Elect. Comm. in Probab. **12** (2007), 36–50

ELECTRONIC COMMUNICATIONS in PROBABILITY

A PROOF OF A NON-COMMUTATIVE CENTRAL LIMIT THEOREM BY THE LINDEBERG METHOD

VLADISLAV KARGIN
Courant Institute of Mathematical Sciences; 109-20 71st Road, Apt. 4A, Forest Hills NY 11375
email: kargin@cims.nyu.edu

Submitted February 16, 2006, *accepted in final form* February 13, 2007

AMS 2000 Subject classification: 46L54, 60F05
Keywords: central limit theorem, Lindeberg method, free probability, free convolution, free independence

Abstract
A Central Limit Theorem for non-commutative random variables is proved using the Lindeberg method. The theorem is a generalization of the Central Limit Theorem for free random variables proved by Voiculescu. The Central Limit Theorem in this paper relies on an assumption which is weaker than freeness.

1 Introduction

One of the most important results in free probability theory is the Central Limit Theorem (CLT) for free random variables ([11]). It was proved almost simultaneously with the invention of free probability theory. Later conditions of the theorem were relaxed ([10]). Moreover, a far-reaching generalization was achieved in [1], which studied domains of attraction of probability laws with respect to free additive convolutions. See also [2].

Freeness is a very strong condition imposed on operators and it is of interest to find out whether the Central Limit Theorem continues to hold if this condition is somewhat relaxed. This problem calls for a different proof of the non-commutative CLT which does not depend on R-transforms or on the vanishing of mixed free cumulants, because both of these techniques are closely connected with the concept of freeness.

In this paper we give a proof of free CLT that avoids using either R-transforms or free cumulants. This allows us to develop a generalization of the free CLT to random variables that are not necessarily free but that satisfy a weaker assumption. An example shows that this assumption is strictly weaker than the assumption of freeness.

The proof that we use is a modification of the Lindeberg proof of the classical CLT ([6]). The main difference is that we use polynomials instead of arbitrary functions from $C_c^3(\mathbb{R})$, and that more ingenuity is required to estimate the residual terms in the Taylor expansion formula.

The closest result to the result in this paper is Theorem 2.1 in ([12]), where the Central Limit Theorem is proved under the conditions on summands that are weaker than the requirement

of freeness. The conditions that we use are somewhat different than those in Voiculescu's paper. In addition, we give an explicit example of variables that are not free but that satisfy conditions of the theorem.
The rest of the paper is organized as follows. Section 2 provides background material and formulates the main result. Section 3 shows by an example that a condition in the main result is strictly weaker than the condition of freeness. Section 4 contains the proof of the main result. And Section 5 concludes.

2 Background and Main Theorem

Before proceeding further, let us establish the background. A *non-commutative random space* $(\mathcal{A}, E)$ is a pair of an operator algebra $\mathcal{A}$ and a linear functional E on $\mathcal{A}$. It is assumed that $\mathcal{A}$ is closed relative to taking the adjoints and contains a unit, and that E is
1) positive, i.e., $E(X^*X) \geq 0$ for every $X \in \mathcal{A}$,
2) finite, i.e., $E(I) = 1$ where I denotes the unit operator, and
3) tracial, i.e., $E(X_1X_2) = E(X_2X_1)$ for every X_1 and $X_2 \in \mathcal{A}$.
This linear functional is called *expectation.* Elements of $\mathcal{A}$ are called *random variables.*
Let X be a self-adjoint random variable (i.e., a self-adjoint operator from algebra $\mathcal{A}$). We can write X as an integral over a resolution of identity:

$$X = \int_{-\infty}^{\infty} \lambda dP_X(\lambda),$$

where $P_X(\lambda)$ is an increasing family of commuting projectors. Then we can define the *spectral probability measure* of interval $(a, b]$ as follows:

$$\mu_X\{(a, b]\} = E[P_X(b) - P_X(a)].$$

We can extend this measure to all measurable subsets in the usual way. We will call μ_X the *spectral probability measure* of random variable X, or simply its *spectral measure.*
We can calculate the expectation of any summable function of a self-adjoint variable X by using its spectral measure:

$$Ef(X) = \int_{-\infty}^{\infty} f(\lambda)\, d\mu_X(\lambda).$$

In particular, the *moments* of the probability measure μ_X equal the expectation values of the powers of X:

$$\int_{-\infty}^{\infty} \lambda^k d\mu_X(\lambda) = E\left(X^k\right).$$

Let us now recall the definition of freeness. Consider sub-algebras $\mathcal{A}_1, ..., \mathcal{A}_n$. Let a_i denote elements of these sub-algebras and let $k(i)$ be a function that maps the index of an element to the index of the corresponding algebra: $a_i \in \mathcal{A}_{k(i)}$.

Definition 1. *The algebras $\mathcal{A}_1, ..., \mathcal{A}_n$ (and their elements) are* free *if $E(a_1...a_m) = 0$ whenever the following two conditions hold:*
(a) $E(a_i) = 0$ for every i, and
(b) $k(i) \neq k(i+1)$ for every $i < m$.
The variables $X_1, ..., X_n$ are called free if the algebras $\mathcal{A}_i$ generated by $\{I, X_i, X_i^*\}$, respectively, are free.

An important property of freeness is that we can compute the moments of the products of the free random variables.

Proposition 2. *Suppose $X_1, ..., X_n$ are free. Then*

$$E\left(X_1...X_n\right) = \sum_{r=1}^{n} \sum_{1 \le k_1 < ... < k_r \le n} (-1)^{r-1} E\left(X_{k_1}\right) ... E\left(X_{k_r}\right) E\left(X_1...\widehat{X}_{k_1}...\widehat{X}_{k_r}...X_n\right), \quad (1)$$

where ˆ denotes terms that are omitted.

