Edgeworth Expansions in Non-regular Cases and Their Applications to Bootstrap¹

GUTTI JOGESH BABU
Pennsylvania State University, University Park, U. S. A.

Abstract

Edgeworth expansions for sums of non-identically distributed random vectors and of random vectors with lattice and non-lattice coordinates are reviewed. Their applications to errors-in-variables models, least absolute deviation estimators etc., are presented. Applications to bootstrap are also discussed.

1 Introduction

Let $f$ be the Fourier transform or characteristic function of a distribution $F$ and $\psi$ be the characteristic function of another distribution $\Psi$. Consider the formal expansion

$$f(t) = \exp \left[ \log f(t) - \log \psi(t) \right] \psi(t)$$

$$= \exp \left( \sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!} \right) \psi(t).$$

¹Research supported in part by NSA Grant MDA904-90-H-1001 and NSF Grants DMS-9007717 and DMS-9209066.

AMS Subject Classifications: Primary 62E20; Secondary 60F05

Key words and Phrases: mean absolute deviation; non-lattice distribution; errors-in-variables; $t$-type statistics
Let $\kappa_r$ and $\gamma_r$ denote cumulants of $F$ and $\Psi$ respectively. The Fourier transform of $(-D)^r\Psi$ is given by the function $g_r$, where $D$ denotes the differential and $g_r(t) = (it)^r\psi(t)$. By formally inverting the right hand side of the above equation term by term one obtains

$$F(x) = \left(\exp\left(\sum_{r=1}^{\infty} \frac{(-D)^r}{r!} \right)\right)^{\kappa_r - \gamma_r} \Psi(x).$$

This is called Charlier differential series (Charlier, 1906). The formal expansion, with normal distribution as the developing function, was considered by Chebyshev (1890), Edgeworth (1905) and Charlier (1905). The choice of $\Psi$ has no effect on convergence, but it clearly has great influence on the quality of approximation by the first few terms. For details on the history of such expansions see Wallace (1958). The approximation by finite number of terms have come to be known as Edgeworth expansions.

In their fundamental paper, Bhattacharya and Ghosh (1978) show the validity of formal Edgeworth expansions for a wide class of statistics. Their paper is to a great extent, responsible for the renewed interest in Edgeworth expansions. Applications to the asymptotic theory of bootstrap can be cited as another reason for the recent spurt in work on Edgeworth expansions. In a series of papers, Babu and Singh (1983, 1984) and Singh and Babu (1990) have shown superiority of the bootstrap method for a wide class of statistics in approximating the sampling distribution, using the techniques and results of Bhattacharya and Ghosh (1978) on Edgeworth expansions. This class includes statistics which can be written as a smooth function of sample mean of i.i.d. random vectors.

To state the main result of Bhattacharya and Ghosh (1978), let $\bar{Z}_n$ denote the sample mean of $n$ independent identically distributed random vectors $Z_1, \ldots, Z_n$ in $\mathbb{R}^k$ with mean $\mu$. Let $P_n$ denote the distribution of $\sqrt{n}(\bar{Z}_n - \mu)$ and let $g$ be a measurable function. Under some conditions on $Z_1$, they show that

$$\int g d(P_n - \Psi) = o(n^{-s/2}) + \delta_n(g),$$

where $\delta_n(g)$ goes to zero at certain rate,

$$\Psi_{\sigma,n}(x) = \Phi_{\Sigma}(x) + \phi_{\Sigma}(x) \left(\sum_{i=1}^{s} n^{-i/2} p_i(x)\right),$$

and $p_i$ are linear combinations of Chebyshev-Hermite polynomials with coefficients depending on the moments of $Z_1$. Using this result they establish that

$$P(\sqrt{n}(H(\bar{Z}_n) - H(\mu)) \leq x) = \Phi_\sigma(x) + \phi_\sigma(x) \left(\sum_{i=1}^{s} n^{-i/2} q_i(x)\right) + o(n^{-s/2}),$$
holds under some conditions on \( Z_1 \) and on the smooth function \( H \), where \( q_i \) are polynomials.

In this paper we review some of the recent work done at Pennsylvania State University on Edgeworth Expansions.

In addition to conditions on the moments of \( Z_1 \), these results are proved assuming:

a) \( Z_1 \) satisfies Cramér's condition

\[
\limsup_{\|t\| \to \infty} |E(e^{it'Z})| < 1,
\]

and

b) the random variables \( Z_1, \ldots, Z_n \) are independent and identically distributed.

