

"Information and Coding theory?"

MA41024

MA60020

→ undergraduate
→ post-graduate

38

25

Information

Coding

— a "lot of theorems".
— nothing to do with
computer programming.

→ "probability" & Statistics.

random variables — discrete

→ "linear algebra".

Matrix (Null space, Range space, Inner Product
I T P)

$$\begin{matrix} \bar{3} + \bar{1} + \bar{0} \\ = \quad = \quad = \\ \quad \quad \quad \chi \end{matrix}$$

T

Mon - 8-9:55 am
Tue - 12-12:55 pm

Fri - 5-5:55 pm

Rohit Sarma Sarker.

rohit.sarma.sarker95@iitkgp.ac.in

Books.

- ① Elements of Information Theory
— Thomas & Thomas
— John Wiley & Sons
- ② A First Course in Information Theory
— Yeung Raymond
— Springer

③

The Theory of Information & Coding — Robert McEliece
— Cambridge Univ. Press.

There will be 3 "class tests".

Viva

Bibhas Adhikari
①

Linear algebra — (V, IF)
finite field.



"IF" → $K, \dots, 0$

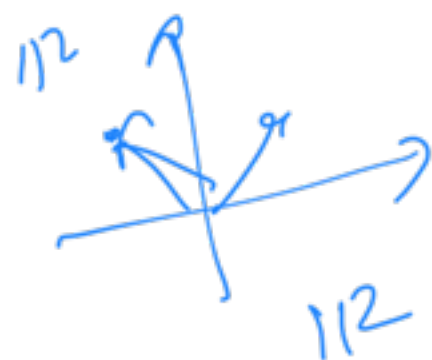
112

$$\mathbb{Z}_2 = \{0, 1\}$$

$$\mathbb{Z}_3 = \{0, 1, 2\}$$

$$a + (-a) = 0$$

$$a \times \frac{1}{a} = 1$$



$$IM: \textcircled{V} \rightarrow \textcircled{W}$$

$$\begin{bmatrix} x \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

$$\mathbb{Z}_2^n, \mathbb{Z}_2$$

$$\mathbb{Z}_2 \times \mathbb{Z}_2 \times \dots \times \mathbb{Z}_2$$

$\underbrace{\hspace{10em}}_n$

$$= \{ (x_1, \dots, x_n) \mid x_i \in \{0, 1\} \}$$

Introduction to Probability and
Statistics for Engineers and Scientists
c M. Ross.

'random'

$\{x_1, x_2, \dots, x_{10} \mid x_i \in \{H, T\}\}$

$\begin{matrix} H \rightarrow 1 \\ T \rightarrow 0 \end{matrix}$

$S \rightarrow$ Sample space

$X: S \rightarrow \mathbb{R}$

finite / countably infinite $\mathbb{N}$

"random variable"

universe



discrete
continuous "X"

$f(x)$

$S \rightarrow$ Sample space.

$X: S \rightarrow \mathbb{R}$

$P(A)$

$A \subseteq S$

$\square \square$

$$A = \{ \omega \in S \mid a \leq X(\omega) \leq b \}$$

$$= [a \leq X \leq b]$$

$\{ (3,4), (1,6), (2,5), (5,2), (6,1), (4,3) \}$

$$P(A) = \frac{6}{36}$$

$$f(x) = P(X=x)$$

pmf.

$$F(n) = P(X \leq n)$$

Latin — "informatio" — art of framing knowledge
Verb — "informare" — to explain
to conceive

— 'measure information'

1948, Claude E. Shannon

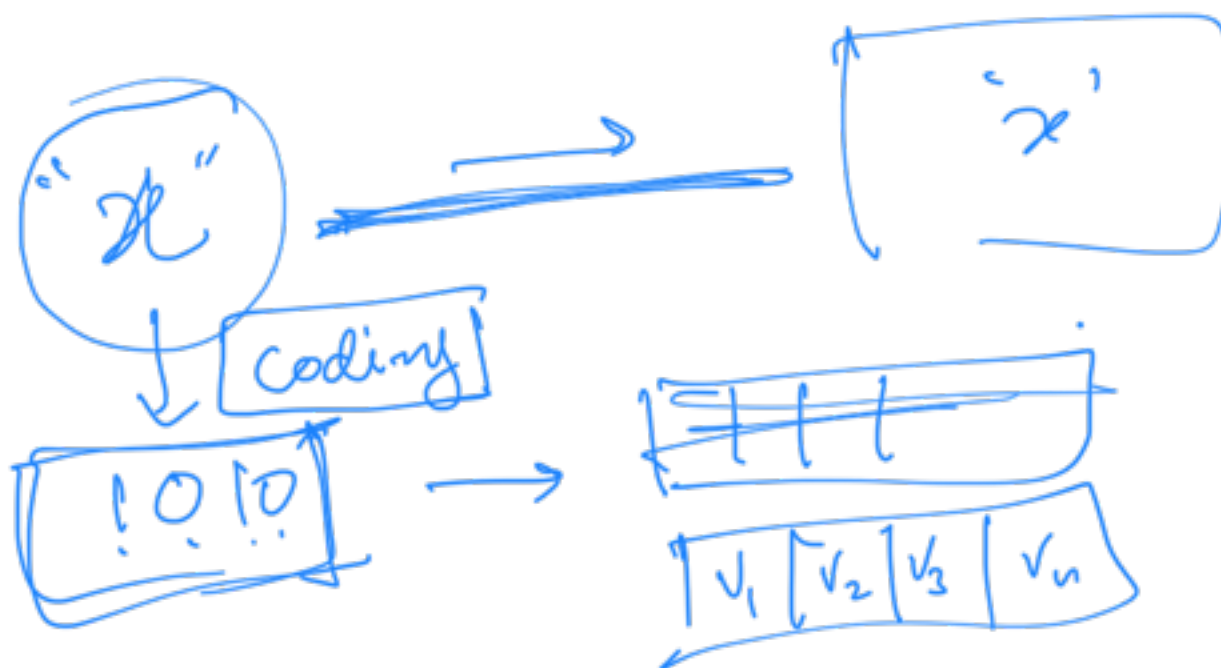
published a paper "The
mathematical theory of communication"

Concepts:

(1) "Information is uncertainty"
— "Information source is"

Information
a "random variable"
or "random process".

② Information has to be
~~process~~ transmitted as digital.





He proved two theorems.

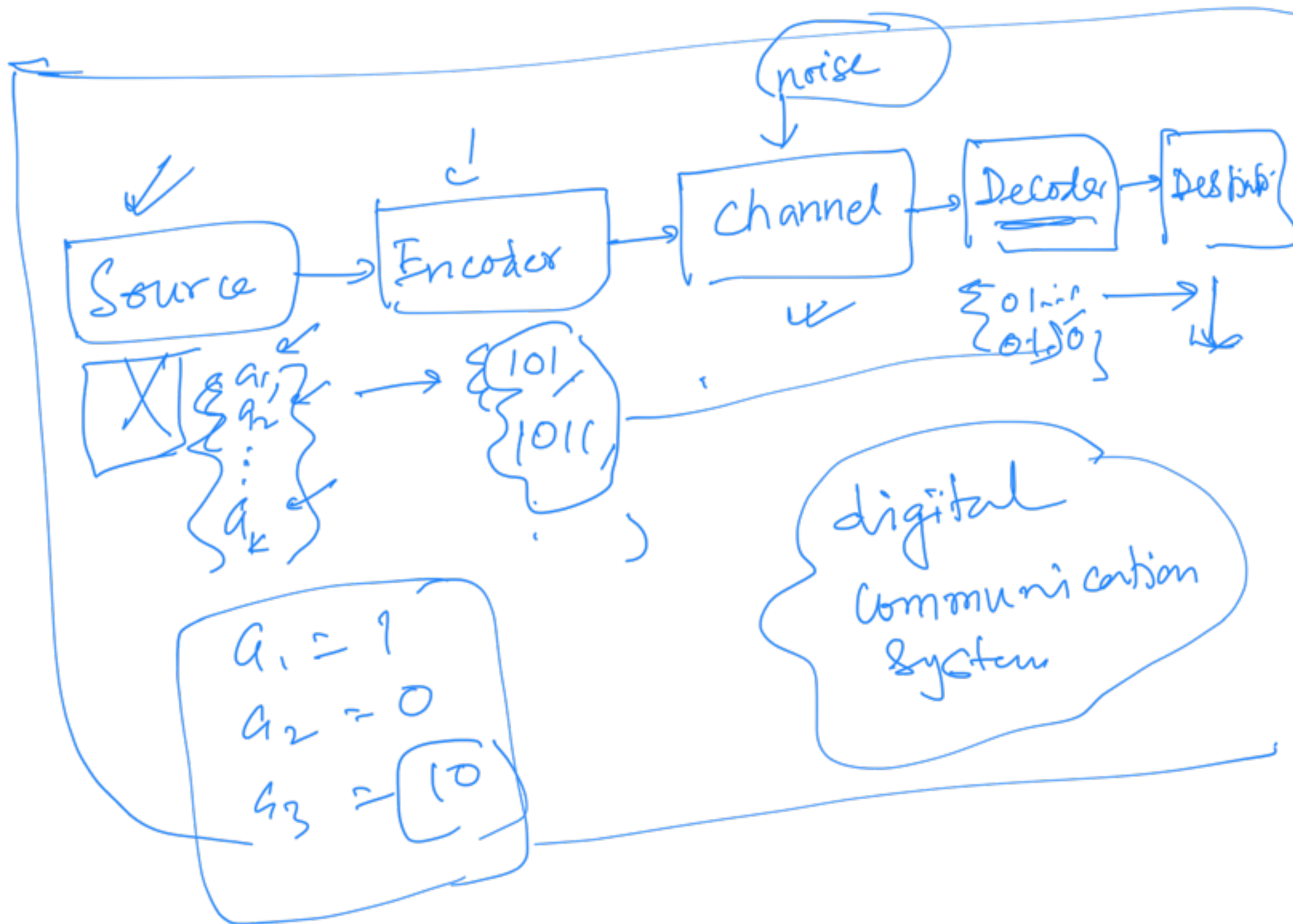
① Source Coding theorem
— 'entropy' as a fundamental measure of information.

