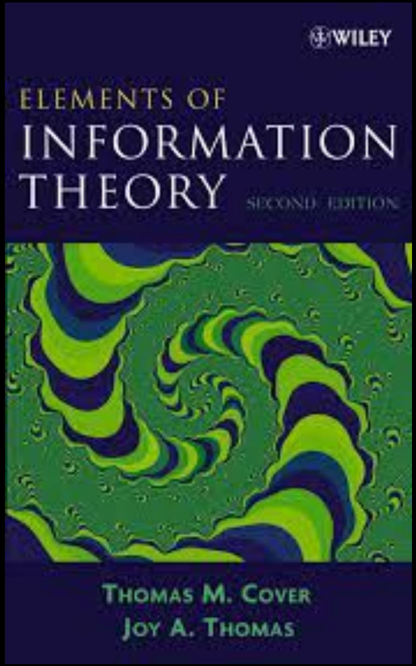
> “Information is the resolution of uncertainty.”
>
> Shannon

Erfan Mirzaei

Information Theory Mini-Course

Nov 2022

References and Acknowledges



Session 4

Fano Inequality,
Weak Law of Large Numbers,
AEP

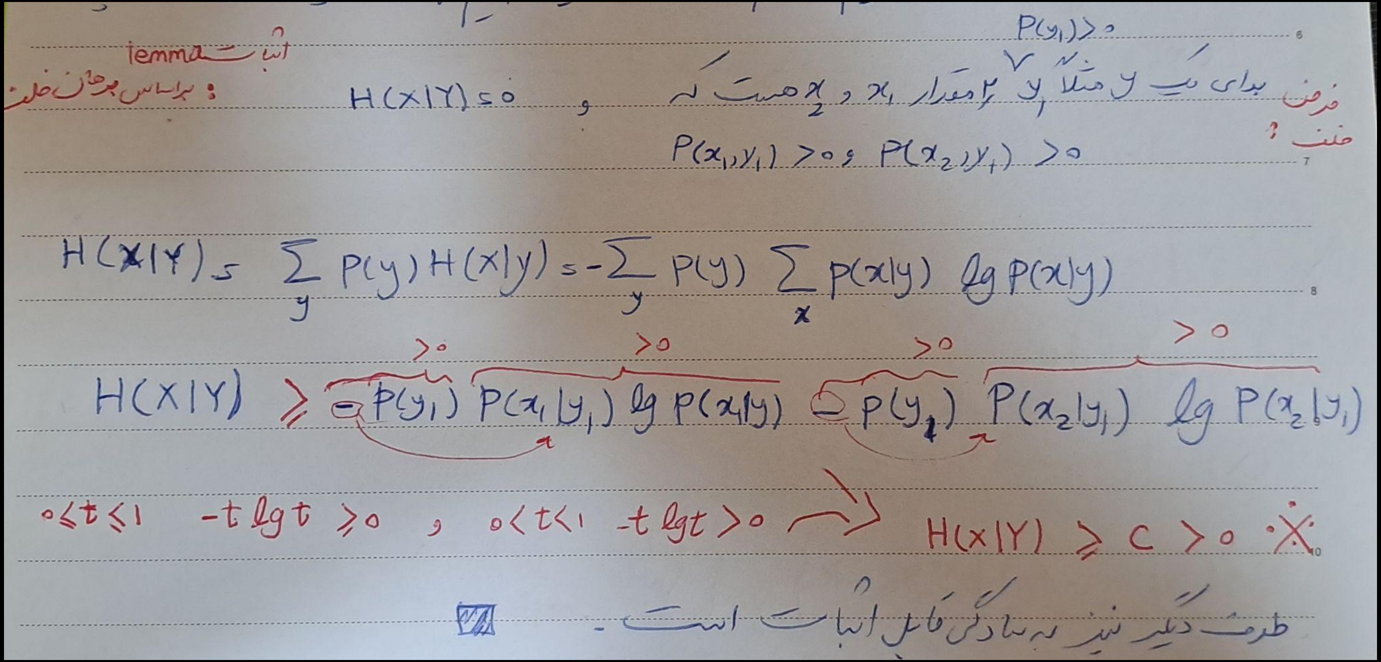
Fano Inequality

- Two correlated random variable X, Y.
- Y is known. We wish to guess the value of X.

