Singapore

4:15 Friday, August 29, 2003

1. Introduction

Let \( Y_1, Y_2, \ldots \) be independently identically distributed random variables with mean 0 and variance 1 and let

\[
\beta_3 = \mathbb{E} |Y_i|^3 < \infty.
\]

Also let

\[
W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i
\]

where the \( X_i \) are independent, identically distributed random variables defined by \( X_i = \frac{Y_i}{\sqrt{n}} \) so that

\[
E X_i = 0
\]
and

\[
E X_i^2 = \frac{1}{n}.
\]

My aim is to develop an identity that should be useful for bounding the error of the standard normal approximation to the distribution of \( W_n \), in particular for proving the Berry-Esseen Theorem, which asserts that, with \( \Phi \) the cumulative distribution of the standard normal distribution, given by

\[
\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-\frac{z^2}{2}} dz,
\]
we have, for all \( w_0 \), all \( n \), and all distributions for \( X_1 \),

\[
P\{W_n \leq w_0 \} - \Phi(w_0) \leq \frac{C \beta_3}{\sqrt{n}}
\]
where \( C \) is an absolute constant. In Theorem 1 of Section 2 an expression is derived for the error in the normal approximation to the expectation of \( h(W_n) \) for a fairly arbitrary function \( h \). When I decided to speak about this at the conference, I thought that I would have some more definitive results. However, I decided to make the manuscript public despite its highly incomplete nature because I hope that other people may develop it further. Observe that, except for some vague remarks, this work consists entirely of identities.

We shall need some notation. For any function \( h \) for which the integral in (7) exists, let

\[
N h = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(z) e^{-\frac{z^2}{2}} dz,
\]
and let \( f \) be the particular solution of the differential equation

\[
f'(w) - w f(w) = h(w) - N h
\]
that is given by

\[
f(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} (h(z) - N h) e^{-\frac{z^2}{2}} dz.
\]
At first thought, in order to obtain an upper bound for \( P\{W_n \leq w_0 \} \), we might wish to try to choose an appropriate \( \delta > 0 \), set \( w_1 = w_0 + \delta \) and work with \( h \) specialized to \( h_{w_1} \), defined by

\[
h_{w_1}(w) = 1 \text{ if } w \leq w_1 \text{ but } 0 \text{ if } w > w_1,
\]
in which case \( f \) is specialized to \( f_{w_1} \), given by

\[
f_{w_1}(w) = \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(\min(w_1, w)) (1 - \Phi(\max(w_1, w)) ),
\]
and \( c = -1 \). I hope to place an improved version of this paper on the web, perhaps by the end of September, but I encourage other people to work on it because I am not sure that I shall succeed, although I am essentially sure that the present approach will eventually be successful in getting sharp bounds on the error of the normal approximation.

\[
\beta_3 = E |Y_i|^3 < \infty.
\]

Also let

\[
W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i
\]

where the \( X_i \) are independent, identically distributed random variables defined by \( X_i = \frac{Y_i}{\sqrt{n}} \) so that

\[
E X_i = 0
\]
and

\[
E X_i^2 = \frac{1}{n}.
\]

My aim is to develop an identity that should be useful for bounding the error of the standard normal approximation to the distribution of \( W_n \), in particular for proving the Berry-Esseen Theorem, which asserts that, with \( \Phi \) the cumulative distribution of the standard normal distribution, given by

\[
\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-\frac{z^2}{2}} dz,
\]
we have, for all \( w_0 \), all \( n \), and all distributions for \( X_1 \),

\[
P\{W_n \leq w_0 \} - \Phi(w_0) \leq \frac{C \beta_3}{\sqrt{n}}
\]
where \( C \) is an absolute constant. In Theorem 1 of Section 2 an expression is derived for the error in the normal approximation to the expectation of \( h(W_n) \) for a fairly arbitrary function \( h \). When I decided to speak about this at the conference, I thought that I would have some more definitive results. However, I decided to make the manuscript public despite its highly incomplete nature because I hope that other people may develop it further. Observe that, except for some vague remarks, this work consists entirely of identities.

