THE LINDEBERG-FELLER CENTRAL LIMIT THEOREM VIA ZERO BIAS TRANSFORMATION

JAINUL VAGHASIA

Contents

1. Introduction

In 1738, Abraham de Moivre published his book *The Doctrine of Chances* where he provided techniques for solving gambling problems using a version of the Central Limit theorem for Bernoulli trials with $p = 1/2$. Roughly a century later, Pierre Simon Laplace published an extension of this proof, although not a rigorous one, for $p \neq 1/2$ in his work *Theorie Analytique des Probabilities*. A more general statement of the Central Limit Theorem appeared only in 1922 when Lindeberg suggested that the sequence of random variables need not be identically distributed given certain condition known as the Lindeberg condition. Feller has explained that the Lindeberg condition "requires the individual variances be due mainly to masses in an interval whose length is small in comparison to the overall variance".

In his paper [3], Goldstein offers an equivalent, seemingly simpler condition through Stein's zero bias transformation introduced in [4]. As a result of equivalency, this condition can be used to prove the Lindeberg-Feller Central Limit theorem and its partial converse (independently due to Feller and Lévy).

This paper will outline the properties of zero bias transformation, and describe its role in the proof of the Lindeberg-Feller Central Limit Theorem and its Feller-Lévy converse. In light of completeness, we shall also offer an application of the Central Limit theorem using the small zero bias condition to the number

Date: June 3, 2018.

of cycles in a random permutation pooled from all permutations of the set $\{1, 2, \ldots, n\}$, and show that it is asymptotically Normal in n.

2. Notations

This paper uses the following notations:

- We write $X \sim P$ to mean that X has distribution P.
- $Y_n \to_d Y$ means convergence in distribution. $Y_n \to_p Y$ is convergence in probability.
- $\mathbb{E}$ means expected value and Var means variance, while $\mathbb{P}$ is the probability.
- $\mathbf{1}$ denotes the indicator function
- X^* denotes the Random Variable with X-zero biased distribution

3. Background in Probability Theory

We begin by highlighting some elementary facts and definitions that will be used later in this paper.

Let Ω be a sample space with a probability distribution (also called a probability measure) $\mathbb{P}$. A random variable is a map $X : \Omega \to \mathbb{R}$. We write

$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

3.1. Expectation and Variance.

Definition 3.1. (*Expected value*) The expected value of $g(X)$ is

$$\mathbb{E}(g(X)) = \int g(x)\, d\mathbb{P}(x) = \begin{cases} \int_{\infty}^{-\infty} g(x)p(x)\, dx & \text{if } X \text{ is continuous,} \\ \sum_j g(x_j)p(x_j) & \text{if } X \text{ is discrete.} \end{cases}$$

It has following properties:

(1) Linearity: $\mathbb{E}(\sum_{j=1}^{k} c_j g_j(X)) = \sum_{j=1}^{k} c_j \mathbb{E}(g_j(X))$.
(2) If $X_1, \ldots, X_n$ are independent then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_i \mathbb{E}(X_i).$$

Definition 3.2. (*Variance*) Variance is

$$\sigma^2 = \mathrm{Var}\,(X) = \mathbb{E}((X - \mu)^2).$$

It has following properties:

(1) $\mathrm{Var}\,(X) = \mathbb{E}(X^2) - \mu^2$.
(2) If $X_1, \ldots, X_n$ are independent then

$$\mathrm{Var}\left(\sum_{j=1}^{k} a_j X_j\right) = \sum_{j=1}^{k} a_j^2 \mathrm{Var}\,(X_j).$$

3.2. **Convergence.**

Definition 3.3. (*Convergence in Distribution*) A sequence of random variables Y_n is said to converge in distribution to Y if

$$\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) = \mathbb{P}(Y \leq x) \text{ for all continuity points } x \text{ of } \mathbb{P}(Y \leq x).$$

Definition 3.4. (*Convergence in Probability*) A sequence of random variables Y_n converges in probability to Y if

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \geq \epsilon) = 0 \quad \text{for all } \epsilon > 0.$$

We will also need the following equivalence theorems between convergence of expectation and convergence in distribution.