This property is easy to prove by induction. However, we will not need all the power of this property. Below we formulate the conditions that we need to impose on the random variables to prove the CLT. These conditions are consequences of freeness but are likely to be weaker. We will say that a sequence of *zero-mean* random variables $X_1, ..., X_n, ...$ satisfies *Condition A* if:

1. For every k, $E\left(X_k X_{i_1} ... X_{i_r}\right) = 0$ provided that $i_s \neq k$ for $s = 1, ..., r$.
2. For every $k \geq 2$, $E\left(X_k^2 X_{i_1} ... X_{i_r}\right) = E\left(X_k^2\right) E\left(X_{i_1} ... X_{i_r}\right)$ provided that $i_s < k$ for $s = 1, ..., r$.
3. For every $k \geq 2$,

$$E\left(X_k X_{i_1} ... X_{i_p} X_k X_{i_{p+1}} ... X_{i_r}\right) = E\left(X_k^2\right) E\left(X_{i_1} ... X_{i_p}\right) E\left(X_{i_{p+1}} ... X_{i_r}\right)$$

provided that $i_s < k$ for $s = 1, ..., r$.

Intuitively, if we know how to calculate every moment of the sequence $X_1, ..., X_{k-1}$, then using Condition A we can also calculate the expectation of any product of random variables X_1, ..., X_k that involves no more than two occurrences of variable X_k. Part 1 of Condition A is stronger than is needed for this calculation, since it involves variables with indices higher than k. However, we will need this additional strength in the proof of Lemma 13 below, which is essential for the proof of the main result.

Proposition 3. *Every sequence of free random variables X_1, ..., X_n, ... satisfies Condition A.*

This proposition can be checked by direct calculation using Proposition 2.
We will also need the following fact.

Proposition 4. *Let $X_1, ..., X_l$ be zero-mean variables that satisfy Condition A(1), and let $Y_{l+1}, ..., Y_n$ be zero-mean variables which are free from each other and from the algebra generated by variables $X_1, ..., X_l$. Then the sequence $X_1, ..., X_l, Y_{l+1}, ..., Y_n$ satisfies Condition A(1).*

Proof: Consider the moment $E\left(X_k A_{i_1} ... A_{i_s}\right)$, where A_{i_t} is either one of Y_j or one of X_i but it can equal X_k. Then we can use the fact that Y_j are free and write

$$E\left(X_k A_{i_1} ... A_{i_s}\right) = \sum_{\alpha} c_\alpha E\left(X_k X_{i_1(a)} ... X_{i_r(\alpha)}\right),$$

where none of $X_{i_t(\alpha)}$ equals X_k. Then, using the assumption that X_i satisfy Condition A(1), we conclude that $E\left(X_k A_{i_1} ... A_{i_s}\right) = 0$. Also, $E\left(Y_k A_{i_1} ... A_{i_s}\right) = E\left(Y_k\right) E\left(A_{i_1} ... A_{i_s}\right) = 0$, provided that none of A_{i_t} equals Y_k. In sum, the sequence $X_1, ..., X_l, Y_{l+1}, ..., Y_n$ satisfies Condition A(1). QED.

While the freeness of random variables X_i is the same concept as the freeness of the algebras that they generate, Condition A deals only with variables X_i, and not with the algebras that they generate. For example, it is conceivable that a sequence $\{X_i\}$ satisfies condition A but $\{X_i^2 - E(X_i^2)\}$ does not. In particular, this implies that Condition A requires checking a much smaller set of moment conditions than freeness. Below we will present an example of random variables which are not free but which satisfy Condition A.

Recall that the standard semicircle law μ_{SC} is the probability distribution on $\mathbb{R}$ with the density $\pi^{-1}\sqrt{4-x^2}$ if $x \in [-2; 2]$, and 0 otherwise. We are going to prove the following Theorem.

Theorem 5. *Suppose that*
(i) $\{\xi_i\}$ *is a sequence of self-adjoint random variables that satisfies Condition* A;
(ii) every ξ_i *has asbsolute moments of all orders, which are uniformly bounded, i.e.,* $E|\xi_i|^k \leq \mu_k$ *for all* i;
(iii) $E\xi_i = 0$, $E\xi_i^2 = \sigma_i^2$;
(iv) $(\sigma_1^2 + ... + \sigma_N^2)/N \to s$ *as* $N \to \infty$.
Then the spectral measure of $S_N = (\xi_1 + ... + \xi_N)/\sqrt{\sigma_1^2 + ... + \sigma_N^2}$ *converges in distribution to the semicircle law* μ_{SC}.

The contribution of this theorem is twofold. First, it shows that the semicircle central limit holds for a certain class of non-free variables. Second, it gives a proof of the free CLT which is different from the usual proof through R-transforms. However, it is not stronger than a version of the free CLT which is formulated in Section 2.5 in [10].

3 Example

Let us present an example that suggest that Condition A is strictly weaker than the freeness condition.

Let F be the free group with a countable number of generators f_k. Consider the set of relations $R = \{f_k f_{k-1} f_k f_{k-1} f_k f_{k-1} = e\}$, where $k \geq 2$, and define $G = F/\mathcal{R}$, that is, G is the group with generators f_k and relations generated by relations in R.

Here are some consequences of these relationships:

1) $f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k = e$.
(Indeed, $e = f_k^{-1}(f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}) f_k = f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k$.)
2) $f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} = e$ and $f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} = e$.

We are interested in the structure of the group G. For this purpose we will study the structure of $\mathcal{R}$, which is a subgroup of F generated by elements of R and their conjugates. We will represent elements of F by *words*, that is, by sequences of generators. We will say that a word is *reduced* if does not have a subsequence of the form $f_k f_k^{-1}$ or $f_k^{-1} f_k$. It is *cyclically reduced* if it does not have the form of $f_k ... f_k^{-1}$ or $f_k^{-1} ... f_k$. We will call a number of elements in a reduced word w its length and denote it as $|w|$. A set of relations R is *symmetrized* if for every word $r \in R$, the set R also contains its inverse r^{-1} and all cyclically reduced conjugates of both r and r^{-1}.

