07.02.24

$X_k,\ k = 1, 2, \dots, N$. Each X_k is distributed according to X

$$\Pr(x = x_j) = p_j, \quad j = 1, 2, \dots, d$$

Multinomial distribution Event $\to n_j = Np_j$

$$\begin{pmatrix}\text{Probability of}\\ \text{any typical}\\ \text{sequence}\end{pmatrix} = p_1^{n_1} \cdots p_d^{n_d} \xrightarrow{N\to\infty} 2^{-NH(X)}$$

$$e^{N \log e \left(-\sum_j p_j \log p_j\right)}$$

$$-\log(\quad) = NH(X)$$

$$\begin{pmatrix}\#\text{ of typical}\\ \text{sequences}\end{pmatrix} = \frac{N!}{n_1! \cdots n_d!} \xrightarrow{N\to\infty} e^{N\left(-\sum_j p_j \ln p_j\right)}$$

$$\boxed{\log x = \frac{\ln x}{\ln 2}} \Rightarrow \log(\quad) = \frac{\ln e^{N\left(-\sum_j p_j \ln p_j\right)}}{\ln 2}$$

$$= N\left(-\sum_j p_j \frac{\ln p_j}{\ln 2}\right)$$

$$= NH(X)$$

$$\begin{pmatrix}\text{Probability of}\\ \text{all typical sequences}\end{pmatrix} = 2^{NH(X)}\, 2^{-NH(X)} = 1$$

Typical sequences hog all the probability in the $N \to \infty$ limit

$$p(x_1,\dots,x_N) = e^{-NH(X)}$$ concatenation means seq.

$x_1 \cdots x_N \in \tau(N,\epsilon)$

Def$^{\underline{n}}$ A sequence is ϵ-typical if

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

$$2^{-N[H(X)+\epsilon]} \leq p(x_1,\dots,x_N) \leq 2^{-N[H(X)-\epsilon]}$$

Looks like a LOL-type statement

Sample mean $\quad S_N = -\frac{1}{N} \log p(x_1,\dots,x_N) = \frac{1}{N} \sum_{k=1}^{N} \underbrace{-\log p(x_k)}_{J_k}$

$$\langle S_N \rangle = H(X)$$

Easy computation of $\langle (\Delta S_N)^2 \rangle$ by recalling that it is $\frac{1}{N} \operatorname{Var} J$

$$= \frac{1}{N} \left\langle \left[\Delta(-\log p(x)) \right]^2 \right\rangle = \frac{1}{N} \sum_x p_x \left(-\log p_x - H(X) \right)^2$$

Average vs Single-shot

$$H(X) + \log p_x \leq \log d + \log p_x \leq \log d$$

Proven elsewhere

$$\Rightarrow \langle (\Delta S_N)^2 \rangle \leq \frac{(\log d)^2}{N}$$

$$\operatorname{Var} J \leq (\log d)^2$$

Source $\rightarrow$ Channel $\rightarrow$ Receiver

Thm Asyptotic Equipartition (AEP)

① For any $\epsilon, \delta > 0$, $\exists N_0$ s.t. $\forall N \geq N_0$, the probability that a sequence is ϵ-typical is $\geq 1-\delta$.

N that is considered to be "large" depends on the rate of convergence from the LOL

② The number of typical sequences is

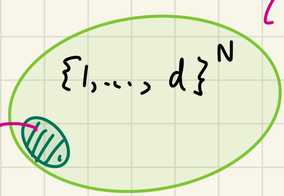
$$(1-\delta)\, 2^{N[H(X)-\epsilon]} \leq |\mathcal{T}(N,\epsilon)| \leq 2^{N[H(X)+\epsilon]}$$

③ Let R_N be any set of sequences of length N, $|R_N| \leq 2^{NR}$, and $R < H(X)$. For any $\delta \geq 0$, $\exists N_0$ s.t. $\forall N \geq N_0$,

