Three Elementary Proofs of the Central Limit Theorem with Applications to Random Sums*

T. Cacoullos, N. Papadatos and V. Papathanasiou

Abstract

Three simple proofs of the classical CLT are presented. The proofs are based on some basic properties of *covariance kernels* or w-functions in conjunction with bounds for the total variation distance. Applications to random sum CLT's are also given.

AMS 1991 *subject classifications.* Primary 60F15; secondary 60F05.
Key words and phrases: covariance kernels, CLT, random sums.

1. Introduction

Consider a (real) absolutely continuous r.v. X with d.f. F, density f, mean μ, and finite variance σ^2. The (characterizing) covariance kernel $w(\cdot)$ is defined for every x in the interval support of X by the relation

$$\sigma^2 w(x) f(x) = \int_{-\infty}^{x} (\mu - t) f(t)\, dt = \int_{x}^{\infty} (t - \mu) f(t)\, dt; \tag{1.1}$$

w appears in the basic *covariance identity* ([4], Lemma 3.1)

$$\mathrm{Cov}[X, g(X)] = \sigma^2 E[w(X) g'(X)], \tag{1.2}$$

provided that the (otherwise arbitrary) absolutely continuous function g satisfies $E|w(X)g'(X)| < \infty$. Interestingly enough, the same w-function appears both in the upper and lower bounds on the variance of $g(X)$.

The purpose of the paper is to give elementary proofs of the CLT in the i.i.d. (univariate) case, by only using properties of the covariance kernel w. It should be noted that these proofs are heavily dependent on some previous

*Work partially supported by the Greek Ministry of Industry, Energy and Technology, under Grant 1369/95.

results ([4],[6], [7] and [3]). Furthermore, multivariate extensions have been given by [9] and [5].

The first proof shows the weak convergence to the standard normal d.f. Φ of the d.f. F_n of the standardized partial sums

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}.$$

The assumption used here is that the covariance kernel w of X (where X has the same d.f. as $X_1, X_2, \ldots$) has finite variance:

$$\text{Var}[w(X)] < \infty. \tag{1.3}$$

The second proof is stronger, since assumption (1.3) is unnecessary and, at the same time, the conclusion is strengthened to

$$\varrho(F_n, \Phi) \to 0, \text{ as } n \to \infty. \tag{1.4}$$

In this paper, $\varrho(F, G)$ denotes the total variation distance between the d.f.'s F and G (or the corresponding r.v.'s X and Y), defined by

$$\varrho(F, G) = \sup_A \{|F(A) - G(A)|, A \text{ Borel}\}, \tag{1.5}$$

where $F(A) = P[X \in A]$, $G(A) = P[Y \in A]$.

In the third proof, the stronger assumption (1.3) is used in order to obtain the bound

$$\varrho(F_n, \Phi) \leq c/\sqrt{n}, \tag{1.6}$$

for some constant c (see (2.10)), depending only upon the d.f. of X. Obviously (1.6) implies (1.4), and furthermore it gives a bound on the rate of convergence.

It is worth noting that the last technique is also applicable to sums of independent (not necessarily identically distributed) r.v.'s, as well as random sums. The first case has been treated in [3], while the random sums are discussed in Section 3.

2. Three elementary proofs of CLT

Without loss of generality, take $E[X_j] = 0$ and $\text{Var}[X_j] = 1$, $j = 1, 2, \ldots$, and let F_n be the d.f. of the standardized sum $S_n = (X_1 + \cdots + X_n)/\sqrt{n}$. Suppose also that w_n is the w-function (covariance kernel) of S_n. Obviously, w_1 is the w-function of $S_1 = X_1$.

In Theorem 2 of [6], it was shown that for any r.v. Y with w-function w_Y,

$$E[w_Y^2(Y)] \geq 1, \tag{2.1}$$

and equality characterizes the normal distribution. The stability of the preceding characterization was also proved in Theorem 3, namely,

$$E[w^2_{Y_n}(Y_n)] \to 1 \text{ implies } \frac{Y_n - E[Y_n]}{\sqrt{\text{Var}[Y_n]}} \xrightarrow{\text{d}} N(0,1) \text{ as } n \to \infty.$$

It should be noted that the condition $E[w^2_{Y_n}(Y_n)] \to 1$ is equivalent to $Var[w_{Y_n}(Y_n)] \to 0$, since $E[w_Y(Y)] = 1$ for any r.v. Y. Thus, the proof of the CLT requires showing that

$$\text{Var}[w_n(S_n)] \to 0, \text{ as } n \to \infty. \tag{2.2}$$

A proof of (2.2) under the assumption (1.3) was given in [6], Theorem 4, by employing certain properties of the w-functions. A simpler proof is given below.

