Journal of Logic & Analysis 7:3 (2015) 1–20
ISSN 1759-9008

Near Equivalence on Metric Spaces and a Nonstandard Central Limit Theorem

CHARLES J. GEYER
BERNARDO B. DE ANDRADE

Abstract: This article proves a nonstandard Central Limit Theorem (CLT) in the sense of Nelson's Radically Elementary Probability Theory [11]. The CLT proved here is obtained by establishing the near equivalence of standardized averages obtained from L_2 IID random variables to the standardized average resulting from a binomial CLT. A nonstandard model for near equivalence on metric spaces replaces conventional results of weak convergence. Statements and proofs remain radically elementary without applying the full Internal Set Theory. A nonstandard notion of normality is discussed.

2010 Mathematics Subject Classification 60F05 (primary); 60F25, 03H05 (secondary)

Keywords: radically elementary probability, weak convergence, Prohorov distance, Levy distance

1 Introduction

Nelson's [11] book "Radically Elementary Probability Theory" (REPT) provides a nonstandard probability model based on an axiomatic subsystem of Internal Set Theory (Nelson [10]) also known as *minimal IST* (Herzberg [8]). Recent advances in IST-based probability theory include diffusions and interacting particle systems (Weisshaupt [15, 16]), Markov chains (Andrade [2]), discrete functions on infinitesimal grids with an application to probability (van den Berg [4]), stochastic calculus including Itô's stochastic integration and Lévy processes [8] and several applications covered in the book edited by Diener and Diener [6]. This paper is restricted to REPT and hence more closely related to recent work by Herzberg [8] and Weisshaupt [16]. The principles of idealization, transfer, standardization and sequence (Nelson [10], Robert [12]) are not invoked here and our results and definitions remain strictly within the scope of REPT.

REPT has a functional central limit theorem (fCLT) [11, Theorem 18.1] where the objects of interest are martingales and the Wiener walk (nonstandard Brownian motion).

Published: August 2015 doi: 10.4115/jla.2015.7.3

This fCLT has been specialized by Andrade [3] to a *classical* CLT for independent and identically distributed (IID) nonstandard L_2 random variables (rvs). However REPT lacks an explicit definition of nonstandard normality and an explicit classical CLT. The main contribution of this paper is to prove a classical CLT (Theorem 3.2) directly without reference to the fCLT. This is done in Section 3. We had to develop a radically elementary model for studying weak convergence in metric spaces based on the metrics of Prohorov and Lévy (Section 2). A different notion of weak convergence (in topological, not simply metric, spaces) based on Robinson's nonstandard analysis [13] and Loeb measure spaces [9] is given by Anderson and Rachid [1]. Appendices A and B contain auxiliary material used in the paper.

To those unfamiliar with REPT we must mention that in radically elementary analysis no new objects, such as hyperreals, are added. Infinitesimals are not new objects in $\mathbb{R}$ but are simply not used by the conventional theory. A property, *limitedness*, and four axioms that govern the use of this property are defined. The axioms in minimal IST apply directly only to the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$. An important goal of REPT is to be within reach by anyone familiar with basic (discrete) probability and the minimal notions about the rigorous use of infinitesimals (Nelson's "tiny bit" analysis [11, Preface]) without any heavy apparatus from nonstandard analysis, whether from IST [10] or from Robinson's nonstandard model [13].

2 Near Equivalence on Metric Spaces

In order to prove our nonstandard version of the classical CLT, Theorem 3.2, we will need tools that do not yet exist in REPT. This section develops such tools.

All basic definitions of nonstandard analysis, such as that of an infinitesimal ($a \simeq 0$), an unlimited number ($\nu \simeq \infty$), an appreciable number ($a \gg 0$) and a near line are taken from Nelson [11]. We also make use of the notation $x \lesssim y$ meaning that $x \leq y$ or $x \simeq y$.

Definition 2.1 Let (S, d) be a metric space with a set S and a metric for d for S. Points $x, y \in S$ are *nearly equal*, $x \simeq y$, when $d(x, y) \simeq 0$.

The usual definition of $x \simeq y$ in $\mathbb{R}$ is a special case when $d(x, y) = |x - y|$.

Definition 2.2 Let (S, d) and (S', d') be metric spaces. A function $h : S \to S'$ is *nearly continuous at a point* $x \in S$ if

(1) $$y \in S \text{ and } x \simeq y \Longrightarrow h(x) \simeq h(y),$$

and *nearly continuous on a set* $T \subset S$ if (1) holds at all $x \in T$.

Definition 2.3 Random elements X and Y of a metric space (S, d) defined on possibly different probability spaces are *nearly equivalent*, $X \overset{\mathrm{w}}{\simeq} Y$, if for every limited nearly continuous function $g : S \to \mathbb{R}$,

$$\mathrm{E}\{g(X)\} \simeq \mathrm{E}\{g(Y)\}.$$

Remark 2.4 Let d_1 and d_2 be two different metrics for S. We say that d_1 and d_2 are *equivalent* if they agree as to the meaning of $x \simeq y$, that is, if

$$d_1(x, y) \simeq 0 \iff d_2(x, y) \simeq 0.$$

In this case, a function $h : S \to S'$ where (S', d') is another metric space is nearly continuous when d_1 is the metric for S if and only if it is continuous when d_2 is the metric. In this sense, near equivalence of random elements of metric spaces does not depend on the metric, only on the external equivalence relation $\simeq$ induced by it.

Notation A random element X of (S, d) induces a probability measure P, $P = \mathcal{L}(X)$ (P is the *law of* X), defined by

$$P(A) = \Pr(X \in A), \qquad A \subset S$$

and probability mass function p, $p = dP$, defined by

$$p(x) = \Pr(X = x), \qquad x \in S.$$

Remark 2.5 In REPT all probability measures have finite support and this is implicit throughout the paper, though we may, sometimes, redundantly say "P with finite support".

