엔트로피 기본정의

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

$$E_p g(X) = \sum_{x \in \mathcal{X}} g(x) p(x),$$

$$g(X) = \log \frac{1}{p(X)}.$$

$$H(X) = E_p \log \frac{1}{p(X)}.$$

따라서

$H(X) = -E \log p(x)$ 로

나타내며

E는 $\sum_X p(x)$를 나타냄

이런 원리로

$H(Y|X)$ 를 구하면

$H(Y|X) = -E \log P(Y|x)$ 이며

$E = \sum_X p(x) \sum_Y P(Y|x)$ (Y는 X조건하이므로)

즉, $H(Y|X) = -\sum_X p(x) \underbrace{\sum_Y P(Y|X) \log P(Y|X)}_{H(Y|X=x)}$

$H(Y|X) = -\sum_x p(x) H(Y|X=x)$

Chain Rule 증명

$H(X,Y) = H(X) + H(Y|X).$

$$H(X,Y) = -E \log P(x,Y)$$

이때 $P(x|Y) = \dfrac{P(x,Y)}{P(x)}$ 이며

$E = \sum_x P(x) \sum_y P(Y)$ 이므로

$$H(X,Y) = -\sum_x P(x) \sum_y P(Y) \log P(x) P(x|Y)$$

$$= -\Big(\sum_x P(x) \underbrace{\sum_y P(Y)}_{=1} \log P(x) + \sum_x P(x) \sum_y P(Y) \log P(x|Y) \Big)$$

$$H(X|Y) = -E \log P(X|Y)$$

$$= -\sum_x P(X|Y) \sum_y P(Y) \log P(X|Y)$$

$$= -\sum_Y P(Y) \sum_X P(X|Y) \log P(X|Y)$$

$$= -\sum_Y P(Y)\, H(X|Y=y)$$

$$= -\Big(\sum_x P(x) \log P(x) + \big(-\underbrace{\sum_x P(x)}_{=1}\big) H(X|Y) \Big)$$

$$= -\big(-H(X) - H(X|Y) \big)$$

$$= H(X) + H(X|Y)$$

Introduction to Computer Science & Engineering

Lecture 11: A Few Basics of Information Systems

Jeonghun Park

KNU KYUNGPOOK NATIONAL UNIVERSITY

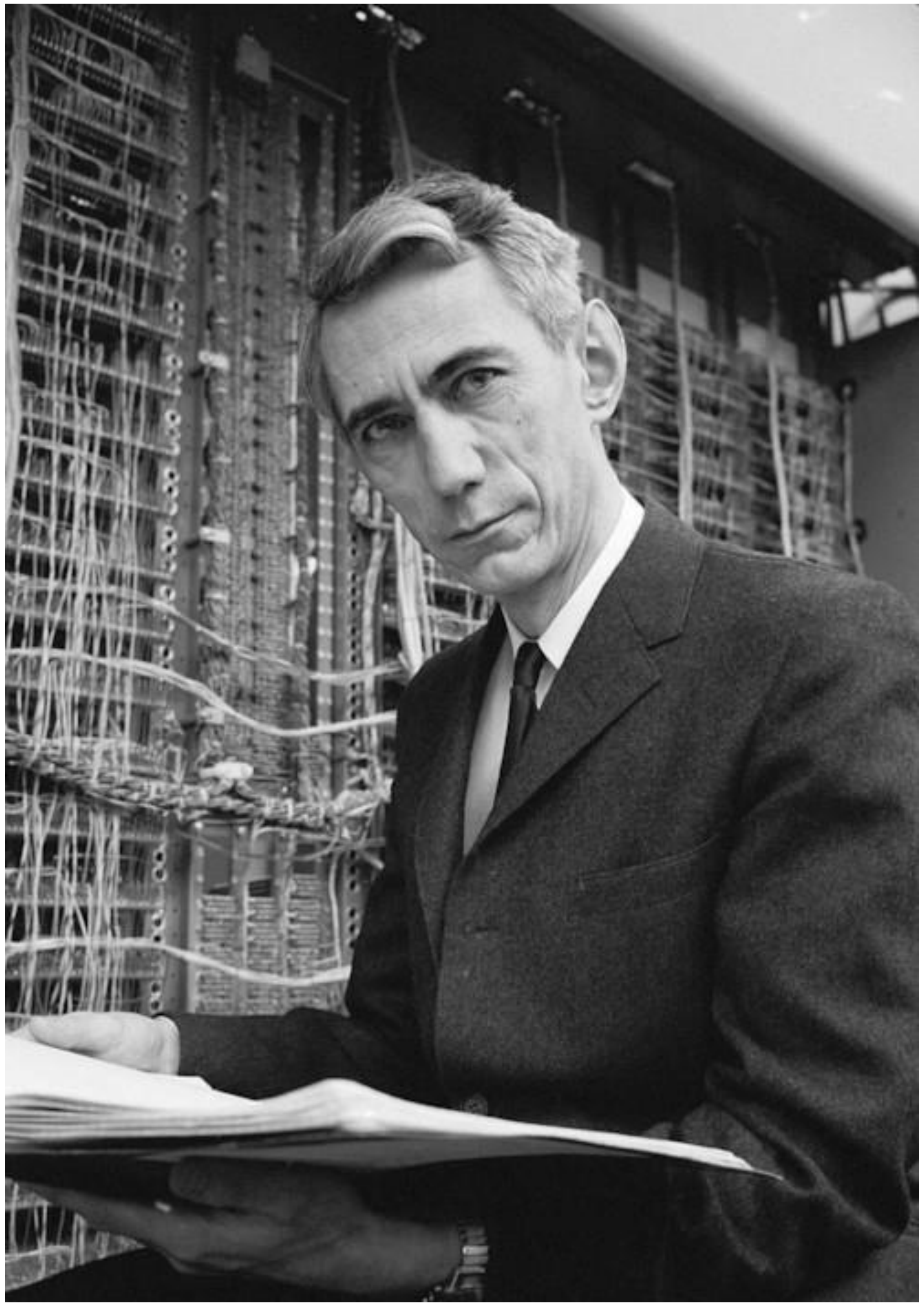
Introduction

- What is information theory?
  - Define and play with the information measure
  - Mathematical answers for two fundamental questions:
    - What is the ultimate data compression?
    - What is the ultimate transmission rate of communication?
  - Beyond this, it has significant contributions to make in statistical physics, computer science, statistical inference, and probability and statistics etc..

