Illinois Journal of Mathematics
Volume 61, Numbers 3–4, Fall and Winter 2017, Pages 355–370
S 0019-2082

A NEW DIRECT PROOF OF THE CENTRAL LIMIT THEOREM

VLADIMIR DOBRIĆ AND PATRICIA GARMIRIAN

Abstract. We prove the central limit theorem from the definition of weak convergence using the Haar basis, calculus, and elementary probability, and we estimate the rate of convergence off the tails. The use of the Haar basis pinpoints the role of $L^2([0,1])$ in the CLT as well as the assumption of finite variance.

1. Introduction

In this paper, we give an elementary proof of the central limit theorem (CLT). The proof is elementary in the sense that it avoids the use of characteristic functions and only requires knowledge of elementary probability and calculus. The general idea of the proof is to expand the random variables in the statement of the CLT with respect to the Haar basis and to approximate these expansions by finite sums having m values. This allows finite sums of independent copies of these random variables to be approximated by multinomial distributions. Calculations of the multinomial coefficients are then made explicit by Stirling's formula and Taylor series approximations. Our main result is a proof of Theorem 1.

Theorem 1. *Let (X_i) be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $f:\mathbb{R}\to\mathbb{R}$ be a bounded, continuous function. Then, for each $\varepsilon>0$, there exists $n_1\in\mathbb{N}$ such that*

$$\left|E\left(f\left(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\right)\right)-E\big(f(Y)\big)\right|<\varepsilon\big(9\|f\|_\infty+2\big)$$

for all $n\geq n_1$, where Y is a standard normal random variable.

Received January 18, 2017; received in final form February 26, 2018.

Garmirian dedicates this paper to the memory of her dear departed mentor, Vladimir Dobrić, and acknowledges the helpful comments of Daniel Conus, Lee Stanley, and Rob Neel, as well as the helpful comments of the anonymous referees.

2010 *Mathematics Subject Classification.* Primary 60F05. Secondary 42C40, 28A33.

We provide a brief history of the CLT. The first major contribution to the CLT was in 1733 by de Moivre. De Moivre proved a version of the CLT for Bernoulli random variables in which the binomial coefficients were approximated via Stirling's formula and Taylor series. In 1820, the work of Laplace laid the groundwork for a more general CLT [2]. Our proof extends the work of de Moivre by establishing a connection between the general CLT and the multinomial distribution.

In 1935, both Lévy and Feller independently proved the CLT. Feller's method of proof was based on characteristic functions. [6] While Lévy had championed characteristic functions in the 1920s by proving his Continuity Theorem, Lévy favored more rigorous analytic methods for proving the CLT. Lévy's proof relied on a decomposition of sums of random variables and the dispersion of these sums. Lévy believed that these analytic methods provided more intuition about the CLT than the method of characteristic functions [7].

Later proofs of the CLT have also avoided the use of characteristic functions. In [1] (1941), Berry calculates exact estimates of the supremum of the absolute value of the difference between the cumulative distribution functions for the i.i.d. sums of random variables and the normal distribution using elementary calculus. In [9] (1959), Trotter restates the definition of weak convergence of a sequence of random variables as the weak-$\star$ convergence of a sequence of operators. The defined operators are contraction operators on the space of functions having continuous second derivatives. These functions are then expanded as second order Taylor polynomials with an error term which is controlled by the assumption of continuous second derivative.

Our proof is similar to those in [1] and [9] as it uses calculus to obtain error estimates. While the proof in [1] makes use of the cumulative distribution functions for the relevant distributions, our proof (as well as the proof in [9]) uses weak-$\star$ convergence of the relevant sequence of measures. Although the proof in [9] does not give a rate of convergence, the Berry–Esseen theorem (which combines the results in [1] along with the work of Esseen in [5] and [4]) gives both a rate and exact constant of convergence assuming finite third moment. Our constant of convergence depends on $\|f\|_\infty$ (where f is given in the definition of weak convergence) and m (which depends on the number of terms in the Haar expansion), however, we only assume a finite second moment.

In [8] (1972), Stein's method set the stage for more generalizations of the CLT. This method establishes a normality criterion for the expectations of differential equations of random variables for a certain class of functions. One important advantage of Stein's method is that it not only applies to sequences of independent random variables but dependent random variables as well [2].

There continues to be interest in new elementary proofs of the CLT. In [3] (2005), Dalang proved the CLT by using weak-$\star$ convergence for a sequence of operators on the space of bounded functions with continuous third derivatives. Dalang expresses these functions as second order Taylor polynomials and calculates their errors. He then calculates the rate of convergence in the CLT by expressing the error as a telescoping sum of incremental errors. In [10] (2013), Zong and Hu prove the CLT by showing that the weak-$\star$ limit of a sequence of normalized sums of i.i.d. random variables satisfies the heat equation.

Each of these proofs is similar to our proof due to the avoidance of characteristic functions and the use of techniques from calculus. Although the proof in [10] does not give a rate of convergence, the proof in [3] does give a rate and constant of convergence. The constant of convergence in [3] depends on the sup norms of the second and third derivatives of f. Although our constant also depends on f, we use the weaker assumption that f is only a bounded, continuous function.

The steps of our proof proceed as follows: given an i.i.d. sequence of random variables on a probability space, we construct an i.i.d. sequence on $[0, 1]$ with the Borel sigma algebra and Lebesgue measure having the same sequence of distributions. As the new sequence of random variables is defined on $[0, 1]$ and also has finite variance, we then expand this sequence with respect to the Haar basis.