Remark 2.6 Since near equivalence is determined by expectations, and expectations are determined by measures, near equivalence really depends only on measures not on rvs (except through their measures). Thus we make the definition: measures P and Q on a metric space (having finite support) are *nearly equivalent*, written $P \overset{\mathrm{w}}{\simeq} Q$ if $Ph \simeq Qh$ for every limited nearly continuous function $h : S \to \mathbb{R}$, where

$$Ph = \sum_{x \in S} h(x) dP(x)$$

is a shorthand for the expectation of the rv $h(X)$ when $P = \mathcal{L}(X)$.

Notation If d is a metric on S we define for any nonempty $A \subset S$,

$$d(x, A) = \inf_{y \in A} d(x, y)$$

and for any $\epsilon > 0$, we define the ϵ*-dilation of* A,

$$A^\epsilon = \{\, x \in S : d(x, A) < \epsilon \,\},$$

with the ϵ-dilation of the empty set being empty.

The triangle inequality implies $(A^\epsilon)^\eta \subset A^{\epsilon+\eta}$.

Definition 2.7 Let S be a finite set with metric d and let $\mathcal{P}(S, d)$ denote the set of all probability measures on S. The *Prohorov metric* on $\mathcal{P}(S, d)$ is the function $\pi : \mathcal{P}(S, d) \times \mathcal{P}(S, d) \to \mathbb{R}$ defined so that $\pi(P, Q)$ is the infimum (attained since P and Q have finite support) over all $\epsilon > 0$ such that

$$P(A) \leq Q(A^\epsilon) + \epsilon \text{ and } Q(A) \leq P(A^\epsilon) + \epsilon, \qquad A \subset S. \tag{2}$$

Note that (2) holds for every $\epsilon > \pi(P, Q)$.

It can be shown that π is indeed a metric but this fact will not be needed here.

A useful fact about the Prohorov metric which will be required to prove Lemma 2.14 is the following.

Lemma 2.8 *The Prohorov distance between P and Q is the infimum over all ϵ such that*

$$P(A) \leq Q(A^\epsilon) + \epsilon, \qquad A \subset S. \tag{3}$$

Proof The same as in Billingsley [5, p. 72, (ii)] because the argument uses no measure theory and it translates directly into REPT. □

Definition 2.9 Let (S, d) and (S', d') be metric spaces. A function $h : S \to S'$ is *nearly Lipschitz continuous* if there exists $L \ll \infty$ such that

$$d'\big(h(x), h(y)\big) \leq L \cdot d(x, y), \qquad x, y \in S.$$

Note that near Lipschitz continuity implies near continuity.

Lemma 2.10 *Let (S, d) be a metric space. For any nonempty $A \subset S$ and any $\epsilon \gg 0$, define $h : S \to \mathbb{R}$ by*

$$h(x) = \max\big(0, 1 - d(x, A)/\epsilon\big).$$

Then h is limited and nearly Lipschitz continuous.

Proof (Sketch) By considering three cases: $d(x,A)=0$ so $h(x)=1$, $d(x,A)\geq\epsilon$ so $h(x)=0$ and $0<d(x,A)<\epsilon$ so $0<h(x)<1$ we can establish

$$|h(x)-h(y)|\leq\frac{|d(x,A)-d(y,A)|}{\epsilon}. \tag{4}$$

It is easy to show that $d(x,y)\geq|d(y,A)-d(x,A)|$ from which we see that (4) implies

$$|h(x)-h(y)|\leq\frac{|d(x,A)-d(y,A)|}{\epsilon}\leq\frac{1}{\epsilon}\cdot d(x,y).$$

With $1/\epsilon$ being limited, the result is established. □

Theorem 2.11 *Assume P and Q are measures on (S,d) and that $Ph\simeq Qh$ holds for every limited nearly Lipschitz continuous function $h:S\to\mathbb{R}$. Then $\pi(P,Q)\simeq 0$.*

Proof For any nonempty $A\subset S$ and any $\epsilon\gg 0$, by the assumptions on h,

$$P(A)\leq Ph\leq P(A^{\epsilon})$$

and similarly with P replaced by Q. Hence $P(A)\leq Ph\simeq Qh\leq Q(A^{\epsilon})$ holds for all A. Thus (3) holds for every $\epsilon\gg 0$ and hence $\pi(P,Q)$ is infinitesimal. □

Since $P\overset{\mathrm{w}}{\simeq}Q$ is a stronger requirement than Lipschitz continuity we have the obvious result below.

Corollary 2.12 $P\overset{\mathrm{w}}{\simeq}Q\Longrightarrow\pi(P,Q)\simeq 0$.

The reverse is true but the proof is long and the result will not be needed here.

Definition 2.13 Let A be any property that may or may not hold at points of S and let P be a measure on (S,d) having finite support. We say A holds P*-almost everywhere* if for every $\epsilon\gg 0$ there exists a set N such that $P(N)\leq\epsilon$ and $A(x)$ holds except for x in N. The next result is similar to Nelson [11, Theorem 17.3] concerning stochastic processes.

