The Annals of Statistics
1995, Vol. 23, No. 2, 363–375

UNIFORM COVERAGE BOUNDS FOR CONFIDENCE INTERVALS AND BERRY–ESSEEN THEOREMS FOR EDGEWORTH EXPANSION

BY PETER HALL AND BING-YI JING

Australian National University

We derive upper bounds for the coverage error of confidence intervals for a population mean uniformly over large classes of populations and different types of confidence intervals. It is shown that the order of these bounds is achieved by the normal approximation method for constructing confidence intervals, uniformly over distributions with finite third moment, and, by an empirical Edgeworth correction of this approach, uniformly over smooth distributions with finite fourth moments. These results have straightforward extensions to higher orders of Edgeworth correction and higher orders of moments. Our upper bounds to coverage accuracy are based on Berry–Esseen theorems for Edgeworth expansions of the distribution of the Studentized mean.

1. Introduction. Work on bootstrap and related methods during the last decade has produced a wide variety of new techniques for constructing confidence intervals, some of them very accurate and sophisticated. However, this accuracy is obtained only at a price, usually expressed in terms of moment assumptions about the sampling distribution. In the present paper we address optimality issues in the study of confidence intervals for a population mean. We prove that in a well-defined sense and for one-sided confidence intervals, the normal approximation method provides the best possible order of coverage accuracy uniformly over a large class of distributions with finite third moments, and an Edgeworth-corrected normal approximation method does the same among smooth distributions with finite fourth moments. Similarly, higher-order Edgeworth corrections may be shown to be optimal among smooth distributions with higher-order moments. These results are based on a new approach to identifying large classes of confidence regions and on new methods for developing high-order bounds to coverage probabilities over large sets of distributions.

Results of this type are sometimes referred to as "minimax," in that they describe the best possible performance (i.e., "min" error) over all possible estimators, across a large class of models (i.e., "max" worst-case performance among different distributions). In particular, Section 2 shows that the maximum of coverage error over all confidence interval types and all distributions must be at least of a certain order, and that certain specific interval types

Received December 1992; revised March 1994.
AMS 1991 *subject classifications*. Primary 62G15; secondary 62E20.
Key words and phrases. Berry–Esseen bound, bootstrap, confidence interval, coverage error, Edgeworth expansion, minimax bound, skewness.

achieve this order of accuracy. Section 3 introduces new theoretical results that provide the basis for these conclusions. These take the form of Berry–Esseen bounds for Edgeworth expansions of the distribution of the Studentized mean and for adjusted versions of that distribution. The proofs of those results employ new ways of incorporating a concise description of distribution smoothness into explicit constants for the bounds. All proofs are deferred to Section 4.

Edgeworth correction methods for constructing confidence intervals have been developed by Pfanzagl (1979), Hall (1983) and Abramovitch and Singh (1985). Edgeworth expansion theory, and its application to bootstrap and other confidence interval methods, have been described by Hall (1992). A Berry–Esseen theorem for normal approximation of the distribution of the Studentized mean has been provided by Slavova (1985). Edgeworth expansions of this distribution, under specific moment conditions, have been discussed by Chibishov (1984) and Hall (1987).

2. Coverage errors of confidence intervals.

2.1. *Introduction and summary.* Section 2.2 introduces upper bounds for the coverage errors of confidence intervals, uniformly over large classes of distributions. The orders of magnitude of these errors are of course not new, but the uniformity is. This is the critical feature of our work, since it provides the essential complement to the lower bounds described in Section 2.3. Together, the upper bounds in Section 2.2 and lower bounds in Section 2.3 show that in the case of one-sided confidence intervals, normal approximation methods are optimal over a large class of distributions with finite third moments, and that Edgeworth-corrected confidence intervals are optimal among smooth distributions with finite fourth moments. Here, "optimality" refers to the order of magnitude of coverage error, which is $n^{-1/2}$ in the former case and n^{-1} in the latter.

Let $X, X_1, X_2, \ldots$ denote independent and identically distributed random variables with finite third moment. Write $\bar{X} = n^{-1} \sum_{i \le n} X_i$, $\hat{\sigma}^2 = n^{-1} \sum_{i \le n} (X_i - \bar{X})^2$ and $\hat{\gamma} = \hat{\sigma}^{-3} n^{-1} \sum_{i \le n} (X_i - \bar{X})^3$ for estimators of $\mu = E(X)$, $\sigma^2 = \operatorname{var}(X)$ and $\gamma = E(X - \mu)^3/\sigma^3$, respectively. It will prove convenient to append the subscript F to μ when the expectation $E(X) = \mu$ is taken under a specific distribution F.

2.2. *Upper bound to coverage error.* First we show that for $j = 1$ or 2 the usual j-sided confidence interval for a mean based on the normal approximation has coverage error of order $n^{-j/2}$ uniformly over a large class of distributions with finite $(j + 1)$th order moments. Then we treat one-sided confidence intervals based on empirical Edgeworth correction and show that coverage errors there are $O(n^{-1})$ in a uniform sense.

