CS286.2 Lecture 12: Classical and Quantum de Finetti theorems

Scribe: Jenish C. Mehta

Classical and Quantum de Finetti Theorems

Over the next two weeks, we will look at de Finetti Theorems, both in the classical and quantum setting. Our goal is to prove the following result, which is one of the strongest "no-go" results on the quantum PCP conjecture (Local Hamiltonian version) known:

Theorem 1. ***(Brandao-Harrow '13)****: Let H be a 2-local Hamiltonian acting on n qubits such that the interaction graph (which has an edge between the qubits q_1 and q_2 iff H has a local term H_i acting on both of them) is regular with degree D. Then there exists a product state $|\psi\rangle = |\psi_1\rangle \otimes ... \otimes |\psi_n\rangle$, such that*

$$\langle\psi|H|\psi\rangle \leq \lambda_0(H) + O\left(\frac{1}{D^{1/3}}\right)\frac{nD}{2}$$

Corollary 2. *Assume* NP $\neq$ QMA*, and if $b - a = \gamma m = \gamma\frac{nD}{2}$, then* QPCP *can only be true only if $D < \frac{1}{\gamma^3}$ (where the interaction graph is regular of degree D).*

Proof. If $D > \frac{1}{\gamma^3}$, then $\langle\psi|H|\psi\rangle < b$, and since $|\psi\rangle$ is a product state, it has a succinct classical description from which its energy can be checked in classical polynomial-time, contradicting the assumption. □

Recall that, as we saw (as an exercise) earlier, one can also show that assuming NP $\neq$ QMA the QPCP is false for Hamiltonians whose interaction graph is a lattice of dimension up to $O(\log n)$. Together with the above, this shows that if QPCP is true the interaction graph must have high "dimension" (i.e. it has very good expansion properties, in the same sense that we defined expander graphs), but it must also have small constant degree. Such graphs exist, but they clearly restrict those sets of instances of LH that could potentially be "hard" instances for QPCP.

Review of Probability Theory

We write $[n]$ for the set $\{1, 2, ..., n\}$. P, Q, R shall denote distributions on finite sets. μ, ν shall denote continuous density functions, usually on $[0, 1]$. The statistical distance between distributions is defined as

$$\|P - Q\|_1 = \sum_i |P(i) - Q(i)| = 2 \sup_{B\subseteq[n]} |P(B) - Q(B)| \tag{1}$$

Multinomial Distribution: The multinomial distribution $M^d_{n,k}$ denotes the distribution obtained by k *independent* draws *with* replacement from an urn containing n elements chosen from a colour class $[d]$. Let n_i be the number of elements in the urn of colour i. Clearly, $n = \sum_{i=1}^{d} n_i$. Let v_i be the number of elements

of color i that were drawn, where $\sum_{i=1}^{d} v_i = k$. If the elements that are drawn are given by $(s_1, s_2, ..., s_k)$ where each $s_i \in [d]$, then the probability mass function is given by

$$M_{n,k}^d(s_1, s_2, ..., s_k) = \prod_{i=1}^{d} \left(\frac{n_i}{n}\right)^{v_i}.$$

Hypergeometric Distribution: The hypergeometric distribution $H_{n,k}^d$ is same as the $M_{n,k}^d$, except that the independent draws are ***without*** replacement. All other notation remaining the same, the probability mass function is given by

$$H_{n,k}^d(s_1, s_2, ..., s_k) = \frac{\prod_{i=1}^{d} \prod_{j=1}^{v_i} (n_i - j + 1)}{\prod_{j=1}^{k} (n - j + 1)}$$

Lemma 3. (i) $\|H_{n,k}^d - M_{n,k}^d\|_1 \leq \frac{2dk}{n}$

(ii) *Up to the constant* 2*, the bound in* (i) *is tight for* $d = 2$ *and any* $k \leq n$.

(iii) *If* $d \geq n$ *and* $\forall i, n_i \leq 1$*, then* $\|H_{n,k}^d - M_{n,k}^d\|_1 \leq \frac{k(k-1)}{n}$

Proof. The proofs of (i) and (ii) are by direct (but tedious, especially for (ii)) calculation and left as exercise. For (iii), note that $M_{n,k}^d$ is uniform over $[n]^k$ and $H_{n,k}^d$ is uniform over the set of all sequences with no repetition. Thus (iii) follows by observing that the set B maximizing the right-hand side in (1) can be taken as the set of all sequences without repetition, whose measure under $M_{n,k}^d$ is $\approx 1 - k^2/n$ by the birthday paradox. □

Definition 4. A distribution P on $[d]^n$ is said to be n-exchangeable (or permutation-invariant), if $\forall (s_1, ..., s_n)$ where $s_i \in [d]$ and for all permutations π on $[n]$, $P(s_1, s_2, ..., s_n) = P(s_{\pi(1)}, s_{\pi(2)}, ..., s_{\pi(n)})$

Some examples are:

- The uniform distribution
- The product distribution, where $P(s_1, ..., s_n) = \prod_{i=1}^{n} Q(s_i)$ (Notation used: $P = Q^{\otimes n}$)
- The uniform distribution on all sequences having exactly $n/2$ ones. Note that the latter is very far from product: the closest product distribution will be $Q^{\otimes n}$ where Q is uniform, but this assigns probability only about $1/\sqrt{2\pi n}$ to sequences with exactly $n/2$ ones.