- P_e = Pr {X_hat != X}
- X —> Y —> X_hat is a markov chain.
- When P_e is zero?

- When X is a function of Y. (for every y p(y) > 0 there is only one x p(x,y) >0)
- H(X | Y) = 0 ←→ X is a function of Y

Fano Inequality



اثبات lemma : براساس برهان خلف

فرض خلف: برای یک y مثلا y_1 ، ۲ مقدار x_1 و x_2 هست که $P(y_1)>0$ و $H(X|Y)=0$

$P(x_1,y_1)>0$ و $P(x_2,y_1)>0$

$$H(X|Y) = \sum_y P(y) H(X|y) = -\sum_y P(y) \sum_x P(x|y) \lg P(x|y)$$

$$H(X|Y) \geq \underbrace{-P(y_1)}_{>0}\,\underbrace{P(x_1|y_1) \lg P(x_1|y_1)}_{>0} \underbrace{-P(y_1)}_{>0}\,\underbrace{P(x_2|y_1) \lg P(x_2|y_1)}_{>0}$$

$0 \leq t \leq 1 \quad -t \lg t \geq 0$ ، $0<t<1 \quad -t \lg t > 0 \rightarrow H(X|Y) \geq c > 0$ ✗

طرف دیگر نیز به سادگی قابل اثبات است. ■

- What we can say when H(X | Y) > 0, and its relation to P_e?

Fano Inequality

Theorem 2.10.1 *(Fano's Inequality)* *For any estimator $\hat{X}$ such that $X \to Y \to \hat{X}$, with $P_e = \Pr(X \neq \hat{X})$, we have*

2

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y). \tag{2.130}$$

$\boxed{1}$

This inequality can be weakened to

$$1 + P_e \log |\mathcal{X}| \geq H(X|Y) \tag{2.131}$$

or

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}. \tag{2.132}$$

Fano Inequality

$$E = \begin{cases} 1 & \text{if } \hat{X} \neq X, \\ 0 & \text{if } \hat{X} = X. \end{cases}$$

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0} \\ &= \underbrace{H(E|\hat{X})}_{H(E|\hat{X}) \leq H(E) = H(P_e)} + \underbrace{H(X|E, \hat{X})}_{\leq P_e \log|\mathcal{X}|}. \end{aligned}$$

E is a function of X and $\hat{X}$

$$\begin{aligned} H(X|E, \hat{X}) &= \Pr(E=0) H(X|\hat{X}, E=0) + \Pr(E=1) H(X|\hat{X}, E=1) \\ &\leq (1 - P_e)0 + P_e \log|\mathcal{X}|, \end{aligned}$$

$$H(P_e) + P_e \log|\mathcal{X}| \geq H(X|\hat{X}).$$

- X —> Y —> X_hat, by data processing inequality I(X ; X_hat) =< I(X;Y) => H(X| X_hat) >= H(X | Y)
- Fano inequality is sharp.

Lemma

Lemma 2.10.1 *If X and X' are i.i.d. with entropy $H(X)$,*

$$\Pr(X = X') \geq 2^{-H(X)},$$

with equality if and only if X has a uniform distribution.

Proof: Suppose that $X \sim p(x)$. By Jensen's inequality, we have

f= 2^Y, Y = log p(X)

$$2^{E \log p(X)} \leq E2^{\log p(X)}, \tag{2.147}$$

which implies that

$$2^{-H(X)} = 2^{\sum p(x) \log p(x)} \leq \sum p(x) 2^{\log p(x)} = \sum p^2(x). \quad \square \tag{2.148}$$

Weak Law of Large Number

n 3 2 1

$$\frac{X1 + X2 + X3 + \ldots\ldots\ldots + Xn}{n \text{ (number of variables)}} \longrightarrow E(X) \quad \text{Where} \quad n \longrightarrow \infty$$

Asymptotic Equipartition property (AEP)

- Imagine you throw n dices.(n >>1) what is the probability of one specific realization?

