OPERATOR PROBABILITY THEORY

Stan Gudder
Department of Mathematics
University of Denver
Denver, Colorado 80208
stan.gudder@nsm.du.edu

Abstract

This article presents an overview of some topics in operator probability theory. We do not strive for generality and only simple methods are employed. To give the reader a flavor of the subject we concentrate on the two most important topics, the law of large numbers and the central limit theorem.

1 Introduction

This article surveys various aspects of operator probability theory. This type of work has also been called noncommutative probability or quantum probability theory. The main applications are in quantum mechanics, statistical mechanics and quantum field theory. Since the framework deals with Hilbert space operators it is also of interest to operator theorists. The article is directed toward mathematicians who are not experts in this field but who want to learn something about it. For this reason we shall not strive for great generality and shall concentrate on the two important topics, the law of large numbers and the central limit theorem.

A lot of work has been devoted to probability theory on operator algebras such as C^*-algebras and von Neumann algebras [1, 2, 3, 5, 6, 7, 8, 10]. However, to avoid various technicalities we shall only consider the full algebra $\mathcal{B}(H)$ of bounded operators on a Hilbert space H. Also, we shall not discuss important but technically difficult recent research such as noncommutative measure and ergodic theory [2, 4, 5] and free probability theory

[2, 11]. Strictly speaking , the results we shall present are not new. However, we think that our methods are simpler and present a clearer picture. Since some of our definitions are not standard we present some new versions of known results.

2 Notation and Definitions

Let H be a complex Hilbert space and let $\mathcal{B}(H)$ be the set of bounded linear operators on H. We denote the set of bounded self-adjoint operators on H by $\mathcal{S}(H)$ and the set of positive trace class operators with unit trace by $\mathcal{D}(H)$. The trace of a trace class operator T is denoted by $\mathrm{tr}(T)$. We think of $\mathcal{B}(H)$ as a set of (noncommutative) complex-valued random variables and $\mathcal{D}(H)$ as a set of states (probability measures). If $A \in \mathcal{S}(H)$ with spectral measure P^A then

$$P_\rho(A \in \Delta) = \mathrm{tr}\left[\rho P^A(\Delta)\right]$$

is interpreted as the probability that A has a value in the Borel set Δ for the state ρ. It is then natural to call the real probability measure $\Delta \mapsto \mathrm{tr}\left[\rho P^A(\Delta)\right]$ the **distribution** of A in the state ρ. It follows that the expectation $E_\rho(A)$ of A in the state ρ is given by

$$E_\rho(A) = \int \lambda \mathrm{tr}\left[\rho P^A(d\lambda)\right] = \mathrm{tr}(\rho A)$$

An arbitrary $A \in \mathcal{B}(H)$ has the unique representation $A = A_1 + iA_2$ for $A_1, A_2 \in \mathcal{S}(H)$ given by $A_1 = (A + A^*)/2$, $A_2 = (A - A^*)/2i$. We then write $A_1 = \mathrm{Re}(A)$ and $A_2 = \mathrm{Im}(A)$. Then A generates two distributions $\Delta \mapsto P_\rho(A_1 \in \Delta)$ and $\Delta \mapsto P_\rho(A_2 \in \Delta)$. It is natural to define the ρ-**expectation** of A by

$$E_\rho(A) = E_\rho(A_1) + iE_\rho(A_2) = \mathrm{tr}\left[\rho(A_1 + iA_2)\right] = \mathrm{tr}(\rho A)$$

We can now define other probabilistic concepts in the usual way. The ρ-**moments** of A are $\mathrm{tr}(\rho A^n)$, $n = 0, 1, \ldots,$ and the ρ-**variance** of A is

$$\mathrm{Var}_\rho(A) = E_\rho\left[(A - E_\rho(A)I)^2\right] = E_\rho(A^2) - E_\rho(A)^2$$

Notice that $\mathrm{Var}_\rho(A)$ may be complex and we have

$$E_\rho(A^*) = \mathrm{tr}(\rho A^*) = \mathrm{tr}\left[(\rho A)^*\right] = \overline{\mathrm{tr}(\rho A)} = \overline{E_\rho(A)}$$

and hence, $\text{Var}_\rho(A^*) = \overline{\text{Var}_\rho(A)}$.

For $A \in \mathcal{B}(H)$ we write $|A| = (A^*A)^{1/2} \in \mathcal{S}(H)$. The ρ-**absolute variance** of A is

$$\begin{aligned} \left|\text{Var}_\rho\right|(A) &= E_\rho\left[|A - E_\rho(A)I|^2\right] = E_\rho\left[(A - E_\rho(A)I)^*(A - E_\rho(A)I)\right] \\ &= E_\rho\left[A^*A + |E_\rho(A)|^2 I - \overline{E_\rho(A)}A - E_\rho(A)A^*\right] \\ &= E_\rho(A^*A) - |E_\rho(A)|^2 = E_\rho\left(|A|^2\right) - |E_\rho(A)|^2 \end{aligned}$$

Since $\left|\text{Var}_\rho\right|(A) \geq 0$ we conclude that $|E_\rho(A)|^2 \leq E_\rho\left(|A|^2\right)$ or

$$|\text{tr}(\rho A)|^2 \leq \text{tr}\left(\rho|A|^2\right) \tag{2.1}$$

Of course, if $A \in \mathcal{S}(H)$, then $\text{Var}_\rho(A) = \left|\text{Var}_\rho\right|(A)$.

