# A Combinatorial Approach to The Free Central Limit Theorem

Jeremy Li, mentored by Quinn Kolt

Department of Mathematics, University of California Santa Barbara (DRP 2024)

## Abstract

Free probability theory is a relatively new area of study, bringing together many different fields of mathematics, such as operator algebras, random matrices, combinatorics, and representation theory of symmetric groups. In this poster, we will explore the basic groundwork for free probability and provide a combinatorial proof of the free central limit theorem.

## Background

### $*$-Probability Space $(\mathcal{A}, \varphi)$

A $*$-probability space consists of a unital $*$-algebra $\mathcal{A}$ and an expectation $\varphi : \mathcal{A} \to \mathbb{C}$.

A **unital $*$-algebra** $\mathcal{A}$ is a vector space over $\mathbb{C}$ equipped multiplication that is associative but not necessarily commutative, a multiplicative identity $1_{\mathcal{A}}$, and an adjoint operation $*$ (e.g., adjoint of linear maps, conjugate transpose of matrices).

- Elements $a \in \mathcal{A}$ are called **(non-commutative) random variables**.
- $\mathcal{A}_0$ is called a **subalgebra** of $\mathcal{A}$ if $\mathcal{A}_0 \subseteq \mathcal{A}$ and it is itself a unital $*$-algebra.
- The expectation $\varphi$ is a **linear functional** with $\varphi(1_{\mathcal{A}}) = 1$ and $\varphi(a^*a) \geq 0, \forall a \in \mathcal{A}$.

### Classical Independence

Let $(\mathcal{A}, \varphi)$ be a $*$-probability space. Elements $a, b \in \mathcal{A}$ are **(classically) independent** if $ab = ba$ and $\varphi(a^n b^m) = \varphi(a^n)\varphi(b^m)$ for all $n, m \in \mathbb{N}$.

### Free Independence

Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $I$ be a fixed index set. For each $i \in I$, let $\mathcal{A}_i \subseteq \mathcal{A}$ be a subalgebra. The subalgebras $(\mathcal{A}_i)_{i \in I}$ are **freely independent (FI)** if

$$\varphi(a_1 a_2 \dots a_k) = 0$$

whenever we have the following:

- $a_j \in \mathcal{A}_{i(j)}$, with $i(j) \in I$, for all $j = 1, 2, \dots, k$ and $k \in \mathbb{N}$.
- $\varphi(a_j) = 0$ for all $j = 1, 2, \dots, k$.
- Neighboring elements are from different subalgebras, i.e., $i(1) \neq i(2)$, ..., $i(k-1) \neq i(k)$.

Elements $a, b \in \mathcal{A}$ are **freely independent** if the subalgebras they generate are FI.

### Convergence in Distribution

Let $(\mathcal{A}_N, \varphi_N)_{N \in \mathbb{N}}$ and $(\mathcal{A}, \varphi)$ be $*$-probability spaces. Consider $a_N \in \mathcal{A}_N$ for each $N \in \mathbb{N}$ and $a \in \mathcal{A}$. We say that $a_N$ **converges in distribution** to $a$ as $N \to \infty$ and denote this by $a_N \xrightarrow{\text{distr}} a$, if we have convergence of all moments: $\lim_{N\to\infty} \varphi_N(a_N^n) = \varphi(a^n)$ for all $n \in \mathbb{N}$

## The Classical Central Limit Theorem

**Theorem 1. (Classical Central Limit Theorem)** Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \dots \in \mathcal{A}$ be a sequence of independent and identically distributed self-adjoint random variables. Assume all random variables are centered: $\varphi(a_r) = 0, \forall r \in \mathbb{N}$ and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance of the random variables. Then, we have

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{\text{distr}} x,$$

where $x$ is a normal random variable with mean 0 and variance $\sigma^2$.

**Remark.** This statement means explicitly

$$\lim_{N\to\infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} t^n e^{-\frac{t^2}{2\sigma^2}} dt, \qquad \forall n \in \mathbb{N}.$$

## The Free Central Limit Theorem

**Theorem 2. (Free Central Limit Theorem)** Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \dots \in \mathcal{A}$ be a sequence of **freely independent and identically distributed (FIID)** self-adjoint random variables. Assume all random variables are centered: $\varphi(a_r) = 0, \forall r \in \mathbb{N}$ and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance of the random variables. Then, we have

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{\text{distr}} s,$$

where $s$ is a semicircular random variable with variance $\sigma^2$.