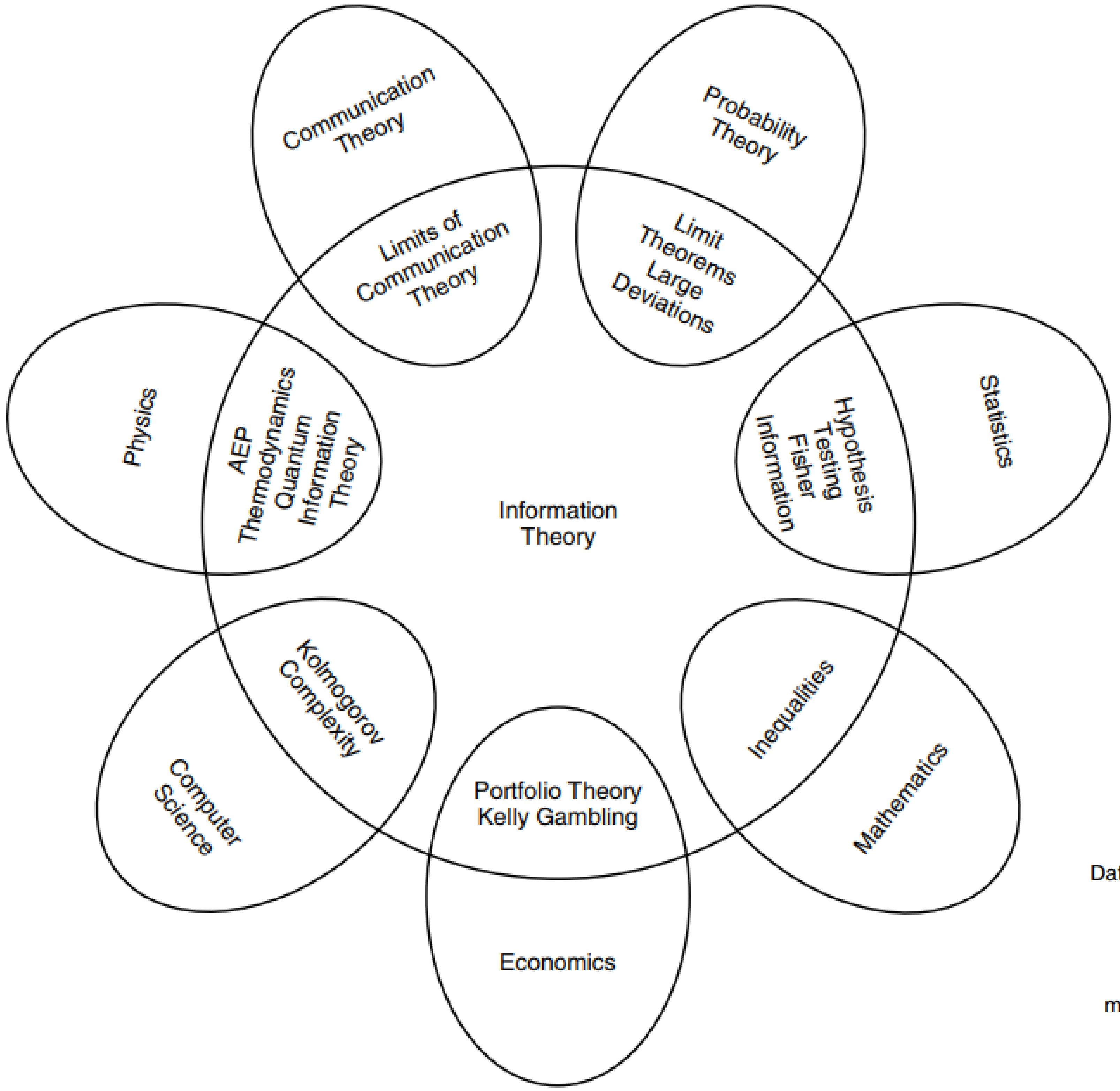
jeonghun.park@knu.ac.kr

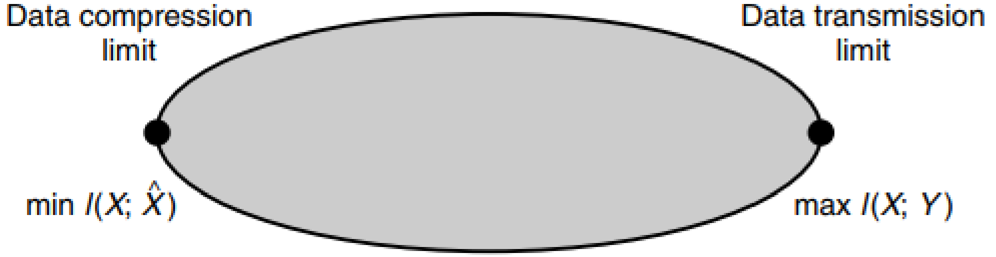
History

- Originated from one genius
  - Claude Shannon (1916.4.30 ~ 2001.2.24)
  - Ph.D. from MIT
  - The author of
    - ***A Mathematical Theory of Communications***
  - The farther of digital comm.
  - Stock master
    - Isaac Newton
    - Warren Buffett



Impact


Communication Theory
Probability Theory
Limits of Communication Theory
Limit Theorems Large Deviations
Physics
AEP Thermodynamics Quantum Information Theory
Information Theory
Hypothesis Testing Fisher Information
Statistics
Kolmogorov Complexity
Computer Science
Portfolio Theory Kelly Gambling
Economics
Inequalities
Mathematics


Data compression limit
Data transmission limit
min $I(X; \hat{X})$
max $I(X; Y)$

Entropy (무질서도)

- Measure of the fundamental uncertainty of randomness
- Let X be a discrte random variable
- 그 때의 확률
  $p(x) = \mathbb{P}[X = x]$
  Random Variable
- Then the entropy is
  - $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$
  - $H(X) = \mathbb{E}[\log_2 p(x)^{-1}]$
    Average.

$H(X) = - p(x) \log_2 p(x)$

$H(X, Y) = -E \log p(X, Y).$

More on Entropy

확률만 안다면 구할 수 있음.

Binary tree 처럼 특별한 method 를 쓰는게 아님.

- The units of the entropy is *“bits”*
- Some properties
  - $H(X) \geq 0$ since $p(x) \leq 1$

    ① 확률의 합은 1이어야 하므로 $H(X) \geq 0$
- Let $X = 1$ with probability p

  $X = 0$ with $1-P$
  - Then we have $H(X) = -p\log_2 p - (1-p)\log_2(1-p)$

$$H(X) = -\sum_{x \in \mathcal{X}} P(x)\log_2 P(x)$$

$$= -P(0)\log_2 P(0) - P(1)\log_2 P(1)$$

$$= -(1-P)\log_2(1-P) - P\log_2 P$$

So, What is Entropy?

- It presents the fundamental uncertainty of the randomness 랜덤에 대한 불확실성을 나타냄.
- Coin toss example

Head가 $\frac{1}{2}$ 이면 tail도 $\frac{1}{2}$ 이므로 엔트로피가 가장큼. 불확실성이 가장큼

- We consider two cases:
  - Head with probability ½
  - Head with probability 0

Which case is more uncertain?

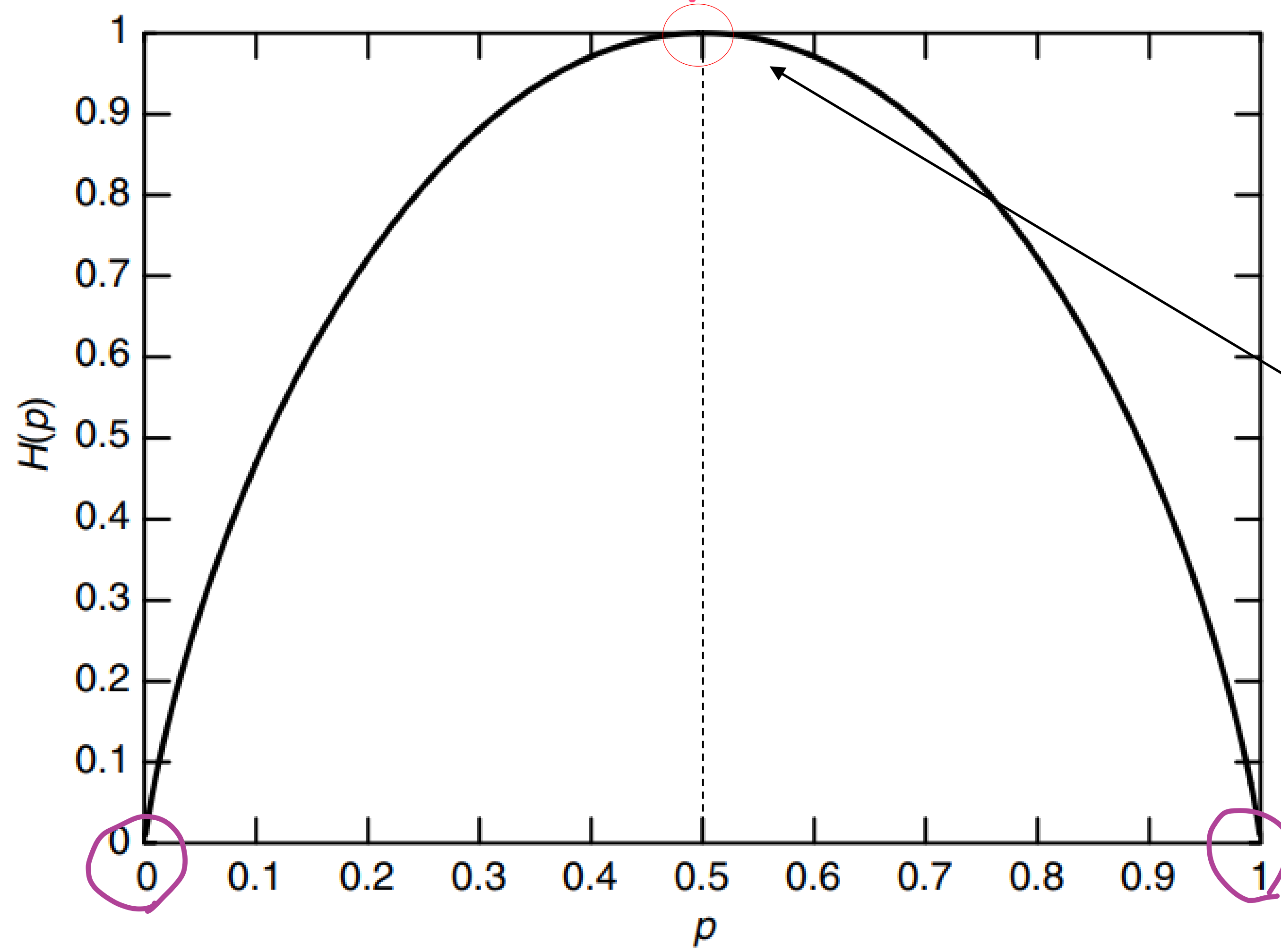
Head가 0 이면 tail은 1 이므로 무조건 tail 이므로 bit 자체가 필요없다.