② Channel Coding theorem
— he introduced the concept of capacity of channel.

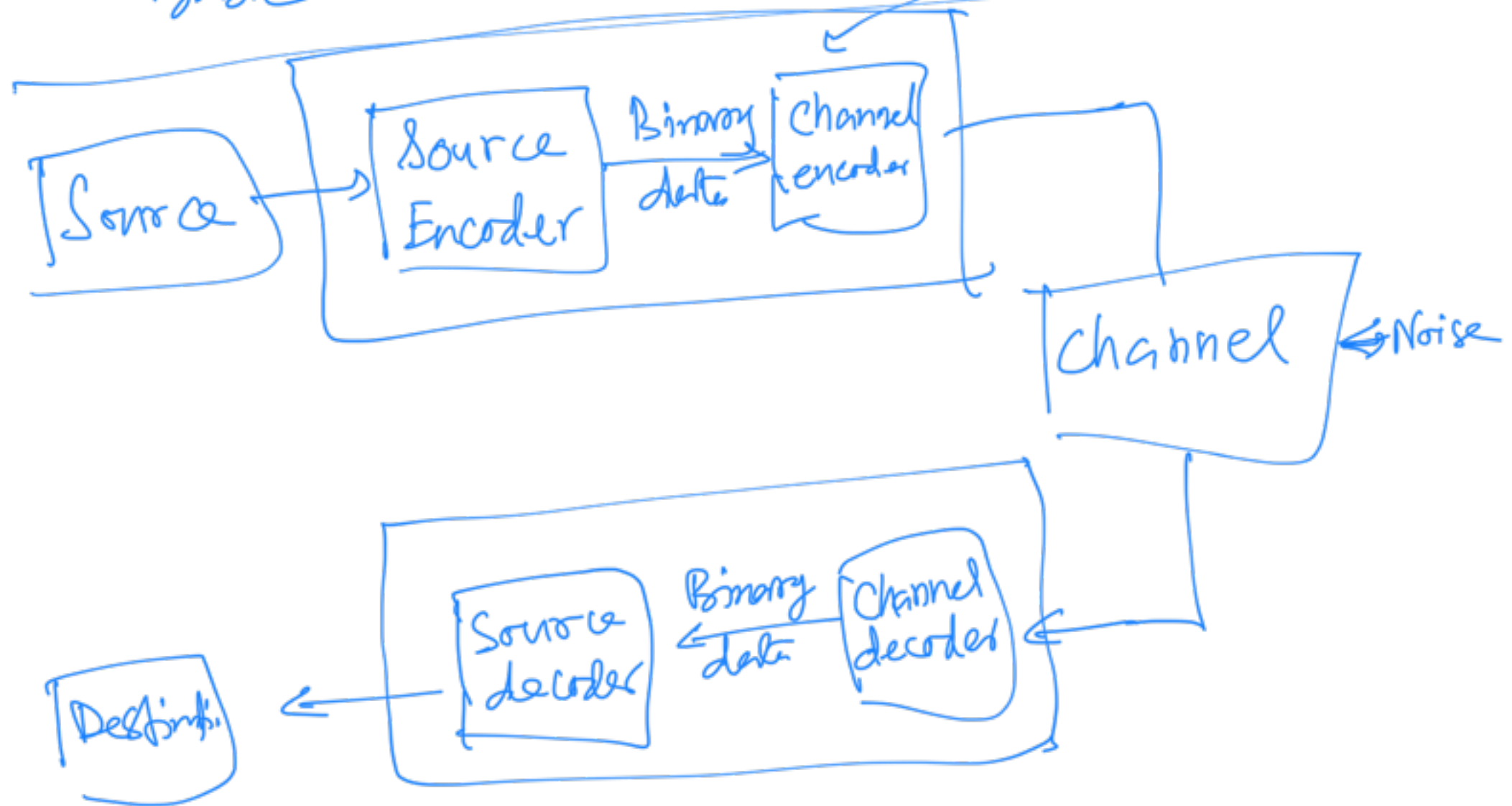


channel

of a crisis.



Basic models of source and channel encoders



$p(8) > p(3)$

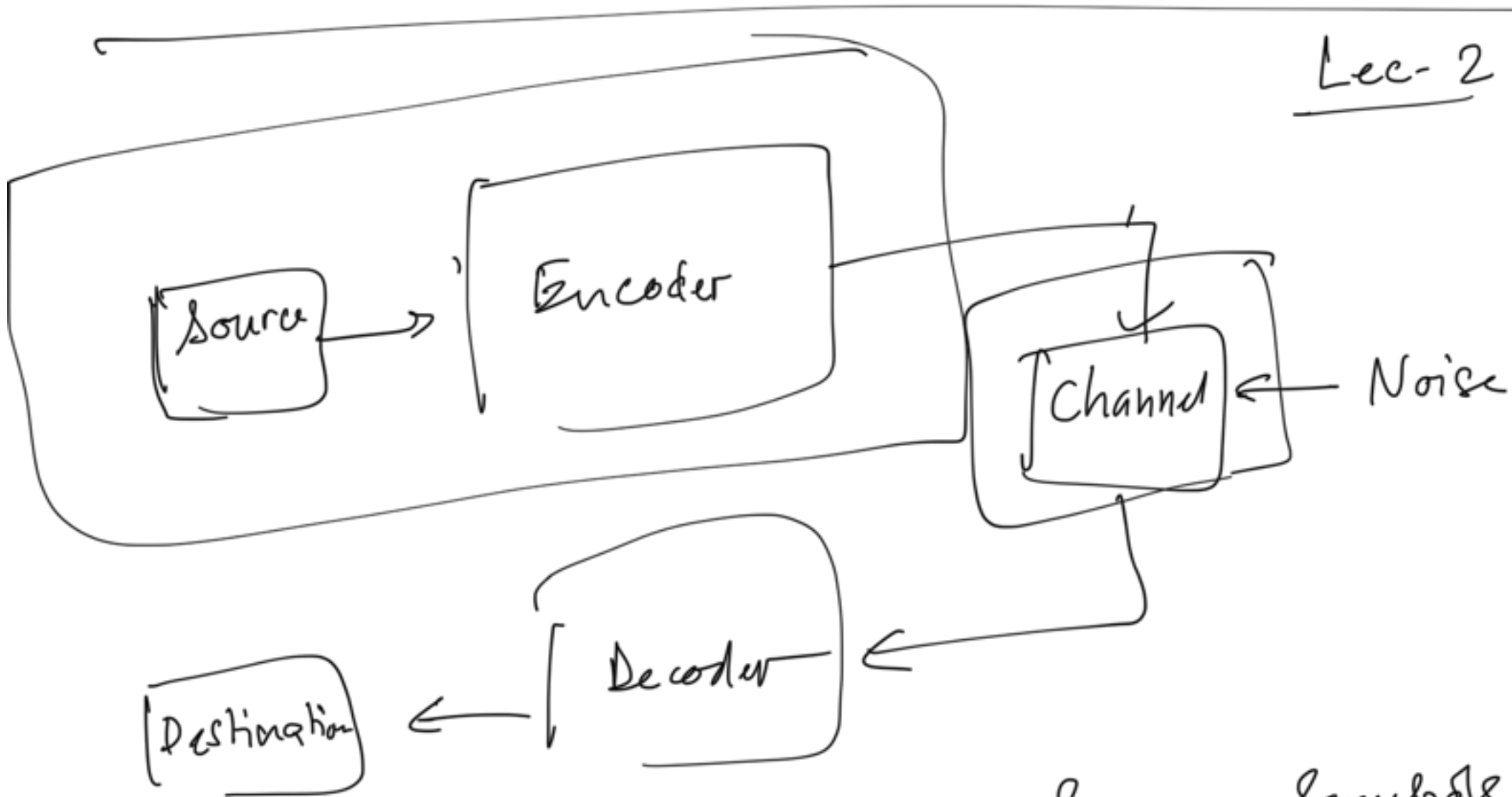
1101

ASCII

$$s = \underbrace{00110}_{\uparrow} \quad \delta = \underbrace{\quad\quad\quad}_{\text{out.}}$$

X

Lec-2

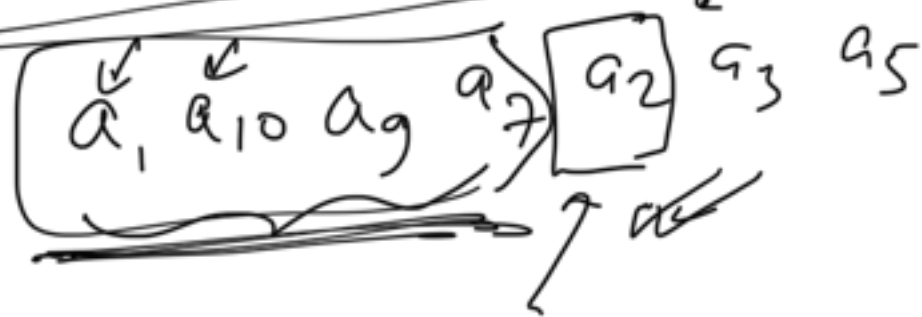


$\underbrace{\{a_1, a_2, \dots, a_k\}}_{\substack{\text{Alphabet}}} \rightarrow \text{Source Symbols}$

channel models

Source models & channel models

① Discrete memoryless source



ensemble $\left\{ a_1, a_2, \dots, a_K \right\}$
 $\left\{ p(a_1), p(a_2), \dots, p(a_K) \right\}$
 $\sum p(a_j) = 1$

X

Encoder

$a_1 = 00$
 $a_2 = 10$
 $a_3 = 01$
 $a_4 = 11$

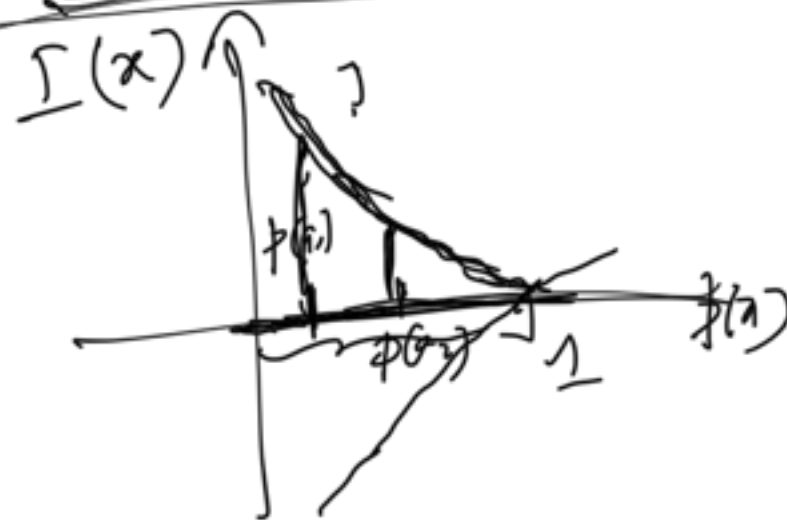
② Source with memory.

