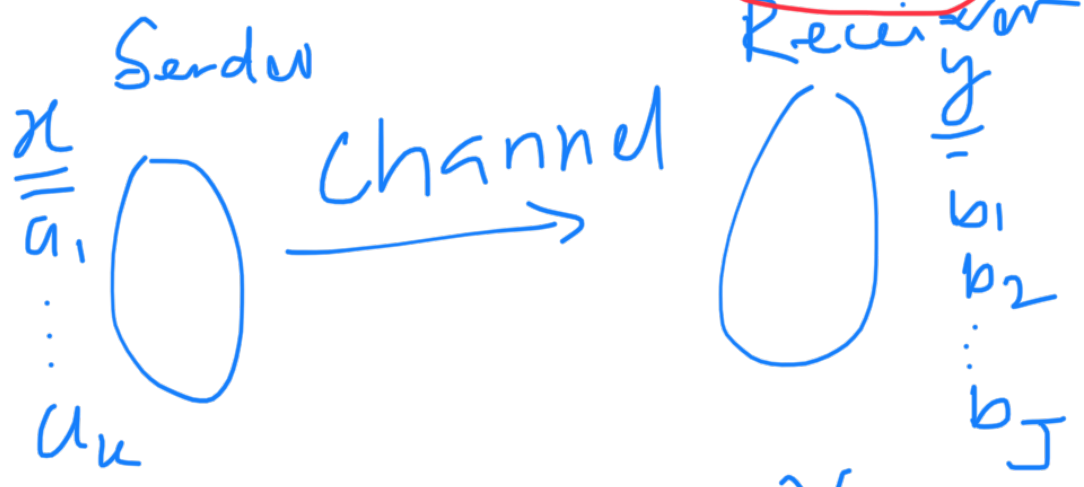
Information & Coding Theory - Lec 3

Mutual Information

$$I(x;y) = \log \frac{P(x|y)}{P(x)} \text{ bits}$$



$$I_{X;Y}(a_k;b_j) = \log \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)} \quad (1)$$

$$= I_{Y;X}(b_j;a_k)$$

Obs. $P_{X|Y}(a_k|b_j) = \frac{P_{XY}(a_k,b_j)}{P_Y(b_j)}$

$$\Rightarrow I_{X;Y}(a_k;b_j) = \log \frac{P_{XY}(a_k,b_j)}{P_X(a_k)P_Y(b_j)}$$

If $X = a_k$, $Y = b_j$ are independent events

then $I_{X;Y}(a_k, b_j) = 0$

Exp. Binary symmetric channel

Average mutual information

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

Self-information

$$I_X(a_k) = \log \frac{1}{p_X(a_k)}$$

(2) $$I(x) = \log \frac{1}{p(x)} = -\log p(x)$$ bits

Obs. This cannot be defined when X is a continuous random variable.

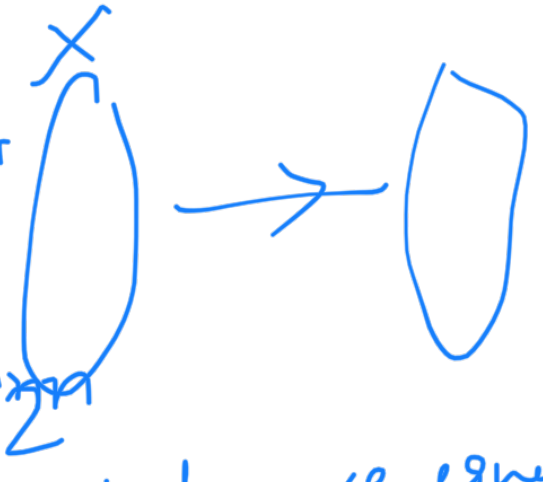
→ amount of surprise.

Exp. Let us consider the random variable X whose sample space is $\{0,1\}^n$

$(\quad) = a_1$

$\vdots$

$(\quad) = a_{2^n}$



Let all the sample points are equally $p(a_i) = \frac{1}{2^n}, \quad 1 \le i \le 2^n$

likely is. $p(x) = \frac{1}{2^n}$

$$I(x) = -\log p(x) = -\log 2^{-n}$$

$$= n \text{ bits}.$$

Remark: ↓ fits with the standard idea about bits.

Consider conditional self-information (because we are living in a joint sample space) of an event $x = a_k$ given the fact that $y = b_j$ has occured. as

$$I_{X|Y}(a_k|b_j) = \log \frac{1}{P_{X|Y}(a_k|b_j)}$$

$$I(x|y) = -\log p(x|y) \quad (3)$$

Then from the following definitions

$$\leftarrow I(x;y) = \log \frac{p(x|y)}{p(x)} \quad \text{w}$$

$$\leftarrow I(x) = \log \frac{1}{p(x)} \quad \text{w}$$

$$I(x|y) = \log \frac{1}{} \quad \text{w}$$

← I(a|o) — γ p(x|z)

Q. What is the observation from the above definitions?

$$I(x;y) = I(x) - I(x|y)$$

$I(x;y)$: amount of information about x when y has occurred

$I(x)$: self information about x

$I(x|y)$: amount of uncertainty about x given y

The entropy of X.

The average value of self information

$$H(X) = \sum_{x} p(x) I(x)$$

$$= \sum_{k=1}^{K} P_X(a_k) \log \frac{1}{P_X(a_k)}$$

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

[Later I will give an example about the entropy in terms of encoding the sample points $a_1, a_2, \ldots a_k$.

The entropy corresponding to conditional self-information on

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

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

↓

average information required to specify X when y is known.

Q. Can we relate $I(X;Y)$, $H(X|Y)$

$$\log \frac{p(x|y)}{p(x)} \qquad \sum p(x,y) \log \frac{1}{p(x|y)}$$

H.W. $I(X;Y) = H(X) - H(X|Y)$

↓

Q. How to interpret it? (VIVA)

Obs. $I(x,y) = -\log p(x,y)$

$= -\log P_{XY}(x,y)$

(Known) $I(x) = -\log p(x)$

$= -\log P_X(x)$

Relate ① $I(x,y)$, $I(x)$, $I(x|y)$

$I(y) + I(x|y) = I(x,y) = I(x) + I(y|x)$

② $I(x;y)$, $I(x)$, $I(y)$, $I(x|y)$

$I(x;y) = I(x) + I(y) - I(x,y)$

③ $H(XY) = H(X) + H(Y|X)$

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

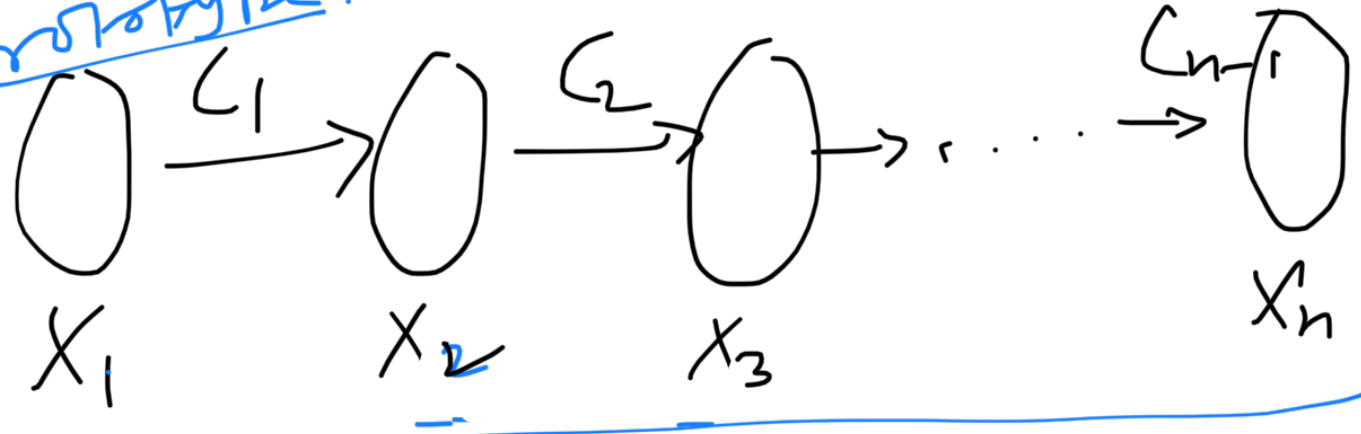
?

Hint: Take average both sides of the relation ① or expression.

④ $I(X;Y) = H(X) + H(Y) - H(XY)$

Hint: Take average values in the expression ②

Prototype.



$$(x_1, x_2, \ldots, x_n) \longrightarrow (X_1, X_2, \ldots, X_n) = \bar{X}$$

Sample space of $\bar{X}$

$$\{(x_1, x_2, \ldots, x_n) \mid x_i \in S_i\}$$

where S_i is the sample space of X_i.

Let (X, Y, Z) be a joint r.v.

$$I(x; y \mid z) = \log \frac{P(x \mid y, z)}{P(x \mid z)}$$

$$= I(x \mid z) - I(x \mid y, z)$$

H.W.

conditional mutual information

Taking the average value of the above expression:

$$I(X; Y \mid Z) = \sum_{x,y,z} P(x, y, z)\, I(x; y \mid z)$$

H.W.

$$= H(X \mid Z) - H(X \mid YZ)$$

$$I(x_1, x_2, \cdots, x_n)$$

$$= I(x_1) + I(x_2|x_1) + \cdots + I(x_n|x_1, x_2, \cdots, x_{n-1})$$

H.W.

Hint: Chain rule of probability.

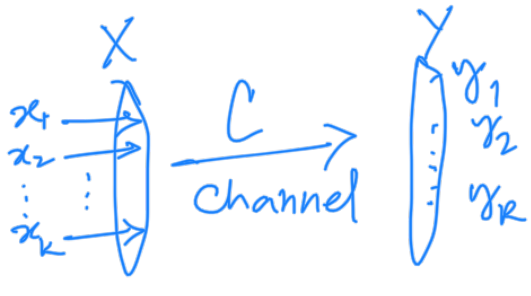
$$p(x_1, x_2, \cdots, x_n) = p(x_1)\, p(x_2|x_1) \cdots p(x_n|x_1, x_2 \cdots x_{n-1})$$

$$H(X_1 X_2 \cdots X_n)$$

$$= H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1 X_2 \cdots X_{n-1})$$

H.W

Lec-4



Let X be a random variable associated with a probability distribution $p(X=x)$, $x \in S_X$

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

↓

a measure of X

entropy of X: $0 \log 0 = 0$

Average — Self-information

$$I(a_i; b_j) = \log \frac{p(a_i|b_j)}{p(a_i)} \quad \text{bits}$$

(mutual information)

$p(a_i|b_j) \le 1$

$= I(b_j; a_i)$

Expectation

$$\text{Ave}(a_1, a_2, \dots a_n) = \frac{a_1 + a_2 + \dots + a_n}{n}$$

$$= \frac{1}{n} a_1 + \frac{1}{n} a_2 + \dots + \frac{1}{n} a_n$$

$$= p_1 a_1 + p_2 a_2 + \dots + p_n a_n$$

$$= \text{Exp}(X)$$

$$I(x) = -\log p(x) \quad \text{bits} \qquad \ge 0$$

"Self-information"

$H(X)$ → average self-information

$I(X;Y)$ → average mutual-information

$$H(X) = I(X;X)$$

Random variable — an amount of "randomness" that the

X — random variable

processes.

