8.372 Quantum Information Science III	**Fall 2024**
Lecture 12: October 17, 2024	
Scribe: Yeongwoo Hwang	*Random states and entanglement*

12.1 Introduction

In this lecture, we will discuss random variables, states and unitaries. First, why do we care about this topic? Well, they are critical for many applications, such as benchmarking of quantum devices and quantum state learning (e.g. state tomography). We'll see two themes in this lecture,

- Concentration of measure
- Probability distributions on larger spaces, such as the complex unit sphere.

12.2 Scalar random variables

A central tool is *Markov's Inequality*, which states that for a random variable $\boldsymbol{X}$, with $\boldsymbol{X} \geq 0$ always, then

$$\Pr_{\boldsymbol{X}}[\boldsymbol{X} \geq c \cdot \mathbb{E}[\boldsymbol{X}]] \leq \frac{1}{c} \tag{12.1}$$

where we use **bold face** to indicate random variables. By simple transformations on the random variable on which we apply Markov's inequality, we can achieve tighter bounds than the above.

Chebyshev's Inequality For $\boldsymbol{X}$ as above, we have

$$\Pr\Big[|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]| \geq c\sqrt{\mathrm{Var}[\boldsymbol{X}]}\Big] \leq \frac{1}{c^2}$$

Proof. Let $\boldsymbol{Y} := (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2$. Then, we see that $\mathbb{E}[\boldsymbol{Y}] = \mathrm{Var}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{X}^2] - \mathbb{E}[\boldsymbol{X}]^2$. By substituting in $\boldsymbol{Y}$ for $\boldsymbol{X}$ into (12.1), we obtain the desired conclusion. □

Chernoff To get the equivalent of a Chernoff inequality, we apply (12.1) to the moment generating functions for $\boldsymbol{X}$, which is $\mathbb{E}[e^{\lambda \boldsymbol{X}}]$. Applying Markov's inequality to $e^{\lambda X}$ then yields

$$\Pr\Big[e^{\lambda \boldsymbol{X}} \geq c\, \mathbb{E}[e^{\lambda \boldsymbol{X}}]\Big] \leq \frac{1}{c} \tag{12.2}$$

A common case is when $\boldsymbol{X} \sim \mathrm{Uniform}[\{+1, -1\}]$. In this case, the RHS of (12.2) becomes

$$\mathbb{E}[e^{\lambda \boldsymbol{X}}] = \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \cosh(\lambda)$$

This becomes particularly useful when $\boldsymbol{X}$ is itself the sum of multiple random variables. Let's define $\boldsymbol{Y} = \sum_{i=1}^{n} \boldsymbol{X}_i$, where eahc $\boldsymbol{X}_i$ is again uniform over $\{+1, -1\}$. Then,

$$\mathbb{E}[e^{\lambda \boldsymbol{Y}}] = \prod_{i=1}^{m} \mathbb{E}[e^{\lambda \boldsymbol{X}_i}] = \prod_{i=1}^{m} \cosh(\lambda) = \cosh(\lambda)^m \leq e^{\lambda^2 \frac{m}{2}} \tag{12.3}$$

Where the final inequality uses that $\cosh(\lambda) \le e^{\frac{\lambda^2}{2}}$. We can use this to get concentration scaling exponentially in n. In particular, setting $\boldsymbol{X}$ to be $\boldsymbol{Y}$ and

$$c = \frac{e^{\delta^2 m}}{\mathbb{E}[e^{\lambda \boldsymbol{Y}}]} = \frac{e^{\delta^2 m}}{e^{\lambda^2 \frac{m}{2}}}$$

in (12.2), we obtain that

$$\Pr[\boldsymbol{Y} \ge \delta m] \le e^{m\lambda^2/2 - m\delta^2} \le e^{-m\delta^2/2}$$

for $\lambda := \delta$.

12.3 Random vectors

We now move on distributions over vectors. Our goal throughout this section will be to compute quantities like $\mathbb{E}_{\boldsymbol{\psi}\sim\text{unif}(S^{d-1})}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$, where S^{d-1} is d-dimensional unit sphere. First, let's consider when $|\boldsymbol{g}\rangle$ is drawn with each coordinate IID from the complex normal distribution $\mathcal{N}_{\mathbb{C}}(0, \frac{1}{d})$, i.e.

$$|\boldsymbol{g}\rangle = \begin{pmatrix} \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_d \end{pmatrix} \qquad \text{with } \boldsymbol{g}_i \sim \mathcal{N}_{\mathbb{C}}(0, \tfrac{1}{d})$$

As a remark, sampling $\boldsymbol{g}_i \sim \mathcal{N}_{\mathbb{C}}(0, \frac{1}{d})$ is equivalent to sampling $\boldsymbol{x}_i, \boldsymbol{y}_i \sim \mathcal{N}(0, \frac{1}{2d})$ and setting $g_i = \boldsymbol{x}_i + \mathrm{i}\boldsymbol{y}_i$. This distribution has several nice properties.

- $\mathbb{E}_{|\boldsymbol{g}\rangle} \langle \boldsymbol{g}|\boldsymbol{g}\rangle = 1$
- $|\boldsymbol{g}\rangle$ has density $\mu(|\boldsymbol{g}\rangle) = \left(\frac{d}{\pi}\right)^d e^{-d\langle \boldsymbol{g}|\boldsymbol{g}\rangle}$
- From the above property, we see that the PDF does not depend on the *direction* of $|\boldsymbol{g}\rangle$, only its magnitude. This implies that the distribution is rotationally invariant, and in particular, it is *unitarily invariant.*

Of course, this distribution has the downside that the vectors are normalized only in expectation. Thus, we define a related distribution where we generate $|\boldsymbol{u}\rangle$ as,

$$|\boldsymbol{u}\rangle = \frac{|\boldsymbol{g}\rangle}{\sqrt{\langle \boldsymbol{g}|\boldsymbol{g}\rangle}}$$

with $|\boldsymbol{g}\rangle$ drawn as before. This distribution is denoted as $\text{Haar}(S^{d-1})$. We claim that this distribution is uniform over the surface of the complex sphere, i.e. unitarily invariant.