Theorem 2.1 (c.f. [6]). *Let $X, X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. and absolutely continuous r.v.'s with $E[X] = 0$, $E[X^2] = 1$ and $E[w^2(X)] < \infty$, where w is the covariance kernel of X. Then, (2.2) holds.*

For the proof we make use of the following Lemma (c.f. relation (3.3) of [5]).

Lemma 2.1. *Under the notations of Theorem 2.1, for $j \le n$ we have*

$$E[w_j(S_j)w_n(S_n)] = E[w_n^2(S_n)].$$

Proof. Set $A_j = E[w_j(S_j)w_n(S_n)]$. From the basic covariance identity (1.2) we get for arbitrary g

$$E[S_j g(S_j)] = E[w_j(S_j)g'(S_j)].$$

On the other hand, using the same identity,

$$\begin{aligned} E[S_j g(S_j)] &= E\left[\left(\sqrt{\frac{j-1}{j}}S_{j-1} + \sqrt{\frac{1}{j}}X_j\right) g(S_j)\right] \\ &= \sqrt{\frac{j-1}{j}}E\left\{E\left[S_{j-1}\, g(S_j)\,|X_j\right]\right\} + \sqrt{\frac{1}{j}}E\left\{E\left[X_j\, g(S_j)\,|S_{j-1}\right]\right\} \\ &= \frac{j-1}{j}E\left[w_{j-1}(S_{j-1})g'(S_j)\right] + \frac{1}{j}E\left[w(X_j)g'(S_j)\right]. \end{aligned}$$

Hence,

$$E[w_j(S_j)g'(S_j)] = \frac{j-1}{j}E\left[w_{j-1}(S_{j-1})g'(S_j)\right] + \frac{1}{j}E\left[w(X_j)g'(S_j)\right]. \tag{2.3}$$

Applying (2.3) to $g'(x) = w_n\left(\sqrt{j/n}\,x + \sqrt{(n-j)/n}\,S^*_{n-j}\right)$, where $S^*_{n-j} = (X_{j+1} + \cdots + X_n)/\sqrt{n-j}$ for $j < n$ and $S^*_0 \equiv 0$, we obtain

$$E\left[w_j(S_j)w_n(S_n)\,\middle|S^*_{n-j}\right]$$
$$= \frac{j-1}{j}E\left[w_{j-1}(S_{j-1})w_n(S_n)\,\middle|S^*_{n-j}\right] + \frac{1}{j}E\left[w(X_j)w_n(S_n)\,\middle|S^*_{n-j}\right].$$

Thus, taking expectations with respect to S^*_{n-j} and using the fact that

$$E[w(X_j)w_n(S_n)] = E[w_1(S_1)w_n(S_n)]$$

for each j, we conclude that $jA_j = (j-1)A_{j-1} + A_1$. This implies that $A_1 = \cdots = A_n$ and the proof is complete. ∎

Proof of Theorem 2.1. Set $\sigma_n = \mathrm{Var}\,[w_n(S_n)]$. It follows from Lemma 2.1 that

$$\sigma_n - \sigma_{n+1} = E[w_n^2(S_n)] - E[w_{n+1}^2(S_{n+1})] = E[w_n(S_n) - w_{n+1}(S_{n+1})]^2.$$

Therefore, σ_n decreases and hence it converges since $\sigma_1 < \infty$ by the assumptions. Thus,

$$\sigma_n - \sigma_{2n} = E[w_{2n}(S_{2n}) - w_n(S_n)]^2 \to 0.$$

Consequently,

$$\begin{aligned} E[w_{2n}(S_{2n}) - w_n(S_n)]^2 &\geq E[w_{2n}(S_{2n}) - E[w_{2n}(S_{2n})|S_n]]^2 \\ &= E\left\{\mathrm{Var}[w_{2n}(S_{2n})|S_n]\right\} \to 0. \end{aligned}$$

