Information Theory & Coding

Oct 25th, 2020

Yesterday:

Channels

$$x \in \mathcal{X} \longrightarrow \boxed{} \longrightarrow Y$$

$$p(Y = y \mid \text{input} = x) = \underline{\underline{p(y \mid x)}}$$

If the channel is stationary & memoryless, & if the channel is used without feedback

then $$p(Y^n = y^n \mid X^n = x^n) = \prod_{i=1}^{n} p(y_i \mid x_i).$$

$$\Rightarrow \quad H(Y^n \mid X^n) = \sum H(Y_i \mid X_i).$$

Given a channel $P = \{p(y \mid x)\}$ we define

$$C(P) \stackrel{\Delta}{=} \max_{p_X} \underbrace{I(X;Y)}_{\text{function of } p_X \quad f(p_X, P)}$$

As a corollary:

$$\begin{tikzcd} U^k \arrow[r] & \boxed{} \arrow[r] & X^n \arrow[r] & \boxed{\text{Channel } P} \arrow[r] & Y^n \arrow[d] \\ & & & & \boxed{} \arrow[d] \\ & & & & V^k \end{tikzcd}$$

$$I(U^k; V^k) \leq I(X^n; Y^n) \leq nC(P)$$

$$\Rightarrow \quad H(U^k \mid V^k) \geq H(U^k) - nC(P).$$

Thm: Suppose $U_1 U_2 \ldots$ is a stationary source with entropy rate $\mathcal{H}$. Suppose we have a communication system that sends s source letters per channel use. Then the "symbol error probability" $\bar{p}$ will satisfy

$$h_2(\bar{p}) + \bar{p}\log(|\mathcal{U}|-1) \geq \mathcal{H} - C(P)/s.$$

Pf:

$$U^k \longrightarrow \square \longrightarrow X^n \longrightarrow \square \longrightarrow Y^n \longrightarrow \square \longrightarrow V^k$$

$k/n \to s$

① $$\frac{1}{k} H(U^k | V^k) \geq \frac{1}{k} H(U^k) - \frac{1}{k} n C$$
$$\geq \mathcal{H} - C/s$$

② By Fano's Ineq. we have

$$h_2(\bar{p}) + \bar{p}\log(|\mathcal{U}|-1) \geq \frac{1}{k} H(U^k | V^k)$$

with $\bar{p} = \frac{1}{k} \sum_{i=1}^{k} \Pr(U_i \neq V_i)$

Put ① & ② together //.

Ex: Suppose $\mathcal{U} = \{0, 1\}$ $\mathcal{H} = 0.9$ bits

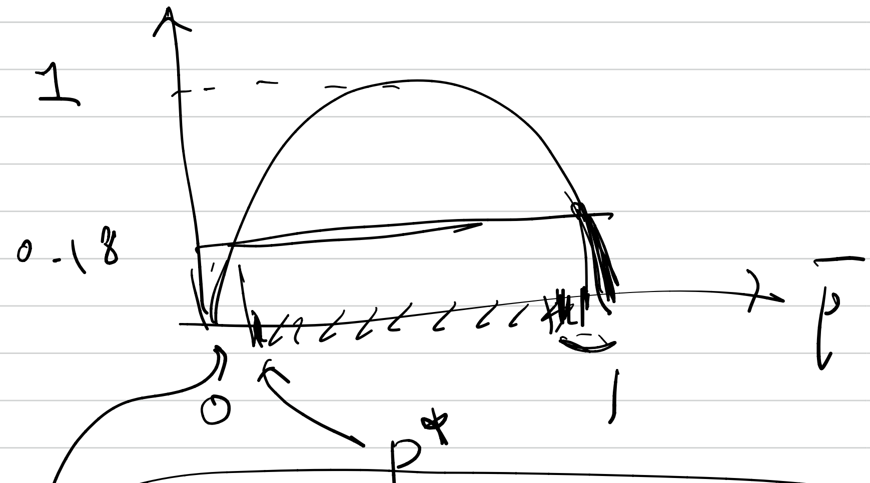
Suppose $C(P) = 0.5$

$s = 0.7$ source letter / channel use

Thm will say $\bar{p}$ will satisfy

$h_2(\bar{p}) + \cancel{\bar{p}\log(2-1)} \geq 0.9 - \frac{0.5}{0.7}$

$h_2(\bar{p}) = 0.18\ldots$



$\bar{p} < p^*$ is impossible regardless of the design of

$u^k \to \boxed{} \to x^n \ \ldots \ y^n \to \boxed{} \to v^k$

Consequence: Reliable systems (the systems that allow for small $\bar{p}$) are possible only when