We then reduce the problem of showing weak convergence of this new sequence of random variables to the case where the Haar expansions are truncated to have only M terms, for some finite M which will be chosen to accomplish certain other objectives (Lemma 1). These truncated Haar expansions each have $m = 2^{M+1}$ possible outcomes. Next, we show that the sum of Haar expansions having only M terms is in fact the projection of a multinomial random variable.

In Lemma 2, we identify the tails of the multinomial random variable. After cutting off these tails, we compute the probabilities for the multinomial distribution using Stirlings's formula and Taylor series approximation (Lemma 3). The appearance of the Gaussian density on the multinomial side can be seen in this step.

On the Gaussian side, we express a standard normal random variable as a sum of m independent normal random variables with coefficients being the outcomes of the truncated Haar expansion. We then apply Fubini's theorem to reduce by one dimension the expression for the expected value on the Gaussian side as an integral over a hyperplane in $\mathbb{R}^m$ (Lemma 4). In Lemma 5, we identify the tails of the Gaussian. After cutting off these tails, we approximate the integral by a Riemann sum. The Riemann and the multinomial sums match perfectly.

In Proposition 6, by bounding the function f by its sup norm, we estimate the sum of the absolute values of the differences between the multinomial and Gaussian probabilities. It is here that we also obtain the rate of convergence of $n^{-1/2}$ and the constant for convergence of $\frac{2m^2}{3\sqrt{2\pi}}$, off the tails. In both instances, the restriction to "off the tails" arises since our truncations (of the Haar expansions, the multinomial sum, and the Gaussian Riemann sum) are based Chebyshev's inequality, in which coarseness is the price of its generality. Finally, in Theorem 1, we pull together the preceding results to prove the CLT.

2. Preliminary estimates

2.1. The Haar expansion and approximation. Let $\varepsilon > 0$. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded, continuous function. Let Z be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. We may assume that $E(Z) = 0$ and $\mathrm{var}(Z) = 1$. Define the quantile of Z to be the function $X : [0,1] \to \mathbb{R}$ defined by

$$X(x) := \inf\{y \in \mathbb{R} | P(Z \le y) \ge x\}.$$

Then, X is a random variable on the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, where $\mathcal{B}([0,1])$ denotes the Borel sets of $[0,1]$ and λ is Lebesgue measure. Further, X has the same distribution as the random variable Z.

For $x \in (0,1)$, let $\varepsilon_i(x)$ be the ith bit in the binary expansion of x (for dyadic rationals, choose the expansion with the tail of 0's). We create the following matrix of binary digits:

$$\begin{pmatrix} \varepsilon_1 & \varepsilon_3 & \varepsilon_6 & \\ \varepsilon_2 & \varepsilon_5 & \varepsilon_9 & \cdots \\ \varepsilon_4 & \varepsilon_8 & \varepsilon_{13} & \\ & \vdots & & \end{pmatrix}.$$

For all $x \in (0,1)$, define $P_i(x)$ to have binary expansion given by the ith column of the matrix. Let $X_i(x) := X(P_i(x))$. Then, (X_i) is an i.i.d. sequence of random variables on $[0,1]$ having the same distribution as X.

Let $L^2([0,1])$ denote the space of all functions $g : [0,1] \to \mathbb{R}$ satisfying

$$\|g\|_2^2 := \int_0^1 \left|g(x)\right|^2 d\lambda(x) < \infty.$$

The Haar basis is the simplest orthonormal system in $L^2([0,1])$ and consists of the set $S = \{H_{j,k}(x) | 0 \le j < \infty, 0 \le k \le 2^j - 1\} \cup \{\chi_{[0,1]}\}$, where

$$H_{j,k}(x) := \begin{cases} 2^{\frac{j}{2}}, & x \in [\frac{k}{2^j}, \frac{k+\frac{1}{2}}{2^j}), \\ -2^{\frac{j}{2}}, & x \in [\frac{k+\frac{1}{2}}{2^j}, \frac{k+1}{2^j}), \\ 0 & \text{otherwise} \end{cases}$$

and

$$\chi_{[0,1]}(x) = \begin{cases} 1, & x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$E(X) = \int_0^1 X(x)\, d\lambda(x) = 0,$$

then

$$\|X\|_2^2 = \int_0^1 \left|X(x)\right|^2 d\lambda(x) = E\left(X^2\right) = \operatorname{Var}(X) = 1 < \infty.$$

Thus, $X \in L^2([0,1])$, and

$$X(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{j,k} H_{j,k}(x),$$

where $c_{j,k} = \int_0^1 X(x) H_{j,k}(x)\,dx$. Then,

$$X(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{j,k} 2^{\frac{j}{2}} (-1)^{\varepsilon_{j+1}(x)} \chi_{\{k\}}\left(\lfloor 2^j x \rfloor\right) = \sum_{j=0}^{\infty} 2^{\frac{j}{2}} c_{j,\lfloor 2^j x \rfloor} (-1)^{\varepsilon_{j+1}(x)},$$

where, as usual, $\lfloor x \rfloor$ denotes the greatest integer $\leq x$, and

$$\chi_{\{k\}}(x) = \begin{cases} 1, & x = k, \\ 0 & \text{otherwise.} \end{cases}$$

For $n \geq 1$, define

$$S_n(x) := \sum_{i=1}^{n} X_i(x) = \sum_{i=1}^{n} \sum_{j=0}^{\infty} 2^{\frac{j}{2}} c_{j,\lfloor 2^j P_i(x) \rfloor} (-1)^{\varepsilon_{j+1}(P_i(x))},$$

and for $M \geq 1$, define

$$S_{n,M}(x) := \sum_{i=1}^{n} X_{i,M}(x) = \sum_{i=1}^{n} \sum_{j=0}^{M} 2^{\frac{j}{2}} c_{j,\lfloor 2^j P_i(x) \rfloor} (-1)^{\varepsilon_{j+1}(P_i(x))} \tag{1}$$

and

$$\sigma_M^2 := \operatorname{Var}(S_{n,M}) = \sum_{j=0}^{M} \sum_{k=0}^{2^j-1} c_{j,k}^2. \tag{2}$$