We begin by defining confidence intervals based on normal approximation. Let $z_\alpha = \Phi^{-1}(\alpha)$ denote the α-level point of the standard normal distribution and put $\beta = (1 + \alpha)/2$. The one-sided interval is $I_{1,\alpha} = (-\infty, \bar{X} + n^{-1/2}\hat{\sigma} z_\alpha)$

or $I_{1,\alpha} = (\bar{X} - n^{-1/2}\hat{\sigma} z_\alpha,\ \infty)$; our main theorem applies equally to either definition. The two-sided interval is $I_{2,\alpha} = (\bar{X} - n^{-1/2}\hat{\sigma} z_\beta,\ \hat{X} + n^{-1/2}\hat{\sigma} z_\beta)$. Assuming only that the sampling distribution has finite variance, these intervals cover the true mean $\mu = \mu_F$ with probability $\alpha + o(1)$ as $n \to \infty$.

Next we describe our classes of distributions. Let $\mathscr{F}_{1,d}$ denote the class of all distributions of X with finite third moment satisfying

$$E|X - EX|^3 (\operatorname{var} X)^{-3/2} \le d.$$

Let $\mathscr{F}_{2,d}$ be the class of all distributions of X that have finite fourth moment and which satisfy:

for some $y \in (-\infty, \infty)$ and $c_1, c_2 > 0$,

$$\liminf_{h \downarrow 0} h^{-1} P(|X - EX + x - y| \le h) \ge c_1, \quad \text{all } |x| \le c_2,$$

and

$$\left(c_1 c_2^3\right)^{-2} E|X - EX - y|^4 + \left(c_1 c_2^3\right)^{-3} \left(E|X - EX - y|^3\right)^2 \le d.$$

Both distribution classes are nonempty if d is sufficiently large.

Our specification of $\mathscr{F}_{1,d}$ is quite straightforward, but $\mathscr{F}_{2,d}$ perhaps requires a little elaboration. To describe an important subclass of $\mathscr{F}_{2,d}$, let F_1 denote any absolutely continuous distribution with finite fourth moment, whose density f_1 is bounded away from zero in a neighborhood of the origin, say, $f_1(x) \ge c_1'$ for $|x| \le c_2$. Fix $\pi_1 \in (0, 1)$ and let $\mathscr{F}_{2,d}'$ denote the class of distributions F which may be represented as convex combinations $F = \pi F_1 + (1 - \pi) F_2$, with $\pi \in [\pi_1, 1]$ and, defining $c_1 = 2c_1'\pi$,

$$\left(c_1 c_2^3\right)^{-2} E_F|X - E_F X|^4 + \left(c_1 c_2^3\right)^{-3} \left(E_F|X - E_F X|^3\right)^2 \le d.$$

Then $\mathscr{F}_{2,d}' \subseteq \mathscr{F}_{2,d}$ and so in the case $j = 2$, the bound at (2.1) below applies to $\mathscr{F}_{2,d}'$ as well as to $\mathscr{F}_{2,d}$.

THEOREM 2.1. *There exists a constant $B_d > 0$, depending only on d, such that*

$$\sup_{F \in \mathscr{F}_{j,d}} \sup_{0 < \alpha < 1} |P_F(\mu_F \in I_{j,\alpha}) - \alpha| \le B_d n^{-j/2} \tag{2.1}$$

for $j = 1, 2$ and $n \ge 2$.

Proofs of all theorems are deferred to Section 4.

Next we treat Edgeworth-corrected confidence intervals of the type studied by Hall (1983). See also Pfanzagl (1979) and Abramovitch and Singh (1985). Nominal α-level one-sided intervals are given by

$$J_{1,\alpha} = \left(-\infty,\ \bar{X} + n^{-1/2}\hat{\sigma}\left\{z_\alpha - n^{-1/2}\tfrac{1}{6}\hat{\gamma}\left(2z_\alpha^2 + 1\right)\right\}\right)$$

and

$$J_{1,\alpha} = \left(\bar{X} - n^{-1/2}\hat{\sigma}\left\{z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\left(2z_\alpha^2 + 1\right)\right\},\ \infty\right);$$

a two-sided interval is

$$J_{2,\alpha} = \Big(\hat{X} - n^{-1/2}\hat{\sigma}\Big\{z_\beta + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(2z_\beta^2 + 1\big)\Big\},$$
$$\bar{X} + n^{-1/2}\hat{\sigma}\Big\{z_\beta - n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(2z_\beta^2 + 1\big)\Big\}\Big).$$

THEOREM 2.2. *There exists a constant* $B_d' > 0$, *depending only on* d *and* $\varepsilon \in (0, \frac{1}{2})$, *such that*

$$\sup_{F \in \mathscr{F}_{j,d}} \sup_{\varepsilon < \alpha < 1-\varepsilon} |P_F(\mu_F \in J_{j,\alpha}) - \alpha| \le B_d' n^{-1} \tag{2.2}$$

for $j = 1, 2$ *and* $n \ge 2$.