jeonghun.park@knu.ac.kr

Entropy of Coin Toss

- Remember?
  - $H(X) = -p \log_2 p - (1-p)\log_2(1-p)$

⇒ 동전 앞,뒷면이 $\frac{1}{2}$ 확률일때 엔트로피가 가장크다



What is the meaning of this?

⇒ 불확정도가 높음

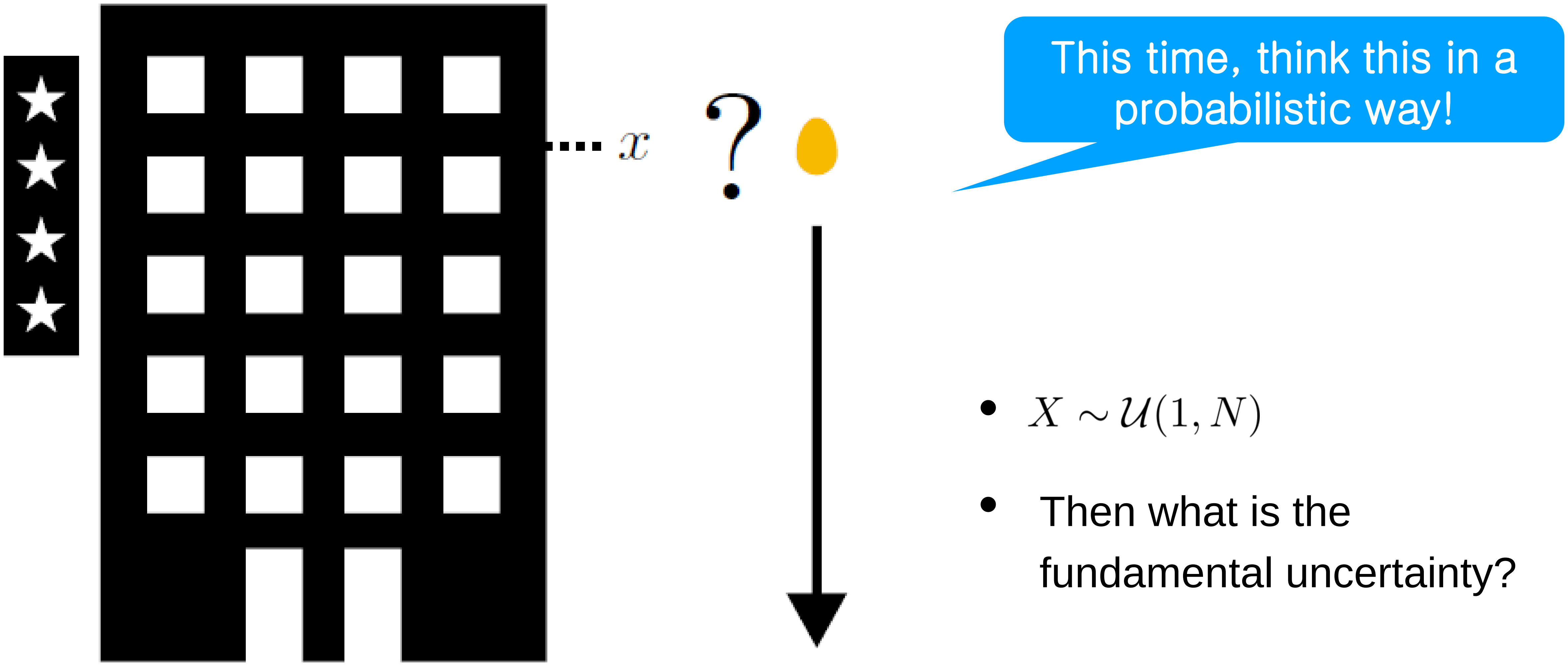
비트를 쓸 필요가 없다.

Twenty-Questions Interpretations

- Entropy can be interpreted as follows:
  - The minimum number of the questions to eliminate the randomness
  - Think about 스무고개
- One important (and surprising) fact:
  - We don’t have to know about the specific questions!

⇒ 확률만 안다면 어떤 질문인지도 모르고 몇번째 질문에서 맞출수 있는지 알 수 있다.

Egg Drop Rerevisit



- $X \sim \mathcal{U}(1, N)$
- Then what is the fundamental uncertainty?

Compute Entropy (Randomless 평가)

만약 스무고개에서 $P[X=x] = \frac{1}{N} \forall x \in [1,N]$
일때 $P(x) = \frac{1}{N}$ 으로 두고
엔트로피를 구해, 되도 질문수를 구할수 있다.

- $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$
- $\boxed{\mathbb{P}[X = x] = \frac{1}{N}}$ $\boxed{\forall x \in [1, N]}$

1 ~ N 사이에 있는 모든 x 에 대해

모든 확률이 Uniform 할때

- $H(X) = \underbrace{-\frac{1}{N}\log_2(\frac{1}{N}) - \frac{1}{N}\log_2(\frac{1}{N}) - \ldots - \frac{1}{N}\log_2(\frac{1}{N})}_{N}$

$H(X) = -\log_2 \frac{1}{N} = \log_2 N$

- $H(X) = -\log_2(N^{-1}) = \log_2(N)$

Compare the complexity of our binary search method
= The fundamental number of the questions to eliminate the randomness

* Binary Search 에서는 $\frac{N}{2^n} \le 1$ 만족하는 $n = \log_2 N$ 을 찾았기는
Entropy 에서는 확률로 구함.

jeonghun.park@knu.ac.kr

H(X)가 크다는건, 불확정성이 높다는것이다.

$H(X) =$ X's entropy

$X_1 = 1, 2, 3$
$P(X_1) = \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ 일때 → 엔트로피가 더 큼 (uniform)

$X_2 = 1, 2, 3, 4, 5, 6, 7$
$P(X_2) = 0.8, 0.1, \ldots, \ldots, \ldots, \ldots, \ldots$ 일때 → 엔트로피가 더 작음

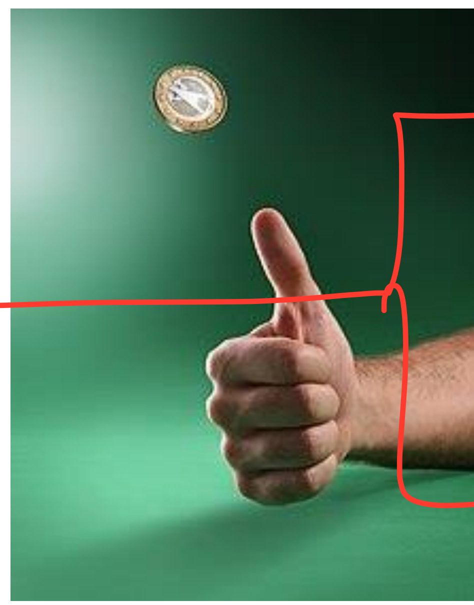
One Surprise

- In entropy computation, we did not use any specific egg drop method
- Try to understand what this means:
  - Without ***any specific question***, we can calculate **the fundamental limit of the number of questions** to find an answer

Conditional Coin Toss

H 나왔을때 4~7을 가짐

이거
확률은
$\frac{1}{2}, \frac{1}{2}$



H

$$X \sim \mathcal{U}(4, 7)$$

4~7까지 유니폼한 $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ 확률

T

$$X \sim \mathcal{U}(1, 10)$$

1~10까지 유니폼한 $\frac{1}{10}, \frac{1}{10}, \frac{1}{10} \cdots$

T 나왔을때
1~10을 가짐.