LEMMA 1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded, continuous function. Given any $\varepsilon > 0$, there exists a positive integer M_0 such that for all $M \geq M_0$:*

$$\left| \int_0^1 f\left(\frac{S_n(x)}{\sqrt{n}}\right) d\lambda(x) - \int_0^1 f\left(\frac{S_{n,M}(x)}{\sigma_M \sqrt{n}}\right) d\lambda(x) \right| < \varepsilon\left(6\|f\|_\infty + 1\right).$$

Proof. Note that $E(\frac{S_n}{\sqrt{n}}) = 0$ and $\operatorname{Var}(\frac{S_n}{\sqrt{n}}) = 1$. Let $\varepsilon > 0$, and define

$$A := \left\{ \left| \frac{S_n}{\sqrt{n}} \right| > L \right\} \quad \text{and} \quad B_M := \left\{ \left| \frac{S_{n,M}}{\sigma_M \sqrt{n}} \right| > L \right\}.$$

By Chebyshev's inequality,

$$\lambda(A) \leq \frac{1}{L^2} < \varepsilon \quad \text{and} \quad \lambda(B_M) \leq \frac{1}{L^2} < \varepsilon$$

for L large. Since f is uniformly continuous on $[-L, L]$, then there exists a $\delta > 0$ such that $x, y \in [-L, L]$ satisfying $|x - y| < \delta$ implies that $|f(x) - f(y)| < \varepsilon$. Now, let

$$C_M := \left\{ \left| \frac{S_n}{\sqrt{n}} - \frac{S_{n,M}}{\sigma_M \sqrt{n}} \right| \geq \delta \right\}.$$

Since $\sigma_M \to 1$ as $M \to \infty$, then there exists an $M_0 \in \mathbb{N}$ such that for all $M \geq M_0$:

$$\operatorname{var}\left(\frac{S_n}{\sqrt{n}} - \frac{S_{n,M}}{\sigma_M \sqrt{n}} \right) \leq \left(1 - {\sigma_M}^2\right) + 2(1 - \sigma_M)\sqrt{1 - \sigma_M^2} + (1 - \sigma_M)^2 < \varepsilon \delta^2,$$

and so

$$\lambda(C_M) \leq \frac{\varepsilon \delta^2}{\delta^2} = \varepsilon.$$

Now, let $S := A^c \cap B_M^c \cap C_M^c$. Then, $\lambda(S^c) < 3\varepsilon$. Hence,

$$\left| E\left(f\left(\frac{S_n}{\sqrt{n}} \right) \right) - E\left(f\left(\frac{S_{n,M}}{\sigma_M \sqrt{n}} \right) \right) \right| \leq 2\|f\|_\infty \lambda\left(S^c\right) + \varepsilon \lambda(S) \leq \varepsilon\left(6\|f\|_\infty + 1\right).$$

$\square$

2.2. The multinomial random variable. Let

$$X_{0,M}(x) := \frac{1}{\sigma_M} \sum_{j=0}^{M} 2^{\frac{j}{2}} c_{j,\lfloor 2^j x \rfloor} (-1)^{\varepsilon_{j+1}(x)}$$

for $x \in [0, 1]$. Below, we investigate the properties of this random variable. Note that $X_{0,M}$ is a random variable which depends on $(\varepsilon_1, \ldots, \varepsilon_{M+1})$. From now on, we will let $m := 2^{M+1}$ for notational convenience. Thus, $X_{0,M}$ is constant on dyadic intervals of length $2^{-(M+1)} (= \frac{1}{m})$. Let $o_1, \ldots, o_m$ denote the m possible values of $X_{0,M}$. It follows that

$$\sum_{i=1}^{m} o_i = 0 \quad \text{and} \quad \sum_{i=1}^{m} o_i^2 = m$$

as $E(X_{0,M}) = 0$ and $\operatorname{var}(X_{0,M}) = 1$.

We will now take a closer look at the random variable $S_{n,M}$ of Equality (1). Each $X_{i,M}$ is a random variable with the m possible outcomes $o_1, \ldots, o_m$. Let K_i be the random variable which denotes the number of times the outcome o_i is observed among n independent trials. Then,

$$S_{n,M}(x) = K_1(x) o_1 + \cdots + K_m(x) o_m, \tag{3}$$

where $K_1 + \cdots + K_m = n$.

Note that $S_{n,M}$ is a scalar product of an m-nomial random variable and the vector of outcomes. Since each outcome has probability $\frac{1}{m}$ and the trials are independent,

$$\lambda\left(\left\{x \in (0, 1) : K_1(x) = k_1, \ldots, K_m(x) = k_m\right\}\right) = \binom{n}{k_1, \ldots, k_m} \left(\frac{1}{m}\right)^n$$

and

$$E\left(f\left(\frac{S_{n,M}(x)}{\sqrt{n}}\right)\right) = \sum_{\substack{k_1=0 \\ k_1+\cdots+k_m=n}}^{n} \cdots \sum_{k_m=0}^{n} \left(\frac{1}{m}\right)^n \binom{n}{k_1,\ldots,k_m} f\left(\frac{\sum_{i=1}^m k_i o_i}{\sqrt{n}}\right).$$

The following lemma allows us to cut off the tails from the multinomial random variable. Consequently, we prepare the ground for the usage of Taylor's formula. The tails of the multinomial random variable consist of all $(k_1,\ldots,k_m) \in \{0,1,\ldots,n\}^m$ such that $k_1+\cdots+k_m = n$ and $k_i \notin [\lfloor\frac{n}{m}\rfloor - \lfloor b\sqrt{n}\rfloor, \lfloor\frac{n}{m}\rfloor + \lfloor b\sqrt{n}\rfloor]$ for some $1 \le i \le m-1$.