Theorems 2.1 and 2.2 imply that the orders of coverage error of the two-sided intervals $I_{2,\alpha}$ and $J_{2,\alpha}$ are identical. However, the one-sided intervals $I_{1,\alpha}$ and $I_{2,\alpha}$ have coverage errors of orders $n^{-1/2}$ and n^{-1}, respectively. In the next subsection we show that these convergence rates are optimal over the respective classes of distributions.

2.3. *Lower bound to coverage error.* In this section we derive inequalities that represent converses to (2.1) and (2.2). Our results apply not just to confidence intervals based on normal approximation, discussed in Section 2.2, but to a very wide range of intervals.

However, some restrictions are necessary. To appreciate this point, observe that it is clearly possible to construct rather uninformative confidence intervals with very high coverage accuracy. For example, if the interval $I = (U, V)$ is taken equal to $(-\infty, \infty)$ with probability α and to $(\bar{X}, \bar{X})$ with probability $1 - \alpha$, then I covers μ with probability precisely equal to α. This practically useless interval differs from more informative approaches in that both its endpoints have very unsmooth distributions.

We may exclude this pathological case by asking that if the confidence interval I is of the form $(-\infty, U)$ or (U, ∞), then for some reference distribution F_0, constants $C_1, C_2 > 0$ and n sufficiently large,

$$P_{F_0}\big(|U - \mu_{F_0}| \le C_2 n^{-(j+1)/2}\big) \ge C_1 n^{-j/2}, \tag{2.3}$$

where $j = 1$ or 2. This condition is no more than an assumption about the smoothness of the distribution of U, coupled with the assumption that U is distant order $n^{-1/2}$ from μ_{F_0}. For example, if

$$P_{F_0}\big(n^{1/2}|U - \mu_{F_0}| \le x\big) \ge C_3 x$$

for all sufficiently small x and large n, then any values of C_1 and C_2 which satisfy $C_1 = C_2 C_3$ will produce (2.3), for all large n.

We should stress that F_0 may be chosen almost arbitrarily to facilitate checking (2.3). For example, F_0 may be chosen to be the standard normal distribution or a uniform distribution. If F_0 is the standard normal distribution, then one may readily check that for appropriate choices of C_1 and C_2, (2.3) holds for intervals $(-\infty, U)$ or (U, ∞) based on the normal approximation, the percentile or percentile-t bootstrap, the double bootstrap, the bootstrap calibrated normal approximation, the Edgeworth-corrected normal approximation and so forth. Thus, the class of confidence interval I for which (2.3) holds for an appropriately chosen reference distribution F_0 is very large.

We now state the main result in this subsection. Let $\mathscr{F}_{j,d}$ denote the class of distributions introduced in Section 2.2, suppose that $F_0 \in \mathscr{F}_{j,d_0}$ for some $d_0 > 0$ and let $\mathscr{C}_j = \mathscr{C}_j(F_0, C_1, C_2, n_0)$ denote the class of all interval sequences $I = I_n = (-\infty, U)$ or (U, ∞) such that (2.3) holds for $n \geq n_0$.

THEOREM 2.3. *If $d \geq d_0$ is chosen sufficiently large (depending on F_0, C_2 and n_0) and for $j = 1$ or 2,*

$$\liminf_{n \to \infty} n^{j/2} \inf_{I \in \mathscr{C}_j} \inf_{0 < \alpha < 1} \sup_{F \in \mathscr{F}_{j,d}} |P_F(\mu_F \in I) - \alpha| > 0. \tag{2.4}$$

To compare the upper bounds (2.1) and (2.2) with the lower bound (2.4), let us suppose that a given one-sided confidence interval $I = I_n$ has nominal coverage α. Theorems 2.1 and 2.2 declare that if I is based on the normal approximation (the case $j = 1$) or on the empirically Edgeworth-corrected normal approximation (for $j = 2$), then

$$\sup_{F \in \mathscr{F}_{j,d}} |P_F(\mu_F \in I) - \alpha| = O(n^{-j/2}).$$

Theorem 2.3 states that for a much more general class of confidence intervals,

$$\sup_{F \in \mathscr{F}_{j,d}} |P_F(\mu_F \in I) - \alpha| \geq Cn^{-j/2}$$

for large n and a constant $C > 0$. In this sense the confidence intervals discussed in Section 2.2 produce levels of coverage accuracy which cannot be bettered by other intervals [satisfying (2.3)], uniformly over distributions in $\mathscr{F}_{j,d}$.

