8.372 Quantum Information Science III **Fall 2024**

Lecture 6: September 24, 2024

Scribe: John Blue *Quantum Relative Entropy*

6.1 Chernoff-Stein Lemma

Let's start by proving the Chernoff-Stein Lemma from the last lecture. The setup: we have a string x^n, which was sampled either from p^n or q^n, and we want to know which. To do this, we will look at the likelihood ratio test (LRT). To perform this test, we first compute

$$W(x^n) = \log \frac{p^n(x^n)}{q^n(x^n)}. \tag{6.1}$$

Note that W is a random variable, and that

- $\mathbb{E}_{x^n \sim p^n}[W] = nD(p||q)$
- $\mathbb{E}_{x^n \sim q^n}[W] = -nD(q||p)$

Then, to make the decision, we define some value T such that if $W \geq T$, we guess p^n, and if $W < T$, we guess q^n. Let $A = \{x^n | W(x^n) \geq T\}$ be the "acceptance region".

We're interested in asymmetric hypothesis testing: we need $p^n(A) \geq 1 - \epsilon$ (i.e. the probability that we guess q when it was actually p should be less than ϵ), and then $q^n(A) \leq e^{-nR}$ for some R (the probability that we guess p when it was actually q should grow exponentially small with n).

To decide where to set T, observe that if you set the threshold above $nD(p||q)$, then (in the limit of large sample sizes), we will never guess p. On the other hand, if we set the threshold below $-nD(q||p)$, we will never guess q. This suggests we should set T somewhere inside this range. Since we want to minimize $q^n(A)$, we'll pick T to be closer to this upper bound: $T = n(D(p||q) - \delta)$.

We will first show that this T achieves the desired bound for $p^n(A)$. Consider that

$$p^n(A) = \Pr_{x^n \sim p^n}\left[\log \frac{p^n(x^n)}{q^n(x^n)} > nD(p||q)\right] \tag{6.2}$$

$$= \Pr_{x^n \sim p^n}\left[D(p||q) - \frac{1}{n}\sum_{i=1}^{n} W[x_i] < \delta\right] \tag{6.3}$$

Since $\mathbb{E}_{x \sim p}[W[x]] = D(p||q)$, and each of the x_i are independent and identically drawn from the source, by the law of large numbers, this quantity approaches 1 as n goes to infinity. Thus, for any ϵ and δ we can take n large enough that $p^n(A) \geq 1 - \epsilon$.

Now to show that $q^n(A)$ is small. If $x^n \in A$, then $q^n(x^n) \leq e^{-T} p^n(x^n)$. Then,

$$q^n(A) \leq e^{-T} p^n(A) \tag{6.4}$$

$$\leq e^{-T} \tag{6.5}$$

so $R = D(p||q) - \delta$ (for any $\delta > 0$).

6.1.1 Multiple hypothesis testing

We briefly mention another form of hypothesis testing: multiple hypothesis testing. Here, we have $Q \subseteq \Delta_d = \{\text{prob dists on } [d]\}$. Now, we want to distinguish between the two cases $x^n \sim p^n$ or q^n for some $q \in Q$. Intuitively, it makes sense that distinguishing p from Q should be at least as hard as distinguishing p from q^*, where q^* is the distribution in Q closest to p (see figure 6.1). It turns out that it is actually equally as hard - you can distinguish with the exponential rate

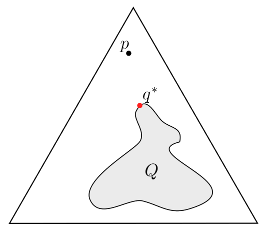

Figure 6.1: An example of multiple hypothesis testing. Given a sample x^n, we want to distinguish between two cases, $x^n \sim p^n$, or $x^n \sim q^n$ where $q \in Q$, a subset of the probability simplex. Here, q^* is the point in Q closest to p.

$R = \min_{q \in Q} D(p||q)$.

6.2 Quantum Relative Entropy and Quantum Chernoff-Stein

We will now turn to the quantum analogue of hypothesis testing. First, we define the quantum relative entropy.

Definition 1 (Quantum Relative Entropy)**.** *The quantum relative entropy, $D(\rho||\sigma)$, is defined as*

$$D(\rho||\sigma) = Tr[\rho(\log \rho - \log \sigma)].$$

Note that if $[\rho, \sigma] = 0$, this reduces to the classical relative entropy. Just like in the classical case, $D(\rho||\sigma) \geq 0$. From this, we get that

- $S(\rho) \leq d$
- $I(A; B) \geq 0$
- $S(A) \geq S(A|B)$

  We also have a Quantum Pinsker's Inequality.