"Entropy measures uncertainty"

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

$H(X)$: # of possible outcome could be countably infinite

$\log p(x)$: $I(x)$

$\sum$: series

Exp. Entropy could be infinite.

Sample space $\leftarrow X = \{1, 2, 3, \cdots\}$

$$p(j) = \frac{c}{(j+1)(\log(j+1))^2}$$

where c is a constant whose value is decided by the condition $\sum_j p(j) = 1$

$$H_a(X) = -\sum p(x) \log_a p(x), \quad a > 1$$

then $H_b(X) = \log_b a \; H_a(X)$ (H.W.)

$\therefore$ Exp. $X \rightarrow$ is a random variable

$$\chi = \{H, T\} = \{0, 1\}$$

" X ": $\chi \to \mathbb{R}$

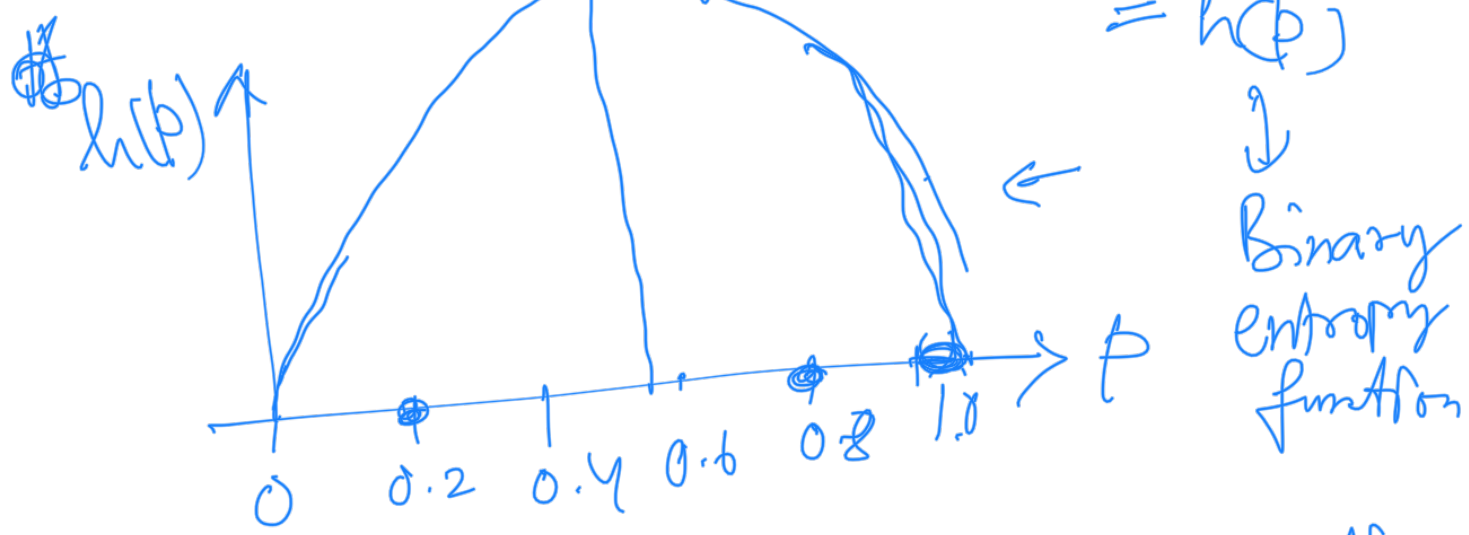
By $p(0) = p$, $p(1) = 1-p$

$$H(X) = -\sum_x p(x) \log p(x)$$

$$= -p(0)\log p(0) - p(1)\log p(1)$$

$$= -p\log p - (1-p)\log(1-p)$$

$$= h(p)$$



↓ Binary entropy function

From this exp, entropy is concave, continuous and the maximum value occurs at $p = \frac{1}{2}$.

— Entropy of a random vector

$$\vec{X} = (X_1, X_2, \cdots, X_n)$$

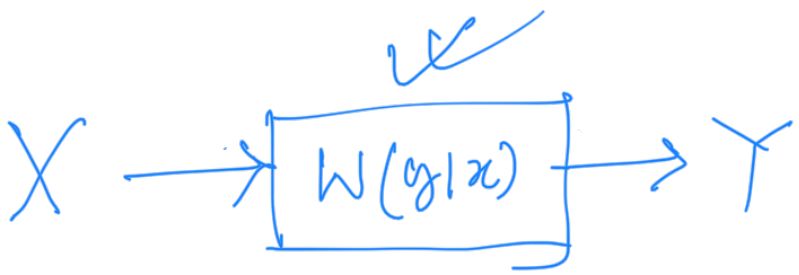
$$H(X_1, X_2, \cdots, X_n) = -\sum_{(x_1,\cdots,x_n)} p(x_1,\cdots,x_n) \log p(x_1,\cdots,x_n)$$

Conditional entropy

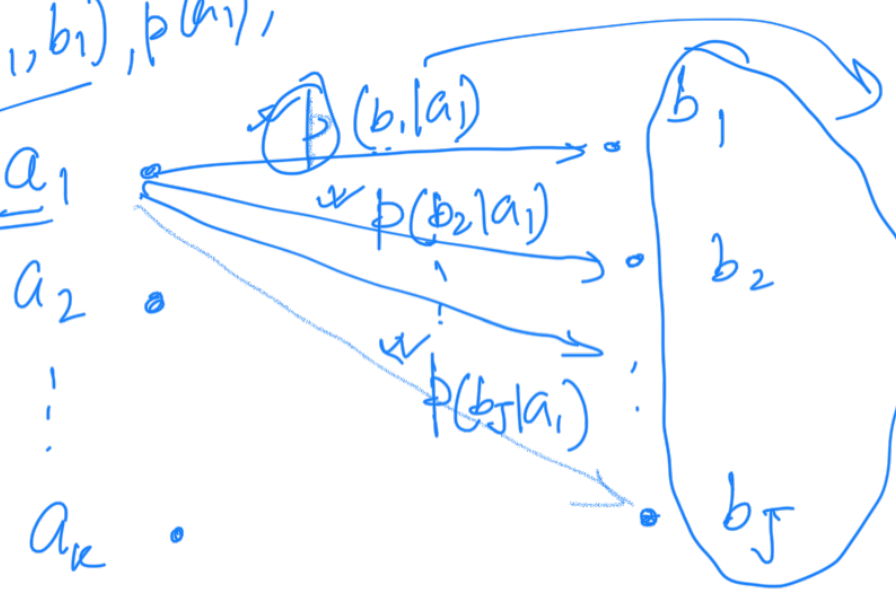
?? $H(Y|X) = \sum_{(x,y)} p(x,y) I(y|x)$

$= -\log p(y|x)$.

=



$p(a_1, b_1)$, $p(a_1)$,



induces

$p(y|x)$

$\forall y \in \{b_1, \ldots, b_J\}$.

$X = \{a_1, \dots, a_K\}$

$Y = \{b_1, \dots, b_J\}$.

Then $\sum_y W(y|x) = 1$, for any x

$W(y|x)$ → these conditional probabilities define the channel.

Due to the definition of conditional probability

$$P_{XY}(x,y) = P_X(x)\, W(y|x).$$

gives the joint distribution of (X,Y). We can calculate Joint entropy $H(X,Y) = H(XY)$.

Then we define a probability distribution on Y as follows:

$$P_Y(y) = \sum_x P_X(x)\, W(y|x).$$

If an input x is fixed then

the output Y has the distribution $W(y|x)$, $y \in \mathcal{Y}$

Then $H(Y|x) = -\sum_{y \in \mathcal{Y}} W(y|x) \log W(y|x)$

↓

"given x"

Now if we take the expectation over the distribution of X:

$$H(Y|X) = \sum_x \tilde{P}_X(x)\, \tilde{H}(Y|x)$$

(H.W.) $\overset{??}{=}$??

$$= -\sum P_{XY}(x,y) \log W(y|x)$$

$$= \sum_{x,y} P_{XY}(x,y)\, I(y|x)$$

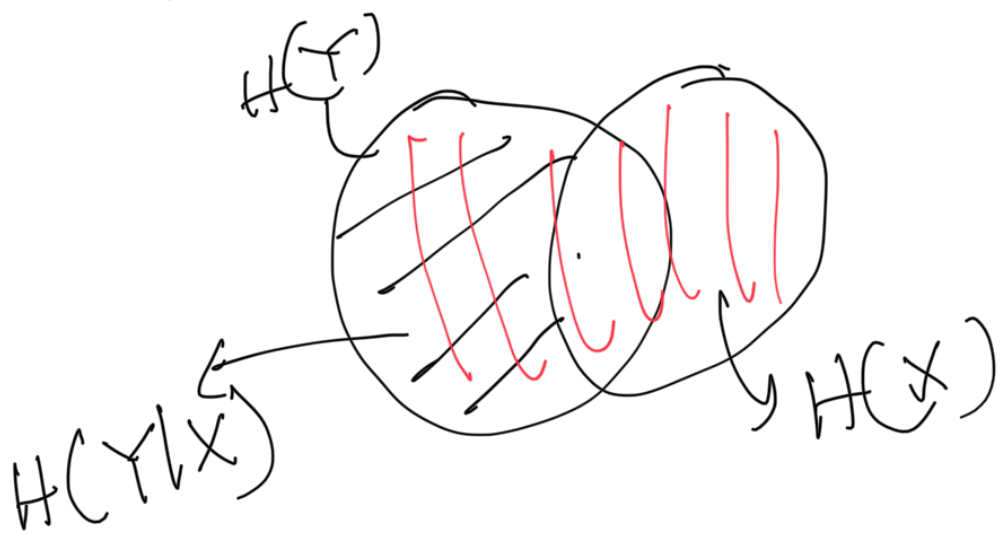
Obs. ① $H(Y|X) \geq 0$ (H.W.)

② $H(Y|X) \overset{?}{=} 0$

(H.W.)

