PHY 4105: Quantum Information Theory
Lecture 4

Anil Shaji
School of Physics, IISER Thiruvananthapuram
(Dated: August 13, 2013)

A. Classical information theory

A sequence is ϵ-typical if

$$\left| -\frac{1}{N}\log p(x_1, \ldots x_N) - H(\vec{p})\right| \leq \epsilon.$$

Equivalently

$$2^{-N(H(\vec{p})+\epsilon)} \leq p(x_1, \ldots, x_N) \leq 2^{-N(H(\vec{p})-\epsilon)}.$$

We denote the set of all ϵ-typical sequences by $T(N, \epsilon)$.

Now consider the variable

$$S \equiv -\frac{1}{N}\log p(x_1, \ldots, x_N) = \frac{1}{N}\sum_{l=1}^{N} - \log p(x_l).$$

We can think of this variable as the sample mean of $-\log p(x)$. The mean of S is

$$\langle s\rangle = \frac{1}{N}\sum_{l=1}^{N}\Big(- \sum_{x_1,\ldots x_N} p(x_1)\cdots p(x_N)\log p(x_l)\Big) = \frac{1}{N}N(-\sum_{x_j} p(x_j)\log p(x_j)) = H(\vec{p}).$$

and

$$\langle(\Delta s)^2\rangle = \frac{1}{N}\langle(\Delta(-\log p(x)))^2\rangle = \frac{1}{N}\sum_x p(x)\Big(- \log p(x) - H(\vec{p})\Big)^2.$$

The asymptotic equipartition theorem or Typical sequences theorem

(i) For any ϵ, $\delta > 0$, there exists N_0 such that for all $N > N_0$, the probability that a sequence is ϵ-typical is $\geq 1 - \delta$

Proof:

$$\begin{aligned} p\left(\left| -\frac{1}{N}\log p(x_1, \ldots, x_N) - H(\vec{p})\right| \leq \epsilon\right) &= 1 - p\left(\left| -\frac{1}{N}\log p(x_1, \ldots, x_N) - H(\vec{p})\right| > \epsilon\right) \\ &\leq 1 - \frac{\langle(\Delta(-\log p(x)))^2\rangle}{N\epsilon^2}, \end{aligned}$$

The inequality coming from the weak law of large numbers of the form,

$$p(|s - \langle x\rangle| > \epsilon) \leq \frac{\langle(\Delta s)^2\rangle}{\epsilon^2} = \frac{\langle(\Delta x)^2\rangle}{N\epsilon^2},$$

We choose

$$N_0 = \frac{\langle (\Delta(-\log p(x)))^2 \rangle}{\delta \epsilon^2},$$

so that

$$p\left(\left| -\frac{1}{N} \log p(x_1, \ldots, x_N) - H(\vec{p}) \right| \leq \epsilon \right) \geq 1 - \delta.$$

(ii) The number of ϵ-typical sequences, $[T(N, \epsilon)]$, satisfies

$$(1-\delta) 2^{N(H(\vec{p})-\epsilon)} \leq [T(N, \epsilon)] \leq 2^{N(H(\vec{p})+\epsilon)}, \quad N \geq N_0.$$

Proof:

$$\begin{aligned} 1 &\geq \sum_{\epsilon\text{-typical sequences}} p(x_1, \ldots, x_N) \\ &\geq [T(N, E)] \min p(x_1, \ldots, x_N) = [T(N, E)] 2^{-N(H(\vec{p})+\epsilon)} \\ \Rightarrow [T(N, E)] &\leq 2^{N(H(\vec{p})+\epsilon)}. \end{aligned}$$

$$\begin{aligned} 1 - \delta &\leq \sum_{\epsilon\text{-typical sequences}} p(x_1, \ldots, x_N) \\ &\leq [T(N, E)] \max p(x_1, \ldots, x_N) = [T(N, E)] 2^{-N(H(\vec{p})-\epsilon)} \\ \Rightarrow [T(N, E)] &\geq (1-\delta) 2^{N(H(\vec{p})-\epsilon)}. \end{aligned}$$

(iii) Let S_N be *any* set of sequences of length N, containing at most 2^{NR} sequences, where $R < H(\vec{p})$. Given any $\delta > 0$, there exists N_0 such that for all $N \geq N_0$,

$$\sum_{x_1, \ldots x_N \in S_N} p(x_1, \ldots, x_N) \leq \delta.$$

Proof: Let $\epsilon < H(\vec{p}) - R$. For part (i), choose $\delta' = \delta/2$ with corresponding N_0' $(= 2N_0)$. Now

$$\sum_{x \in S_N} p(x) = \sum_{\epsilon-\text{typ}, x \in S_N} p(x) + \sum_{\epsilon-\text{atyp}, x \in S_N} p(x).$$