Notation Let (S,d) be a metric space and define for $A\subset S$ the ϵ*-erosion of* A,

$$A_{\epsilon}=S\setminus(S\setminus A)^{\epsilon}.$$

Lemma 2.14 *Let P and Q be elements of $\mathcal{P}(S,d)$ such that $P\overset{\mathrm{w}}{\simeq}Q$, and let A be any property (internal or external) such that $x,y\in S$ and $x\simeq y$ implies $A(x)\Longleftrightarrow A(y)$. Then A holds P-almost everywhere if and only if it holds Q-almost everywhere.*

Proof By Corollary 2.12 we have $\pi(P, Q) \simeq 0$. By Lemma 2.8 the infinitesimality of $\pi(P, Q)$ is equivalent to the existence of an $\epsilon \simeq 0$ such that (3) holds, which implies that for some $\epsilon \simeq 0$ and for all $A \subset S$ we have $P(A) \lesssim Q(A^\epsilon)$. Let $B = S \setminus A$. By the complement rule we immediately have $Q(A^\epsilon) = 1 - Q(B_\epsilon)$ and thus

$$P(A) \lesssim Q(A^\epsilon) \iff P(B) \gtrsim Q(B_\epsilon)$$

hence if the left hand side holds for all A, then the right hand side holds for all B and vice versa. Thus from $P \overset{\text{w}}{\simeq} Q$ we have concluded that for some $\epsilon \simeq 0$ and for all $A \subset S$ we have $P(A) \gtrsim Q(A_\epsilon)$.

Suppose A holds P-almost everywhere. Fix $\epsilon \gg 0$ and choose $N \subset S$ such that $P(N) \leq \epsilon/2$ and $A(x)$ holds for all $x \in S \setminus N$. By the previous paragraph there is a $\delta \simeq 0$ such that $P(N) \gtrsim Q(N_\delta)$. So $Q(N_\delta) \leq \epsilon$. By definition $S \setminus N_\delta = (S \setminus N)^\delta$. Hence $y \in S \setminus N_\delta$ if and only if there exists an x in $S \setminus N$ such that $d(x, y) < \delta$. This implies $x \simeq y$, and hence $A(y)$ holds. Hence A holds on $S \setminus N_\delta$. Since $\epsilon \gg 0$ was arbitrary, A holds Q-almost everywhere. □

Remark 2.15 Let P be a measure on (S, d) having finite support. A function $h : S \to S'$ is *nearly continuous P-almost everywhere* if the property $A(x)$ in the definition of "almost everywhere" is "h is nearly continuous at x." Note that by the lemma above, when $P \overset{\text{w}}{\simeq} Q$ we have h nearly continuous P-almost everywhere if and only if h is nearly continuous Q-almost everywhere.

Definition 2.16 The *Lévy metric* on the set of all cumulative distribution functions (cdfs) (with finite support) on $\mathbb{R}$ is the function λ defined so that $\lambda(F, G)$ is the infimum of all $\epsilon > 0$ such that

$$F(x) \leq G(x + \epsilon) + \epsilon \text{ and } G(x) \leq F(x + \epsilon) + \epsilon, \qquad x \in \mathbb{R}. \tag{5}$$

It is easy to see that λ actually is a metric.

Definition 2.17 We say cdfs F and G are *nearly equal*, written $F \simeq G$, if $\lambda(F, G) \simeq 0$.

Remark 2.18 Near equality of cdfs does not necessarily imply the corresponding rvs are nearly equivalent. See Theorem 2.21 and following remark.

Lemma 2.19 *If λ is the Lévy metric and π the Prohorov metric, F and G are cdfs and P and Q are the corresponding measures, then*

$$\pi(P, Q) \geq \lambda(F, G).$$

Proof Fix $\epsilon > \pi(P, Q)$. Then

$$\begin{aligned} F(x) &= P\{(-\infty, x]\} \\ &\le Q\{(-\infty, x]^\epsilon\} + \epsilon \\ &= Q\{(-\infty, x+\epsilon)\} + \epsilon \\ &= \max_{y<x} G(y+\epsilon) + \epsilon \end{aligned}$$

holds for all x. In particular we have

$$F(x) \le G(x+\epsilon) + \epsilon \tag{6}$$

whenever $x + \epsilon$ is not a jump of G. However, even if $x + \epsilon$ is a jump of G, there exists a $\delta > 0$ sufficiently small so that G has no jump in $(x+\epsilon, x+\epsilon+\delta]$, and, applying (6) with ϵ replaced by $\epsilon + \delta$, we have

$$F(x) \le G(x+\epsilon+\delta) + \epsilon + \delta = G(x+\epsilon) + \epsilon + \delta,$$

which, since $\delta > 0$ was arbitrary, implies (6) even when $x + \epsilon$ is a jump of G. The same argument with F and G swapped finishes the proof. □

Corollary 2.20 *If X and Y are rvs having cdfs F and G, then $X \overset{\mathrm{w}}{\simeq} Y$ implies $F \simeq G$.*

Theorem 2.21 *If X and Y are rvs having cdfs F and G, either X or Y is limited almost surely, and $F \simeq G$, then $X \overset{\mathrm{w}}{\simeq} Y$.*

Proof Fix arbitrary appreciable ϵ_1 and ϵ_2. By Nelson [11, Theorem 7.3] if one of X and Y is limited almost surely and $X \overset{\mathrm{w}}{\simeq} Y$, then the other is also limited almost surely. Hence, by Theorem B.1, there exist limited a and b such that

$$F(a) \le \epsilon_1 \tag{7a}$$
$$F(b) \ge 1 - \epsilon_1 \tag{7b}$$

and similarly with F replaced by G. Let h be a nearly continuous function with limited bound M. Then

$$|\mathrm{E}\{h(X) I_{(-\infty,a]}(X)\}| \le M F(a) \le M\epsilon_1 \tag{8a}$$

and

$$|\mathrm{E}\{h(X) I_{(b,\infty)}(X)\}| \le M\big[1 - F(b)\big] \le M\epsilon_1 \tag{8b}$$

and similarly with X replaced by Y and F replaced by G.