What is the entropy?

KNU KYUNGPOOK NATIONAL UNIVERSITY

Joint and Conditional Entropy

Only one Conditional Entropy $H(X) = -\sum_x p(x) \log_2 p(x)$

- Let Y = coin.
- The joint entropy is
  - $H(X,Y) = -\sum_x \sum_y p(x,y) \log_2 p(x,y)$
- The conditional entropy is 조건부 엔트로피
  - $H(X|Y=H) = -\sum_x p(x|Y=H) \log_2 p(x|Y=H)$

$$H(X|Y) = \mathbb{P}[Y=H]H(X|Y=H) + \mathbb{P}[Y=T]H(X|Y=T)$$

Joint and Conditional Entropy

- Consider the following joint distribution:

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

$\rightarrow \sum = 1$

→ Joint Entropy는 주어진 확률 그대로 써도됨.

단, Y=1 일때로 Conditional entropy 구하면

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$

다 합쳐도 1이 아니므로 다른 방법을 써야함

- Calculate the joint and conditional entropy
  - ▶ Understand the marginal distribution

Y \ X	1	2	3	4	$P(Y)$
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
4	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$
$P(X)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	

$$H(Y) = -\sum_y P(y)\log P(y) = \frac{2}{4}+\frac{2}{4}+\frac{2}{4}+\frac{2}{4} = 2$$

$$H(X) = -\sum_x P(x)\log P(x) = \frac{1}{2}+\frac{2}{4}+\frac{3}{8}+\frac{3}{8} = \frac{7}{4}$$

① $H(Y|X)$

$$= -E\log_2 P(y|x)$$

$$= -\sum_x p(x)\sum_y P(y|x)\log_2 P(y|x)$$

$$= \sum_x p(x) H(Y|X=x)$$

$$= \sum_{i=1}^{4} P(X=i)\, H(Y|X=i)$$

③ $H(X,Y)$

$$= \sum_{x,y} P(x,y)\log P(x,y)$$

② $H(X|Y) = \sum_{i=1}^{4} P(Y=i)\, H(X|Y=i)$

$$\frac{1}{4}\times H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{8}\right) = \frac{1}{4}\left(\frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{3}{8}\right) = \frac{1}{4}\cdot\frac{7}{4}$$

$+$

$$\frac{1}{4}\times H\left(\frac{1}{4},\frac{1}{2},\frac{1}{8},\frac{1}{8}\right) = \frac{1}{4}\left(\frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{3}{8}\right) = \frac{1}{4}\cdot\frac{7}{4}$$

$+$

$$\frac{1}{4}\times H\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right) = \frac{1}{4}\left(\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\right) = \frac{1}{4}\cdot 2$$

$+$

$$\frac{1}{4} H(1,0,0,0) = \frac{1}{4}(0+0+0+0) = 0$$

$$= \frac{1}{4}\left(\frac{7}{4}+\frac{7}{4}+2\right) = \frac{11}{8}$$

H, T 나올 확률이 확률은 $\frac{1}{2}$이면

Math Fun: Complicated Coin Toss

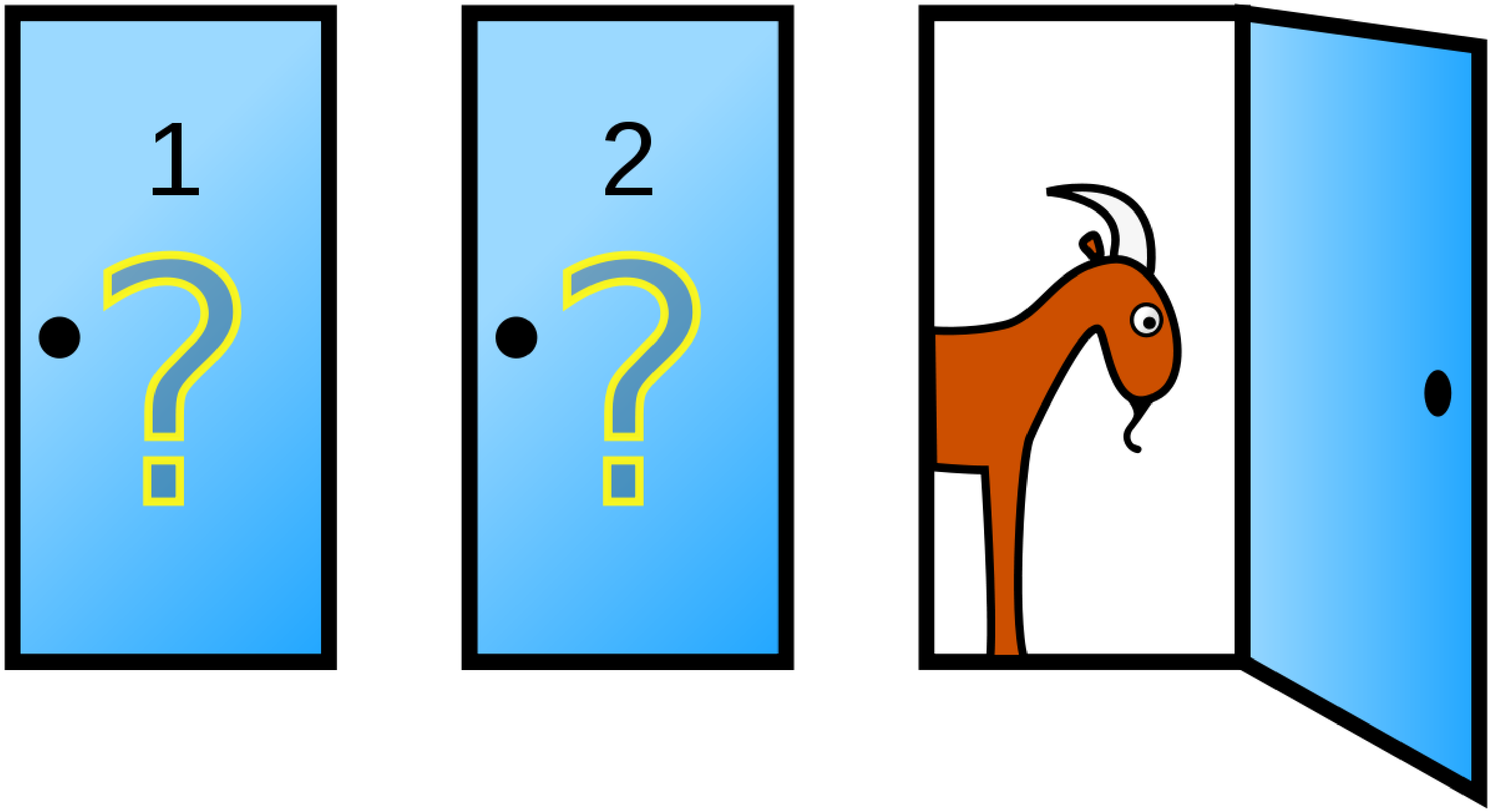
① H $(X=1)$ $P=\frac{1}{2}$
② T H $(X=2)$ $P=\frac{1}{4}$
③ T T H $(X=3)$ $P=\frac{1}{8}$
④ T T T H $(X=4)$ $P=\frac{1}{16}$
Ⓝ T T T ⋯ H $(X=N)=\frac{1}{2^N}$