Furthermore, if $S^*_n = (X_{n+1} + \cdots + X_{2n})/\sqrt{n}$, we have

$$\begin{aligned} &E[w_{2n}(S_{2n}) - E[w_{2n}(S_{2n})|S_n]]^2 \\ &\quad = E\left\{E\left([w_{2n}(S_{2n}) - E[w_{2n}(S_{2n})|S_n]]^2\,|S^*_n\right)\right\} \\ &\quad \geq E\left\{E^2\left([w_{2n}(S_{2n}) - E[w_{2n}(S_{2n})|S_n]]\,|S^*_n\right)\right\} \\ &\quad = E\left[E[w_{2n}(S_{2n})|S^*_n] - 1\right]^2 \\ &\quad = E\left[E[w_{2n}(S_{2n})|S_n] - 1\right]^2 \\ &\quad = \mathrm{Var}\left\{E[w_{2n}(S_{2n})|S_n]\right\} \to 0. \end{aligned}$$

Hence, $\sigma_{2n} = \mathrm{Var}\left\{E[w_{2n}(S_{2n})|S_n]\right\} + E\left\{\mathrm{Var}[w_{2n}(S_{2n})|S_n]\right\} \to 0$ by the above arguments, which completes the proof. ∎

The second proof of CLT is based on the bound (see [7], Theorem 1.1)

$$\varrho(F,\Phi) \leq 2E|w_X(X) - 1| \leq 2\sqrt{\mathrm{Var}[w_X(X)]}, \tag{2.4}$$

where X is any standardized r.v. with d.f. F. In fact, the factor 2 in (2.4) can be replaced by 3/2, as shown in the following

Lemma 2.2. *If X has mean zero, variance one, density f and d.f. F,*

$$\varrho(F, \Phi) \leq (3/2)E|w_X(X) - 1| \leq (3/2)\sqrt{\mathrm{Var}[w_X(X)]}. \tag{2.5}$$

Proof. For an arbitrary Borel set A, consider the function $\psi_A(x)$, $x \in R$, defined by (see [7], [2], and references therein)

$$\psi_A(x) = \exp(x^2/2) \int_{-\infty}^{x} (I(t \in A) - \Phi(A)) \exp(-t^2/2)\, dt.$$

Applying the basic covariance identity (1.2) with $g = \psi_A$ and using relation (1.3) of [7], we get

$$\begin{aligned} F(A) - \Phi(A) &= E\{\psi_A'(X) - X\psi_A(X)\} \\ &= E\{\psi_A'(X)[1 - w_X(X)]\} \\ &= E\{I(X \in A)[1 - w_X(X)]\} + E\{X\psi_A(X)[1 - w_X(X)]\}. \end{aligned}$$

Observe that for $A_0 = \{x \in R : w_X(x) \leq 1\}$,

$$\begin{aligned} E\{I(X \in A)[1 - w_X(X)]\} &= \int_A [1 - w_X(x)]\, dF(x) \qquad (2.6) \\ &\leq \int_{A_0} [1 - w_X(x)]\, dF(x) = \frac{1}{2} E|w_X(X) - 1|. \end{aligned}$$

Moreover,

$$\begin{aligned} E\{X\psi_A(X)[1 - w_X(X)]\} &\leq E|X\psi_A(X)||1 - w_X(X)| \\ &\leq E|1 - w_X(X)|, \qquad (2.7) \end{aligned}$$

since for each x and A,

$$|x\psi_A(x)| \leq (|x|/\varphi(x)) \min\{\Phi(x), 1 - \Phi(x)\} \leq 1.$$

Now, the first inequality in (2.5) follows from (2.7) and (2.7), in view of the fact that

$$\sup_A |F(A) - \Phi(A)| = \sup_A [F(A) - \Phi(A)],$$

and the second one is obvious since $E[w_X(X)] = 1$. ∎

We now state the following CLT (for a proof see [7], Theorem 5.1).