Theorem 3.5. *$Y_n \to_d Y$ implies*

$$\lim_{n\to\infty} \mathbb{E}(h(Y_n)) = \mathbb{E}(h(Y)) \qquad \text{for all } h \in C_b,$$

where C_b is the collection of all bounded, continuous functions.

Since the set of all functions with compact support which integrate to zero and are infinitely differentiable $C^\infty_{c,0} \subset C_b$, Theorem 3.5 holds for $C^\infty_{c,0}$. Following converse holds for $C^\infty_{c,0}$ as well.

Theorem 3.6. *If*

$$\lim_{n\to\infty} \mathbb{E}(h(Y_n)) = \mathbb{E}(h(Y))$$

for all $h \in C^\infty_{c,0}$, then $Y_n \to_d Y$.

The proofs of Theorems 3.5 and 3.6 are too involved to be worth the digression, and therefore shall not be presented here.

4. Zero Bias Transformation

Definition 4.1. Let X be a random variable with mean zero and finite, nonzero variance σ^2. We say that X^* has the $X-$ *zero biased distribution* if for all differentiable f for which these expectations exist,

$$\sigma^2 \mathbb{E}(f'(X^*)) = \mathbb{E}(Xf(X)).$$

In [3], the proof of existence of the zero biased distribution for any such X is omitted for simplicity. We provide a brief Ideas of the argument here. For any given continuous function with compact support g, let $G = \int_0^x g$. Then $Tg = \sigma^{-2}\mathbb{E}(XG(X))$ exists since the variance $\mathbb{E}(X^2) < \infty$. Note that T is linear as inherited by the linearity of expectation. Now take $g \geq 0$, so G is increasing, and therefore X and $G(X)$ are positively correlated. Hence, $Tg = \mathbb{E}(XG(X)) \geq \mathbb{E}(X)\mathbb{E}(G(X)) = 0$. This shows that T is positive. The existence of a unique probability measure ν with $Tg = \int_0^x g\, d\nu$ would then be obtained by Riesz-Markov-Kakutani representation theorem (see eg. [2]; [5] provides a category theoretic proof).

Next, we show that zero bias transformation enjoys the following continuity property. This will be instrumental later in proving the Feller-Lévy converse.

Theorem 4.2. *Let Y and $Y_n, n = 1, 2, \ldots$ be mean zero random variables with finite, nonzero variances $\sigma^2 = Var(Y_n)$, respectively. If*

$$Y_n \to_d Y \qquad \textit{and} \qquad \lim_{n\to\infty} \sigma_n^2 = \sigma^2,$$

then

$$Y_n^* \to_d Y^*.$$

Proof. We shall use the alternate definition of convergence in distribution provided by Theorems 3.5 and 3.6. Let $f \in C_{c,0}^\infty$ and $F(y) = \int_{-\infty}^y f(t)\,dt$. Since Y and Y_n have mean zero and finite variances, their zero bias distributions exists. In particular,

$$\sigma_n^2 \mathbb{E}(f(Y_n^*)) = \mathbb{E}[Y_n F(Y_n)] \quad \text{ for all } n.$$

By Theorem 3.5, since $yF(y)$ is in C_b, we obtain

$$\sigma^2 \lim_{n\to\infty} \mathbb{E}(f(Y_n^*)) = \lim_{n\to\infty} \mathbb{E}[Y_n F(Y_n)] = \mathbb{E}[YF(Y)] = \sigma^2 \mathbb{E}(f(Y^*)).$$

Hence, $\mathbb{E}(f(Y_n^*)) \to \mathbb{E}(f(Y^*))$ for all $f \in C_{c,0}^\infty$, so $Y_n^* \to_d Y^*$ by Theorem 3.6. □

The main utility of the zero bias transforms is that Normal distribution is the unique fixed point of the transformation. This is stated formally below:

Theorem 4.3. *Let X be a zero-meaned random variable with nonzero, finite variance σ^2. Then X has distribution $\mathcal{N}(0, \sigma^2)$ if and only if*

$$\sigma^2 \mathbb{E}(f'(X)) = \mathbb{E}(Xf(X))$$

for all absolutely continuous functions f for which these expectations exist.