LEMMA 2. *Let*

$$q(n,k_1,\ldots,k_m) := \left(\frac{1}{m}\right)^n \binom{n}{k_1,\ldots,k_m} f\left(\frac{\sum_{i=1}^m k_i o_i}{\sqrt{n}}\right).$$

Then, there exists a b_0 such that for all $b \ge b_0$:

$$\begin{aligned}&\Bigg|\sum_{\substack{k_1=1 \\ k_1+\cdots+k_m=n}}^{n} \cdots \sum_{k_m=0}^{n} q(n,k_1,\ldots,k_m) \\ &\qquad - \sum_{\substack{k_1=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor \\ k_1+\cdots+k_m=n}}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} \cdots \sum_{k_{m-1}=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} q(n,k_1,\ldots,k_m)\Bigg| \\ &\quad < \varepsilon\|f\|_\infty.\end{aligned}$$

Proof. Recall that K_i is the random variable which denotes the number of times the outcome o_i is observed, having values k_i. Since each K_i is a binomial random variable, we have $E(K_i) = \frac{n}{m}$ and $\operatorname{var}(K_i) = n(\frac{1}{m})(1-\frac{1}{m})$. By Chebyshev's inequality,

$$\lambda\left(\left|K_i - \frac{n}{m}\right| \ge b\sqrt{n}\right) \le \frac{(1-\frac{1}{m})}{b^2 m} \le \frac{1}{b^2 m}.$$

Then, there exists a b_0 such that for all $b \ge b_0$:

$$\begin{aligned}&\lambda\left(\left|K_i - \frac{n}{m}\right| \ge b\sqrt{n} \text{ for some } 1 \le i \le m-1\right) \\ &\qquad \le \sum_{i=1}^{m-1} \lambda\left(\left|K_i - \frac{n}{m}\right| \ge b\sqrt{n}\right) \le \frac{1}{b^2} < \varepsilon.\end{aligned}$$

Let

$$L_{n,m} := \sum_{\substack{k_1=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor \\ k_1+\cdots+k_m=n}}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} \cdots \sum_{k_{m-1}=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} \binom{n}{k_1,\ldots,k_m}\left(\frac{1}{m}\right)^n.$$

Thus,

$$\begin{aligned}
&\Bigg|\sum_{\substack{k_1=1\\k_1+\cdots+k_m=n}}^{n}\cdots\sum_{k_m=0}^{n} q(n,k_1,\dots,k_m)\\
&\qquad - \sum_{k_1=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor}\cdots\sum_{\substack{k_{m-1}=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor\\k_1+\cdots+k_m=n}}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} q(n,k_1,\dots,k_m)\Bigg|\\
&\quad\le \|f\|_\infty\Bigg|\sum_{\substack{k_1=1\\k_1+\cdots+k_m=n}}^{n}\cdots\sum_{k_m=0}^{n}\binom{n}{k_1,\dots,k_m}\left(\frac{1}{m}\right)^n - L_{n,m}\Bigg|\\
&\quad= \|f\|_\infty\lambda\left(\left|K_i-\frac{n}{m}\right|\ge b\sqrt{n}\text{ for some }1\le i\le m-1\right) < \|f\|_\infty\varepsilon. \qquad\square
\end{aligned}$$

In the following lemma, we will use Stirling's formula and Taylor series to approximate the probabilities for the multinomial distribution. For use here and in the proof of Theorem 1, we define some functions. For $n>0$, we let:

$$d_n := \sqrt{m}\left(\frac{m}{2n\pi}\right)^{\frac{m-1}{2}}. \tag{4}$$

For $n>0$ and integers, $j_1,\dots,j_m$ whose sum is 0, we let:

$$H(n,j_1,\dots,j_m) := -\frac{m}{2n}\sum_{i=1}^{m} j_i^2 + \frac{m^2}{6n^2}\sum_{i=1}^{m} j_i^3 \tag{5}$$

and

$$p(n,j_1,\dots,j_m) := d_n e^{H(n,j_1,\dots,j_m)}. \tag{6}$$

LEMMA 3. *Let $j_i = k_i - \lfloor\frac{n}{m}\rfloor$, and suppose that $-\lfloor b\sqrt{n}\rfloor \le j_i \le \lfloor b\sqrt{n}\rfloor$ for $1\le i\le m-1$ and $j_m = -j_1-\cdots-j_{m-1}$ and $n\ge b^2m^2$. Then,*

$$\frac{1}{m^n}\binom{n}{k_1,\dots,k_m} = p(n,j_1,\dots,j_m) + O\left(\frac{1}{n}\right).$$

The proof of this lemma is given in the Appendix.

2.3. The Gaussian side. Now we consider the Gaussian side. Let $Y_1,\dots,Y_m$ be i.i.d. standard normal random variables. Then, by the properties of i.i.d. normal random variables,

$$Y := \frac{1}{\sqrt{m}}\sum_{i=1}^{m} o_i Y_i$$

is a standard normal random variable.