Theorem 3.1.1 (*AEP*) *If $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$, then*

$$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(X) \qquad \text{in probability.}$$

Proof: Functions of independent random variables are also independent random variables. Thus, since the X_i are i.i.d., so are $\log p(X_i)$. Hence, by the weak law of large numbers,

$$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) = -\frac{1}{n}\sum_i \log p(X_i) \tag{3.3}$$

$$\to -E\log p(X) \quad \text{in probability} \tag{3.4}$$

$$= H(X), \tag{3.5}$$

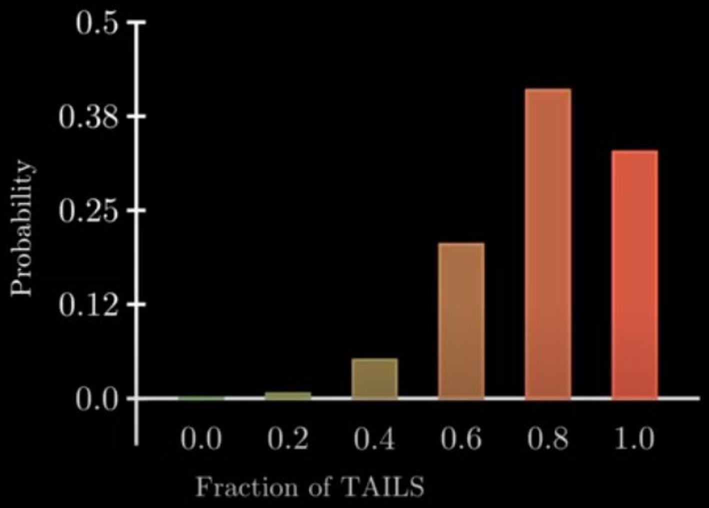
which proves the theorem. □

Asymptotic Equipartition property (AEP)

- When n —> inf, all sequence have the same probability.

P(Head) = 1/5

P(Tail) = 4/5



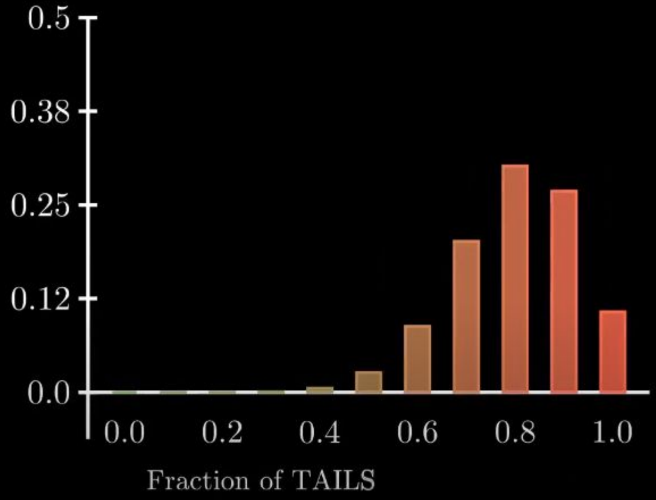
L = 5

Asymptotic Equipartition property (AEP)

- When n —> inf, all sequence have the same probability.

P(Head) = 1/5

P(Tail) = 4/5



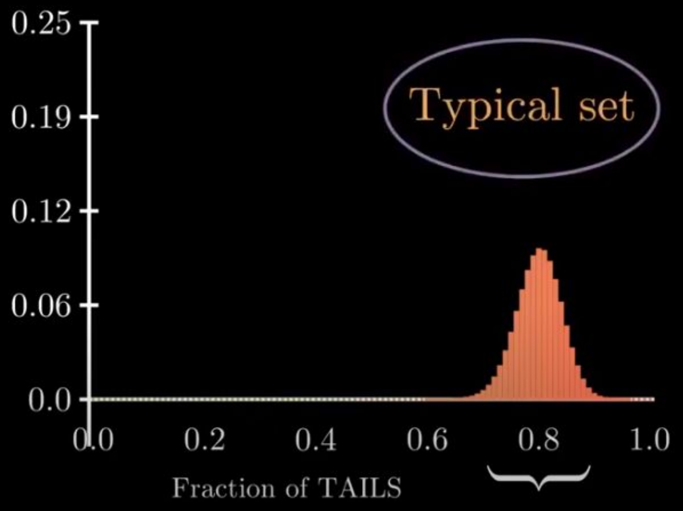
L = 10

Asymptotic Equipartition property (AEP)

- When n —> inf, all sequence have the same probability.