By near continuity of h there exists $\delta \gg 0$ such that $|h(x) - h(y)| \leq \epsilon_2$ whenever $|x - y| \leq \delta$. There exists a limited natural number n such that $(b - a)/n \leq \delta/2$. Define $c_k = a + (k/n)(b - a)$ for integer k, noting that $c_0 = a$ and $c_n = b$. Then

$$|h(x) - h(c_k)| \leq \epsilon_2, \qquad c_{k-2} \leq x \leq c_{k+2}, \tag{9}$$

and this together with (8a) and (8b) implies

$$\left| \mathrm{E}\{h(X)\} - \sum_{k=1}^{n+1} h(c_k)\bigl[F(c_k) - F(c_{k-1})\bigr] \right| \leq \epsilon_2 + 2M\epsilon_1. \tag{10}$$

The assumption $\lambda(F, G) \simeq 0$ implies that there exist b_k and d_k such that $b_k \leq c_k \leq d_k$ and $b_k \simeq c_k \simeq d_k$ and

$$G(b_k) \lesssim F(c_k) \lesssim G(d_k), \qquad \text{for all } k.$$

The same reasoning that lead to (10) implies that (10) holds with X replaced by Y and F replaced by G. But

$$\begin{aligned} \sum_{k=1}^{n+1} g(c_k)\bigl[F(c_k) - F(c_{k-1})\bigr] - \sum_{k=1}^{n+1} g(c_k)\bigl[G(b_k) - G(b_{k-1})\bigr] \\ = \sum_{k=1}^{n} \bigl[g(c_k) - g(c_{k+1})\bigr] \cdot \bigl[F(c_k) - G(b_k)\bigr] \\ + g(c_{n+1})\bigl[F(c_{n+1}) - G(b_{n+1})\bigr] - g(c_1)\bigl[F(c_0) - G(b_0)\bigr] \end{aligned} \tag{11}$$

and $0 \lesssim F(c_k) - G(b_k) \lesssim G(d_k) - G(b_k)$, and the latter sum to less or equal to one. Hence, a limited sum of infinitesimals being infinitesimal, the sum in (11) is less than or nearly equal to ϵ_2 and greater than or nearly equal to zero. The other terms are less than or equal to $4M\epsilon_1$ in absolute value. That is,

$$\left| \sum_{k=1}^{n+1} g(c_k)\bigl[F(c_k) - F(c_{k-1})\bigr] - \sum_{k=1}^{n+1} g(c_k)\bigl[G(b_k) - G(b_{k-1})\bigr] \right| \leq \epsilon_2 + 4M\epsilon_1.$$

Hence by the triangle inequality

$$|\mathrm{E}\{g(X)\} - \mathrm{E}\{g(Y)\}| \leq 3\epsilon_3 + 8M\epsilon_1.$$

Since ϵ_1 and ϵ_2 were arbitrary appreciable numbers and M is limited, we actually have $\mathrm{E}\{g(X)\} \simeq \mathrm{E}\{g(Y)\}$. □

Remark 2.22 The limited almost surely condition cannot be suppressed. Let X have the uniform distribution on the even integers between 1 and 2ν and Y have the uniform distribution on the odd integers between 1 and 2ν. Then if F and G are the corresponding cdfs, then we have

$$0 \leq G(x) - F(x) \leq \frac{1}{\nu}$$

for all x. So $\lambda(F, G)$ is infinitesimal whenever ν is unlimited. But, if h is defined by $h(x) = \sin^2(\pi x)$, then h is a limited continuous function, and $h(X) = 0$ and $h(Y) = 1$ (for all ω).

3 The Central Limit Theorem

The following radically elementary De Moivre-Laplace CLT (Theorem 3.1) plays a central role in the proof of the more general classical CLT (Theorem 3.2). It can be proved by careful translation of the (lengthy) treatment in Feller [7, Sec.7.2] starting with (a nonstandard version of) Stirling's approximation and using some integration concepts described in Appendix A.3. A similar proof using Robinson's model is found in Section 0.3 of Stroyan and Bayod [14]. The proof of the next result is thus only sketched.

Theorem 3.1 (Radically Elementary De Moivre-Laplace) *Suppose X has the Binomial(ν, p) distribution with $\nu \simeq \infty$, $0 \ll p \ll 1$ and*

$$Z = \frac{X - \nu p}{\sqrt{\nu p q}}.$$

Then there exists a near line T such that

$$\Pr(Z \leq z) \simeq \frac{1}{\sqrt{2\pi}} \sum_{\substack{t \in T \\ t \leq z}} e^{-t^2/2}\, dt.$$

Furthermore,

$$F(z) \equiv \frac{1}{c} \sum_{\substack{t \in T \\ t \leq z}} e^{-t^2/2}\, dt, \quad c \equiv \sum_{t \in T} e^{-t^2/2}\, dt,$$

satisfies, for all $z \in \mathbb{R}$,

$$F(z) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt.$$

Proof (Sketch) The essence of the proof is a careful adaptation of Feller [7, Sec.7.2] combined with radically elementary integration (Appendix A.3) specially Corollary A.6.

We will only indicate some landmarks in the proof to which we will refer later. It can be quickly argued that a nonstandard version of Stirling's approximation is

$$\nu! \sim (2\pi)^{1/2}\nu^{\nu+1/2}e^{-\nu}, \tag{12}$$

where the notation $x \sim y$ is explained in the beginning of Appendix A.3. Trivial but lengthy calculations lead us to

$$\begin{aligned} f(k) &\equiv \binom{\nu}{k} p^k q^{\nu-k} \\ &\sim \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \equiv \varphi(z) dz \end{aligned}$$

where z lies in the near line with regular spacing $dz = (1/\sqrt{\nu pq}) \simeq 0$,

$$T = \{\, (k - \nu p)dz : k = 0, \dots, \nu \,\}.$$

Now for any limited numbers a and b with $a < b$ we have by Lemmas A.3 and A.2,

$$\Pr(a < Z < b) \simeq \sum_{\substack{z \in T \\ a<z<b}} \varphi(z)\, dz. \tag{13}$$

Since $\mathrm{E}(Z) = 0$ and $\mathrm{var}(Z) = 1$ it follows from Chebyshev's inequality that

$$\Pr(|Z| \geq a) \leq \frac{1}{a^2}.$$

Hence, by Theorem A.4, Z is limited almost surely, and by Theorem B.1 for any $\epsilon \gg 0$ there exists $a \ll \infty$ such that $\Pr(|Z| \geq a) \leq \epsilon$. Hence

$$\left| \Pr(Z < b) - \sum_{\substack{z \in T \\ z<b}} \varphi(z)\, dz \right| \lesssim \epsilon$$

and, since $\epsilon \gg 0$ was arbitrary the claim about $\Pr(Z \leq z)$ follows. The second part of the theorem, $F(z) \simeq \int_{-\infty}^{z} \varphi(t)dt$, follows from Corollary A.6.