LEMMA 4. *Define*

$$g(z_1,\dots,z_m) := f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^m o_i z_i\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m z_i^2\right).$$

Then,

$$\mathbb{E}\big(f(Y)\big) = \frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g\left(z_1,\dots,z_{m-1},-\sum_{i=1}^{m-1} z_i\right) dz_1\cdots dz_{m-1}.$$

Proof. Set

$$o = \begin{bmatrix} o_1 \\ \vdots \\ o_m \end{bmatrix}, \qquad \tilde{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}, \qquad u = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Define the hyperplane $S \subseteq \mathbb{R}^m$ as follows:

$$S := \left\{ (y_i)_{i=1}^m \in \mathbb{R}^m : \sum_{i=1}^m y_i = 0 \right\}.$$

Since the vector u is orthogonal to S, then the orthogonal projection of $\tilde{Y}$ onto S is given by:

$$Z := \tilde{Y} - V, \quad \text{where } V := \left(\frac{1}{\sqrt{m}}\tilde{Y}\cdot u\right)u.$$

Since $\sum_{i=1}^m o_i = 0$, then $o^T V = 0$, and so

$$\mathbb{E}\big(f(Y)\big) = \mathbb{E}\left(f\left(\frac{1}{\sqrt{m}}o^T\tilde{Y}\right)\right) = \mathbb{E}\left(f\left(\frac{1}{\sqrt{m}}o^T Z\right)\right).$$

As the orthogonal projection of the random variable $\tilde{Y}$ onto the hyperplane S, Z is an $(m-1)$-dimensional standard normal random variable with values in S. Thus:

$$\begin{aligned}\mathbb{E}\big(f(Y)\big) &= \mathbb{E}\left(f\left(\frac{1}{\sqrt{m}}o^T Z\right)\right)\\ &= \frac{1}{(\sqrt{2\pi})^{m-1}}\int_S f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^m o_i y_i\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m y_i^2\right) dS.\end{aligned}$$

Evaluating the surface integral:

$$\mathbb{E}\big(f(Y)\big) = \frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g\left(z_1,\dots,z_{m-1},-\sum_{i=1}^{m-1} z_i\right) dz_1\cdots dz_{m-1}.$$

□

For the following results, let $j_i = k_i - \lfloor \frac{n}{m} \rfloor$. Then, define the sum "off the tails" as follows:

$$\sum_{A(n,m,b)} := \sum_{\substack{j_1=-\lfloor b\sqrt{n}\rfloor \\ j_1+j_2+\cdots+j_m=0}}^{\lfloor b\sqrt{n}\rfloor} \cdots \sum_{j_{m-1}=-\lfloor b\sqrt{n}\rfloor}^{\lfloor b\sqrt{n}\rfloor}.$$

LEMMA 5. *Let*

$$I := \frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}} \int_{y_1+y_2+\cdots+y_m=0}\cdots\int f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^m o_i y_i\right) e^{(-\frac{1}{2}\sum_{i=1}^m y_i^2)}\, dy_1\cdots dy_{m-1}.$$

Then, there exist $n_0 \in \mathbb{N}$ and $b_1 > 0$ such that for all $n \geq n_0$ and for all $b \geq b_1$,

$$\left| I - \sum_{A(n,m,b)} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^m o_i j_i\right) e^{-(\frac{m}{2}\sum_{i=1}^m \frac{j_i^2}{n})} \right| < \varepsilon\big(\|f\|_\infty + 1\big).$$

Proof. Let

$$C := \left\{ y \in \mathbb{R}^m : y_i \in [-b\sqrt{m}, b\sqrt{m}]\ \forall 1 \leq i \leq m-1 \text{ and } y_m = -\sum_{i=1}^{m-1} y_i \right\}.$$

We have

$$\begin{aligned}
&\frac{1}{(\sqrt{2\pi})^{m-1}} \int_S \exp\left(-\frac{1}{2}\sum_{i=1}^m y_i^2\right) dS \\
&\quad = \frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}} \int_{y_1+y_2+\cdots+y_m=0}\cdots\int \exp\left(-\frac{1}{2}\sum_{i=1}^m y_i^2\right) dy_1\cdots dy_{m-1} < \infty.
\end{aligned}$$

Let

$$I_b := \frac{\sqrt{m}}{(2\pi)^{\frac{m-1}{2}}} \int_{y\in C}\cdots\int f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^m o_i y_i\right) e^{(-\frac{1}{2}\sum_{i=1}^m y_i^2)}\, dy_1\cdots dy_{m-1}.$$

Then, there exists a b_1 such that for all $b \geq b_1$,

$$\frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}} \left(\int_{y\in C^c}\cdots\int e^{-\frac{1}{2}\sum_{i=1}^m y_i^2}\, dy_1\cdots dy_{m-1} \right) < \varepsilon,$$

and so

$$|I - I_b| < \varepsilon \|f\|_\infty.$$

Suppose that $b \geq b_1$. Then,

$$I_b = \lim_{n\to\infty} \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} \sum_{A(n,m,b)} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^m o_i j_i\right) e^{(-\frac{m}{2}\sum_{i=1}^m \frac{j_i^2}{n})}.$$

Here I_b is a limit of Riemann sums with points of discretization $(\frac{\sqrt{m}j_i}{\sqrt{n}})_{i=1}^{m-1}$, where $-\lfloor b\sqrt{n}\rfloor \le j_i \le \lfloor b\sqrt{n}\rfloor$, and step size $(\frac{m}{n})^{\frac{m-1}{2}}$. Hence, there exists an $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$,

$$\left| I - \sum_{A(n,m,b)} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{m} o_i j_i\right) e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \right| < \varepsilon\big(\|f\|_\infty + 1\big).$$

□

3. Main results

Theorem 1 gives our proof of the CLT. The proof appeals to Proposition 6, which provides a comparison of the multinomial and Gaussian sides; the rate is $n^{-1/2}$ and the constant of convergence is $\frac{2m^2}{3\sqrt{2\pi}}$. These only hold off the tails as we have truncated the Haar expansions, the multinomial sum, and the Gaussian Riemann sum. From now on, let $b \ge \max\{b_0, b_1\}$, where the former is as in Lemma 2 and the latter is as in the proof of Lemma 5.