$P(\text{Head}) = 1/5$

$P(\text{Tail}) = 4/5$



$L = 100$

Typical Set

$k = np \in \mathbb{Z}$

رشته‌ی n بیتی $\quad Pr(1) = p \ ,\ Pr(0) = 1-p$

$Pr(\text{k بیت ۱ باشند}) = \binom{n}{k} \underbrace{p^k (1-p)^{n-k}}$

$2^{\lg_2 r} = r$

$2^{\lg_2 p^k (1-p)^{n-k}} = 2^{\lg_2 p^k + \lg_2 (1-p)^{n-k}} = 2^{k \lg_2 p + n-k \lg_2 1-p}$

$= \binom{n}{k} 2^{np \lg_2 p + n(1-p) \lg_2 (1-p)} = \binom{n}{k} 2^{n \overbrace{[p \lg p + (1-p) \lg (1-p)]}^{-H(p)}}$

Definition The *typical set* $A_\epsilon^{(n)}$ with respect to $p(x)$ is the set of sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with the property

$$2^{-n(H(X)+\epsilon)} \le p(x_1, x_2, \ldots, x_n) \le 2^{-n(H(X)-\epsilon)}. \tag{3.6}$$

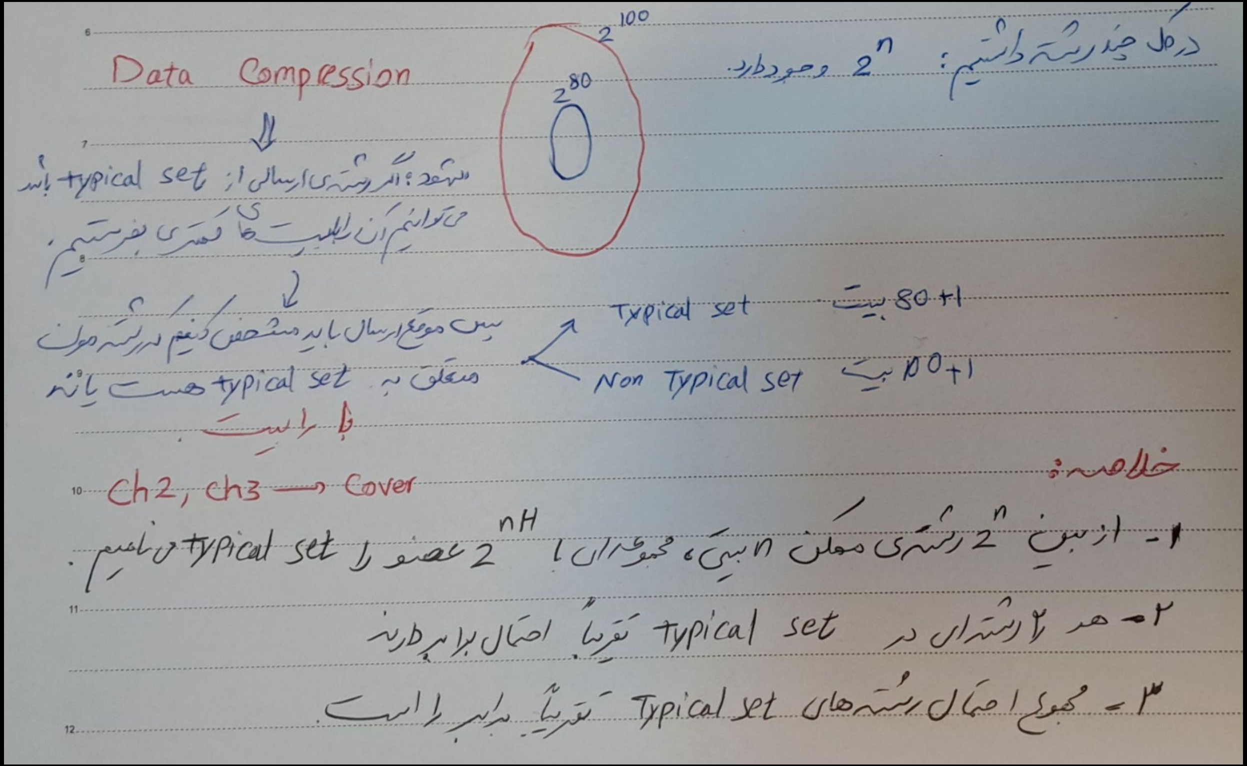
Typical Set

Theorem 3.1.2

1. *If* $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \le -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \le H(X) + \epsilon$.
2. $\Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon$ *for* n *sufficiently large.*
3. $|A_\epsilon^{(n)}| \le 2^{n(H(X)+\epsilon)}$, *where* $|A|$ *denotes the number of elements in the set* A.