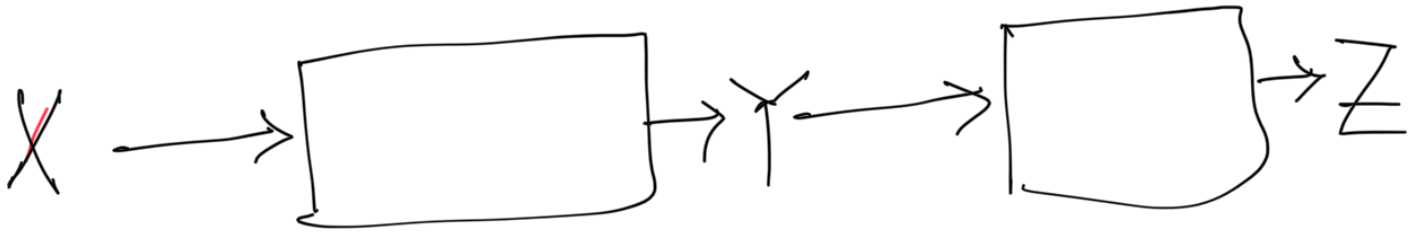
(iff f a function s.t. $f(X) = Y$.)

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



(4) Chain rule:

$$H(\bar{X}) = H(X_1, X_2, \cdots X_n)$$
$$= H(X_1) + H(X_2|X_1) + H(X_3|X_1X_2) + \cdots + H(X_n|X_1X_2 \cdots X_{n-1}).$$



$H(X, Y, Z) = ??$

Lec-5.

$X \rightarrow$ [channel]

① $H(X)$ → entropy of the r.v. X

② $H(Y)$ → entropy of the r.v. Y

③ $H(Y|X)$ → conditional entropy.

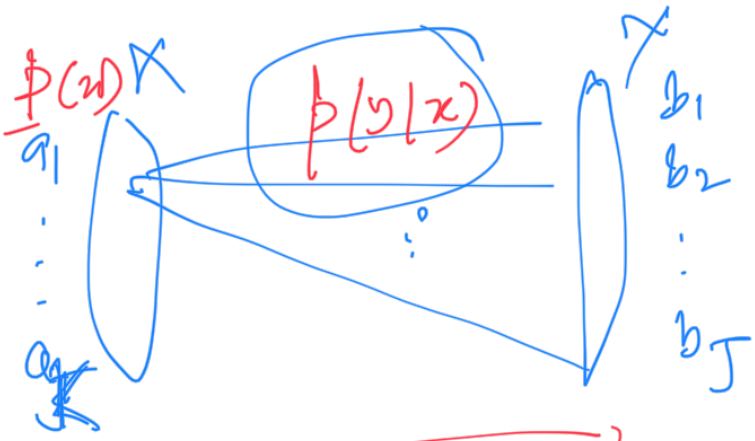
— Note: "Entropy is a function of the pmf associated with X".

Input

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

Channel

$p(b_j|a_K)$



Using $p(x)$ and $p(y|x)$ $\rightarrow$ $pW(y|x)$ we defined a pmf for Y, and here showed that

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

$$H(X_1, X_2, \ldots X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1X_2) + \ldots + H(X_n|X_1X_2\ldots X_{n-1})$$

Mathematical properties of entropy.

- $H(X)$ depends on the pmf of a random

Entropy depends ...

variable

a set of vectors

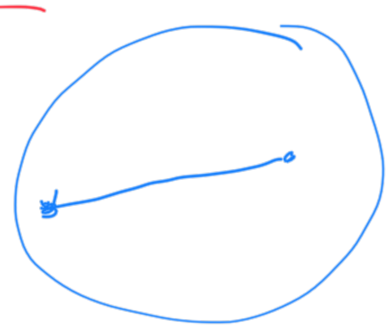
Let X be the sample space of a discrete random variable. $X : S \to \mathbb{R}$, $X(s_i) = x_i$, $i = 1, 2, \ldots$

where $S = \{s_1, s_2, \ldots\}$

$p(X = x) = f(x) \leftarrow$ pmf.

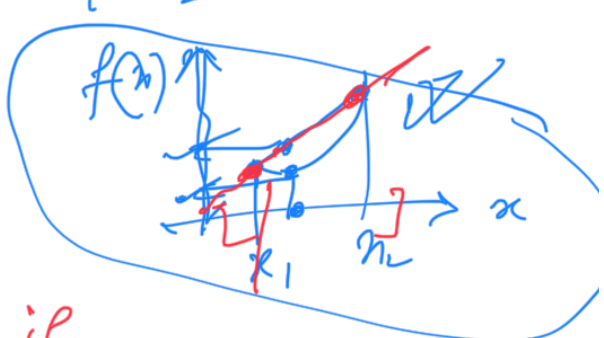
$$= \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \end{bmatrix} = \begin{bmatrix} P(X = x_1) \\ P(X = x_2) \\ \vdots \end{bmatrix}.$$

Convex function: Let f be a real-valued function defined on k-dimensional real vectors which form a "convex set". Then f is said to be a convex function if



$$f(\lambda_1 X_1 + \lambda_2 X_2) \leq \lambda_1 f(X_1) + \lambda_2 f(X_2)$$

where $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$.



$\lambda_1 X_1 + (1 - \lambda_1) X_2$

A function f is called concave if

$-f$ is convex.

Jensen's inequality:

Let $X = (X_1, X_2, \ldots, X_k)$ be a k-dimensional real-valued random variable. If $f(x)$ is a convex function then

$$E[f(X)] \geq f(EX).$$

Can be an equality only when $Y = EX$ w/p 1, for a strictly convex f.

Theorem.

(1) The entropy $H(X) \equiv H(P_X)$ is a strictly concave function.

(2) The entropy $H(Y) = H(P_X W_{Y|X})$ is a concave function with respect to the P_X.

(3) The conditional entropy $H(Y|X) = H(W_{Y|X} | P_X)$ is concave with respect to the channel $W_{Y|X}$, and linear w.r.t. the P_X.

Pf. (1) Let P_1 and P_2 be two pmfs corresponding to random variables X_1 and X_2 associated with a sample space $\mathcal{X}$.

$$H(\underbrace{\lambda P_1 + (1-\lambda) P_2}_{P}) \geq \lambda H(P_1) + (1-\lambda) H(P_2)$$

We have:

$$\sum_x P(x) \log \frac{1}{P(x)} \geq \lambda \sum_x P_1(x) \log \frac{1}{P_1(x)} + (1-\lambda) \sum_x P_2(x) \log \frac{1}{P_2(x)}$$

$f(y) = \boxed{y \log y}$. $f: [0,1] \to \mathbb{R}$

↳ is a "convex function"

↳ strictly convex fn.

H.W.

↳ achieved by using Jensen's inequality

$$E(f(X)) \geq f(EX)$$

Then taking sum of both the sides over all x gives the desired result.

(2) Let P_1 & P_2 be two probs.

$$P = \lambda P_1 + (1-\lambda) P_2$$

$$H(PW) = H((\lambda P_1 + (1-\lambda) P_2) W)$$

$$= H(\lambda P_1 W + (1-\lambda) P_2 W)$$

$$\geq \lambda H(P_1 W) + (1-\lambda) H(P_2 W)$$

Divergence.

Let X and Y be two random variables over the same sample space.

$\downarrow P_X$ & Q_X

Then the divergence, Kullback–Leibler information, or Boltzmann relative entropy of X & Y is defined as

$$D_a(X \| Y) = D_a(P_X \| Q_X)$$

$$= \sum P(x) \log \frac{P_X(x)}{}$$

$\frac{1}{T} \sum_{x \in S} k^{(x)} \log q_x(x)$ -

with the convention that

$0 \log(0/b) = 0, \quad b \geq 0$

$a \log(a/0) = +\infty, \quad a > 0.$

Observation - $D(X \| Y) \neq D(Y \| X)$

Log-Sum Inequality. For nonnegative numbers $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ we let $a = \sum_{i=1}^{n} a_i < \infty$, $b = \sum_{i=1}^{\infty} b_i < \infty$.

Then $\sum_{i=1}^{\infty} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}$

The equality holds if and only if a_i / b_i is constant for all i.

Result: $D(X \| Y) = D(P \| Q_x) \geq 0$ and equality holds if and only if $P = Q$.

Pf: Let $a_i = P(i)$, $b_i = Q(i)$ where $i \in$ Sample space corresponding to P and Q. Then $\sum Q_i = 1 = \sum b_i$ (the Q_i are a, the b_i are b)

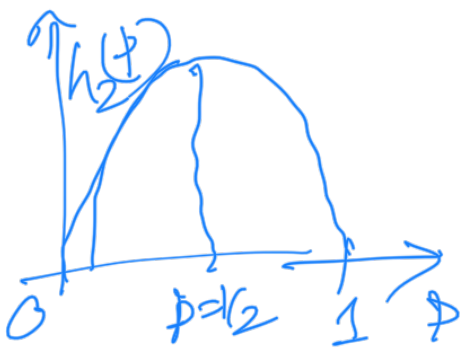
Alt. proof: Let $Y = Q_x / P_x$

$D(P \| Q) = \sum_i P(i) \left(-\log \frac{Q(i)}{P(i)} \right)$

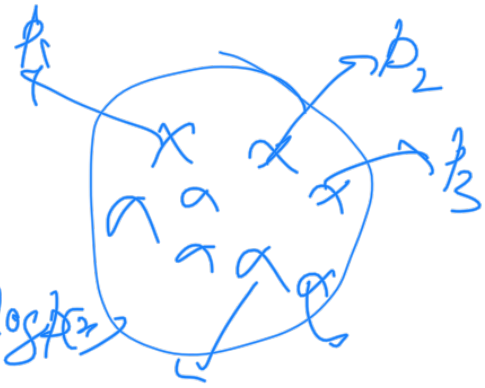
$$\geq -\log \sum_i p(i)\frac{q(i)}{p(i)}$$

$-\log y$ is a convex function and using the Jensen's inequality.

Q. What is the maximum value of entropy.



$$H(X) = H(p)$$
$$= -\sum_x p(x)\log p(x)$$



Result: $H(X) \leq \log |S_X|$ with equality only if X obeys the uniform distribution.

Pf. Let $P(x) = P_X(x)$

$$Q(x) = 1/|S_X|$$

Then
$$D(P\|Q) = \log|S_X| - H(P_X)$$
$$\geq 0$$

** The uniform distribution is defined only when the sample space is finite.

Result: If $X_1, X_2, \cdots, X_n$ then

$$\sum_{i=1}^{n} H(X_i) \geq H(X_1, X_2 \cdots X_n)$$

where equality holds only if X_i, $i=1,\cdots,n$ are independent

Pf. Let $P(x) = P(x_1, x_2, \dots x_n)$

$Q(x) = P(x_1) P(x_2) \dots P(x_n)$

$$D(P \| Q) = \sum_n P_n \log \frac{P(x)}{Q(x)} \geq 0$$

H.W. Give an example of a situation where this inequality is useful.

Result: $I(X;Y) \geq 0$

This implies $H(X|Y) \leq H(X)$ — Imp ??