- A fair coin is flipped until the first head occurs. We are interested in X, the number of toss required.
  - Find the entropy

$$H(X) = -\sum_x P(x)\log_2 P(x)$$
$$= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \cdots - \frac{1}{2^N}\log_2\frac{1}{2^N}$$
$$= +\frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 2^2 + \frac{1}{8}\log_2 2^3 + \cdots + \frac{1}{2^N}\log_2 2^N$$
$$= \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \cdots + \frac{1}{2^N}N$$
$$= \sum_{n=1}^{\infty} 2^{-n}\cdot n$$

$$S = r + 2r^2 + 3r^3 + \cdots + nr^n$$
$$rS = r^2 + 2r^3 + 3r^4 + \cdots + nr^{n+1}$$
$$S(1-r) = r + r^2 + r^3 + \cdots + r^n - nr^{n+1}$$

Math Fun: Monty-Hall Problem



Math Fun: Monty-Hall Problem

- Can we use information theory to analyze this problem?
- The intuition in the previous lecture is WRONG! (sorry for confusion)
- We can interpret that $\overbrace{H(X)}^{\text{그냥고른경우}} \geq \overbrace{H(X|Y)}^{\text{염소를 보여준 경우}}$
  - Where the equality holds iff Y is independent to X

Conditional Entropy and Marginal Distribution

다 합쳐서 1이 안됨

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

$$H(X|Y=1)$$
$$= \sum_{x\in\mathcal{X}} \boxed{p(x|Y=1)} \log_2 p(x|Y=1)$$

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

- Then what is $p(Y=1)$?
- This is called "marginal distribution," and this can be obtained as:

$$p(Y=1) = \sum_{x\in\mathcal{X}} p(Y=1,x)$$
$$= \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4}$$

Chain Rule

- The following chain rule is satisfied:

$$H(X,Y) = H(X) + H(Y|X)$$

H(Y|X)

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

- Proof:

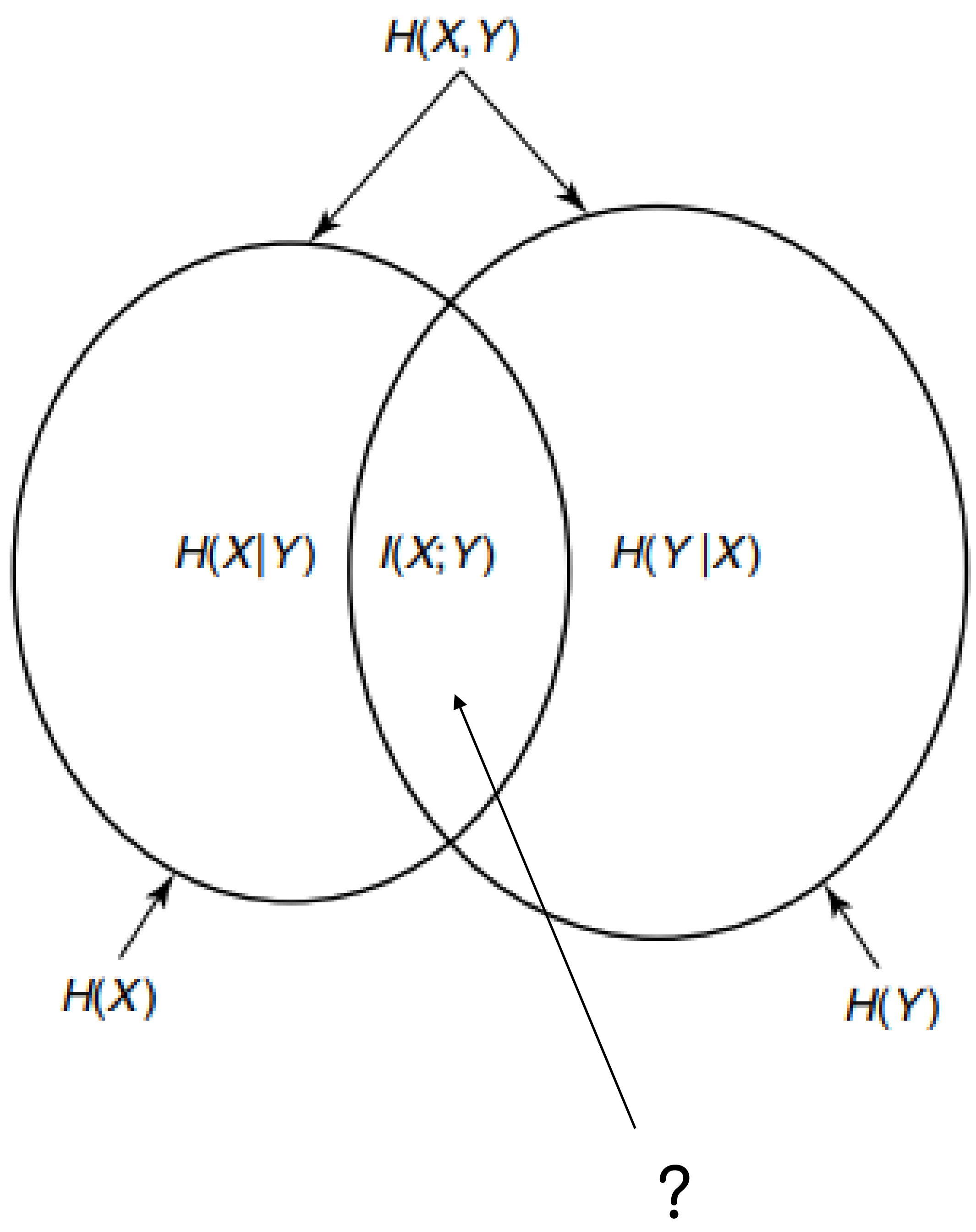
$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 \boxed{p(x,y)}$$

⇒ Joint Entropy

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 p(x)p(y|x)$$

$$= \underbrace{-\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)}_{H(X)} - \underbrace{-\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 p(y|x)}_{H(Y|X)}$$

Intuitive Description



Relations of Multiple Randomness

- Entropy
  - Measure of the ***uncertainty*** of the randomness
  - Measure of the amount of information required on the average to describe the random variable
- Now we focus on some relationships of two randomness

jeonghun.park@knu.ac.kr

The *relative entropy* is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(p||q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. For example, if we knew the true distribution p of the random variable, we could construct a code with average description length $H(p)$. If, instead, we used the code for a distribution q, we would need $H(p) + D(p||q)$ bits on the average to describe the random variable.

Definition The *relative entropy* or *Kullback–Leibler distance* between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{2.26}$$

$$= E_p \log \frac{p(X)}{q(X)}. \tag{2.27}$$

Relative Entropy

- Measure of ***distance*** between two distributions
  - Denote as $D(p||q)$
- It can be interpreted as:
  - The amount of inefficiency of assuming that the distribution is q when the true distribution is p
  - For example, assume that we use the distribution q, while the true distribution is p
  - Then the code length is $H(p) + D(p||q)$
  - (Think our example)

Relative Entropy

- The relative entropy (or denoted as *Kullback-Leibler distance*) between two probabilities p(x) and q(x) is defined as

엔트로피와 다른점

  - $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}$

- 가 없고 $\frac{p(x)}{q(x)}$

- If it is a distance, the following should be satisfied:
  - $D(p||q) = D(q||p)$ => False

P, Q는 기본값으로

Is this true?

measure of space 가 같다.

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$
$$= D(p(x,y)||p(x)p(y))$$
$$= E_{p(x,y)}\log\frac{p(X,Y)}{p(X)p(Y)}.$$

$$D(p||q) = p(0)\log\frac{p(0)}{q(0)} + p(1)\log\frac{p(1)}{q(1)}$$
$$+ \cdots$$

Example 2.3.1 Let $\mathcal{X} = \{0, 1\}$ and consider two distributions p and q on $\mathcal{X}$. Let $p(0) = 1 - r$, $p(1) = r$, and let $q(0) = 1 - s$, $q(1) = s$. Then

$$D(p||q) = (1-r)\log\frac{1-r}{1-s} + r\log\frac{r}{s} \tag{2.31}$$

and

$$D(q||p) = (1-s)\log\frac{1-s}{1-r} + s\log\frac{s}{r}. \tag{2.32}$$

If $r = s$, then $D(p||q) = D(q||p) = 0$. If $r = \frac{1}{2}$, $s = \frac{1}{4}$, we can calculate

$$D(p||q) = \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{3}{4}} + \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{4}} = 1 - \frac{1}{2}\log 3 = 0.2075 \text{ bit}, \tag{2.33}$$

whereas

$$D(q||p) = \frac{3}{4}\log\frac{\frac{3}{4}}{\frac{1}{2}} + \frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{2}} = \frac{3}{4}\log 3 - 1 = 0.1887 \text{ bit}. \tag{2.34}$$

Note that $D(p||q) \neq D(q||p)$ in general.

