IEICE TRANS. FUNDAMENTALS, VOL.E102–A, NO.5 MAY 2019

LETTER

A Family of Counterexamples to the Central Limit Theorem Based on Binary Linear Codes

Keigo TAKEUCHI[†a)], ***Member***

SUMMARY The central limit theorem (CLT) claims that the standardized sum of a random sequence converges in distribution to a normal random variable as the length tends to infinity. We prove the existence of a family of counterexamples to the CLT for d-tuplewise independent sequences of length n for all $d = 2, \ldots, n-1$. The proof is based on $[n, k, d+1]$ binary linear codes. Our result implies that d-tuplewise independence is too weak to justify the CLT, even if the size d grows linearly in length n.

key words: *central limit theorem, dependent random variables, counterexamples, binary linear codes*

1. Introduction

Let $\boldsymbol{X} = \{X_i\}_{i=1}^n$ denote a zero-mean and unit-variance random sequence of length $n \in \mathbb{N}$. The central limit theorem (CLT) claims that, under some assumptions of $\boldsymbol{X}$, the sum $S_n = n^{-1/2} \sum_{i=1}^n X_i$ converges in distribution to a standard normal random variable as $n \to \infty$. The CLT is useful in the field of information theory, communications, and signal processing. For example, it provides a foundation for the additive white Gaussian noise (AWGN) channel in information theory, and was utilized to prove the asymptotic convergence property of message-passing algorithms in communications or compressed sensing [1].

Since Etemadi's pioneering proof [2] on the strong law of large numbers (SLLN) under *pairwise* independence, mathematicians have considered the CLT for dependent random sequences, such as martingale difference sequences [3], exchangeable sequences [4], symmetric sequences [5], or stationary and ergodic sequences [6]. Existing CLTs require *global* sufficient conditions over the whole sequence, while the SLLN needs local conditions such as pairwise independence. In fact, local assumptions may be too weak to justify the CLT. Janson [7] and Bradley [8] constructed pairwise independent sequences for which the CLT fails. Their results were generalized to the case of d-tuplewise independence for fixed integers d in [9]. However, it is open whether the CLT holds for the case of $O(n)$-tuplewise independence as the length n tends to infinity.

The purpose of this letter is to present a negative answer to this open problem. We claim that d-tuplewise independence is too weak to justify the CLT, even if d grows linearly in the length n. More precisely, we prove the following:

Manuscript received October 17, 2018.
Manuscript revised January 25, 2019.

[†]The author is with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Toyohashi-shi, 441-8580 Japan.
a) E-mail: takeuchi@ee.tut.ac.jp

DOI: 10.1587/transfun.E102.A.738

Theorem 1: There is a family of counterexamples to the CLT such that $\boldsymbol{X}$ is d-tuplewise independent for all n and $d = 2, \ldots, n-1$.

Theorem 1 implies that it is impossible to prove the CLT only under local assumptions on the sequence $\{X_i\}_{i=1}^n$. We cannot provide a fully explicit construction of counterexamples, since our proof is based on the existence of a family of binary linear codes.

2. Proof of Theorem 1

The proof strategy is as follows: We first construct a random sequence $\boldsymbol{X}$ based on $[n, k, d+1]$ binary linear codes from independent symmetric random variables with unbounded supports. We next classify the moments of $\boldsymbol{X}$ into two groups: non-trivial codewords and the other sequences. The moments are shown to be positive for non-trivial codewords. Otherwise, they are equal to the corresponding moments of the underlying random variables. Finally, we use this classification to prove that a higher-order moment of the sum S_n is different from the corresponding one of the standard normal distribution, and that $\boldsymbol{X}$ is d-tuplewise independent.