For $x \in H$ with $\|x\| = 1$, we denote the one-dimensional projection onto the span of x by P_x. Of course, $P_x \in \mathcal{D}(H)$ and if $\rho = P_x$ we have $E_\rho(A) = \langle Ax, x\rangle$ and

$$|\text{Var}_\rho(A)| = \left\langle |A|^2x, x\right\rangle - |\langle Ax, x\rangle|^2 = \|Ax\|^2 - |\langle Ax, x\rangle|^2$$

Elements of $\mathcal{D}(H)$ that have the form P_x are called **pure states**. It is well known that any state is a convex combination of pure states, that is, if $\rho \in \mathcal{D}(H)$ then

$$\rho = \sum \lambda_i P_{x_i} \tag{2.2}$$

where $\lambda_i > 0$ with $\sum \lambda_i = 1$. In this case, x_i is an eigenvector of ρ with corresponding eigenvalue λ_i. Our first result shows that $\left|\text{Var}_\rho\right|(A) = 0$ only under rare circumstances.

Lemma 2.1. (a) $\left|\text{Var}_\rho\right|(A) = 0$ *for every* $\rho \in \mathcal{D}(H)$ *if and only if* $A = \lambda I$ *for some* $\lambda \in \mathbb{C}$. (b) *Suppose* ρ *has the form (2.2). Then* $\left|\text{Var}_\rho\right|(A) = 0$ *if and only if* $Ax_i = E_\rho(A)x_i$, $i = 1, 2, \ldots$. (c) *Suppose* $\rho = P_x$ *is a pure state. Then* $\left|\text{Var}_\rho\right|(A) = 0$ *if and only if* $Ax = E_\rho(A)x$.

Proof. (a) If $A = \lambda I$ then $\lambda = E_\rho(A)$ so $\left|\text{Var}_\rho\right|(A) = 0$. Conversely, if $\left|\text{Var}_\rho\right|(A) = 0$ for all $\rho \in \mathcal{D}(H)$ then $|A - E_\rho(A)I| = 0$. It follows that $A = E_\rho(A)I$. (b) If $Ax_i = E_\rho(A)x_i$, $i = 1, 2, \ldots$, then

$$\left\langle |A - E_\rho(A)I|^2 x_i, x_i\right\rangle = \|(A - E_\rho(A)I)\, x_i\|^2 = 0$$

Hence,

$$|\mathrm{Var}_\rho|(A) = \mathrm{tr}\left[\rho\,|A - E_\rho(A)I|^2\right] = \sum \lambda_i \left\langle |A - E_\rho(A)I|^2 x_i, x_i \right\rangle = 0 \quad (2.3)$$

Conversely, if $|\mathrm{Var}_\rho|(A) = 0$, then as in (2.3) we have

$$\sum \lambda_i \left\langle |A - E_\rho(A)I|^2 x_i, x_i \right\rangle = 0$$

It follows that

$$\|(A - E_\rho(A)I)\, x_i\|^2 = \left\langle |A - E_\rho(A)|^2 x_i x_i \right\rangle = 0$$

for $i = 1, 2, \ldots$. Hence, $Ax_i = E_\rho(A)x_i$, $i = 1, 2, \ldots$.
(c) This follows directly from (b). □

If we write ρ in the form (2.2), then $\{x_i\}$ forms an orthonormal set (not necessarily a basis) in H. The range of ρ is $\mathrm{Ran}(\rho) = \overline{\mathrm{span}}\,\{x_i\}$. We say that ρ is **faithful** if $\mathrm{tr}(\rho A) = 0$ for $A \geq 0$ implies $A = 0$.

Lemma 2.2. *The following statements are equivalent*
(a) *ρ is faithful.* (b) $\mathrm{Ran}(\rho) = H$. (c) *$\{x_i\}$ forms an orthonormal basis for H.* (d) *ρ is invertible.*

Proof. (a)⇒(b) If $\mathrm{Ran}(\rho) \neq H$, then there exists an $x \in H$ with $\|x\| = 1$ such that $x \perp x_i$, $i = 1, 2, \ldots$. But then $P_x \geq 0$ and $P_x \neq 0$, but $\mathrm{tr}(\rho P_x) = 0$. Hence, ρ is not faithful. (b)⇒(c) If $\mathrm{Ran}(\rho) = H$ then $\overline{\mathrm{span}}\,\{x_i\} = H$ so that $\{x_i\}$ is an orthonormal basis. (c)⇒(d) If $\{x_i\}$ forms an orthonormal basis, then all the eigenvalues of ρ are positive. It follows that ρ is invertible. (d)⇒(a) Suppose ρ is invertible. If $A \geq 0$ and $\mathrm{tr}(\rho A) = 0$, then $\mathrm{tr}(\rho^{1/2} A \rho^{1/2}) = \mathrm{tr}(\rho A) = 0$. Since $\rho^{1/2} A \rho^{1/2} \geq 0$, we conclude that $\rho^{1/2} A \rho^{1/2} = 0$. Since $\rho^{1/2}$ is invertible, it follows that $A = 0$. Hence, ρ is faithful. □

It follows from Lemmas 2.1 and 2.2 that if ρ is faithful then $|\mathrm{Var}_\rho|(A) = 0$ if and only if $A = E_\rho(A)I$.