□

The classical CLT can be seen as a mild extension of the above binomial CLT.

Theorem 3.2 (Classical CLT and definition of normality) *Suppose* $X_1, X_2, \dots, X_\nu$, $\nu \simeq \infty$, *are independent and identically distributed* L_2 *rvs with mean* μ *and variance* $\sigma^2 = \mathrm{E}(X_i - \mu)^2 \gg 0$. *Let*

$$\bar{X}_\nu = \frac{1}{\nu} \sum_{i=1}^{\nu} X_i.$$

Then the rv

$$Z = \frac{\bar{X}_\nu - \mu}{\sigma/\sqrt{\nu}} \tag{14}$$

is L_2 and limited almost surely. Any rv nearly equivalent to Z is defined to be normal. Furthermore, if X is an L_2 normal rv and $Y = \alpha + \beta X$, where α and β are limited and $\beta \geq 0$, then Y is L_2, $\mathrm{E}(Y) \simeq \alpha$, and $\mathrm{var}(Y) \simeq \beta^2$.

Finally, the distribution function of Z is nearly equal to

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt. \tag{15}$$

Proof Following similar argument in the proof of Nelson [11, Theorem 18.1], the rv Z is L_2 if and only if for all $\epsilon \gg 0$ there is a $c \ll \infty$ such that $\mathrm{E}[f_c(Z)] \geq \mathrm{E}(Z) - \epsilon$ where

$$f_c(u) = \begin{cases} u^2, |u| \leq c \\ c^2, |u| > c \end{cases}$$

It is clear that $\mathrm{E}(Z^2) \simeq 1$ and therefore for all $\epsilon \gg 0$ there is a $c \ll \infty$ such that $\mathrm{E}[f_c(Z)] \gtrsim 1 - \epsilon$, which establishes that Z is L_2.

Consider the special case where each X_i is Bernoulli(p), $0 \ll p \ll 1$. Then the associated normal rv Z according to (14) is limited almost surely (this is shown near the end of the proof of Theorem 3.1). Hence by [11, Theorem 7.3] every normal rv is limited almost surely.

Let X be an L_2 normal rv and $Y = \alpha + \beta X$, with $\alpha \ll \infty$ and $0 \leq \beta \ll \infty$. As just shown every normal rv that arises in the CLT is L_2. Moreover, such an rv, equation (14), has mean zero and standard deviation one. Since the map $x \mapsto x^{(a)}$ defined by (23) is limited and continuous for limited a, it follows by Lemma B.3 that nearly equivalent L_2 rvs have nearly equal mean and variance. Hence $\mathrm{E}(X) \simeq 0$ and $\mathrm{var}(X) \simeq 1$. By (22a) and (22b), Y is L_2 and clearly $\mathrm{E}(Y) \simeq \alpha$ and $\mathrm{var}(Y) \simeq \beta^2$.

Finally, Theorem 3.1 asserts that one particular normal random variable has cdf nearly equal to Φ. Hence by Corollary 2.20, Theorem 2.21, and the fact that Z is limited almost surely, an rv is normal if and only if its distribution function is nearly equal to Φ, equation (15). □

Remark 3.3 The assertion of the CLT is that every rv of the form (14) defined with an IID L_2 sequence is nearly equivalent to an rv of the same form defined starting with some other IID L_2 sequence and we have labeled any such rv as "normal". In REPT the term “normal” cannot name a distribution. It is an *external property* that distributions may or may not have. As with any external property, it is illegal set formation to try to form the set of all normal distributions.

Definition 3.4 If X is a normal rv, μ is a limited real number, and σ is a positive appreciable real number, then we say $Y = \mu + \sigma X$ is *general normal* and we also apply this terminology to the distribution of Y.

Remark 3.5 Like normality, general normality is an external property. It is tempting to call μ the mean and σ the standard deviation but in REPT this does not make sense. A (general) normal distribution, as we have defined the concept, need not have moments anywhere close to those of a conventional normal distribution. What the theorem shows is that if X is L_2 then $\mathrm{E}(Y) \simeq \mu$ and $\mathrm{var}(Y) \simeq \sigma^2$. If a rv X is normal, in our sense, then for g limited nearly continuous we have $\mathrm{E}[g(X)] \simeq \mathrm{E}[g(Z)]$ and $F_X \simeq \Phi$ which does not imply (cf. Weisshaupt [16, Definition 4.18])

$$\mathrm{E}[g(X)] \simeq \int g(x)\Phi(dx).$$

Remark 3.6 It can be shown that if X is a normal rv and $Y = \mu + \sigma X$, $\mu \ll \infty$ and $0 \leq \sigma \ll \infty$, then the median$(Y) \simeq \mu$ and the $\Phi(1)$-quantile of Y is nearly equal to $\mu + \sigma$, where Φ is defined by (15).

4 Final Remarks

In conventional probability the binomial CLT, the classical CLT and any fCLT are three well separated results. The binomial case can be proved with Stirling’s formula and combinatorics and it is not used in the proof of the classical CLT which is typically demonstrated by means of characteristic functions. fCLTs require the apparatus of advanced stochastic analysis.