$\hat{a}_i / a_i$
Source output $\rightarrow$ Binary data

Self-information of a_i

$$I(a_i) = -\log_2 p(a_i) \text{ bits}$$

$$p(a_k) \rightarrow 1$$



$$\Rightarrow I(a_1) > I(a_2).$$

$$p(H) = .7$$

$$p(T) = .3$$

$$A = \{a_1, a_2, \dots, a_K\}$$

or expected self-information
 "average value of self-information"

$$\sum_{k=1}^K p(a_k) I(a_k) = \text{"entropy"}$$



$u_{12} / 1 \rightarrow$ Channel.

inputs for the channels

1

2

3

outputs for the channel.

transitional/conditional probabilities

$p(0) \quad 0$

$p(1) \quad 1$

$p(0|0)$

$p(1|0)$

$p(0|1)$

$p(1|1)$

0

1

$\Delta = \{0, 1\}$

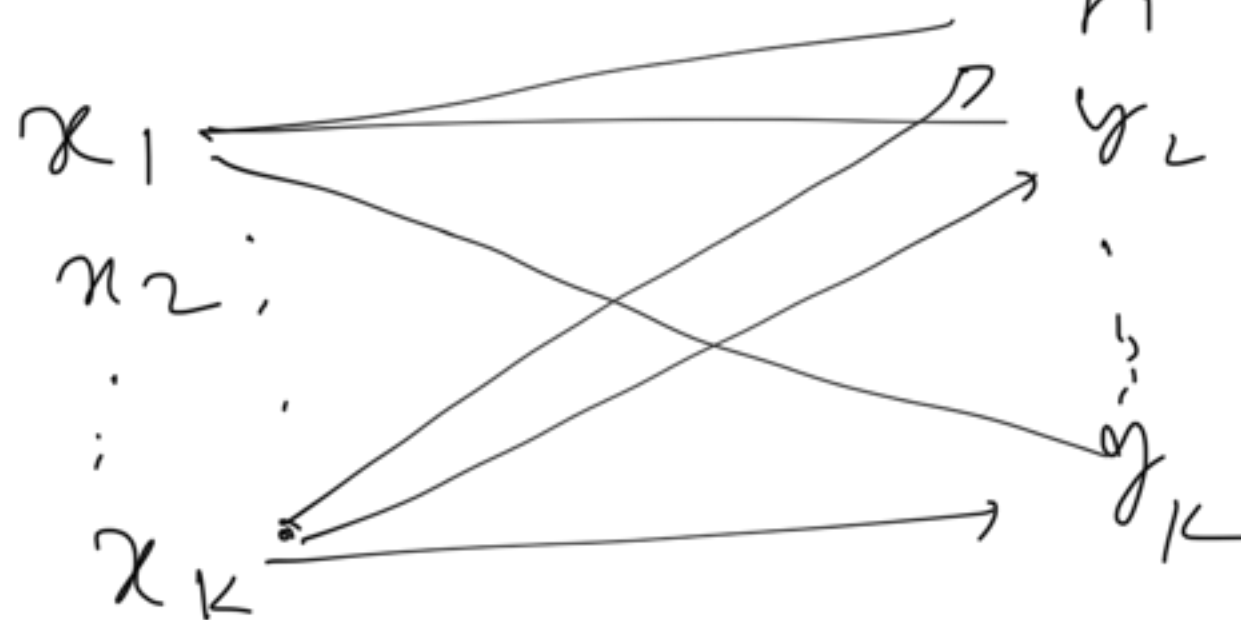
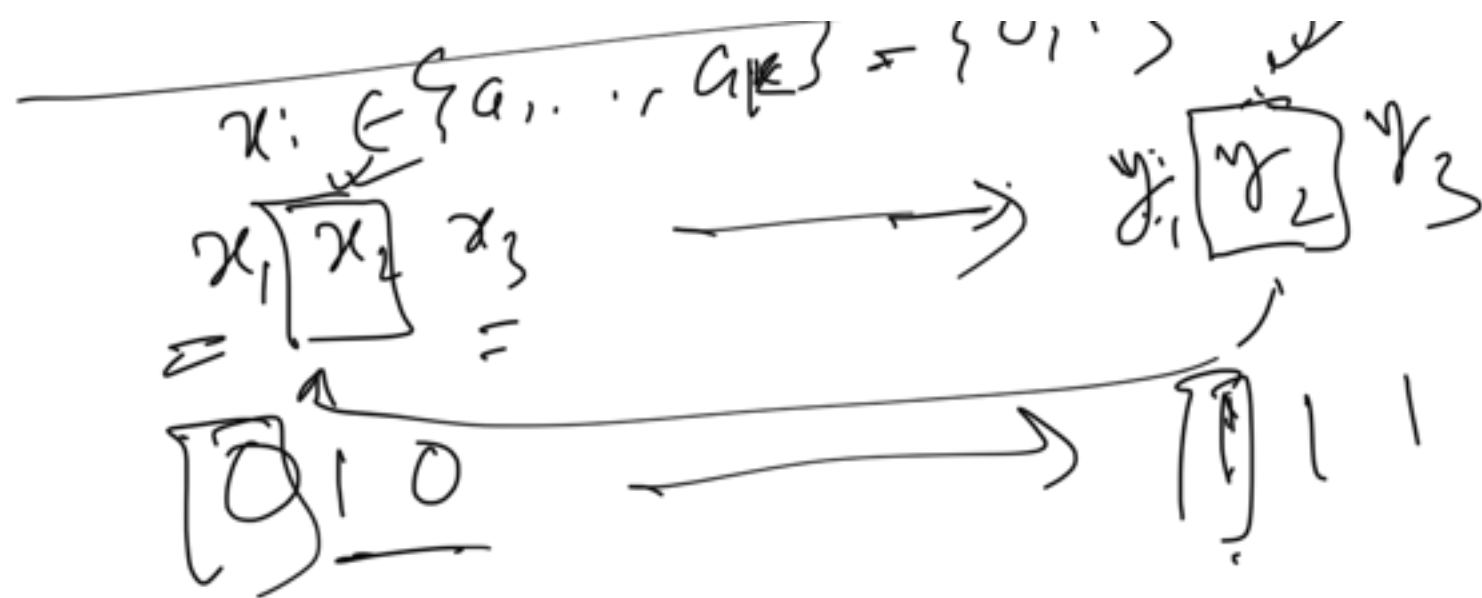
$$\begin{aligned} p(0|0) + p(1|0) &= 1 \\ p(0|1) + p(1|1) &= 1 \end{aligned}$$

$a_1 a_2 \dots a_k$

τ_{a_1}

$b_1 b_2 \dots b_k$

b_1
 b_2

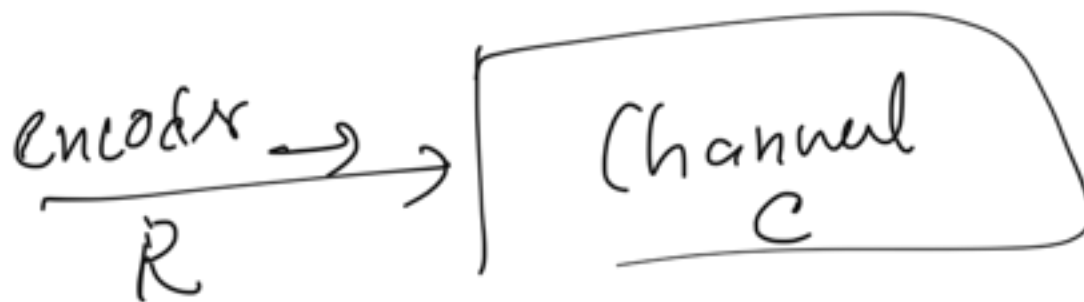


(i) memoryless channel in memory.

- channel with noise
- ② Gaussian noise channel.
 - ③

Capacity of a channel.

