ACCURATE CONFIDENCE INTERVALS FOR DISTRIBUTIONS
WITH ONE PARAMETER

C. S. Withers
(Received Aug. 6, 1981; revised Feb. 18, 1982)

Summary

Let $\hat{\theta}_n$ be an estimate of a real parameter $\theta$. Suppose that for some function $c(\cdot)$ and some random variable (r.v.) $\tau_n$, the distribution of

$$Z_n=(c(\theta)-c(\hat{\theta}_n))/\tau_n$$

is continuous and depends only on $\theta$ and $n$ and that the cumulants of $Z_n$ can be expanded in the form

$$K_i(Z_n) \approx \sum_{i=r-1}^{\infty} a_r(\theta)n^{-i}.$$ 

Then a confidence interval for $\theta$ can be constructed with level $1-\alpha+O(n^{-j/2})$ for any given value of $\alpha$ and $j$.

1. Introduction

This paper offers ways of improving the accuracy of approximate confidence intervals (C.I.'s) for one parameter problems when the cumulants of the parameter estimate have a very commonly occurring type of asymptotic expansion.

Section 2 summarises the usual first-order approximations to C.I.'s based on an asymptotically normal estimate, and indicates the magnitude of their error. Section 3 shows how to reduce this error, and Section 4 gives some examples.

2. First-order confidence intervals

Let $\theta$ be an unknown real parameter known to lie in an interval $[a, b]$, where $-\infty \leq a < b \leq \infty$. Let $c(\cdot)$ be a one to one increasing func-

---

Keywords and phrase: Accurate confidence intervals, Cornish-Fisher expansions, cumulants, percentiles.
tion on \([a, b]\). Let \(\Phi, \phi\) be the distribution function and density of \(\mathcal{N}(0, 1)\). Suppose that \(\tau_n > 0\) is an r.v. bounded in probability away from 0 and \(\infty\) and that the distribution of

\[ Z_n = (c(\theta) - c(\hat{\theta}_n))/\tau_n \]

depends only on \(\theta\) and \(n\). When

(1) \[ Z_n \text{ is asymptotically } \mathcal{N}(0, \nu(\theta)/n), \]

then a confidence interval for \(\theta\) of level approximately \(1 - \alpha\) is

(2) \[ V_{1a}(\hat{\theta}_n, x_2) \leq \theta \leq V_{1a}(\hat{\theta}_n, x_1) \]

where \(V_{1a}(\theta, x) = c^{-1}(c(\theta) + n^{-1/2}x\nu(\theta)^{1/2})\) and \(x_1, x_2\) are chosen so that

(3) \[ \phi(x_1) - \phi(x_2) = 1 - \alpha. \]

For a one-sided test one chooses \(x_1 = \infty\) or \(x_2 = -\infty\), and for a two-sided test, the usual choice is

(4) \[ x_1 = -x_2 = \Phi^{-1}(1 - \alpha/2). \]

The C.I. (2) has level \(1 - \alpha + e_n\), where generally speaking the error \(e_n\) has magnitude \(n^{-1/2}\) as \(n \to \infty\), unless either (4) holds—i.e. the tails are equal—or the distribution of \(Z_n\) is symmetric; in either of these events the magnitude of the error reduces to \(n^{-1}\). For more precise conditions see Withers [9].

Example 2.1. Suppose \(\hat{\theta}_n \sim \mathcal{N}(\theta, V(\theta)/n)\). Then \(c(\hat{\theta}_n) \sim \mathcal{N}(c(\theta), V_c(\theta)/n)\) where \(V_c(\theta) = c''(\theta)^2V(\theta)\). (We use \(f^{(r)}(\theta)\) to denote the \(r\)th derivative of \(f(\theta)\).) The choice \(\tau^2_n = V_c(\theta)\) implies \(\nu(\theta) = 1\). Generally \(c(\cdot)\) is chosen either

(i) for simplicity—such as \(c(\theta) = \theta\); or
(ii) to satisfy \(c(a) = -\infty, c(b) = \infty\)—so that the interval (2) contains no points outside \([a, b]\); or
(iii) so that \(V_c(\theta) = 1\)—that is \(c(\theta) = \int_0^\theta V(x)^{1/2}dx\); or
(iv) to reduce the bias or skewness of \(c(\hat{\theta}_n)\).

However none of these choices reduce the magnitude of the error of (2).