  Theorem 1.

$$D(\rho||\sigma) \geq \frac{1}{2 \ln 2} ||\rho - \sigma||_1^2.$$

Now for asymmetric hypothesis testing. Our distributions now will be two quantum states, ρ and σ, and the test will be a set of measurement operators $\{M, I - M\}$ where an outcome of M means we say the state is ρ, and an outcome of $I - M$ means we say the state is σ. We now want to find

$$\beta_\epsilon^n = \min \left\{ \text{Tr} \left[M \sigma^{\otimes n} \right] \mid \text{Tr} M \rho^{\otimes n} \geq 1 - \epsilon \right\}.$$

Theorem 2 (Quantum Chernoff-Stein Theorem)**.**

$$\lim_{n \to \infty} \frac{-1}{n} \log \beta_\epsilon^n = D(\rho||\sigma).$$

Before looking at the proof, we will examine the case when ρ and σ are pure and $D(\rho||\sigma) = \infty$, i.e. $\text{supp}(\rho) \nsubseteq \text{supp}(\sigma)$. Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$. The measurement that achieves the desired rate is $M = I - B^{\otimes n}$, where $A = |\phi^{\perp}\rangle\langle\phi^{\perp}|$, and $B = I - A$. (A result of M is saying that you measured $|\phi^{\perp}\rangle$ at least once).

Then

$$\text{Tr}(M\sigma^{\otimes n}) = 0$$

while $\text{Tr}(M\rho^{\otimes n}) \to 1$ as $n \to \infty$ (in every register you get some probability of $|\phi^{\perp}\rangle$, so as $n \to \infty$ you are increasing your chances)

We will now prove the theorem.

Proof. We want an M such that $\text{Tr}(\rho^n M) \geq \alpha$ and $\text{Tr}(\sigma^n M) \leq e^{-nR}$. The idea will be to construct something similar to the LRT, but we will have to be careful about eigenbases.

Let $\rho = \sum_x r_x |\alpha_x\rangle\langle\alpha_x|$ and $\sigma = \sum_x s_x |\beta_x\rangle\langle\beta_x|$.

Recall the definition of a typical projector:

$$\Pi^n_{p,\delta} = \sum_{x^n : |\frac{1}{n}\sum_{i=1}^n \log r_{x_i} + \text{Tr}(\rho\log\rho)| \leq \delta} |\alpha_{x^n}\rangle\langle\alpha_{x^n}|.$$

Next, define

$$\Pi^n_{\rho||\sigma,\delta} = \sum_{x^n : |\frac{1}{n}\sum_{i=1}^n \log s_{x_i} - \text{Tr}(\rho\log\sigma)| \leq \delta} |\beta_{x^n}\rangle\langle\beta_{x^n}|.$$

We note that both of the subspaces defined by these projectors are typical under ρ, *i.e.*, $\text{Tr}(\rho^n \Pi^n_{p,\delta}) \geq 1-\epsilon$ and $\text{Tr}(\rho^{\otimes n}\Pi^n_{\rho||\sigma,\delta}) \geq 1-\epsilon$. We also have that $[\Pi^n_{\rho,\delta}, \rho^{\otimes n}] = 0$ and $[\Pi^n_{\rho||\sigma,\delta}, \sigma^{\otimes n}] = 0$.

If we sandwich $\rho^{\otimes n}$ between the typical projectors, we cut off the "atypical" eigenvalues:

$$e^{-n(S(\rho)+\delta)}\Pi^n_{p,\delta} \leq \Pi^n_{\rho,\delta}\rho^{\otimes n}\Pi^n_{\rho,\delta} \leq e^{-n(S(\rho)-\delta)}\Pi^n_{\rho,\delta}.$$

Similarly, if you do the conditional projection, it squishes the eigenvalues of σ into the following range:

$$e^{n(\text{Tr}(\rho\log\sigma)-\delta)}\Pi_{\rho||\sigma} \leq \Pi_{\rho||\sigma}\sigma^n\Pi_{\rho||\sigma} \leq e^{n(\text{Tr}(\rho\log\sigma)+\delta)}\Pi_{\rho||\sigma}. \tag{6.6}$$

(Note that from here on we will drop the δ and n on the typical projectors).

To get some intuition for equation 6.6, suppose you measure $\log\sigma = \sum_x \log s_x \beta_x$ on ρ. Then

$$\Pr[\log s_x] = \text{Tr}[\rho\beta_x],$$

and the expectation is $\text{Tr}\rho\log\sigma$. If you do this n times, the law of large numbers says that the average will approach the expectation.

