A TRACIAL QUANTUM CENTRAL LIMIT THEOREM

GREG KUPERBERG

ABSTRACT. We prove a central limit theorem for non-commutative random variables in a von Neumann algebra with a tracial state: Any non-commutative polynomial of averages of i.i.d. samples converges to a classical limit. The proof is based on a central limit theorem for ordered joint distributions together with a commutator estimate related to the Baker-Campbell-Hausdorff expansion. The result can be considered a generalization of Johansson's theorem on the limiting distribution of the shape of a random word in a fixed alphabet as its length goes to infinity.

1. INTRODUCTION

One of the most important results in mathematics for science is the central limit theorem. But there is only an incomplete theory of central limits in the setting of quantum probability theory, even though the quantum rules of probability describe physical reality just as the classical rules do. In this paper we will prove a quantum central limit theorem for tracial states. Our result is not as sharp or as general as what one may conjecture, but it is sharper than previous central limit theorems in the same setting [3, 4, 5, 12].

The difference between quantum and classical central limits only appears in the multivariate case. In quantum probability theory, a real-valued random variable is expressed by a self-adjoint operator in a von Neumann algebra (see below). Quantum behavior only arises with non-commuting operators. In particular independent variables do commute, so the proof of the classical central limit theorem applies to independent, identically distributed (i.i.d.) real-valued quantum random variables.

The classical theorem does not apply to i.i.d. samples of a vector of random variables with non-commuting components. For example, let A and B be two non-commuting measurements in a quantum probability space $\mathcal{M}$, and let $\widetilde{A}$ and $\widetilde{B}$ be the corresponding averaged measurements in the larger space $\mathcal{M}^{\otimes N}$ consisting of N independent copies of $\mathcal{M}$. We would like to say that the joint distribution of $\widetilde{A}$ and $\widetilde{B}$ converges to a Gaussian distribution. But since $\widetilde{A}$ and $\widetilde{B}$ are not simultaneously measurable, it is not clear what this means. One physically observable interpretation is that any suitable self-adjoint expression, such as the anti-commutator $\widetilde{A}\widetilde{B} + \widetilde{B}\widetilde{A}$, should approach a classical distribution.

Our result follows this interpretation. To state it precisely we introduce some notation. Let $(\mathcal{M}, \rho)$ be a quantum probability space consisting of a von Neumann

algebra $\mathcal{M}$ and a normal state ρ. Then $(\mathcal{M}^{\otimes N}, \rho^{\otimes N})$ denotes N independent copies of $(\mathcal{M}, \rho)$. If $A \in \mathcal{M}$, then

$$A^{(k)} = I^{\otimes k-1} \otimes A \otimes I^{\otimes N-k}$$

denotes the kth copy of A in $\mathcal{M}^{\otimes N}$, while

$$\widetilde{A} = \frac{A^{(1)} + A^{(2)} + \cdots + A^{(N)}}{\sqrt{N}}$$

is the rescaled sum of N independent samples of A. Finally if A is a quantum random variable, $\mathcal{L}[A]$ denotes its distribution, or spectral measure.

Theorem 1. *Let $(\mathcal{M}, \rho)$ be a quantum probability space with a tracial state ρ, and let $A_1, A_2, \ldots, A_a$ be self-adjoint elements with mean 0. Let $p \in \mathbb{C}\langle A_1, A_2, \ldots, A_a \rangle$ be a self-adjoint non-commutative polynomial in k variables. Then*

$$\lim_{N \to \infty} \mathcal{L}[p(\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_a)] = \mathcal{L}[p(X_1, X_2, \ldots, X_a)],$$

where $X_1, X_2, \ldots, X_a$ are classical Gaussian random variables with covariance matrix

$$E[X_j X_k] = \rho(A_j A_k).$$

We briefly review the terminology and notation in Theorem 1 and the discussion above [8, 9]. A *von Neumann algebra* is an algebra of bounded operators on some Hilbert space which is both $*$-closed and closed in the strong (equivalently weak) operator topology. A *state* ρ is a bounded, $*$-invariant functional on $\mathcal{M}$ which is non-negative on positive elements of $\mathcal{M}$; it is called *normal* if it is continuous with respect to the strong (equivalently weak) operator topology. A state ρ is *tracial* if

$$\rho(AB) = \rho(BA)$$

for all $A, B \in M$. If $A \in \mathcal{M}$ is self-adjoint, then any normal state ρ on $\mathcal{M}$ induces a measure on the spectrum of A, the (scalar-valued) *spectral measure* on A. Thus we can interpret A as a real-valued random variable; this is the Copenhagen or Born interpretation in physics. If $\mathcal{M}$ is commutative, then it is analogous to a σ-algebra, while ρ is analogous to a measure on the algebra.

In physics terminology, $\mathcal{M}$ is a suitably complete calculus of operators on a quantum system. The state ρ then corresponds to a density operator or an ensemble, although the mathematical convention is to express it as a linear functional on operators rather than as an operator; that is, $\rho(A)$ denotes the expectation or mean of A with respect to ρ, often written $\text{Tr}(\rho A)$ in physics.

The ring $\mathbb{C}\langle A_1, \ldots, A_a \rangle$ of non-commutative polynomials admits a unique $*$-involution which fixes each A_n and which, as usual, is anti-linear and an algebra anti-automorphism. A polynomial p is self-adjoint if $p^* = p$. For example, the commutator $[A_1, A_2]$ is anti-self-adjoint, but $i[A_1, A_2]$ is self-adjoint.