We shall need some notation. For any function \( h \) for which the integral in (7) exists, let

\[
N h = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(z) e^{-\frac{z^2}{2}} dz,
\]
and let \( f \) be the particular solution of the differential equation

\[
f'(w) - w f(w) = h(w) - N h
\]
that is given by

\[
f(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} (h(z) - N h) e^{-\frac{z^2}{2}} dz.
\]
At first thought, in order to obtain an upper bound for \( P\{W_n \leq w_0 \} \), we might wish to try to choose an appropriate \( \delta > 0 \), set \( w_1 = w_0 + \delta \) and work with \( h \) specialized to \( h_{w_1} \), defined by

\[
h_{w_1}(w) = 1 \text{ if } w \leq w_1 \text{ but } 0 \text{ if } w > w_1,
\]
in which case \( f \) is specialized to \( f_{w_1} \), given by

\[
f_{w_1}(w) = \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(\min(w_1, w)) (1 - \Phi(\max(w_1, w)) ),
\]
and \( c = -1 \). I hope to place an improved version of this paper on the web, perhaps by the end of September, but I encourage other people to work on it because I am not sure that I shall succeed, although I am essentially sure that the present approach will eventually be successful in getting sharp bounds on the error of the normal approximation.
2. A basic identity.

The proof of Theorem 1, below, is based on two simple ideas. The first, Lemma 1, is very old, only integration by parts together with the solution of a first order ordinary differential equation. Mallows[] and I used the necessity independently and nearly simultaneously, he apparently first, in evaluating the expected sum of squares of errors for a fairly large class of estimates of the mean of a multivariate normal distribution. The second is an identity, given as (20) below, which is related to an approach of Bolthausen[1984].

Lemma 1. A real random variable $Z$ has a standard normal distribution if and only if, for all continuous and piecewise continuously differentiable functions $f$ for which $E | f'(Z) | < \infty$,

$$(12) \quad E(f'(Z) - Z f(Z)) = 0.$$ 

Of course much less is required for normality to hold. For example it is well known that it is sufficient that (13) hold for all polynomials or for all complex exponentials $f(x) = e^{itx}$, and in my first application of exchangeable pairs I first used complex exponentials. After some time, it occurred to me that the argument I was using worked just as well with an arbitrary function, which saved me the trouble of inverting the characteristic function.

Theorem 1. Let $h$ be a function that does not increase too fast at infinity and is continuous except possibly at one point $w_1$, and let

$$(13) \quad c_1 = h(w_1+) - h(w_1-).$$

Then, in the situation described in the first section, with $f$ defined by (9)

$$(14) \quad E h(W_n) = N h + T_1 - T_2 - T_3,$$

where

$$(15) \quad T_1 = E W_n X_{n+1} \int_0^1 (f'((W_n + t X_{n+1}) - f'(W_n))) dt,$$

$$(16) \quad T_2 = E W_n X_{n+1} \int_0^1 (1-u) f''(W_n + u X_{n+1}) du,$$

and

$$(17) \quad T_3 = c_1 E n X_{n+1}^2(\{W_n \leq w_1 < W_n + X_{n+1}\} - \{W_n + X_{n+1} \leq w_1 < W_n\})(1 - \frac{w_1-W_n}{X_{n+1}}).$$

The proof is based on an identity, (20) below, which uses an argument similar to that in the formula (2.6) of a paper by Bolthausen [1984]. Because the $X_i$ are exchangeable, for every continuous function $f$ that is continuously differentiable except possibly at one point $w_0$ and for which the required expectations exist,

$$(18) \quad 0 = E(X_{n+1} - X_i) f(\sum_{i=1}^{n} X_i) = E(X_{n+1} - X_i) f(W_n + X_{n+1}).$$

Summing from $i = 1, \ldots, n$, and using the fact that $X_{n+1}$ is independent of $W_n$ and $E X_{n+1} = 0$ and $E X_{n+1}^2 = 1$

$$(19) \quad 0 = E(n X_{n+1} - W_n) f(W_n + X_{n+1})
\quad = E(n X_{n+1} f(W_n + X_{n+1} - f(W_n)) - W_n f(W_n + X_{n+1}))
\quad = E(n X_{n+1}^2 \int_0^1 f'(W_n + t X_{n+1}) dt - W_n f(W_n + X_{n+1}))
\quad = E n X_{n+1}^2 \int_0^1 f'(W_n + t X_{n+1}) dt + E f'(W_n - W_n f(W_n)) + E W_n f(W_n) - f(W_n + X_{n+1}).$$

At the second equality sign I have subtracted $n E X_{n+1} f(W_n)$, which is 0 because $E X_{n+1} = 0$ and $W_n$ is independent of $X_{n+1}$. At the third equality sign I have expressed the difference of the values of $f$ at two points as the integral of its derivative. At the fourth equality sign I have added and subtracted both $f'(W_n)$ and $E W_n f(W_n)$ and used the fact that $E X_{n+1}^2 = 1$ and $W_n$ is independent of $X_{n+1}$. Using the differential equation (8) to evaluate the second term on the extreme right hand side of (19) and rearranging terms, we obtain an expression for the error in the normal approximation for $E h(W_n)$.