For our particular example, a symmetrized set of relations is given by the following list:

$$R=\left\{\begin{array}{cc} f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}, & f_{k-1} f_k f_{k-1} f_k f_{k-1} f_k, \\ f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1}, & f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} \end{array}\right\},$$

where k are all integers ≥ 2.
A word b is called a *piece* (relative to a symmetrized set R) if there exist two elements of R, r_1 and r_2, such that $r_1 = bc_1$ and $r_2 = bc_2$. In our case, each f_k and f_k^{-1} with index $k \geq 2$ is a piece because f_k is the initial part of relations $f_k f_{k-1} f_k f_{k-1} f_k f_{k-1}$ and $f_k f_{k+1} f_k f_{k+1} f_k f_{k+1}$, and f_k^{-1} is the initial part of relations $f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1} f_k^{-1} f_{k-1}^{-1}$ and $f_k^{-1} f_{k+1}^{-1} f_k^{-1} f_{k+1}^{-1} f_k^{-1} f_{k+1}^{-1}$. There is no other pieces.
Now we introduce the condition of *small cancellation* for a symmetrized set R:

Condition 6 ($C'(\lambda)$)**.** *If $r \in R$ and $r = bc$ where b is a piece, then $|b| < \lambda |r|$.*

Essentially, the condition says that if two relations are multiplied together, then a possible cancellation must be relatively small. Note that if R satisfies $C'(\lambda)$ then it satisfies $C'(\mu)$ for all $\mu \geq \lambda$.
In our example R satisfies $C'(1/5)$.
Another important condition is the triangle condition.

Condition 7 (T)**.** *Let r_1, r_2, and r_3 be three arbitrary elements of R such that $r_2 \neq r_1^{-1}$ and $r_3 \neq r_2^{-1}$ Then at least one of the products r_1r_2, r_2r_3, or r_3r_1, is reduced without cancellation.*

In our example, Condition (T) is satisfied.
If s is a word in F, then $s > \lambda R$ means that there exists a word $r \in R$ such that $r = st$ and $|s| > \lambda |r|$. An important result from small cancellation theory that we will use later is the following theorem:

Theorem 8 (Greendlinger's Lemma)**.** *Let R satisfy $C'(1/4)$ and T. Let w be a non-trivial, cyclically reduced word with $w \in \mathcal{R}$. Then either*
(1) $w \in R$,
or some cyclycally reduced conjugate w^* of w contains one of the following:
(2) two disjoint subwords, each $> \frac{3}{4}R$, or
(4) four disjoint subwords, each $> \frac{1}{2}R$.

This theorem is Theorem 4.6 on p. 251 in [7].
Since in our example R satisfies both $C'(1/4)$ and T, we can infer that in our case the conclusion of the theorem must hold. For example, (2) means that we can find two disjoint subwords of w, s_1 and s_2, and two elements of R, r_1 and r_2, such that $r_i = s_i t_i$ and $|s_i| > (3/4)|r_i| = 9/2$. In particular, we can conclude that in this case $|w| \geq 10$. Similarly, in case (4), $|w| \geq 16$. One immediate application is that G does not collapse into the trivial group. Indeed, f_i are not zero.
Let $L^2(G)$ be the functions of G that are square-summable with respect to the counting measure. G acts on $L^2(G)$ by left translations:

$$(L_g x)(h) = x(gh).$$

Let $\mathcal{A}$ be the group algebra of G. The action of G on $L^2(G)$ can be extended to the action of $\mathcal{A}$ on $L^2(G)$. Define the expectation on this group algebra by the following rule:

$$E(h) = \langle \delta_e, L_h \delta_e \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L^2(G)$. Alternatively, the expectation can be written as follows:

$$E(h) = a_e,$$

where $h = \sum_{g \in G} a_g g$ is a representation of a group algebra element h as a linear combination of elements $g \in G$. The expectation is clearly positive and finite by definition. It is also tracial because $g_1 g_2 = e$ if and only if $g_2 g_1 = e$.

If $L_h = \sum_{g \in G} a_g L_g$ is a linear operator corresponding to the element of group algebra $h = \sum_{g \in G} a_g g$, then its adjoint is $(L_h)^* = \sum_{g \in G} \overline{a_g} L_{g^{-1}}$, which corresponds to the element $h^* = \sum_{g \in G} \overline{a}_g g^{-1}$.

Consider elements $X_i = f_i + f_i^{-1}$. They are self-adjoint and $E(X_i) = 0$. Also we can compute $E\left(X_i^2\right) = 2$. Indeed it is enough to note that $f_i^2 \neq e$ and $f_i^{-2} \neq e$, and this holds because insertion or deletion of an element from R changes the degree of f_i by a multiple of 3. Therefore, every word equal to zero must have the degree of every f_i equal to 0 modulo 3.

Proposition 9. *The sequence of variables $\{X_i\}$ is not free but satisfies Condition A.*

Proof: The variables X_k are not free. Consider $X_2 X_1 X_2 X_1 X_2 X_1$. Its expectation is 2, because $f_2 f_1 f_2 f_1 f_2 f_1 = e$ and $f_2^{-1} f_1^{-1} f_2^{-1} f_1^{-1} f_2^{-1} f_1^{-1} = e$, and all other terms in the expansion of $X_2 X_1 X_2 X_1 X_2 X_1$ are different from e. Indeed, the only terms that are not of the form above but still have the degree of all f_i equal to zero modulo 3 are $f_2 f_1^{-1} f_2 f_1^{-1} f_2 f_1^{-1}$ and $f_2^{-1} f_1 f_2^{-1} f_1 f_2^{-1} f_1$, but they do not equal zero by application of Greendlinger's lemma. Therefore, $E(X_2 X_1 X_2 X_1 X_2 X_1) = 2$. This contradicts the definition of freeness of variables X_2 and X_1.

Let us check Condition A. For A(1), we have to prove that $E(X_k X_{i_1} ... X_{i_n}) = 0$, where $k \neq i_s$ and $i_s \neq i_{s+1}$ for every s. Consider $E(f_k f_{i_1} ... f_{i_n})$, where $k \neq i_s$ and $i_s \neq i_{s+1}$ for every s. Note $f_k f_{i_1} ... f_{i_n} \neq e$, as can be seen from the fact that the degree of f_k does not equal zero modulo 3. Therefore $E(f_k f_{i_1} ... f_{i_n}) = 0$. A similar argument works for $E\left(f_k^{-1} f_{i_1} ... f_{i_n}\right) = 0$ and more generally for the expectation of every element of the form $f_k^{\varepsilon} f_{i_1}^{n_1} ... f_{i_n}^{n_2}$, where $\varepsilon = \pm 1$ and n_s are integer.

