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 \(X = \{X_i\}_{i=1}^n\) 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 \(X\), 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 theory, and was utilized to prove the asymptotic convergence of algorithms in communications theory.

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 independence. In fact, local assumptions may be too weak to justify the CLT, even if the size \(d\) grows linearly in length \(n\). More precisely, we prove the following:

\[ \sum_{i=1}^d X_i = \sum_{i=1}^d Y_i, \]

where \(Y_i\) denotes the sign of \(Y_i\), i.e. \(Y_i = 1, 0, -1\) 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 \(H\) 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 \(H\) with length \(n\), dimension \(k\), and minimum weight (number of odd elements) \(d\). If \(Hx\) has no odd elements, a vector \(x \in \mathbb{F}_2^n\) 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 \(H\) with length \(d + 1\). However, Pruss \([9]\) investigated another

Theorem 1: There is a family of counterexamples to the CLT such that \(X\) 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 \(X\) based on \([n, k, d + 1]\) binary linear codes from independent symmetric random variables with unbounded supports. We next classify the moments of \(X\) 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 \(S_n\) different from the corresponding one of the standard normal distribution, and that \(X\) 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 \(H = \{h_{ij}\} \in \{0, 1\}^{n-k} \times n\) with \(k < n\), define \(X\) as

\[ X_j = |Y_j| \prod_{i=1}^{n-k} \tilde{Y}_{h_{ij}}, \]

\[ (1) \]

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 \(H\) 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 \(H\) with length \(n\), dimension \(k\), and minimum weight (number of odd elements) \(d\). If \(Hx\) has no odd elements, a vector \(x \in \mathbb{F}_2^n\) 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 \(H\) 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 $\mu(m) = \mathbb{E}[\prod_{j=1}^{m} X_j^{m_j}]$ for a sequence of non-negative integers $m = \{m_j \in \mathbb{N}_0\}_{j=1}$. Then,

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

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

**Proof:** It is straightforward to confirm the last statement. We shall evaluate the moment $\mu(m)$. Using $\prod_{i \in [n]} Y_i^{\sum_{k=0}^{i-1} \mu_i} = \prod_{k=0}^{n} Y_i^{\sum_{k=0}^{i-1} \mu_i}$ and $\prod_{j=1}^{n} Y_j^{m_j}$, we have

$$
\mu(m) = \mathbb{E} \left[ \prod_{j=1}^{n} Y_j^{m_j} \right] = \prod_{j=1}^{n} \mathbb{E} \left[ Y_j^{m_j} \right], \tag{3}
$$

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

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

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

**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 $S_n = n^{-1/2} \sum_{i=1}^{n} X_i$ does not coincide with that of $\tilde{S}_n$ for all $n$ and $d \leq n$.

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

$$
\mathbb{E}[S_n^{2m+d}] = \frac{1}{n^{m+d/2}} \sum_{i_1, \ldots, i_{2m+d}} \mathbb{E}[X_{i_1} \cdots X_{i_{2m+d}}]
$$

$$
= \frac{1}{n^{m+d/2}} \sum_{m \in \mathbb{N}_0, \sum_{j=1}^{n} m_j = 2m+d} c(m) \mu(m), \tag{4}
$$

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

$$
D_m = \frac{1}{n^{m+d/2}} \sum_{m \in \mathbb{N}_0, \sum_{j=1}^{n} m_j = 2m+d} c(m) [\mu(m) - \hat{\mu}(m)], \tag{5}
$$

where the summation is over all possible non-trivial codewords $m$ satisfying $\sum_{j=1}^{n} m_j = 2m+d$.

In particular, we focus on the non-trivial codeword $m_0$ with $2m+1$, 1, and 0 in the $i$th elements for $i = 1, 2, \ldots, d$, and $i > d$, respectively. Without loss of generality, we can assume the existence of the codeword, by rearranging the columns of $H$. Since $c(m) \geq 1$ holds, we obtain

$$
D_m > \frac{\mu(m_0)}{n^{m+d/2}} = \frac{\mathbb{E}[|Y_{i_0}|^{2m+1}]}{n^{m+d/2}} \prod_{i=2}^{d} \mathbb{E}[|Y_{i}|]. \tag{6}
$$

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

$$
\mathbb{E}[|Y_{i_0}|^{2m+1}] > \mathbb{E}[|Y_{i_0}|^{2m+1} I(|Y_{i_0}| > n)] \mathbb{E}[|Y_{i}|] \to \infty \tag{7}
$$

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

**Proof of Theorem 1:** For any $n \geq 1$ and $2 \leq d < n$, let $H$ be a parity-check matrix of an $[n, k, d+1]$ linear code with some $1 \leq k < n$. The existence of $H$ is guaranteed for any $d = 2, \ldots, n-1$ if the Gilbert-Varshamov (GV) bound [10, p. 33] $\sum_{i=0}^{d} (n-1)^i < 2^{-k}$ 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

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

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

From Lemma 2, we need to prove that $X$ is $d$-tuplewise independent. In other words, it is sufficient to prove that $\mu(m)$ coincides with $\hat{\mu}(m)$ for all $m$ that have weights smaller than or equal to $d$. By definition, such a vector $m$ is not a non-trivial codeword of $H$, 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**