Theorem 2.2 (CPU [7]). *Let $X, X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. and absolutely continuous r.v.'s with mean zero and variance one. Then, under the notations of Theorem 2.1,*

$$E|w_n(S_n) - 1| \to 0. \tag{2.8}$$

Note that (2.8), combined with (2.5), implies a strengthened CLT (L_1 convergence of the densities). However, $w(X)$ need not have finite variance, since (2.8) is an immediate consequence of the (weak) law of large numbers applied to $\overline{w_n} = (w(X_1) + \cdots + w(X_n))/n$ (see CPU [7] for more details).

The third elementary proof of the CLT is based on the following lemma, which may be of some interest in itself.

Lemma 2.3 (CPP [3], Lemma 2.1.) *Let X, Y be two independent and standardized absolutely continuous r.v.'s and consider the r.v. $S = aX+bY$, where a, b are real constants such that $a^2 + b^2 = 1$. Then*

$$\mathrm{Var}[w_S(S)] \le a^4 \mathrm{Var}[w_X(X)] + b^4 \mathrm{Var}[w_Y(Y)],$$

where w_X, w_Y, w_S, are the w-functions of X, Y, S, respectively.

As an immediate consequence of the above lemma we have the following

Theorem 2.3. *If $X, X_1, X_2, \ldots, X_n, \ldots$ are as in Theorem 2.1, then*

$$\varrho(F_n, \Phi) \le c/\sqrt{n}, \tag{2.9}$$

where the constant c can be taken as

$$c = (3/2)\sqrt{Var[w(X)]}. \tag{2.10}$$

Proof. If we apply Lemma 2.3 to $S_n = \sqrt{(n-1)/n}\, S_{n-1} + \sqrt{1/n}\, X_n$, we get

$$\sigma_n \le \left(\frac{n-1}{n}\right)^2 \sigma_{n-1} + \frac{1}{n^2}\sigma_1,$$

where $\sigma_n = \mathrm{Var}[w_n(S_n)]$. Thus, by induction on n, we conclude that $\sigma_n \le \sigma_1/n$, and (2.9) follows from (2.5). ∎

3. Applications to Random Sums

In this section, we give some applications of the above results to random sums (i.e., when the sample size n is no longer constant, but is a discrete nonnegative integer-valued random variable N, independent of the sequence

$\{X_i\}$). It is proved that under appropriate conditions (in fact, when N is likely to be large), the CLT continues to hold for the standardized sums of N i.i.d. standardized r.v.'s. It should be noted however, that here the term 'standardized' sums has a somewhat different meaning in the sense that

$$S_N = \frac{1}{\sqrt{N}}(X_1 + \cdots + X_N) \ , \tag{3.1}$$

(where $S_0 \equiv 0$ by definition) need not have variance one.

We first prove the following auxiliary result:

Lemma 3.1. *Suppose $X, X_1, \ldots, X_i, \ldots$ are i.i.d. r.v.'s as in Theorem 2.1. Then, for any nonnegative integer-valued r.v. N, independent of $\{X_i\}$,*

$$\varrho(F_N, \Phi) \leq P[N \leq m] + c/\sqrt{m+1}, \ m = 0, 1, \ldots \ , \tag{3.2}$$

where F_N is the d.f. of S_N and c can be taken as in Theorem 2.3.

Proof. For an arbitrary Borel set A we have, by (2.9),

$$\begin{aligned} |P[S_N \in A] - \Phi(A)| &\leq \sum_{n=0}^{\infty} P[N = n] \, |P[S_N \in A | N = n] - \Phi(A)| \\ &\leq \sum_{n=0}^{m} P[N = n] + \sum_{n=m+1}^{\infty} cP[N = n]/\sqrt{n}, \end{aligned}$$

and the proof is complete. ∎

Using the above Lemma, we can easily establish the following:

Theorem 3.1. *Let $\{X_i\}$ satisfy the assumptions of Lemma 3.1 and suppose that the nonnegative integer-valued r.v.'s N_n in probability tend to infinity, in the sense that for any $m > 0$, $P[N_n > m] \to 1$ as $n \to \infty$. Then,*

$$\varrho(F_n, \Phi) \to 0, \ \text{as } n \to \infty, \tag{3.3}$$

where F_n is the d.f. of S_{N_n}.