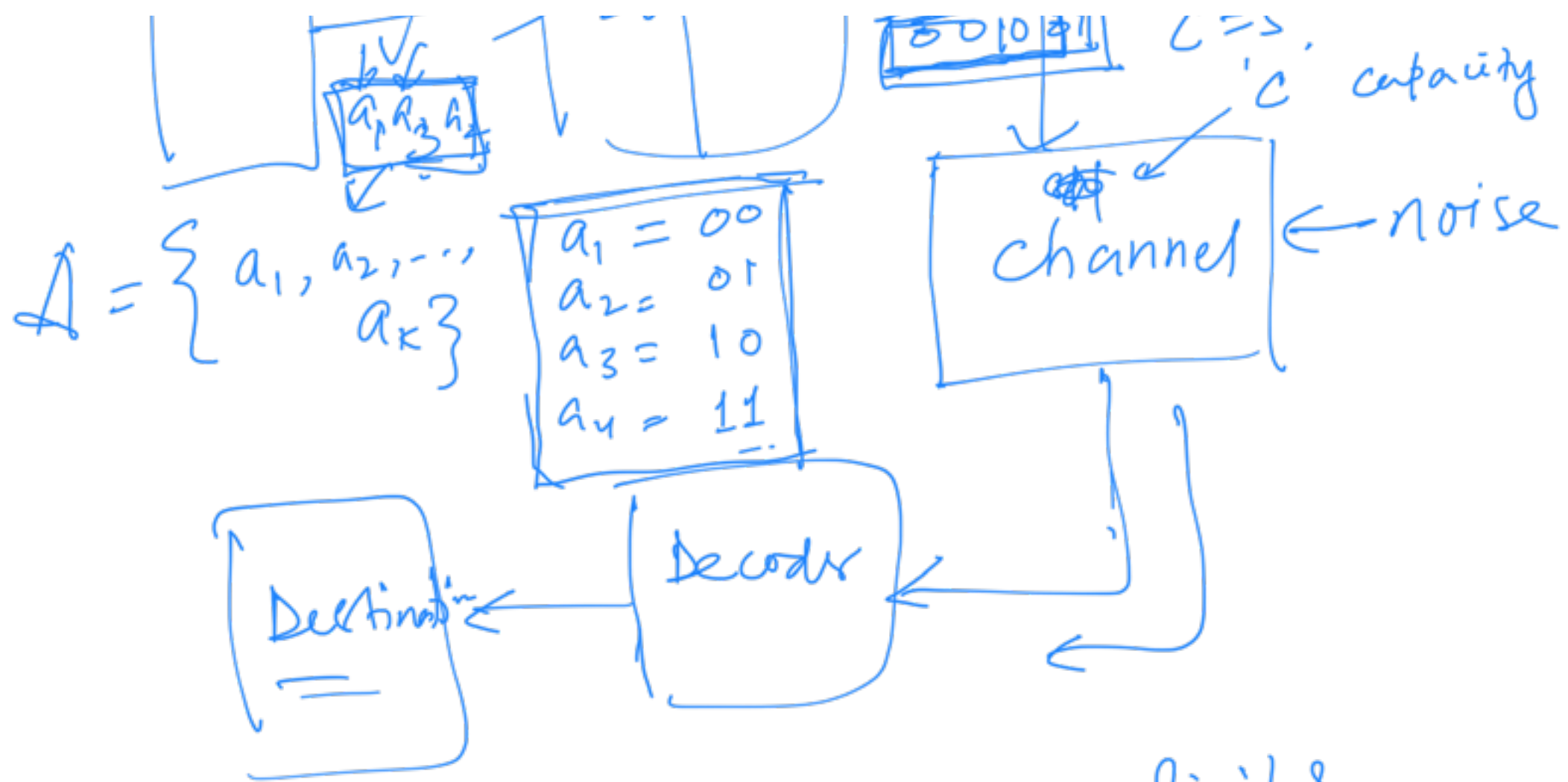
"C' bits per second"



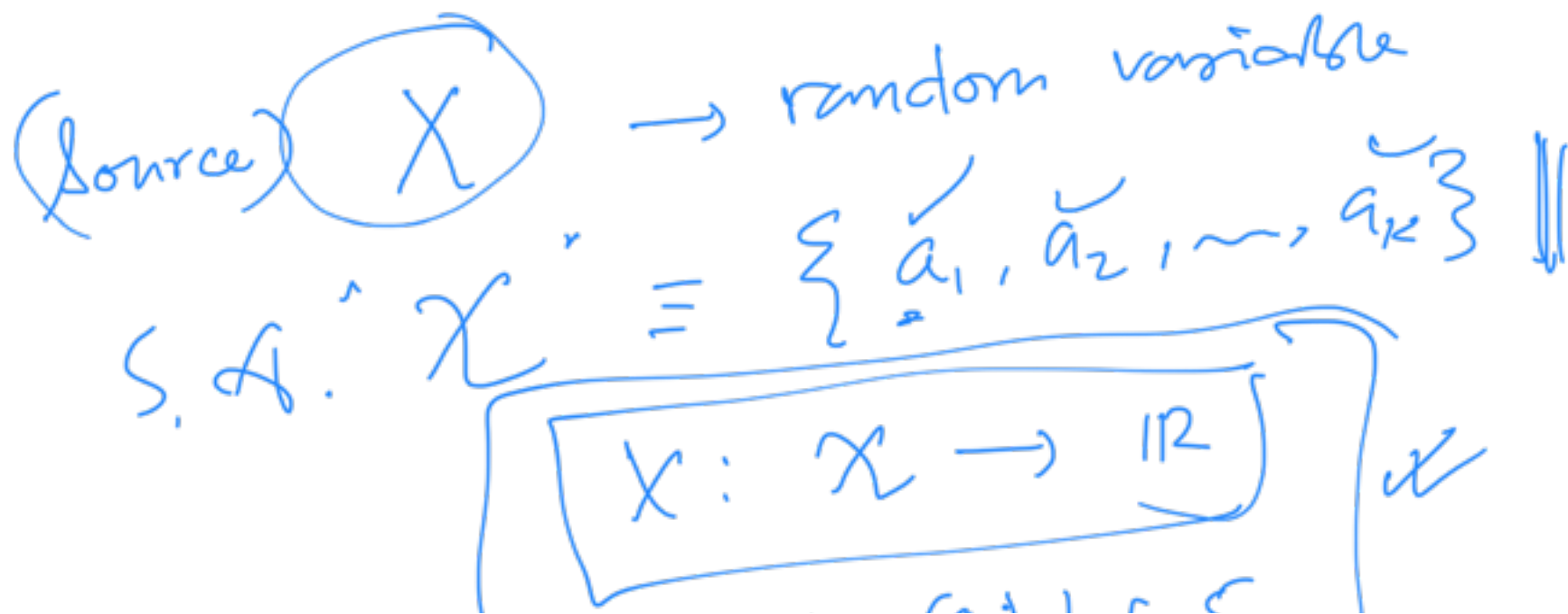
if $R < C$



⓪:
Lec-3



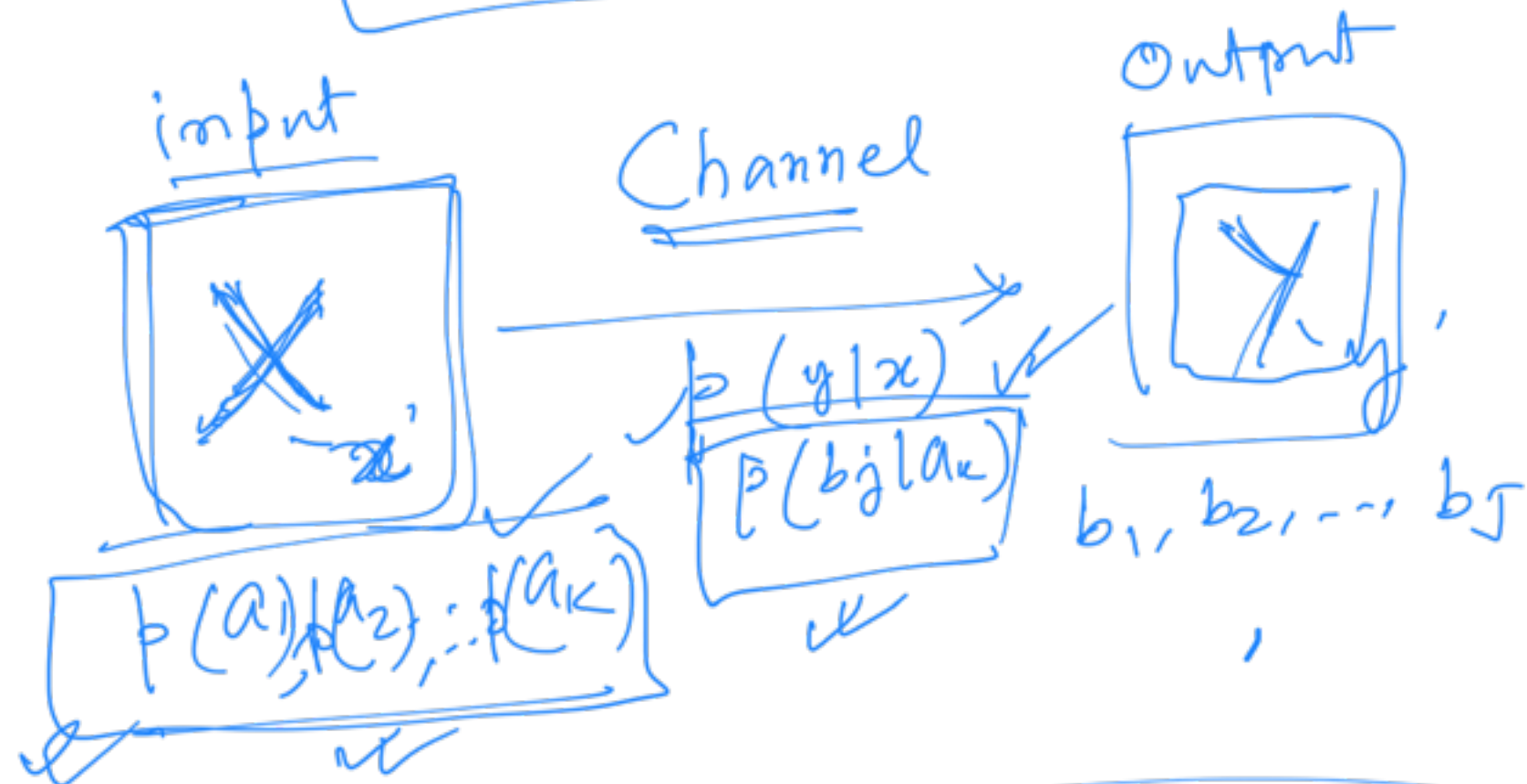
bit - binary digits



$$X (a_i \mapsto)$$

pmf.