3 Types of Convergence

Various types of convergence have been discussed in the literature [2, 4, 8, 10] and we shall now consider four of them. In the following definitions $A_n \in$

$\mathcal{B}(H)$ will be a sequence of operators and $\rho \in \mathcal{D}(H)$ will be a fixed state. A sequence A_n converges to A **almost uniformly** $[\rho]$ if for any $\varepsilon > 0$ there exists a projection P such that $\mathrm{tr}\,[\rho(I-P)] < \varepsilon$ and $\lim\limits_{n\to\infty} \|(A_n - A)P\| = 0$. Projections are thought of as (noncommutative) events and the trace condition is equivalent to $\mathrm{tr}(\rho P) > 1-\varepsilon$. The definition says that for small ε the event P is likely to occur and A_n converges to A uniformly on the range of P.

A sequence A_n converges to A **almost surely** $[\rho]$ if

$$\lim_{n\to\infty} E_\rho\left(|A_n - A|^2\right) = \lim_{n\to\infty} \mathrm{tr}\left(\rho\,|A_n - A|^2\right) = 0$$

We frequently write $A_n \to A$ a.s. $[\rho]$ for almost sure convergence. We say that A_n **converges in ρ-mean** to A if

$$\lim_{n\to\infty} E_\rho(A_n - A) = \lim_{n\to\infty} \mathrm{tr}\,[\rho(A_n - A)] = 0$$

and that A_n **converges in ρ-probability** to A if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P_\rho\left(|A_n - A| \geq \varepsilon\right) = 0$$

In this definition, $P_\rho\left(|A_n - A| \geq \varepsilon\right)$ is shorthand for $P_\rho\left(|A_n - A| \in [\varepsilon, \infty)\right)$. A fifth type of probabilistic convergence will be discussed in Section 5.

We say that a sequence A_n converges to A **strongly** on $K \subseteq H$ if $\lim\limits_{n\to\infty} \|(A_n - A)x\| = 0$ for all $x \in K$. If $K = H$ then we just say that A_n converges strongly to A. We now compare these various types of convergence.

Lemma 3.1. *If $A_n \to A$ a.s. $[\rho]$ then $A_n \to A$ in ρ-mean.*

Proof. Applying Equation (2.1) gives

$$|E_\rho(A_n - A)|^2 \leq E_\rho\left(|A_n - A|^2\right)$$

and the result follows. □

Theorem 3.2. (a) *$A_n \to A$ a.s. $[\rho]$ if and only if $A_n \to A$ strongly on $\mathrm{Ran}(\rho)$.* (b) *If ρ is faithful, then $A_n \to A$ a.s. $[\rho]$ if and only if $A_n \to A$ strongly.*

Proof. (a) Assuming ρ has the form (2.2) we have

$$\operatorname{tr}\left(\rho\,|A_n - A|^2\right) = \sum_j \lambda_j \big\langle |A_n - A|^2\, x_j, x_j \big\rangle = \sum_j \lambda_j \left\|(A_n - A)x_j\right\|^2$$

Hence, if $A_n \to A$ a.s. $[\rho]$ then $A_n x_j \to A x_j$ for every j so $A_n x \to Ax$ for all x in a dense subset of $\operatorname{Ran}(\rho)$. It follows that $A_n \to A$ strongly on $\operatorname{Ran}(\rho)$. Conversely, suppose that $A_n \to A$ strongly on $\operatorname{Ran}(\rho)$. Let $A'_n, A'\colon \operatorname{Ran}(\rho) \to H$ denote the restrictions of A_n and A to $\operatorname{Ran}(\rho)$. By the uniform boundedness theorem there is an $M > 0$ such that $\left\|A'_n - A'\right\|^2 \le M$ for all n. Let $\varepsilon > 0$ be given. Then there exists an integer N such that

$$\sum_{i=N+1}^{\infty} \lambda_i < \frac{\varepsilon}{2M}$$

Moreover, there is an integer K such that $n \ge K$ implies

$$\sum_{i=1}^{N} \left\|(A'_n - A')x_i\right\|^2 < \frac{\varepsilon}{2}$$

Hence, $n \ge K$ implies

$$\begin{aligned} \operatorname{tr}\left(\rho\,|A_n - A|^2\right) &= \sum_{i=1}^{N} \lambda_i \left\|(A'_n - A')x_i\right\|^2 + \sum_{i=N+1}^{\infty} \lambda_i \left\|(A'_n - A')x_i\right\|^2 \\ &< \frac{\varepsilon}{2} + M \sum_{i=N+1}^{\infty} \lambda_i < \varepsilon \end{aligned}$$

Hence, $A_n \to A$ a.s. $[\rho]$. Part (b) follows from Lemma 2.2. □

It follows from Theorem 3.2 that if $A_n \to A$ and $B_n \to B$ a.s. $[\rho]$ and $\alpha, \beta \in \mathbb{C}$, then $\alpha C A_n + \beta D B_n \to \alpha C A + \beta D B$ a.s. $[\rho]$. The next result is a version of Egoroff's theorem.