For one term expansions condition a) can be relaxed to,

a') the distribution of \( Z_1 \) satisfies strongly non-lattice condition,

\[
|E(e^{it'Z})| < 1 \quad \text{for all} \quad t \neq 0,
\]

i.e., the distribution of \( Z_1 \) is not concentrated on countably many parallel hyperplanes.

For several statistics of interest, some of these conditions are violated. For example, condition a) or a') is violated, if a statistic is normalized using \( L_1 \)-estimate of scale rather than the traditional \( L_2 \)-estimate. Condition b) does not hold in the context of errors-in-variables regression.

For such statistics the standard results on Edgeworth expansions are not applicable. These non-regular cases are discussed in the next few sections.

2 Presence of Lattice Structure

Ratio Estimator

The ratio estimator \( \hat{\theta} = \bar{Y}_n / \bar{X}_n \) of \( \theta = E(Y)/E(X) \), is often encountered in sample surveys, where \( \bar{V}_n \) denotes the sample mean of \( V_1, \ldots, V_n \). Quite often the auxiliary variable \( X \) has lattice distribution and the variable of interest \( Y \) has a continuous distribution. The difference \( \hat{\theta} - \theta \) can be rewritten as

\[
\hat{\theta} - \theta = H(\bar{Z}_n) = \frac{\bar{W}_n}{E(X)} \left( 1 + \frac{E(X) - \bar{X}_n}{\bar{X}_n} \right)
\]

where \( Z_i = (W_i, X_i) \), \( H \) is a smooth function and the random variable \( W_i = Y_i - \theta X_i \) has continuous distribution. Notice that the leading term has
continuous distribution and the lattice variable $X$ occurs only in quadratic
or higher order terms in the Taylor series expansion
\[ H(\bar{Z}_n) - H(\mu) = (\bar{Z}_n - \mu)^\ell(\mu) + (\bar{Z}_n - \mu)^L(\bar{Z}_n - \mu) + \cdots, \]
where $\mu = (E(W), E(X))$. In this case Edgeworth expansions are valid
by the Corollary of Babu and Singh (1989). Babu (1991) generalized the
results of Babu and Singh (1989) to s-term Edgeworth expansions, when
the lattice structure does not appear in the first few terms of the Taylor
series expansion of the function $H$ at $\bar{Z}_n$ around $\mu$. This is achieved by
convoluting the lattice variable with a ‘small smooth’ random variable and
removing its effect later.

Mean absolute deviations

Easy access to inexpensive computing facilities of late, has renewed interest
in the statistical methodology based on $L_1$-norm, instead of the traditional
methods based on $L_2$-norm or least squares. Recent developments in robust
procedures in statistical inference, is also to some extent responsible for
this trend. The use of $L_1$-norm in estimation and tests of significance is not
new, the idea goes back to Galileo, but the complexity of the distributions
involved stood in the way of its use in practical applications. The least-
squares (or $L_2$) estimation, which also has its roots in celestial mechanics
gained popularity due to its mathematical tractability. In this section we
consider $t$-type statistics, where the usual least-squares scaling factor is
replaced by the one based on mean absolute deviations.

To be precise, let $X_1, \ldots, X_n$ be independent random variables with a
common continuous distribution $F$ having mean $\mu$ and variance $\sigma^2$. For
scale parameter we use the mean absolute deviation
\[ M_n = \frac{1}{n} \sum_{i=1}^{n} |X_i - X_n|, \]
instead of the classical root mean square deviation
\[ s_n = \left( \frac{1}{n} \sum_{i=1}^{n} (X_i - X_n)^2 \right)^{1/2}. \]

Consider the robust student’s $t$-type statistic, $H_n = \sqrt{n}(\bar{X}_n - \mu)/M_n$.
Suppose $F$ has a derivative $f$ in a neighborhood of $\mu$ and suppose $f(\mu)$ is
positive. If for some $a > 0$ and $b > 0$,
\[ |f(x) - f(\mu)| \leq a|x - \mu|^b, \]
for $x$ near $\mu$, then it can be shown (see Babu and Rao, 1992) that
\[ H_n = \sqrt{n}(H(\bar{Z}_n) - H(E(Z_1))) + R_n, \]
where
\[ Y_i = \sigma^{-1}(X_i - \mu), \quad \gamma^{-1} = E|Y_1|, \]
\[ Z_i = (Y_i, |Y_i| - \gamma^{-1}, I(Y_i \leq 0) - F(\mu)), \]
\[ H(x, y, z) = \alpha x (1 - y_2 - x_2 (2F(\mu) - 1)) \]
\[-x_2 (2x + x_2 f(\mu) + \gamma^2 (y_2 + x_2 (2F(\mu) - 1))^2), \]

and for some \( K > 0 \) and \( \epsilon > 0 \),
\[ P(|R_n| > Kn^{-1-\epsilon}) = o(n^{-1}). \]