Proof. Recall the probability density function for the standard normal distribution:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

We evaluate $\mathbb{E}(f'(X))$ using integration by parts:

$$\begin{aligned}
\mathbb{E}(f'(X)) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(x) e^{-x^2/2}\,dx \\
&= \frac{1}{\sqrt{2\pi}} \left[f(x) e^{-x^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x f(x) e^{-x^2/2}\,dx \right] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x f(x) e^{-x^2/2}\,dx \\
&= \mathbb{E}(Xf(X)),
\end{aligned}$$

as claimed. Since the solution to the differential equation is unique upto adding a constant, we have the result precisely for the family of normal distribution. □

This property can be exploited to reason that a distribution that gets mapped to one nearby is close to being a fixed point of the transformation. Hence, normal approximation can be applied whenever the distribution of a random variable is close to that of its zero bias transformation. With this in hand, consider the distribution of a sum W_n of comparably sized independent random variables. We show that this can be zero biased by randomly selecting a single summand chosen with probability proportional to its variance, and replacing it with comparable size random variable. This would imply that W_n and W_n^* are close, and therefore approximately Normal. This is the gist of the proof of the Lindeberg-Feller Central Limit Theorem under the condition that Goldstein calls 'small zero bias condition'. We shall now formalize this argument.

5. Setting up the conditions

This section serves to list the conditions that go together in some combination into creating the statement of the Lindeberg-Feller Central Limit Theorem.

Condition 5.1. *For every $n = 1, 2, \ldots$, the random variables making up the collection $\mathbf{X}_n = \{X_{i,n} : 1 \leq i \leq n\}$ are independent with mean zero and finite variances $\sigma_{i,n}^2 = Var(X_{i,n})$, standardized so that*

$$W_n = \sum_{i=1}^{n} X_{i,n} \text{ has variance } Var(W_n) = \sum_{i=1}^{n} \sigma_{i,n}^2 = 1.$$

This is a non-negotiable condition because it is necessary with both the Lindeberg condition and the small zero bias condition.

Condition 5.2. *(Lindeberg Condition)*

$$\forall \epsilon > 0 \lim_{n\to\infty} L_{n,\epsilon} = 0 \text{ where } L_{n,\epsilon} = \sum_{i=1}^{n} \mathbb{E}[X_{i,n}^2 \mathbf{1}(|X_{i,n}| \geq \epsilon)].$$

As described in the introduction, this condition forces the random variables $X_{i,n}$ to have individual variances small compared to their sum.

Condition 5.3. *(Small zero bias condition) Given $\mathbf{X}_n$ satisfying condition 5.1, let $\mathbf{X}_n^* = \{X_{i,n}^* : 1 \leq i \leq n\}$ be a collection of random variables so that $X_{i,n}^*$ has the $X_{i,n}$ zero biased distribution and is independent of $\mathbf{X}_n$. Further let I_n be a random index, independent of $\mathbf{X}_n$ and $\mathbf{X}_n^*$, with distribution*

$$\mathbb{P}(I_n = i) = \sigma_{i,n}^2, \tag{5.4}$$

and define

$$X_{I_n,n} = \sum_{i=1}^{n} \mathbf{1}(I_n = i) X_{i,n} \quad \text{and} \quad X_{I_n,n}^* = \sum_{i=1}^{n} \mathbf{1}(I_n = i) X_{i,n}^*. \tag{5.5}$$