Theorem 3.3. *If $A_n \to A$ a.s. $[\rho]$, then $A_n \to A$ almost uniformly $[\rho]$.*

Proof. Since $A_n \to A$ a.s. $[\rho]$, $A_n x \to Ax$ for every $x \in \operatorname{Ran}(\rho)$. Let ρ have the form (2.2) and let Q_n be the projection onto $\overline{\operatorname{span}}\{x_1, \dots, x_n\}$. Then

$Q_n = \sum_{i=1}^{n} P_{x_i}$ and $Q = \sum P_{x_i}$ is the projection onto $\mathrm{Ran}(\rho)$. Then letting $Q^\perp = I - Q$ we have

$$\begin{aligned}\mathrm{tr}\left[\rho(I-Q_n)\right] &= \mathrm{tr}\left[\rho(Q+Q^\perp-Q_n)\right] = \mathrm{tr}\left[(Q-Q_n)\right]\\ &= \mathrm{tr}\left(\rho\sum_{i=n+1}^{\infty}P_{x_i}\right) = \mathrm{tr}\left(\sum_{i=n+1}^{\infty}\lambda_i P_{x_i}\right) = \sum_{i=n+1}^{\infty}\lambda_i\end{aligned}$$

Given $\varepsilon > 0$ there exists an n such that $\mathrm{tr}\left[\rho(I-Q_n)\right] < \varepsilon$. Letting $P = Q_n$ we have that $\mathrm{tr}\left[\rho(I-P)\right] < \varepsilon$. Let $\varepsilon' > 0$. Now there exists an integer N such that $m \geq N$ implies $\|(A_m - A)x_i\| < \frac{\varepsilon'}{n}$, $i = 1, \ldots, n$. Then for any $x \in H$ with $\|x\| \leq 1$ we have $\|Px\| \leq 1$ and $Px = \sum_{i=1}^{n} c_i x_i$ where $\sum |c_i|^2 = \|Px\|^2 \leq 1$. Hence, $|c_i| \leq 1$, $i = 1, \ldots, n$. Thus, for $m \geq N$ we have

$$\begin{aligned}\|(A_m - A)Px\| &= \left\|(A_m - A)\sum c_i x_i\right\| \leq \sum |c_i|\;\|(A_m - A)x_i\|\\ &< \frac{\varepsilon'}{n}\sum |c_i| \leq \varepsilon'\end{aligned}$$

Therefore, $m \geq N$ implies that $\|(A_m - A)P\| \leq \varepsilon'$ so $A_n \to A$ almost uniformly $[\rho]$. □

The converse of Theorem 3.3 does not hold. However, it is shown in [5] that if A_n are uniformly bounded then $A_n \to A$ uniformly $[\rho]$ implies that $A_n \to A$ a.s. $[\rho]$.

4 Law of Large Numbers

We first present some standard probabilistic results.

Lemma 4.1. (Markov) *If $A \geq 0$ and $a > 0$, then*

$$P_\rho(A \geq a) \leq \frac{E_\rho(A)}{a}$$

Proof. By the spectral theorem we have

$$A = \int_0^\infty \lambda P^A(d\lambda) \geq \int_a^\infty \lambda P^A(d\lambda) \geq a\int_a^\infty P^A(d\lambda) = aP^A\left([a,\infty)\right)$$

Taking expectations gives

$$E_\rho(A) \geq aE_\rho\left[P^A\left([a,\infty)\right)\right] = aP_\rho(A \geq a) \qquad \square$$

Corollary 4.2. *If $A_n \to A$ a.s. $[\rho]$, then $A_n \to A$ is ρ-probability.*

Proof. By Lemma 4.1 and Equation (2.1) we have

$$P_\rho\left(|A_n - A| \geq \varepsilon\right) \leq \frac{E_\rho\left(|A_n - A|\right)}{\varepsilon} \leq \frac{1}{\varepsilon}\left[E_\rho\left(|A_n - A|^2\right)\right]^{1/2}$$

The result now follows $\square$

Corollary 4.3. (Chebyshev) *If $A \in \mathcal{B}(H)$ with $\mu = E_\rho(A)$, $\sigma = \left|\mathrm{Var}_\rho\right|(A)$, then for any $k > 0$ we have*

$$P_\rho\left(|A - \mu I| \geq k\right) \leq \frac{\sigma^2}{k^2}$$

Proof. By Lemma 4.1 we have

$$P_\rho\left(|A - \mu I|^2 \geq k^2\right) \leq \frac{E_\rho\left(|A - \mu I|^2\right)}{k^2} = \frac{\sigma^2}{k^2}$$

But $P_\rho\left(|A - \mu I| \geq k\right) = P_\rho\left(|A - \mu I|^2 \geq k^2\right)$. $\square$

The following results are called one-sided Chebyshev inequalities.