Within REPT we have used the binomial CLT as an important stepping stone for the classical CLT. We needed, however, to develop the apparatus of Section 2 which boils down to Lemma 2.14 and the two subsequent results. Lemma 2.14 could have been replaced by a more general result regarding trajectories of stochastic processes [11, Theorem 17.3] and the associated notion of near equivalent processes but we chose to develop a separate model for independent random variables. REPT’s fCLT [11, Theorem 18.1] is just a page away from [11, Theorem 17.3]. Andrade [3] argues that deriving the classical CLT as a special case of the much more general fCLT is not an absurd circumvention as it would be in conventional probability. The results proved in this paper make clear the close connection between the binomial and the classical CLT (within REPT) and also the notion of normality which is only implicit in REPT's fCLT [11, Theorem 18.1]. Therefore the present paper can be seen as a complement to Andrade [3] interconnecting the binomial, the classical and the martingale CLTs within REPT.

Appendices

The following appendices summarize some definitions and results from Nelson [11, Chapters 7-8] and bring some new results in infinitesimal summation or integration (see also Robert [12, Chapter 6]).

A Radically Elementary Nonstandard Analysis

A.1 Convergence of a Real Sequence

A sequence x_1, x_2, …, x_ν of nonrandom real numbers *nearly converges* to a real number x if

(16a) $$x_n \simeq x, \qquad n \simeq \infty.$$

Equivalently, a sequence nearly converges if

(16b) $$\forall \varepsilon \gg 0\, \exists N \ll \infty\, \forall n \geq N\, (|x_n - x| \leq \varepsilon).$$

For example, if $x_1, \ldots, x_\nu$, $\nu \simeq \infty$, nearly converges to zero and $M = \max_{1 \leq n \leq \nu} |x_n| \ll \infty$, then it can be shown that the partial sums

$$S_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

nearly converge to zero. However, allowing for $M \simeq \infty$ creates a number of cases. Two trivial situations with $M \simeq \infty$ are: (i) $x_1 = M$, $x_2 = -M$ and $x_i = 0, i > 2$ then S_n nearly converges to zero but (ii) if $x_1 = M$, $x_i = 0, i > 1$, S_n would not converge since $\nexists \varepsilon \gg 0$ such that $|M/N| \leq \varepsilon$ for some $N \ll \infty$.

A.2 Continuity

Let $S \subset \mathbb{R}$. The function $f : S \to \mathbb{R}$ is *nearly continuous at the point* $x \in S$ if

$$\forall y \in \mathbb{R} \left(x \simeq y \Longrightarrow f(x) \simeq f(y) \right) \tag{17}$$

and *nearly continuous on* S if (17) holds for all $x \in S$. Near continuity at a point x is analogous to conventional continuity at x, but near continuity on a set S is analogous to conventional *uniform continuity on* S. For instance, the function $x \mapsto 1/x$ is continuous (but not uniformly continuous) on $S = (0, \infty)$, but it is not nearly continuous on S. Another example is the function $x \mapsto e^x$ which is *not* nearly continuous on $\mathbb{R}$. The next result from Robert [12] establishes near continuity for the common functions used in calculus.

Lemma A.1 *([12, Chapter 4]) Suppose f is a differentiable function $S \to \mathbb{R}$, where S is an open interval, and f' is limited on S. Then f is nearly continuous on S.*

A.3 Integration

Within REPT integrals must be replaced by unlimited finite sums. Let T be a finite subset of $\mathbb{R}$. For $t \in T' = T \setminus \{\max(T)\}$ we write dt to mean the difference between t and its successor in T, that is,

$$dt = \min\{s \in T : s > t\} - t. \tag{18}$$

For $t = \max(T)$, we set $dt = 0$. We refer to dt collectively as the *spacings of* T. If all the spacings are infinitesimal and each $t \in T$ is limited, then we call T a *near interval*. We say that T is a *near line* if all the spacings dt are infinitesimal with $\min(T) \simeq -\infty$ and $\max(T) \simeq \infty$.

If x and y are real numbers such that $y \neq 0$ and $x/y \simeq 1$, we say that x is *asymptotic to* y, denoted by $x \sim y$. The external relation $\sim$ is an equivalence on $\mathbb{R} \setminus \{0\}$. We use this notion in the following study of sums with an unlimited number of terms which are the analogues of conventional infinite series and integrals.

Lemma A.2 *([11, Chapter 5]) Suppose x or y is appreciable. Then $x \sim y$ if and only if $x \simeq y$.*

Lemma A.3 *([11, Chapter 5]) If $x_i > 0$, $y_i > 0$, and $x_i \sim y_i$ for each i, then*

$$\sum_{i=1}^{\nu} x_i \sim \sum_{i=1}^{\nu} y_i. \tag{19}$$

Thus, if one side of (19) is appreciable, there is no difference between asymptotic equivalence and near equality ($\sim$ and $\simeq$ respectively). We now present the following "integration" results.

Theorem A.4 *Suppose T is a near interval and suppose f and g are limited-valued functions on T such that $f(t) \simeq g(t)$ for all $t \in T$. Then*

$$\sum_{t \in T} f(t)\,dt \simeq \sum_{t \in T} g(t)\,dt.$$

Proof Let L be the maximum of all $|f(t)|$ and $|g(t)|$ for $t \in T$. The maximum is achieved because T is finite and hence is limited by assumption. Fix $\varepsilon \gg 0$ and define

$$\begin{aligned} T_+ &= \{\, t \in T : f(t) \geq \varepsilon \,\} \\ T_0 &= \{\, t \in T : |f(t)| < \varepsilon \,\} \\ T_- &= \{\, t \in T : f(t) \leq -\varepsilon \,\} \end{aligned}$$