3. A Berry–Esseen theorem for Edgeworth expansions. Adopt notation from Section 2. Without loss of generality, $\mu = 0$ and $\sigma = 1$. Put $T_0 = n^{1/2}\bar{X}/\hat{\sigma}$, representing the so-called "Studentized mean." If X has a nonsingular distribution, then a one-term Edgeworth expansion is valid:

$$P(T_0 \leq x) = \Phi(x) + n^{-1/2}\tfrac{1}{6}\gamma(2x^2 + 1)\phi(x) + o(n^{-1/2}) \tag{3.1}$$

uniformly in x, as $n \to \infty$. In this formula, Φ and ϕ denote the standard normal distribution and density functions, respectively, and $\gamma = E(X^3) \times (\operatorname{var} X)^{-3/2}$ equals the skewness of the sampling distribution. See Chibishov (1984) and Hall (1987). It is of particular interest to determine factors that influence the size of the remainder term $o(n^{-1/2})$ in (3.1), as this would indicate the extent to which the expansion depends on properties of the sampling distribution. In the event that fourth moments are finite and the sampling distribution is nonsingular, (3.1) may be extended to a longer expansion:

$$
\begin{aligned}
&P(T_0 \le x) \\
&\quad = \Phi(x) + n^{-1/2}\tfrac{1}{6}\gamma(2x^2+1)\phi(x) \\
&\qquad + n^{-1}\left\{\tfrac{1}{12}\kappa(x^2-1) - \tfrac{1}{18}\gamma^2(x^4+2x^2-3) - \tfrac{1}{4}(x^2+3)\right\}\phi(x) \\
&\qquad + o(n^{-1}),
\end{aligned} \tag{3.2}
$$

uniformly in x, where $\kappa = E(X^4)(\text{var } X)^{-2} - 3$ denotes kurtosis. Result (3.2) suggests the following "error bound" to formula (3.1):

$$
\begin{aligned}
&\sup_{-\infty<x<\infty} \left| P(T_0 \le x) - \Phi(x) - n^{-1/2}\tfrac{1}{6}\gamma(2x^2+1)\phi(x) \right| \\
&\qquad \le Cn^{-1}\left\{E(X^4) + \left(E|X|^3\right)^2\right\},
\end{aligned} \tag{3.3}
$$

where C is an appropriate constant.

Inequality (3.3) is reminiscent of the Berry–Esseen theorem, which declares that

$$
\sup_{-\infty<x<\infty} \left|P\left\{n^{1/2}\bar{X} \le (\text{var } X)^{1/2}x\right\} - \Phi(x)\right| \le An^{-1/2}\nu,
$$

where $\nu = E|X|^3(\text{var } X)^{-3/2}$ and A denotes an *absolute* constant. (In particular, A does not depend in any way on the sampling distribution and of course not on n.) A version of this result has been proved by Slavova (1985) for the case of the Studentized mean:

$$
\sup_{-\infty<x<\infty} |P(T_0 \le x) - \Phi(x)| \le B(\nu)n^{-1/2}, \tag{3.4}
$$

where $B(\nu)$ depends only on ν. A careful examination of Slavova's proof shows that $B(\nu)$ has the property that for each $\nu_0 > 0$,

$$
\sup_{0<\nu<\nu_0} B(\nu) < \infty, \tag{3.5}
$$

and so without loss of generality $B(\nu)$ is increasing in ν.

Despite these encouraging precedents to (3.3), that inequality must fail as a general result since even the milder expansion (3.1) can be invalid for lattice-valued sampling distributions. To overcome this problem we impose a smoothness condition on the sampling distribution as follows. Suppose that for some $y \in (-\infty, \infty)$ and $c_1, c_2 > 0$,

$$
\liminf_{h \downarrow 0} h^{-1}P(|X + x - y| \le h) \ge c_1, \quad \text{all } |x| < c_2. \tag{3.6}
$$

This condition may be equivalently expressed by saying that "the distribution of X contains a uniform distribution on $(y - c_2, y + c_2)$ with weight c_1c_2." As before, define γ to equal skewness.

To appreciate the implications of (3.6), let us suppose that the distribution function F of X is nonsingular. Then we may write $F = \pi F_1 + (1 - \pi)F_2$, where $0 < \pi \le 1$, F_1 and F_2 are distribution functions and F_1 is absolutely continuous. Write $f_1 = F_1'$ for the density function corresponding to F_1 and let $I = (a, b)$ denote a nonempty interval on which f_1 is bounded away from zero. Then (3.6) holds with $y = \frac{1}{2}(a + b)$, $c_1 = \pi \inf_{x \in I} f_1(x)$ and $c_2 = \frac{1}{2}(b - a)$.

Theorem 3.1 is a Berry–Esseen theorem for Edgeworth expansions and Theorem 3.2 is a variant of that result needed to derive Theorem 2.2.

THEOREM 3.1. *There exists an absolute constant $A > 0$ such that for all distributions of X that satisfy $E(X^4) < \infty$, $E(X) = 0$ and condition* (3.6),

$$\sup_{-\infty < x < \infty} |P(T_0 \le x) - \Phi(x) - n^{-1/2}\tfrac{1}{6}\gamma(2x^2 + 1)\phi(x)| \le An^{-1}\Big\{\big(c_1c_2^3\big)^{-2}E|X - y|^4 + \big(c_1c_2^3\big)^{-3}\big(E|X - y|^3\big)^2\Big\} \tag{3.7}$$

for $n \ge 2$.

It is straightforward to develop versions of inequality (3.7) for higher-order expansions.