Similarly, we can prove that $E\left(f_k^{\pm 2} f_{i_1}^{n_1} ... f_{i_n}^{n_2}\right) = 0$ and this suffices to prove A(2).

For A(3) we have to consider elements of the form $f_k^{\varepsilon_1} f_{i_1} ... f_{i_p} f_k^{\varepsilon_2} f_{i_{p+1}} ... f_{i_q}$. Assume that neither $f_{i_1} ... f_{i_p}$ nor $f_{i_{p+1}} ... f_{i_q}$ can be reduced to e. Otherwise we can use property A2. Then the claim is that $E\left(f_k^{\varepsilon_1} f_{i_1} ... f_{i_p} f_k^{\varepsilon_2} f_{i_{p+1}} ... f_{i_q}\right) = 0$. This is clear when ε_1 and ε_2 have the same sign since in this case the degree of f_k does not equal 0 modulo 3. A more difficult case is when $\varepsilon_1 = 1$ and $\varepsilon_2 = -1$. (The case with opposite signs is similar.) However, in this case we can conclude that $f_k f_{i_1} ... f_{i_p} f_k^{-1} f_{i_{p+1}} ... f_{i_q} \neq e$ by an application of Greendlinger's lemma. Indeed, the only subwords that this word can contain and which would also be subwords of an element of R, are subwords of length 1 and 2. But these subwords fail to satisfy the requirement of either (2) or (4) in Greendlinger's lemma. Therefore, we can conclude that $f_k f_{i_1} ... f_{i_p} f_k^{-1} f_{i_{p+1}} ... f_{i_q} \neq e$, and therefore A(3) is also satisfied. Thus Condition A is satisfied by random variables $X_1, ..., X_k, ...$ in algebra $\mathcal{A}$, although these variables are not free. QED.

4 Proof of the Main Result

Outline of Proof: Our proof of the free CLT proceeds along the familiar lines of the Lindeberg method. We take a family of functions, $\{f\}$, and compare $Ef(S_N)$ with $Ef\left(\widetilde{S}_N\right)$, where $S_N = X_1 + ... + X_N$ and $\widetilde{S}_N = Y_1 + ... + Y_N$, and Y_i are free semicircle variables chosen in such a way that $\mathrm{Var}(S_N) = \mathrm{Var}\left(\widetilde{S}_N\right)$. To estimate $\left|Ef(S_N) - Ef\left(\widetilde{S}_N\right)\right|$, we substitute the elements in S_N with free semicircle variables, one by one, and estimate the corresponding change in the expected value of $f(S_N)$. After that, we show that the total change, as all elements in the sum are substituted with semicircle random variables, is asymptotically small as $N \to \infty$. Finally, the tightness of the selected family of functions allows us to conclude that the distribution of S_N must converge to the semicircle law as $N \to \infty$.

The usual choice of functions f in the classical case are functions from $C_c^3(\mathbb{R})$, that is, functions with a continuous third derivative and compact support. In the non-commutative setting this family of functions is not appropriate because the usual Taylor series formula is difficult to apply. Intuitively, it is difficult to develop $f(X+h)$ in a power series of h if variables X and h do not commute. Since the Taylor formula is crucial for estimating the change in $Ef(S_N)$, we will still use it but we will restrict the family of functions to polynomials.

To show that the family of polynomials is sufficiently rich for our purposes, we use the following Proposition:

Proposition 10. *Suppose there is a unique distribution function F with the moments $\left\{m^{(r)}, r \geq 1\right\}$. Suppose that $\{F_N\}$ is a sequence of distribution functions, each of which has all its moments finite:*

$$m_N^{(r)} = \int_{-\infty}^{\infty} x^r dF_N.$$

Finally, suppose that for every $r \geq 1$:

$$\lim_{n\to\infty} m_N^{(r)} = m^{(r)}.$$

Then $F_N \to F$ vaguely.

See Theorem 4.5.5.on page 99 in [3] for a proof. Note that Chung uses words "vague convergence" to denote that kind of convergence which is more often called the weak convergence of probability measures.

Since the semicircle distribution is bounded and therefore is determined by its moments (see Corollary to Theorem II.12.7 in [8]), therefore the assumption of Proposition 10 is satisfied, and we only need to show that the moments of S_n converge to the corresponding moments of the semicircle distribution.

Proof of Theorem 5: Define η_i as a sequence of random variables that are freely independent among themselves and also freely independent from all ξ_i. Suppose also that η_i have semicircle distributions with $E\eta_i = 0$ and $E\eta_i^2 = \sigma_i^2$. We are going to accept the fact that the sum of free semicircle random variables is semicircle, and therefore, the spectral distribution of $(\eta_1 + ... + \eta_N)/\left(s\sqrt{N}\right)$ converges in distribution to the semicircle law μ_{SC} with zero expectation and unit variance. Let us define $X_i = \xi_i/s_N$ and $Y_i = \eta_i/s_N$. We will proceed by proving that moments of $X_1 + ... + X_N$ converge to moments of $Y_1 + ... + Y_N$ and applying Proposition 10. Let

$$\Delta f = Ef(X_1 + ... + X_N) - Ef(Y_1 + ... + Y_N),$$

where $f(x) = x^m$. We want to show that this difference approaches zero as N grows.

By assumption, $EY_i = EX_i = 0$ and $EY_i^2 = EX_i^2 = \sigma_i^2/s_N^2$.