The equality holds if X & Y are indepen.

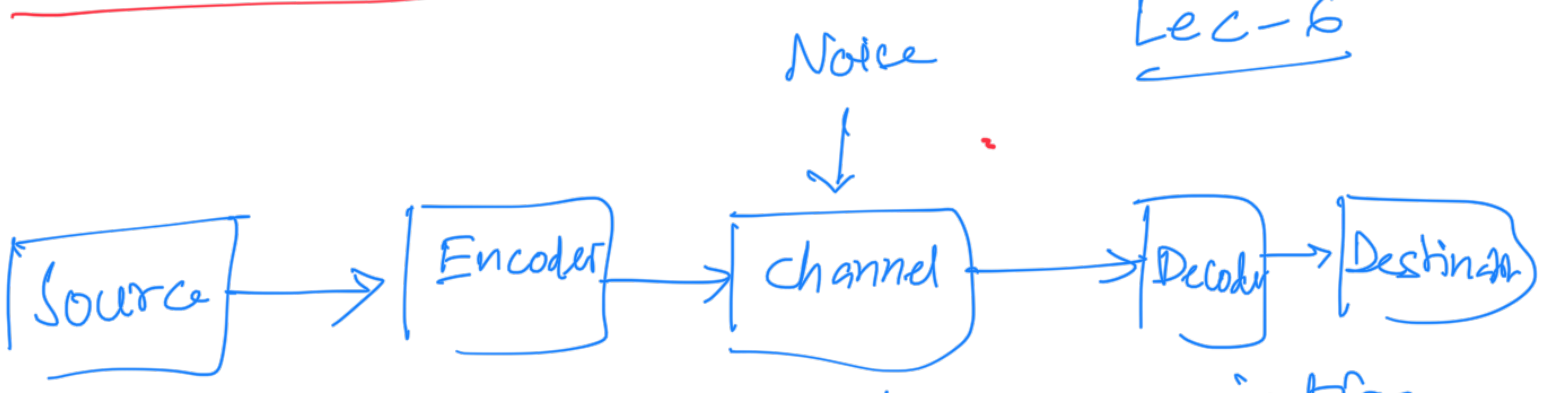
Pf. $P(x,y) = P_{XY}(x,y)$

$Q(x,y) = P_X(x)\, P_Y(y)$.

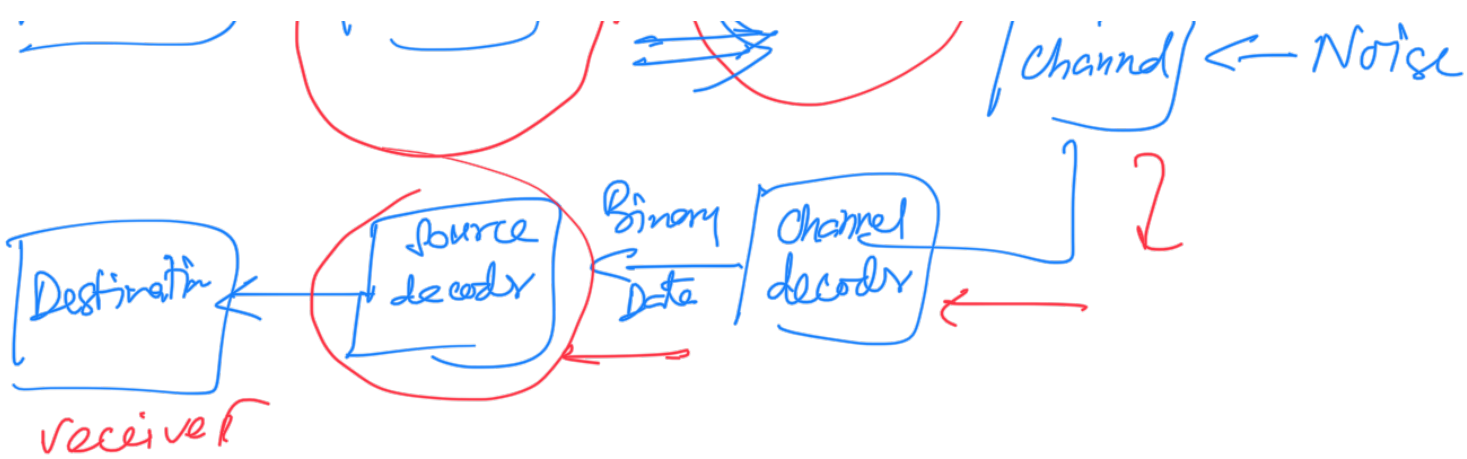
Then $D(P\|Q) = I(X;Y) \geq 0$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A|B) \neq P(A)$$

Lec-6



Block diagram of a communication system.



Let $a_1, a_2, \ldots, a_K$ be the source letters which the source produces.

Let $p(a_1), p(a_2), \ldots, p(a_K)$ be the probabilities of each of those possible outcomes, and these letters are independent, which we call as "memoryless source"

Morse Code - ".", "-"

$D = \{., -\}$

Def. A source code C for a random variable X is a mapping from $\mathcal{X}$, the range of X, to $\mathcal{D}^*$, the finite length string of symbols from a D-ary alphabet.

$$C: \mathcal{X} \to \{0, 1, \sim, D-1\} = \mathcal{D}^*$$

$C(x)$ = codeword corresponding to x.

$l(x)$ = the length of $C(x)$

$\sum_{x \in \mathcal{X}} p(x) l(x) :=$ ~~expected length of codeword.~~

Q. How to decide $C(a_i)$ and $l(a_i)$!!

→ variable-length codewords

Codewords → fixed-length.

Let a_1, a_2, a_3, a_4.

	Coding Sch-1	Coding Sch.
a_1	00	0
a_2	01	10
a_3	10	110
a_4	11	111

fixed-length ✓ | variable-length ✗

a a a a

Operational meaning of entropy.

Source letters a, b, c, d.

Alice → Bob.

$p(a) = 1/2$, $p(b) = 1/8$, $p(c) = 1/4$, $p(d) = 1/8$

Q. What should be the coding scheme for Alice?

$a = 00$, $b = 01$, $c = 10$, $d = 11$ ✓

Expected length of a codeword = 2 bits

a d b b d b b c a b
"00 11 01 01 11 10 10 11 00 10"

Uniquely decodable?

Is it ...?

Bigger question: What is the minimum value of expected length of a codeword? (among all possible codes)

$a \equiv 0$

$b \equiv 110$

$c \equiv 10$

$d \equiv 111$

Expected length

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

$= \frac{1}{2} + \frac{3}{8} + \frac{1}{2} + \frac{3}{8}$

$= \boxed{\frac{7}{4}} < 2$ bits.

Obs.

= Entropy of the source random variable

$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8}$

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

$= \boxed{\frac{7}{4}}$ bits.

—— o ——

Lec-08

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

$$\mathcal{A} = \{a_1, a_2, \ldots, a_K\}$$

Fixed length codewords for each a_i

variable length codewords.

Fixed length codewords.

How to decide the length?

Assume that the sender wants to send a finite sequence of source letters.

$$\bar{U}_L = (U_1, U_2, \ldots, U_L)$$

$p(a_1)$
$p(a_2)$
$\vdots$
$p(a_K)$.

$U_i \in \{a_1, a_2, \ldots a_K\}$

Q. There are K^L different sequences of source letters.

Suppose each source letter is encoded using alphabets from a set with D elements, and each codeword has length N.

Then # of codewords is D^N.

It is desirable that $K^L \leq D^N$

$$\Rightarrow L \log K \leq N \log D$$

$$\Rightarrow \frac{N}{L} \geq \frac{\log K}{\log D} = C$$

$\Rightarrow$ $\log K / \log D$ many code alphabets are required for each source letter.

Ex: "Teletype/teleprinter"

If we want to relax this then there will be situations i.e. we would not be able to decode the original sequence of source letters from the code sequence.

$\longrightarrow$

Q. What do we mean by saying L is large?

$\longrightarrow$

Assumption: The source is memoryless

$\bar{U}_L = (U_1, U_2, \ldots, U_L)$, U — i.i.d. random variables.

$$Pr(\bar{U}_L) = \prod_{l=1}^{L} P(U_l)$$

Exp. Suppose we have a source with letters a_1, a_2.

Let $P(a_1) = 0.7$, $P(a_2) = 0.3$.

$\bar{U}_3 = (U_1, U_2, U_3) = (a_2, a_1, a_1)$

$$P(\bar{U}_3) = p(a_2) \times p(a_1) \times p(a_1) = .3 \times .7 \times .7$$

Self-information of $\bar{U}_L$:

$$I(\bar{U}_L) = -\log P(\bar{U}_L) = -\log \prod_{l=1}^{L} P(u_l) = \sum_{l=1}^{L} -\log P(u_l) = \sum_{l=1}^{L} I(u_l)$$

Law of large numbers. (LLN)

For independent, identically distributed (i.i.d.) random variables $X_1, \dots, X_n$.

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to EX, \quad n \to \infty.$$

(sample mean)

For any $\delta > 0$ $\exists$ $\varepsilon(L, \delta) > 0$ s.t.

$$P\left[\left| \frac{I(\bar{U}_L)}{L} - H(U) \right| > \delta \right]$$

≥ 1 $\leq \varepsilon(L,\delta)$

and $\lim_{L\to\infty} \varepsilon(L,\delta) = 0$ (*)

Recall: we do not want to consider all seqs (of length L) of source letters.

Consider the set of sequences, denoted by T (Typical set)

for which

$$\left| \frac{I(\bar{U}_L)}{L} - H(U) \right| \leq \delta, \quad U_L \in T \qquad (**)$$

for any δ and 'L'

Then, from (*) we have

$$P(T) \geq 1 - \varepsilon(L,\delta).$$

From (**) $-\delta \leq \frac{I(\bar{U}_L)}{L} - H(U) \leq \delta$

$$\Rightarrow H(U) - \delta \leq \frac{I(\bar{U}_L)}{L} \leq H(U) + \delta$$

$$\Rightarrow L(H(U)-\delta) \leq I(\bar{U}_L) \leq L(H(U)+\delta)$$

$$\Rightarrow L(H(U)-\delta) \le -\log_2 P(\bar{U}_l) \le L(H(U)+\delta)$$

$$\Rightarrow \boxed{2^{-L(H(U)+\delta)} \le P(\bar{U}_l) \le 2^{-L(H(U)-\delta)}} \quad (***)$$

$\bar{U}_l \in T$

Q. What is the size of the typical set? i.e. What is the # of elements in T?

$$1 \ge P(T) = \sum_{\bar{U}_l \in T} P(\bar{U}_l)$$

$$\overset{??}{\ge} |T| \min_{\bar{U}_l \in T} P(\bar{U}_k)$$

since $P(\bar{U}_l) \ge \min_{\bar{U}_l \in T} P(\bar{U}_l)$

$$\Rightarrow |T| \min_{\bar{U}_l \in T} P(\bar{U}_l) \le 1$$

$$\Rightarrow |T| \le \frac{1}{\min_{\bar{U}_l \in T} P(\bar{U}_l)}$$

since $P(\bar{U}_l) \ge 2^{-L(H(U)+\delta)}$

$$\Rightarrow \frac{1}{P(\bar{U}_l)} \le 2^{L(H(U)+\delta)}$$

$$\Rightarrow \boxed{|T| \le 2^{L(H(U)+\delta)}}$$

$|T| \geq$ ~~$2^{L(H(U)+\delta)}$~~

H.W. Find a lower bound for $|T|$.

$$|T| \geq \left[1 - \varepsilon(L,\delta)\right] 2^{L(H(U)-\delta)}$$

Q. What is the value of N? i.e. = the fixed length of all codewords.

First, Let us choose N which satisfies: $N \log_2 D \geq L(H(U)+\delta)$ //

Then $\log_2 D^N \geq L(H(U)+\delta)$

i.e., $D^N \geq 2^{L(H(U)+\delta)}$

$\Rightarrow$ D^N is larger than the # of sequences in the typical set T

$\therefore$ a separate codeword is possible for each element $U_L \in T$.