Readers who are not interested in general von Neumann algebras can consider the special case $\mathcal{M} = \mathcal{M}_k$, the $k \times k$ matrices, and

$$\rho(A) = \frac{\text{Tr}(A)}{k}.$$

In this case each A_n is a $k \times k$ Hermitian matrix. We interpret a Hermitian matrix A as a random variable by the formula

$$P[A = \lambda] = \frac{m}{k}$$

if λ is an eigenvalue of A with multiplicity m. This special case captures the difficulties of the general case. We refer readers to Nielsen and Chuang [11] or Sakurai [13] for introductions to quantum probability theory.

With these definitions we can discuss the limitations of Theorem 1. Possibly the most serious one is that ρ must be tracial. We previously conjectured Theorem 1 without assuming that ρ is tracial, only that the covariance matrix of $A_1, A_2, \ldots, A_a$ is symmetric [10]. Goderis, Verbeure, and Vets found a quantum central limit for any state ρ, namely a quasi-free state (or, in physics terminology, a product of thermal states of harmonic oscillators) [5]. Thus we conjecture that in Theorem 1 we can let ρ be any normal state if we replace "classical Gaussian random variables" by "quasi-free variables".

Of course Theorem 1 applies to the case where ρ does not symmetrize all of $\mathcal{M}$, only the subalgebra generated by $A_1, \ldots, A_a$, since we can replace $\mathcal{M}$ by this subalgebra. In physics language it suffices for $A_1, \ldots, A_a$ to commute with the density matrix. This occurs, for example, if ρ is a thermal state (also called a Gibbs state or the Boltzmann distribution) of a Hamiltonian H and each A_n is a conserved quantity. It also occurs in the infinite-temperature limit of any system, because in this limit the density matrix approaches the identity.

Another limitation of Theorem 1 is the requirement that each A_n is bounded. By contrast the classical central limit theorem requires only that the covariance matrix is finite. Thus a more satisfying version would allow each A_n to be an unbounded operator affiliated with $\mathcal{M}$, although with enough restrictions that $p(\widetilde{A}_1, \ldots, \widetilde{A}_a)$ is still well defined. Finally the polynomial ring $\mathbb{C}\langle A_1, \ldots, A_a \rangle$ could be replaced by some von Neumann algebra to which $\mathbb{C}\langle A_1, \ldots, A_a \rangle_{\mathrm{sa}}$ is affiliated. This algebra would model measurable non-polynomial expressions in the variables $A_1, \ldots, A_a$; it would be a non-commutative analogue of the algebra $L^\infty(\mathbb{R}^a)$ of bounded, measurable functions on $\mathbb{R}^a$.

Despite limitations, Theorem 1 is useful. For instance, we previously showed that it implies Johansson's theorem on the limiting distribution of the shape of a random word in k letters as the length N goes to infinity [2, 7, 10]. Indeed Theorem 1 can be taken as a generalization of Johansson's theorem.

1.1. **Some previous results.** Many results in the literature are called or could be called quantum central limit theorems. Here we discuss a few that apply to a discrete set of independent, non-commuting random variables, as Theorem 1 does.

Cushen and Hudson proved the quantum central limit theorem for a pair of conjugate variables P and Q, i.e., such that $[P, Q] = iI$ [3]. Quaegebeur later generalized this result to arbitrary CCR algebras [12].

Giri and von Waldenfels proved that the general non-commutative moment $\rho(A_1 A_2 \cdots A_a)$ converges to a Gaussian or quasi-free value [4]. This can be considered an "algebraic" quantum central limit theorem because it applies to any $\ast$-algebra.

Goderis, Verbeure, and Vets established the convergence of $\rho^{\otimes N}(e^{i\widetilde{A}})$ for all $A \in \mathcal{M}$ to the characteristic functions of a quasi-free state of a universal CCR-algebra over $\mathcal{M}$ [5].

If we compare these three results and Theorem 1 to each other, none is eclipsed by the others. As implied above, we conjecture that there is a mutual generalization of all four results in the von Neumann algebra setting.

Finally Voiculescu and many followers have developed a non-commutative probability theory in which conventional statistical independence is replaced by free independence in the sense of non-commutative free products [14]. This is a very interesting theory which may yet be useful in physics, but the results in this paper are not part of it.

2. Outline of the proof

We use $\mathcal{M}_{\text{sa}}$ to denote the space of self-adjoint elements of $\mathcal{M}$. We abbreviate $\vec{A} = (A_1, A_2, \ldots, A_a)$. We also abbreviate

$$\vec{A}^{\vec{\alpha}} = (A_1^{\alpha_1}, A_2^{\alpha_2}, \ldots, A_a^{\alpha_a}),$$

and we use $\vec{x} \cdot \vec{y}$ to denote the standard inner product on $\mathbb{R}^a$.

Our proof of Theorem 1 is based on a multivariate generalization of the spectral measure $\mathcal{L}[A]$ that we call the ordered joint distribution

$$\mathcal{L}[A_1, A_2, \ldots, A_a] = \mathcal{L}[\vec{A}].$$

It can be defined by its Fourier transform

$$\widehat{\mathcal{L}}[\vec{A}](\vec{t}) = \rho(e^{it_1 A_1} e^{it_2 A_2} \cdots e^{it_a A_a}).$$

The ordered joint distribution is not directly observable, among other reasons because it allows negative and even non-real "probabilities".