Let $\{Y_i\}_{i=1}^n$ denote a sequence of independent symmetric random variables with unit variance, all finite moments, and unbounded supports, i.e. $-Y_i \sim Y_i$, $\mathbb{E}[Y_i^m] < \infty$ for all $m \in \mathbb{N}$, and $\mathbb{P}(|Y_i| \geq a) > 0$ for all $a > 0$. For a binary matrix $\boldsymbol{H} = \{h_{ij}\} \in \{0,1\}^{(n-k)\times n}$ with $k < n$, define $\boldsymbol{X}$ as

$$X_j = |Y_j| \prod_{i=1}^{n-k} \tilde{Y}_i^{h_{ij}}, \tag{1}$$

where $\tilde{Y}_i$ denotes the sign of Y_i, i.e. $\tilde{Y}_i = 1, 0, -1$ for $Y_i > 0$, $Y_i = 0$, and $Y_i < 0$, respectively. By definition, we have $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2] = 1$.

One may regard $\boldsymbol{H}$ as a parity-check matrix on the binary field $\mathbb{F}_2$. Rather, we focus on the set $\mathbb{N}_0$ of non-negative integers. Consider an $[n, k, d]$ linear code defined by $\boldsymbol{H}$ with length n, dimension k, and minimum weight (number of odd elements) d. If $\boldsymbol{Hx}$ has no odd elements, a vector $\boldsymbol{x} \in \mathbb{N}_0^n$ is referred to as a *codeword*. In particular, a codeword is said to be trivial if it has no odd elements. Otherwise, it is said to be non-trivial and has at least d odd elements.

Remark 1: The sequence (1) reduces to that proposed in [9], by selecting a $[d+1, 1, d+1]$ repetition code as $\boldsymbol{H}$ with length $d+1$. However, Pruss [9] investigated another

longer sequence such that the sum S_n for the longer sequence converges in distribution to that for the sequence based on the repetition code with finite d. As a result, the size d of d-tuplewise independence could not be increased as $n \to \infty$.

Lemma 1: Let $\mu(\boldsymbol{m}) = \mathbb{E}[\prod_{j=1}^{n} X_j^{m_j}]$ for a sequence of non-negative integers $\boldsymbol{m} = \{m_j \in \mathbb{N}_0\}_{j=1}^{n}$. Then,

$$\mu(\boldsymbol{m}) = \prod_{j=1}^{n} \mathbb{E}\left[|Y_j|^{m_j}\right] \tag{2}$$

if $\boldsymbol{m}$ is a non-trivial codeword of $\boldsymbol{H}$. Otherwise, $\mu(\boldsymbol{m})$ is equal to the corresponding moment $\tilde{\mu}(\boldsymbol{m}) = \mathbb{E}[\prod_{j=1}^{n} Y_j^{m_j}]$. In particular, $\tilde{\mu}(\boldsymbol{m}) = 0$ holds if $\boldsymbol{m}$ is not a trivial codeword.

Proof: It is straightforward to confirm the last statement. We shall evaluate the moment $\mu(\boldsymbol{m})$. Using $(\prod_i a_i)^{k_0} = \prod_i a_i^{k_0}$ and $\prod_j \tilde{Y}_i^{k_j} = \tilde{Y}_i^{\sum_j k_j}$ for $\{k_j \in \mathbb{N}_0\}_{j=0}^{n}$, from (1) we obtain

$$\begin{aligned}\mu(\boldsymbol{m}) =& \mathbb{E}\left[\prod_{j=1}^{n} |Y_j|^{m_j} \cdot \prod_{j=1}^{n}\prod_{i=1}^{n-k} \tilde{Y}_i^{h_{ij}m_j}\right] \\ =& \prod_{i=1}^{n-k} \mathbb{E}\left[|Y_i|^{m_i}\tilde{Y}_i^{s_i}\right] \prod_{j=n-k+1}^{n} \mathbb{E}\left[|Y_j|^{m_j}\right],\end{aligned} \tag{3}$$

where $s_i = \sum_{j=1}^{n} h_{ij}m_j$ denotes the ith *syndrome*.