Example 2.1(a). Let \(\hat{\theta}_n\) be the sample correlation of a sample of size \(N = n\) from a bivariate normal population with correlation \(\theta\). Let \(c(\theta) = \tanh^{-1} \theta\). This choice satisfies (ii), (iii), and (iv). But since its bias is still \(O(n^{-1})\), the C.I. (2) still has error \(e_n = O(n^{-1/2})\) or \(O(n^{-1})\) if the tails are equal, the same as for the choice \(c(\theta) = \theta\), \(\tau_n = 1 - \hat{\theta}_n^2\). (The same is true with choices of \(n\) such as \(N-1\) or \(N-3\)).
Example 2.2. Let \( \{X_1, \ldots, X_n\} \) be a random sample from \( F_\theta((x-\theta)/\sigma) \) where \( F_\theta \) is a given distribution with variance 1. Choose \( c(\theta) = \theta \) and \( \hat{\theta}_n, \tau_n \) such that the distribution of \( (\hat{\theta}_n-\theta)/\tau_n \) does not depend on \( (\theta, \sigma) \). (This is true for a wide class of estimates \( (\hat{\theta}_n, \tau_n) \).) Then in general the interval (2) has error \( O(n^{-1/2}) \) unless either the tails are equal or \( F_\theta \) is symmetric, in which case the error is \( O(n^{-1}) \).

3. Improved approximations

We now give a method for obtaining a C.I. for \( \theta \) with error \( O(n^{-1/2}) \) for any given \( j \). We replace (1) by the stronger condition that the cumulants of \( Z_n \) have expansions of the form

\[
K_r(Z_n) \approx \sum_{i=1}^{\infty} a_{r_i} n^{-i}, \quad r \geq 1, \quad a_{10} = 0
\]

where \( \{a_{r_i} = a_{r_i}(\theta)\} \) are functions of \( \theta \), and we assume that the distribution of \( Z_n \) is absolutely continuous. By Fisher and Cornish [3], (5) implies (1) with \( v(\theta) = a_{2i}(\theta) \).

The assumption (5) holds for a wide class of \( (\hat{\theta}_n, \tau_n) \); see for example Withers [9] where formulas for \( \{a_{r_i}\} \) are given when \( (\hat{\theta}_n, \tau_n) \) are regular functionals of the empirical distribution of a random sample of size \( n \).

Let

\[
P_n(x) = \Pr \left( n^{1/2} a_{2i}^{-1/2} Z_n \leq x \right).
\]

Upon substitution of (5) into the expansions of Cornish and Fisher, one obtains the asymptotic expansion

\[
P_n^{-1}(\Phi(x)) \approx x + \sum_{r=1}^{\infty} n^{-r/2} g_r(x)
\]

where \( g_r(x) \) is a polynomial of degree \( r+1 \) given in terms of

\[
A_{r_i} = a_{2i}^{-1/2} a_{r_i},
\]

in Appendix 1.

Set

\[
g_r(x, \theta) = \begin{cases} x, & r = 0 \\ g_r(x), & r \geq 1 \end{cases}
\]

and for a given value of \( x \), set

\[
p_i(\theta) = P_i(c(\theta)) = -\tau_i a_{2i}(\theta)^{1/2} g_{i-1}(x, \theta), \quad i \geq 1
\]
and

\[
R_{j*}(\theta) = c(\theta) + \sum_{i=1}^{j} n^{-1/2} p_i(\theta), \quad j \geq 1.
\]

**Theorem 3.1.** Suppose that (6) holds for \(x = x_1, x_2\) satisfying (3) and that for some \(j \geq 1\), \(R_{j*}(\theta)\) is one to one increasing in a suitably large neighbourhood of \(\hat{\theta}_n\). Then a confidence interval of level \(1 - \alpha\) with error \(O(n^{-1/2})\) is given by

\[
(8) \quad V_{jn}(\hat{\theta}_n, x_1, x_2) \leq \theta \leq V_{jn}(\hat{\theta}_n, x_1)
\]

where

\[
(9) \quad V_{jn}(\theta, x) = c^{-1}\left(c(\theta) + \sum_{i=1}^{j} n^{-1/2} q_i(\theta)\right),
\]

\[
q_i(\theta) = \tau_{i} x a_{2i}(\theta)^{1/2},
\]

\[
q_4(\theta) = \tau_{4} x a_{11}(\theta)^{1/2} g(x, \theta) + \tau_{5} x c^{(1)}(\theta)^{-1} a_{12}^{(1)}(\theta)/2,
\]

\[
q_5(\theta) = \tau_{6} a_{12}(\theta)^{1/2} g_2(x, \theta) + \tau_{7} x c^{(1)}(\theta)^{-1} \left\{ a_{21}(\theta) g(x, \theta) \right\}/\theta
\]

\[
+ \tau_{8} x c^{(2)}(\theta)^{-1} a_{22}(\theta)^{1/2} (2a_{12}^{(2)}(\theta) + a_{11}^{(1)}(\theta)^2 a_{21}^{(1)}(\theta)^{-1})/8
\]

\[- \tau_{9} x c^{(1)}(\theta)^{-1} a_{22}(\theta)^{1/2} a_{11}^{(1)}(\theta)^{-1}/4,
\]