As a first step, an essentially classical central limit theorem holds for ordered joint distributions. Consequently for any positive exponents $\alpha_1, \alpha_2, \ldots, \alpha_a$, the distribution

$$\mathcal{L}[\widetilde{A}_1^{\alpha_1}, \widetilde{A}_2^{\alpha_2}, \ldots, \widetilde{A}_a^{\alpha_a}] = \mathcal{L}[\widetilde{\vec{A}}^{\vec{\alpha}}]$$

also approaches a classical limit. When ρ is tracial, the difference between an ordered joint characteristic function,

$$\widehat{\mathcal{L}}[\vec{A}](\vec{t}) = \rho(e^{it_1 \widetilde{A}_1^{\alpha_1}} e^{it_2 \widetilde{A}_2^{\alpha_2}} \cdots e^{it_a \widetilde{A}_a^{\alpha_a}}),$$

and the characteristic function of a linear combination of powers of $\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_a$,

$$\rho(e^{i(t_1 \widetilde{A}_1^{\alpha_1} + t_2 \widetilde{A}_2^{\alpha_2} + \cdots + t_a \widetilde{A}_a^{\alpha_a})}) = \rho(e^{i\vec{t} \cdot \widetilde{\vec{A}}^{\vec{\alpha}}}),$$

is bounded by a decaying commutator estimate. Thus the spectral characteristic functions converge pointwise to a classical limit when $p(\vec{A})$ is a linear combination of powers. This implies many special cases of Theorem 1, indeed every case up to a change of variables.

3. The proof

The proof is more natural in the C^*-algebra setting than in the von Neumann algebra setting. Recall that a C^*-algebra is a special kind of complex Banach algebra and that every von Neumann algebra is a C^*-algebra [8]. C^*-algebras have states, but they do not have normal states in the absence of weak and strong operator topologies. By the GNS construction, given any state ρ on a C^*-algebra $\mathcal{A}$, there exists a von Neumann algebra $\mathcal{M} \supseteq \mathcal{A}$ and an extension of ρ which is normal on $\mathcal{M}$. Henceforth we replace the von Neumann algebra $\mathcal{M}$ by a C^*-algebra $\mathcal{A}$ and drop the inessential condition that ρ is normal. (This is analogous to considering probability distributions on topological spaces instead of measure spaces.)

We call a norm $\|\cdot\|_S$ on a C^*-algebra $\mathcal{A}$ *spectral* if it is $*$-invariant and if

$$\|UA\|_S = \|A\|_S$$

for any unitary U. We consider the GNS norm

$$\|A\|_\rho = \sqrt{\rho(A^*A)}$$

on $\mathcal{A}$, which is spectral when ρ is tracial. (If ρ is not faithful, it is only a semi-norm.) By the Russo-Dye theorem [9],

$$\|AB\|_S \leq \min\{\|A\|_S\|B\|, \|A\|\,\|B\|_S\}$$

in any spectral norm $\|\cdot\|_S$. In particular

$$\|A\|_S \leq \|1\|_S\|A\|,$$

so the topology induced by $\|A\|_S$ is at least as coarse as the norm topology.

If $\mathcal{A}$ happens to be a von Neumann algebra and $A \in \mathcal{A}_{\mathrm{sa}}$ has a discrete spectrum, then its spectral measure $\mathcal{L}[A]$ is given by the rule

$$P[A = \lambda] = \rho(A_\lambda),$$

where λ is an eigenvalue of A and A_λ denotes projection onto its λ-eigenspace. If $A_1, A_2, \ldots, A_a \in \mathcal{A}_{\mathrm{sa}}$ all have discrete spectra, the *ordered joint distribution* $\mathcal{L}[\vec{A}]$ is likewise given by

$$P[\vec{A} = \vec{\lambda}] = \rho((A_1)_{\lambda_1}(A_2)_{\lambda_2}\cdots(A_a)_{\lambda_a}).$$

This is generally a complex-valued measure rather than a non-negative measure. Without the discrete spectrum assumption, $\mathcal{L}[\vec{A}]$ might not strictly be a measure at all. Recall that a probability measure on $\mathbb{R}^a$ can be regarded as a bounded functional on the space $C_0(\mathbb{R}^a)$ of continuous functions on $\mathbb{R}^a$ that vanish at infinity. In this paper we define $\mathcal{L}[\vec{A}]$ as a function on the space of products

$$f(\vec{x}) = f_1(x_1)f_2(x_2)\cdots f_a(x_a)$$

of decaying, continuous, univariate functions by the rule

$$\mathcal{L}[\vec{A}](f) = \rho(f_1(A_1)f_2(A_2)\cdots f_a(A_a)).$$

Since ρ is bounded, $\mathcal{L}[\vec{A}]$ is a bounded functional on the projective Banach tensor product $C_0(\mathbb{R})^{\otimes a}$ (i.e., the minimal completion of the algebra tensor product, or the completion with respect to the greatest cross norm). In order to be a genuine measure, $\mathcal{L}[\vec{A}]$ would have to extend from $C_0(\mathbb{R})^{\otimes a}$ to the much larger Banach space $C_0(\mathbb{R}^a)$. This is not always possible, although $\mathcal{L}[\vec{A}]$ might naturally extend to some intermediate Banach space. The distribution $\mathcal{L}[\vec{A}]$ is also sensitive to permutations of the variables. But it otherwise reasonably generalizes the usual joint distribution of classical random variables.