We will first show achievability. Our measurement will be the product of both projectors - first measure $\{\Pi_{\rho||\sigma}, I - \Pi_{\rho||\sigma}\}$, and if you get the positive outcome $\Pi_{\rho||\sigma}$, then measure $\{\Pi_\rho, I - \Pi_\rho\}$.

More rigorously, define

$$M = \Pi_{\rho||\sigma}\Pi_\rho\Pi_{\rho||\sigma}.$$

Then

$$\text{Tr}\left(\rho^{\otimes n}M\right) = \text{Tr}\left(\Pi_\rho\Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}\Pi_\rho\right).$$

We need to show that the probability of ρ accepting is large. As we saw above, the probability of ρ accepting for each individual measurement is large. To show the combination works, we can use the Gentle Measurement lemma, which says that the state after accepting $\Pi_{\rho||\sigma}$ is still very close to ρ:

$$||\rho^{\otimes n} - \Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}||_1 \leq 2\sqrt{\epsilon}.$$

We then see that

$$\mathrm{Tr}(\Pi_\rho(\rho^{\otimes n} - \Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma})) \leq \frac{1}{2}||\rho^{\otimes n} - \Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}||_1 \leq \sqrt{\epsilon}$$

from which it follows that

$$\mathrm{Tr}(\Pi_\rho\Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}\Pi_\rho) \geq \mathrm{Tr}(\Pi_\rho\rho^{\otimes n}) - \sqrt{\epsilon} \geq 1 - \epsilon - \sqrt{\epsilon}.$$

So we have now showed that type one error is small enough, and now we need to bound type two error. Consider that

$$\begin{aligned} \mathrm{Tr}(M\sigma^{\otimes n}) &= \mathrm{Tr}\left(\Pi_\rho\Pi_{\rho||\sigma}\sigma^{\otimes n}\Pi_{\rho||\sigma}\right) && (6.7)\\ &\leq \mathrm{Tr}\left(\Pi_\rho e^{n(\mathrm{Tr}(\rho\log\sigma)+\delta)}\Pi_{\rho||\sigma}\right) && (6.8)\\ &\leq e^{n(S(\rho)+\delta)}e^{n(\mathrm{Tr}(\rho\log\sigma)+\delta)} && (6.9)\\ &= e^{-n(D(\rho||\sigma)-2\delta)} && (6.10) \end{aligned}$$

where we used the operator inequality from equation 6.6, the fact that $\Pi_{\rho||\sigma} \leq I$, and that $\mathrm{Tr}(\Pi_\rho) = |T_p| \leq e^{n(S(\rho)+\delta)}$.

We have now shown achievability. Next, we will use similar arguments to show the converse, i.e., you cannot do better than the rate $D(\rho||\sigma)$.

Suppose $\mathrm{Tr}\left(M\rho^{\otimes n}\right) \geq \alpha$. Our goal is to show that $\mathrm{Tr}\left(M\sigma^{\otimes n}\right)$ is "not too small". We will use the following operator inequalities:

$$\begin{aligned} \sigma^{\otimes n} &\geq \Pi_{\rho||\sigma}e^{n(\mathrm{Tr}(\rho\log\sigma)-\delta)} && (6.11)\\ \Pi_\rho\rho^{\otimes n}\Pi_\rho &\leq e^{-n(S(\rho)-\delta)}\Pi_\rho. && (6.12) \end{aligned}$$

Now

$$\begin{aligned} \mathrm{Tr}\left(M\sigma^{\otimes n}\right) &\geq \mathrm{Tr}\left(\Pi_{\rho||\sigma}M\right)e^{n(\mathrm{Tr}(\rho\log\sigma)-\delta)} && (6.13)\\ &\geq (\alpha - \sqrt{2\epsilon})e^{-n(D(\rho||\sigma)-2\delta)} && (6.14) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Tr}\left(\Pi_{\rho||\sigma}M\right) &\geq \mathrm{Tr}\left(\Pi_{\rho||\sigma}M\Pi_{\rho||\sigma}\Pi_\rho\right) && (6.15)\\ &\geq \mathrm{Tr}\left(M\Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}e^{-n(S(\rho)-\delta)}\right). && (6.16) \end{aligned}$$

We can again use Gentle Measurement to show that $\Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}$ is close to $\rho^{\otimes n}$, and using a similar argument as before, we get that

$$\mathrm{Tr}\left(M\Pi_{\rho||\sigma}\rho^{\otimes n}\Pi_{\rho||\sigma}\right) \geq \alpha - \sqrt{2\epsilon}.$$

Putting it all together, we find

$$\mathrm{Tr}\left(M\sigma^{\otimes n}\right) \geq (\alpha - \sqrt{2\epsilon})e^{-n(D(\rho||\sigma)-\delta)} \quad (6.17)$$

which completes the proof. □