**Remark.** This statement means explicitly, $\forall n \in \mathbb{N}$ and $\sigma := \sqrt{\sigma^2}$,

$$\lim_{N\to\infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \int_{-2\sigma}^{2\sigma} \frac{t^n}{2\pi\sigma^2} \sqrt{4\sigma^2 - t^2}\, dt = \begin{cases} \frac{\sigma^{2k}}{k+1}\binom{2k}{k}, & \text{if } n = 2k \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

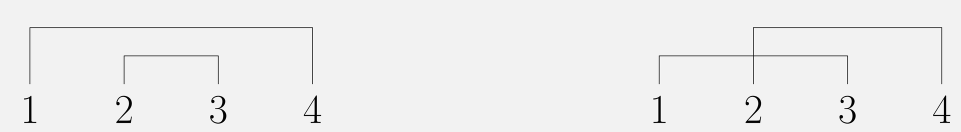
We provide a combinatorial proof of **Theorem 2 (Free CLT)** using **Lemmas 1-3**.

**Lemma 1.** Under the condition of **Theorem 2**,

$$\lim_{N\to\infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = D_n \sigma^n,$$

where $D_n := \left|\{\pi : \pi \text{ non-crossing pair partition of } \{1, \dots, n\}\}\right|$.

**Figure 1.** Non-crossing / crossing pair partitions of $\{1, 2, 3, 4\}$



a **non-crossing** pair partition a **crossing** pair partition

**Proof of Lemma 1.** By definition of FIID, we have for large $N$,

$$\varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \sum_{1 \leq r(1), \dots, r(n) \leq N} N^{-n/2} \varphi(a_{r(1)} \cdots a_{r(n)}) \approx \sum_{\substack{\pi \text{ partition} \\ \text{of } \{1,\dots,n\}}} N^{|\pi| - n/2} \kappa_\pi$$

where $\pi$ has index tuple $(r(1), \dots, r(n))$ and $\kappa_\pi := \varphi(a_{r(1)} \cdots a_{r(n)})$.

If $\pi$ contains a subset with only one element, then the resulting $\kappa_\pi$ must be zero, since the $\varphi(a_{r(m)}) = 0, \forall a_{r(m)}$. Moreover, if $\pi$ contains a subset with at least three elements, then $|\pi| < n/2$, so $N^{|\pi| - n/2}$ vanishes as $N \to \infty$. Therefore, we have

$$\lim_{N\to\infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \sum_{\substack{\pi \text{ pair partition} \\ \text{of } \{1,\dots,n\}}} \kappa_\pi$$

**Case 1.** Consider the case when all consecutive indices are different: $r(1) \neq \ldots \neq r(n)$. Since $\varphi(a_{r(m)}) = 0$ for all $m = 1, \dots, n$, we have by the definition of free independence

$$\kappa_\pi = \varphi(a_{r(1)} \cdots a_{r(n)}) = 0$$

**Case 2.** Consider the case when two consecutive indices coincide, i.e., $r(m) = r(m+1)$, for some $m = 1, \dots, n-1$. Since $a_{r(m)} a_{r(m+1)}$ is freely independent from the subalgebra generated by $\{a_{r(1)}, \dots, a_{r(m-1)}, a_{r(m+2)}, \dots, a_{r(n)}\}$, we have the following factorization

$$\kappa_\pi = \varphi(a_{r(1)} \cdots a_{r(m-1)} a_{r(m+2)} \cdots a_{r(n)}) \cdot \cancelto{\sigma^2}{\varphi(a_{r(m)} a_{r(m+1)})}$$

We repeat this iteration until we either get zero in one of the steps or arrive at the moment $\varphi(1_{\mathcal{A}}) = 1$, in which the corresponding partition will give a contribution of $\sigma^n$. This occurs precisely when $\pi$ is a non-crossing pair partition of $\{1, \dots, n\}$. $\square$