Remark. For conjugate variables P and Q, the ordered joint distribution is related to the Wigner distribution. The Fourier transform of the Wigner distribution is given by

$$\widehat{W}(q,p) = \rho(e^{iPq/2}e^{iQp}e^{iPq/2}).$$

We can then compare

$$\widehat{\mathcal{L}}[P,Q](q,p) = \rho(e^{iPq}e^{iQp}) = e^{-ipq/2}\widehat{W}(q,p),$$
$$\widehat{\mathcal{L}}[Q,P](p,q) = \rho(e^{iQp}e^{iPq}) = e^{ipq/2}\widehat{W}(q,p).$$

Thus the Fourier transform of the Wigner distribution is a phase-corrected form of the Fourier transform of the ordered joint distribution.

Theorem 2. *If $\mathcal{A}$ is a C^*-algebra with a tracial state ρ and $A_1, A_2, \ldots, A_a \in \mathcal{A}_{\mathrm{sa}}$ with mean zero, then their ordered joint distribution $\mathcal{L}[\vec{A}]$ obeys the central limit theorem*:

$$\lim_{N \to \infty} \mathcal{L}[\tilde{\vec{A}}] = \mathcal{L}[\vec{X}]$$

weakly as functionals on $C_0(\mathbb{R})^{\otimes a}$, and $\vec{X}$ is Gaussian with covariance matrix

$$M_{j,k} = E[X_j X_k] = \rho(A_j A_k).$$

Proof. We follow standard proofs of the central limit theorem by the method of characteristic functions [1], but the argument must be applied carefully because $\mathcal{L}[\vec{A}]$ is a complex measure. We will argue convergence of $\mathcal{L}[\vec{A}]$ on four classes of functions in turn:

1. products of sinusoids,
2. products of smooth functions with bounded support,
3. products of bounded, smooth functions, and
4. products of bounded, continuous functions.

Observe that if $B_1, B_2, \ldots, B_a$ are self-adjoint elements in another C^*-algebra $\mathcal{B}$ with a state τ and

$$A_n^{(1)} = A_n \otimes I \in \mathcal{A} \otimes \mathcal{B}, \qquad B_n^{(2)} = I \otimes B_n \in \mathcal{A} \otimes \mathcal{B},$$

then

$$\mathcal{L}[\vec{A}^{(1)} + \vec{B}^{(2)}] = \mathcal{L}[\vec{A}] * \mathcal{L}[\vec{B}],$$

where the operation $*$ denotes convolution. Thus

$$(1) \qquad \mathcal{L}[\tilde{\vec{A}}] = T_{1/\sqrt{N}}(\mathcal{L}[\vec{A}] * \mathcal{L}[\vec{A}] * \cdots * \mathcal{L}[\vec{A}]),$$

where T_x denotes the operator on functionals induced by rescaling space by a factor of x. Equation (1) implies convergence of characteristic functions:

$$\widehat{\mathcal{L}}[\tilde{\vec{A}}](\vec{t}) = \rho^{\otimes N}(e^{i\vec{t} \cdot \tilde{\vec{A}}}) \to e^{-\vec{t} \cdot M \vec{t}/2}$$

locally uniformly in $\vec{t}$.

Suppose that

$$f(\vec{t}) = f_1(t_1) f_2(t_2) \cdots f_a(t_a)$$

is a product of smooth univariate functions with bounded support. Its Fourier transform $\hat{f}$ decays super-polynomially. Consider the joint expectation

$$\mathcal{L}[\tilde{\vec{A}}](f) = (2\pi)^{-a/2} \int_{\mathbb{R}^a} \widehat{\mathcal{L}}[\tilde{\vec{A}}](\vec{t}) \overline{\hat{f}(\vec{t})} d\vec{t}$$

as an integral in Fourier space. For any $R > 0$, the integral converges inside the box $[-R, R]^a$ because there the integrand converges uniformly in $\vec{t}$. On the other hand, the integrand vanishes outside of the box uniformly in N as $R \to \infty$ because $\hat{f}$ decays and, for all $\vec{t}$,

$$|\widehat{\mathcal{L}}[\tilde{\vec{A}}](\vec{t})| = |\rho^{\otimes N}(e^{i\vec{t} \cdot \tilde{\vec{A}}})| \leq 1.$$

(Both $\rho^{\otimes N}$ and its argument have norm 1.) Thus for these f,

$$\mathcal{L}[\tilde{\vec{A}}](f) \to \mathcal{L}[\vec{X}](f).$$

Now suppose that $f_1, f_2, \ldots, f_a$ have bounded support but are merely continuous. By the Weierstrass approximation theorem, for each $\varepsilon > 0$, we can let

$$f_n(t) = f_n^{(1)}(t) + f_n^{(2)}(t),$$

where $f_n^{(1)}(t)$ is smooth and

$$\|f_n^{(2)}\| \leq \varepsilon.$$

Then

$$\mathcal{L}[\widetilde{\vec{A}}](f) = \sum_{\vec{\sigma} \in \{1,2\}^a} \rho^{\otimes N}(f^{(\sigma_1)}(\widetilde{A}_1) f_2^{(\sigma_2)}(\widetilde{A}_2) \cdots f_a^{(\sigma_a)}(\widetilde{A}_a)).$$

