Analyzing variance in central limit theorems

Mynbayev, Kairat and Darkenbayeva, Gulsim

Kazakh-British University, Satbayev University, International IT University, Al-Farabi Kazakh National University, Institute of Mathematics and Mathematical Modeling

2019
Analyzing variance in central limit theorems

Kairat T. Mynbaev¹,²,a, Gulsim S. Darkenbayeva²,³,⁴,b

¹Kazakh-British University and Satbayev University, Almaty, Kazakhstan
²Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan
³International IT University, Almaty, Kazakhstan
⁴Al-Farabi Kazakh National University, Almaty, Kazakhstan
a e-mail: kairat_mynbayev@yahoo.com, b e-mail: g.spandiyar@gmail.com

Communicated by: Mukhtarbai Otelbayev

Received: 19.08.2019 * Final Version: 19.09.2019 * Accepted/Published Online: 04.10.2019

Abstract. Central limit theorems deal with convergence in distribution of sums of random variables. The usual approach is to normalize the sums to have variance equal to 1. As a result, the limit distribution has variance one. In most papers, existence of the limit of the normalizing factor is postulated and the limit itself is not studied. Here we review some results which focus on the study of the normalizing factor. Applications are indicated.

Keywords. Central limit theorems, convergence in distribution, limit distribution, variance.

1 Introduction

In this paper we review some results concerning central limit theorems (CLTs). The references are by no means comprehensive; in all cases the reader is advised to see the bibliography in the papers we cite. As a point of departure, we use the Lindeberg CLT.

Consider a triangular array \{X_{nt}, t = 1, ..., n, n \in \mathbb{N}\} of random variables defined on the same probability space \((\Omega, \mathcal{F}, P)\), having zero mean \(EX_{nt} = 0\) and variances \(\sigma^2_{nt} = EX_{nt}^2\).

Then the sums \(S_n = \sum_{t=1}^{n} X_{nt}\) under independence have variances \(s^2_n = ES_n^2 = \sum_{t=1}^{n} \sigma^2_{nt}\).

Lindeberg theorem [1]. Let the array \{X_{nt}\} be independent and satisfy

\[\sum_{t=1}^{n} \sigma^2_{nt} = 1.\] (1)
If
\[ \lim_{n \to \infty} \sum_{t=1}^{n} \int_{|X_{nt}| > \varepsilon} X_{nt}^2 dP = 0, \quad \text{for all } \varepsilon > 0, \quad (2) \]
then \( S_n \) converges in distribution to a standard normal variable (with mean 0 and variance \( \sigma^2 = 1 \)).

The main advantage of the Lindeberg theorem, in comparison with previous results, is that it allows for heterogeneity (variances \( \sigma_{nt}^2 \) may be different). Since the publication of this result in 1922 many different developments took place. 1) The independence condition has been relaxed and replaced by various notions of dependence (mixing and linear processes, among others). 2) For (2), weaker versions have been suggested, including the conditional version. 3) Certain applications required the study of expressions that depend on \( X_{nt} \) in a nonlinear fashion, quadratic forms \( \sum_{n,t=1}^{n} a_{nt} X_{nt} X_{ns} \) being the most important case. There are also results on functionals of stochastic processes where the analytical form of the functional is not specified. 4) Finally, for many CLTs their continuous-time analogues have been obtained, which are called functional CLTs or invariance principles. These have been left out completely in our review.

From the applied point of view, the normalization condition (1) is one of the main obstacles. One can argue that if it is not satisfied, then one can consider \( S_n / s_n \) instead of \( S_n \). Convergence in distribution of \( S_n / s_n \) can be achieved in this way but the question about the convergence of \( S_n \) and asymptotic behavior of \( s_n \) remains. It is particularly important to make sure that \( s_n \) does not tend to zero or infinity. In the next section we indicate some researches where the behavior of \( s_n \) is controlled and the limit \( \sigma^2 = \lim_{n \to \infty} \sum_{t=1}^{n} \sigma_{nt}^2 \) is found explicitly.

2 Analyzing variance

For the purpose of analyzing \( s_n \), it is convenient to normalize \( X_{nt} \) by their standard deviations: \( X_{nt} = \sigma_{nt} e_{nt} \). Then \( S_n \) becomes
\[ S_n = \sum_{t=1}^{n} \sigma_{nt} e_{nt}, \quad (3) \]
where the sigmas are deterministic and \( e_{nt} \) are stochastic. In the Lindeberg-Lévy theorem (see [2]) \( \sigma_{nt} \) are of order \( n^{-1/2} \) (which we call classical). The following papers are focussed on relaxing the independence condition and maintain the classical order: [3]–[23]. Davidson [24], [25] does not analyze directly \( s_n \) but allows variances going to zero or infinity.