For $N \geq N_0'$, we have

$$\sum_{\epsilon-\text{typ}, x \in S_N} p(x) \leq 2^{NR} 2^{-N(H(\vec{p})-\epsilon)} = 2^{-N(H(\vec{p})-R-\epsilon)}$$

and

$$\sum_{\epsilon-\text{atyp}, x \in S_N} p(x) \leq \sum_{\epsilon-\text{atop}} p(x) = 1 - \sum_{\epsilon-\text{typ}} p(x) \leq \frac{\delta}{2},$$

the last inequality following from (i). So we have

$$\sum_{x \in S_N} p(x) \leq 2^{-N(H(\vec{p})-R-\epsilon)} + \frac{\delta}{2}.$$

Choose $N_0 \geq N_0'$ such that $2^{-N(H(\vec{p})-R-\epsilon)} \leq \delta/2$ so that

$$\sum_{x \in S_N} p(x) \leq \delta.$$

Shannon's noiseless coding theorem: The theorem is essentially a re-phrasing of the typical sequences theorem as applied to data compression. The theorem may be stated as:

> N i.i.d. random variables each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as N tends to infinity; but conversely, if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

In other words it says that typical sequences can be coded into a block code of "rate" $H(\vec{p})$, but not smaller.

The Shannon Entropy

The Shannon entropy or Shannon information we have seen, is defined as

$$H(\vec{p}) = -\sum_i p_i \log p_i = -\sum_x p(x) \log p(x) \equiv H(X).$$

Since we extensively deal with bits and qubits, a particular instance of the Shannon entropy that shows up frequently is the binary entropy,

$$H_2(x) = -x \log x - (1-x) \log(1-x).$$

The graph of the function is plotted below in Fig. 1:

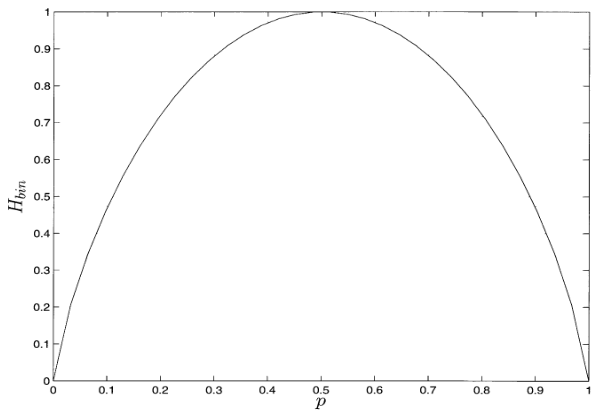

FIG. 1: The binary entropy function H_2

Note that $H_2(x) = H_2(1-x)$. Let us now list a few properties of $H(X)$

1. $0 \leq H(x) \leq \log D$, where D is the number of alternatives (dimension) for the random variable X.

2. We can define a **relative entropy** or the Kullback-Liebler distance between two probability distributions $\vec{p}$ and $\vec{q}$ as

$$H(\vec{p}\|\vec{q}) \equiv \sum_x p(x)\Big(- \log \frac{p(x)}{q(x)}\Big) = -H(\vec{p}) - \sum_x p(x) \log q(x) \geq 0.$$

We can use the convexity of the $-\log$ function to prove the last inequality:

$$H(\vec{p}\|\vec{q}) = \sum_x p(x)\Big(- \log \frac{p(x)}{q(x)}\Big) \geq - \log \Big(\sum_x p(x) \frac{q(x)}{p(x)}\Big) = 0.$$

In the above we have used Jensen's inequality which states that for a concave function $f(x)$,

$$\langle f(x) \rangle = \sum_x p(x) f(x) \leq f\Big(\sum_x p(x) x\Big) = f(\langle x \rangle),$$

and for a convex function $f(x)$

$$\langle f(x) \rangle = \sum_x p(x) f(x) \geq f\Big(\sum_x p(x) x\Big) = f(\langle x \rangle).$$

These inequalities follow in a simple manner from the definition of concave and convex functions as

$$\begin{aligned} f(\lambda x_1 + (1-\lambda) x_2) &\geq \lambda f(x_1) + (1-\lambda) f(x_2) \quad \text{(Concave)} \\ f(\lambda x_1 + (1-\lambda) x_2) &\leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad \text{(Convex)}. \end{aligned}$$

Jensen's inequality is a way of writing these definitions in terms of averages.

From the positivity of the relative entropy we can show that

$$0 \leq H(\vec{p}\|\vec{q}) = -H(\vec{p}) + \sum_x p(x) \log q(x) = -H(\vec{p}) + \log D,$$

when $q(x) = 1/D$ is uniformly distributed. Then

$$H(\vec{p}) = H(X) \leq \log D.$$

3. Concavity of the Shannon entropy:

$$H(\lambda \vec{p} + (1-\lambda)\vec{q}) \geq \lambda H(\vec{p}) + (1-\lambda) H(\vec{q}).$$

This means that mixing two probability distributions increases the entropy. We have equality when either $\lambda = 0$ or $\vec{q} = \vec{p}$.

Proof:

$$\begin{aligned} H(\lambda \vec{p} + (1-\lambda)\vec{q}) &= \sum_x -(\lambda p(x) + (1-\lambda) q(x)) \log(\lambda p(x) + (1-\lambda) q(x)) \\ &\geq -\lambda p(x) \log(\lambda p(x)) - (1-\lambda) \log((1-\lambda) q(x)) \\ &\geq -\lambda p(x) \log p(x) - (1-\lambda) q(x) \log q(x) \\ &\geq \lambda H(\vec{p}) + (1 - \lambda H(\vec{q})). \end{aligned}$$