Each term of the sum other than the first one is bounded by ε, while the first term converges to $\mathcal{L}[\vec{X}](f)$ if we take $\varepsilon \to 0$. Thus for continuous f with bounded support,

$$\mathcal{L}[\widetilde{\vec{A}}](f) \to \mathcal{L}[\vec{X}](f).$$

Finally, suppose that $f_1, f_2, \ldots, f_a$ are bounded but do not have bounded support. Without loss of generality we suppose that

$$\|f_n\| \leq 1$$

for each n. For each $R > 0$, we can choose a continuous partition

$$f_n(t) = f_n^{(1)}(t) + f_n^{(2)}(t),$$

where

$$f_n^{(1)}(t) = \min\{\max\{0, |t| - R\}, 1\} f_n(t).$$

Thus

$$f_n(t) = \begin{cases} f_n^{(1)}(t), & |t| \leq R, \\ f_n^{(2)}(t), & |t| \geq R + 1. \end{cases}$$

By the univariate central limit theorem,

$$\lim_{R \to \infty} \|f_n^{(2)}(\widetilde{A}_n)\|_{\rho^{\otimes N}} = 0$$

uniformly in N. Let $\varepsilon > 0$ and choose R such that for each n,

$$\|f_n^{(2)}(\widetilde{A}_n)\|_{\rho^{\otimes N}} < \varepsilon.$$

Since for each n,

$$\|f_1^{(2)}(\widetilde{A}_n)\| \leq 1,$$

and since $\| \cdot \|_{\rho^{\otimes N}}$ is spectral, each term other than the first in the expansion

$$\mathcal{L}[\widetilde{\vec{A}}](f) = \sum_{\vec{\sigma} \in \{1,2\}^a} \rho^{\otimes N}(f_1^{(\sigma_1)}(\widetilde{A}_1) f_2^{(\sigma_2)}(\widetilde{A}_2) \cdot f_a^{(\sigma_a)}(\widetilde{A}_a))$$

is bounded by ε. (The $\rho^{\otimes N}$-norm of the argument in each term is bounded by ε.) As before, the first term converges to $\mathcal{L}[\vec{X}](f)$ if we take $\varepsilon \to 0$. Thus for all f described in the theorem,

$$\mathcal{L}[\widetilde{\vec{A}}](f) \to \mathcal{L}[\vec{X}](f). \qquad \square$$

Lemma 3. *If A, B are elements of a C^*-algebra $\mathcal{A}$, then*

$$e^A e^B - e^{A+B} = \int_0^1 \int_0^{1-t} e^{t(A+B)} e^{(1-t-s)A} [A, B] e^{sA} e^{(1-t)B} ds \, dt.$$

Proof. Let

$$X = e^{A/n}, \qquad Y = e^{B/n}.$$

By elementary calculation,

$$(2) \qquad X^n Y^n - (XY)^n = \sum_{t=1}^{n-1} \sum_{s=0}^{n-t-1} (XY)^{t-1} X^{n-t-s} [X,Y] X^s Y^{n-t}.$$

Since

$$\lim_{n\to\infty} X^{tn} = e^{tA}, \qquad\qquad\qquad \lim_{n\to\infty} Y^{tn} = e^{tB},$$
$$\lim_{n\to\infty} (XY)^{tn} = e^{t(A+B)}, \qquad\qquad \lim_{n\to\infty} n^2 [X,Y] = [A,B],$$

equation (2) converges to the statement of the lemma as n goes to infinity. $\qquad\square$

Lemma 4. *If $A_1, A_2, \ldots, A_a$ are self-adjoint elements of a C^*-algebra $\mathcal{A}$ and $\|\cdot\|_S$ is a spectral semi-norm, then*

$$\|e^{iA_1} e^{iA_2} \cdots e^{iA_a} - e^{i(A_1 + A_2 + \cdots + A_a)}\|_S \le \sum_{1 \le j < k \le a} \frac{\|[A_j, A_k]\|_S}{2}.$$

Proof. If $a = 2$, then Lemma 3 implies that

$$\|e^{iA_1} e^{iA_2} - e^{i(A_1 + A_2)}\|_S \le \frac{\|[A_1, A_2]\|_S}{2}$$

by taking the norm inside the integral. Note that the integral is defined by convergence in the algebra norm, but we can exchange it with a weaker norm or semi-norm. Every spectral semi-norm $\|\cdot\|_S$ is at least as weak as the algebra norm.

The general case follows by induction. $\qquad\square$

Let $\mathbb{C}\{\vec{A}\}$ denote the Lie algebra freely generated by the symbols $A_1, A_2, \ldots, A_a$. Recall that its universal enveloping algebra is the ring of non-commutative polynomials in the same variables:

$$U(\mathbb{C}\{\vec{A}\}) = \mathbb{C}\langle \vec{A} \rangle.$$

Lemma 5. *If X and Y commute, $\alpha, \beta \in \mathbb{Z}_{\ge 0}$, and $q > 1$, then*

$$X^\alpha Y^\beta = \sum_{n=0}^{\alpha+\beta} t_n (X + q^n Y)^{\alpha+\beta}$$

for some $t_0, t_1, \ldots, t_{\alpha+\beta} \in \mathbb{R}$.