In [26] the normalizing factor is classical but the expression for \( \sigma^2 \) is not trivial (see Corollary 1). Let \( X_j \) be a linear process
\[ X_j = \sum_{r} c_{j-r} \xi_r, \quad \xi_r \text{ are i.i.d. with mean zero and variance } 1, \sum_{r} c_{r}^2 < \infty. \quad (4) \]
The cumulant $cum(X_{j_1}, ..., X_{j_k})$ is given by $cum(X_{j_1}, ..., X_{j_k}) = d_k \sum c_{j_1-i}...c_{j_k-i}$, where $d_k$ denotes the $k$-th cumulant of $\xi_i$. Letting $c(x)$ denote the Fourier transform of the sequence $c_j$, one finds the $k$-th cumulant spectral function as $f(k)(x_1, ..., x_{k-1}) = d_k c(x_1)...c(x_{k-1})c(-x_1-...-x_{k-1}).$ Consider the CLT for $Y_n = \sum_{j=1}^{n} :X_j^{(n)} :$, where $X_j^{(n)}$ denotes the Wick power of $X_j$ (it is a polynomial of degree $n$). Corollary 1 states that $n^{-1/2}Y_n$ converges in law to the normal distribution with mean 0 and variance

$$\sigma^2 = \sum_{G \in G_2} \int T f^{(n)}(yM^*)dy_1...dy_N.$$

See the definitions of $T$, $G_2$, $n_t$ and $M^*$ in the paper.

Giraitis L. and Taqqu M.S. [27] consider quadratic forms of bivariate Appell polynomials and give $\sigma^2$ in terms of these polynomials. Consider quadratic forms

$$Q_N = \sum_{s,t=1}^{N} b(t-s)P_{m,n}(X_t, X_s),$$

where $P_{m,n}(X_t, X_s)$ is a bivariate Appell polynomial of $X_t, X_s$. Giraitis L. and Taqqu M.S. [27] prove the next theorem:

**Theorem.** Suppose

$$\sum_{l,k,t \in \mathbb{Z}} |b(l)b(k)Cov(P_{m,n}(X_{t},X_{t+t}),P_{m,n}(X_{0},X_{k}))| < \infty.$$

If $b(0) = 0$, suppose in addition that $\sum_{t} |EX_tX_0|^{m+n} < \infty$. Then $N^{-1/2}Q_N$ converges in distribution to a normal variable with mean zero and variance

$$\sigma^2 = \sum_{l,k,t \in \mathbb{Z}} b(l)b(k)Cov(P_{m,n}(X_{t},X_{t+t}),P_{m,n}(X_{0},X_{k})).$$

Ho H.C. and Sun T.C. [28] in a nonlinear situation (non-instantaneous filter) give $\sigma^2$ in terms of the spectral distribution function of a normal stationary process. For a normal stationary process such that $EX_t = 0$ the autocovariances $r_t = EX_nX_{n+t}$ are represented as $r_t = \int_{-\pi}^{\pi} e^{itx}dG(x)$, where $G(x)$ is the spectral distribution function. The process itself is represented as $X_t = \int_{-\pi}^{\pi} e^{itx}Z_G(dx)$, where $Z_G$ is a random Gaussian measure corresponding to $G(x)$. Consider a non-instantaneous filter (a functional) $H$ such that $EH(X_{t1}, ..., X_{t_k}) = 0$.
and $EH(X_{t_1},...,X_{t_d})^2 < \infty$. Put $Y_N = A_N^{-1} \sum_{t=1}^{N} H(X_{t+t_1},...,X_{t+t_d})$. Ho and Sun find conditions for CLT to hold, the normalizing factor $A_N$ being of classical order. Under some conditions they prove that the limits

$$
\sigma_j^2 = \lim_{n \to \infty} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \int \exp \left[ i(m-n)(x_1 + ... + x_j) \right] |\alpha_j(x_1, ..., x_j)|^2 dG(x_1) ... dG(x_j)
$$

exist for each $j \geq k$ and $\sigma^2 = \sum_{j=k}^{\infty} \sigma_j^2 < \infty$ is the variance of the limit normal distribution. The functions $\alpha_j$ arise from Wiener-Ito expansions of $H(X_{t_1},...,X_{t_d})$.