Proposition 6. *Let*

$$D_n := \sum_{A(n,m,b)} \left| \left(\frac{1}{m^n} \binom{n}{\lfloor \frac{n}{m}\rfloor + j_1, \ldots, \lfloor \frac{n}{m}\rfloor + j_m} - \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \right) \right|.$$

Then,

$$D_n \le \frac{2m^2}{3\sqrt{2\pi n}} + O\big(n^{-1}\big).$$

Proof. By Lemma 3,

$$D_n = \sum_{A(n,m,b)} \left| e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \big(e^{G(n,j_1,\ldots,j_m)+O(n^{-1})} - 1\big) \right|,$$

where $G(n, j_1, \ldots, j_m) = (\frac{m^2}{4n^2})\sum_{i=1}^{m} j_i^2 + (\frac{m^2}{6n^2} - \frac{m^3}{6n^3})\sum_{i=1}^{m} j_i^3 - \frac{m^3}{3n^3}\sum_{i=1}^{m} j_i^4$. Then,

$$\begin{aligned} D_n &= \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \left| e^{G(n,j_1,\ldots,j_m)+O(n^{-1})} - 1 \right| \\ &\le \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \left| G(n,j_1,\ldots,j_m) + O\big(n^{-1}\big) \right|. \end{aligned}$$

All of the terms which decay at a rate of n^{-1} or faster are absorbed into the error term $O(n^{-1})$. We let

$$\begin{aligned} E_n &:= \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} \left(\frac{m^2}{6n^2}\right) \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} \left(\sum_{i=1}^{m-1} |j_i|^3 \right) \\ &\le \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} \left(\frac{m^2}{6n^2}\right) \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m-1}\frac{j_i^2}{n})} \left(\sum_{i=1}^{m-1} |j_i|^3 \right). \end{aligned}$$

Approximating the sum by an integral,

$$E_n < (m-1)\sqrt{m} \cdot \frac{m^2}{6n^2} \cdot \left(\frac{n}{m}\right)^{3/2} E\big(|X|^3\big) + O\big(n^{-1}\big),$$

where X is a standard normal random variable. Thus,

$$E_n < \frac{2}{\sqrt{2\pi}} \frac{m(m-1)}{3\sqrt{n}} + O\big(n^{-1}\big).$$

Now, consider

$$F_n := \left(\frac{m^2}{6n^2}\right) \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} |j_m|^3.$$

By maximizing $e^{-\frac{1}{2}x^2}|x|^3$,

$$F_n \le \left(\frac{m^2}{6n^2}\right) \frac{m^{\frac{m}{2}}}{(2\pi n)^{\frac{m-1}{2}}} \left(\frac{n}{m}\right)^{3/2} \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m-1}\frac{j_i^2}{n})} e^{-3/2} 3^{3/2}.$$

Approximating the sum by an integral,

$$F_n < \sqrt{m} \cdot \left(\frac{m^2}{6n^2}\right) \cdot \left(\frac{n}{m}\right)^{3/2} e^{-3/2} 3^{3/2} + O\big(n^{-1}\big).$$

Thus,

$$F_n < \frac{m}{6\sqrt{n}} e^{-3/2} 3^{3/2} + O\big(n^{-1}\big).$$

Hence, we have

$$E_n + F_n < \frac{2}{\sqrt{2\pi}} \cdot \frac{m(m-1)}{3\sqrt{n}} + \frac{m}{6\sqrt{n}} e^{-3/2} 3^{3/2} + O\big(n^{-1}\big) < \frac{2m^2}{3\sqrt{2\pi n}} + O\big(n^{-1}\big).$$

Since $-\lfloor b\sqrt{n}\rfloor \le j_i \le \lfloor b\sqrt{n}\rfloor$ for $1 \le i \le m-1$, then the terms

$$\left(\frac{m^2}{4n^2}\right) \sum_{i=1}^{m-1} |j_i|^2 \quad \text{and} \quad \left(\frac{m^3}{3n^3}\right) \sum_{i=1}^{m-1} |j_i|^4$$

can be absorbed in the error term $O(n^{-1})$. Maximizing $e^{-\frac{1}{2}x^2}|x|^2$ and $e^{-\frac{1}{2}x^2}|x|^4$ and approximating the sums by integrals,

$$\begin{aligned} &\frac{m^{\frac{m}{2}}}{(2\pi)^{\frac{m-1}{2}}} \left(\frac{m^2}{4n^2}\right) \sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^{m}\frac{j_i^2}{n})} |j_m|^2 \\ &\quad \le \sqrt{m}\left(\frac{m^2}{4n^2}\right)\left(\frac{n}{m}\right) e^{-1} \cdot 2 + O\big(n^{-1}\big) = \frac{m^{3/2}e^{-1}}{2n} + O\big(n^{-1}\big) \end{aligned}$$

and

$$\frac{m^{\frac{m}{2}}}{(2\pi)^{\frac{m-1}{2}}}\left(\frac{m^3}{3n^3}\right)\sum_{A(n,m,b)} e^{-(\frac{m}{2}\sum_{i=1}^m \frac{j_i^2}{n})}|j_i|^4$$
$$\leq \sqrt{m}\left(\frac{m^3}{3n^3}\right)\left(\frac{n}{m}\right)^2 e^{-2}\cdot 16 + O\big(n^{-1}\big) = \frac{16m^{3/2}e^{-2}}{3n} + O\big(n^{-1}\big).$$