Proof. The coefficients $t_0, t_1, \ldots, t_{\alpha+\beta}$ must satisfy the linear system

$$\sum_{n=0}^{\alpha+\beta} q^{nk} = \begin{cases} \frac{1}{\binom{\alpha+\beta}{\alpha}}, & k = \alpha, \\ 0, & k \ne \alpha, \end{cases}$$

for $0 \le k \le \alpha + \beta$. The matrix of this system is a Vandermonde matrix, invertible when $q > 1$, so the system has a solution. $\qquad\square$

Lemma 6. *Any $p \in \mathbb{C}\langle \vec{A} \rangle$ can be expressed as a linear combination of power sums*

$$p(\vec{A}) = t_1 B_1^{\beta_1} + t_2 B_2^{\beta_1} + \cdots + t_b B_b^{\beta_b} = \vec{t} \cdot \vec{B}^{\vec{\beta}}$$

with each $B_n \in \mathbb{C}\{A_1, A_2, \ldots, A_a\}$ and each $\beta_n \in \mathbb{Z}_{\ge 0}$. If p is self-adjoint, then we can take each $t_n \in \mathbb{R}$ and each B_n to be self-adjoint.

Proof. The Poincaré-Birkhoff-Witt theorem provides a vector space isomorphism

$$\Phi : S(L) \to U(L)$$

from the symmetric algebra to the universal enveloping algebra of any Lie algebra L [6]. The map Φ is given by symmetrization:

$$\Phi(X_1 X_2 \cdots X_n) = \frac{1}{n!} \sum_{\pi \in S_n} X_{\pi(1)} X_{\pi(2)} \cdots X_{\pi(n)}$$

for any $X_1, X_2, \ldots, X_n \in L$. If L has an anti-involution $*$, then it extends to both $S(L)$ and $U(L)$ and Φ intertwines it. The map Φ also preserves powers of elements of L:

$$\Phi(X^n) = X^n.$$

It therefore also preserves linear combinations of powers.

We claim that if L is a vector space, then every element of the symmetric algebra $S(L)$ is a linear combination of powers. Lemma 5 establishes the special case that the product of two powers is a linear combination of powers. The set P of linear combinations of powers is therefore closed under multiplication; it is a (complex) subalgebra of $S(L)$. Since $L \subset P$ trivially and L generates $S(L)$, we conclude that $P = S(L)$. If L has an anti-linear $*$-involution, then the same argument applies to $S(L)_{\mathrm{sa}}$, which is generated as a real algebra by the real vector space L_{sa}.

In conclusion, for any Lie algebra L, powers of elements of L span $U(L)$ as a complex vector space. If L has an anti-involution $*$, then powers of elements of L_{sa} span $U(L)_{\mathrm{sa}}$ as a real vector space. The lemma is the special case

$$L = \mathbb{C}\{A_1, A_2, \ldots, A_a\}. \qquad \square$$

The proof of Theorem 1 is simpler if the vector of Lie elements $\vec{B}$ provided by Lemma 6 is linear in $\vec{A}$. (This includes many interesting choices for p, for instance the anti-commutator $A_1 A_2 + A_2 A_1$.) Indeed in this case

$$p(\vec{\widetilde{A}}) = \vec{t} \cdot \vec{\widetilde{B}}^{\vec{\beta}}$$

is a polynomial in $\vec{\widetilde{B}}$, so we could replace $\vec{A}$ with $\vec{B}$ in the statement of the theorem.

Lemma 7. *If $A, B \in \mathcal{A}_{\mathrm{sa}}$ have mean 0 and $\alpha, \beta \in \mathbb{Z}_{\geq 0}$, then*

$$\|[\vec{A}^\alpha, \vec{B}^\beta]\|_{\rho^{\otimes N}} = O(N^{-1/2})\|A\|\,\|B\|.$$

Before proving Lemma 7 in full generality, we motivate it with a simple proof when $\alpha = \beta = 1$. In this case

$$[\widetilde{A}^\alpha, \widetilde{B}^\beta] = \frac{1}{\sqrt{N}}\widetilde{[A, B]}.$$

Since $[A, B]$ has mean 0 (because ρ is tracial), the typical eigenvalue of $\widetilde{[A, B]}$ is $O(1)$. Thus the norm of the right side is $O(N^{-1/2})$.

Proof. In brief, since

$$\rho(A) = \rho(B) = \rho([A, B]) = 0,$$

and since ρ is tracial, the multilinear expansion of

$$\rho^{\otimes N}([\widetilde{A}^\alpha, \widetilde{B}^\beta]^2) = \pm\|[\widetilde{A}^\alpha, \widetilde{B}^\beta]\|^2_{\rho^{\otimes N}}$$

has $O(N^{\alpha+\beta-1})$ non-cancelling terms. Since the expansion also has a factor of $N^{\alpha+\beta}$ in the denominator, the square of the norm is $O(N^{-1})$.

In detail, the expansion is

(3)

$$N^{\alpha+\beta}\rho^{\otimes N}([\widetilde{A}^\alpha, \widetilde{B}^\beta]^2) = \sum_{\vec{X},\vec{Y},\vec{Z},\vec{W}} \left(\prod_n \rho(X_n Y_n Z_n W_n) - \prod_n \rho(Y_n X_n Z_n W_n) \right.$$

$$\left. - \prod_n \rho(X_n Y_n W_n Z_n) + \prod_n \rho(Y_n X_n W_n Z_n) \right),$$

where each vector $\vec{X}$ and $\vec{Z}$ consists of α copies of A and $N-\alpha$ copies of I, and each vector $\vec{Y}$ and $\vec{W}$ consists of β copies of B and $N-\beta$ copies of I. Momentarily fix $\vec{X},\vec{Y},\vec{Z},\vec{W}$ and consider the corresponding four terms in equation (3). If for some n, exactly one of X_n, Y_n, Z_n, W_n is A or B, then all four terms vanish. Likewise if for all n, not all four of X_n, Y_n, Z_n, W_n are A or B, then the four terms cancel. Thus in a non-cancelling choice of the vectors $\vec{X},\vec{Y},\vec{Z},\vec{W}$, the four components X_n, Y_n, Z_n, W_n are non-trivial for most $\alpha + \beta - 1$ values of n. There are only $O(N^{\alpha+\beta-1})$ such terms, each bounded by $\|A\|^2\|B\|^2$.   □