\[
q_7(\theta) = -p(\theta) + \frac{1}{2} \sum_{i=1}^{3} \tilde{P}^{(i)} q_i(\theta) - \tilde{P}^{(2)} q_i(\theta) q_i(\theta) q_i(\theta)
\]

\[- \tilde{P}^{(3)} q_i(\theta)^3/6,
\]

and

\[
q_i(\theta) = -p(\theta) - \sum_{i=1}^{4} P^{(i)} q_i(\theta) - \tilde{P}^{(2)} q_i(\theta) q_i(\theta) q_i(\theta)
\]

\[- \tilde{P}^{(3)} q_i(\theta)^3/2 + q_i(\theta) q_i(\theta) q_i(\theta)
\]

\[- \tilde{P}^{(4)} q_i(\theta)^4/24,
\]

where \(\{\tilde{P}^{(r)} = P^{(r)}(c(\theta)), 1 \leq r \leq 4\}\) are given by

\[
\tilde{P}^{(1)} = c^{(1)}(\theta)^{-1} p^{(1)}(\theta)
\]

\[
\tilde{P}^{(2)} = c^{(1)}(\theta)^{-2} p^{(2)}(\theta) - c^{(1)}(\theta)^{-3} c^{(2)}(\theta) p^{(3)}(\theta),
\]

\[
\tilde{P}^{(3)} = c^{(1)}(\theta)^{-4} p^{(3)}(\theta) - 3 c^{(1)}(\theta)^{-4} c^{(2)}(\theta) p^{(4)}(\theta)
\]

\[+ \{3 c^{(1)}(\theta)^{-4} c^{(2)}(\theta)^2 - c^{(1)}(\theta)^{-4} c^{(3)}(\theta)\} p^{(4)}(\theta),
\]

\[
\tilde{P}^{(4)} = c^{(1)}(\theta)^{-5} p^{(4)}(\theta) - 6 c^{(1)}(\theta)^{-4} c^{(2)}(\theta) p^{(5)}(\theta)
\]

\[+ \{15 c^{(1)}(\theta)^{-4} c^{(2)}(\theta)^2 - 4 c^{(1)}(\theta)^{-4} c^{(3)}(\theta)\} p^{(3)}(\theta)
\]

\[+ \{-15 c^{(1)}(\theta)^{-4} c^{(2)}(\theta)^3 + 10 c^{(1)}(\theta)^{-4} c^{(3)}(\theta) c^{(3)}(\theta)
\]

\[\quad - c^{(1)}(\theta)^{-4} c^{(4)}(\theta)\} p^{(4)}(\theta).
\]
PROOF. By (6) with probability \( \Phi(x + O(n^{-j/3})) \), \( c(\hat{\theta}_n) \geq R_{j,n}(\theta) \), which for \( R_{j,n} \) one to one is equivalent to

\[
c(V_j, \hat{\theta}_n, x) + O_p(n^{-j/3}) \geq c(\theta),
\]

where \( q_i(\theta) = Q_i(c(\theta)) \), and \( g(x) = x + \sum_{i=1}^\infty n^{-r/2} Q_i(x) \) is the inverse of \( x(g) = g + \sum_{i=1}^\infty n^{-r/2} P_i(g) \). Now apply Appendix 2.

Note 1. For Example 2.1 for a given \( c(\cdot) \), \( \{q_i(\hat{\theta}_n)\} \) are independent of the choice of \( \tau_n \), provided \( \tau_n \) is a function of \( \hat{\theta}_n \) independent of \( n \); thus \( \{q_i\} \) are most easily computed choosing \( \tau_n = 1 \).

Note 2. If \( j = 2 \) and \( \tau_n = 1 \) or \( V_j(\hat{\theta}_n)^{1/2} \), then for Example 2.1 the interval (8) is just the same as that given by Withers [6] for

\[
Y_n(\theta) = n^{1/2}(c(\theta) - c(\hat{\theta}_n))V_c(\hat{\theta}_n)^{-1/2}.
\]

4. Some examples

Example 4.1. Returning to Example 2.1(a) we have

Case 1. \( c(\theta) = \theta \) and \( \tau_n = 1 \): then \( a_{21}(\theta) = (1 - \theta^2)^2 \) and by (4.12) of Withers [6], \( g_1(x) = \theta(x^2 - 1/2) \), \( g_2(x) = (3x - x^3)/4 + \theta^2(-5x + 4x^3)/4 \), so that

\[
q_1(\theta) = x(1 - \theta^2), \quad q_3(\theta) = -\theta(\theta - 3/4)(1/2 + x^2),
\]

and

\[
q_3(\theta) = (1 - \theta^2)(x - x^3 + \theta^2(5x + 4x^3))/4.
\]

Case 2. \( c(\theta) = \tanh^{-1} \theta \) and \( \tau_n = 1 \): then \( a_{21}(\theta) = 1 \) and by (4.13) of Withers [6], \( g_1(x) = -\theta/2 \), \( g_2(x) = (9x + x^3)/12 - \theta^2 x/4 \), so that