It then follows that

$$D_n \leq \frac{2m^2}{3\sqrt{2\pi n}} + O\big(n^{-1}\big).$$

□

Finally, we prove the CLT using Lemmas 1–5. Recall from Lemma 2 that

$$q(n,k_1,\ldots,k_m) := \left(\frac{1}{m}\right)^n \binom{n}{k_1,\ldots,k_m} f\left(\frac{\sum_{i=1}^m k_i o_i}{\sqrt{n}}\right).$$

Recall from Lemma 3 that

$$d_n := \sqrt{m}\left(\frac{m}{2n\pi}\right)^{\frac{m-1}{2}}.$$

For $n > 0$ and integers, $j_1,\ldots,j_m$ whose sum is 0:

$$H(n,j_1,\ldots,j_m) := \frac{-m}{2n}\sum_{i=1}^m j_i^2 + \frac{m^2}{6n^2}\sum_{i=1}^m j_i^3$$

and

$$p(n,j_1,\ldots,j_m) := d_n e^{H(n,j_1,\ldots,j_m)}.$$

We now give a proof of Theorem 1.

Proof of Theorem 1. Let $M > M_0$, $b > b_0, b_1$, and $n > n_0$. By Lemma 1, we reduce the problem to dealing with the projection of a multinomial random variable and we have

$$\begin{aligned}\Delta_n &:= \left|E\left(f\left(\frac{S_n}{\sqrt{n}}\right)\right) - E\big(f(Y)\big)\right| \\ &< \left|E\left(f\left(\frac{S_{n,M}}{\sigma_M\sqrt{n}}\right)\right) - E\big(f(Y)\big)\right| + \varepsilon\big(6\|f\|_\infty + 1\big) \\ &= \left|\sum_{\substack{k_1=0\\ k_1+\cdots+k_m=n}}^{n}\cdots\sum_{k_m=0}^{n} q(n,k_1,\ldots,k_m) - E\big(f(Y)\big)\right| + \varepsilon\big(6\|f\|_\infty + 1\big).\end{aligned}$$

By Lemma 2, we cut off the tails of the multinomial random variable to obtain

$$\begin{aligned}\Delta_n < &\left|\sum_{k_1=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor}\cdots\sum_{\substack{k_{m-1}=\lfloor\frac{n}{m}\rfloor-\lfloor b\sqrt{n}\rfloor\\ k_1+\cdots+k_m=n}}^{\lfloor\frac{n}{m}\rfloor+\lfloor b\sqrt{n}\rfloor} q(n,k_1,\ldots,k_m) - E\big(f(Y)\big)\right| \\ &+ \varepsilon\big(7\|f\|_\infty + 1\big).\end{aligned}$$

By Lemma 3, we further simplify the multinomial sum to obtain

$$\Delta_n < \left| \sum_{A(n,m,b)} p(n, j_1, \ldots, j_m) f\left(\frac{\sum_{i=1}^m j_i o_i}{\sqrt{n}}\right) - E\big(f(Y)\big) \right| + \varepsilon\big(7\|f\|_\infty + 1\big).$$

Writing Y as a sum of m independent standard normal random variables, it follows that $\frac{1}{\sqrt{m}} \sum_{i=1}^m o_i Y_i \overset{d}{=} N(0,1)$. By Lemma 4,

$$\mathbb{E}\big(f(Y)\big) = \frac{\sqrt{m}}{(\sqrt{2\pi})^{m-1}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g\left(z_1, \ldots, z_{m-1}, -\sum_{i=1}^{m-1} z_i\right) dz_1 \cdots dz_{m-1},$$

where

$$g(z_1, \ldots, z_m) := f\left(\frac{1}{\sqrt{m}} \sum_{i=1}^m o_i z_i\right) \exp\left(-\frac{1}{2} \sum_{i=1}^m z_i^2\right).$$

By Lemma 5, we approximate the Gaussian integral by a Riemann sum. This approximation allows us to match the multinomial side and the Gaussian side and apply Proposition 6:

$$\begin{aligned}\Delta_n &< d_n \left| \sum_{A(n,m,b)} f\left(\frac{\sum_{i=1}^m j_i o_i}{\sqrt{n}}\right) \left(e^{H(n,j_1,\ldots,j_m)} - e^{-(\frac{m}{2} \sum_{i=1}^m \frac{j_i^2}{n})}\right) \right| \\ &\quad + \varepsilon\big(8\|f\|_\infty + 2\big) \\ &< \varepsilon\big(8\|f\|_\infty + 2\big) + \frac{2m^2 \|f\|_\infty}{3\sqrt{2\pi n}} + O\big(n^{-1}\big) < \varepsilon\big(9\|f\|_\infty + 2\big). \qquad \square\end{aligned}$$

REMARK 1. The i.i.d. assumption is necessary so that sums of the approximating random variables have multinomial distributions. The assumption in the dependent case requires that sums of the random variables are sufficiently close the multinomial distributions: In order for the proof to carry through in the dependent case, the following is a sufficient condition:

$$\sum_{A(n,m,b)} \left| \frac{1}{m^n} \binom{n}{\frac{n}{m} + j_1, \ldots, \frac{n}{m} + j_m} - P\left(S_{n,M} = \sum_{i=1}^m k_i o_i\right) \right| \leq \frac{C}{\sqrt{n}}$$

for some $C > 0$.