Probability of error $\rightarrow$ denote it as P_e $\rightarrow$ there is a

seq of codewords for which the original
seq of source letters is not possible to
derive.

$$P_e \leq \; ??$$ (H.W.)

Otherwise. Let

$$N \log D \leq L(H(U) - 2\epsilon)$$

$$\Rightarrow D^N \leq 2^{L(H(U)-2\epsilon)}$$

From (***) $\quad p(\bar{U}_L) \leq 2^{-L(H(U)-\epsilon)}$

Q. what is the probability that we can provide separate codeword for elements in the Typical set?

Lec-8

Tomorrow – 2pm - 3pm
45 mins.
Topics - up to last lecture.
22 marks

Fixed length code

$\{a_1, a_2, \ldots, a_K\}$

$\mathcal{A} = \{a_1, a_2, \ldots\}$

Decide codewords for these alphabets.

N → length for each of the codeword.

$\mathcal{D}$ → the set code letters

$\mathcal{D} = \{0, 1\}$

Restriction — the receiver must decode the received seqn of source code letters uniquely.

For each seqn of source alphabets there must be a separate seqn of code letters.

$\bar{u}_L = (u_1, u_2, \ldots, u_L)$, $u_i \in \mathcal{A}$

U → random variable

For any $\delta > 0$, $\exists$ $\varepsilon(\delta, L)$ s.t.

$$P\left(\left| \frac{I(\bar{u}_L)}{L} - H(U) \right| > \delta \right) \leq \varepsilon(L, \delta)$$

$$\lim_{L \to \infty} \varepsilon(L, \delta) = 0$$

"Typical set of sequences" of alphabets.

$$T = \left\{ \bar{u}_L : \left| \frac{I(\bar{u}_L)}{L} - H(U) \right| \leq \delta \right\}.$$

$I(\bar{u}_L) = -\log_2 P(\bar{u}_L) = \sum I(u_k)$

① $P(T) \geq 1 - \varepsilon(\delta, L)$

② $2^{-L(H(U)+\delta)} \leq P(\bar{u}_L) \leq 2^{-L(H(U)-\delta)}$, $\bar{u}_L \in T$

③ $(1 - \varepsilon(\delta, L)) 2^{L(H(U)-\delta)} \leq |T| \leq 2^{L(H(U)+\delta)}$

Q. How to decide the value of N?

$$N \log_2 D \geq L[H(U) + \delta]$$

$$\Rightarrow \log_2 D^N \geq L(H(U) + \delta)$$

$$\Rightarrow D^N \geq 2^{L(H(U)+\delta)}$$

$\#$ of elements of the Typical set

If P_e denote the probability of error (meaning: for each received code letters there is a unique Asymptotic sequence)

$$P_e \leq \varepsilon(\delta, L)$$

If $L \to \infty$ then $P_e \to 0$

and this happens when

$$\frac{N}{L} \geq \frac{H(U) + \delta}{\log_2 D}$$

lower bound.

Q. Is there any upper bound for $\frac{N}{L}$?

$N \log_2 D \leq L\,(H(U) - 2\delta)$

$\Rightarrow D^N \leq 2^{L(H(U)-2\delta)}$ — (A)

$P(E) = P(T) \cdot P(\bar{U}_L)$

$P(\bar{U}_L) \leq 2^{-L(H(U)-\delta)}$ — (B)

What is the probability that there is a codeword of the typical set for an element of the typical set?

$\leq 2^{L(H(U)-2\delta)} \cdot 2^{-L(H(U)-\delta)}$

$= 2^{-L\delta}$

Recall: $P(T) \geq 1 - \varepsilon(\delta, L)$

The probability that $\bar{U}_L$ is a codeword is chosen from outside T is at most: $\varepsilon(\delta, L)$

The probability that there is a codeword of the sequence.

$1 - P_e \leq 2^{-L\delta} + \varepsilon(\delta, L)$

$\Rightarrow P_e \geq 1 - 2^{-L\delta} - \varepsilon(\delta, L)$

?? $P_e \geq 1 - \varepsilon(\delta, L) - 2^{-L\delta}$

?? ~~x~~ $\Rightarrow \quad 1 - P_e \leq \varepsilon(\delta, L) + 2^{-L(*)}$

$\delta > 0$,

Then if $L \to \infty$ then $P_e \to 1$

And this happens when

$$\frac{N}{L} \leq \frac{H(U) - 2\delta}{\log_2 D}$$

Source coding Theorem.

Suppose we have a discrete memoryless source with entropy $H(X)$. Consider coding a ~~set~~ sequence of length L into sequences of ~~code lengths~~ codewords each of length N from a set of code letters D.

Then for any $\delta > 0$, if

$$\frac{N}{L} \geq \frac{H(X) + \delta}{\log_2 D} \Rightarrow N \geq L H(X)$$

then the probability of error goes to 0 if L is large.

Conversely if $\frac{N}{L} \leq \frac{H(X) - 2\delta}{\log_2 D}$

then probability of error is close to 1 when $L \to \infty$.

... length codewords.

Variable - length ...

$$\mathcal{A} = \{a_1, a_2, \ldots, a_K\}$$

$$P(a_1)\quad P(a_2)\ldots P(a_K)$$

Suppose n_K be the length of the codeword for a_K, $K = 1, 2, \ldots, K$.

Estimate. $\bar{n} = \sum_{K=1}^{K} P(a_K) n_K$

↙ average # of code letters per source letter/alphabet !!

Source letters	$P(a_K)$	Code I	Code II	Code III
a_1	0.5	0	0	0
a_2	0.25	0	1	10
a_3	0.125	1	00	110
a_4	0.125	10	11	111
		X	X	✓

Code II

$a_1 a_1 \rightarrow$ "00"

$a_2 \rightarrow$ "11"

Defn. A code is called uniquely decodable if for each seq of source letter, The seq of code letters corresponding that source seq is separate from any other source sequence.

does not give any clue about how to define or generate a uniquely decodable code.

Def. This is a code in which

Prefix Code: [illegible]
no codeword is a prefix of any
other codeword.

Prob 6. of Assignment - 1.

Q: $P \equiv (p_1, p_2, \dots)$.

If $\boxed{\sum_{n=1}^{\infty} p_n \log n} < \infty$ then $H(P) < \infty$.

Ans. $H(P) = \sum_{n=1}^{\infty} p_n \log \frac{1}{p_n} < \infty$

$= -\sum_{n=1}^{\infty} p_n \log p_n$

Weierstrass M-test.

$\sum f_n(x) < \infty$, and $|f_n(x)| \le M_n$, $\sum M_n < \infty$

$H(P) = -\sum_{n=1}^{\infty} p_n \log p_n$ $\le$ $\sum p_n \log n$

Archimedian property

$x \in \mathbb{R} \Rightarrow \exists\, n \in \mathbb{N}$ s.t. $n > x$

for $\frac{1}{p_n}$ $\exists\, N_{p_n} \in \mathbb{N}$ s.t. $N_{p_n} > \frac{1}{p_n}$.

$H(P) = \sum_{n \ge 1} p_n \log \frac{1}{p_n} \le \sum p_n \log N_{p_n} < \infty$

Converse is not True:

Find an exp of "P" $P = (p_1, p_2, \dots)$ s.t.

$$H(P) = -\sum p_n \log p_n < \infty$$

but

$$\sum_{n=1} p_n \log n \not< \infty$$

meaning: $p_n \geq$ s.t. $\sum_{n=1}^{\infty} p_n = 1$

$$p_n = \begin{cases} \frac{1}{2^k} & \text{if } n = 2^{2^k} \\ 0 & \text{otherwise} \end{cases}$$

!!

Lec 9.

(1) Memoryless Channel.

$\bar{U}_L = (u_1, u_2, \dots, u_L)$

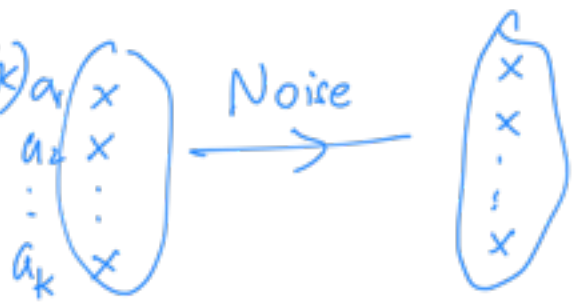
(2) Codeword length is fixed

Lec-10

Variable-Length codewords.

$p(a_1), p(a_2), \dots, p(a_K)$

$\mathcal{D} = \{0, 1\}$



Code — set of codewords

For each symbol a_k there is a codeword from the set $\mathcal{D}^*$, which contains strings of 0, 1's.

code

$a_1 \rightarrow (10\ldots1)$, $a_2 \rightarrow (01\ldots0)$ ... a_K

If the receiver receives a Source Sequence
then the receiver must be able to "decode" it

Coding → Fixed length codewords
→ variable length codewords.

If the sender sends a Source Seq. of length L

$$\text{length} \left(\frac{N}{L}\right) \geq \left(H(U) + \delta\right) / \log D, \quad D = |\mathcal{B}|$$

$H(U)$ → entropy
δ → Typical set

Variable-length codewords.

$a_1, a_2, \ldots, a_K$, $b(a_1), b(a_2), \ldots, b(a_K)$

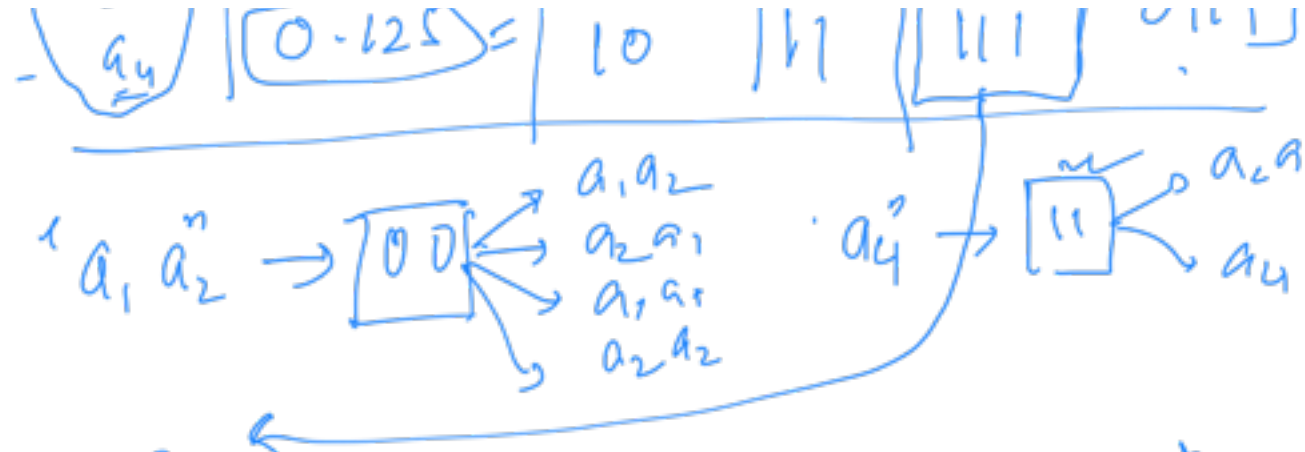
$n_1 \quad n_2 \quad n_K$

$$\bar{n} = \sum_{k=1}^{K} n_k\, b(a_k)$$

Q. What is the min value of $\bar{n}$

$\min \bar{n}$
the code must be uniquely decodable

Source letters	$P(a_k)$	Code$_1$	Code$_2$	Code$_3$	Code-4
a_1	0.5	0	0	0	0
a_2	0.25	0	1	10	01
a_3	0.125	1	00	110	011



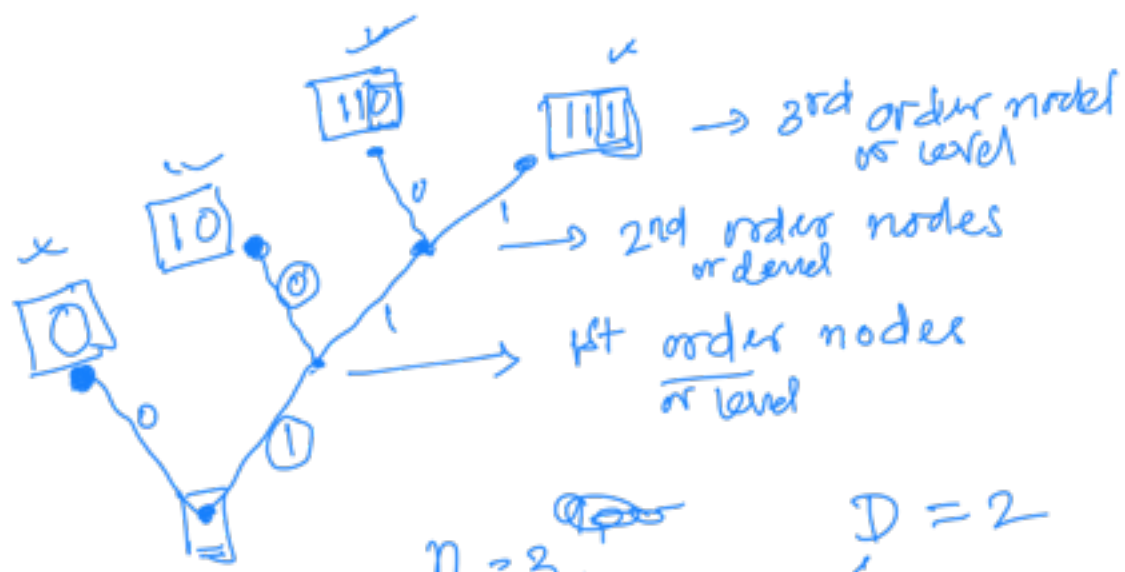
Code-3

No codeword is a prefix of another codeword

Prefix Code :- → It is called "instantaneous code."
→ uniquely decodable.

Code-4. is uniquely decodable but not prefix code.

$a_1 a_3 a_4 \rightarrow$ 0 | 110 | 111
a_1 a_3 a_4

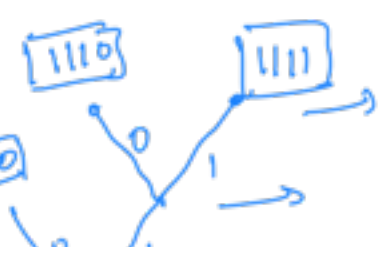


$n = 3$ → order

$D = 2$ → Code alphabet size

$a_1 = 0$
$a_2 = 10$
$a_3 = 11$

$a_1 = 0$

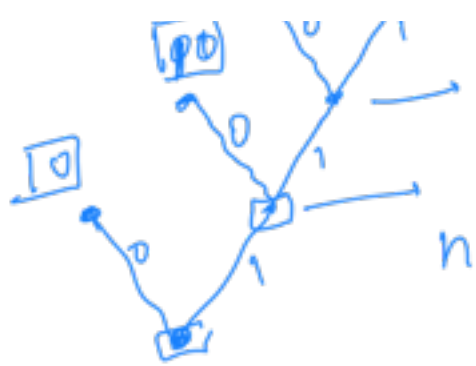


$a_2 = 10$

$a_3 = 110$

$a_4 = 1110$

$a_5 = 11111$



$n = 4$ $\boxed{D = 2}$

$$\mathcal{A} = \{0, 1, \dots, 100\}$$

H.W. Show that this construction of generation of prefix code works for any D. For example $D = 3$.

Check: For Code-3, $\bar{n} = 1.75$ and $H(U) = 1.75$

Kraft inequality, 1949

Suppose $n_1, n_2, \dots, n_K$ are positive integers that satisfy

$$(*) \quad \sum_{k=1}^{K} D^{-n_k} \le 1$$

Then $\exists$ a prefix code with alph code alphabet size D s.t. n_i is the length of the codeword for a_i, $1 \le i \le K$

Conversely, the lengths of codewords of a prefix code satisfies the inequality $(*)$.

Pf. "D" need not be 2.

Suppose we consider the tree. Then if the tree has order n the # of terminal nods is D^n with D number of nodes of order j originating from each node of

"$a_1, a_2, \dots, a_K$"

$n_1 \quad n_2 \quad n_K$

order $i-1$.

# of terminal nodes
$= 8 = 2^3$

$D = 2, n = 3$

Obs. 1 There is fraction of $D^{-1} = \frac{1}{D}$ number of nodes of each order $j \geq 1$ originate from each of the D nodes of order 1.

Obs. 2 There is a fraction of $D^{-2} = \frac{1}{D^2}$ number of nodes of each order $j \geq 2$ originate from each of the D^2 nodes of order 2

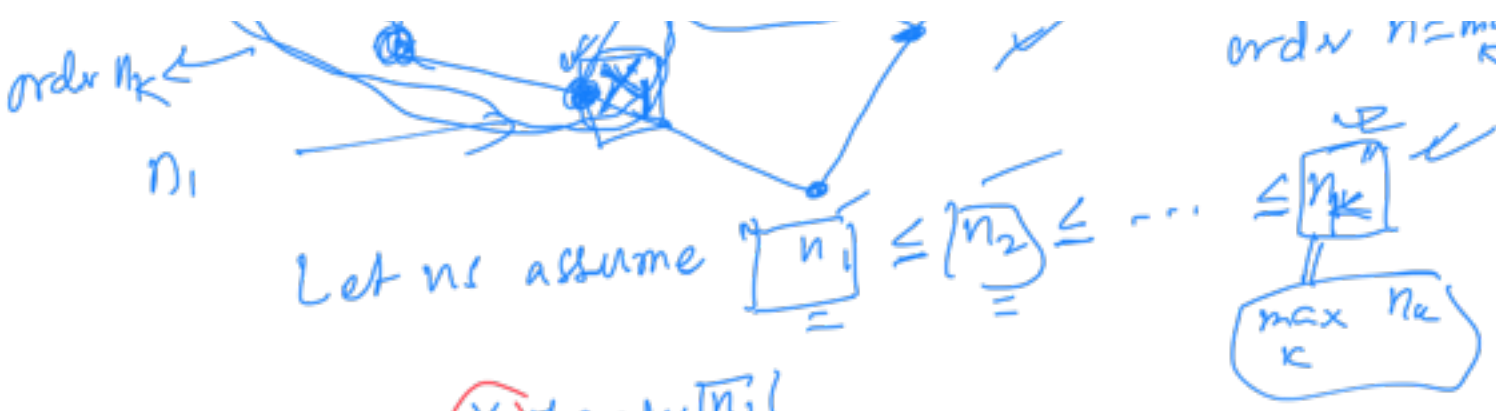
Final observation: There is a fraction of $D^{-j} = \frac{1}{D^j}$ number of nodes of each order greater or equal to j originate from each of the D^j nodes of order j.

Suppose $n_1, n_2, \dots, n_K$ satisfy the Kraft inequality (*)

Now consider the tree of order n which equals to the largest n_k.

Goal: To identity the nodes of the tree to define a prefix code.

→ Tree of



Let us assume $n_1 \le n_2 \le \cdots$

X_1 of order n_1

Pick any node of the trace.

Then there are the fraction D^{-n_1} nodes that originate from X_1.

Pick another node X_2 of order n_2 that does not belong to the branch originated from X_1

H.W. what happens where there are multiples $n_K = n_1$.

there are fraction of D^{-n_2} nodes that originate from X_2.

Then the total # of nodes that can not be chosen for next codeworde is

$$D^{-n_1} + D^{-n_2} \le 1$$

Then for a_l, $l < K$ there will be still a fraction of

$$D^{-n_1} + D^{-n_2} + \cdots + D^{-n_l} \le 1$$

available for the next step, as long as $l < K$

Therefore we can always be able to identify nodes from the tree of order $\max_k n_k$ that will be the beginning for a prefix

provide a unique
code.