Classical de Finetti Theorem

Theorem 5. ***(de Finetti '37, Diaconis-Freedman '80s):*** *Let P be an n-exchangeable distribution on $[d]^n$ and $Tr_{n-k}(P)$ be the marginal distribution of P on k elements. Then there exists a measure μ on the set of all distributions on $[d]$ (i.e. the d-simplex) such that $\forall k \leq n$,*

$$\|Tr_{n-k}(P) - \int Q^{\otimes k} d\mu(Q)\| \leq \min\left\{\frac{2kd}{n}, \frac{k(k-1)}{n}\right\}$$

Proof. Note that any n-exchangeable distribution P is a convex combination with weights w_i of distributions P_{U_i}, where P_{U_i} is the distribution on urn U_i defined as the uniform distribution over all sequences with $n_{i,j}$ copies of the element $j \in [d]$. Thus $Tr_{n-k}(P) = \sum_i w_i Tr_{n-k}(P_{U_i})$, with $Tr_{n-k}(P_{U_i}) = H_{n,k,U_i}^d$ (U_i is used to denote the urn). Now note that the multinomial distribution on a sample of size k is just the product of multinomial distributions on samples of size 1 — it is a product distribution! And we saw that M_{n,k,U_i}^d is a good approximation to H_{n,k,U_i}^d: $\|H_{n,k,U_i}^d - M_{n,k,U_i}^d\|_1 \leq \frac{2dk}{n}$, completing the theorem with μ the measure that assigns weight w_i to the distribution $Q_{U_i} = \mathrm{Tr}_{n-1}(P_{U_i})$ on $[d]$ associated with U_i. □

Density Matrices

Before we discuss the quantum de Finetti theorem, we will need a generalization of pure states, called density matrices. Density matrices are distributions over pure states. Consider the distribution over pure states $\{p_i, |\psi_i\rangle\}$, where p_i is the probability of state $|\psi_i\rangle$. The associated density matrix ρ on $\mathbb{C}^d \times \mathbb{C}^d$ is defined as follows:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

An important thing to note is that this representation does *not* keep all the information from the distribution, but only that information which can be observed by measurement! Consider the following example.

Example 6. For $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$, the density matrix is $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{I}_2/2$, and for $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$, the density matrix is $\frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \mathbb{I}_2/2$! Now consider a measurement $\{P, \mathbb{I} - P\}$. The probability of obtaining the outcome 'P' in either case is

$$\frac{1}{2}\|\sqrt{P}|0\rangle\|^2 + \frac{1}{2}\|\sqrt{P}|1\rangle\|^2 = \frac{1}{2}\langle 0|P|0\rangle + \frac{1}{2}\langle 1|P|1\rangle = \mathrm{Tr}\big(P(\mathbb{I}_2/2)\big) = \frac{1}{2}\|\sqrt{P}|+\rangle\|^2 + \frac{1}{2}\|\sqrt{P}|-\rangle\|^2$$

Thus, the density matrix exactly captures the information that can be observed!

More generally, for any measurement $\{X, \mathbb{I} - X\}$,

$$\begin{aligned} Pr(\text{outcome } X \text{ on distribution } \{p_i, |\psi_i\rangle\}) &= \sum_i p_i \|\sqrt{X}|\psi_i\rangle\|^2 \\ &= \sum_i p_i \langle\psi_i|X|\psi_i\rangle \\ &= \sum_i p_i \mathrm{Tr}(X|\psi_i\rangle\langle\psi_i|) \\ &= Tr(X\rho) \end{aligned}$$

The above formula gives the general measurement rule for density matrices.

Note that ρ as defined will always be a positive semidefinite matrix with $Tr(\rho) = 1$. Moreover, any ρ satisfying these two conditions can be expressed as a probability distribution using the spectral theorem. Thus, we can give an equivalent definition of a density matrix.

Definition 7. A density matrix is $\rho \in \mathbb{C}^d \times \mathbb{C}^d$ such that ρ is positive semidefinite and $\mathrm{Tr}(\rho) = 1$.

Now that we have a definition of density matrices, we want to extend the notion of statistical distance to distance between two matrices. The most natural generalization is that of trace distance, which corresponds to the maximum probability with which the two states can be distinguished.

Definition 8. (Trace Distance): The trace distance between two density matrices ρ and σ is defined as:

$$\|\rho - \sigma\|_{tr} = \max_{0 \le X \le \mathbb{I}} |\mathrm{Tr}(X\rho) - \mathrm{Tr}(X\sigma)| = \frac{1}{2}\|\rho - \sigma\|_1,$$

where for a matrix A, we denote $\|A\|_1 = \mathrm{Tr}(\sqrt{AA^\dagger})$ the sum of the singular values of A.