Proof. Let U be any unitary. Then by the third property above, we see that $U|\boldsymbol{g}\rangle$ is distributed the same as $|\boldsymbol{g}\rangle$. Thus,

$$U|\boldsymbol{u}\rangle = \frac{U|\boldsymbol{g}\rangle}{\|U|\boldsymbol{g}\rangle\|_2} = \frac{|\boldsymbol{g}\rangle}{\|U|\boldsymbol{g}\rangle\|_2} = \frac{|\boldsymbol{g}\rangle}{\||\boldsymbol{g}\rangle\|_2}$$

where the equalities are equalities as distributions. Since the quantity on the RHS is the definition of $|\boldsymbol{u}\rangle$, this shows that $|\boldsymbol{u}\rangle$ is rotationally invariant. □

12.3.1 Computation of moments

Given this, how do we compute $\mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$? We will,

1. Compute $\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|^{\otimes m}]$
2. Relate to $\mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$

Step 1 Let's start with $\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|]$. It suffices to compute the expectations of each entry, e.g. $\mathbb{E}[\boldsymbol{g}_i\boldsymbol{g}_j^* \, |i\rangle\langle j|]$. Since each index of $|\boldsymbol{g}\rangle$ is sampled IID, this is equal to $\mathbb{E}[\boldsymbol{g}_i\boldsymbol{g}_j^*] \, |i\rangle\langle j| = \frac{\delta_{i,j}}{d} \, |i\rangle\langle j|$. This implies that,

$$\mathbb{E}_{|\boldsymbol{g}\rangle}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|] = \frac{1}{d}\mathbb{I}$$

For $\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|^{\otimes 2}]$, we encounter slightly more complicated expressions. In particular,

$$\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|^{\otimes 2}] = \sum_{\substack{i_1,i_2,\\ j_1,j_2}} \mathbb{E}[\boldsymbol{g}_{i_1}\boldsymbol{g}_{i_2}\boldsymbol{g}_{j_1}^*\boldsymbol{g}_{j_2}^*] \, |i_1 i_2\rangle\langle j_1 j_2|$$

To evaluate each of these expectations, we use *Wick's Theorem*. In the $m = 2$ case, this theorem tells us that

$$\mathbb{E}[\boldsymbol{g}_{i_1}\boldsymbol{g}_{i_2}\boldsymbol{g}_{j_1}^*\boldsymbol{g}_{j_2}^*] = \mathbb{E}[\boldsymbol{g}_{i_1}\boldsymbol{g}_{j_1}*] \, \mathbb{E}[\boldsymbol{g}_{i_2}\boldsymbol{g}_{j_2}^*] + \mathbb{E}[\boldsymbol{g}_{i_1}\boldsymbol{g}_{j_2}^*] \, \mathbb{E}[\boldsymbol{g}_{i_2}\boldsymbol{g}_{j_1}^*] = \frac{\delta_{i_1,j_1}\delta_{i_2,j_2} + \delta_{i_1,j_2}\delta_{i_2,j_1}}{d^2}$$

where the middle sum is over all possible pairs of (i_1, i_2) with (j_1, j_2). For general m, we obtain expectations like

$$\mathbb{E}[\boldsymbol{g}_{i_1}, \ldots \boldsymbol{g}_{i_m}\boldsymbol{g}_{j_1}^* \ldots \boldsymbol{g}_{j_m}^*] = \frac{1}{d^n} \prod_{\pi \in S_n} \prod_{\ell=1}^{m} \delta_{i_\ell, j_{\pi(\ell)}}$$

After some simplification, which was omitted in the lecture, we find that,

$$\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|^{\otimes m}] = \frac{1}{d^n} \sum_{\pi \in S_n} P_\pi \qquad \text{with } P_\pi = \sum_{i_1,\ldots,i_m} |i_1, \ldots, i_m\rangle\langle i_{\pi(1)}, \ldots, i_{\pi(m)}| \tag{12.4}$$

Step 2 Finally, we want to relate (12.4) to $\mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$. Recall that $|\boldsymbol{g}\rangle = \boldsymbol{r}\,|\boldsymbol{u}\rangle$ and, moreover, $\boldsymbol{r}$ and $|\boldsymbol{u}\rangle$ are *uncorrelated* due to the rotational invariance of $|\boldsymbol{g}\rangle$. This means that,

$$\mathbb{E}[|\boldsymbol{g}\rangle\langle\boldsymbol{g}|^{\otimes m}] = \mathbb{E}[\boldsymbol{r}^{2n}] \, \mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$$

We saw that the LHS is $\Pi_{\text{sym}} = \frac{1}{d^n}\sum_{\pi \in S_n} P_\pi$. Since we know $\mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}]$ is a *normalized* quantum state, this implies that the scaling factor is just $\text{tr}[\Pi_{\text{sym}}]$. Thus, we conclude that,

$$\mathbb{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|^{\otimes m}] = \frac{\Pi_{\text{sym}}}{\text{tr}[\Pi_{\text{sym}}]}$$

To make this more concrete, when $m = 2$, it turns out that Π_{sym} (which is the projector onto the *symmetric subspace*) is spanned by $|00\rangle, |11\rangle$, and $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$. These are all states invariant under the SWAP operation. What is $\text{tr}[\Pi_{\text{sym}}]$, i.e. the dimension of the symmetric subspace? We'll see in the next lecture this is $\binom{d+m-1}{m}$.