3. $X \in \{1, 2, \cdots, m\}$

$p_1, p_2, \cdots, p_m$

$Y \in \{2, 3, \cdots, m\}$

$p_1 + p_2, p_3, \cdots, p_m$

$Z \in \{0, 1, 2\}$.

$Z = 1$, if $X = 1$.

$= 0$, if $X = 2$.

$= 2$, otherwise

Apply Chain Rule

---

* Mutual Information & its properties

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

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

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

$$= \sum_{x,y} p(x,y) \log \left[\frac{p(x,y)}{p(x)p(y)} \right]$$

$(I(X,Y) = I(Y,X))$

Symmetric ⇒ Makes Sense to call it mutual information

$D(p(x,y) \| p(x)p(y))$

$I(X;Y) \geq 0$ with $=$ iff $X \perp\!\!\!\perp Y$

$I(X;Y|Z) := H(X|Z) - H(X|Y,Z)$

Chain Rule

$$I(X_1^n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_1^{i-1})$$

Thm: a) $I(X;Y)$ is a concave function of $p(x)$, for $p(y|x)$ fixed

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

b) $I(X;Y)$ is a convex fn of $p(y|x)$ for $p(x)$ fixed.

Proof: a) $I(X;Y) = H(Y) - H(Y|X)$

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

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

[Communication of Noisy Channel]

linear in $p(x)$

$H(Y)$ is concave in $p(y)$.

$p(y) = \sum_x p(x)\, p(y|x)$

fixed.

(linear change of variable)

$\therefore H(Y)$ is concave fn of $p(x)$.

---

b) $p_0(x,y) = p(x)\, p_0(y|x)$,

$p_1(x,y) = p(x)\, p_1(y|x)$.

$p_\lambda(x,y) = (1-\lambda)\, p_0(x,y) + \lambda\, p_1(x,y)$

$p_\lambda(y|x) = (1-\lambda)\, p_0(y|x) + \lambda\, p_1(y|x)$

$\therefore p_\lambda(x,y) = p(x)\, p_\lambda(y|x)$.

- $p_\lambda(y) = (1-\lambda)\, p_0(y) + \lambda\, p_1(y)$.

$q_\lambda(x,y) = p(x)\, p_\lambda(y) = (1-\lambda)\, q_0(x,y) + \lambda\, q_1(x,y)$

$I(X;Y) = D\left(p_\lambda(x,y) \,\|\, q_\lambda(x,y)\right)$

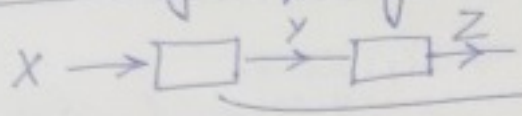
$\leq (1-\lambda) D\left(p_0(x,y) \,\|\, q_0(x,y)\right)$

$+ \lambda D\left(p_1(x,y) \,\|\, q_1(x,y)\right)$.

[Lossy Compression → Construct $p(y|x)$]

---

* Data Processing Inequality



Defn: X, Y, Z form a Markov chain (denoted by $X \to Y \to Z$) if X & Z are mutually independent given Y.

$p(x,z|y) = p(x|y)\, p(z|y)$

$p(x,y,z) = p(x)\, p(y|x)\, p(z|y)$ ∵ given y it does n't depend upon x.

$Z = g(Y)$.

$X \to Y \to g(Y)$ ← This could be a decoder

Thm: If $X \to Y \to Z$ then $I(X;Y) \geq I(X;Z)$.

with = iff $X \to Z \to Y$ (form a Markov chain)

Proof: $I(X;Y,Z) = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0 \text{ (for indep r.v. m. is 0)}}$

$= I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0}$

$\therefore I(X;Y,Z) \geq I(X;Z)$ | = holds iff $I(X;Y|Z) \geq 0$

$\Leftrightarrow X \to Z \to Y$.

$I(Y;Z) \geq I(X,Z)$.

$X \to Y \to Z$

* **Channel Coding:**

Source Code is a Code that Compress
that Code

A discrete (time) → alphabets memoryless and stationary if

$p(y_1^n | x_1^n) = \prod_{i=1}^{n} p(y_i | x_i)$

↳ what we recieve at time i is same as wh we send at time i

for each input we have p.m.f. of output. (DMSC)

Input alphabet $= \mathcal{X}$, output alphabet $= \mathcal{Y}$.

Channel characterized by: $p(y|x)$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$.

M = # of possible messages.

$\{1, 2, 3, \ldots, M\}$

A Code C_n (block of length n) is a mapping

$C_n : \{1, 2, \ldots, M\} \longrightarrow \mathcal{X}^n$

$i \longmapsto x_1^n(i)$

A decoder $\mathcal{D}_n$ is a mapping

$\mathcal{D}_n : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$.