Corollary 4.4. *If $A \in \mathcal{B}(H)$ is self-adjoint, $\mu = E_\rho(A)$, $\sigma^2 = \mathrm{Var}_\rho(A)$ and $a > 0$, then*

$$P_\rho\left(A \geq \mu + a\right) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \quad P_\rho\left(A \leq \mu - a\right) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Proof. First assume that $\mu = 0$. For $b > 0$ we have

$$P_\rho(A \geq a) = P_\rho(A + bI \geq a + b) = P_\rho\left[(A + bI)^2 \geq (a + b)^2\right]$$

Applying Lemma 4.1 gives

$$P_\rho(A \geq a) \leq \frac{E_\rho\left[(A + bI)^2\right]}{a^2 + b^2} = \frac{\sigma^2 + b^2}{(a + b)^2}$$

Letting $b = \sigma^2/a$ we obtain

$$P_\rho(A \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \tag{4.1}$$

Now suppose that $\mu \neq 0$. Since $A - \mu I$ and $\mu I - A$ have mean 0 and variance σ^2, by (4.1) we have

$$P_\rho(A - \mu I \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \quad P_\rho(\mu I - A \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

The result now follows □

For $A, B \in \mathcal{B}(H)$ we define the ρ-**correlation coefficient** by

$$\begin{aligned} \mathrm{Cor}_\rho(A, B) &= E_\rho\left[(A - E_\rho(A)I)^* (B - E_\rho(B)I)\right] \\ &= E_\rho(A^* B) - E_\rho(A^*) E_\rho(B) \end{aligned}$$

Notice that $\mathrm{Cor}_\rho(A, A) = |\mathrm{Var}_\rho|(A)$. We say that A and B are **uncorrelated** if $\mathrm{Cor}_\rho(A, B) = 0$. Of course, A and B are uncorrelated if and only if $E_\rho(A^* B) = E_\rho(A^*) E_\rho(B)$. Also, A and B are uncorrelated if and only if B and A are uncorrelated. It is easy to check that

$$\mathrm{Cor}_\rho(A + aI, B + bI) = \mathrm{Cor}_\rho(A, B) \tag{4.2}$$

for all $a, b \in \mathbb{C}$. It follows from (4.2) that if A and B are ρ-uncorrelated then $A + aI$ and $B + bI$ are ρ-uncorrelated.

It is clear that $\mathrm{Var}_\rho(\lambda A) = \lambda^2 \mathrm{Var}_\rho(A)$ and that

$$|\mathrm{Var}_\rho|(\lambda A) = |\lambda|^2 |\mathrm{Var}_\rho|(A)$$

In general

$$|\mathrm{Var}_\rho|(A + B) \neq |\mathrm{Var}_\rho|(A) + |\mathrm{Var}_\rho|(B)$$

but additivity does hold in the uncorrelated case.

Lemma 4.5. (a) *For $A, B \in \mathcal{B}(H)$ we have*

$$|\mathrm{Var}_\rho|(A + B) = |\mathrm{Var}_\rho|(A) + |\mathrm{Var}_\rho|(B) + 2\mathrm{Re}\,\mathrm{Cor}_\rho(A, B)$$

(b) *If A and B are uncorrelated, then* $|\mathrm{Var}_\rho|(A+B) = |\mathrm{Var}_\rho|(A) + |\mathrm{Var}_\rho|(B)$.

Proof. (a) The result follows from

$$\begin{aligned}
|\mathrm{Var}_\rho|(A+B) &= \mathrm{tr}\left[\rho(A+B)^*(A+B)\right] - \left|\mathrm{tr}\left[\rho(A+B)\right]\right|^2\\
&= \mathrm{tr}(\rho A^*A) + \mathrm{tr}(\rho B^*B) + \mathrm{tr}(\rho A^*B) + \mathrm{tr}(\rho B^*A)\\
&\quad - \left[\mathrm{tr}(\rho A) + \mathrm{tr}(\rho B)\right]\left[\mathrm{tr}(\rho A^*) + \mathrm{tr}(\rho B^*)\right]\\
&= |\mathrm{Var}_\rho|(A) + |\mathrm{Var}_\rho|(B) + \mathrm{Cor}_\rho(A,B) + \overline{\mathrm{Cor}_\rho(A,B)}\\
&= |\mathrm{Var}_\rho|(A) + |\mathrm{Var}_\rho|(B) + 2\mathrm{Re}\,\mathrm{Cor}_\rho(A,B)
\end{aligned}$$

(b) this follows directly from (a). □

The next result gives the core of the proof of the law of large numbers.