Then by Lemma A.2 we have

$$f(t) \sim g(t), \qquad t \in T_+ \cup T_-$$

and hence by Lemmas A.3 and A.2 again we have

$$\sum_{t \in T_+} f(t)\,dt \simeq \sum_{t \in T_+} g(t)\,dt$$

and the same with T_+ replaced by T_-. Also

$$\left| \sum_{t \in T_0} f(t) \right| \leq \sum_{t \in T_0} |f(t)| \leq \varepsilon(b-a)$$

where a and b are the endpoints of T, and the same with f replaced by g and ε replaced by 2ε, because $|f(x)| \leq \varepsilon$ implies $|g(x)| \leq 2\varepsilon$. Since $\varepsilon \gg 0$ was arbitrary and a and b are limited (by the near interval assumption), we have

$$\sum_{t \in T_0} f(t) \simeq 0$$

and the same with f replaced by g. Thus from the triangle inequality and the sum of infinitesimals being infinitesimal we obtain

$$\left| \sum_{t \in T} f(t) - \sum_{t \in T} f(t) \right| \lesssim 3\varepsilon(b-a)$$

Since $\varepsilon \ll 0$ was arbitrary and a and b are limited (by the near interval assumption), we obtain the desired result. □

Theorem A.5 *Suppose $T = \{a, \ldots, b\}$ is a near interval and suppose $f : [a, b] \to \mathbb{R}$ a Riemann integrable function with a limited bound and nearly continuous on T. Then*

$$\sum_{t \in T} f(t)\, dt \simeq \int_a^b f(t)\, dt.$$

Proof Fix $\varepsilon \gg 0$. By definition of Riemann integrability, there exists a subset S of $\mathbb{R}$ with endpoints a and b such that

$$\left| \sum_{s \in S} f(s)\, ds - \int_a^b f(t)\, dt \right| \leq \varepsilon$$

(where ds is the spacing of S at s), and the same holds when S is replaced by a finer partition, in particular,

$$\left| \sum_{u \in U} f(u)\, du - \int_a^b f(t)\, dt \right| \leq \varepsilon$$

where $U = S \cup T$ (and where du is the spacing of U at u).

Define $h : U \to \mathbb{R}$ by

$$h(u) = f(t), \qquad t \in T \text{ and } t \leq u < t + dt.$$

By near continuity of f we have $f(u) \simeq h(u)$ for $u \in U$, and hence by Theorem A.4

$$\sum_{u \in U} f(u)\, du \simeq \sum_{u \in U} h(u)\, du = \sum_{t \in T} f(t)\, dt.$$

Hence by the triangle inequality, we have

$$\left| \sum_{t \in T} f(t)\, dt - \int_a^b f(t)\, dt \right| \lesssim \varepsilon.$$

Since $\varepsilon \gg 0$ was arbitrary, this finishes the proof. □

Corollary A.6 *Suppose T is a near line and suppose $f : \mathbb{R} \to \mathbb{R}$ is absolutely Riemann integrable with a limited bound and nearly continuous at each point of T. Also suppose*

$$\int_{-\infty}^{a} |f(t)|\, dt + \int_a^{\infty} |f(t)|\, dt \simeq 0, \qquad a \simeq \infty.$$

Then

$$\sum_{t \in T} f(t)\, dt \simeq \int_{-\infty}^{\infty} f(t)\, dt.$$

Proof Fix $\varepsilon \gg 0$. Then

$$\int_{-\infty}^{a} |f(t)|\, dt + \int_{a}^{\infty} |f(t)|\, dt \le \varepsilon$$

holds for every $a \simeq \infty$ hence, by overspill, for some limited a. Applying previous theorem we get

$$\sum_{\substack{t\in T\\ |t|\le a}} f(t)\, dt \simeq \int_{-a}^{a} f(t)\, dt,$$

and by the triangle inequality and the arbitrariness of ε we establish the desired result. □

Remark A.7 Consider the following inequality obtained by integration by parts

$$\int_{x}^{\infty} e^{-t^2/2}\, dt \le \frac{e^{-x^2/2}}{x}$$

This shows that the integral is infinitesimal for unlimited x and since $t \mapsto e^{-t^2/2}$ is nearly continuous for all t and bounded by 1, the above corollary gives

$$\sum_{t\in T} e^{-t^2/2}\, dt \simeq \sqrt{2\pi} \tag{20}$$

whenever T is a near line.

For a different treatment of integration using the full apparatus of IST see Diener and Diener [6, Chapter 9].

B Random Variables

B.1 Almost Sure Infinitesimality and Limitedness

In REPT, a random variable X is *infinitesimal almost surely* if for every $\epsilon \gg 0$ there exists an event N (which may depend on ϵ) such that $\Pr(N) \le \epsilon$ and $X(\omega) \simeq 0$ holds except for $\omega \in N$.

We say that X is *limited almost surely* if for every $\epsilon \gg 0$ there exists an event N (which may depend on ϵ) such that $\Pr(N) \le \epsilon$ and $|X(\omega)| \ll \infty$ holds except for $\omega \in N$.

In REPT there exists three characterizations of almost sure infinitesimality (Nelson [11, Theorem 7.1]). We prove similar characterizations for almost sure limitedness.

Theorem B.1 *The following three conditions are equivalent.*

(i) *X is limited almost surely (i.e $|X| \ll \infty, a.s.$).*

(ii) *If $\eta \simeq \infty$, then $\Pr(|X| \geq \eta) \simeq 0$.*

(iii) *For every $\epsilon \gg 0$ there exists a limited x such that $\Pr(|X| \geq x) \leq \epsilon$.*

Proof Assume (i). Then for every $\epsilon \gg 0$ there exists an event N, $\Pr(N) \leq \epsilon$, such that $X(\omega)$ is limited except when $\omega \in N$. Hence if $x \simeq \infty$ the event $|X| \geq x$ is contained in N, and $\Pr(|X| \geq x) \leq \epsilon$. Since ϵ was arbitrary, (ii) holds. Thus (i) $\Rightarrow$ (ii).