THEOREM 3.2. *Let x and x_0 denote arbitrary fixed positive constants. Then*

$$P\big\{T_0 \le x - n^{-1/2}\tfrac{1}{6}\hat{\gamma}(2x^2 + 1)\big\} = \Phi(x) + O(n^{-1})$$

uniformly in $|x| \le x_0$ and distributions of X that satisfy (3.6) *and are such that $E(X) = 0$, $E|X - y|^4 \le C(c_1c_2^3)^2$ and $E|X - y|^3 \le C(c_1c_2^3)^{3/2}$.*

4. Proofs. Theorem 2.1 follows from (3.4) and (3.5) when $j = 1$, and from Theorem 3.1 when $j = 2$. Theorem 2.2 is implied by Theorem 3.2.

PROOF OF THEOREM 3.1. We outline the proof up to the point where details coincide relatively closely with those described by Hall (1987). It is permissible for us to assume that $y = 0$ in condition (3.6), although of course we are now not permitted to suppose that X has zero mean. In this new notation, $T_0 = n^{1/2}(\bar{X} - EX)(n^{-1}\sum X_i^2 - \bar{X}^2)^{-1/2}$ and result (3.7) is equivalent to

$$\sup_{-\infty < x < \infty} |P(T_0 \le x) - \Phi(x) - n^{-1/2}\tfrac{1}{6}\gamma(2x^2 + 1)\phi(x)| \le An^{-1}\Big\{\big(c_1c_2^3\big)^{-2}E|X|^4 + \big(c_1c_2^3\big)^{-3}\big(E|X|^3\big)^2\Big\}, \tag{4.1}$$

for an absolute constant A. We shall outline a proof of (4.1).

Write $\mathscr{F}$ for the σ-field generated by $|X_1|, \dots, |X_n|$ and put $S_j = \operatorname{sgn}(X_j)$, $p_j = P(X_j > 0 \| X_j|)$, $Y_j = X_j - E(X_j \| X_j|)$, $\beta_{kj} = E(Y_j^k \| X_j|)$ for $k \ge 2$, $\nu_k = E(\beta_{k1})$, $s^2 = \sum \beta_{2j}$, $T = \sum Y_j$ and $\psi_j(t) = E\{\exp(itY_j) \| X_j|\}$. We preface our proof with two lemmas.

LEMMA 4.1. *Under condition* (3.6),

$$E|\psi_1(t)| \le 1 - (c_1c_2/28)\min\{1, (2c_2t)^2\}$$

for all real t.

PROOF. [Recall that we suppose $y = 0$ in (3.6).] Let F denote the distribution function of X, write F_1 for the distribution function of the uniform distribution on $(-c_2, c_2)$, put $\pi = c_1c_2 \le 1$ and let $F_2 = (1-\pi)^{-1}(F - \pi F_1)$ if $\pi < 1$. Then F_2 is a distribution function and, by construction, $F = \pi F_1 + (1-\pi)F_2$. Let χ, χ_1 or χ_2 denote the characteristic function of X conditional on $|X|$ in the case where X has distribution function F, F_1 or F_2, respectively. Then $\chi = \pi\chi_1 + (1-\pi)\chi_2$ and so $|\chi| \le \pi|\chi_1| + 1 - \pi$. Also, $\chi_1(t) = \cos(t|X|)$ and so

$$E|\psi_1(t)| = E_F|\chi(t)| \le \pi E_{F_1}|\chi_1(t)| + 1 - \pi = \pi c_2^{-1}\int_0^{c_2}|\cos(tx)|\,dx + 1 - \pi.$$

However,

$$\begin{aligned}\left\{\int_0^{c_2}|\cos(tx)|\,dx\right\}^2 &\le c_2\int_0^{c_2}\cos^2(tx)\,dx = (1/2)c_2^2\left\{1 + (2c_2t)^{-1}\sin(2c_2t)\right\}\\ &\le c_2^2\left[1 - (1/14)\min\{1, (2c_2t)^2\}\right]\\ &\le c_2^2\left[1 - (1/28)\min\{1, (2c_2t)^2\}\right]^2,\end{aligned}$$

the second-to-last line following since $t^{-1}\sin t \le 1 - (1/7)\min(1, t^2)$. Therefore,

$$E|\psi_1(t)| \le \pi\left[1 - (1/28)\min\{1, (2c_2t)^2\}\right] + 1 - \pi,$$

which implies the lemma. □

Our next result is a portion of Lemma 2.3 of Hall (1987). Here and below, $A_1, A_2, \ldots$ denote positive absolute constants.