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( j = 1 )</th>
<th>( j = 2 )</th>
<th>( j = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( -.9 )</td>
<td>-.202</td>
<td>-.09(6)</td>
<td>-.05(3)^2</td>
</tr>
<tr>
<td>( -.5 )</td>
<td>-.154</td>
<td>-.06(6)</td>
<td>-.03(2)</td>
</tr>
<tr>
<td>0</td>
<td>-.09(0)</td>
<td>-.02(8)</td>
<td>-.007</td>
</tr>
<tr>
<td>( .5 )</td>
<td>-.01(7)</td>
<td>.00(9)</td>
<td>.010</td>
</tr>
<tr>
<td>( .9 )</td>
<td>.043</td>
<td>-.19(2)</td>
<td>.01(4)</td>
</tr>
<tr>
<td>( \text{Case 1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( -.9 )</td>
<td>-.07(1)</td>
<td>-.04(1)^a</td>
<td>-.01(5)^2</td>
</tr>
<tr>
<td>( -.5 )</td>
<td>-.05(8)</td>
<td>-.03(3)</td>
<td>-.01(0)</td>
</tr>
<tr>
<td>0</td>
<td>-.04(1)</td>
<td>-.024</td>
<td>-.005</td>
</tr>
<tr>
<td>( .5 )</td>
<td>-.021</td>
<td>-.017</td>
<td>-.001</td>
</tr>
<tr>
<td>( .9 )</td>
<td>-.00(1)</td>
<td>-.01(5)</td>
<td>.00(2)</td>
</tr>
</tbody>
</table>

Table 1. Error in exact probability of one-sided nominally 95% confidence interval given by (8) with \( \tau_n = 1 \) for Example 2.1(a)
Table 1. (Continued)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$j=1$</th>
<th>$j=2$</th>
<th>$j=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.9</td>
<td>-.12(8)</td>
<td>-.04(5)</td>
<td>-.01(9)</td>
</tr>
<tr>
<td>-.5</td>
<td>-.09(3)</td>
<td>-.02(7)</td>
<td>-.00(9)</td>
</tr>
<tr>
<td>0</td>
<td>-.04(5)</td>
<td>-.00(5)</td>
<td>.00(1)</td>
</tr>
<tr>
<td>.5</td>
<td>.00(6)</td>
<td>.00(8)</td>
<td>.00(0)</td>
</tr>
<tr>
<td>.9</td>
<td>.04(6)</td>
<td>-.07(1)</td>
<td>.01(3)$^2$</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.9</td>
<td>-.03(5)</td>
<td>-.01(6)</td>
<td>-.00(4)</td>
</tr>
<tr>
<td>-.5</td>
<td>-.028</td>
<td>-.013</td>
<td>-.002</td>
</tr>
<tr>
<td>0</td>
<td>-.018</td>
<td>-.01(0)</td>
<td>-.00(1)</td>
</tr>
<tr>
<td>.5</td>
<td>-.00(7)</td>
<td>-.00(8)</td>
<td>-.00(0)</td>
</tr>
<tr>
<td>.9</td>
<td>.00(2)$^2$</td>
<td>-.00(8)$^3$</td>
<td>.00(1)</td>
</tr>
</tbody>
</table>

| $n=20$   |         |         |         |
| Case 1   |         |         |         |
| -.9      | -.08(4) | -.02(2) | -.00(7) |
| -.5      | -.05(7) | -.01(1) | -.00(2) |
| 0        | -.02(3) | .00(0)  | .00(0)$^3$ |
| .5       | .01(3)  | .00(2)$^2$ | -.00(0)$^2$ |
| .9       | .04(2)  | -.03(0) | .00(6)  |
| Case 2   |         |         |         |
| -.9      | -.01(9) | -.00(5) | -.00(1) |
| -.5      | -.015   | -.00(5) | -.00(0) |
| 0        | -.00(8) | -.00(4)$^2$ | -.00(0) |
| .5       | -.00(1) | -.00(3)$^3$ | .00(0)$^2$ |
| .9       | .00(4)  | -.00(4) | .00(0)$^5$ |

Tables 1 and 2 were calculated by quadratic interpolation on the nearest three points in David's 'Tables of the Correlation Coefficient' (1938). The error of this formula was found using the tables on a point outside this range and is indicated by the brackets and superscripts:

$.00(8)$ means $.008\pm.001$,  $.00(6)^3$ means $.006\pm.003$.