## Non-crossing Pair Partitions and Dyck Paths

The $n$-th **Catalan number** $C_n$ for $n \geq 0$ is given by

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}$$

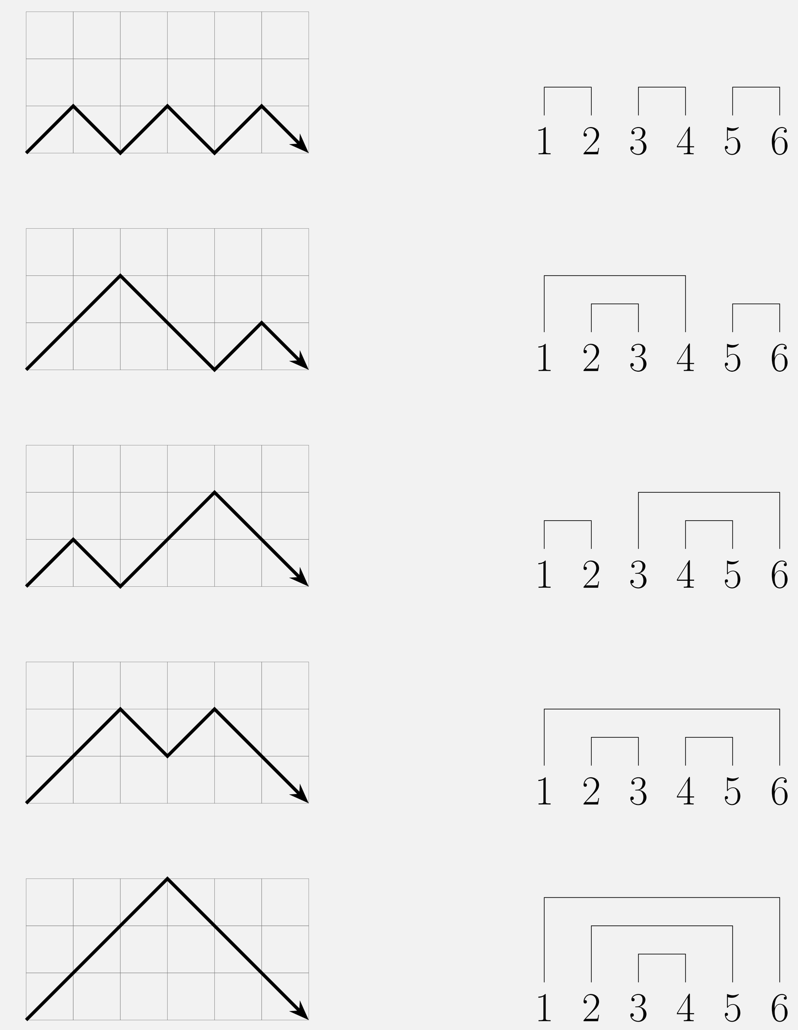
An equivalent representation of the Catalan numbers is the following recurrence relation

$$C_0 = C_1 = 1, \quad C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}, \ n \geq 2$$

We define **Dyck paths** in the lattice $\mathbb{Z}^2$ to be walks from $(0, 0)$ to $(n, 0)$, for $n \in \mathbb{N}$ even, with steps either of the form $(+1, +1)$ or $(+1, -1)$, keeping $y$-coordinate nonnegative.

**Lemma 2.** There is a **one-to-one correspondence (bijection)** between Dyck paths on $2n$ steps and non-crossing pair partition of $\{1, 2, ..., 2n\}$.

**Figure 2.** Dyck paths on 6 steps / non-crossing pair partition of $\{1, 2, 3, 4, 5, 6\}$



**Bijection** between Dyck paths and non-crossing pair partitions

**Lemma 3.** $C_n = D_{2n}$, the number of Dyck paths on $2n$ steps.

**Proof of Lemma 3.** Consider a Dyck path on $2n$ steps and suppose it first hits $y = 0$ at the $2k$-th step with $1 \leq k \leq n$. The path has to move $(+1, +1)$ (go up) on the first step, and it must move $(+1, -1)$ (go down) at the $2k$-th step. The middle $2(k-1)$ steps are arbitrary as long as $y \geq 0$, so the first $2k$ steps have $D_{2(k-1)}$ combinations. The remaining $2(n-k)$ steps also have no restriction and have $D_{2(n-k)}$ combinations. Thus, this case (first hits at $2k$) contributes $D_{2(k-1)} D_{2(n-k)}$ to $D_{2n}$. By induction, we have

$$D_{2n} = \sum_{k=1}^{n} D_{2(k-1)} D_{2(n-k)} = \sum_{k=1}^{n} C_{k-1} C_{n-k} = C_n \quad (\text{with the convention } D_0 = 1).$$

This shows Catalan numbers count the number of Dyck paths. Together with **Lemma 2**, we proved that Catalan numbers count the number of non-crossing pair partitions. $\square$

## Reference

A. Nica, R. Speicher, Lectures on the Combinatorics of Free Probability Theory