$\mathcal{H} \cdot s \leq C.$

Bad news

Given a channel P; input alphabet $\mathcal{X}$
output " $\mathcal{Y}$

Def: an encoder is

a mapping: $\text{enc}: \{1, \ldots, M\} \to \mathcal{X}^n$

M: # of different messages the encoder accepts

n: block length of the encoder

$\frac{\log_2 M}{n}$: rate of the encoder bits/channel use

Def: a decoder for the same channel is a map: dec: $\mathcal{Y}^n \to \{1, .., M\}$

Let $P_{e,m} = \Pr(\text{dec}(Y^n) \neq m \mid \underbrace{X^n = \text{enc}(m)}).$

$P_e(\text{enc}, P, \text{dec}) = \frac{1}{M}\sum_{m=1}^{M} P_{e,m}$

$\hat{P}_e(\text{enc}, P, \text{dec}) = \max_m P_{e,m}.$

Def: given a channel P, a rate R is said to be achievable if $\forall\ \underline{\underline{\varepsilon > 0}}$ $\exists$ an enc, dec s.t

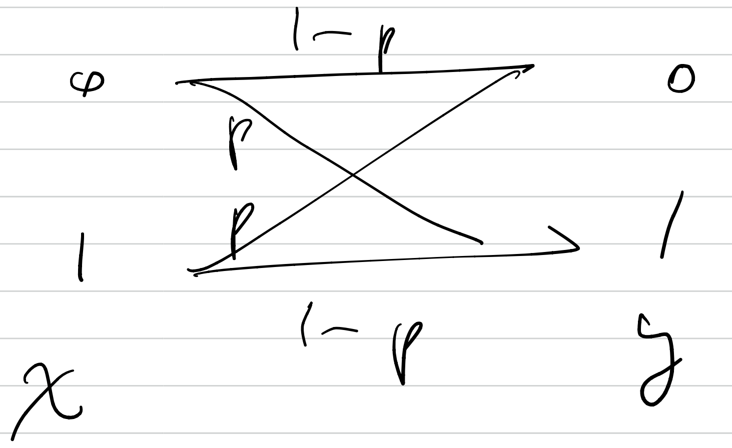
$\hat{P}_e(\text{enc}, P, \text{dec}) < \varepsilon$ &

rate(enc) $\geq R$.

Thm: given a channel P any $R < C(P)$ is achievable.

Examples for $C(P)$.

BSC(p):



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

$H(Y|X) = H(Y|X=0)\,p_X(0) + H(Y|X=1)\,p_X(1)$

$$\left.\begin{array}{l} H(Y|X=0) = h_2(p) \\ H(Y|X=1) = h_2(p) \end{array}\right\} \Rightarrow H(Y|X) = h_2(p).$$

$$\max_{p_X} I(X;Y) = \underbrace{\max_{p_X} H(Y)} - h_2(p)$$

since $\mathcal{Y} = \{0,1\}$ $\quad \underline{H(Y) \leq \log_2 |\mathcal{Y}| = 1}$