The condition requires that

$$X_{I_n,n}^* \to_p 0.$$

Theorem 5.6. *Condition 5.2 and Condition 5.3 are equivalent.*

Proof. Since the random index I_n is independent of $\mathbf{X}_n$ and $\mathbf{X}_n^*$, we use (5.4) and (5.5) to obtain

$$\mathbb{E}(f(X_{I_n,n})) = \sum_{i=1}^{n} \sigma_{i,n}^2 \mathbb{E}(f(X_{i,n})) \text{ and } \mathbb{E}(f(X_{I_n,n}^*)) = \sum_{i=1}^{n} \sigma_{i,n}^2 \mathbb{E}(f(X_{i,n}^*)) \tag{5.7}$$

Showing that 5.2 implies 5.3 relies on the function

$$f(x) = |x-\epsilon|\mathbf{1}(|x| \geq \epsilon) = \begin{cases} x+\epsilon & \text{if } x \leq -\epsilon \\ 0 & \text{if } -\epsilon < x < \epsilon \\ x-\epsilon & \text{if } x \geq \epsilon \end{cases}$$

Notice that

$$f'(x) = \mathbf{1}(|x| \geq \epsilon) \qquad a.e.$$

Then, using $\mathbb{E}(\mathbf{1}(x)) = \mathbb{P}(x)$ for the first equality and zero bias relation for the second equality, we obtain

$$\sigma_{i,n}^2 \mathbb{P}(|X_{i,n}^*| \geq \epsilon) = \sigma_{i,n}^2 \mathbb{E}(f'(X_{i,n}^*)) = \mathbb{E}\Big[(X_{i,n}^2 - \epsilon|X_{i,n}|)\mathbf{1}(|X_{i,n}| \geq \epsilon)\Big] \tag{5.8}$$

Bounding this quantity from above,

$$\begin{aligned} \sigma_{i,n}^2 \mathbb{P}(|X_{i,n}^*| \geq \epsilon) &\leq \mathbb{E}\Big[(X_{i,n}^2 + \epsilon|X_{i,n}|)\mathbf{1}(|X_{i,n}| \geq \epsilon)\Big] \\ &\leq 2\mathbb{E}\Big[X_{i,n}^2 \mathbf{1}(|X_{i,n}| \geq \epsilon)\Big]. \end{aligned}$$

Applying (5.7) on the indicator function $\mathbf{1}(|x| \geq \epsilon)$ and using $\mathbb{E}(\mathbf{1}(x)) = \mathbb{P}(x)$, we obtain

$$\mathbb{P}(|X_{I_n,n}^*| \geq \epsilon) = \sum_{i=1}^n \sigma_{i,n}^2 \mathbb{P}(|X_{i,n}^*| \geq \epsilon)$$

Using the bound derived above,

$$\mathbb{P}(|X_{I_n,n}^*| \geq \epsilon) \leq 2L_{n,\epsilon}.$$

Hence, small bias condition is satisfied given Lindeberg condition.

For the reverse implication, observe that

$$x^2\mathbf{1}(|x| \geq \epsilon) \leq 2\left(x^2 - \frac{\epsilon}{2}|x|\right)\mathbf{1}\left(|x| \geq \frac{\epsilon}{2}\right).$$

Then

$$\begin{aligned} L_{n,\epsilon} = \sum_{i=1}^n \mathbb{E}(X_{i,n}^2 \mathbf{1}(|X_{i,n}| \geq \epsilon)) &\leq 2\sum_{i=1}^n \left(X_{i,n}^2 - \frac{\epsilon}{2}|X_{i,n}|\right)\mathbf{1}\left(|X_{i,n}| \geq \frac{\epsilon}{2}\right) \\ &= 2\sum_{i=1}^n \sigma_{i,n}^2 \mathbb{P}\left(|X_{i,n}^*| \geq \frac{\epsilon}{2}\right) \\ &= 2\mathbb{P}\left(|X_{i,n}^*| \geq \frac{\epsilon}{2}\right), \end{aligned}$$

where we used (5.8) and (5.5) for second and third equalities, respectively. This finishes the equivalence argument. □

Now that we know that the two conditions are equivalent, it is worthwhile to see how small bias condition can be used to prove the Central Limit theorems.