LEMMA 4.2. *Let $Z_1, \ldots, Z_n$ be independent random variables with finite third moments, zero means and $\sum_{j=1}^n E(Z_j^2) = 1$. Set*

$$\chi_j(t) \equiv E(e^{itZ_j}) \quad \text{and} \quad \beta_j(t) \equiv E\left\{\exp(itZ_j) - \sum_{r=0}^{3}\frac{1}{r!}(itZ_j)^r\right\}$$

and choose l so large that

$$\sum_{j=1}^{n} E\{Z_j^2 I(|Z_j| > l)\} \le \tfrac{1}{8}.$$

Then whenever $|t| \le 1/12l$,

$$\left|\prod_{j=1}^{n} \chi_j(t) - \left\{1 + \sum_{j=1}^{n} \beta_j(t) + \tfrac{1}{6}(it)^3 \sum_{j=1}^{n} E\left(Z_j^3\right)\right\} e^{-t^2/2}\right|$$
$$\le A_1 |t| e^{-t^2/6} \left[\sum_{j=1}^{n} \left\{E\left(Z_j^2\right)\right\}^2 + \left\{\sum_{j=1}^{n} E\left(|Z_j|^3\right)\right\}^2\right].$$

Put $\lambda = \{1024 E(X^4)/\nu_2\}^{1/2}$ and $\nu(\lambda) = E\{Y_1^2 I(|Y_1| > \lambda)\}$. Let $\mathscr{E}$ denote the event that $|s^2 - n\nu_2| \le \frac{1}{2} n \nu_2$ and

$$\left|\sum_{j=1}^{n} E\left\{Y_j^2 I\left(|Y_j| > \lambda\right) | \mathscr{F}\right\} - n\nu(\lambda)\right| \le (1/32)\, n\nu_2.$$

Write $\tilde{\mathscr{E}}$ for the complement of $\mathscr{E}$. We may prove from Markov's inequality that

$$P(\tilde{\mathscr{E}}) \le A_2 n^{-1} \nu_2^{-2} E(X^4). \tag{4.2}$$

Similarly, since $|Y_j| \le 2|X_j|$ then the following results may be established on $\mathscr{E}$:

$$s^{-2} \sum_{j=1}^{n} E\left\{Y_j^2 I\left(|Y_j| > \lambda\right) | \mathscr{F}\right\} \le \left(\tfrac{1}{2} n\nu_2\right)^{-1} n \lambda^{-2} 16 E(X^4) + \tfrac{1}{16} = \tfrac{1}{8},$$

$$s^{-4} \sum_{j=1}^{n} \left\{E\left(Y_j^2 | \mathscr{F}\right)\right\}^2 \le \left(\tfrac{1}{2} n\nu_2\right)^{-2} \sum_{j=1}^{n} \left(4X_j^2\right)^2 = 64 (n\nu_2)^{-2} \sum_{j=1}^{n} X_j^4,$$

$$s^{-6} \left\{\sum_{j=1}^{n} E\left(|Y_j|^3 | \mathscr{F}\right)\right\}^2 \le s^{-4} \sum_{j=1}^{n} E\left(Y_j^4 | \mathscr{F}\right) \le 64 (n\nu_2)^{-2} \sum_{j=1}^{n} X_j^4,$$

$$s^{-3} \sum_{j=1}^{n} E\left(|Y_j|^3 | \mathscr{F}\right) \le (512)^{1/2} (n\nu_2)^{-3/2} \sum_{j=1}^{n} |X_j|^3,$$

$$\left|\sum_{j=1}^{n} \psi_j(t/s) - \sum_{r=0}^{3} \frac{1}{r!} (it)^r \sum_{j=1}^{n} \beta_{rj}\right| \le 64 (n\nu_2)^{-2} t^4 \sum_{j=1}^{n} X_j^4.$$

Therefore, taking $Z_j = s^{-1} Y_j$ and $l = s^{-1}\lambda$ in Lemma 4.2, and evaluating all distributions and expectations conditional on $\mathscr{F}$, we deduce that if the event $\mathscr{E}$ obtains,

$$\left|\prod_{j=1}^{n} \psi_j(t/s) - \left\{1 + \tfrac{1}{6}(it)^3 (n\nu_2)^{-3/2} \sum_{j=1}^{n} E\left(Y_j^3 | \mathscr{F}\right)\right\} e^{-t^2/2}\right|$$
$$\le A_3 |t| e^{-t^2/6} \left\{(n\nu_2)^{-2} \sum_{j=1}^{n} X_j^4 + (n\nu_2)^{-5/2} \left(\sum_{j=1}^{n} |X_j|^3\right) |s^2 - n\nu_2|\right\}$$

uniformly in $|t| \le s/(12\lambda)$. Now apply the smoothing lemma for characteristic functions [e.g., Petrov (1975), Theorem 2, page 109, with $T = c_2^{-4} nE(X^4)$ in Petrov's notation], obtaining on $\mathscr{E}$,

$$\sup_{-\infty<x<\infty} \left| P(T \le sx|\mathscr{F}) - \left\{ \Phi(x) + \tfrac{1}{6}(1-x^2)\phi(x)(n\nu_2)^{-3/2} \sum_{j=1}^{n} E\left(Y_j^3|\mathscr{F}\right) \right\} \right| \tag{4.3}$$
$$\le A_4 \left[(n\nu_2)^{-2} \sum_{j=1}^{n} X_j^4 + (n\nu_2)^{-5/2} \left(\sum_{j=1}^{n} |X_j|^3 \right) |s^2 - n\nu_2| + \int_{s/(12\lambda)}^{c_2^{-4} nE(X^4)} \left\{ \prod_{j=1}^{n} |\psi_j(t/s)| \right\} dt + c_2^4 \{nE(X^4)\}^{-1} \right].$$