Table 2. Error in exact probability of two-sided nominally 90% confidence interval given by (8) with $r_n=1$ for Example 2.1(a)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$j=1$</th>
<th>$j=2$</th>
<th>$j=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-.18(0)</td>
<td>-.05(6)</td>
<td>-.01(5)</td>
</tr>
<tr>
<td>$\pm .5$</td>
<td>-.17(1)</td>
<td>-.05(7)</td>
<td>-.02(2)</td>
</tr>
<tr>
<td>$\pm .9$</td>
<td>-.15(9)</td>
<td>-.28(9)</td>
<td>-.03(9)$^3$</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-.08(2)$^2$</td>
<td>-.04(9)</td>
<td>.01(0)</td>
</tr>
<tr>
<td>$\pm .5$</td>
<td>-.07(9)$^3$</td>
<td>-.05(0)</td>
<td>-.01(2)$^3$</td>
</tr>
<tr>
<td>$\pm .9$</td>
<td>-.07(2)$^2$</td>
<td>-.06(6)$^4$</td>
<td>-.01(2)$^4$</td>
</tr>
</tbody>
</table>
\begin{align*}
\text{ACCURATE CONFIDENCE INTERVALS} & \quad 55 \\

n=10 & \\
\text{Case 1} & \quad 0 \quad -0.09(0)^2 \quad -0.01(1) \quad 0.00(2) \\
& \quad \pm 0.5 \quad -0.08(7)^2 \quad -0.02(1)^2 \quad -0.00(8) \\
& \quad \pm 0.9 \quad -0.08(2)^2 \quad -0.11(7)^2 \quad -0.00(6)^2 \\
\text{Case 2} & \quad 0 \quad -0.03(3)^2 \quad -0.02(0) \quad -0.00(2)^2 \\
& \quad \pm 0.5 \quad -0.03(5) \quad -0.02(1) \quad -0.00(3) \\
& \quad \pm 0.9 \quad -0.03(3)^2 \quad -0.02(5)^2 \quad -0.00(3)^2 \\

n=20 & \\
\text{Case 1} & \quad 0 \quad -0.04(6)^2 \quad -0.00(0)^2 \quad 0.00(1)^4 \\
& \quad \pm 0.5 \quad -0.04(4)^2 \quad -0.00(8)^2 \quad -0.00(2)^4 \\
& \quad \pm 0.9 \quad -0.04(2)^2 \quad -0.05(2) \quad -0.00(1)^2 \\
\text{Case 2} & \quad 0 \quad -0.01(6)^2 \quad -0.00(8)^2 \quad -0.00(0)^4 \\
& \quad \pm 0.5 \quad -0.01(6) \quad -0.00(9)^4 \quad -0.00(0)^4 \\
& \quad \pm 0.9 \quad 0.01(5)^2 \quad -0.01(0) \quad -0.00(0)^4 \\

(10) \quad q_1(\theta) = x, \quad q_4(\theta) = -\theta/2, \quad q_2(\theta) = \frac{3x + x^2}{12} + \frac{\theta^2 x}{4}.

When \( p_n(\hat{\theta}_n; \theta) \), the density of \( \hat{\theta}_n \), satisfies \( p_n(\hat{\theta}_n; -\theta) = p_n(-\hat{\theta}_n; \theta) \), as is true for Example 2.1(a), then the error of the two-sided confidence interval given by (4), (8) is symmetric in \( \theta \).

\textbf{Example 4.2.} Let \( X_1, \ldots, X_n \) be i.i.d. \( \chi^2_\nu \), \( S = \sum_{i=1}^n X_i \), \( \hat{\theta}_n = S/n \). Taking \( c(\theta) = \theta \) and \( \tau_n = 1 \) gives \( a_2(\theta) = 2\theta \) and

\[
(\theta - \hat{\theta}_n)(n/2\theta)^{1/2} = (m - S)(2m)^{-1/2},
\]

where \( m = n\theta \). Since \( S \sim \chi^2_m \),

\[
g_r(x, \theta) = \theta^{-r/2}(-1)^r g_{r0}(x)
\]

where \( g_{r0}(x) \) denotes \( g_r(x) \) for \( Z_m = (\chi^2_n - m)/m \), given, for example, for \( 1 \leq r \leq 6 \), by (3a) of Fisher and Cornish [3].

Hence

\[
q_1(\theta) = (2\theta)^{1/2} x, \quad q_4(\theta) = (2 + x^2)/3,
\]

\[
q_2(\theta) = -(2\theta)^{-1/2}(2x + x^3)/18,
\]

\textbf{Table 3.} Error in exact probability of one-sided nominally 95\% confidence interval given by (8) with \( \tau_n = 1 \) for Example 4.2