In [29] $s^2_n$ is related to the spectral density of the innovations of the linear process at zero. For the process in (4) put $S_n = \sum_{k=1}^{n} X_k$, $b_{n,j} = c_{j-1} + ... + c_{j-n}$, $b_n^2 = \sum_{j \in \mathbb{Z}} b_{n,j}^2$. Under some conditions

$$
\lim_{n \to \infty} \text{Var}(S_n)/b_n^2 = 2\pi f(0)
$$

and the sequence $S_n/b_n$ converges in distribution to $\sqrt{\eta z}$ where $z$ is standard normal and $\eta$ is defined in terms of innovations $\xi_k$ and independent of $z$.

To model the behavior of the sigmas in (3), Mynbaev K.T. [30] introduced the $L_p$-approximability notion. The idea is to represent converging sequences of deterministic vectors with functions of a continuous argument. It is realized as follows. Let $1 \leq p < \infty$. The interpolation operator $\Delta_{np} : \mathbb{R}^n \to L_p(0,1)$ is defined by

$$
(\Delta_{np} w)(x) = n^{\frac{1}{p}} \sum_{t=1}^{n} w_1 \left[ \frac{t-1}{n} , \frac{t}{n} \right)(x), \ w \in \mathbb{R}^n.
$$

(5)

If $w_n \in \mathbb{R}^n$ for each $n$ and there exists a function $W \in L_p(0,1)$ such that

$$
\|\Delta_{np} w_n - W\|_{L_p(0,1)} \to 0, \ n \to \infty,
$$

then we say that $\{w_n\}$ is $L_p$-approximable and also that it is $L_p$-close to $W$. Suppose, for simplicity, that the $e_{nt}$ in (3) are i.i.d. with mean zero and variance 1. If the sequence $\sigma_n = (\sigma_{n1},...,\sigma_{nn})$ is $L_2$-close to a function $F \in L_2(0,1)$, then (3) converges in law to a normal variable with variance

$$
V = \int_0^1 F^2(x)dx.
$$

(6)

This result extends to the case when $e_{nt}$ are linear processes with short memory. It would be interesting to obtain something similar in case of processes with long memory.
P.C.B. Phillips and many of his followers use properties of Brownian motion to establish convergence results for regression estimators. Mynbaev K.T. [31] showed that some problems solved using Brownian motion are easier handled applying $L_p$-approximability.

To state the result from [32] on quadratic forms $Q_n(k_n) = \sum k_{nat}X_sX_t$ we need more notation.

Let $A$ be a compact linear operator in a Hilbert space with a scalar product $(\cdot, \cdot)$. The operator $H = (A^*A)\frac{1}{2}$ is called the modulus of $A$, here $A^*$ is the adjoint operator of $A$. The eigenvalues of $H$, denoted $s_i$, $i = 1, 2, \ldots$, and counted with their multiplicity, are called $s$-numbers of $A$. $U$ denotes a partially isometric operator that isometrically maps the range $R(A^*)$ onto the range $R(A)$. Then we have the polar representation $A = UH$. Denote by $r(A)$ the dimension of the range $R(A)$ ($r(A) \leq \infty$).

Let $\{\phi_i\}$ be an orthonormal system of eigenvectors of $H$ which is complete in $R(H)$. Then, we have the representation

$$Ax = \sum_{i=1}^{r(A)} s_i(x, \phi_i)U\phi_i$$

or, denoting $\psi_i = U\phi_i$,

$$Ax = \sum_{i=1}^{r(A)} s_i(x, \phi_i)\psi_i,$$

where $\{\phi_i\}$ and $\{\psi_i\}$ are orthonormal systems, $H\phi_i = s_i\phi_i$, $\lim_{i \to \infty} s_i = 0$. In particular, when $A$ is selfadjoint, $\phi_i$ are eigenvectors of $A$ and $s_i = |\lambda_i|$, where $\lambda_i$ are eigenvalues of $A$.