6. The Lindeberg-Feller Central Limit Theorem

Lemma 6.1. *Let* $\mathbf{X}_n, n = 1, 2, \ldots$ *satisfy Condition 5.1 and* $m_n = \max_{1\le i\le n} \sigma^2_{i,n}$. *Then*

$$X_{I_n,n} \to_p 0 \quad \textit{whenever} \quad \lim_{n\to\infty} m_n = 0.$$

Proof. Using (5.7) with $f(x) = x$, we obtain $\mathbb{E}(X_{I_n,n}) = 0$, and hence $\mathrm{Var}\,(X_{I_n,n}) = \mathbb{E}(X^2_{I_n,n})$. Again with $f(x) = x^2$,

$$\mathrm{Var}\,(X_{I_n,n}) = \sum_{i=1}^{n} \sigma^4_{i,n}.$$

Since $\sigma^4_{i,n} \le \sigma^2_{i,n} \max_{1\le j\le n} \sigma^2_{j,n} = \sigma^2_{i,n} m_n$, we have for all $\epsilon > 0$

$$\mathbb{P}(|X_{I_n,n} \ge \epsilon) \le \frac{\mathrm{Var}\,(X_{I_n,n})}{\epsilon^2} \le \frac{1}{\epsilon^2} m_n \sum_{i=1}^{n} \sigma^2_{i,n} = \frac{1}{\epsilon} m_n,$$

where we used the Chebyshev's inequality. □

Lemma 6.2. *If* $\mathbf{X}_n, n = 1, 2, \ldots$ *satisfies Condition 5.1 and the small bias condition 5.3, then*

$$X_{I_n,n} \to_p 0.$$

Proof. For all n, $1 \le i \le n$, and $\epsilon > 0$,

$$\sigma^2_{i,n} = \mathbb{E}(X^2_{i,n}\mathbf{1}(|X_{i,n}| < \epsilon)) + \mathbb{E}(X^2_{i,n}\mathbf{1}(|X_{i,n}| \ge \epsilon)) \le \epsilon^2 + L_{n,\epsilon}.$$

It follows that

$$m_n \le \epsilon^2 + L_{n,\epsilon}, \quad \text{and therefore} \quad \limsup_{n\to\infty} m_n \le \epsilon^2,$$

where we used $L_{n,\epsilon} \to 0$ by Theorem 5.6. The claim now follows by Lemma 6.1. □

Theorem 6.3. *If* $\mathbf{X}_n, = 1, 2, \ldots$ *satisfies Condition 5.3, then* $W_n \to_d Z$, *a standard normal random variable*

Proof. (*Ideas*) Let $h \in C^{\infty}_{c,0}$. Stein has shown that there exists a twice differentiable solution f with bounded derivatives to the 'Stein equation':

$$f'(w) - wf(w) = h(w) - \mathbb{E}(h(Z))$$

Then using zero bias relation,

$$\mathbb{E}[h(W_n) - \mathbb{E}(h(Z))] = \mathbb{E}[f'(W_n) - f'(W^*_n)]$$

where $W^*_n = W_n + X^*_{I_n,n} - X_{I_n,n}$. Since f', f'' are bounded, we have $\eta(\delta) = \sup_{|x-y|\le\delta} |f'(x) - f'(y)| \to 0$ as $\delta \to 0$. By the small bias condition, $W^*_n - W_n \to 0$. Hence, $\eta(|W^*_n - W_n|) \to 0$. By the triangle inequality, $\mathbb{E}[h(W_n) - \mathbb{E}(h(Z))] \le \mathbb{E}(\eta(|W^*_n - W_n|))$. This shall give us $\lim_{n\to\infty} \mathbb{E}(h(W_n)) = \mathbb{E}(h(Z))$. The result would then follow by Theorem 3.6. □

7. The Feller-Lévy Partial Converse

Lemma 7.1. *(Slutsky's Lemma) Let U_n and $V_n, n = 1, 2, \ldots$ be two sequences of random variables. Then*

$$U_n \to_d U \text{ and } V_n \to_p 0 \quad \text{implies} \quad U_n + V_n \to_d U.$$

The proof of Slutsky's Lemma is an exercise in applying definitions and is left to the reader. When independence holds, we have the following converse.