<table>
<thead>
<tr>
<th>( m = n\theta )</th>
<th>( j=1 )</th>
<th>( j=2 )</th>
<th>( j=3 )</th>
<th>( j=4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-0.0795</td>
<td>0.0085</td>
<td>-0.0034</td>
<td>0.0012</td>
</tr>
<tr>
<td>10</td>
<td>-0.0502</td>
<td>0.0034</td>
<td>-0.0007</td>
<td>0.0012</td>
</tr>
<tr>
<td>15</td>
<td>-0.0388</td>
<td>0.0021</td>
<td>-0.00032</td>
<td>0.0004</td>
</tr>
<tr>
<td>20</td>
<td>-0.0324</td>
<td>0.0015</td>
<td>-0.00019</td>
<td>0.0002</td>
</tr>
<tr>
<td>100</td>
<td>-0.0128</td>
<td>0.00026</td>
<td>-0.00007</td>
<td>0.00000(5)</td>
</tr>
</tbody>
</table>
\[ q_i(\theta) = 2\theta^{-1}(-16+7x^2+3x^4)/405 . \]

**Example 4.3.** Returning to Example 2.2, \( a_{21}(\theta) = 1 \) and \( g_\ell(x, \theta) = g_\ell(x) \) do not depend on \( \theta \), so that
\[ q_i(\theta) = \tau_{a_{21}(\theta)}(x) \]
and
\[ V_{j\alpha}(\hat{\theta}_n, x) = \hat{\theta}_n + \tau_{a_{21}(\theta)} n^{-1/2} \left\{ x + \sum_{r=1}^{j'_\alpha} n^{-r/\alpha} g_\ell(x) \right\} . \]

Fisher and Cornish [3] give \( \{ g_\ell(x) \} \) for the case of Student's t-statistic.

**Acknowledgements**

The first draft of this paper was written in 1978 while an Exchange Visitor at the Division of Mathematics and Statistics, C.S.I.R.O., Canberra. I wish to thank D.M.S. for its support, and also David Harte for assistance with Tables 1 and 2.

**Addendum**

Since this paper was written, Winterbottom [5] has published formulae (A.1)-(A.3) equivalent to ours for the case when \( c(\theta) = \theta \), \( \tau_\alpha = 1 \), and \( j = 5 \).

<table>
<thead>
<tr>
<th>His notation ( \hat{\theta}(\xi) )</th>
<th>( \xi )</th>
<th>( \theta )</th>
<th>( T )</th>
<th>( \nu(\theta) )</th>
<th>( (-1)^{r_\xi} \kappa_{r_\xi} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our notation ( V_{j\alpha}(\hat{\theta}_n, x) )</td>
<td>( x )</td>
<td>( \theta )</td>
<td>( \hat{\theta}_n )</td>
<td>( a_{21}(\theta) )</td>
<td>( A_{r_\xi} )</td>
</tr>
</tbody>
</table>

(Equivalently, taking \( c(\theta) = -\theta \), his \( \kappa_{r_\xi} \) is our \( A_{r_\xi} \).)

He applied it to \( \hat{\theta}_n = (\chi^2(\lambda) - n)/n \) with \( \theta = \lambda^2/n \) (NOT \( \lambda/n \) as stated), also to the maximum likelihood estimate, and to Case 2 of Example 4.1, for which he obtains \( q_i(\theta) \) and \( q_3(\theta) \).

**APPENDIX 1**

From Corollary 3.1 of Withers [6] we have

**Lemma 1.** (5), (7) implies (6) with
\[ g_1(x) = A_{11} + A_{32}(x^2 - 1)/6, \]
\[ g_2(x) = A_{22}x/2 + A_{43}(x^3 - 3x)/24 + A_{23}(2x^3 + 5x)/36, \]
\[ g_3(x) = A_{12} + A_{32}(x^2 - 1)/6 + A_{22}A_{43}(-x^3 + 1)/6 \]

\[ + A_{45}(x^4 - 6x^3 + 3)/120 + A_{32}A_{43}(-x^4 + 5x^3 - 2)/24 \\
+ A_{25}(12x^4 - 53x^2 + 17)/324 , \]

and

\[ g_r(x) = \sum_{0 \leq k \leq \lceil (r-1)/2 \rceil} G_r^{k} g_r^{k}(x) \]

where for \( 1 \leq r \leq 6 \), \( G_{r,0} \) and \( g_0(x) \) are given on pages 214, 215 of Fisher and Cornish [3] with \( a = A_{11} \), \( b = A_{22} \), \( c = A_{32} \), \( d = A_{43} \), \( e = A_{44} \), \( f = A_{65} \), \( g = A_{76} \), \( h = A_{87} \): for example,

\[ (G_{4,0})_3 = \text{coefficient in line 3 of IV} = A_{23}A_{33}^2 , \]

and

\[ (g_0(x))_3 = \text{polynomial/divisor in line 3 of IV} = 5(2x^3 - 5x)/72 , \]

while the other \( \{ G_{r,k} \} \) needed for \( 4 \leq r \leq 6 \) are as follows.