Mutual Information

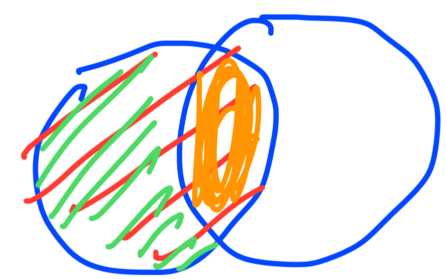
- The mutual information is defined as follows:
  - $I(X;Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} = D(p(x,y)||p(x)p(y))$

$p(x|y) = \frac{p(x,y)}{p(y)}$

$$= \sum_{x,y} p(x,y) \log_2 \frac{p(x|y)}{p(x)}$$

$$= \sum_{x,y} p(x,y) \log_2 p(x) + \sum_{x,y} p(x,y) \log_2 p(x|y)$$

$$= H(X) - H(X|Y)$$



- By symmetry, we also have
  - $I(X;Y) = I(Y;X) = H(Y) - H(Y|X)$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)}$$

$$= \sum_{x,y} p(x,y) \log p(x|y) - \sum_{x,y} p(x,y) \log p(x)$$

$$= -H(X|Y) - \sum_{x} p(x) \log p(x)$$

$$= -H(X|Y) + P(X)$$

Mutual Information

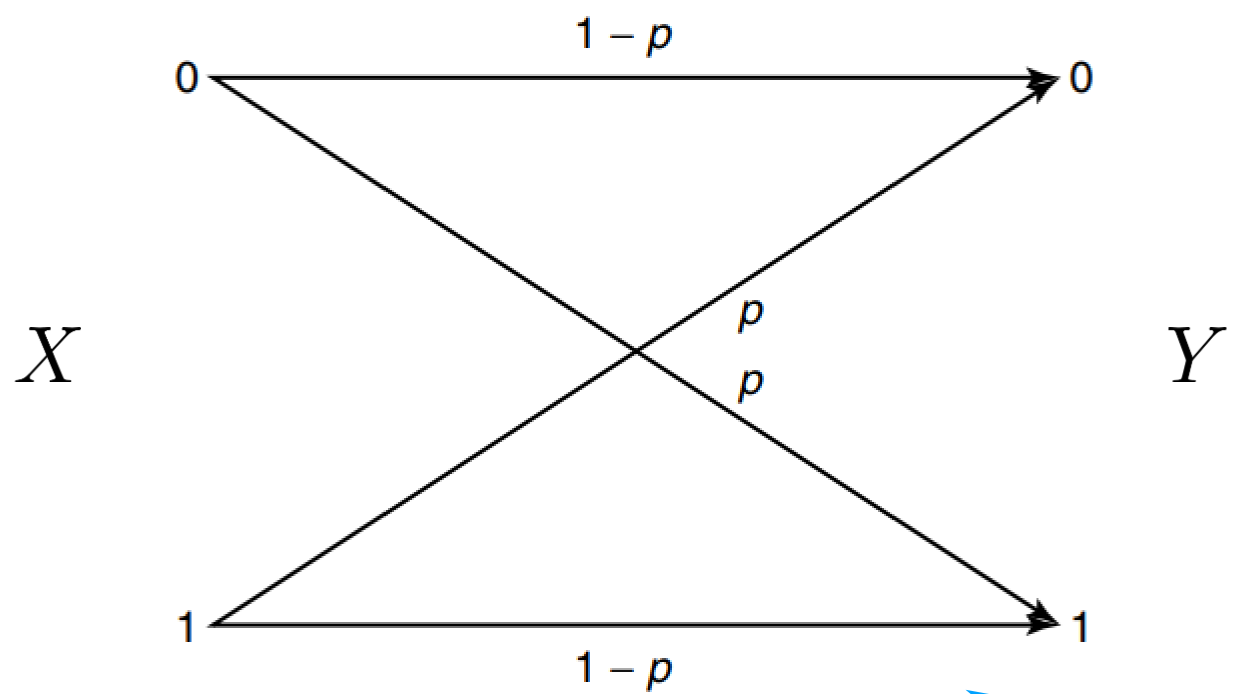
주파수 효율, 에너지 효율

- Fundamental of communications

어떤 measure 도 가능

- X is random, and Y is also random
  - ▸ They are not same (because of the channel), but “related” each other
  - ▸ How to define the maximum information amount that can be successfully delivered through the channel?

Toy Example



$p =$ 에러

Binary symmetric channel

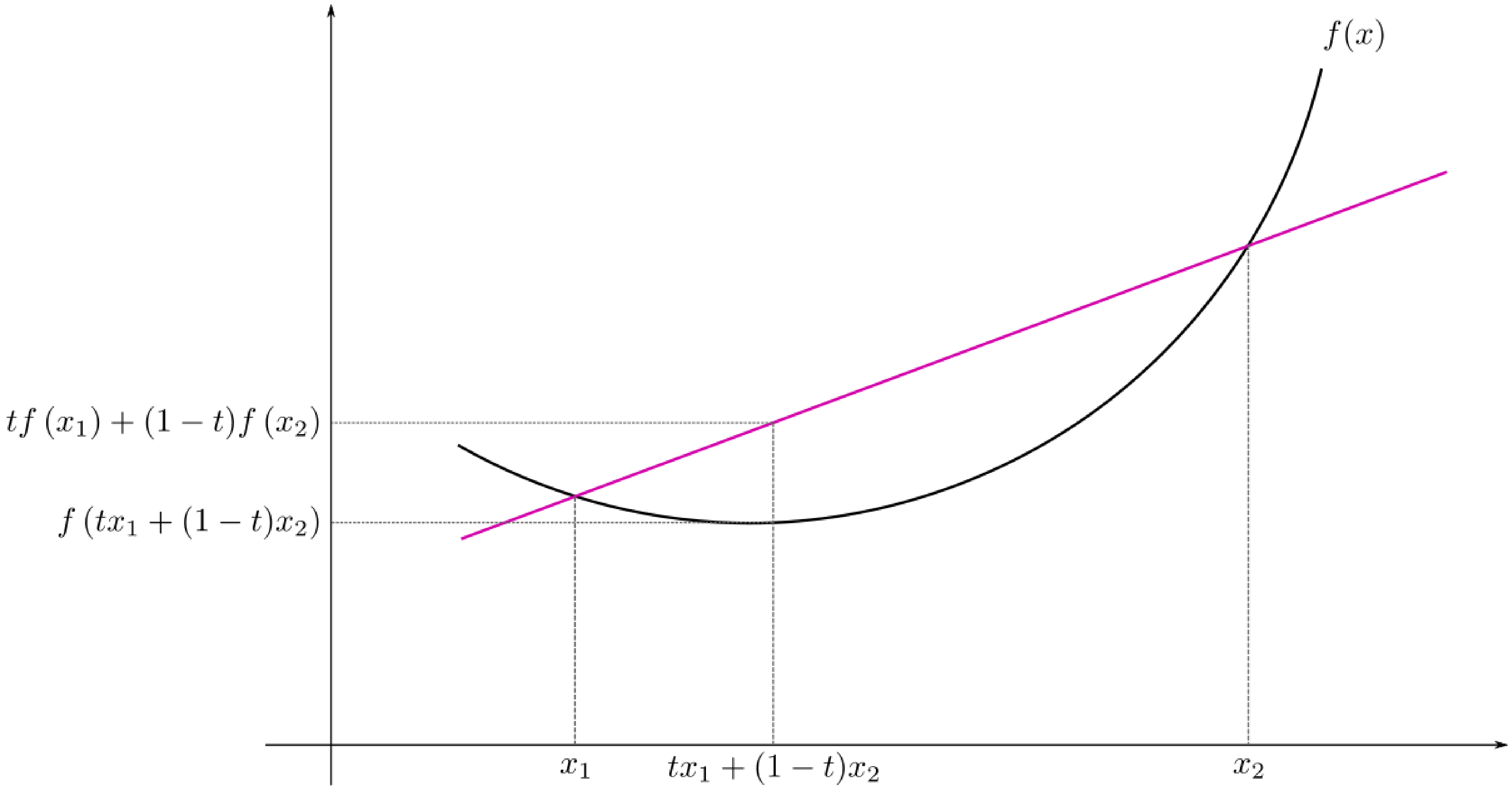
$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum p(x) H(Y|X=x) \\
&= H(Y) - H(p) \\
&\leq 1 - H(p)
\end{aligned}
$$

Maximum information rate =
Channel capacity

Math Fun

- Jensen's inequality

정보량은 0 이하가 될 수 없다.