Strict inequality

$$\sum_{x_1 \cdots x_N \in R_N} p(x_1, \ldots, x_N) \leq \delta$$

Pr $\leq \delta$



Sets smaller than $\mathcal{T}(N,\epsilon)$ (Doesn't have to be inside $\mathcal{T}(N,\epsilon)$) become negligible for some large N Can't be used to compress as good as $\mathcal{T}(N,\epsilon)$

① ■ As a consequence of the LOL,

$$\Pr\left\{\left|-\frac{1}{N}\log p(x_1, \ldots, x_N) - H(X)\right| \leq \epsilon\right\} \geq 1 - \frac{\operatorname{Var} J}{N\epsilon^2}$$

$$\geq 1 - \frac{(\log d)^2}{\underbrace{N\epsilon^2}_{\delta}} \qquad \square$$

Then choose $N_0 = \dfrac{(\log d)^2}{\delta \epsilon^2}$

If I want $\Pr \geq 0.9999$ $(\delta = 10^{-4})$, $\epsilon = 10^{-4}$, $d = 2 \Rightarrow N_0 = 10^{12}$

② **Upper bound**

$$1 \geq \sum_{\epsilon\text{-typical}} p(x_1,\dots,x_N) \geq |\tau(N,\epsilon)| \overbrace{\min p(x_1,\dots,x_N)}^{2^{-N[H(X)+\epsilon]}}$$

$$\Rightarrow \quad |\tau(N,\epsilon)| \leq 2^{N[H(X)+\epsilon]}$$

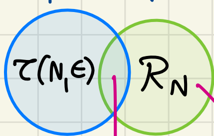
Lower bound

Part ①

$$1-\delta \leq \sum_{\epsilon\text{-typical}} p(x_1,\dots,x_N) \leq |\tau(N,\epsilon)| \overbrace{\max p(x_1,\dots,x_N)}^{2^{-N[H(X)-\epsilon]}}$$

$$\Rightarrow \quad |\tau(N,\epsilon)| \geq (1-\delta)\, 2^{N[H(X)-\epsilon]} \qquad \square$$

③



$$\sum_{x\in R_N} p(x_1,\dots,x_N) = \sum_{\text{typical}} p + \sum_{\text{atypical}} p$$

$$\left(N_0' = \frac{(\log d)^2}{\delta' \epsilon^2} \right)$$

$N \geq N_0'$

$$\leq 1 - 1 + \delta' = \delta'$$

Define this to be $\delta/2$ and try to bound the left term by $\delta/2$ as well.

$$\leq 2^{NR} \max p(x_1,\dots,x_N)$$

$$= 2^{NR}\, 2^{-N[H(X)-\epsilon]}$$

$$= 2^{-N[H(X)-R-\epsilon]}$$

Choose $N_0 \geq N_0'$ s.t. $2^{-N[H(X)-R-\epsilon]} \leq \delta'$

Then for $N \geq N_0$, $\displaystyle\sum_{x_1\cdots x_N \in R_N} p(x_1,\dots,x_N) \leq 2\delta' = \delta$

Contrapositive: if any set has an appreciable probability (bounded away from zero), then it is "as large as $\tau(N,\epsilon)$"

Shannon source coding theorem

A (lossy) (N,ϵ)-block code for i.i.d. random variables $X_1,\dots,X_N$ is a set $S \subset \underbrace{\Omega \times \cdots \times \Omega}_{N \text{ times}}$ such that

$$\Pr(x_1 x_2 \cdots x_N \in S) \geq 1-\delta$$

$\frac{\log|S|}{N}$ is called the rate of the code because it is the number of bits ($\log|S|$) per symbol (x_j for each j) used for encoding

base 2

The minimum possible asyptotic rate is the entropy $H(X)$ of the source

In information theory, the source coding theorem (Shannon 1948)[2] informally states that (MacKay 2003, pg. 81,[3] Cover 2006, Chapter 5[4]):

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

Similar statements & proofs can be found in Mackay's (free) electronic book: Information Theory, Inference, and Learning Algorithms