Theorem 4.6. *Let $A_i \in \mathcal{B}(H)$ be mutually ρ-uncorrelated with a common mean $\mu = E_\rho(A_i)$, $i = 1, 2, \ldots$. If $S_n = \frac{1}{n}\sum_{i=1}^{n} A_i$, then for any $\lambda \in \mathbb{C}$ we have*

$$\mathrm{tr}\left[\rho\,|S_n - \lambda I|^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} |\mathrm{Var}_\rho|(A_i) + |\mu - \lambda|^2$$

Proof. First assume that $\mu = 0$. Then

$$\begin{aligned}
\mathrm{tr}\left[\rho\,|S_n - \lambda I|^2\right] &= \mathrm{tr}\left[\rho\left(S_n^*S_n + |\lambda|^2 I - \overline{\lambda}S_n - \lambda S_n^*\right)\right]\\
&= \mathrm{tr}(\rho S_n^*S_n) + |\lambda|^2 - \overline{\lambda}\mathrm{tr}(\rho S_n) - \lambda\,\overline{\mathrm{tr}(\rho S_n)}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{tr}(\rho A_i^*A_i) + \frac{1}{n^2}\sum_{i\neq j=1}^{n}\mathrm{tr}(\rho A_i^*A_j) + |\lambda|^2\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{tr}(\rho A_i^*A_i) + \frac{1}{n^2}\sum_{i\neq j=1}^{n}\mathrm{tr}(\rho A_i^*)\mathrm{tr}(\rho A_j) + |\lambda|^2\\
&= \frac{1}{n^2}\sum_{i=1}^{n}|\mathrm{Var}_\rho|(A_i) + |\lambda|^2
\end{aligned}$$

If $\mu \neq 0$, then by Equation (4.2), $A_i - \mu I$ are mutually uncorrelated with mean 0 and

$$S_n' = \frac{1}{n}\sum_{i=1}^{n}(A_i - \mu I) = S_n - \mu I$$

Hence, by our preceding work

$$\begin{aligned}\operatorname{tr}\left[\rho\,|S_n-\lambda I|^2\right] &= \operatorname{tr}\left[\rho\,|S_n'+(\mu-\lambda)I|^2\right]\\ &= \frac{1}{n^2}\sum_{i=1}^{n}|\mathrm{Var}_\rho|\,(A_i-\mu I)+|\mu-\lambda|^2\\ &= \frac{1}{n^2}\sum_{i=1}^{n}|\mathrm{Var}_\rho|\,(A_i)+|\mu-\lambda|^2\end{aligned}$$ □

The law of large numbers says that under certain conditions, if A_i are mutually ρ-uncorrelated with common mean μ, then their average $S_n = \frac{1}{n}\sum_{i=1}^{n} A_i$ converges to the mean operator μI. We now present several versions of this law.

Theorem 4.7. (Strong law of large numbers)
Let A_i be mutually ρ-uncorrelated with common mean $\mu = E_\rho(A_i)$, $i = 1, 2, \dots,$ and suppose there exist real numbers $M > 0$, $0 < r < 2$ such that $\sum_{i=1}^{n}|\mathrm{Var}_\rho|\,(A_i) \le Mn^r$ for all n. If $S_n = \frac{1}{n}\sum_{i=1}^{n} A_i$, then $S_n \to \lambda I$ a.s. $[\rho]$ if and only if $\lambda = \mu$.

Proof. Since

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}|\mathrm{Var}_\rho|\,(A_i) \le \lim_{n\to\infty} Mn^{r-2} = 0$$

applying Theorem 4.6 gives

$$\lim_{n\to\infty}\operatorname{tr}\left[\rho\,|S_n-\lambda I|^2\right] = |\mu-\lambda|^2$$

Hence, $\lim_{n\to\infty} E_\rho\left(|S_n-\lambda I|^2\right) = 0$ if and only if $\lambda = \mu$. □

A simple condition that implies the variance property of Theorem 4.7 is the uniform boundedness condition $|\mathrm{Var}_\rho|\,(A_i) \le M$ for all i. Applying Corollary 4.2 and Theorem 4.7 we obtain the following result.

Corollary 4.8. (Weak Law of large numbers) *Under the same assumptions as in Theorem 4.7 we have that $S_n \to \mu I$ in ρ-probability.*

We can even relax the common mean condition in Theorem 4.7 to obtain the following stronger version.

Theorem 4.9. *Let A_i be mutually ρ-uncorrelated, let $\mu_i = E_\rho(A_i)$, $i = 1, 2, \ldots,$ and suppose there exist real numbers $M > 0$, $0 < r < 2$ such that $\sum_{i=1}^{n} \left|\mathrm{Var}_\rho\right|(A_i) \le Mn^r$. If $S_n = \frac{1}{n}\sum_{i=1}^{n} A_i$ then*

$$S_n - \left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right) I \to 0 \text{ a.s. } [\rho]$$

Proof. Let $A_i' = A_i - \mu_i I$ and $S_n' = \frac{1}{n}\sum_{i=1}^{n} A_i'$. Then A_i' are mutually ρ-uncorrelated and $E_\rho(A_i') = 0$, $i = 1, 2, \ldots$. As in the proof of Theorem 4.6 we have

$$\mathrm{tr}\left(\rho\left|S_n'\right|^2\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left|\mathrm{Var}_\rho\right|(A_i') \le Mn^{r-2}$$

Hence, $\lim_{n\to\infty} \mathrm{tr}\left(\rho\left|S_n'\right|^2\right) = 0$. But

$$S_n' = \frac{1}{n}\sum_{i=1}^{n}(A_i - \mu_i I) = \frac{1}{n}\sum_{i=1}^{n} A_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i I = S_n - \left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right) I$$