Jensen's Inequality

함수의 생김새로 구분 ① convex ② X convex 오목 vs 볼록

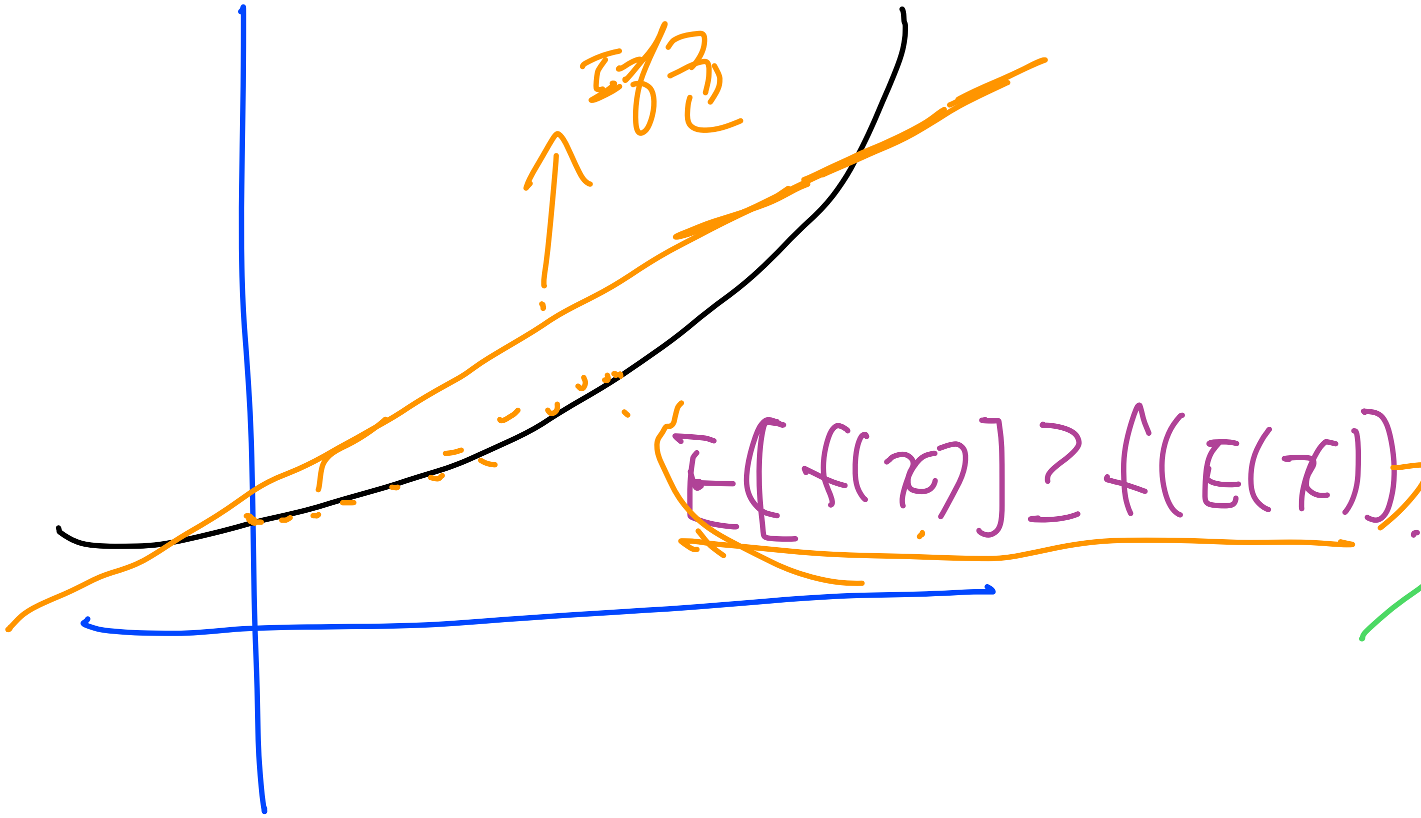
- If f is a convex function and X is a random variable, then $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$

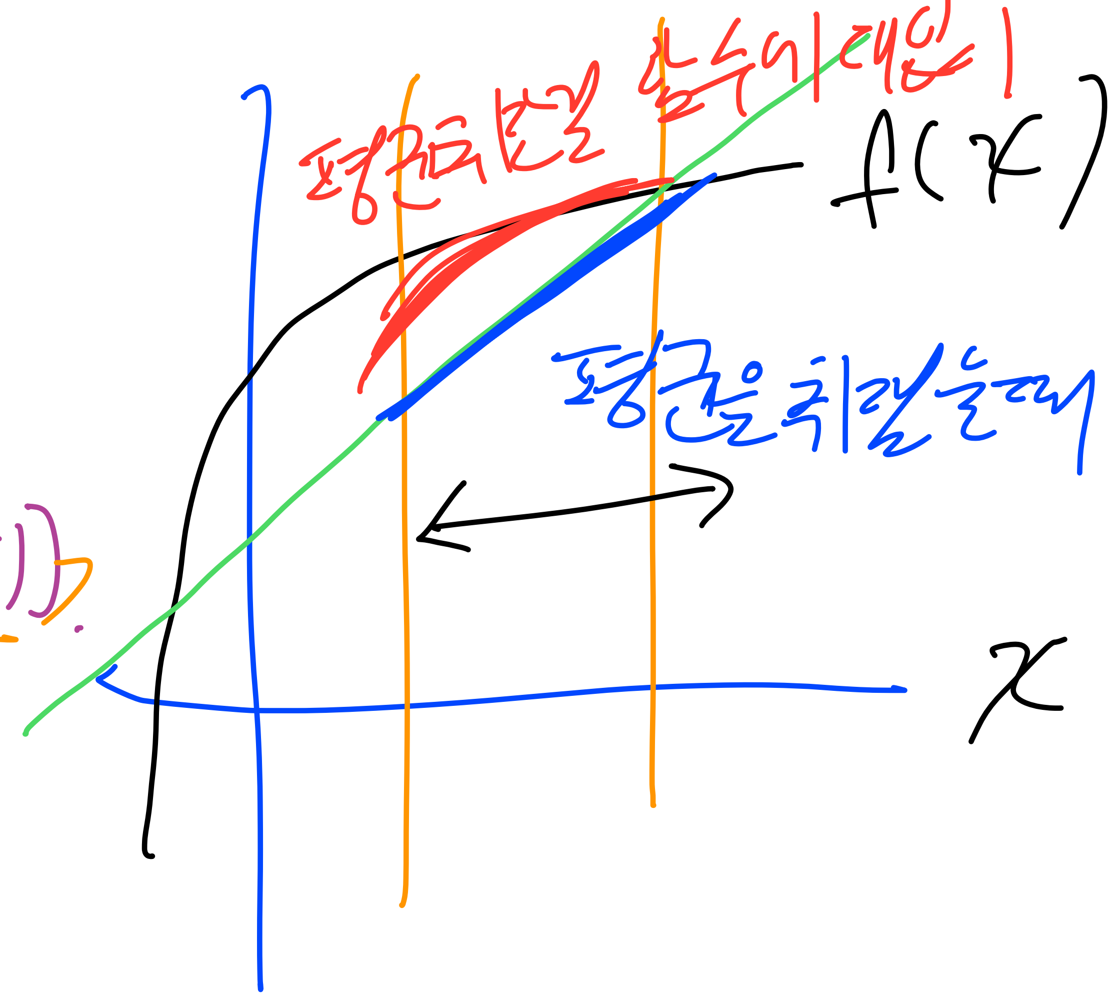
- Then we naturally have $I(X;Y) \geq 0$