Let $K \in L_2 \left((0, 1)^2\right)$. For each natural $n$, we define an $(n \times n)$-matrix

$$(\delta_nK)_{ij} = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} K(s, t)dsdt, \quad 1 \leq i, \ j \leq n.$$  

We say that the sequence $\{k_n\}$ is $L_2$-close to $K$ if

$$\left(\sum_{i,j} (k_n - \delta_nK)_{ij}^2\right)^{\frac{1}{2}} = ||k_n - \delta_nK||_2 \to 0.$$  

Unlike the one-dimensional case, where $L_2$-approximability of $\{\sigma_n\}$ is enough to have convergence in distribution, in the two-dimensional case one has to impose a stronger condition on the rate of approximation. One version of such a condition is

$$||k_n - \delta_nK||_2 = o\left(\frac{1}{\sqrt{n}}\right). \quad (7)$$
Define an integral operator by
\[
(Kf)(s) = \int_0^1 K(s,t) f(t) \, dt, \quad f \in L_2(0,1).
\]

**Theorem [32].** Let \( X_j \) from (4) satisfy \( \sum_j |c_j| < \infty \) and let (7) hold. If \( K \) is nuclear, then
\[
Q_n(k_n) \xrightarrow{d} \left( \sum_i c_i \right)^2 \sum_{i \geq 1} s_i u_i^{(1)} u_i^{(2)},
\]
where \( \{u_i^{(1)}\}, \{u_i^{(2)}\} \) are systems of independent (within a system) standard normals, \( s_i \) are \( s \)-numbers of \( K \) and
\[
\text{cov}(u_i^{(1)}, u_j^{(2)}) = (\psi_i, \phi_j) \quad \text{for all } i, j.
\]
If \( K \) is symmetric, then \( u_i^{(1)} = u_i^{(2)} \) for all \( i \).

For more information about history of these results, see [33], [34] and [32]. Note the difference between the limit in (8), which is not a normal variable, and the above results, where the limit of quadratic forms is normal. This is due to the centering in the above results. Centering requires knowledge of means and may be problematic in applications.

Wu W. and Shao X. [35] prove asymptotic normality of
\[
\sum_{1 \leq s < t \leq n} a_{nst} X_s X_t / \sigma_n, \quad \text{where } \sigma_n^2 = \sum_{t=2}^n \sum_{j=1}^{t-1} a_{nst}^2,
\]
and \( X_s \) is a real stationary process with mean zero and finite covariances.

### 3 Some applications

Here we list a couple of applications that illustrate the following point. With expressions of type (6) and (8) at hand one can study the limit distribution further. We call this analysis at infinity.

[36] initiated the study of regressions with slowly varying regressors. The limit variance matrix of the OLS estimator for such regressions is degenerate. The analysis at infinity comes in very handy, see [37].

The main technical problem with a spatial model \( Y_n = \rho W Y_n + X_n \beta + \varepsilon_n \) is that in its reduced form \( Y_n = (I - \rho W_n)^{-1} (X_n \beta + \varepsilon_n) \) there is an inverse matrix \( (I - \rho W_n)^{-1} \) and one has to deduce the properties of the inverse from the assumptions on \( W_n \). Many researchers have been unable to do that and instead imposed high level conditions involving the inverse. Mynbaev K.T. and Ullah A. [38] and Mynbaev K.T. [39] gave the first derivation
of the asymptotic distribution of the OLS estimator for spatial models (without and with exogenous regressors, resp.) that does not rely on high level conditions.

Most of K.T. Mynbaev's contributions are collected in [40]. In particular, for the purely spatial model in Chapter 5 it is shown that the said model violates the habitual notions in several ways:

1. the OLS asymptotics is not normal,
2. the limit of the numerator vector is not normal,
3. the limit of the denominator matrix is not constant,
4. the normalizer is identically 1 (that is, no scaling is necessary) and
5. there is no consistency.

References


Кыпшылық теоремалардың дисперсиялық қолданылуын ұсыну үшін қолданылады. Олардың қолданылуын құрастыру үшін қолданылады.

Құлтық сөздер. Орталық шектік теоремалар, үлестірім бойынша жинақталу, дисперсия.
Мынбаев К.Т., Даркенбаева Г.С. АНАЛИЗ ДИСПЕРСИИ В ЦЕНТРАЛЬНЫХ ПРЕДЕЛЬНЫХ ТЕОРЕМАХ

Центральные предельные теоремы связаны со сходимостью по распределению сумм случайных величин. Обычный подход заключается в нормализации сумм так, чтобы иметь дисперсию, равную единице. В результате этого предельное распределение имеет дисперсию, равную единице. Во многих работах существование нормализующего фактора постулируется, а сам предел не изучен. Здесь мы рассмотрим некоторые результаты, которые сосредоточены на изучении коэффициента нормализации. Указаны их области применения.

Ключевые слова. Центральные предельные теоремы, сходимость по распределению, предельное распределение, дисперсия.