For \( r = 4 \):
\[ G_{12} = (A_{23}, A_{44}, 2A_{32}, A_{33}) . \]

For \( r = 5 \):
\[ G_{13} = (A_{11}, A_{44}) , \]
\[ G_{14} = (A_{23}A_{33} + A_{23}A_{33}, A_{44}A_{33} + A_{33}A_{44}, 3A_{33}^2A_{33}) . \]

For \( r = 6 \):
\[ G_{15} = (A_{42}, A_{43}, A_{33} + 2A_{32}, A_{33}) , \]
\[ G_{16} = (2A_{22}A_{23}, A_{22}A_{43} + A_{43}A_{23}, A_{32}A_{33} + 2A_{32}A_{23}, A_{33}A_{43}, A_{33}A_{54} \\
+ A_{32}A_{33}, 2A_{32}A_{43}, 2A_{32}A_{33}A_{43}, 4A_{32}^2A_{33}) . \]

APPENDIX 2

A1. Summary

This contains some formulas for inverting series. Section 2 gives the inverse of

\[ y(\epsilon) = \sum_{r=1}^{\infty} \epsilon^r P_r , \]

as a power series in \( \epsilon \). Section 3 gives the inverse of

\[ x(g) = g + \sum_{r=1}^{\infty} \epsilon^r P_r(g) \]

as a power series in \( \epsilon \), and gives some statistical applications.

A2. The first inversion series

Expressions for the inverse of (A1.1) are well known, e.g. §3.6.25
of Abramowitz and Stegun [1]. However these expressions only give the first few terms of the inverse, as a series in $\varepsilon$. The general term may easily be expressed using the notation of the following lemma.

**Lemma A2.** For $j=0,1,2,\cdots$

$$\left(\sum_{i=1}^{\infty} \varepsilon^i Q_i\right)^j = \sum_{r=0}^{\infty} \varepsilon^r C_{r,j}([Q_i])$$

where

$$C_{r,0}([Q_i]) = \begin{cases} 1, & r=0 \\ 0, & r>0 \end{cases}$$

and for $r\geq j\geq 0$, $C_{r,j}([Q_i]) = \sum Q_{k_{i}} \cdots Q_{k_{j}}$ summed over $k_{1} + \cdots + k_{j} = r$, $k_{1} \geq 1, \ldots, k_{j} \geq 1$, or equivalently

$$C_{r,j}([Q_i]) = \sum (k_{1}, \ldots, k_{r}) Q_{k_{1}}^1 \cdots Q_{k_{r}}^r$$

summed over $k_{1} + \cdots + k_{r} = j$, $k_{1} + 2k_{2} + \cdots + rk_{r} = r$, $k_{1} \geq 0, \ldots, k_{r} \geq 0$, where $(k_{1}, \ldots, k_{r})$ is the multinomial coefficient $j!/(k_{1}! \cdots k_{r}!)$.

For example $C_{r,1}([Q_i]) = Q_r$, $C_{j,j}([Q_i]) = Q_{j}^1$, $C_{j+1,j}([Q_i]) = jQ_{j+1}^1Q_{j}^1$,

$$C_{j+2,j}([Q_i]) = \binom{j}{2} Q_{j-1}^1Q_{2}^1 + \binom{j}{1} Q_{j}^1,$$

$$C_{j+3,j}([Q_i]) = \binom{j}{2} Q_{j-1}^1Q_{3}^1 + \binom{j}{1} jQ_{j-2}^1Q_{3}^1 + j(j-1) Q_{j-2}^2Q_{2}^1Q_{3}.$$

**Theorem A1.** When both series converge, the inverse of

$$y(\varepsilon) = \sum_{r=1}^{\infty} \varepsilon^r P_{r},$$

is given for $P_{1} \neq 0$ by

$$\varepsilon(y) = \sum_{r=1}^{\infty} y^r Q_{r},$$

where $Q_{r}$ is defined recursively by $Q_{1} = P_{1}^{-1}$,

$$Q_{r} = -P_{1}^{-1} \sum_{i=1}^{r} P_{i} C_{r-i}([Q_i]), \quad r>1.$$

**Proof.** Set $y = y(\varepsilon)$, $\varepsilon = \varepsilon(y)$. Then

$$P_{1} \varepsilon = y - \sum_{r=2}^{\infty} P_{r} \varepsilon^{r}.$$ But

$$\varepsilon^{r} = \sum_{r=4}^{\infty} g^{r} C_{r}([Q_i]).$$
An alternative formula for $P_r$ was given by McMahon in 1894. An extension of his result to the problem of expressing a power of $e(y)$ as a series in $\{y^r\}$ is given in Part IX of David, et al. [2]. Their Table 9 may be used as an alternative to Theorem 1 to obtain $Q_r$ for $r \leq 11$.

A3. The second inversion series

Let $\{P_r\}$ be functions on $R$ with derivatives $\{P_r^{(j)}\}$.