Since $S_n' \to 0$ a.s. $[\rho]$, the result follows □

5 Central Limit Theorem

We say that $A, B \in \mathcal{B}(H)$ are **independent** in the state ρ if

$$\mathrm{tr}(\rho A^{n_1} B^{m_1} \cdots A^{n_r} B^{m_r}) = \mathrm{tr}(\rho A^{n_1 + \cdots n_r})\mathrm{tr}(\rho B^{m_1 + \cdots + m_r})$$

for all $n_1, \ldots, n_r, m_1, \ldots, m_r \in \mathbb{N}$. For example, if $A \in \mathcal{B}(H_1)$, $\rho_1 \in \mathcal{D}(H_1)$, $B \in \mathcal{B}(H_2)$, $\rho_2 \in \mathcal{D}(H_2)$ then $A \otimes I \in \mathcal{B}(H_1 \otimes H_2)$ and $I \otimes B \in \mathcal{B}(H_1 \otimes H_2)$ are independent in the state $\rho_1 \otimes \rho_2$. In this case the operators commute and it is not surprising that they are independent. However, there are simple examples of noncommuting independent operators. For example, suppose

$Ax = \alpha x$, $Bx = \beta x$ $\|x\| = 1$. Then A and B are independent in the pure state P_x. Indeed,

$$\begin{aligned}\langle A^{n_1} B^{m_1} \cdots A^{n_r} B^{n_r} x, x\rangle &= \alpha^{n_1+\cdots+n_r} \beta^{m_1+\cdots+m_r} \\ &= \left\langle A^{n_1+\cdots+n_r} x, x\right\rangle \left\langle B^{m_1+\cdots+m_r} x, x\right\rangle\end{aligned}$$

The **moment generating function** of $A \in \mathcal{B}(H)$ relative to $\rho \in \mathcal{D}(H)$ is the function $M_{\rho,A} \colon \mathbb{R} \to \mathbb{R}$ given by $M_{\rho,A}(t) = E_\rho(e^{tA})$ the terminology comes from the fact that

$$\frac{d^n}{dt^n} M_{\rho,A}(t) \mid_{t=0} = E_\rho(A^n)$$

which is the n-the moment.

Lemma 5.1. *If A and B are independent in the state ρ, then $M_{\rho,A+B} = M_{\rho,A} M_{\rho,B}$.*

Proof. Since A and B are independent we have

$$E_\rho\left[(A+B)^n\right] = \sum_{k=0}^{n} \binom{n}{k} E_\rho(A^{n-k}) E_\rho(B^k)$$

Hence,

$$\begin{aligned}&E_\rho\left[e^{t(A+B)}\right] \\ &\quad = E_\rho\left[I + t(A+B) + \frac{t_2}{2!}(A+B)^2 + \cdots\right] \\ &\quad = 1 + t\left[E_\rho(A) + E_\rho(B)\right] + \cdots + \frac{t^n}{n!} \sum_{k=0}^{n} \binom{n}{k} E_\rho(A^{n-k}) E_\rho(B^k) + \cdots \\ &\quad = \left[1 + tE_\rho(A) + \cdots + \frac{t^n}{n!} E_\rho(A^n) + \cdots\right] \\ &\qquad\quad \left[1 + tE_\rho(B) + \cdots + \frac{t^n}{n!} E_\rho(B^n) + \cdots\right] \\ &\quad = E_\rho(e^{tA}) E_\rho(e^{tB})\end{aligned}$$

We conclude that $M_{\rho,A+B}(t) = M_{\rho,A}(t) M_{\rho,B}(t)$. □

We say that a sequence $\{A_i\}$ is **independent** in the state ρ if A_{i+1} is independent of $A_1+\cdots+A_i$ for all $i\in\mathbb{N}$. Notice that if $\{A_i\}$ is independent, then by Lemma 5.1 we have

$$\begin{aligned} M_{\rho,A_1+\cdots+A_n}(t) &= E_\rho\left[e^{t(A_1+\cdots+A_1)}\right] \\ &= E_\rho\left[e^{t(A_1+\cdots+A_{n-1})}\right]E_\rho(e^{tA_n}) \\ &= \cdots = E_\rho(e^{tA_1})E_\rho(e^{tA_2})\cdots E_\rho(e^{tA_n}) \\ &= M_{\rho,A_1}(t)\cdots M_{\rho,A_n}(t) \end{aligned} \tag{5.1}$$

We say that A and B are **identically distributed** in the state ρ if $E_\rho(A^n)=E_\rho(B^n)$ for all $n\in\mathbb{N}$. A sequence $\{A_i\}$ **converges in distribution relative to** ρ if $\lim M_{\rho,A_n}(t)=M_\rho(t)$ for all $t\in\mathbb{R}$, where $M_\rho(t)$ is the moment generating function of a classical random variable. We do not require that $M_\rho(t)=M_{\rho,A}(t)$ for an $A\in\mathcal{B}(H)$. It is well known that the moment generating function for the classical normal distribution with zero mean and variance one is $M(t)=e^{t^2/2}$. The central limit theorem says that if $\{A_i\}$ is an independent, identically distributed sequence then suitably normalized averages of the $\{A_i\}$ converge in distribution to the normal distribution.