Let J denote the integral on the right-hand side. On $\mathscr{E}$,

$$J \le \left(\tfrac{3}{2} n\nu_2\right)^{1/2} \int_{1/(12\lambda)}^{c_2^{-4}((1/2)n\nu_2)^{-1/2} nE(X^4)} \left\{ \prod_{j=1}^{n} |\psi_j(t)| \right\} dt.$$

From this point, making use of Lemma 4.1, one may show that

$$E\{JI(\mathscr{E})\} \le A_5 \left\{ \left(c_1 c_2^3\right)^{-3/2} \nu_2^{-1/2} + \left(c_1 c_2^3\right)^{-2} \right\} n^{-1} E(X^4).$$

Adopting notation introduced during the proof of Lemma 4.1, we have

$$\nu_2 = 4E_F\{p_1(1-p_1)X_1^2\} \ge 4\pi E_{F_1}\{p_1(1-p_1)X_1^2\}$$
$$= 4 \cdot c_1 c_2 \cdot \tfrac{1}{4} c_2^{-1} \int_0^{c_2} x^2\, dx = \tfrac{1}{3} c_1 c_2^3.$$

Therefore, $E\{JI(\mathscr{E})\} \le A_6 (c_1 c_2^3)^{-2} n^{-1} E(X^4)$. Replacing x in (4.3) by an arbitrary $\mathscr{F}$-measurable random variable U_x and taking expectation, we deduce that

$$\delta \equiv \sup_{-\infty<x<\infty} \left| P(T \le sU_x) - E\left\{ \Phi(U_x) + \tfrac{1}{6}(1-U_x^2)\phi(U_x)(n\nu_2)^{-3/2} \sum_{j=1}^{n} E\left(Y_j^3|\mathscr{F}\right) \right\} \right| \tag{4.4}$$
$$\le A_7 \Big[\left(c_1 c_2^3\right)^{-2} n^{-1} E(X^4) + n^{-3/2} \nu_2^{-5/2} E\left\{ |X_1|^3 |s^2 - n\nu_2| I\left(|s^2 - n\nu_2| \le \tfrac{1}{2} n\nu_2\right) \right\}$$
$$+ P(\tilde{\mathscr{E}}) + n^{-1/2} \nu_2^{-3/2} E\left\{ |E\left(Y_1^3|\mathscr{F}\right)| I(\tilde{\mathscr{E}}) \right\} + c_2^4 n^{-1} \left(EX^4\right)^{-1} \Big].$$

Now

$$E\{|X_1|^3|s^2 - n\nu_2|I(|s^2 - n\nu_2| \le \tfrac{1}{2}n\nu_2)\}$$
$$\le A_8\{(n\nu_2)^{1/2}E(X^4) + (n/\nu_2)^{1/2}(E|X|^3)^2\}$$

and for $n \ge 2$,

$$E\{|X_1|^3 I(\tilde{\mathscr{E}})\} \le A_9(n\nu_2)^{-1/2}E(X^4).$$

Combining the results from (4.4) down and noting (4.2), we conclude that

$$\delta \le A_{10}\{(c_1c_2^3)^{-2}E(X^4) + c_2^4(EX^4)^{-1} + (c_1c_2^3)^{-3}(E|X|^3)^2\}n^{-1}. \tag{4.5}$$

In notation from the proof of Lemma 4.1, $\pi E_{F_1}(X^4) \le E_F(X^4)$ or, equivalently,

$$c_1c_2 \cdot c_2^{-1}\int_0^{c_2} x^4\,dx \le E(X^4), \quad \text{that is, } c_1c_2^5 \le 5E(X^4).$$

Therefore, $c_2^4E(X^4) \le 25(c_1c_2^3)^{-2}E(X^4)$. Hence, by (4.5),

$$\delta \le A_{11}\{(c_1c_2^3)^{-2}E(X^4) + (c_1c_2)^{-3}(E|X|^3)^2\}n^{-1}. \tag{4.6}$$

The argument in Hall [(1987), pages 925–930] may be employed to show that, with

$$t_1(x) = E\left\{\tfrac{1}{6}(1 - U_x^2)\phi(U_x)(n\nu_2)^{-3/2}\sum_{j=1}^{n}E(Y_j^3|\mathscr{F})\right\},$$
$$t_2(x) = n^{-1/2}\tfrac{1}{6}\nu_2^{-3/2}\nu_3E\{(1 - U_x^3)\phi(U_x)\},$$

we have

$$\sup_{-\infty<x<\infty}|t_1(x) - t_2(x)| \le A_{12}\{(c_1c_2^3)^{-2}E(X^4) + (c_1c_2^3)^{-3}(E|X|^3)^2\}n^{-1}. \tag{4.7}$$

Result (4.1) follows from (4.6) and (4.7). This completes the proof of Theorem 3.1. □

The proof of Theorem 3.2 is similar to that of Theorem 3.1 and so is not given here.