The first step is to write the difference Δf as follows:

$$\begin{aligned}\Delta f &= [Ef(X_1+...+X_{N-1}+X_N)-Ef(X_1+...+X_{N-1}+Y_N)] \\ &\quad +[Ef(X_1+...+X_{N-1}+Y_N)-Ef(X_1+...+Y_{N-1}+Y_N)] \\ &\quad +[Ef(X_1+Y_2+...+Y_{N-1}+Y_N)-Ef(Y_1+Y_2+...+Y_{N-1}+Y_N)].\end{aligned}$$

We intend to estimate every difference in this sum. Let

$$Z_k = X_1+...+X_{k-1}+Y_{k+1}+...+Y_N. \tag{2}$$

We are interested in

$$Ef(Z_k+X_k)-Ef(Z_k+Y_k).$$

We are going to apply the Taylor expansion formula but first we define directional derivatives. Let $f'_{X_k}(Z_k)$ be the *derivative of f at Z_k in direction X_k*, defined as follows:

$$f'_{X_k}(Z_k) =: \lim_{t\downarrow 0}\frac{f(Z_k+tX_k)-f(Z_k)}{t}.$$

The higher order directional derivatives can be defined recursively. For example,

$$f''_{X_k}(Z_k) =: \left(f'_{X_k}\right)'_{X_k}(Z_k) = \lim_{t\downarrow 0}\frac{f'_{X_k}(Z_k+tX_k)-f'_{X_k}(Z_k)}{t}. \tag{3}$$

For polynomials, this definition is equivalent to the following definition:

$$f''_{X_k}(Z_k) = 2\lim_{t\downarrow 0}\frac{f(Z_k+tX_k)-f(Z_k)-tf'_{X_k}(Z_k)}{t^2}. \tag{4}$$

Example 11. *Operator directional derivatives of $f(x)=x^4$*

Let us compute $f'_X(Z)$ and $f''_X(Z)$ for $f(x)=x^4$. Using definitions we get

$$f'_X(Z) = Z^3X+Z^2XZ+ZXZ^2+XZ^3$$

and

$$f''_X(Z) = 2\left(Z^2X^2+ZXZX+XZ^2X+ZX^2Z+XZXZ+X^2Z^2\right), \tag{5}$$

and the expression for $f''_X(Z)$ does not depend on whether definition (3) or (4) was applied.

The derivatives of f at $Z_k+\tau X_k$ in direction X_k are defined similarly, for example:

$$\begin{aligned}& f'''_{X_k}(Z_k+\tau X_k) \\ =\;& 6\lim_{t\downarrow 0}\frac{f(Z_k+(\tau+t)X_k)-f(Z_k+\tau X_k)-tf'_{X_k}(Z_k+\tau X_k)-\frac{1}{2}t^2f''_{X_k}(Z_k+\tau X_k)}{t^3}.\end{aligned}$$

Next, let us write the Taylor formula for $f(Z_k+X_k)$:

$$f(Z_k+X_k) = f(Z_k)+f'_{X_k}(Z_k)+\frac{1}{2}f''_{X_k}(Z_k)+\frac{1}{2}\int_0^1(1-\tau)^2 f'''_{X_k}(Z_k+\tau X_k)\,d\tau. \tag{6}$$

Formula (6) can be obtained by integration by parts from the expression

$$f(Z_k + X_k) - f(Z_k) = \int_0^1 f'_{X_k}(Z_k + \tau X_k)\, d\tau.$$

For polynomials it is easy to write the explicit expressions for $f^{(r)}_{X_k}(Z_k)$ or $f^{(r)}_{X_k}(Z_k + \tau X_k)$ although they can be quite cumbersome for polynomials of high degree. Very schematically, for a function $f(x) = x^m$, we can write

$$f'_{X_k}(Z_k) = X_k Z_k^{m-1} + Z_k X_k Z_k^{m-2} + ... + Z_k^{m-1} X_k, \tag{7}$$

and

$$f''_{X_k}(Z_k) = 2\left(X_k^2 Z_k^{m-2} + X_k Z_k X_k Z_k^{m-3} + ... + Z_k^{m-2} X_k^2\right), \tag{8}$$

Similar formulas hold for $f'_{Y_k}(Z_k)$ and $f''_{Y_k}(Z_k)$, with the change that Y_k should be used instead of X_k.

Using the assumptions that sequence $\{X_k\}$ satisfies Condition A and that variables Y_k are free, we can conclude that $Ef'_{Y_k}(Z_k) = Ef'_{X_k}(Z_k) = 0$ and that $Ef''_{Y_k}(Z_k) = Ef''_{X_k}(Z_k)$. Indeed, consider, for example, (8). We can use expression (2) for Z_k and the free independence of Y_i to expand (8) as

$$Ef''_{X_k}(Z_k) = \sum_\alpha c_\alpha P_\alpha\left(E\left(X_k \overline{X_1} X_k \overline{X_2}\right), E\left(X_k \overline{X_3} X_k \overline{X_4}\right), ...\right), \tag{9}$$

where $\overline{X_i}$ denotes certain monomials in variables $X_1, ..., X_{k-1}$ (i.e., $\overline{X_i} = X_{i_1} ... X_{i_p}$ with $i_k \in \{1, ..., k-1\}$), and where α indexes certain polynomials P_α. In other words, using the free independence of Y_i and X_i we expand the expectations of polynomial $f''_{X_k}(Z_k)$ as a sum over polynomials in joint moments of variables X_j and Y_i where $j = 1, ..., k$ and $i = k+1, ..., N$. By freeness, we can reduce the resulting expression so that the moments in the reduced expression are either joint moments of variables X_j or joint moments of variables Y_i but never involve both X_j and Y_i. Moreover, we can explictly calculate the moments of Y_i (i.e., expectations of the products of Y_i) because their are mutually free. The resulting expansion is (9).