Lemma 7.2. *Let U_n and $V_n, n = 1, 2, \ldots$ be two sequences of random variables such that U_n and V_n are independent for every n. Then*

$$U_n \to_d U \text{ and } U_n + V_n \to_d U \quad \text{implies} \quad V_n \to_p 0.$$

Proof. (Ideas)
Assume the special case that $U_n =_d U$ for all n. Then assume for contradiction that V_n does not tend to zero in probability. Without loss of generality, there exists an infinite subsequence $\mathbf{K}$ such that for all $n \in \mathbf{K}$,

$$\mathbb{P}(V_n \geq \epsilon) \geq p.$$

Use density of U to obtain continuous $s(x) = \mathbb{P}(x \leq U \leq x + 1)$. Hence, $s(x)$ can be showed to attain its maximum value in a bounded region. Then using the independence of U and V_n, we have

$$\mathbb{P}(y \leq U + V_n \leq y + 1 | V_n) = s(y - V_n) \quad \text{for all } n.$$

On one hand,

$$\mathbb{P}(y \leq U + V_n \leq y + 1 | V_n \geq \epsilon) \leq \sup_{x \leq y - \epsilon} s(x) \quad \text{for all } n \in \mathbf{K},$$

but on the other hand by conditioning on $V_n \geq \epsilon$ and its complement, and the fact that U is absolutely continuous, we obtain a contradiction

$$\liminf_{n\to\infty} \mathbb{P}(y \leq U + V_n \leq y + 1) < \lim_{n\to\infty} \mathbb{P}(y \leq U + V_n \leq y + 1).$$

Generalizing the situation requires more sophisticated constructs for which we refer to the Appendix of [3]. □

Theorem 7.3. *If $\mathbf{X}_n, = 1, 2, \ldots$ satisfies Condition 5.1 and*

$$\lim_{n\to\infty} m_n = 0, \qquad m_n = \max_{1 \leq i \leq n} \sigma^2_{i,n},$$

then the small zero bias condition is necessary for $W_n \to_d Z$.

Proof. Since $W_n \to_d Z$ and $\operatorname{Var}(W_n) \to \operatorname{Var}(Z)$ with limit identically one, Theorem 4.2 implies $W_n^* \to Z$. Note that the limit is technically Z^* but $Z^* = Z$ for the fixed point of zero bias transformation.

Since $m_n \to 0$, Lemma 6.1 yields that $X_{I_n,n} \to_p 0$. It follows from Slutsky's Lemma 7.1 that

$$W_n + X^*_{I_n,n} = W^*_n + X_{I_n,n} \to_d Z.$$

Hence,

$$W_n \to_d Z \quad \text{and} \quad W_n + X^*_{I_n,n} \to_d Z.$$

Since W_n is a function of $\mathbf{X}_n$, which is independent of I_n and X^*_n and therefore $X^*_{I_n,n}$, Lemma 7.2 yields that $X^*_{I_n,n} \to_p 0$. □

8. Cycles in a Random Permutation: Application of Lindeberg-Feller CLT

Random permutations play an important role in many areas of probability and statistics. Here we provide one application that uses the cycles in random permutations. Consider a party with n people, and everyone writes their name down on a piece of paper and puts it into a bag. The bag is mixed up, and each person draws one piece of paper. If you draw the name of someone else, you are considered to be in his “group”. We want to show that the distribution of total number of “groups” is asymptotically Normal. This construct is an extension of that of the classical hat-check problem where each “group” is only allowed at most two people, and the motive is to find the distribution of the number of singleton “groups”.