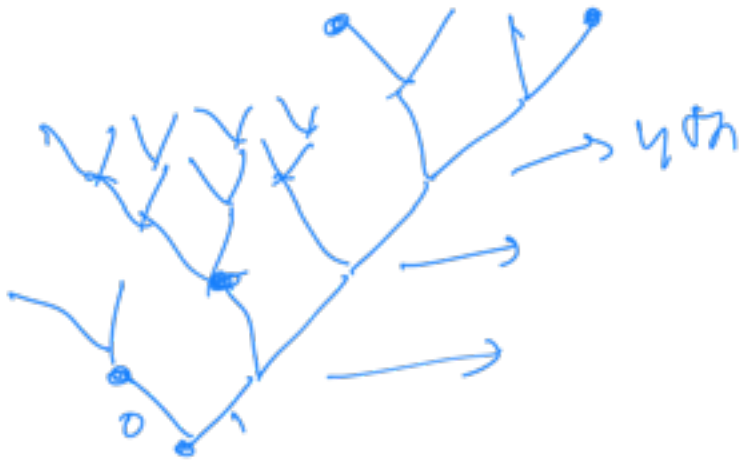
Conversely, If there is a prefix code with lengths of the codewords as $n_1, \dots, n_K$

Goal: Show that Kraft inequality is true.

$n_1 = 2$

$n_2 = 4$

$n_3 = 7$



Remark. The Kraft inequality is an if and only if condition for a prefix code.

Q. Is it true for uniquely decodable codes?

Thm If C is a uniquely decodable code then the Kraft inequality is always true.

i.e. C is a code on the code alphabet of size $D = |\mathcal{A}|$ with codewords of lengths $n_1, n_2, \dots, n_K$

Then $\sum_{k=1}^{K} D^{-n_k} \leq 1$

$\min \bar{n} = \sum_{k=1}^{K} n_k p_k$

s.t. $\sum_{k=1}^{K} D^{-n_k} \le 1$

Lec-11

Pf. Let L be a positive integer then

$$(*) \quad \left(D^{-n_1} + D^{-n_2} + \cdots + D^{-n_K}\right)^L = \sum_{K_1=1}^{K}\sum_{K_2=1}^{K}\cdots\sum_{K_L=1}^{K} D^{-(n_{K_1}+n_{K_2}+\cdots+n_{K_L})}$$

H.W. Using mathematical induction

For instance if $K=2$ and $L=2$ then

$$\left(D^{-n_1} + D^{-n_2}\right)^2 = 1\left(D^{-n_1}\right)^2 + 1\left(D^{-n_2}\right)^2 + 2D^{-n_1}D^{-n_2}$$

$$= D^{-n_1-n_1} + D^{-n_2-n_2} + 2D^{-n_1-n_2}$$

$$= \sum_{K_1=1}^{2}\sum_{K_2=1}^{2} D^{-(n_{K_1}+n_{K_2})}$$

$D^{-(n_1+n_2)} + D^{-(n_2+n_1)}$

$a_{i_1}a_{i_2}$ → symbols

n_1 n_2

all 2 length "sequence" of codewords with 2 at

a_1a_1, a_2a_2, a_1a_2, a_2a_1

$(\ )(\ )$ n_1 n_1 n_2 n_2 n_1 n_2 n_2 n_1

Corresponding to each sequence of L codewords there is a term in the right hand side of the equality.

total length of L codewords for a sequence of L symbols/codewords → $n_{K_1} + n_{K_2} + \cdots + n_{K_L}$

For exp. the symbols a_1, a_2, a_3, a_4 with n_1, n_2, n_3, n_4

Then the sequence of 3 symbols

$a_1a_1a_1$, $a_1a_2a_4$, $a_2a_3a_4$, $a_3a_1a_2$, ...

$A_i = \{(c_1, c_2, \dots, c_L)$ | $i = n_{k_1}+n_{k_2}+\dots+n_{k_L}$ where n_{k_1} is the $|C_1|$, n_{k_2} ... $|C_L|\}$

and $A_i = |A_i|$ = the total number of sequences of L codewords that have total length i.

From (*)

$$\left(\sum_{k=1}^{K} D^{-n_k}\right)^L = \sum_{i=1}^{L n_{max}} A_i D^{-i} \quad (**)$$

where n_{max} is the longest of the n_k.

Then $A_i \le D^i$ since the code is uniquely decodable.

Then from (**)

$$\left(\sum_{k=1}^{K} D^{-n_k}\right)^L \le \sum_{i}^{L n_{max}} D^i D^{-i} = L n_{max}$$

$$\Rightarrow \sum_{k=1}^{K} D^{-n_k} \le \left(L n_{max}\right)^{1/L} = L^{1/L} \cdot n_{max}^{1/L}$$

If we consider $L \to \infty$ then

$$\Rightarrow \sum_{k=1}^{K} D^{-n_k} \le \lim_{L\to\infty} L^{1/L} = 1$$

H.W.

$|G_n| \le n^{1/n} \le |b_n|$ use Sandwich theorem.

H.W. Given an exp. of a code which is NOT uniquely decodable but the Kraft inequality is satisfied.

The Source Coding Theorem

D → the # of code alphabet.

Let U → random variable corresponds to the information source.

$X \to \{a_1, \dots, a_K\} \xrightarrow{U} \{p(a_1), p(a_2), \dots, p(a_K)\}$

Then the expected length of a uniquely decodable code satisfies.

$$\bar{n} \geq \frac{H(U)}{\log_2 D} \quad (*)$$

It is possible to define a prefix code such that

$$\frac{H(U)}{\log_2 D} \leq \bar{n} < \frac{H(U)}{\log_2 D} + 1$$

Pf: $p(a_1), \dots, p(a_K)$

Let $n_1, \dots, n_K$ be the length of the code words corresponding to $a_1, a_2, \dots, a_K$.

In order to prove (*) it is enough to show $H(U) - \bar{n} \log_2 D \leq 0$

L.H.S. $H(U) - \bar{n} \log_2 D$

$$= \sum_{k=1}^{K} p(a_k) \log \frac{1}{p(a_k)} - \sum_{k=1}^{K} p(a_k) n_k \log_2 D$$

$$= \sum_{k=1}^{K} p(a_k) \log \frac{1}{p(a_k)} - \sum_{k=1}^{K} p(a_k) \log_2 D^{n_k}$$

$$= \sum_{k=1}^{K} p(a_k) \log \frac{D^{-n_k}}{p(a_k)}$$

(H.W.) $\log x \leq x - 1$ for $x > 0$

Note that $\log_2(x) \le (x-1)\log_2 e$

$$\text{Ian } H(U) - \bar{n}\log_2 D \le \log_2 e \left[\sum_{k=1}^{K} D^{-n_k} - \sum_{k=1}^{K} p(a_k)\right]$$

$$\le 0$$

$$\Rightarrow \boxed{\bar{n} \ge \frac{H(U)}{\log_2 D}}$$

Q. When can this be an equality?

Ans. "$p(a_k) = D^{-n_k}$" $\forall k$

2nd part of the theorem.

Recall that $p(a_1), p(a_2), \ldots, p(a_K)$.

If we are able to choose n_k that satisfy $\boxed{p(a_k) = D^{-n_k}}$ $\forall\ 1 \le k \le K$ (***)

If we are not able to find true integers $n_1, n_2, \ldots, n_K$ that satisfy (***) then we choose n_k s.t.

$$\boxed{D^{-n_k} < p(a_k) < D^{-n_k+1}} \quad \forall\ 1 \le k \le K$$

$$\Rightarrow \boxed{\sum_{k=1}^{K} D^{-n_k} \le \sum_{k=1}^{K} p(a_k)} \rightarrow \text{Kraft inequality}$$

From the left hand inequality

From the right hand inequality $\Leftarrow$ $\log_2 p(a_k) < \log D^{-n_k+1}$

$\Rightarrow \log p(a_k) < (-n_k+1)\log D$

$$\Rightarrow \log p(a_k) < -n_k \log D + \log D$$

$$\Rightarrow \frac{\log p(a_k)}{\log D} < -n_k + 1$$

$$\Rightarrow n_k < -\frac{\log p(a_k)}{\log D} + 1$$

Taking expectation

$$\Rightarrow p(a_k) n_k < -\frac{p(a_k) \log p(a_k)}{\log D} + p(a_k)$$

$$\Rightarrow \sum_{k=1}^{K} p(a_k) n_k < \frac{1}{\log D} \sum p(a_k) \log \frac{1}{p(a_k)} + \sum p(a_k)$$

$$\Rightarrow \bar{n} = \frac{H(U)}{\log(D)} + 1$$

∴ We can construct a prefix code such that

$$\frac{H(U)}{\log D} \leq \text{"}\bar{n}\text{"} < \frac{H(U)}{\log D} + 1$$

If the source is memoryless we have the following bound for a prefix code.

If the receiver receive a sequence of codewords, the receiver must be able to decode uniquely.

decode it

$a_1 \to$ "n_1", $a_2 \to$ "n_2", ..., $a_K \to$ "n_K".

Theorem. "Assume that we have a memoryless source. And we know D, $a_1, \ldots, a_K, p(a_1), \ldots, p(a_K)$. We will be able to construct a prefix code such that.

$$\frac{H(U)}{\log D} \le \bar{n} < \frac{H(U)}{\log D} + \frac{1}{L}$$

Pf

$U^L = (U_1, U_2, \ldots, U_L) = \bar{U}$

$U_1, U_2, \ldots, U_L$ i.i.d

$$H(\bar{U}) = H(U_1, U_2, \ldots, U_L) = L\,H(U) \quad \text{(from the chain rule)}$$

~~$H(X_1, X_2) = H(X_1) + H(X_2|X_1)$~~

The average length of the code words corresponding to $\bar{U}$ is $L\bar{n}$ where $\bar{n}$ is the average length for codewords corresponding to U.

$p(u_1, \ldots, u_k) \prod_{i=1}^{k}$

From the previous theorem we have

$$\frac{H(\bar{U})}{\log D} \leq L\bar{n} < \frac{H(\bar{U})}{\log D} + 1$$

i.e. $$\frac{L\,H(U)}{\log D} \leq L\bar{n} < \frac{L\,H(U)}{\log D} + 1$$

i.e. $$\frac{H(U)}{\log D} \leq \text{“}\bar{n}\text{”} < \frac{H(U)}{\log D} + \frac{1}{L}$$

$$\frac{H(U)}{\log_2 D} \leq \boxed{\bar{n}} = \sum n_k p_k \leq \frac{H(U)}{\log_2 D} + \frac{1}{L}$$

Lec-12

Q. How to construct a uniquely decodable code such that the minimum value of $\bar{n}$ is achieved.

“a_1”, “a_2”, …, “a_k”

$p(a_1)$ $p(a_2)$ … $p(a_k)$

$l(c_1)$ $l(c_2)$ … $l(c_k)$

“n_1” “n_2” … “n_k”

Q. Determine $n_1, n_2, \ldots$

$\max_{1 \le k \le K} n_k$

$\bar{n} = \sum n_k p_k$

Q. Does there exist any other uniquely decodeable code s.t. the average/expected codeword length is less than ?!

Huffman code, 1952

There is no other code whose average codeword length is smaller than the average codeword length of the Huffman code.