Applying Theorem 2 to $\vec{B}$, we learn not only that

$$\lim_{N\to\infty} \mathcal{L}[\widetilde{\vec{B}}] = \mathcal{L}[\vec{Y}]$$

with $\vec{Y} = (Y_1, Y_2, \ldots, Y_b)$ Gaussian, but also that

$$\lim_{N\to\infty} \mathcal{L}[\widetilde{B}^{\vec{\beta}}] = \mathcal{L}[\vec{Y}^{\vec{\beta}}].$$

Thus for every $\vec{t}$,

(4) $$\widehat{\mathcal{L}}[\widetilde{B}^{\vec{\beta}}](\vec{t}) = \rho^{\otimes N}(e^{t_1 \widetilde{B}_1^{\beta_1}} e^{it_2 \widetilde{B}_2^{\beta_2}} \cdots e^{it_b \widetilde{B}_b^{\beta_b}}) \to \widehat{\mathcal{L}}[\vec{Y}^{\vec{\beta}}](\vec{t})$$

approaches a classical limit as $N \to \infty$. Combining Lemma 4 with Lemma 7,

(5) $$\|e^{it_1 \widetilde{B}_1^{\beta_1}} e^{it_2 \widetilde{B}_2^{\beta_2}} \cdots e^{it_b \widetilde{B}_b^{\beta_b}} - e^{i\vec{t}\cdot\widetilde{\vec{B}}^{\vec{\beta}}}\|_{\rho^{\otimes N}} = O(N^{-1/2}).$$

Finally combining equations (4) and (5) with

$$\|\rho^{\otimes N}\|_{\rho^{\otimes N}} = 1,$$

we obtain

$$\rho^{\otimes N}(e^{i\vec{t}\cdot\widetilde{\vec{B}}^{\vec{\beta}}}) \to \widehat{\mathcal{L}}[\vec{Y}^{\vec{\beta}}](\vec{t}).$$

Replacing $\vec{t}$ by $z\vec{t}$, we obtain

$$\widehat{\mathcal{L}}[p(\widetilde{\vec{A}})](z) = \rho^{\otimes N}(e^{iz\vec{t}\cdot\widetilde{\vec{B}}^{\vec{\beta}}}) \to \widehat{\mathcal{L}}[\vec{t}\cdot\vec{Y}^{\vec{\beta}}](z).$$

Since $p(\vec{X}) = \vec{t}\cdot\vec{Y}^{\vec{\beta}}$ and pointwise convergence of characteristic functions implies weak convergence of measures, we have established Theorem 1 when $\vec{B}$ is linear.

The idea behind the general case is that the non-linear terms in each B_n decay as $N \to \infty$. We expand each B_n as a sum of homogeneous terms:

$$B_n = B_{n,1} + B_{n,2} + \cdots + B_{n,d_n},$$

where $B_{n,d}$ is the degree d term of B_n. Then

$$p(\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_a) = t_1 \widetilde{C}_1^{\beta_1} + t_2 \widetilde{C}_2^{\beta_2} + \cdots + t_b \widetilde{C}_b^{\beta_b},$$

where
$$C_n = B_{n,1} + N^{-1/2} B_{n,2} + \cdots + N^{(1-d_n)/2} B_{n,d_n}.$$

Furthermore
$$\rho(B_{n,1}) = \rho(B_{n,2}) = 0$$

in the first case because $B_{n,1}$ is linear, and in the second case because $B_{n,2}$ is a commutator and ρ is tracial. It follows that
$$\lim_{N\to\infty} \rho(\widetilde{C}_n) = 0.$$

Lemma 8.
$$\lim_{N\to\infty} \mathcal{L}[\vec{\widetilde{C}}] = \mathcal{L}[\vec{X}],$$

where $\vec{X}$ is Gaussian with covariance matrix
$$M_{j,k} = E[X_j X_k] = \rho(B_{j,1} B_{k,1}).$$

In other words,
$$\mathcal{L}[\widetilde{C}_1, \widetilde{C}_2, \ldots, \widetilde{C}_b]$$

approaches the same classical limit as
$$\mathcal{L}[\widetilde{B}_{1,1}, \widetilde{B}_{2,1}, \ldots, \widetilde{B}_{b,1}].$$

Proof. Observe that
$$\lim_{N\to\infty} C_n = B_{n,1}.$$

The lemma follows from the proof of Theorem 2, where it was left unstated that all estimates are locally uniform in $\vec{A}$. We substitute $\vec{C} - \rho(\vec{C})$ for $\vec{A}$ in the theorem and use the fact that $\rho(\widetilde{C}_n) \to 0$. $\qquad\square$

The rest of the proof of Theorem 1 follows the same argument as the case when $\vec{B}$ is linear by substituting $\vec{C}$ for $\vec{B}$.