Let us try to make this process clearer by an example. Suppose that $f(x) = x^4$, $N = 4$, $k = 2$ and $Z_k = Z_2 = X_1 + Y_3 + Y_4$. We aim to compute $Ef''_{X_2}(Z_2)$. Using formula (5), we write:

$$\begin{aligned} Ef''_{X_2}(Z_2) &= 2E\left(Z_2^2 X_2^2 + ...\right) \\ &= 2E\left((X_1 + Y_3 + Y_4)^2 X_2^2 + ...\right) \\ &= 2\{E\left(X_1^2 X_2^2\right) + E\left(X_1 Y_3 X_2^2\right) + E\left(X_1 Y_4 X_2^2\right) \\ &\quad + E\left(Y_3 X_1 X_2^2\right) + E\left(Y_3^2 X_2^2\right) + E\left(Y_3 Y_4 X_2^2\right) \\ &\quad + E\left(Y_4 X_1 X_2^2\right) + E\left(Y_4 Y_3 X_2^2\right) + E\left(Y_4^2 X_2^2\right) + ...\}. \end{aligned}$$

Then, using the freeness of Y_3 and Y_4 and the facts that $E(Y_i) = 0$ and $E\left(Y_i^2\right) = \sigma_i^2$, we continue as follows:

$$Ef''_{X_2}(Z_2) = 2\{E\left(X_1^2 X_2^2\right) + \sigma_3^2 E\left(X_2^2\right) + \sigma_4^2 E\left(X_2^2\right) + ...\},$$

which is the expression we wanted to obtain.

It is important to note that the coefficients c_α do not depend on variables X_j but only on Y_j, $j > k$, and on the locations which Y_j take in the expansion of $f''_{X_k}(Z_k)$. Therefore, we can substitute Y_k for X_k and develop a similar formula for $Ef''_{Y_k}(Z_k)$:

$$Ef''_{Y_k}(Z_k) = \sum_\alpha c_\alpha P_\alpha \left(E\left(Y_k \overline{X_1} Y_k \overline{X_2}\right), E\left(Y_k \overline{X_3} Y_k \overline{X_4}\right), ...\right). \tag{10}$$

In the example above, we will have

$$Ef''_{Y_2}(Z_2) = 2\{E\left(X_1^2 Y_2^2\right) + \sigma_3^2 E\left(Y_2^2\right) + \sigma_4^2 E\left(Y_2^2\right) + ...\}.$$

Formula (10) is exactly the same as formula (9) except that all X_k are substituted with Y_k. Finally, using Condition A we obtain that for every i:

$$\begin{aligned} E\left(Y_k \overline{X_i} Y_k \overline{X_{i+1}}\right) &= E\left(Y_k^2\right) E\left(\overline{X_i}\right) E\left(\overline{X_{i+1}}\right) \\ &= E\left(X_k^2\right) E\left(\overline{X_i}\right) E\left(\overline{X_{i+1}}\right) \\ &= E\left(X_k \overline{X_i} X_k \overline{X_{i+1}}\right), \end{aligned}$$

and therefore $Ef''_{Y_k}(Z_k) = Ef''_{X_k}(Z_k)$.
Consequently,

$$\begin{aligned} & Ef(Z_k + X_k) - Ef(Z_k + Y_k) \\ = & \frac{1}{2}\int_0^1 (1-\tau)^2 Ef'''_{X_k}(Z_k + \tau X_k)\, d\tau - \frac{1}{2}\int_0^1 (1-\tau)^2 Ef'''_{Y_k}(Z_k + \tau Y_k)\, d\tau. \end{aligned}$$

Next, note that if f is a polynomial, then $f'''_{X_k}(Z_k + \tau X_k)$ is the sum of a finite number of terms which are products of $Z_k + \tau X_k$ and X_k. The number of terms in this expansion is bounded by C_1, which depends only on the degree m of the polynomial f.
A typical term in the expansion looks like

$$E\left(Z_k + \tau X_k\right)^{m-7} X_k^3 \left(Z_k + \tau X_k\right)^3 X_k.$$

In addition, if we expand the powers of $Z_k + \tau X_k$, we will get another expansion that has the number of terms bounded by C_2, where C_2 depends only on m. A typical element of this new expansion is

$$E\left(Z_k^{m-7} X_k^3 Z_k^2 X_k^2\right).$$

Every term in this expansion has a total degree of X_k not less than 3, and, correspondingly, a total degree of Z_k not more than $m-3$. Our task is to show that as $n \to \infty$, these terms approach 0.
We will use the following lemma to estimate each of the summands in the expansion of $f'''_{X_k}(Z_k + \tau X_k)$.

Lemma 12. *Let X and Y be self-adjoint. Then*

$$\begin{aligned} & \left|E\left(X^{m_1} Y^{n_1} ... X^{m_r} Y^{n_r}\right)\right| \\ \leq & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}. \end{aligned}$$

Proof: For $r = 1$, this is the usual Cauchy-Schwartz inequality for traces:

$$|E(X^{m_1}Y^{n_1})|^2 \le E\left(X^{2m_1}\right) E\left(Y^{2n_1}\right).$$

See, for example, Proposition I.9.5 on p. 37 in [9].
Next, we proceed by induction. We have two slightly different cases to consider. Assume first that r is even, $r = 2s$. Then, by the Cauchy-Schwartz inequality, we have:

$$\begin{aligned} & |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\ \le \; & E(X^{m_1}Y^{n_1}...X^{m_s}Y^{n_s}Y^{n_s}X^{m_s}...Y^{n_1}X^{m_1}) E(Y^{n_r}X^{m_r}...Y^{n_{s+1}}X^{m_{s+1}}X^{m_{s+1}}Y^{n_{s+1}}...X^{m_r}Y^{n_r}) \\ = \; & E\left(X^{2m_1}Y^{n_1}...X^{m_s}Y^{2n_s}X^{m_s}...Y^{n_1}\right) E\left(Y^{2n_r}X^{m_r}...Y^{n_{s+1}}X^{2m_{s+1}}Y^{n_{s+1}}...X^{m_r}\right). \end{aligned}$$

Applying the inductive hypothesis, we obtain:

$$\begin{aligned} & |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\ \le \; & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r+1}} \left[E\left(Y^{2^r n_s}\right)\right]^{2^{-r+1}} \left[E\left(Y^{2^{r-1} n_1}\right)\right]^{2^{-r+2}} ... \left[E\left(X^{2^{r-1} m_s}\right)\right]^{2^{-r+2}} \\ & \times \left[E\left(X^{2^r m_{s+1}}\right)\right]^{2^{-r+1}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r+1}} \left[E\left(Y^{2^{r-1} n_{s+1}}\right)\right]^{2^{-r+2}} ... \left[E\left(X^{2^{r-1} m_r}\right)\right]^{2^{-r+2}}. \end{aligned}$$