Proof. It follows from Lemma 3.1 that for any $\epsilon > 0$ (arbitrarily small) and $m > 0$ (arbitrarily large),

$$\varrho(F_n, \Phi) \leq \epsilon + c/\sqrt{m+1} \ \text{ when } n > n_0(\epsilon, m)$$

and the assertion follows from the arbitrariness of m and ϵ. ∎

Corollary 3.1. *If $N_n/a_n \to \Theta$ in probability, where Θ is an arbitrary positive r.v. and $a_n \to \infty$,*

$$\varrho(F_n, \Phi) \to 0, \quad \text{as } n \to \infty. \tag{3.4}$$

Proof. It is easy to verify the conditions of Theorem 3.1. ∎

Remark 3.1. Similar conditions imposed on $\{N_n\}$ can also be found in [1]; Theorem 17.2, p. 147 and [8], p. 258, where the sums are scaled by $\sqrt{a_n} = \sqrt{n}$ instead of $\sqrt{N_n}$ and Θ equals one. Billingsley considers the general case of $\{N_n\}$ not necessarily independent of $\{X_i\}$. He notes that the assumption that N_n in probability tends to infinity *does not suffice alone* for the convergence in distribution of F_n to Φ. For a counterexample see [1], pp. 143–144. It is therefore a crucial assumption that the sequences $\{N_n\}$ and $\{X_i\}$ are independent.

Remark 3.2. The restriction here to r.v.'s having w-functions with finite variance makes it possible not only to relax the condition on $\{N_n\}$, but also to obtain estimates for the rate of convergence. In fact, the following theorem provides uniform estimates for the rate of convergence, under somewhat stronger conditions imposed on $\{N_n\}$.

Theorem 3.2. *Let the sequences $\{X_i\}$, $\{a_n\}$ and $\{N_n\}$ be as in Corollary 3.1, and suppose that* [1]

$$\sup_n \left\{ \sqrt{a_n}\, E\left|\frac{N_n}{a_n} - \Theta\right| \right\} = C < \infty,$$

where Θ is a positive r.v. such that $P[\Theta \geq \delta] = 1$ for some $\delta > 0$. Then,

$$\varrho(F_n, \Phi) \leq O(a_n^{-1/2}), \quad \text{as } n \to \infty. \tag{3.5}$$

Proof. By using Markov's inequality, it is easily verified that

$$P[2N_n \leq \delta a_n] \leq P\left[\left|\frac{N_n}{a_n} - \Theta\right| \geq \frac{\delta}{2}\right] \leq \frac{2C}{\delta\sqrt{a_n}}.$$

Applying Lemma 3.1 with $m = [\delta a_n/2]$, we get

$$\varrho(F_n, \Phi) \leq \frac{2C/\delta + c\sqrt{2}/\sqrt{\delta}}{\sqrt{a_n}}$$

and the proof is complete. ∎

---

[1] Thanks are due to the referee for pointing out the insufficiency of an earlier condition for concluding (3.5); also that the proof of the CLT via w-functions, in the case of a noninterval support, is not straightforward, and hence omitted.

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379–386.

[3] Cacoullos, T., Papadatos, N., and Papathanasiou, V. (1996). Variance inequalities for covariance kernels and applications to central limit theorems. *Theory Probab. Appl.* **71**, 195–201.

[4] Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7**, 351–356.

[5] Cacoullos, T. and Papathanasiou, V. (1992). Lower variance bounds and a new proof of the central limit theorem. *J. Multivariate Anal.* **43**, 173–184.

[6] Cacoullos, T., Papathanasiou, V., and Utev, S. (1992). Another characterization of the normal law and a proof of the central limit theorem connected with it. *Theory Probab. Appl.* **37**, 648–657 (in Russian).

[7] Cacoullos, T., Papathanasiou, V., and Utev, S. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

[8] Feller, W. (1966). *An introduction to Probability Theory and its Applications.* Vol. 2, Wiley, New York.

[9] Papathanasiou, V. (1996). Multivariate variational inequalities and the central limit theorem. *J. Multivariate Anal.* **58**, 189–196.

T. Cacoullos and V. Papathanasiou
University of Athens
Department of Mathematics
Panepistemiopolis, 157 84 Athens
Greece

N. Papadatos
University of Cyprus
Department of Mathematics and Statistics
P.O. Box 537, Nicosia 1678
Cyprus