Assumption:

Source letters/alphabets: $a_1, a_2, \ldots, a_K$

Assume that $p(a_i)$ are ordered i.e.

$p(a_1) \ge p(a_2) \ge \cdots \ge p(a_{K-1}) \ge p(a_K)$

$\mathcal{D} = \{0, 1\}$, $D = 2$

Goal: Determine $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_K$

where $\bar{x}_j \equiv$ the codeword for a_j, $1 \le j \le K$

and $n_j \equiv$ the length of codeword $\bar{x}_j$.

with possible minimum value of $\bar{n} = \sum_{k=1}^{K} n_k p(a_k)$.

$k=1$

Thm. An optimal binary code exists such that the least likely code words $\bar{x}_{K-1}$ and $\bar{x}_K$ have the same length and they differ only at the last digit, $\bar{x}_K$ ends with '1' and $\bar{x}_{K-1}$ ends with '0'.

$\bar{x}_{K-1} = 11\ldots10$
$\bar{x}_K = 11\ldots11$

Pf. "$n_K \geq n_i$" true for optimal code

if not then there must exist a code in which $n_K < n_i$ for some i.

The we can interchange the code words $\bar{x}_i$ and $\bar{x}_K$ and the effect of this in the expected length is as follows:

$$\Delta = [p(a_i) n_K + p(a_K) n_i] - [p(a_i) n_i + p(a_K) n_K]$$

$$= [p(a_i) - p(a_K)] n_K + [p(a_i) - p(a_K)] n_i$$

$$= [p(a_i) - p(a_K)] [n_K - n_i]$$

$$\leq 0$$

optimal code.

So in the optimal code there

$n_1 \leq n_2 \leq \cdots \leq n_{k-1} \leq n_k$

Obs. in the optimal code there must be another codeword which will differ for $\bar{x}_k$ only in the last digit.

Otherwise we can discard the last digit of $\bar{x}_k$ can be dropped.

Finally, if $\boxed{\bar{x}_i}$ is the codeword which differs from $\bar{x}_k$ in only one position then we should have $n_i = n_{k-1}$

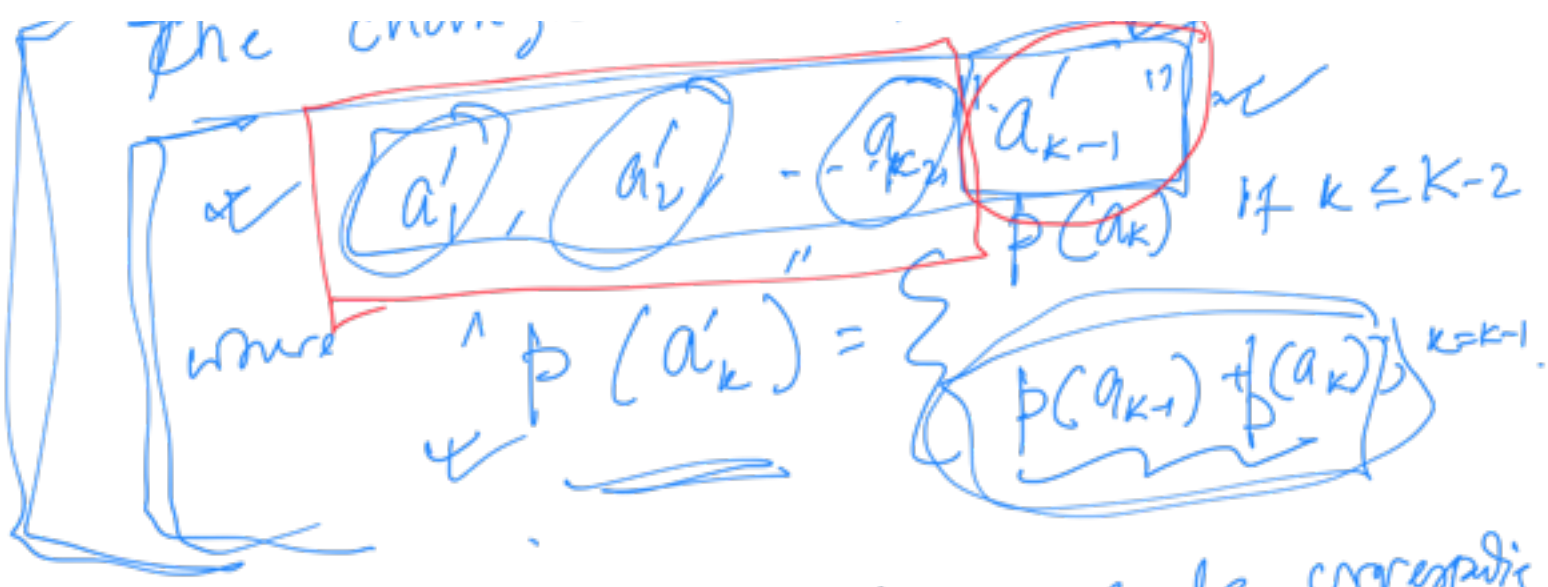
then assign the codewords $\bar{x}_i$ and $\bar{x}_{k-1}$ can be interchanged without increasing the average $\bar{n}$.

Then we have reduced the original problem of construction of the optimal code into following:

The reduced problem:

we need codewords $\bar{x}_1, \ldots, \bar{x}_{k-2}$, and the first n_{k-1} digits of $\bar{x}_{k-1}$ or $\bar{x}_k$.

changed setup:



Thm. If the prefix code corresponding to the reduced problem is optimal then the prefix code corresponding to the original problem is optimal.

$\bar{n}'$ → $\bar{n} = \bar{n}' + "\alpha"$

$y = f(x) = "x" + "3"$

Pf. We want to compute n'_k.

First let us find out

$$n_k = \begin{cases} n'_k, & k \le K-2 \\ n'_k + 1, & k = K-1, \\ & k = K-2 \end{cases}$$

Then $\bar{n} = \sum_{k=1}^{K} p(a_k) n_k$

$$= \sum_{k=1}^{K} p(a_k) n'_k + [p(a_{K-1}) + p(a_K)]$$

$$= \sum_{k=1}^{k-2} p_r(a_k') n_k' + p(a_{k-1}')(n_{k-1}'+1)$$

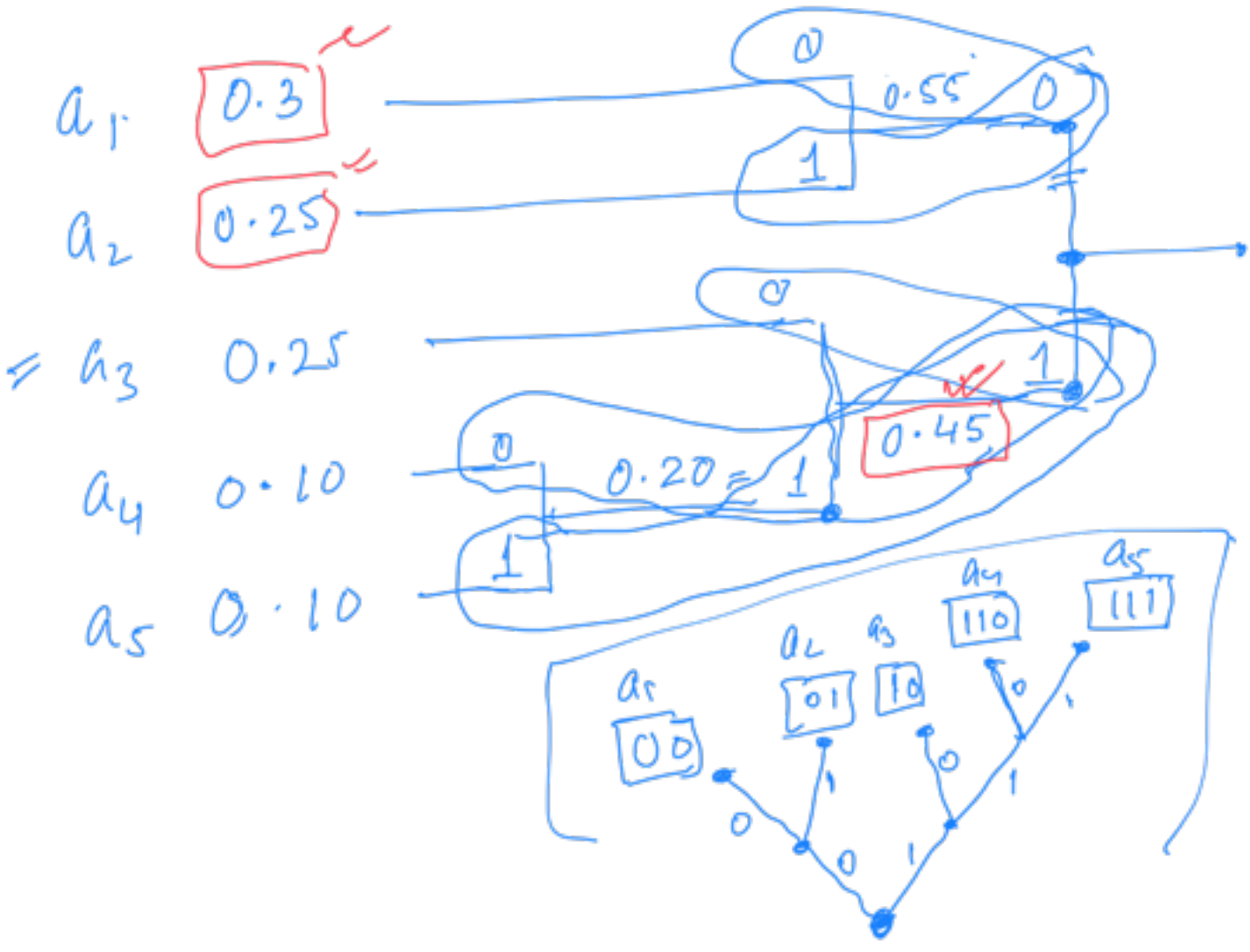
$$= \bar{n}' + p(a_{k-1}')$$

Exp. 1.

$a_1 = 0.3$, $a_2 = 0.25$, $a_3 = 0.25$,

$a_4 = 0.10$, $a_5 = 0.10$

$p(a_1) \geq p(a_2) \geq \cdots \geq p(a_5)$.



Huffman's code.

Algorithm for Huffman

(1) order the source letters $a_1, \dots, a_K$ such that $p(a_1) \geq p(a_2) \geq \dots \geq p(a_K)$

(2) Assign to letter a_{K-1} code symbol 0 and to letter a_K code symbol 1.

(3) Construct a reduced source alphabet $\{a_1, a_2, \dots, a_{K-2}, (a_{K-1}, a_K)\}$ with probabilities $p(a_1), p(a_2), \dots, p(a_{K-2}), p(a_{K-1}) + p(a_K)$

Repeat the steps (1) and (2).