PROOF OF THEOREM 2.3. It suffices to derive the following result, which is substantially more general than Theorem 2.3.

THEOREM 4.1. *Fix $r > 2$ and let F_0 denote any distribution with finite rth moment. Write $\mathscr{C} = \mathscr{C}(F_0, C_1, C_2, n_0, r)$ for the class of all interval sequences $I = I_n$ such that*

$$|P_{F_0}(\mu_{F_0} \in I + C_2n^{-(r-1)/2}) - P_{F_0}(\mu_{F_0} \in I - C_2n^{-(r-1)/2})| \ge C_1n^{-(r-2)/2} \tag{4.8}$$

for $n \geq n_0$. *Let* C_3 *denote any point of support of* F_0, *let* $0 < \varepsilon < 1$ *and for* $i = 1, 2$ *let* $F_{i,n}$ *denote the distribution obtained by taking mass* $\frac{1}{4}\varepsilon C_1 n^{-r/2}$ *from the vicinity of* C_3 *and placing it at* $x_n \sim (-1)^i 4\varepsilon^{-1} C_1^{-1} C_2 n^{1/2}$. *We may choose* x_n, *satisfying this last asymptotic relation, such that*:

(a) $$E_{F_{i,n}}|X|^s - E_{F_0}|X|^s = O(n^{-(r-s)/2}) \quad \text{for } 0 < s \leq r,$$

$$\mu_{F_{i,n}} - \mu_{F_0} = (-1)^i C_2 n^{-(r-1)/2},$$

$$E_{F_{i,n}}|X|^r - E_{F_0}|X|^r \to (4/\varepsilon)^r C_1^{1-r} C_2^r$$

as $n \to \infty$ *and*

(b) $$\liminf_{n\to\infty} n^{(r-2)/2} \inf_{I \in \mathscr{C}} \inf_{0<\alpha<1} \max_{F \in \{F_{1,n}, F_{2,n}, F_0\}} |P_F(\mu_f \in I) - \alpha| > \tfrac{1}{4}(1-\varepsilon)C_1.$$

PROOF. Result (a) follows by direct calculation. To prove (b), note that in view of (4.8) we have either

$$(4.9) \qquad |P_{F_0}(\mu_{F_0} \in I + C_2 n^{-(r-1)/2}) - P_{F_0}(\mu_{F_0} \in I)| \geq \tfrac{1}{2} C_1 n^{-(r-2)/2}$$

or

$$(4.10) \qquad |P_{F_0}(\mu_{F_0} \in I - C_2 n^{-(r-1)/2}) - P_{F_0}(\mu_{F_0} \in I)| \geq \tfrac{1}{2} C_1 n^{-(r-2)/2}.$$

The chance that some data value in an n-sample drawn from $F_{i,n}$ come from the region of probability $\frac{1}{2}\varepsilon C_1 n^{-r/2}$, where F_0 and $F_{i,n}$ differ, is less than $\frac{1}{2}\varepsilon C_1 n^{-(r-2)/2}$. Therefore, if inequality $(4.8 + i)$ holds for $i = 1$ or 2, $|P_{F_{i,n}}(\mu_{F_{i,n}} \in I) - P_{F_0}(\mu_{F_0} \in I)| \geq \frac{1}{2}(1 - \varepsilon)C_1 n^{-(r-2)/2}$. Taking F_n to equal $F_{1,n}$ if (4.9) holds and to equal $F_{2,n}$ if (4.10) holds but (4.9) fails, we deduce that $|P_{F_n}(\mu_{F_n} \in I) - P_{F_0}(\mu_{F_0} \in I)| \geq \frac{1}{2}(1 - \varepsilon)C_1 n^{-(r-2)/2}$. Therefore, for any $0 < \alpha < 1$, $|P_F(\mu_F \in I) - \alpha| > \frac{1}{4}(1 - \varepsilon)C_1 n^{-(r-2)/2}$ holds for $F = F_0$ or $F = F_n$. This establishes part (b) of the theorem. □

Acknowledgments. We are grateful for the constructive comments of two referees, whose suggestions encouraged this particularly succinct revision.

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

CHIBISHOV, D. M. (1984). Asymptotic expansions and deficiencies of tests. In *Proceedings of the International Congress of Mathematicians, August 16–24, 1983, Warsaw* (Z. Ciesielski and C. Olech, eds.) **2** 1063–1079. North-Holland, Amsterdam.

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

HALL, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *Ann. Probab.* **15** 920–931.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, Berlin.

PETROV. V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

PFANZAGL, J. (1979). Nonparametric minimum contrast estimators. *Selecta Statistica. Canadiana* **5** 105–140.

SLAVOVA, V. V. (1985). *On the Berry–Esseen Bound for Student's Statistic. Lecture Notes in Math.* **1155** 335–390. Springer, Berlin.

CENTRE FOR MATHEMATICS AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA