^{-1/2})$, and we may deduce directly from (2.2) that with $\hat{s} = \{(n-1)^{-1}\sum_1^n (Y_i - \bar{Y})^2\}^{1/2}$, we have, under appropriate moment and smoothness conditions,

$$P\{n^{1/2}(\bar{Y} - m)/\hat{s} \le x - n^{-1/2}\tfrac{1}{6}\lambda_3(1 + 2x^2)\} = \Phi(x) + o(n^{-1/2})$$

uniformly on compact intervals, where $\lambda_3 = E(Y - m)^3/s^3$. A convenient estimate of λ_3 is $\hat{\lambda}_3 = n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^3/\hat{s}^3$. If $E|Y|^{9/2} < \infty$ then it may be proved using results of Baum and Katz (1965) that for all $\varepsilon > 0$,

$$P(|n^{-1}\textstyle\sum_1^n Y_i^3 - EY^3| > \varepsilon) + P(|n^{-1}\sum_1^n Y_i^2 - EY^2| > \varepsilon) + P(|n^{-1}\sum_1^n Y_i - m| > \varepsilon) = o(n^{-1/2})$$

as $n \to \infty$, and consequently (2.3) holds. We may now deduce from Theorem 1 that if $E|Y|^6 < \infty$ and $\limsup_{|t_1|+|t_2|\to\infty} |E\exp(it_1Y + it_2Y^2)| < 1$,

$$P\{n^{1/2}(\bar{Y} - m)/\hat{s} \le x - n^{-1/2}\tfrac{1}{6}\hat{\lambda}_3(1 + 2x^2)\} = \Phi(x) + o(n^{-1/2}).$$

Our next result refines the remainder in this expression. Let

$$\xi_{11}(x) = \tfrac{1}{6}\lambda_3(1 + 2x^2)$$

and

$$\xi_{32}(x) = x\{\tfrac{1}{2}(x^2 + 2) - \lambda_4 \tfrac{1}{12}(3x^2 + 5) + \lambda_3^2 \tfrac{1}{72}(16x^2 + 23)\},$$

where $\lambda_4 = E(Y - m)^4/s^4$, and let $\hat{\xi}_{11}$ and $\hat{\xi}_{32}$ denote the same functions but with λ_3 and λ_4 replaced by $\hat{\lambda}_3$ and $\hat{\lambda}_4 = n^{-1}\sum_1^n (Y_i - \bar{Y})^4/\hat{s}^4$, respectively.

THEOREM 2. *If* $E(Y^{12}) < \infty$ *and* $\limsup_{|r|+|s|+|t|\to\infty} |E\exp(irY + isY^2 + itY^3)| < 1$ *then*

$$P\{n^{1/2}(\bar{Y} - m)/\hat{s} \le x - n^{-1/2}\hat{\xi}_{11}(x)\} = \Phi(x) + n^{-1}\xi_{32}(x)\phi(x) + o(n^{-1})$$

and

$$P\{n^{1/2}(\bar{Y} - m)/\hat{s} \le x - n^{-1/2}\hat{\xi}_{11}(x) - n^{-1}\hat{\xi}_{32}(x)\} = \Phi(x) + o(n^{-1}) \tag{2.5}$$

uniformly on compact intervals.

It does not seem possible to state a single result which covers the majority of inverse expansions like (2.5). The form of the estimates $\hat{a}_1$ and $\hat{a}_3$ in (2.4) will vary, depending on the amount of information available about the quantities ℓ_u, ℓ_{uv}, α_{uv} and α_{uvw}. This leads to a wide range of possible forms for (2.5). Our aim in the next result is to provide some idea of the broad scope of the method of Edgeworth inversion. The conditions imposed in this theorem, and the proof we shall give in Section 4, are typical of those in a great many similar situations.

Let us assume for the sake of definiteness that the derivatives ℓ_u and ℓ_{uv} are known, but the quantities α_{uv} and α_{uvw} are unknown. Let $k_1 = k(k+1)/2$, and $\{(u_{11}, v_{11}), \cdots, (u_{1k_1}, v_{1k_1})\}$ denote the set of k_1 *unordered* pairs (u, v) with $1 \le u, v \le k$. Suppose there are k_2 *unordered* triples (u, v, w) such that $\ell_u\ell_v\ell_w \ne 0$, and let these be $\{(u_{21}, v_{21}, w_{21}), \cdots, (u_{2k_2}, v_{2k_2}, w_{2k_2})\}$. We extend the k-vector Z to a vector $Z^* = (Z^{(j)})$ of length $k_3 = k + k_1 + k_2$, whose elements are $Z^{(j)} = f_j(Y)$ for $1 \le j \le k$,

$$Z^{(k+j)} = \{f_{u_{1j}}(Y) - \mu^{(u_{1j})}\}\{f_{v_{1j}}(Y) - \mu^{(v_{1j})}\}, \qquad 1 \le j \le k_1,$$

and

$$Z^{(k+k_1+j)} = \{f_{u_{2j}}(Y) - \mu^{(u_{2j})}\}\{f_{v_{2j}}(Y) - \mu^{(v_{2j})}\}\{f_{w_{2j}}(Y) - \mu^{(w_{2j})}\}, \qquad 1 \le j \le k_2.$$

This type of extension will be denoted by the symbol *. Thus, if $Z_i = (Z_i^{(j)})$, $1 \le i \le n$, is a sequence of k-vectors, then $(\bar{Z}^*)^{(1)} = n^{-1}\sum_1^n Z_i^{(1)}$ and $(\bar{Z}^*)^{(k+j)} = n^{-1}\sum_{i=1}^n \{Z_i^{(u_{1j})} - \mu^{(u_{1j})}\} \cdot \{Z_i^{(v_{1j})} - \mu^{(v_{1j})}\}$ for $1 \le j \le k_1$.

An estimator of $\alpha_{u_1,\cdots,u_p}$ is given by

$$\hat{\alpha}_{u_1,\cdots,u_p} = n^{-1}\sum_{i=1}^n \prod_{j=1}^p \{f_{u_j}(Y_i) - \bar{Z}^{(u_j)}\},$$

and we may define $\hat{a}_1$ and $\hat{a}_3$ by replacing α_{uv} and α_{uvw} by $\hat{\alpha}_{uv}$ and $\hat{\alpha}_{uvw}$, respectively, in the formulae for a_1 and a_3. The statistic $H(\bar{Z})$ and the correction terms $\hat{a}_1$ and $\hat{a}_2$ are all functions of $\bar{Z}^*$.

Define the quantities

$$a_2 = a_2(x) = \tfrac{1}{2}\ell_{pq}\ell_{rs}\alpha_{pr}\alpha_{qs} + \ell_p\ell_{qrs}\alpha_{pq}\alpha_{rs} + (x^2 - 1)$$
$$- (x^2 - 1)\{\tfrac{1}{3}\ell_p\ell_q\ell_r\ell_s\alpha_{pqrs} + \ell_p\ell_{qr}\alpha_{pqr} + (\ell_{pq}\alpha_{pq})(\ell_p\ell_q\ell_r\alpha_{pqr})\},$$

$$a_4 = \ell_p\ell_q\ell_r\ell_s\alpha_{pqrs} - 3 + 12\ell_p\ell_q\ell_r\ell_{st}\alpha_{ps}\alpha_{qrt} + 12\ell_p\ell_q\ell_{rs}\ell_{tu}\alpha_{pr}\alpha_{qt}\alpha_{su} + 4\ell_p\ell_q\ell_r\ell_{stu}\alpha_{ps}\alpha_{qt}\alpha_{ru}.$$

These may be estimated by replacing the α's by their estimates.

THEOREM 3. *If H has four continuous derivatives in a neighbourhood of μ, if $\sigma^2 = 1$, $E\|Z^*\|^4 < \infty$ and the characteristic function β of Z^* satisfies* $\limsup_{\|t^*\|\to\infty} |\beta(t^*)| < 1$, *then*

$$P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x + n^{-1/2}\{\hat{a}_1 + \tfrac{1}{6}\hat{a}_3(x^2 - 1)\}]$$
$$= \Phi(x) - n^{-1}x\{\tfrac{1}{2}a_2 + \tfrac{1}{24}a_4(x^2 - 3) \tag{2.6}$$
$$+ \tfrac{1}{3}a_1a_3(x^2 - 3) + \tfrac{1}{36}a_3^2(2x^4 - 10x^2 + 11)\} + o(n^{-1})$$

uniformly in x on compact intervals. If in addition $E|f_p(Y)f_q(Y)f_r(Y)|^2 < \infty$ for all triples (p, q, r), and $E|f_p(Y)f_q(Y)f_r(Y)f_s(Y)|^2 < \infty$ for all quadruples (p, q, r, s) with $\ell_p\ell_q\ell_r\ell_s \ne 0$, then

$$P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x + n^{-1/2}\{\hat{a}_1 + \tfrac{1}{6}\hat{a}_3(x^2 - 1)\}$$
$$+ n^{-1}x\{\tfrac{1}{2}\hat{a}_2 + \tfrac{1}{24}\hat{a}_4(x^2 - 3) + \tfrac{1}{3}\hat{a}_1\hat{a}_3(x^2 - 3) + \tfrac{1}{36}\hat{a}_3^2(2x^4 - 10x^2 + 11)\}] \tag{2.7}$$
$$= \Phi(x) + o(n^{-1})$$

uniformly in x on compact intervals.

3. Monte Carlo trials. We shall report on a series of Monte Carlo trials involving the χ_4^2 and exponential distributions. If we take $x = -1.644854$ (the lower 5% point of the standard normal distribution) in the formulae of Theorems 1 and 2, we may compute the following end points of confidence intervals:

$$\hat{\ell}_1 = \bar{Y} + n^{-1/2}1.64485\hat{s}; \qquad \hat{\ell}_2 = \bar{Y} + n^{-1/2}\hat{s}(1.64485 + n^{-1/2}1.065\hat{\lambda}_3);$$

$$\hat{\ell}_3 = \bar{Y} + n^{-1/2}\hat{s}\{1.64485 + n^{-1/2}1.0685\hat{\lambda}_3 - n^{-1}(3.8700 + 1.5144\hat{\lambda}_3^2 - 1.7979\hat{\lambda}_4)\}.$$

The interval $(-\infty, \hat{\ell}_j)$ covers the mean, m, with probability $0.95 + O(n^{-j/2})$.

The Monte Carlo trials are summarised in Table 1. They suggest that long inverse expansions may lead to over-correction when the sample size is small, but can provide an improvement when the sample is large.

4. Proofs.

PROOF OF LEMMA 1. The existence of the expansion follows from Theorem 2 of Bhattacharya and Ghosh (1978). It is necessary only to compute the form of the term of order $n^{-1/2}$. Let us define

$$V_{u_1,\ldots,u_p} = n^{-1/2}\sum_{i=1}^{n}[\prod_{j=1}^{p}\{f_{u_j}(Y_i) - \mu^{(u_j)}\} - \alpha_{u_1,\ldots,u_p}].$$

Bhattacharya and Ghosh's "Taylor expansion" of $n^{1/2}\{H(\bar{Z}) - H(\mu)\}$ is given by

$$W_n' = \ell_u V_u + n^{1/2}(1/2)\ell_{uv}V_uV_v,$$

TABLE 1

Simulated coverage probabilities. n = sample size and p_i = proportion of simulations in which the interval $(-\infty, \hat{\ell}_i)$ covered the mean m. (5×10^3 simulations were used in the χ_4^2 case, with 10^3 simulations in the exponential case.)

	χ_4^2 ($m = 4$)				exponential ($m = 1$)			
n	10	15	20	50	10	15	20	50
p_1	0.874	0.898	0.904	0.928	0.849	0.873	0.877	0.906
p_2	0.895	0.915	0.922	0.945	0.860	0.885	0.896	0.924
p_3	0.886	0.915	0.923	0.947	0.853	0.879	0.900	0.929

and if sufficiently many moments are finite then

$$E(W'_n) = n^{-1/2}(1/2)\ell_{uv}\alpha_{uv}, \qquad E(W'^2_n) = \sigma^2 + O(n^{-1})$$

and

$$E(W'^3_n) = n^{-1/2}\{\ell_u\ell_v\ell_w\alpha_{uvw} + \tfrac{3}{2}\ell_u\ell_v\ell_{wx}(\alpha_{uv}\alpha_{wx} + \alpha_{uw}\alpha_{vx} + \alpha_{ux}\alpha_{vw})\} + O(n^{-3/2})$$

as $n \to \infty$. Therefore the cumulants of W'_n are given by $\kappa_{1n} = n^{-1/2}a_1 + o(n^{-1/2})$, $\kappa_{2n} = \sigma^2 + o(n^{-1/2})$ and $\kappa_{3n} = n^{-1/2}a_3 + o(n^{-1/2})$. The first term in the Edgeworth expansion (2.2) is computed by inverting the first term in an expansion of the characteristic function of W'_n.

PROOF OF THEOREM 1. In view of (2.3) we may choose $\varepsilon(n)$ decreasing slowly to zero with n and such that (2.3) continues to hold if the constant ε is replaced by the function $\varepsilon(n)$. Therefore on the interval $[-\lambda, \lambda]$ the left hand side of (2.4) is dominated by

$$\begin{aligned}
&P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x + n^{-1/2}\{a_1 + \tfrac{1}{6}a_3(x^2 - 1)\} + n^{-1/2}\varepsilon(n)(\lambda^2 + 2)] \\
&\quad + P\{|\hat{a}_1 - a_1| > \varepsilon(n)\} + P\{|\hat{a}_3 - a_3| > \varepsilon(n)\} \\
&\qquad = P[n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x(n) + n^{-1/2}\{a_1 + \tfrac{1}{6}a_3(x^2(n) - 1)\}] + o(n^{-1/2}) \\
&\qquad = \Phi\{x(n)\} + o(n^{-1/2}) = \Phi(x) + o(n^{-1/2}),
\end{aligned}$$

using (2.2), where $x(n)$ satisfies $\sup_{|x|\le\lambda}|x - x(n)| = o(n^{-1/2})$. A lower bound may be obtained in the same way, and so the theorem is proved.

Theorem 2 is proved in the manner of Theorem 3.

PROOF OF THEOREM 3. In the notation preceding the theorem, define $U^*_n = n^{1/2}(\bar{Z}^* - \mu^*)$. It may be deduced from the definitions of the estimators $\hat{a}_1$ and $\hat{a}_2$ that we may write

$$\hat{a}_1 = a_1 + n^{-1/2}A_{11}(U^*_n) + n^{-1}A_{12}(U^*_n)$$

and

$$\hat{a}_2 = a_2 + n^{-1/2}A_{21}(U^*_n) + n^{-1}A_{22}(U^*_n) + n^{-3/2}A_{23}(U^*_n) + n^{-2}A_{24}(U^*_n),$$

where the functions A_{11}, A_{12}, A_{21}, A_{22}, A_{23} and A_{24} are of k_3 variables, are infinitely differentiable everywhere and do not depend on n. Given a real number x, define the function $g_n(\cdot) = g_n(\cdot, x)$ of k_3 variables by

$$\begin{aligned}
g_n(z^*) &= n^{1/2}\{H(\mu + n^{-1/2}z) - H(\mu)\} - [x + n^{-1/2}\{a_1 + \tfrac{1}{6}a_3(x^2 - 1)\}] \\
&\quad - n^{-1}\{A_{11}(z^*) + A_{21}(z^*)\tfrac{1}{6}(x^2 - 1)\} - n^{-3/2}\{A_{12}(z^*) + A_{22}(z^*)\tfrac{1}{6}(x^2 - 1)\} \\
&\quad - n^{-2}\{A_{23}(z^*) + n^{-1/2}A_{24}(z^*)\}\tfrac{1}{6}(x^2 - 1).
\end{aligned}$$

Then the inequality

$$n^{1/2}\{H(\bar{Z}) - H(\mu)\} \le x + n^{-1/2}\{\hat{a}_1 + \tfrac{1}{6}\hat{a}_3(x^2 - 1)\}$$

may be rewritten as $g_n(U^*_n) \le 0$.

Our function g_n replaces that defined on the line above (2.4) in Bhattacharya and Ghosh (1978). Bhattacharya and Ghosh's proof of their Theorem 2, part (b), may be reworked with only minor modifications. The uniformity derived by Bhattacharya and Ghosh over Borel sets must in our case be transferred to a uniformity over the functions g_n, indexed by $x \in [-\lambda, \lambda]$. As far as *existence* of the expansion goes, this transferral is readily accomplished by using Corollary 20.2, page 214 of Bhattacharya and Rao (1976), where the function f should be taken equal to the indicator of the event $\{z^* : g_n(z^*; x) \le 0\}$. The bound on the right hand side of Bhattacharya and Rao's (20.44) may be shown to equal

$o(n^{-1})$ uniformly in $x \in [-\lambda, \lambda]$, using analogues of (2.18) and (2.19) of Bhattacharya and Ghosh (1978). *Identification* of the expansion may be achieved by invoking an analogue of Bhattacharya and Ghosh's Lemma 2.1, and using an algebraic argument which we shall now describe.

We first obtain a "Taylor Expansion" of $g(U_n^*)$. This contains an analogue of Bhattacharya and Ghosh's W_n', which we call W_n''. It can be shown that

$$g(U_n^*) = W_n'' - [x + n^{-1/2}\{a_1 + \tfrac{1}{6}a_3(x^2 - 1)\}] + o_p(n^{-1}) \tag{4.1}$$

where

$$\begin{aligned} W_n'' = {} & \ell_p V_p + n^{-1/2}(\tfrac{1}{2})\ell_{pq} V_p V_q + n^{-1}(\tfrac{1}{6})\ell_{pqr} V_p V_q V_r \\ & - n^{-1}\{\tfrac{1}{2}\ell_{pq} V_{pq} + \tfrac{1}{6}(x^2 - 1) \times \ell_p \ell_q \ell_r (V_{pqr} - V_p \alpha_{qr} - V_q \alpha_{pr} - V_r \alpha_{pq}) \\ & + \tfrac{1}{2}(x^2 - 1)\ell_p \ell_q \ell_{rs}(\alpha_{pq} V_{rs} + \alpha_{rs} V_{pq})\}. \end{aligned}$$

If Z has sufficiently many moments finite then

$$E(V_p V_q) = \alpha_{pq}, \qquad E(V_p V_q V_r) = n^{-1/2}\alpha_{pqr},$$

$$E(V_p V_q V_r V_s) = (\alpha_{pq}\alpha_{rs} + \alpha_{pr}\alpha_{qs} + \alpha_{ps}\alpha_{qr}) + n^{-1}\{\alpha_{pqrs} - (\alpha_{pq}\alpha_{rs} + \alpha_{pr}\alpha_{qs} + \alpha_{ps}\alpha_{qr})\},$$

$$E(V_p V_q V_r V_s V_t) = n^{-1/2}(\alpha_{pq}\alpha_{rst} + \cdots + \alpha_{st}\alpha_{pqr}) + O(n^{-3.2})$$

(a total of ten terms covering all permutations within the brackets),

$$E(V_p V_q V_r V_s V_t V_u) = (\alpha_{pq}\alpha_{rs}\alpha_{tu} + \cdots + \alpha_{pt}\alpha_{qr}\alpha_{su}) + O(n^{-1})$$

(a total of fifteen terms within the brackets),

$$E(V_p V_{qr}) = \alpha_{pqr}, \qquad E(V_p V_{qrs}) = \alpha_{pqrs},$$

$$E(V_p V_q V_r V_{st}) = \alpha_{pq}\alpha_{rst} + \alpha_{pr}\alpha_{qst} + \alpha_{qr}\alpha_{pst} + O(n^{-1/2}),$$

$$E(V_p V_q V_r V_{stu}) = \alpha_{pq}\alpha_{rstu} + \alpha_{pr}\alpha_{qstu} + \alpha_{qr}\alpha_{pstu} + O(n^{-1/2})$$

and

$$| E(V_p V_q V_r V_s V_t V_u V_v)| + | E(V_p V_q V_{rs})| + | E(V_p V_q V_{rs} V_{rst})| = O(n^{-1/2}),$$

for *all* values of p, q, r, s, t, u, v. It may now be proved after some tedious algebra that the cumulants of W_n'' are given by $\kappa_{1n} = n^{-1/2}a_1 + o(n^{-1})$, $\kappa_{2n} = 1 + n^{-1}a_2 + o(n^{-1})$, $\kappa_{3n} = n^{-1/2}a_3 + o(n^{-1})$ and $\kappa_{4n} = n^{-1}a_4 + o(n^{-1})$. Consequently W_n'' has characteristic function given by

$$\begin{aligned} e^{-t^2/2}[1 & + n^{-1/2}\{ita_1 + \tfrac{1}{6}(it)^3 a_3\} \\ & + n^{-1}\{\tfrac{1}{2}(it)^2(a_2 + a_1^2) + \tfrac{1}{24}(it)^4(a_4 + 4a_1a_3) + \tfrac{1}{72}(it)^6 a_3^2\} + o(n^{-1})], \end{aligned}$$

and Edgeworth expansion given by

$$\begin{aligned} P(W_n'' \le y) = {} & \Phi(y) - n^{-1/2}\{a_1 + \tfrac{1}{6} a_3(y^2 - 1)\}\phi(y) \\ & - n^{-1}y\{\tfrac{1}{2}(a_2 + a_1^2) + \tfrac{1}{24}(a_4 + 4a_1a_3)(y^2 - 3) \\ & + \tfrac{1}{72}a_3^2(y^4 - 10y^2 + 15)\}\phi(y) + o(n^{-1}). \end{aligned} \tag{4.2}$$

On noting (4.1) we see that the identification of the Edgeworth expansion of $P\{g(U_n^*) \le 0\}$ is completed by setting $y = x + n^{-1/2}\{a_1 + \tfrac{1}{6}a_3(y^2 - 1)\}$ in (4.2). This closes the proof of the first part of the theorem.

For the second part of the theorem, it follows from a result of Baum and Katz (1965) that under the given conditions,

$$\textstyle\sum_{i=1}^4 P(| \hat{a}_i - a_i | > \varepsilon) = o(n^{-1})$$

uniformly in $|x| \leq \lambda$, for each $\varepsilon > 0$. This represents an analogue of (2.3), and the proof may now be completed as was that of Theorem 1.

Acknowledgment. I am grateful to Professor C. R. Heathcote and Dr. C. C. Heyde for helpful discussions.

REFERENCES

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with Discussion). *J. Royal Statist. Soc.* **41** 279–312.

BARTLETT, M. S. (1935). The effect of non-normality on the *t* distribution. *Proc. Cambridge Philos. Soc.* **31** 223–231.

BAUM, L. E. and KATZ, M. (1965). Convergence rates in the law of large numbers. *Trans. Amer. Math Soc.* **120** 108–123.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. Corrigendum *ibid* **8** 1399.

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

BOWMAN, K., BEAUCHAMP, J. and SHENTON, L. (1977). The distribution of the *t*-statistic under non-normality. *Internat. Statist. Rev.* **45** 233–242.

CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54** 367–383.

CHIBISHOV, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theor. Probability Appl.* **17** 620–630.

CHIBISHOV, D. M. (1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theor. Probability Appl.* **18** 295–303.

CRESSIE, N. (1980). Relaxing assumptions in the one-sample *t*-test. *Aust. J. Statist.* **22** 143–153.

GAYEN, A. K. (1949). The distribution of "Student's" *t* in random samples of any size drawn from non-normal universes. *Biometrika* **36** 353–369.

GEARY, R. C. (1936). The distribution of Student's ratio for non-normal samples. *J. Royal Statist. Soc.* Suppl. **3** 178–184.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1979). *Asymptotic Theory of Estimation.* Nauka, Moscow.

PEARSON, E. S. and ADYANTHAYA, N. K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika* **21** 259–286.

PEDERSEN, B. V. (1979). Approximating conditional distributions by the mixed Edgeworth-saddlepoint expansion. *Biometrika* **66** 597–604.

WALLACE, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
BOX 4
CANBERRA ACT 2600
AUSTRALIA