Let us formalize this problem now. Define the symmetric group $\mathcal{S}_n$ to be the set of all permutations π on the set $\{1, 2, \ldots, n\}$. We can represent the permutation using the cycle notation. For example, $\pi \in \mathcal{S}_7$ may be represented as

$$\pi = (1, 4, 6)(2, 3, 7)(5)$$

with the meaning that π maps 1 to 4, 4 to 6, 6 to 1, and so forth. Notice that π has two cycles of length 3 and one of length 1, for a total of three cycles. Hence, in the problem described above, each cycle would represent a single “group” of people. We will denote $C_{r,n}(\pi)$ to be the number of cycles of length r in permutation $\pi \in \mathcal{S}_n$. Define $K_n(\pi)$ to be the total number of cycles in permutation $\pi \in \mathcal{S}_n$.

To begin, we show that $C_{1,n}$ converges in distribution to a random variable having the Poisson distribution with mean 1. This should solve the simplest of the hat-check problems.

Proposition 8.1. *For $k = 0, 1, \ldots, n$,*

$$\mathbb{P}(C_{1,n} = k) \to_d \frac{e^{-1}}{k!} \text{ as } n \to \infty$$

Proof. (Ideas)

Using an inclusion-exclusion argument, we can show an elementary equality that

$$\mathbb{P}(C_{1,n} = k) = \frac{1}{k!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}.$$

Then the proposition holds true as we pass to limit $n \to \infty$. □

Now, we handle the extension of the problem. That is, we show that $K_n(\pi)$ is asymptotically Normal. We will employ Feller coupling [1], which constructs a random permutation π uniformly from $\mathcal{S}_n$ with help of n independent Bernoulli variables $X_{1:n}$ with distributions

$$\mathbb{P}(X_i = 0) = 1 - \frac{1}{i} \quad \text{and} \quad \mathbb{P}(X_i = 1) = \frac{1}{i}, \quad i = 1, \ldots, n. \tag{8.2}$$

Begin the first cycle at step 1 with the element 1. At step $i, i = 1, \ldots, n$, if $X_{n-i+1} = 1$ close the current cycle and begin a new one starting with the smallest number not yet in any cycle, and otherwise choose an element uniformly from those yet unused. In this way, at step i, we complete a cycle with probability $1/(n - i + 1)$.

As the total number $K_n(\pi)$ of cycles of π is precisely the number of times an element closes the loop upon completing its cycle,

$$K_n(\pi) = \sum_{i=1}^{n} X_i.$$

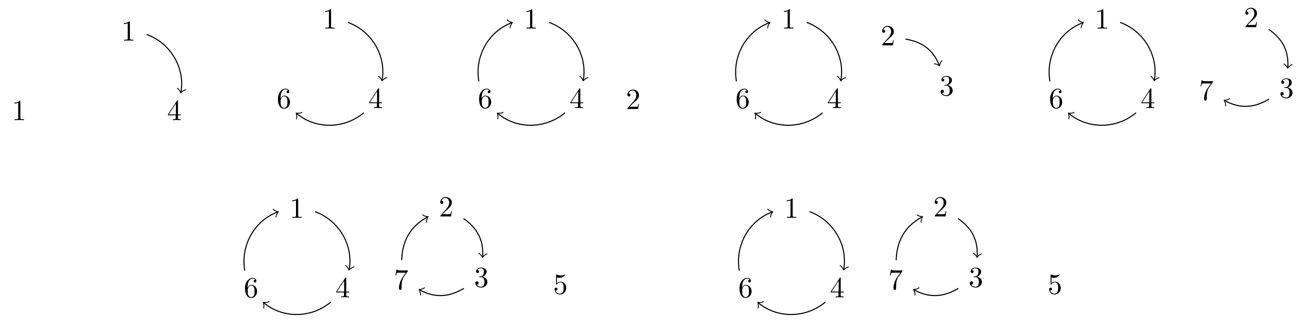

Figure: An example of Feller Coupling for $\pi = (1,4,6)(2,3,7)(5)$ with $X_1 = X_2 = X_5 = 1$.