Note that the last component of \( Z_1 \) is a lattice variable. The standard results on Edgeworth expansions are not applicable. The parameter \( \gamma^{-1} \) is called the shape parameter and it is generally known. For example, its value for the normal family is \( \sqrt{2/\pi} \) and is \( 1/\sqrt{2} \) for the double exponential family. The results of Babu and Singh (1989) lead to valid one term Edgeworth expansion for \( H_n \). A similar result holds for bootstrapped version of \( H_n \). These results establish that the bootstrap automatically corrects for skewness of the sampling distribution of \( H_n \) provided the shape parameter is known. The case when the shape parameter is unknown is presently under investigation.

In the same spirit, Bai and Rao (1991) obtained \( s \)-term Edgeworth expansions for \( \sqrt{n} (\ell \sigma \ell')^{-1/2} (H(\bar{Z}_n) - H(\mu)) \) through a conditional argument, where \( \ell = \text{grad}(H(\mu)) \neq 0 \). In addition to the conditions on the moments \( E|Z_1|^s < \infty, \quad s \geq 3, \) and \( \ell_1 > 0 \), they assume the partial (conditional) Cramér’s condition
\[ \limsup_{i \to \infty} \mathbb{E} \left| \mathbb{E}(e^{itZ_1} | Z_{21}, \ldots, Z_{k1}) \right| < 1, \]
where \( Z_1 = (Z_{11}, \ldots, Z_{k1}) \) and \( \ell_1 \) is the first coordinate of \( \ell \).

However, this method does not lead to expansions for the multivariate means. Another disadvantage of using the partial Cramér’s condition is that it does not hold in general, even in the limit, for the bootstrapped sample. This prevents applications to bootstrap.

**Conditions on the Moments**

In all the results mentioned above the assumption on the moments is too restrictive, needing existence of moments higher than those required to define the expansions. Consider the simple example of Student’s \( t \)-statistic,
\[ t = \sqrt{n}(\bar{X}_n - \mu)/s_n, \]
which can be written as
\[ t = H \left( (\bar{X}_n - \mu, (X_n - \mu)^2) \right), \]
where \( H \) is a smooth function on \( \mathbb{R}^2 \). Standard proof of one term expansion requires at least three moments of \( (X_1, X_1^2) \), i.e., six moments of \( X_1 \). But only three moments of \( X_1 \) appear in the actual expansion. Hall (1987) showed that Edgeworth expansions for Student’s \( t \)-statistic are valid under minimal moment conditions. Babu and Bai (1993) generalized this
and obtained s-term expansions under minimal moment conditions and conditional Cramér's condition for statistics which are smooth functions of means of random vectors. See also Bhattacharya and Ghosh (1988). Chibisov (1980 and 1981) obtained similar results under Cramér's condition, for statistics which can be written as a polynomial in sample means. However, there is a major flaw in the proof given by Chibisov, which we could not fix.

3 Errors in Variables Regression

In this section we consider Edgeworth expansions for statistics, based on independent but not identically distributed random vectors. Second order correctness for bootstrap methods for sums of independent but not identically distributed random variables are considered by Liu (1988) in the univariate case, using Edgeworth expansions. This situation occurs in errors-in-variables (EIV) regression context, even when the residuals are known to be i.i.d. The EIV models have been studied extensively in the literature. See Deeming (1968), Fuller (1987), Gleser (1985) and Jones (1979) among others. Consider the simple linear EIV model \((X_i, Y_i)\):

\[ X_i = u_i + \delta_i, \quad Y_i = v_i + \epsilon_i, \]

where \((\delta_i, \epsilon_i)\) are independent mean zero random vectors and \(u_i\) and \(v_i\) are unknown nuisance parameters. If \(u_i\) are known, then the analysis becomes simpler, though it leads to consideration of sums of independent but not identically distributed random vectors. In many applications to fields like ecology and astronomy, very little is known about the values of \(u_i\) except that on the average they behave well. In this paper \(u_i\) are assumed to be unknown. We are interested in the linear relation \(v_i = \alpha + \beta u_i\). Let \((\delta_i, \epsilon_i)\) be independent copies of \((\delta, \epsilon)\), and \(\sigma_{\delta}\) and \(\sigma_{\epsilon}\) respectively denote the standard deviations of \(\delta_1\) and \(\epsilon_1\). Further let \(\lambda = \sigma_{\epsilon}^2/\sigma_{\delta}^2\). It is well known (see for example Jones, 1979) that the least squares estimators of \(\beta\) and \(\alpha\) are given by

\[ \hat{\beta}_1 = h + \text{sign}(S_{XY})(\lambda + h^2)^{1/2} \quad \text{and} \quad \hat{\alpha}_1 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n, \]

when \(\lambda\) is known, where \(h = (S_{YY} - \lambda S_{XX})/2S_{XY}\),

\[ S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n), \]

\[ S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \quad \text{and} \quad S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2. \]