We recall that by the Lyapunov inequality, $\left[E\left(Y^{2^{r-1}n_1}\right)\right]^{2^{-r+2}} \le \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r+1}}$ and we get the desired inequality:

$$\begin{aligned} & |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})| \\ \le \; & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}. \end{aligned}$$

Now let r be odd, $r = 2s + 1$. Then

$$\begin{aligned} & |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})|^2 \\ \le \; & E(X^{m_1}Y^{n_1}...Y^{n_s}X^{m_{s+1}}X^{m_{s+1}}Y^{n_s}...Y^{n_1}X^{m_1}) E(Y^{n_r}X^{m_r}...X^{m_{s+2}}Y^{n_{s+1}}Y^{n_{s+1}}X^{m_{s+2}}...X^{m_r}Y^{n_r}) \\ = \; & E\left(X^{2m_1}Y^{n_1}...Y^{n_s}X^{2m_{s+1}}Y^{n_s}...Y^{n_1}\right) E\left(Y^{2n_r}X^{m_r}...X^{m_{s+2}}Y^{2n_{s+1}}X^{m_{s+1}}...X^{m_r}\right). \end{aligned}$$

After that we can use the inductive hypothesis and the Lyapunov inequality and obtain that

$$\begin{aligned} & |E(X^{m_1}Y^{n_1}...X^{m_r}Y^{n_r})| \\ \le \; & \left[E\left(X^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(Y^{2^r n_r}\right)\right]^{2^{-r}}. \end{aligned}$$

QED.
We apply Lemma 12 to estimate each of the summands in the expansion of $f'''_{X_k}(Z_k + \tau X_k)$. Consider a summand $E\left(Z_k^{m_1}X_k^{n_1}...Z_k^{m_r}X_k^{n_r}\right)$. Then by Lemma 12, we have

$$\begin{aligned} & |E(Z_k^{m_1}X_k^{n_1}...Z_k^{m_r}X_k^{n_r})| \\ \le \; & \left[E\left(Z_k^{2^r m_1}\right)\right]^{2^{-r}} \left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(Z_k^{2^r m_r}\right)\right]^{2^{-r}} \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}}. \end{aligned} \tag{11}$$

Next step is to estimate the absolute moments of the variable Z_k.

Lemma 13. *Let* $Z = (v_1 + ... + v_N)/N^{1/2}$, *where* v_i *are self-adjoint and satisfy condition A(1) and let* $E|v_i|^k \leq \mu_k$ *for every* i. *Then, for every integer* $r \geq 0$

$$E(|Z|^r) = O(1) \text{ as } N \to \infty.$$

Proof: We will first treat the case of even r. In this case, $E(|Z|^r) = E(Z^r)$. Consider the expansion of $(v_1 + ... + v_N)^r$. Let us refer to the indices 1, ..., N as colors of the corresponding v. If a term in the expansion includes more than $r/2$ distinct colors, then one of the colors must be used by this term only once. Therefore, by the first part of condition A the expectation of such a term is 0.
Let us estimate a number of terms in the expansion that include no more than $r/2$ distinct colors. Consider a fixed combination of $\leq r/2$ colors. The number of terms that use colors only from this combination is $\leq (r/2)^r$. Indeed, consider the product $(v_1 + ... + v_N)(v_1 + ... + v_N)...(v_1 + ... + v_N)$ with r product terms. We can choose an element from the first product term in $r/2$ possible ways, an element from the second product term in $r/2$ possible ways, etc. Therefore, the number of all possible choices is $(r/2)^r$. On the other hand, the number of possible different combinations of $k \leq r/2$ colors is

$$\frac{N!}{(N-k)!k!} \leq N^{r/2}.$$

Therefore, the total number of terms that use no more than $r/2$ colors is bounded from above by

$$(r/2)^r N^{r/2}.$$

Now let us estimate the expectation of an individual term in the expansion. In other words we want to estimate $E\left(v_{i_1}^{k_1}...v_{i_s}^{k_s}\right)$, where $k_t \geq 1$, $k_1 + ... + k_s = r$, and $i_t \neq i_{t+1}$. First, note that

$$\left|E\left(v_{i_1}^{k_1}...v_{i_s}^{k_s}\right)\right| \leq E\left(\left|v_{i_1}^{k_1}...v_{i_s}^{k_s}\right|\right).$$

Indeed, using the Cauchy-Schwartz inequality, for any operator X we can write

$$\begin{aligned} |E(X)|^2 &= \left|E\left(U|X|^{1/2}|X|^{1/2}\right)\right|^2 \leq E\left(|X|^{1/2}U^*U|X|^{1/2}\right)E\left(|X|^{1/2}|X|^{1/2}\right) \\ &= E(|X|P)E(|X|), \end{aligned}$$

where U is a partial isometry and $P = U^*U$ is a projection. Note that from the positivity of the expectation functional it follows that $E(|X|P) \leq E(|X|)$. Therefore, we can conclude that $|E(X)| \leq E(|X|)$.
Next, we use the Hölder inequality for traces of non-commutative operators (see [4], especially Corollary 4.4(iii) on page 324, for the case of the trace in a von Neumann algebra and Section III.7.2 in [5] for the case of compact operators and the usual operator trace). Note that

$$\underbrace{\frac{1}{s} + ... + \frac{1}{s}}_{s\text{-times}} = 1,$$

therefore, the Hölder inequality gives

$$E\left(\left|v_{i_1}^{k_1}...v_{i_s}^{k_s}\right|\right) \leq \left[E\left(|v_{i_1}|^{k_1 s}\right)...E\left(|v_{i_s}|^{k_s s}\right)\right]^{1/s}.$$

Using this result and the uniform boundedness of the moments (from assumption of the lemma), we get:

$$\log \left| E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq \frac{1}{s}\sum_{i=1}^{s} \log \mu_{k_i s}.$$