4. Not the proof

To understand the proof of Theorem 1, it may help to see why some alternative lines of argument do not suffice.

The Giri-von-Waldenfels central limit theorem implies that the moments of $p(\vec{\widetilde{A}})$ converge to the moments of $p(\vec{X})$ by multilinear expansion. However, unless p is either linear or positive-definite quadratic, $p(\vec{X})$ is not uniquely determined by its moments because the tail of $\mathcal{L}[p(\vec{X})]$ is too thick. However, it is yet possible that Theorem 1 would follow from the Giri-von-Waldenfels theorem together with an analytic theory of non-commutative moments.

The Goderis-Verbeure-Vets central limit theorem establishes the convergence of any
$$\rho^{\otimes N}(e^{i\vec{t}\cdot\vec{\widetilde{A}}}),$$

which can be interpreted as a joint characteristic of $\vec{\widetilde{A}}$. Indeed Theorem 2 more generally establishes the convergence of
$$\rho^{\otimes N}(e^{i\vec{t}_1\cdot\vec{\widetilde{A}}} e^{i\vec{t}_2\cdot\vec{\widetilde{A}}} \cdots e^{i\vec{t}_b\cdot\vec{\widetilde{A}}})$$

for any sequence of vectors $\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_b$. However, absent an analytic theory of non-commutative characteristic functions, this does not imply Theorem 1.

The Baker-Campbell-Hausdorff expansion also shows that

$$e^{it_1 \widetilde{B}_1^{\beta_1}} e^{it_2 \widetilde{B}_2^{\beta_2}} \cdots e^{it_b \widetilde{B}_b^{\beta_b}}$$

approximates

$$e^{i(t_1 \widetilde{B}_1^{\beta_1} + t_2 \widetilde{B}_2^{\beta_2} + \cdots + t_b \widetilde{B}_b^{\beta_b})}.$$

Unfortunately the infinite sum of the BCH expansion does not (as far as we know) commute with the limit $N \to \infty$.

If ρ is not tracial, the proof of Lemma 4 fails for the non-spectral norm

$$\|A\|_\rho = \rho(A^* A).$$

Alternatively, if σ is a tracial state, then

$$\|\rho^{\otimes N}\|_{\sigma^{\otimes N}} = \|\rho\|_\sigma^N \to \infty$$

exponentially, assuming that $\mathcal{A}$ even has a tracial state and that $\|\rho\|_\sigma$ is finite. Lemma 3 generalizes to higher-order commutators which decay more and more quickly, but these suffer from the same exchange-of-limits problem as the BCH expansion when combined with the norm of $\rho^{\otimes N}$. Theorem 2 also depends on the assumption that ρ is tracial, but there the assumption may be unnecessary.

Finally when ρ is not tracial, then $\|\cdot\|_\rho$ is at least left-invariant under unitary multiplication. If each commutator $[A_j, A_k]$ is central, then we can move the unitary factors in Lemma 3 to the left and consequently establish Lemma 4 for $\|\cdot\|_\rho$. This proves a slight generalization of Theorem 1. However, it is equivalent to the Cushen-Hudson-Quaegebeur central limit theorem [3, 12], except that they do not require bounded random variables.

Acknowledgments

References

1. Rabi N. Bhattacharya and R. Ranga Rao, *Normal approximation and asymptotic expansions*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York-London-Sydney, 1976. MR **55:**9219

2. Philippe Biane, *Quantum random walk on the dual of* SU(n), Probab. Theory Related Fields **89** (1991), 117–129. MR **93a:**46119

3. C. D. Cushen and R. L. Hudson, *A quantum-mechanical central limit theorem*, J. Appl. Probability **8** (1971), 454–469. MR **44:**6277

4. Narayan C. Giri and Wilhelm von Waldenfels, *An algebraic version of the central limit theorem*, Z. Wahrscheinlichkeitstheorie **42** (1978), 129–134. MR **57:**7731a

5. D. Goderis, A. Verbeure, and P. Vets, *Non-commutative central limits*, Probab. Theory Related Fields **82** (1989), 527–544. MR **91b:**46057

6. James E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics, vol. 9, Springer-Verlag, New York-Heidelberg-Berlin, 1972. MR **48:**2197

7. Kurt Johansson, *Discrete orthogonal polynomial ensembles and the Plancherel measure*, Ann. of Math. (2) **153** (2001), 259–296, arXiv:math.CO/9906120. MR **2002g:**05188

8. Richard V. Kadison and John R. Ringrose, *Fundamentals of the theory of operator algebras*, vol. I, Academic Press, 1983. MR **85j:**46099

9. ———, *Fundamentals of the theory of operator algebras*, vol. II, Academic Press, 1986. MR **88d:**46106

10. Greg Kuperberg, *Random words, quantum statistics, central limits, random matrices*, Methods Appl. Anal. **9** (2002), 99–118.

11. Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.
12. Johan Quaegebeur, *A noncommutative central limit theorem for CCR-algebras*, J. Funct. Anal. **57** (1984), 1–20. MR **85m:**46061
13. Jun John Sakurai, *Modern quantum mechanics*, 2nd ed., Benjamin/Cummings, 1985.
14. Dan Voiculescu, *Lectures on free probability theory*, Lectures on probability theory and statistics (Saint-Flour, 1998), Springer, Berlin, 2000, pp. 279–349. MR **2001g:**46121

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA DAVIS, DAVIS, CALIFORNIA 95616
E-mail address: greg@math.ucdavis.edu