An important consequence of this definition is that random variables $X_i, i = 1,\dots,n$, even though independent, are not identically distributed. As a result, the classical Central Limit does not apply to their sum. However, we show that the small zero bias conditions hold. This gives us the required asymptotic Normality of $K_\pi(n)$.

First, we standardize $K_n(\pi)$ in order to satisfy Condition (5.1). Since Bernoulli variables with defining probability p have expected value p and variance $p(1-p)$, we use the linearity of expectation and that of variance for independent variables to obtain that

$$\mathbb{E}(K_n(\pi)) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} \frac{1}{i} = h_n \quad \text{and} \quad \mathrm{Var}\,(K_n(\pi)) = \sigma_n^2 = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i^2}\right). \tag{8.3}$$

Then the standardized variable is defined as

$$W_n = \frac{K_n(\pi) - h_n}{\sigma_n} = \sum_{i=1}^{n} X_{i,n} \quad \text{where} \quad X_{i,n} = \frac{X_i - i^{-1}}{\sigma_n}. \tag{8.4}$$

It is trivial to verify that W_n satisfies Condition (5.1).

Now we verify the small zero bias condition. Notice that

$$\lim_{n\to\infty} \frac{h_n}{\log n} = 1, \text{ and hence } \lim_{n\to\infty} \frac{\sigma_n^2}{\log n} = 1. \tag{8.5}$$

As a consequence of (8.2) and (8.4), we have $W_n = \sum_{i=2}^{n} X_{i,n}$ as $X_1 = 1$ identically makes $X_{1,n} = 0$ for all n.

X_i^* should be uniformly distributed: Let B be a Bernoulli random variable with success probability p and let X be the centered Bernoulli variable $B - p$, having variance $p(1-p)$. Then

$$\begin{aligned}
\mathbb{E}(Xf(X)) &= \mathbb{E}((B-p)f(B-p)) \\
&= p(1-p)f(1-p) - (1-p)pf(-p) \\
&= \sigma^2[f(1-p) - f(-p)] \\
&= \sigma^2 \int_{-p}^{1-p} f'(u)\,du \\
&= \sigma^2 \mathbb{E}(f'(U)),
\end{aligned}$$

for U having uniform density over $[-p, 1-p]$. Thus, by

$$(\alpha X)^* =_d \alpha X^* \quad \text{for all } \alpha \neq 0,$$

and (8.4), we have

$$X_{i,n}^* =_d \frac{U_i}{\sigma_n}, \text{ where } U_i \text{ has distribution } \mathcal{U}\left[-\frac{1}{i}, 1-\frac{1}{i}\right], i = 2, \ldots, n.$$

It follows that $|U_i| \leq 1$ for all $i = 2, \ldots, n$, and therefore for uniformly random index I_n

$$|X_{I_n,n}^*| \leq \frac{1}{\sigma_n} \to 0$$

by (8.5). Hence, small bias condition is satisfied and the Lindeberg-Feller Central Limit Theorem gives us the result we want.

References

[1] W. Feller. The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51** (1945) 800–832.

[2] G. Folland. Real Analysis: Modern Techniques and Their Applications. John Wiley and Sons (1984).

[3] L. Goldstein. A Probabilistic Proof of the Lindeberg-Feller Central Limit Theorem. The American Mathematical Monthly. Vol. 116, No. 1 (Jan., 2009) 45–60.

[4] L. Goldstein and G. Rienhart. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** (1997) 935–952.

[5] D. G. Hartig. The Riesz Representation Theorem Revisited. The American Mathematical Monthly. Vol. 90, No. 4 (Apr., 1983) 277–280.