REMARK 2. The assumption of "off the tails" can be removed and the same rate is attained provided that there exists a $C > 0$ such that

$$\left| f\left(\frac{\sum_{i=1}^m k_i o_i}{\sqrt{n}}\right) \right| \leq \frac{C}{\sqrt{n}}$$

for all $(k_1, \ldots, k_m) \in \{0, 1, \ldots, n\}^m$ such that $k_1 + \cdots + k_m = n$ and $k_i \notin [\lfloor \frac{n}{m} \rfloor - \lfloor b\sqrt{n} \rfloor, \lfloor \frac{n}{m} \rfloor + \lfloor b\sqrt{n} \rfloor]$ for some $1 \leq i \leq m-1$.

Appendix

In this section, we give a proof of Lemma 3.

Proof. Set

$$l(n,k_1,\dots,k_m) := \frac{1}{m^n}\binom{n}{k_1,\dots,k_m}.$$

By Stirling's Formula, we have

$$l(n,k_1,\dots,k_m) = \frac{(1+O(\frac{1}{n}))(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}m^n k_1^{(k_1+\frac{1}{2})}\cdots k_m^{(k_m+\frac{1}{2})}}.$$

Letting $k_i = \lfloor\frac{n}{m}\rfloor + j_i$ for $1\le i\le m$,

$$\begin{aligned} l(n,k_1,\dots,k_m) &= \frac{(1+O(\frac{1}{n}))(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}m^n(\lfloor\frac{n}{m}\rfloor+j_1)^{(\lfloor\frac{n}{m}\rfloor+j_1+\frac{1}{2})}\cdots(\lfloor\frac{n}{m}\rfloor+j_m)^{(\lfloor\frac{n}{m}\rfloor+j_m+\frac{1}{2})}} \\ &= \frac{(1+O(\frac{1}{n}))m^{\frac{m}{2}}}{(2\pi)^{\frac{m-1}{2}}n^{\frac{m-1}{2}}(1+j_1\lfloor\frac{m}{n}\rfloor)^{(\lfloor\frac{n}{m}\rfloor+j_1+\frac{1}{2})}\cdots(1+j_m\lfloor\frac{m}{n}\rfloor)^{(\lfloor\frac{n}{m}\rfloor+j_m+\frac{1}{2})}}. \end{aligned}$$

For all $1\le i\le m$, we set

$$a(n,m,i) := \left(1+j_i\left\lfloor\frac{m}{n}\right\rfloor\right)^{\lfloor\frac{n}{m}\rfloor+j_i+\frac{1}{2}} = e^{(\lfloor\frac{n}{m}\rfloor+j_i+\frac{1}{2})\ln(1+j_i\lfloor\frac{m}{n}\rfloor)}.$$

Using a Taylor series approximation, we have the following for n large enough (as j_i is bounded by $O(\sqrt{n})$):

$$\begin{aligned} a(n,m,i) &= \exp\left(\left(\frac{n}{m}+j_i+\frac{1}{2}\right)\left(j_i\frac{m}{n}-\frac{m^2j_i^2}{2n^2}+\frac{m^3j_i^3}{3n^3}+O\big(n^{-1}\big)\right)\right) \\ &= \exp\left(j_i+\frac{mj_i^2}{2n}+\frac{mj_i}{2n}-\frac{m^2j_i^3}{6n^2}-\frac{m^2j_i^2}{4n^2}+\frac{m^3j_i^4}{3n^3}+\frac{m^3j_i^3}{6n^3}+O\big(n^{-1}\big)\right). \end{aligned}$$

Therefore, we have

$$\left(1+j_1\left\lfloor\frac{m}{n}\right\rfloor\right)^{\lfloor\frac{n}{m}\rfloor+j_1+\frac{1}{2}}\cdots\left(1+j_m\left\lfloor\frac{m}{n}\right\rfloor\right)^{\lfloor\frac{n}{m}\rfloor+j_m+\frac{1}{2}} = e^{-H(n,j_1,\dots,j_m)},$$

and so

$$l(n,j_1,\dots,j_m) = p(n,j_1,\dots,j_m) + O\left(\frac{1}{n}\right),$$

as required. ∎

References

[1] A. C. Berry, *The accuracy of the Gaussian approximation to the sum of independent variates*, Trans. Amer. Math. Soc. **49** (1941), no. 1, 122–136. MR 0003498

[2] L. H. Y. Chen, L. Goldstein and Q. Shao, *Normal approximation by Stein's method*, Springer, Heidelberg, 2011. MR 2732624

[3] R. C. Dalang, *Une démonstration élémentaire du théorème central limite*, Elem. Math. **60** (2005), no. 1, 1–9. MR 2225670

[4] C.-G. Eseen, *A moment inequality with an application to the central limit theorem*, Skand. Aktuarietidskr. **39** (1956), 160–170. MR 0090166

[5] C.-G. Esseen, *On the Liapunoff limit of error in the theoretical probability*, Ark. Mat. Astr. Fys. **28A** (1942), 1–19. MR 0011909

[6] W. Feller, *Uber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung*, Math. Z. **40** (1935), 521–559. MR 1545580

[7] H. Fischer, *A history of the central limit theorem: From classical to modern probability theory*, Springer, New York, 2011. MR 2743162

[8] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, vol. 2, 1972, pp. 586–602. MR 0402873

[9] H. F. Trotter, *An elementary proof of the central limit theorem*, Arch. Math. **10** (1959), 226–234. MR 0108847

[10] Z. Zong and F. Hu, *A new proof of the central limit theorem for i.i.d. random variables*, Abstr. Appl. Anal. **2013** (2013), Article ID 294910. MR 3147846

Vladimir Dobrić, Department of Mathematics, Lehigh University, 14 East Packer Avenue, Bethlehem, PA 18015, USA

E-mail address: vd00@lehigh.edu

Patricia Garmirian, Department of Mathematics, Tufts University, Boston Avenue, Medford, MA 02155, USA

E-mail address: patricia.garmirian@tufts.edu