Without loss of generality we can assume that bounds μ_k are increasing in k. Using the facts that $s \leq r$ and $k_i \leq r$, we obtain the bound:

$$\log \left| E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq \log \mu_{r^2},$$

or

$$\left| E\left(v_{i_1}^{k_1} ... v_{i_s}^{k_s}\right)\right| \leq \mu_{r^2}.$$

Therefore,

$$E\left(v_1 + ... + v_N\right)^r \leq (r/2)^r \mu_{r^2} N^{r/2},$$

and

$$E\left(Z^r\right) \leq (r/2)^r \mu_{r^2}. \tag{12}$$

Now consider the case of odd r. In this case, we use the Lyapunov inequality to write:

$$\begin{aligned} E\left|Z\right|^r &\leq \left(E\left|Z\right|^{r+1}\right)^{\frac{r}{r+1}} \qquad (13) \\ &\leq \left(\left(\frac{r+1}{2}\right)^{r+1} \mu_{(r+1)^2}\right)^{\frac{r}{r+1}} \\ &= \left(\frac{r+1}{2}\right)^{r} \left(\mu_{(r+1)^2}\right)^{\frac{r}{r+1}}. \end{aligned}$$

The important point is that the bounds in (12) and (13) do not depend on N. QED.
By definition $Z_k = \left(\xi_1 + ... + \xi_{k-1} + \eta_{k+1} + ... + \eta_N\right)/s_N$ and by assumption ξ_i and η_i are uniformly bounded, and $s_N \sim \sqrt{N}$. Moreover, $\xi_1, ..., \xi_{k-1}$ satisfy Condition A by assumption, and $\eta_{k+1}, ..., \eta_N$ are free from each other and from $\xi_1, ..., \xi_{k-1}$. Therefore, by Proposition 4, $\xi_1, ..., \xi_{k-1}, \eta_{k+1}, ..., \eta_N$ satisfy condition A(1). Consequently, we can apply Lemma 13 to Z_k and conclude that $E\left|Z_k\right|^r$ is bounded by a constant that depends only on r but does not depend on N.
Using this fact, we can continue the estimate in (11) and write:

$$\begin{aligned} &\left|E\left(Z_k^{m_1} X_k^{n_1} ... Z_k^{m_r} X_k^{n_r}\right)\right| \qquad (14) \\ \leq\ & C_4 \left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}}, \end{aligned}$$

where the constant C_4 depends only on m.
Next we note that

$$\left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} \leq C\left(\frac{\mu_{2^r n_1}}{N^{2^{r-1} n_1}}\right)^{2^{-r}} = C\frac{\left(\mu_{2^r n_1}\right)^{2^{-r}}}{N^{n_1/2}}.$$

Next note that $n_1 + ... + n_r \geq 3$; therefore we can write

$$\left[E\left(X_k^{2^r n_1}\right)\right]^{2^{-r}} ... \left[E\left(X_k^{2^r n_r}\right)\right]^{2^{-r}} \leq C' N^{-3/2}.$$

In sum, we obtain the following Lemma:

Lemma 14.

$$\left|Ef'''_{X_k}(Z_k + \tau X_k)\right| \leq C_5 N^{-3/2},$$

where C_5 depends only on the degree of polynomial f and the sequence of constants μ_k.

A similar result holds for $\left|Ef'''_{X_k}(Z_k + \tau Y_k)\right|$ and we can conclude that

$$|Ef(Z_k + X_k) - Ef(Z_k + Y_k)| \leq C_6 N^{-3/2}.$$

After we add these inequalities over all $k = 1, ..., N$ we get

$$|Ef(X_1 + ... + X_N) - Ef(Y_1 + ... + Y_N)| \leq C_7 N^{-1/2}.$$

Clearly this estimate approaches 0 as N grows. Applying Proposition 10, we conclude that the measure of $X_1 + ... + X_N$ converges to the measure of $Y_1 + ... + Y_N$ in distribution. This finishes the proof of the main theorem.

5 Concluding Remarks

The key points of this proof are as follows: 1) We can substitute each random variable X_i in the sum S_N with a free random variable Y_i so that the first and the second derivatives of any polynomial with S_N in the argument remain unchanged. The possibility of this substitution depends on Condition A being satisfied by X_i. 2)We can estimate a change in the third derivative as we substitute Y_i for X_i by using the first part of Condition A and several matrix inequalities, valid for any collection of operators. Here Condition A is used only in the proof that the k-th moment of $(\xi_1 + ... + \xi_N)/N^{1/2}$ is bounded as $N \to \infty$.
It is interesting to speculate whether the ideas in this proof can be generalized to the case of the multivariate CLT.

References

[1] Hari Bercovici, Vittorino Pata, and Philippe Biane. Stable laws and domains of attraction in free probability theory. *Annals of Mathematics*, 149:1023–1060, 1999. with an appendix by Philippe Biane. MR1709310

[2] G. Chistyakov and F. Gotze. Limit theorems in free probability theory. I. http://arxiv.org/abs/math.OA/0602219, 2006.

[3] Kai Lai Chung. *A Course in Probability Theory*. Academic Press, third edition, 2001. MR1796326

[4] Thierry Fack. Sur la notion de valeur caracteristique. *Journal of Operator Theory*, 7:307–333, 1982. MR0658616

[5] I. C. Gohberg and M. G. Krein. *Introduction to the Theory of Linear Nonselfadjoint Operators in Hilbert Space*, volume 18 of *Translations of Mathematical Monographs*. American Mathematical Society, 1969. MR0246142

[6] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225, 1922. MR1544569

[7] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*, volume 89 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, New York, 1977. MR0577064

[8] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Text in Mathematics*. Springer, second edition, 1995. MR0737192

[9] Masamichi Takesaki. *Theory of Operator Algebras*, volume 1. Springer-Verlag, New York, 1979. MR0548728

[10] D. Voiculescu. Lectures on free probability theory. In P. Bernard, editor, *Lectures on Probability Theory and Statistics*, volume 1738 of *Lecture Notes in Mathematics*, pages 279–349. Springer, 1998. MR1775641

[11] Dan Voiculescu. Symmetries of some reduced free product C^*-algebras. In *Lecture Notes in Mathematics*, volume 1132, pages 556–588. Springer-Verlag, New York, 1983. MR0799593

[12] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104:201–220, 1991. MR1094052