Information amount is non-negative

- This leads to $H(X|Y) \leq H(X)$

Information cannot hurt





$f(x)$ 를 평균취한것 $E[f(x)]$

$\leq f(E(x))$ 평균취한것을 사용한것 출중에

오가크다

Proof

- The mutual information can be written as

$$I(X;Y) = D(p(x,y)||p(x)p(y)) = \mathbb{E}[\log_2 \frac{p(X,Y)}{p(X)p(Y)}]$$

log2 log2

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

이므로

$E \log_2 p(x) \geq \log_2 E p(x)$

- This also can be

$-\sum_x p(x)$

$$\begin{aligned} -D(p||q) &= -\sum_x \log_2 \frac{p(x)}{q(x)} = \mathbb{E}[\log_2 \frac{q(x)}{p(x)}] \\ &\leq \log_2 \mathbb{E}[\frac{q(x)}{p(x)}] \\ &= \log_2 \sum p(x) \frac{q(x)}{p(x)} \\ &= \log_2 1 \end{aligned} \quad (1)$$

jeonghun.park@knu.ac.kr

Gambling and Information Theory

- The world's best gambler is playing the horse race

100만원 (b) 가 있다하고.

$b_i \longrightarrow$ **A bet for the i-th horse**

1번마 확률 30
2번마 확률 20
3번마 확률 20
⋮
이런식일때

나는 1등하는데 얼마씩 하는게 좋을까?

이기는 말은 하나씩

$$\sum b_i = 1$$

If the horse i wins, the gambler win $\times o_i$

The gambler can reinvest the money that earned in the previous stages

Since the gambler can reinvest his money, his wealth is the product of the gains for each race

$$S_n = \prod_{i=1}^{n} S(X_i)$$

* $b = [b_1, b_2, b_3, \cdots, b_k]$

k개의 말이 있다하자.

<u>1000만원 가진사람</u>

b_1 은 1번말에 투자하는 금액 (O_1)

b_2 는 2번말에 투자하는 금액 (O_2)

⋮

b_k는 k번째말 (배율은 O_k 라 할때)

이중에 나머지 다 지고 b_k 번째말이 이긴다면?

(1000만원 $\times b_k \times O_k$) 를 얻게될

그 다음 라운드에서 이 돈으로 투자해서 b_1 번째 말이 이긴다면

(1000만원 $\times b_k \times 5k) \times b_1 \times O_1$

그 다음 라운드에서 b_2 번째 말이 이긴다면

((1000만원 $\times b_k \times 5k) \times b_1 \times O_1) \times b_2 \cdot O_2$

Gambling and Information Theory

투자량X배율 의 평균값을 크게해야함

- The wealth relatives $S(X) = b(X)o(X)$ is the factor by which the gambler's wealth grows if horse X wins the race

로그를 한 이유는 로그의 더하기는 곱하기라서..?

- The doubling rate of a horse race is

이길확률

$$W(\mathbf{b}, \mathbf{p}) = \mathbb{E}[\log S(X)] = \underbrace{\sum_{k=1}^{K} p_k}_{E} \log \underbrace{b_k o_k}_{S(X)}$$

$$\frac{1}{n} \log S_n = \frac{1}{n} \sum_{i=1}^{n} \log S(X_i) \to \mathbb{E}[\log S(X)] \quad \therefore S_n = 2^{\boxed{nW(\mathbf{b},\mathbf{p})}}$$

jeonghun.park@knu.ac.kr

Gambling and Information Theory

$$\frac{b_i}{p_i} \cdot p_i o_i$$

- The optimum doubling rate

$$W^\star(\mathbf{p}) = \max_{\mathbf{b}} W(\mathbf{b},\mathbf{p}) = \max_{\sum b_i = 1} \sum_{i=1}^{K} p_i \log b_i o_i$$

- Solving the above problem, we have

이게 가능은 도박판을 가라

$$W^\star(\mathbf{p}) = \sum p_i \log o_i - H(\mathbf{p})$$

위험한 도박일수록 잃는게 커짐!

- The above is achievable by the proportional gambling scheme $\mathbf{b}^\star = \mathbf{p}$

optimal 할때는 이길확률이 너 돈만큼 투자하면됨 (도박판이 할땐)

* 배율과는 상관이 없다

Proof

$w(b,p) = E \log S(x)$, $S(x) = b_i O_i$

$E = \sum p_i$

$\Rightarrow \sum p_i \log b_i O_i$

$$W(\mathbf{b}, \mathbf{p}) = \sum p_i \log b_i o_i$$

$$= \sum p_i \log \frac{b_i}{p_i} p_i o_i$$

$$= \sum p_i \log o_i - H(\mathbf{p}) - D(\mathbf{p}||\mathbf{b})$$

$$\leq \sum p_i \log o_i - H(\mathbf{p})$$

$\sum p_i \log \frac{b_i}{p_i} + \sum p_i \log p_i + \sum p_i \log o_i$

$\Rightarrow D(b||q) - H(p) + \sum p_i \log o_i$

$\Rightarrow -D(q||b) - H(p) + \sum p_i \log o_i$

Another case

- If we allow that the gambler the option of retaining some of his wealth as case, the solution becomes different

Side Information

더블레이트 $W(X) = E \log S(X) = \sum p(x) \log b(x) o(x)$

- Consider two doubling rate:

$$W(X) = \max \sum p(x) \log b(x) o(x)$$

$$W(X|Y) = \max \sum p(x, y) \log b(x|y) o(x)$$

$$\Delta W = W(X|Y) - W(X)$$

The side information is helpful?

Side Information

- We have:

$$\Delta W = I(X;Y)$$

- Why?

$$W(X) = \sum p(x) \log b(x) o(x)$$

$$= \sum p(x) \log \frac{b(x)}{p(x)} \cdot p(x) o(x)$$

$$= -D(p \| b) - H(x) + \sum p(x) \log o(x)$$

$$W^*(X) = \sum p(x) \log o(x) - H(X)$$

$$W^*(X|Y) = \sum p(x,y) \log o(x) - H(X|Y)$$

$$\Delta W = W(X|Y) - W(X) = H(X) - H(X|Y)$$

$$= I(X;Y)$$

- When is the side information helpful?

Theorem 6.2.1 *The increase ΔW in doubling rate due to side information Y for a horse race X is*

$$\Delta W = I(X;Y). \tag{6.25}$$

Proof: With side information, the maximum value of $W^*(X|Y)$ with side information Y is achieved by conditionally proportional gambling [i.e., $b^*(x|y) = p(x|y)$]. Thus,

$$W^*(X|Y) = \max_{\mathbf{b}(x|y)} E\left[\log S\right] = \max_{\mathbf{b}(x|y)} \sum p(x,y) \log o(x) b(x|y) \tag{6.26}$$

$$= \sum p(x,y) \log o(x) p(x|y) \tag{6.27}$$

$$= \sum p(x) \log o(x) - H(X|Y). \tag{6.28}$$

Without side information, the optimal doubling rate is

$$W^*(X) = \sum p(x) \log o(x) - H(X). \tag{6.29}$$

Thus, the increase in doubling rate due to the presence of side information Y is

$$\Delta W = W^*(X|Y) - W^*(X) = H(X) - H(X|Y) = I(X;Y). \quad \square \tag{6.30}$$

Hence, the increase in doubling rate is equal to the mutual information between the side information and the horse race. Not surprisingly, independent side information does not increase the doubling rate.

This relationship can also be extended to the general stock market (Chapter 16). In this case, however, one can only show the inequality $\Delta W \leq I$, with equality if and only if the market is a horse race.

Example

- Consider the case of betting on the color of the next card in a deck of 26 red and 26 black
- Bets are placed on whether the next card will be red or black
- The game pays 2-for-1 (2 times payoff for win and nothing for lose)
- What will be the best betting schemes?

Two Different Approaches

$$S_{52}^{*} = \frac{2^{52}}{\binom{52}{26}} = 9.08.$$

- Sequential bet
- One-shot bet (Let each bet ride)

$$\binom{52}{26}$$

Consider their differences!