$$(20) \quad E h(W_n) - N h = E f'(W_n) - W_n f(W_n))
\quad = E W_n f(W_n + X_{n+1} - f(W_n)) - E n X_{n+1}^2 \int_0^1 f'(W_n + t X_{n+1}) - f'(W_n)) dt = T_1 - B$$

where
(21) \[ T_1 = E(W_n(f(W_n + X_{n+1}) - f(W_n))) \]
and
(22) \[ B = E n X_{n+1}^2 \int_0^1 f'(W_n + t X_{n+1}) - f'(W_n) \, dt \]

In order to obtain the expression (15) for \( T_1 \) we express the difference in the values of \( f \) at two points as the integral of \( f' \) between those two points,
(23) \[ T_1 = E W_n(f(W_n + X_{n+1}) - f(W_n)) = E W_n X_{n+1} \int_0^1 f'(W_n + t X_{n+1}) \, dt \]

Differentiating (8),
(24) \[ f''(w) = w f'(w) + f(w) + h'(w) = w(h(w) - N h) + h'(w) + (1 + w^2) f(w) \]
for all \( w \neq w_1 \). Thus \( B \) in (22) is given by
(25) \[ B = E n X_{n+1}^2 \int_0^1 (f'(W_n + t X_{n+1}) - f'(W_n)) \, dt \]
\[ = E n X_{n+1}^2 \left( \int_0^1 f''(W_n + u X_{n+1}) \, du + c_1(\{ W_n \leq w_1 < W_n + t X_{n+1} \} - \{ W_n + t X_{n+1} \leq w_1 < W_n \}) \right) \]
\[ = E n X_{n+1}^2 \int_0^1 (1 - u) f''(W_n + u X_{n+1}) \, du \]
\[ + c_1 E n X_{n+1}^2 (\{ W_n \leq w_1 < W_n + X_{n+1} \} - \{ W_n + X_{n+1} \leq w_1 < W_n \}) (1 - \frac{w_1 - W_n}{X_{n+1}}) \]

At the first equality sign, I have replaced the difference of \( f' \)at two values by the integral of \( f'' \) between those two values plus the term due to the discontinuity of \( f' \) at \( w_1 \). At the second equality sign, I have exchanged the order of integration over \( t \) and \( u \) and then carried out the integration over \( t \). Consequently
(26) \[ B = T_2 + T_3 \]
where \( T_2 \) and \( T_3 \) are given by (16) and (17).

3. The problem of obtaining good bounds for the error of the normal approximation.

We hope eventually to find a good upper bound for the cumulative distribution function of \( W_n \) minus the normal approximation to that c.d.f. Of course, to the extent that we succeed, a lower bound will follow by the essential symmetry of the problem, using the fact that
(27) \[ P(W \leq w_0) = 1 - P(W > w_0) = 1 - P(-W < -w_0), \]
where the distinction between \(< \) and \( \leq \) will not affect the bound. I separate the cases because we may wish to use different functions \( f \) for the two bounds.

Let us look at the remainder term \( R \) in Theorem 1, given by
(28) \[ R = E h(W_n) - N h = T_1 - T_2 - T_3, \]
of Theorem 1, where
(29) \[ T_1 = E W_n X_{n+1} \int_0^1 f'(W_n + t X_{n+1}) - f'(W_n) \, dt, \]
(30) \[ T_2 = E W_n X_{n+1} \int_0^1 (1 - u) f''(W_n + u X_{n+1}) \, du, \]
and
(31) \[ T_3 = c_1 E n X_{n+1}^2 (\{ W_n \leq w_1 < W_n + X_{n+1} \} - \{ W_n + X_{n+1} \leq w_1 < W_n \}) (1 - \frac{w_1 - W_n}{X_{n+1}}). \]
The term \( T_1 \) is typically of the exact order of \( 1/n \). The term \( T_2 \) is typically of the exact order of \( 1/\sqrt{n} \) but could be evaluated approximately, presumably with an error \( O(1/n) \) by a repeated application of the identity (14). The term \( T_3 \) is typically of the exact order of \( 1/\sqrt{n} \) and cannot be evaluated further without detailed quantitative information about the distribution of the \( Y_i \).

References:
E. Bolthausen: An estimate of the remainder in a combinatorial central limit theorem. Zeitschrift fuer Wahrscheinlichkeits-
und verwandte Gebiete. v. 66 pp 379-386.
C. Mallows: Some remarks on \( C_p \), Technometrics