$$p(a_i), 1 \leq i \leq K$$



~~12/1~~

$$(X, Y)$$

$$p(b_j)$$

$$\bar{X} = (X_1, X_2, \dots, X_n)$$

$$i \dots k$$

Sample Space = $\{ (a_k, b_j) \mid 1 \leq k \leq 10, 1 \leq j \leq 5 \}$

If pmf $(X, Y) \equiv P_{XY}(a_k, b_j)$

$= P(X = a_k, Y = b_j)$

is known then $P_X(a_k), P_Y(b_j)$

$$P(X) = \sum_y P(X, y)$$

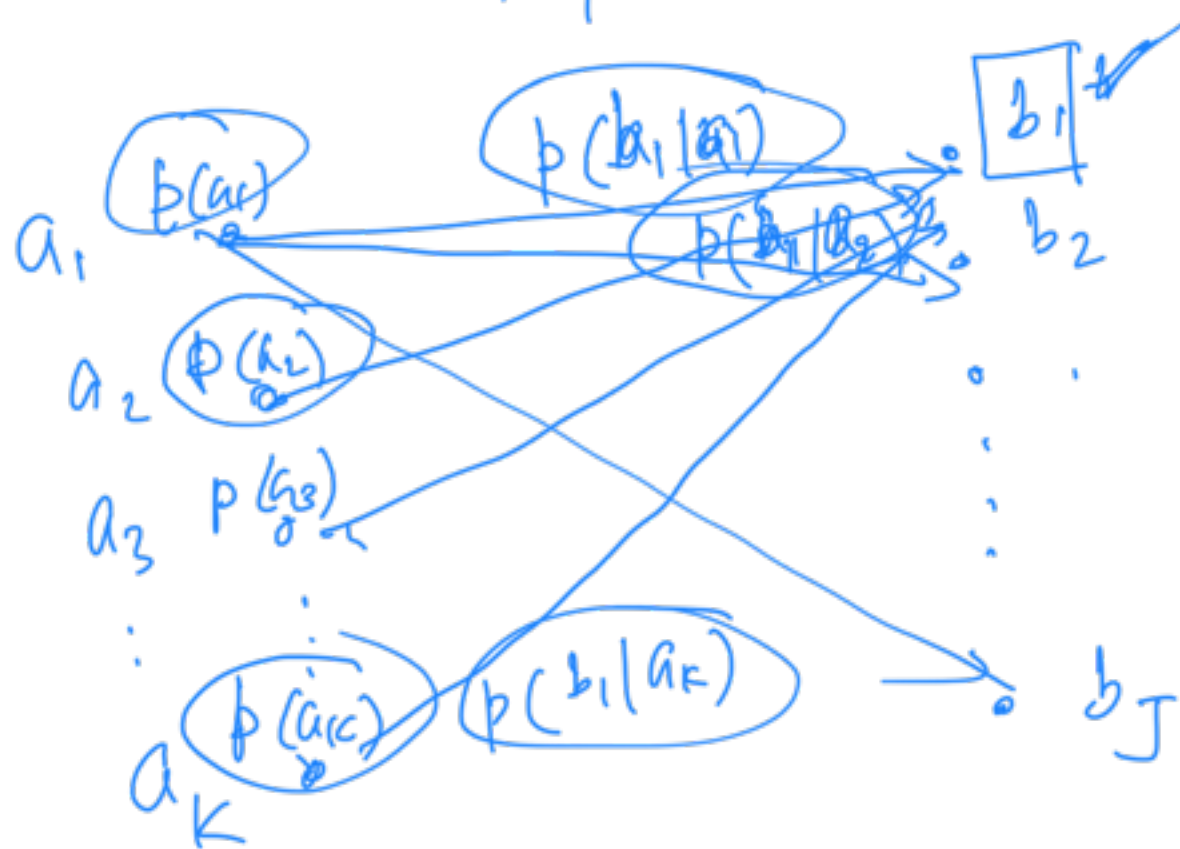
$$P(Y) = \sum_x P(x, y)$$

Recall: $P(b_j | a_k) = \frac{P_{XY}(a_k, b_j)}{P_X(a_k)}$

$$p(y|x) =$$

$$\frac{p(x, y)}{p_{\text{tot}}(x)}$$

$$p_{XY}(a_k, b_j) = p_{Y|X}(y|a_k) p_X(a_k)$$



$$p(b_1) = \sum_{k=1}^K p(a_k) p(b_1|a_k)$$

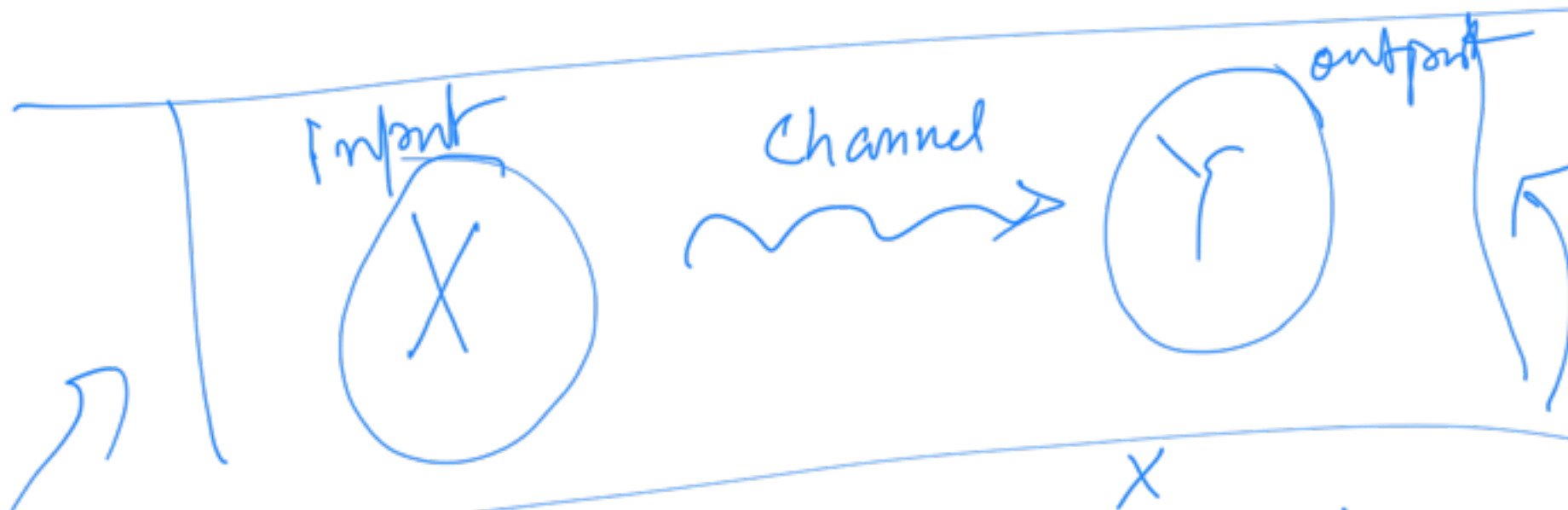
$$\sum_{k=1}^K p(a_k) p(b_1|a_k)$$

$$p(b_1) = \sum_{k=1}^K p(b_1 | a_k)$$

$$= \begin{bmatrix} p(b_1 | a_1) & p(b_1 | a_2) & \dots & p(b_1 | a_K) \end{bmatrix} \begin{bmatrix} p(a_1) \\ p(a_2) \\ \vdots \\ p(a_K) \end{bmatrix}$$

H.W.

$$= \begin{bmatrix} p(x) \\ p(y|x) \end{bmatrix}$$



Self-information: $p(x) = p(X=x)$

$$I(x) = -\log_2 p(x)$$

Discrete random variable



$$p(X=x)$$

$$x \in \mathcal{X}$$

sample.

entropy of $X \rightarrow$ a measure of uncertainty of X

$$H_b(X) = \sum_{x \in \mathcal{X}} p(x) I(x)$$

$$= - \sum_{x \in \mathcal{X}} p(x) \log_b p(x)$$

$$S = \{1, 2, \dots, n\}$$

$$= \sum p(x) \log \frac{1}{p(x)}$$

[X]

$$\begin{aligned} p(6) &= 0 \\ p(1) &= 1/5 \\ &\vdots \\ p(5) &= 1/5 \end{aligned}$$

$$= E\left(\log \frac{1}{p(x)}\right)$$

Convention: $0 \log 0 = 0$
Depending on the 'base' 'b'

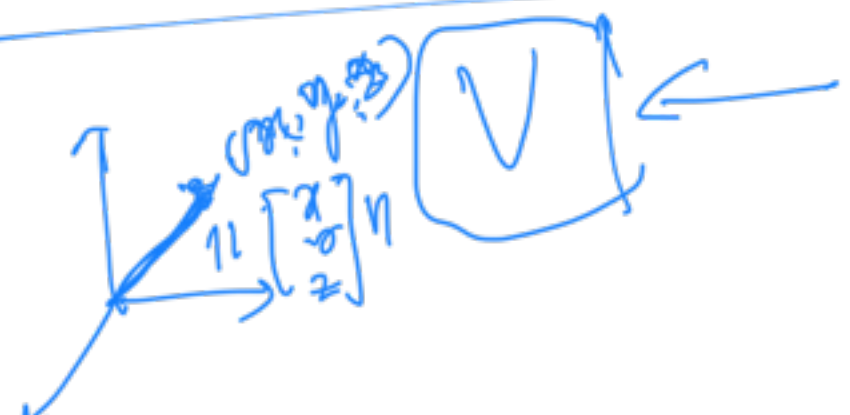
if $b=10$, $H_b(X)$ nat8.
 $b=2$, $H_b(X)$ bits.



X: $S \rightarrow \mathbb{R}$

$$E(g(X)) = \sum_{x \in X} g(x) p(x)$$

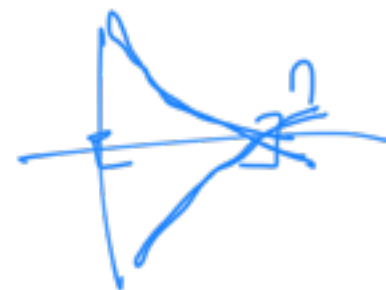
$$E(X) = \sum_{x \in X} x p(x)$$


 $\|\cdot\|: V \rightarrow [0, \infty)$

$$H_b(X) = - \sum_{x \in X} p(x) \log p(x)$$

Result:

① $H_b(X) \geq 0.$



$H_b(X) = H_b(Y)$

$$\textcircled{2} \quad H_b(X) = \lim_{n \rightarrow \infty} \frac{1}{n} H_b(X^n)$$

$$H_b : \text{[Diagram of a space } X \text{ with an 'X' inside]} \rightarrow [0, \infty)$$

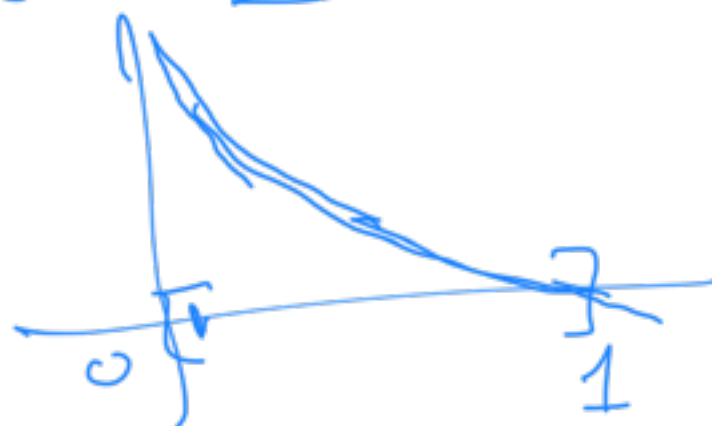
H.W.

$$a \in H_b(X) \leq \boxed{1?} ??$$

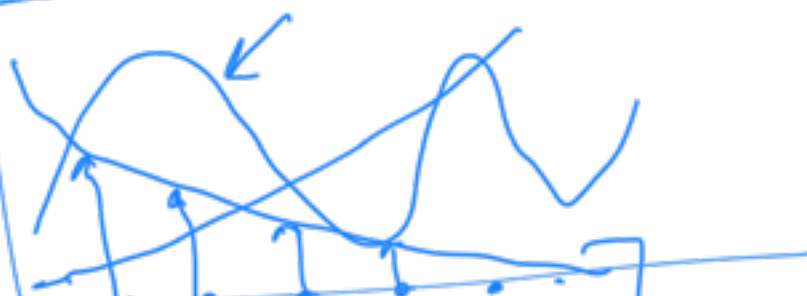
$$p(a) \rightarrow 0$$

$$-\log p(a) \rightarrow \infty$$

$$-\sum p(x) \log p(x)$$



$$H_b(X)$$



$$p(x) \leq 1$$

Exp.

Coin toss $S = \{H, T\}$

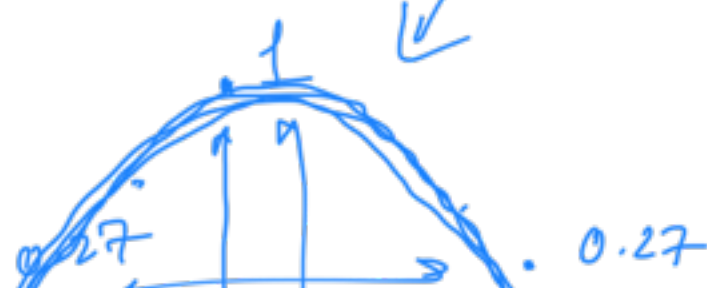
$$P(H) = p \quad P(T) = 1-p$$

$$X: S \rightarrow \mathbb{R}$$

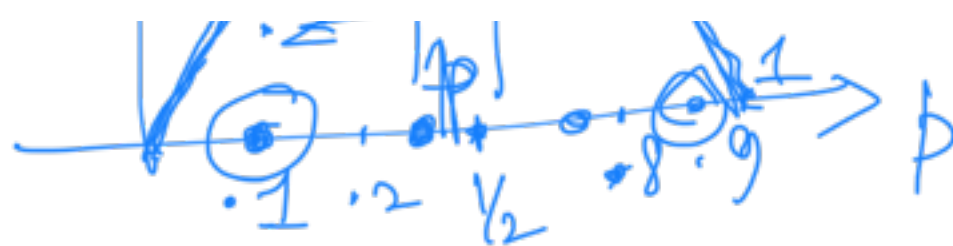
$$X(H) = 3 \quad X(T) = 4$$

$$\begin{aligned} H_2(X) &= -\sum p(x) \log_2(p(x)) \\ &= -p \log_2 p - (1-p) \log_2 (1-p) \\ &= f(p) \end{aligned}$$

$$p = 1/2 \quad f(p) = H(p)$$

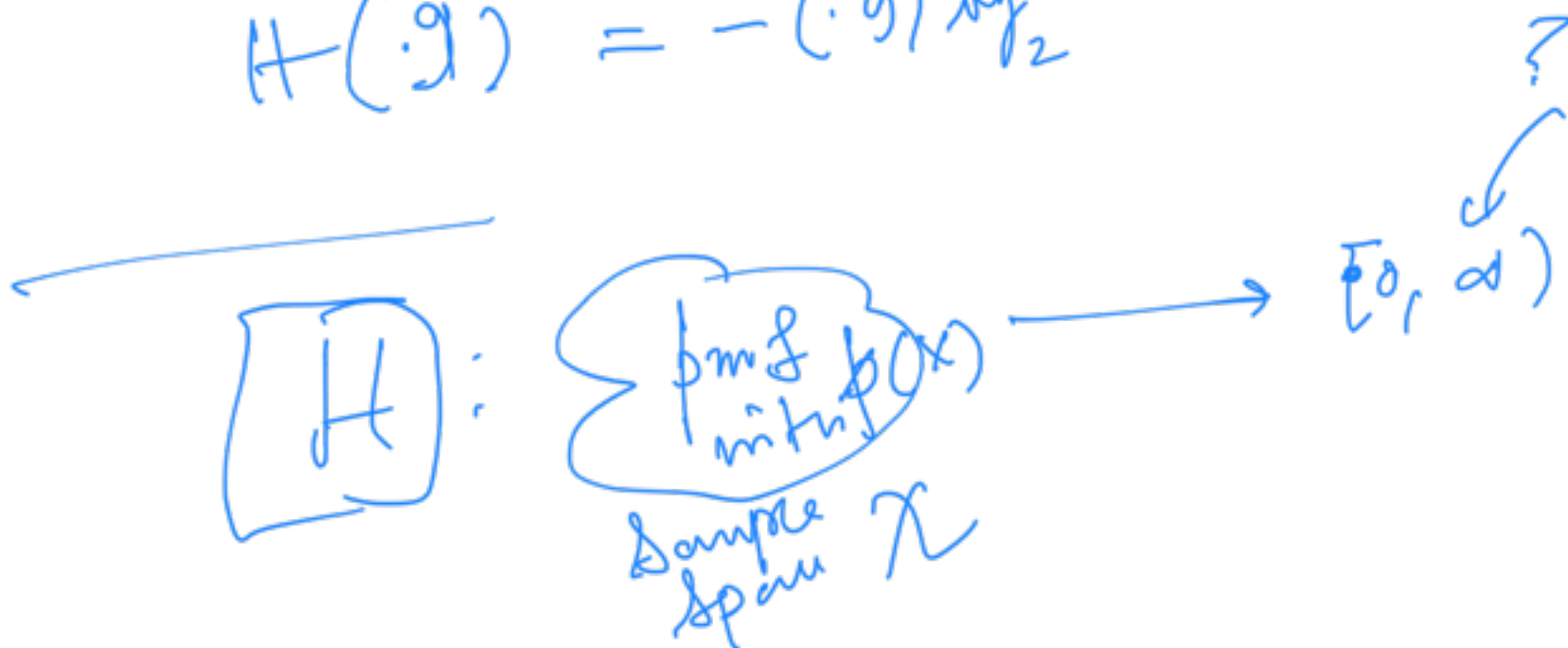


$$p = .9$$



$$H(.1) = -(.1) \log_2(.1) - (1-.1) \log_2(1-.1)$$

$$H(.9) = -(.9) \log_2(.9)$$



← 'Concave'

"X" $\Rightarrow$ "Y"
 $\sigma_L(X, T)$

$$p(x, y) = p_{xy}$$

Joint entropy: (X, Y)

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

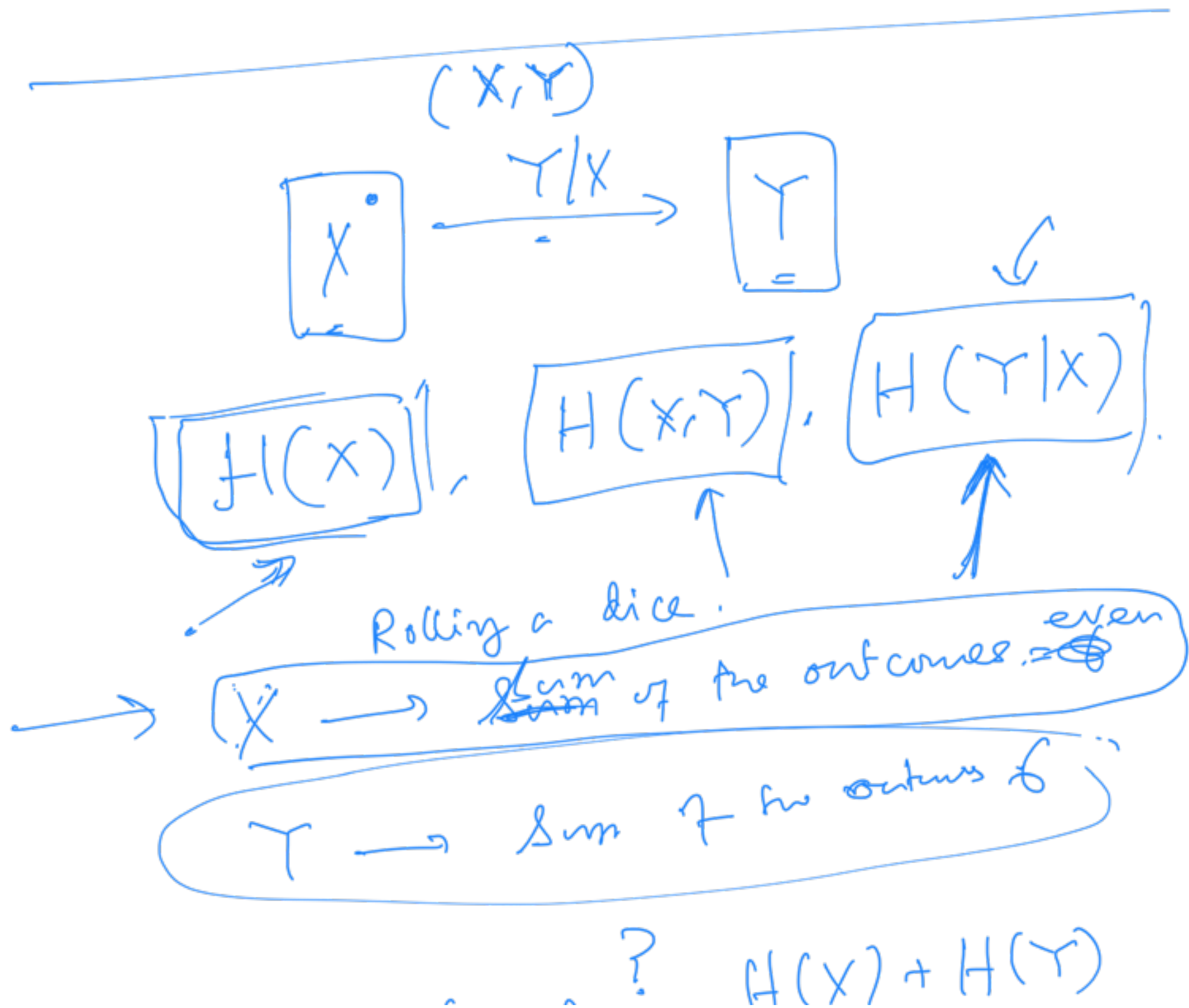
Conditional entropy: $(Y|X=x)$

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

$$= - \sum_{x \in X} p(x) \left(\sum_{y \in Y} p(y|x) \log_b p(y|x) \right)$$

$$\leftarrow \leftarrow \boxed{p(x, y)} \log_b p(y|x)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log P(x, y)$$



Q.

$$H(X, Y) =$$

H.W.

Thm. (Chain rule)

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

$$\text{pf. } H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y)$$

$$= - \sum_x \sum_y p(x, y) \log p(x, y)$$

$$= - \sum_x \left(\sum_y p(x, y) \right) \log p(x) - \sum_x \sum_y p(x, y) \log p(y|x)$$

$$\hookrightarrow b(x) \log p(x) - \sum_x \sum_y p(x, y) \log p(y|x)$$

$$= -\frac{1}{2} \log \frac{1}{2}$$

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

Thus,

$$\underline{H(X,Y) = H(Y) + H(X|Y)}$$

Corollary.

(chain rule)

$$H(\underbrace{X,Y}_Z | Z)$$

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

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

$$\textcircled{1} \quad H(Y|X) \neq H(X|Y)$$

(H.W.)

~~is not true~~

iff

2

$$\phi m f (X) \rightarrow Y$$

$$H(X) \stackrel{?}{=} H(Y)$$

H.W.

X

$$\begin{cases} p(H) = p \\ p(T) = 1-p \end{cases}$$

$$\begin{cases} p(1) = \dots \\ p(2) = \dots \\ p(3) = \dots \\ p(4) = \dots \\ p(5) = \dots \\ p(6) = \dots \end{cases} Y$$

' $\phi m f s$ '

$$X \rightsquigarrow Y$$

Relative entropy -
(Kullback - Leibler)
distance.

induces a notion of
distance b/w two
random variables.

Def.

let $p(x)$ & $q(x)$ be two pmfs

$$D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Then,

$D(\cdot)$

$\boxed{x \in X}$

$$0 \log \frac{0}{0} = 0, \quad 0 \log \frac{0}{2} = 0, \\ p \log \frac{p}{0} = \infty$$

- Q. ① $D(p \parallel q) \geq 0$ ✓
② $D(p \parallel q) = D(q \parallel p)$? X
③ triangle inequality. X

Mutual Information

Let X & Y be

two random variables.

$p(x, y)$

$p(x)$

$p(y)$

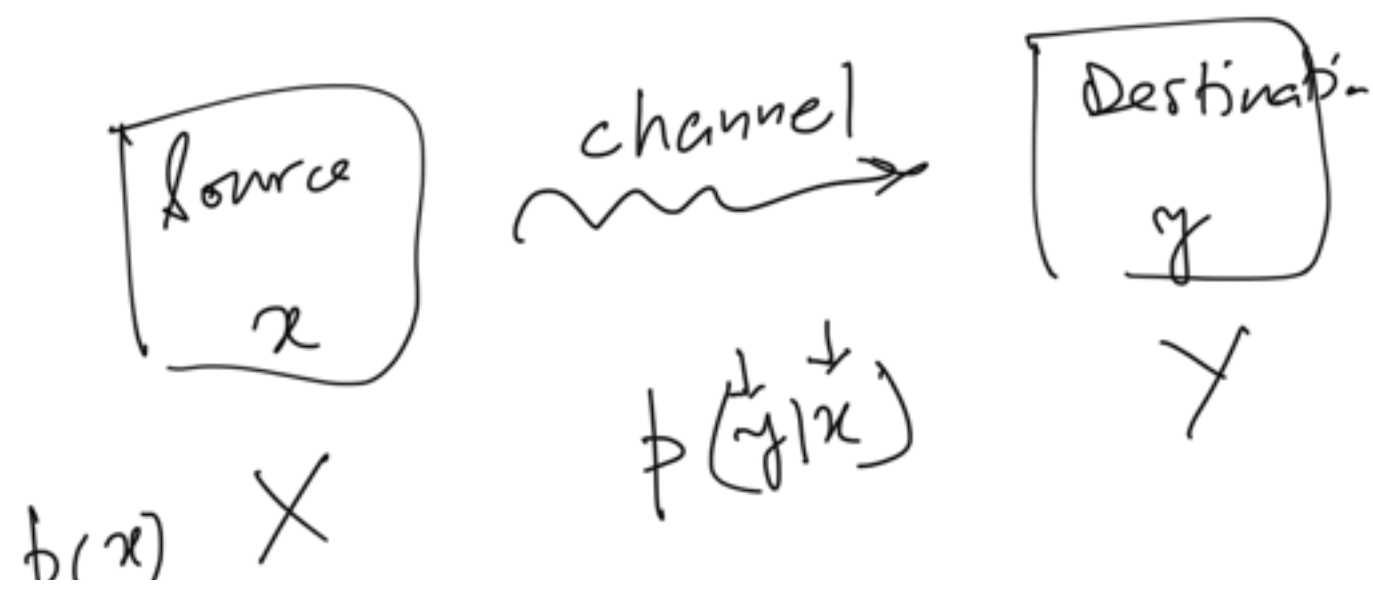
$$D(\underbrace{p(x, y)}_{f(x, y)} \parallel \underbrace{p(x)p(y)}_{g(x, y)})$$

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

$$= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

$$H(X) \quad H(X|Y) \quad H(X, Y)$$

Lec-4.



for,

$H(X) \rightarrow$ entropy of X

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

$\rightarrow$ Self-information of X

$H(X, Y)$

$H(Y|X)$

Chain rule: $H(X, Y) = H(X) + H(Y|X)$

Q. Intuition?

KL-distance:

$$D(p \parallel q) = \sum_{x \in X} p(x) \left\{ \log \frac{p(x)}{q(x)} \right\}$$

? $D(p \parallel q) \neq D(q \parallel p)$

$$0 \leq \dots$$

"Mutual" Information:

$$I(X; Y) = D \left(p(x, y) \parallel p(x) p(y) \right)$$

joint pmf
product
of marginals.

$$p(x, y) = p(x) p(y)$$

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

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y) \end{aligned}$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

e

$$\stackrel{?}{=} H(X|Y) + H(Y|X)$$

$$\stackrel{?}{=} H(X) + H(Y) - H(X|Y) - H(Y|X)$$

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

$$\stackrel{?}{=} \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)}$$

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

$$\stackrel{?}{=} \left(\sum_{x,y} p(x,y) \log p(x|y) \right) - \left(\sum_{x,y} p(x,y) \log p(x) \right)$$

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

$$= - \sum_{x,y} \frac{1}{n \cdot y} \dots$$

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

$$\Rightarrow \underbrace{H(X)} - H(X|Y)$$

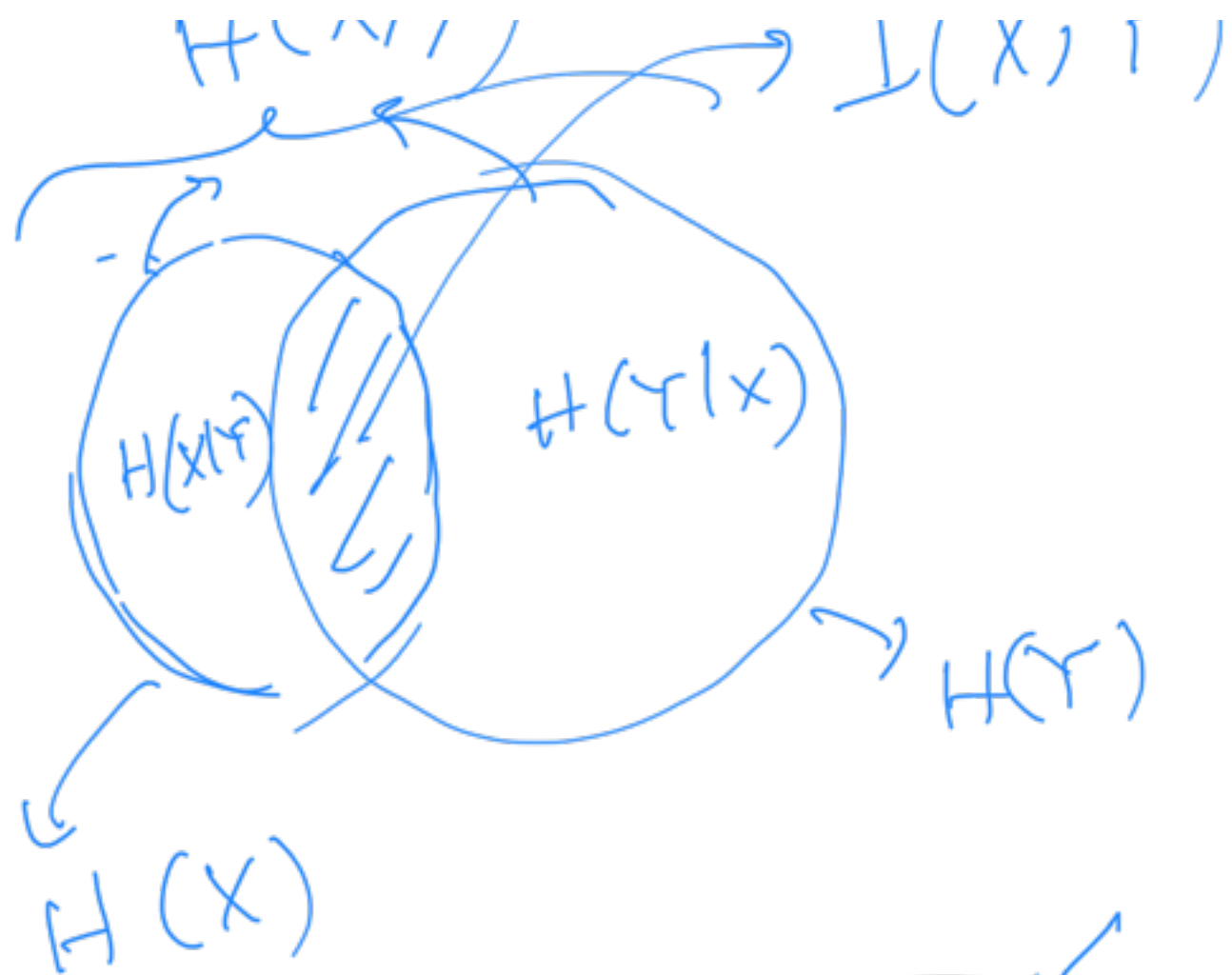
$$H(X, Y) = H(X) + H(Y|X) \\ = H(Y) + H(X|Y)$$

$$\Rightarrow H(X|Y) = H(X, Y) - H(Y)$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \\ = I(Y; X)$$

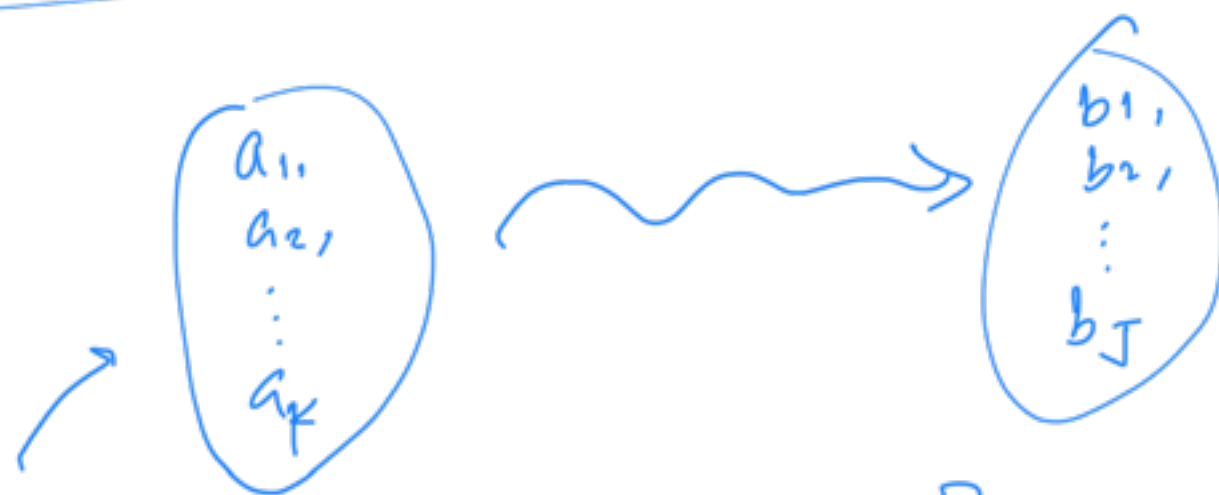
$$I(X; Y)$$

$$I(Y; X)$$



$$I(X; X) = H(X) - \underbrace{H(X|X)}$$

$$\underbrace{I(X; X)} = \underbrace{H(X)}$$



$$I(a_k | b_j)$$

? If I know that b_j is received what is the probability that some a_k was sent?

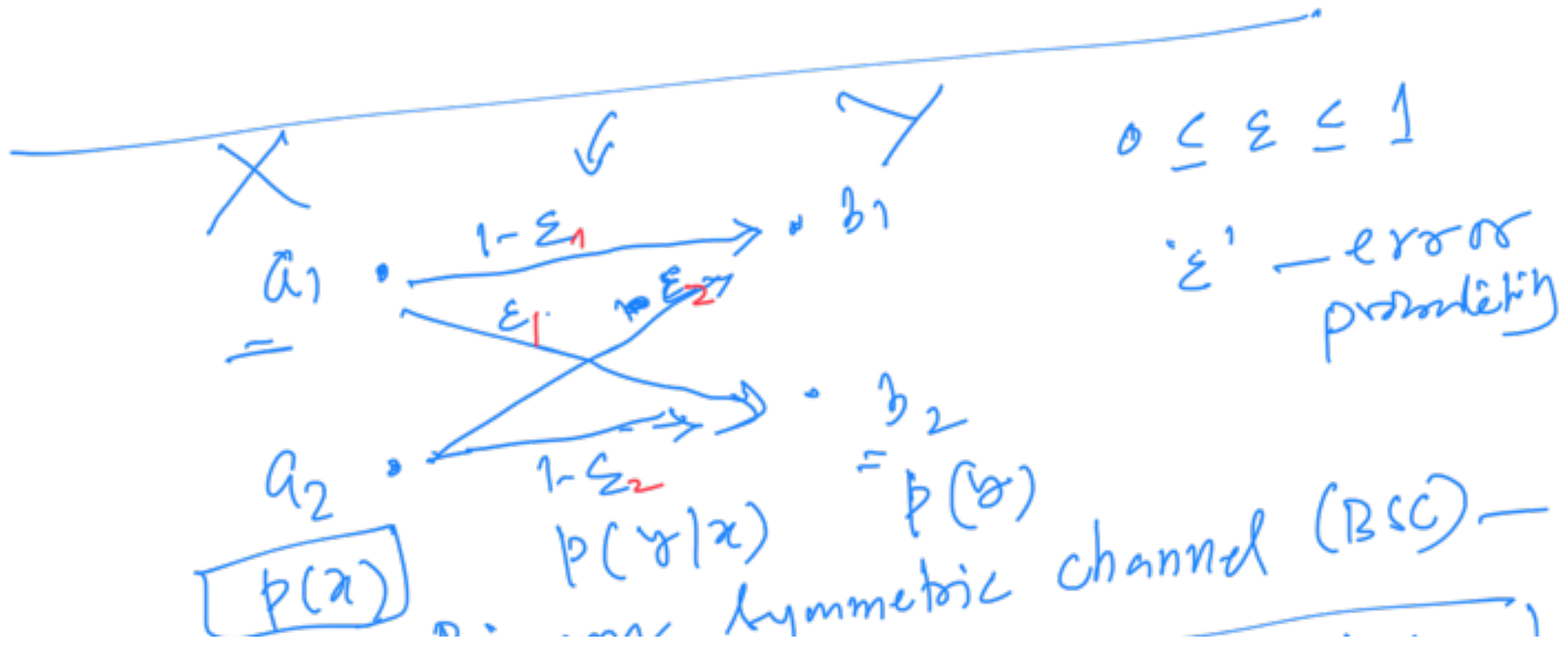
i.e. if ' b_j ' is known, how much I know about ' a_k '?

$$I(a_k | b_j) = \log \frac{p(a_k | b_j)}{p_x(a_k)} = \log \frac{p(a_k, b_j)}{p_x(a_k) p(b_j)}$$

the information about the event
 $X = a_k$ when $Y = b_j$ has
 occurred.

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

$$= \sum_n \sum_Y p(n, y) \log \frac{p(n, y)}{p(n)p(y)}$$



Binary

$$p(a_1) = \frac{1}{2} = p(a_2)$$

[Let us assume that

channel $\left[\begin{array}{l} p(b_1|a_1) = 1-\epsilon = p(b_2|a_2) \\ p(b_2|a_1) = \epsilon = p(b_1|a_2) \end{array} \right]$

$$p(a_1, b_1) = \frac{1-\epsilon}{2} = p(a_2, b_2)$$

$$p(a_1, b_2) = \frac{\epsilon}{2} = p(a_2, b_1)$$

$$I(a_1; b_1) = \log(2(1-\epsilon)) = I(a_2; b_2)$$

$$I(a_1; b_2) = \log(2\epsilon) = I(a_2, b_1)$$

Catch:

then

$$\varepsilon < \frac{1}{2}$$

$$I(a_1; b_2) < 0 !!!$$