From the symmetry of Y_i, we have $\mathbb{E}[|Y_i|^{m_i}\tilde{Y}_i^{s_i}] = 0$ for odd s_i. This implies that if $\boldsymbol{m}$ is not a codeword of $\boldsymbol{H}$, we have $\mu(\boldsymbol{m}) = 0$, which is equal to $\tilde{\mu}(\boldsymbol{m})$. If $\boldsymbol{m}$ is a codeword, $\mu(\boldsymbol{m})$ reduces to (2). In particular, (2) is equal to $\tilde{\mu}(\boldsymbol{m})$ if $\boldsymbol{m}$ is a trivial codeword. Thus, Lemma 1 holds. □

Lemma 2: Suppose that $\boldsymbol{H}$ is a parity-check matrix of an $[n, k, d]$ binary linear code, and consider the sequence $\boldsymbol{X}$ defined in (1). Then, the CLT fails for all $d \le n$.

Proof: Let $\tilde{S}_n = n^{-1/2}\sum_{i=1}^{n} Y_i$. The classical CLT implies that $\tilde{S}_n$ converges in distribution to a standard normal random variable as $n \to \infty$. Thus, it is sufficient to prove that the moment sequence of the sum $S_n = n^{-1/2}\sum_i X_i$ does not coincide with that of $\tilde{S}_n$ for all n and $d \le n$.

We shall evaluate the difference $D_m = |\mathbb{E}[S_n^{2m+d}] - \mathbb{E}[\tilde{S}_n^{2m+d}]|$ for some $m \in \mathbb{N}_0$. By definition, we have

$$\begin{aligned}\mathbb{E}[S_n^{2m+d}] =& \frac{1}{n^{m+d/2}} \sum_{i_1,\ldots,i_{2m+d}} \mathbb{E}[X_{i_1}\cdots X_{i_{2m+d}}] \\ =& \frac{1}{n^{m+d/2}} \sum_{\boldsymbol{m}\in\mathbb{N}_0^n:\sum_j m_j = 2m+d} c(\boldsymbol{m})\mu(\boldsymbol{m}),\end{aligned} \tag{4}$$

where $c(\boldsymbol{m}) \ge 1$ is a coefficient originating from duplication in the summation. From Lemma 1, we find the difference $\mu(\boldsymbol{m}) - \tilde{\mu}(\boldsymbol{m}) = \mu(\boldsymbol{m}) \ge 0$—given by (2)—if $\boldsymbol{m}$ is a non-trivial codeword of $\boldsymbol{H}$. Otherwise, the difference is equal to zero. Thus, we obtain

$$D_m = \frac{1}{n^{m+d/2}} \left| \sum_{\boldsymbol{m}\in\mathbb{N}_0^n:\sum_j m_j = 2m+d} c(\boldsymbol{m})\{\mu(\boldsymbol{m}) - \tilde{\mu}(\boldsymbol{m})\} \right|$$

$$= \frac{1}{n^{m+d/2}} \sum_{\boldsymbol{m}} c(\boldsymbol{m})\mu(\boldsymbol{m}), \tag{5}$$

where the summation is over all possible non-trivial codewords $\boldsymbol{m}$ satisfying $\sum_j m_j = 2m + d$.

In particular, we focus on the non-trivial codeword $\boldsymbol{m}_0$ with $2m+1$, 1, and 0 in the ith elements for $i = 1$, $i = 2, \ldots, d$, and $i > d$, respectively. Without loss of generality, we can assume the existence of the codeword, by rearranging the columns of $\boldsymbol{H}$. Since $c(\boldsymbol{m}) \ge 1$ holds, we obtain

$$D_m > \frac{\mu(\boldsymbol{m}_0)}{n^{m+d/2}} = \frac{\mathbb{E}[|Y_1|^{2m+1}]}{n^{m+d/2}} \prod_{i=2}^{d} \mathbb{E}[|Y_i|]. \tag{6}$$