Assume (ii). Fix $\epsilon \gg 0$. Then $\Pr(|X| \geq x) \leq \epsilon$ holds for all unlimited positive x. Hence by overspill it holds for some limited x, which is (iii). Thus (ii) $\Rightarrow$ (iii) and (iii) $\Rightarrow$ (i) is obvious. □

B.2 L_2 Random Variables

In REPT the mere existence of the expectation is vacuous since every random variable has expected value given by a finite sum so a stronger property is imposed. Thus *definition of* L_1 is as follows. A random variable X is L_1 if

$$E\left\{|X|\mathbb{I}_{\{|X|>a\}}\right\} \simeq 0, \qquad a \simeq \infty, \tag{21}$$

and this is stronger than limited absolute expectation because X is L_1 if and only if $E|X| \ll \infty$ and $\Pr(M) \simeq 0 \implies E\{|X|\mathbb{I}_M\} \simeq 0$ (Nelson [11, Theorem 8.1]).

An rv X is L_2 if X^2 is L_1.

Remark B.2 In REPT, we must always write "X is L_1" and never "$X \in L_1$" to emphasize that L_1 is a property, not a set. Forming the set (of random variables) $\{X \in \mathbb{R}^\Omega : X \text{ is } L_1\}$ would be an instance of *illegal set formation* [11, Chapter 4] because L_1 is an *external* property. The following useful properties hold.

$$X \text{ and } Y \text{ are } L_1 \implies X + Y \text{ is } L_1 \tag{22a}$$

$$X \text{ is } L_1 \text{ and } |a| \ll \infty \implies aX \text{ is } L_1 \tag{22b}$$

$$Y \text{ is } L_1 \text{ and } |X| \leq |Y| \implies X \text{ is } L_1 \tag{22c}$$

For any random variable X and any positive real number a define another random variable

$$X^{(a)}(\omega) = \begin{cases} -a, & X(\omega) < -a \\ X(\omega), & -a \leq X(\omega) \leq a \\ a, & X(\omega) > a \end{cases} \tag{23}$$

Lemma B.3 (Approximation) *Suppose X and Y are L_1 random variables and $EX^{(a)} \simeq EY^{(a)}$ for all limited a. Then $EX \simeq EY$.*

Proof For every $\epsilon \gg 0$ we have

$$|EX - EX^{(a)}| \leq \epsilon \quad \text{and} \quad |EY - EY^{(a)}| \leq \epsilon \tag{24}$$

for all $a \simeq \infty$ by (21) and hence by overspill for some $a \ll \infty$. But for this a we have $EX^{(a)} \simeq EY^{(a)}$ and hence by the triangle inequality

$$|EX - EY| \leq 3\epsilon. \tag{25}$$

Since (25) holds for every $\epsilon \gg 0$, the left hand side must be infinitesimal. □

Acknowledgments: We owe much of the readability of this article to an anonymous referee. The second author would like to thank the Graduate Program in Applied Mathematics and Statistics at UFRN and its group of probabilists for the hospitality and discussions.

References

[1] **R M Anderson**, **S Rashid**, *A nonstandard characterization of weak convergence*, Proc. Amer. Math. Soc. 69 (1978) 327–332

[2] **B B Andrade**, *Nonstandard Central Limit Theorems for Markov chains*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 19 (2011) 251–274

[3] **B B Andrade**, *A note on special cases derived from Nelson's martingale central limit theorem*, Positivity (forthcoming)

[4] **I van den Berg**, *Discretisations of higher order and the theorems of Faà di Bruno and DeMoivre-Laplace*, Journal of Logic & Analysis 5 (2013) 1–35

[5] **P Billingsley**, *Convergence of Probability Measures*, 2nd edition, John Wiley & Sons, New York (1999) MR233396

[6] **F Diener**, **M Diener**, *Nonstandard Analysis in Practice*, Springer-Verlag (Universitext), New York (1995) MR1396794

[7] **W Feller**, *An Introduction to Probability Theory and its Applications*, volume 1, 3rd edition, John Wiley & Sons, New York (1968)

[8] **F S Herzberg**, *Stochastic Calculus with Infinitesimals*, Springer-Verlag, New York (2013) Lecture Notes in Mathematics Volume 2067

[9] **P Loeb**, *Conversion from nonstandard to standard measure spaces and applications to probability theory*, Transactions of the American Mathematical Society 211 (1975) 113–122

[10] **E Nelson**, *Internal set theory: a new approach to nonstandard analysis*, Bulletin of the American Mathematical Society 83 (1977) 1165–1198 MR0469763

[11] **E Nelson**, *Radically Elementary Probability Theory*, Princeton University Press, New York (1987) MR906454

[12] **A M Robert**, *Nonstandard Analysis*, Dover, New York (1988)(translated by the author from the 1985 French original)

[13] **A Robinson**, *Nonstandard Analysis*, Princeton University Press, New York (1996) MR1373196(reprint of the 1974 revised edition)

[14] **K D Stroyan**, **J M Bayod**, *Foundations of Infinitesimal Stochastic Analysis*, North-Holland, Amsterdam (1986)

[15] **H Weisshaupt**, *Diffusion processes via parabolic equations: an infinitesimal approach to Lindeberg's limit theorem*, Journal of Logic & Analysis 1 (2009) 1–29

[16] **H Weisshaupt**, *Radically elementary analysis of an interacting particle system at an unstable equilibrium*, Journal of Logic & Analysis 3 (2011) 1–47

School of Statistics, University of Minnesota, Minneapolis, MN, USA

Departamento de Estatística, Universidade de Brasília, Brasília, DF, Brazil

charlie@stat.umn.edu, bbandrade@unb.br

Received: 7 June 2014 Revised: 6 July 2015