Theorem 5.2. *Let $\{A_i\}$ be independent, identically distributed in the state ρ with common mean $E_\rho(A_i)=\mu$ and common variance $\operatorname{Var}_\rho(A_i)=\sigma^2$. If*

$$T_n=\sum_{i=1}^n\left(\frac{A_i-\mu I}{\sigma\sqrt{n}}\right)$$

then $\lim M_{\rho,T_n}(t)=e^{t^2/2}$ for all $t\in\mathbb{R}$.

Proof. First assume that $\mu=0$ and $\sigma^2=1$ and let $T_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n A_i$. Then by Equation (5.1) we have

$$M_{\rho,T_n}(t)=\left[M_{\rho,A_1}(t/\sqrt{n}\,)\right]^n \tag{5.2}$$

Let $L(t)=\ln M_{\rho,A_1}(t)$. Note that $L(0)=0$ and

$$L'(0)=\frac{M'_{\rho,A_1}(0)}{M_{\rho,A_1}(0)}=\mu=0$$

$$L''(0)=\frac{M_{\rho,A_1}(0){M''}_{\rho,A_1}(0)-M'_{\rho,A_1}(0)^2}{M_{\rho,A_1}(0)}=E_\rho(A_1^2)=1 \tag{5.3}$$

We want to prove that

$$\lim_{n\to\infty} M_{\rho,T_n}(t) = e^{t^2/2} \tag{5.4}$$

Applying (5.2), Equation (5.4) is equivalent to

$$\lim_{n\to\infty} \left[M_{\rho,A_1}(t/\sqrt{n}\,)\right]^n = e^{t^2/2} \tag{5.5}$$

and Equation (5.5) is equivalent to

$$\lim_{n\to\infty} nL(t/\sqrt{n}\,) = \frac{t^2}{2} \tag{5.6}$$

To verify Equation (5.6) we apply (5.3) and L'Hospital's rule to obtain

$$\lim_{n\to\infty} \frac{L(t/\sqrt{n}\,)}{n^{-1}} = \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n}\,)t}{2n^{-1/2}}\right] = \lim_{n\to\infty} \left[L''(t/\sqrt{n}\,)\frac{t^2}{2}\right] = \frac{t^2}{2}$$

Now suppose that μ and σ^2 are arbitrary and let $A_i' = \frac{A_i - \mu I}{\sigma}$. Then $\{A_i\}$ is an independent, identically distributed sequence with $E_\rho(A_i') = 0$,$\mathrm{Var}_\rho(A_i') = 1$. Then letting $T_n' = \frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} A_i'$, by Equation (5.4) we have $\lim M_{\rho,T_n'}(t) = e^{t^2/2}$. However,

$$T_n = \sum_{i=1}^{n} \left(\frac{A_i - \mu I}{\sigma\sqrt{n}}\right) = \sum_{i=1}^{n} \frac{\sigma A_i' + \mu I - \mu I}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i' = T_n'$$

Hence, $\lim M_{\rho,T_n}(t) = e^{t^2/2}$. □

The next result gives a version of the usual central limit theorem.

Corollary 5.3. *If $\{A_i\}$ is an independent, identically distributed sequence in the state ρ and $A_i \in \mathcal{S}(H)$, $i = 1, 2, \ldots,$ then*

$$\lim_{n\to\infty} P_\rho \left(\frac{A_1 + \cdots + A_n - n\mu}{\sigma\sqrt{n}} \le a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

Proof. The distribution functions of T_n converge to the distribution function of the standard normal random variable Z because the moment generating functions M_{ρ,T_n} converge to $M_Z(t) = e^{t^2/2}$ [9]. □

References

[1] O. Bratteli and D. Robinson, *Operator Algebras and Quantum Statistical Mechanics*, Vols I, II, Springer, New York, 1979, 1981.

[2] I. Cuculescu and A. Oprea, *Noncommutative Probability*, Kluwer, Dordrecht, 1994.

[3] G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory*, John Wiley, New York, 1972.

[4] J. Hamhalter, *Quantum Measure Theory*, Kluwer, Dordrecht, 2003.

[5] R. Jajte, *Strong Limit Theorems in Noncommutative Probability*, Springer Verlag, New York, 1985.

[6] P. Meyer, *Quantum Probability for Probabilists*, Springer Verlag, New York, 1993.

[7] K. Parthasarathy, *An Introduction to Quantum Stochastic Calculus* Birkhäuser, Basel 1992.

[8] M. Rédei and S. Summers, Quantum probability theory, Arxiv: quant-ph/0601158, 2006.

[9] S. Ross, *A First Course in Probability*, Pearson Prentice Hall, Upper Saddle River, NJ, 2006.

[10] R. Streater, Classical and quantum probability, *J. Math. Phys.* **41** (2000), 3556–3603.

[11] D. Voiculescu, K. Dykema and A. Nica, *Free Random Variables* Amer. Math. Society, Providence, 1992.