To complete the proof, we prove that the lower bound (6) tends to infinity as $m \to \infty$. Using the assumption $\mathbb{P}(|Y_1| \ge n) > 0$ for all $n > 1$ yields

$$\begin{aligned}\frac{\mathbb{E}[|Y_1|^{2m+1}]}{n^{m+d/2}} >& \frac{\mathbb{E}[|Y_1|^{2m+1}1(|Y_1| \ge n)]}{n^{m+d/2}} \\ >& n^{m+1-d/2}\mathbb{P}(|Y_1| \ge n) \to \infty\end{aligned} \tag{7}$$

as $m \to \infty$, where $1(\cdot)$ denotes the indicator function. Thus, Lemma 2 holds. □

Proof of Theorem 1: For any $n \ge 1$ and $2 \le d < n$, let $\boldsymbol{H}$ be a parity-check matrix of an $[n, k, d+1]$ linear code with some $1 \le k < n$. The existence of $\boldsymbol{H}$ is guaranteed for any $d = 2, \ldots, n-1$ if the Gilbert-Varshamov (GV) bound [10, p. 33] $\sum_{i=0}^{d-1}\binom{n-1}{i} < 2^{n-k}$ is satisfied. The left-hand side of the GV bound is monotonically increasing with respect to d. Thus, it is sufficient to consider the maximum weight $d = n - 1$. In this case, we have

$$\sum_{i=0}^{n-2}\binom{n-1}{i} = \sum_{i=0}^{n-1}\binom{n-1}{i} - 1 = 2^{n-1} - 1 < 2^{n-k}, \tag{8}$$

with $k = 1$. In other words, the GV bound holds for $d = n-1$ and $k = 1$. Thus, the existence of $\boldsymbol{H}$ is guaranteed.

From Lemma 2, we need to prove that $\boldsymbol{X}$ is d-tuplewise independent. In other words, it is sufficient to prove that $\mu(\boldsymbol{m})$ coincides with $\tilde{\mu}(\boldsymbol{m})$ for all $\boldsymbol{m}$ that have weights smaller than or equal to d. By definition, such a vector $\boldsymbol{m}$ is not a non-trivial codeword of $\boldsymbol{H}$, since any non-trivial codeword has at least weight $d + 1$. From Lemma 1, we find that the coincidence is correct. Thus, Theorem 1 holds. □

Acknowledgements

The author was in part supported by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI Grant Number 18H01441), Japan.

References

[1] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," Proc. 2017 IEEE Int. Symp. Inf. Theory, pp.501–505, Aachen, Germany, June 2017.

[2] N. Etemadi, "An elementary proof of the strong law of large numbers," Z. Wahrscheinlichkeitstheorie verw. Gebiete, vol.55, no.1, pp.119–122, Feb. 1981.
[3] B.M. Brown, "Martingale central limit theorems," Ann. Math. Stat., vol.42, no.1, pp.59–66, 1971.
[4] N.C. Weber, "A martingale approach to central limit theorems for exchangeable random variables," J. Appl. Prob., vol.17, no.3, pp.662–673, Sept. 1980.
[5] D.H. Hong, "A remark on the C.L.T. for sums of pairwise i.i.d. random variables," Math. Japonica, vol.42, no.1, pp.87–89, July 1995.
[6] W.B. Wu and M. Woodroofe, "Martingale approximations for sums of stationary processes," Ann. Probab., vol.32, no.2, pp.1674–1690, April 2004.
[7] S. Janson, "Some pairwise independent sequences for which the central limit theorem fails," Stochastics, vol.23, no.4, pp.439–448, 1988.
[8] R.C. Bradley, "A stationary, pairwise independent, absolutely regular sequence for which the cental limit theorem fails," Probab. Theory Relat. Fields, vol.81, no.1, pp.1–10, Feb. 1989.
[9] A.R. Pruss, "A bounded n-tuplewise independent and identically distributed counterexample to the CLT," Probab. Theory Relat. Fields, vol.111, no.3, pp.323–332, July 1998.
[10] F.J. Macwilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, 1983.