**Theorem A2.** When both series converge, the inverse of

$$x(g) = g + \sum_{r=1}^{\infty} e^{r}P_{r}(g)$$

is given by

$$g(x) = x + \sum_{r=1}^{\infty} e^{r}Q_{r}(x)$$

where $Q_{r}(x)$ is defined recursively by

(A2.3) $Q_{r}(x) = -\frac{\sum_{j=0}^{r-1} \sum_{k=j}^{r-1} P_{r-k}^{(j)}(x)C_{k}(\{Q_{r}(x)\})}{j!}$

and

$$P_{r}^{(j)}(x) = (d/dx)^{j}P_{r}(x).$$

**Proof.** $Q_{r}(x)$ is the coefficient of $e^{r}$ in the Taylor series expansion for

$$g = x - \sum_{r=1}^{\infty} e^{r}P_{r}(g).$$

The first five $Q_{r}$ are as follows.

$Q_{1} = -P_{1}, \quad Q_{2} = -P_{2} + P_{1}^{(1)}P_{1},$

$Q_{3} = -P_{3} - P_{2}^{(2)}Q_{1} - P_{2}^{(3)}Q_{2} - P_{2}^{(3)}Q_{3}/2$

$= -P_{3} + P_{2}P_{2}^{(1)} + P_{2}^{(1)}P_{1} - P_{2}^{(1)}P_{1} - P_{2}^{(1)}P_{1}/2,$

$Q_{4} = -P_{4} - P_{3}^{(2)}Q_{1} - P_{3}^{(2)}Q_{2} - P_{3}^{(2)}Q_{3} - P_{3}^{(2)}Q_{4}/2 - P_{3}^{(2)}Q_{1}Q_{2} - P_{3}^{(2)}Q_{1}/6,$

$Q_{5} = -P_{5} - P_{4}^{(3)}Q_{1} - P_{4}^{(3)}Q_{2} - P_{4}^{(3)}Q_{3} - P_{4}^{(3)}Q_{4} - P_{4}^{(3)}Q_{5}/2 - P_{4}^{(3)}Q_{1}Q_{2} - P_{4}^{(3)}Q_{1}Q_{4}/24.$

An alternative formula for $Q_{r}(x)$ involving multivariate Bell polynomials is given by (3) of Riordan [4]. His formula seems more difficult for algebraic manipulation.

As an application in statistics, consider the problem investigated
by Fisher and Cornish [3]. Many standardized asymptotically normal random variables $Y_n$ have $r$th cumulant of the form

$$l_{rn} = O(n^{-r/2}) , \quad r = 1$$
$$1 + l_{rn} = 1 + O(n^{-1}) , \quad r = 2$$
$$l_{rn} = O(n^{1-r/2}) , \quad r > 2 \text{ as } n \to \infty ,$$

(Here $n$ is usually the sample size or associated degrees of freedom.) Under this assumption they showed that $P_n(x) = \Pr(Y_n \leq x)$ satisfies expansions of the form

$$\Phi^{-1}(P_n(x)) = x - \sum_{1}^{\infty} f_r(x, \mathcal{L}_n)$$

and

$$P_n^{-1}(\Phi(x)) = x + \sum_{1}^{\infty} g_r(x, \mathcal{L}_n)$$

where

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp \left( -\frac{y^2}{2} \right) dy$$

and $f_r, g_r$ are polynomials of degree $r + 1$ involving $\mathcal{L}_n = \{l_{rn}\}$ and having magnitude $O(n^{-r/2})$.

They gave the first four $f_r$ and the first six $g_r$, but no expression for the general term. Expressions for $f_5$ and $f_6$ may be obtained from the following application of Theorem A2.

**Corollary A1.** Let $Q_r(x, \{P_i\})$ denote $Q_r(x)$ of (A2.3). Then

$$g_r(x, \mathcal{L}) = Q_r(x, \{-f_i(x, \mathcal{L})\})$$

and

$$f_r(x, \mathcal{L}) = -Q_r(x, \{g_i(x, \mathcal{L})\}) .$$

An expression for $f_r(x, \mathcal{L})$ for general $r$ was given by (2.8) of Withers [6]. This may be used in Corollary 1 to obtain any desired $g_r(x, \mathcal{L})$.

In most instances $\sum_{1}^{\infty} f_r(x, \mathcal{L}_n)$ and $\sum_{1}^{\infty} g_r(x, \mathcal{L}_n)$ can be rewritten in the form $\sum_{1}^{\infty} n^{-r/2} f_r(x)$ and $\sum_{1}^{\infty} n^{-r/2} g_r(x)$. In this case $\{f_r(x)\}$ and $\{g_r(x)\}$ have the same relationship to each other as do $\{f_r(x, \mathcal{L})\}$ and $\{g_r(x, \mathcal{L})\}$, so that $\{f_r(x), 1 \leq r \leq 6\}$ are obtainable using Appendix 1 when (5) holds.

D.S.I.R.
REFERENCES