Instead, if \(\sigma_{\delta}\) alone is known, \(S_{XX} > n\sigma_{\delta}^2\) and \(S_{YY}(S_{XX} - n\sigma_{\delta}^2) > S_{XY}^2\), then the least squares estimators of \(\beta\) and \(\alpha\) are given by

\[ \hat{\beta}_2 = S_{XY}/(S_{XX} - n\sigma_{\delta}^2), \quad \text{and} \quad \hat{\alpha}_2 = \bar{Y}_n - \hat{\beta}_2 \bar{X}_n. \]
On the other hand, if $\sigma$ alone is known, $S_{XX} (S_{YY} - n\sigma^2) > S_{XY}^2$ and $S_{YY} > n\sigma^2$, then the least squares estimators of $\beta$ and $\alpha$ are given by

$$\hat{\beta}_3 = (S_{YY} - n\sigma^2) / S_{XY}, \quad \hat{\alpha}_3 = \bar{Y} - \hat{\beta}_3 \bar{X}.$$

Simple algebra leads to the fact that $\hat{\beta}_r$ and $\hat{\alpha}_r$, $r = 1, 2, 3$ can be written as smooth functions of the mean of

$$\xi_{jn} = (\varepsilon_j^2, \delta_j^2, \varepsilon_j \delta_j, \varepsilon_j, \delta_j, u_j \varepsilon_j, u_j \delta_j).$$

Edgeworth expansions for $\xi_{jn}$ lead to those of $\hat{\beta}_r$. But standard results on Edgeworth expansions are not applicable for two reasons. The first one being that $\xi_{jn}$ are not identically distributed and the second one being that components of $\xi_{jn}$ are linearly dependent. But on the average, the vectors $\xi_{jn}$ behave very well, provided for some $\nu \geq 3$,

$$\sum_{jn} u_{jn} = 0, \quad \frac{1}{n} \sum_{jn} u_{jn}^2 \to \eta > 0 \quad \text{and} \quad \sup_n \frac{1}{n} \sum |u_{jn}|^{\nu} < \infty.$$

Babu and Bai (1992) have shown that if $\varepsilon$ and $\delta$ are independent and continuous random variables with finite sixth moment, and if the conditions on $u_{jn}$ hold with $\nu = 3$, then $\sqrt{n}(\hat{\beta}_r - \beta)$ has valid one-term Edgeworth expansion, for $r = 1, 2, 3$. In fact, $\varepsilon$ and $\delta$ need not be independent, but satisfy weak continuity assumptions on the conditional distributions of $\varepsilon$ and $\delta$. If the conditions on $u_{jn}$ hold with $\nu > 3$, then higher order Edgeworth expansions can be obtained. EIV model illustrates the case where both conditions a') and b) are violated.

As is demonstrated in Babu and Singh (1983, and 1984), bootstrap approximations work best for studentized statistics in general. To establish such a result, Babu and Bai (1992) have obtained Edgeworth expansions for studentized statistics $\sqrt{n}(\hat{\beta}_r - \beta) / \hat{\sigma}_r$, $r = 1, 2, 3$, and their bootstrapped versions. The scaling factors $\hat{\sigma}_r$ are obtained using jackknife type arguments and are given by,

$$\hat{\sigma}_1^2 = n\hat{\beta}_1^2 \psi \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2 - \lambda (X_i - \bar{X}_n)^2 - 2h(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)^2,$$

$$\hat{\sigma}_2^2 = n(\hat{\beta}_2/S_{XY})^2 \sum_{i=1}^{n} ((X_i - \bar{X}_n)(Y_i - \bar{Y}_n - \hat{\beta}_2(X_i - \bar{X}_n)) + \hat{\beta}_2 \sigma_2^2)^2,$$

$$\hat{\sigma}_3^2 = nS_{XY}^2 \sum_{i=1}^{n} ((Y_i - \bar{Y}_n)(Y_i - \bar{Y}_n - \hat{\beta}_3(X_i - \bar{X}_n)) - \sigma_2^2)^2,$$

where $\psi^{-1} = 4S_{XY}^2 (h^2 + \lambda)$.

REFERENCES


Department of Statistics
319 Classroom Building
Pennsylvania State University
University Park, PA 16802
U. S. A.