$P_{e,m}^{(n)}$

$P_e^{(n)} = \frac{1}{M} \sum_{m=1}^{M} P(\text{error given } m \text{ was transmitted})$

* X — $[\,p(y|x)\,]$ — Y

$X \sim q(x)$

$$I(x;y) = \sum_{x,y} q(x)\, p(y|x) \times \log\left[\frac{p(y|x)}{\sum_{x'} q(x')\, p(y|x')}\right]$$

$$= E\left[\log\left[\frac{p(y|x)}{q_Y(y)}\right]\right].$$

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

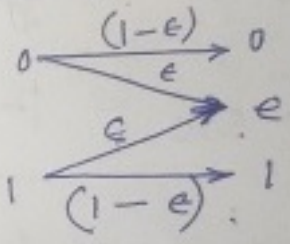
$$\leq H(X) \leq \log|\mathcal{X}|$$

↳ Concave quantity

We can find global maxima

$$C = \max_{q(x)} I(X;Y) \longrightarrow \text{mutual information capacity of the channel.}$$

e.g. ① Erasure channel:



$h(\epsilon) = -\epsilon \log \epsilon - (1-\epsilon)\log(1-\epsilon)$

$$C = \max_{q(x)} I(x;y)$$

$$I(x;y) = H(y) - \underbrace{H(y|x)}_{h(\epsilon)}.$$

$$p(x=1) = \pi, \quad H((1-\pi)(1-\epsilon), \pi(1-\epsilon), \epsilon)$$

$= 0$ otherwise

$H(Y) = H(Y, E)$

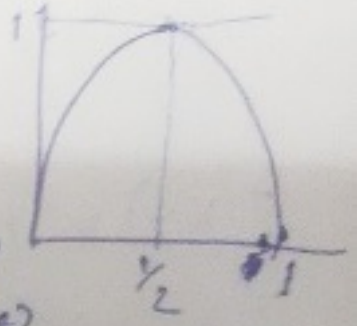
$= H(E) + H(Y|E) = h(\epsilon) + (1-\epsilon)h(\pi) + 0$

$\epsilon = 0$, $\epsilon = 1$ — No Contribution

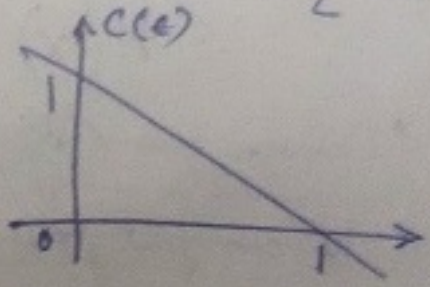
$I(X;Y) = (1-\epsilon)\, h(\pi) \leq (1-\epsilon)$

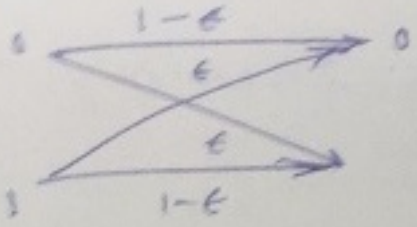
with $=$ for $\pi = \frac{1}{2}$.

$\therefore C_{erasure} = 1 - \epsilon$.



b) BSC (Binary Symmetric Channel)





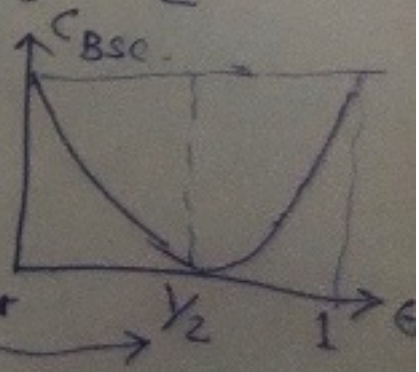
$I(X;Y) = H(Y) - H(Y|X)$, where $H(Y|X) \to h(\epsilon)$

$\leq 1 - h(\epsilon)$

with $=$ iff $P(Y=1) = P(Y=0) = 1/2$.

$C_{BSC} = 1 - h(\epsilon)$.

Input & output are becoming independent.



* Channel Coding Thm:

The rate of a code is $R_n = \frac{\log_2(M)}{n}$ bits/c.u.

Defn: A rate R is achievable if $\exists$ a sequence of codes C_n of rate R such that $P_e^{(n)} \to 0$.

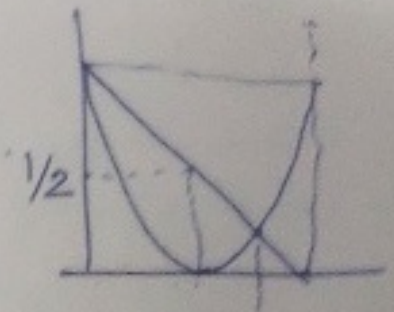
Defn: Let $C_0 := \sup\{R : R \text{ is achievable}\}$.

Thm: $C_0 = \max_{p(x)} I(X;Y) = C$.

(C_0: operation; $I(X;Y)$: depend upon statistics.)

Examples a) $C_{Erasure} = 1 - \epsilon$

b) $C_{BSC} = 1 - h(\epsilon)$



* Joint AEP:

Sequence is typical.

$$A_\epsilon^{(n)} := \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| -\tfrac{1}{n}\log\left(p(x^n)\right) - H(X) \right| < \epsilon, \right.$$

$$\left| -\tfrac{1}{n}\log\left(p(y^n)\right) - H(Y) \right| < \epsilon,$$

y's are also typical.

$$\left. \left| -\tfrac{1}{n}\log\left(p(x^n, y^n)\right) - H(X,Y) \right| < \epsilon \right\}.$$

Thm:

i) $P\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1$ as $n \to \infty$ [WLLN]

ii) $|A_\epsilon^{(n)}| \in 2^{n(H(X,Y)+\epsilon)}$

& for n large enough $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)}$

iii) If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$ (same marginals as X_n, Y_n but indep)

than for n large

$$(1-\epsilon)\, 2^{-n(I(X;Y)+3\epsilon)} \leq P\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \leq 2^{-n(I(X,Y)-3\epsilon)}$$