4. $|A_\epsilon^{(n)}| \ge (1-\epsilon)2^{n(H(X)-\epsilon)}$ *for* n *sufficiently large.*

Thus, the typical set has probability nearly 1, all elements of the typical set are nearly equiprobable, and the number of elements in the typical set is nearly 2^{nH}.

Typical Set



Data Compression

در کل چند رشته داشتیم: 2^n وجود دارد.

2^{100}

2^{80}

مشاهده: اگر رشته‌ی ارسالی از typical set باشد می‌توانیم آن را با بیت‌های کمتری بفرستیم.

پس موقع ارسال باید مشخص کنیم که رشته‌مون متعلق به typical set هست یا نه

با ۱ بیت

Typical set — $80+1$ بیت

Non Typical set — $100+1$ بیت

Ch2, ch3 → Cover

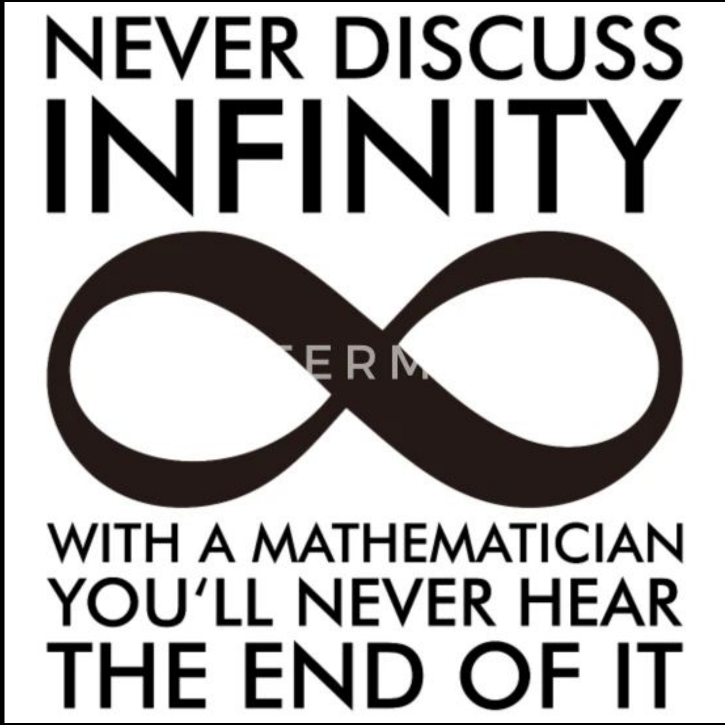
خلاصه:

۱- از بین 2^n رشته‌ی ممکن n بیتی، مجموعه‌ای از 2^{nH} عضو را typical set می‌نامیم.

۲- هر ۲ رشته‌ای در typical set تقریباً احتمال برابر دارند.

۳- مجموع احتمال رشته‌های typical set تقریباً برابر ۱ است.

Infinity


NEVER DISCUSS
INFINITY
WITH A MATHEMATICIAN
YOU'LL NEVER HEAR
THE END OF IT


Everything is nothing
with a twist.
~Kurt Vonnegut

Thanks for your attention

Erfan Mirzaei
erfunmirzaei@gmail.com