$\Rightarrow \quad C(p) \leq 1 - h_2(p)$

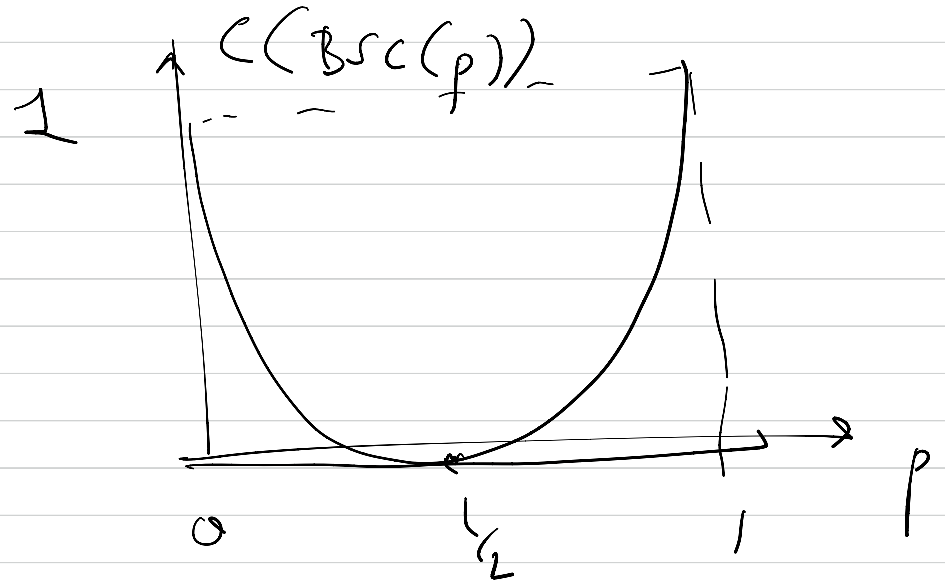
Q: can $H(Y)$ be made $= 1$ by some choice of p_X?

$\equiv$ Can $Pr(Y=0)$ be made $= \frac{1}{2}$ " " " " ?

$$Pr(Y=0) = p_X(0)(1-p) + p_X(1)\,p$$

if $\quad p_X(0) = p_X(1) = \frac{1}{2} \Rightarrow Pr(Y=0) = \frac{1}{2}$

So $\quad C(P) = 1 - h_2(p).$

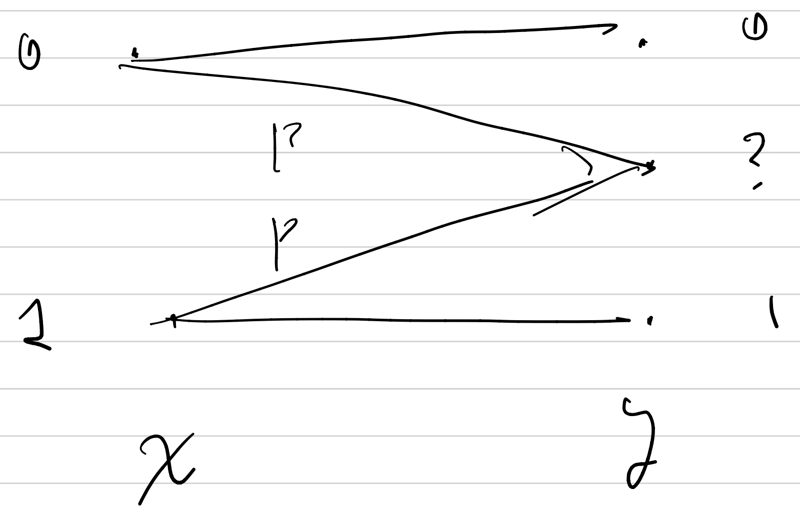


note: $\quad C(BSC(\tfrac{1}{2})) = 0$

for a $BSC(\frac{1}{2})$ $\quad Y$ is independent of X.

$$C(BSC(0)) = C(BSC(1)) = 1$$

Ex: BEC(p)



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

Observe $H(X|Y=0) = H(X|Y=1) = 0$

$$H(X|Y) = H(X|Y=?)\,\underbrace{\Pr(Y=?)}_{p} = \underline{p\,H(X|Y=?)}$$

$$\Pr(Y=?) = \Pr(Y=?|X=0)\,p_X(0) + \Pr(Y=?|X=1)\Pr(X=1)$$
$$= p\,(p_X(0) + p_X(1)) = p.$$

$$\underline{\Pr(X=0|Y=?)} = \frac{\Pr(X=0, Y=?)}{\Pr(Y=?)}$$

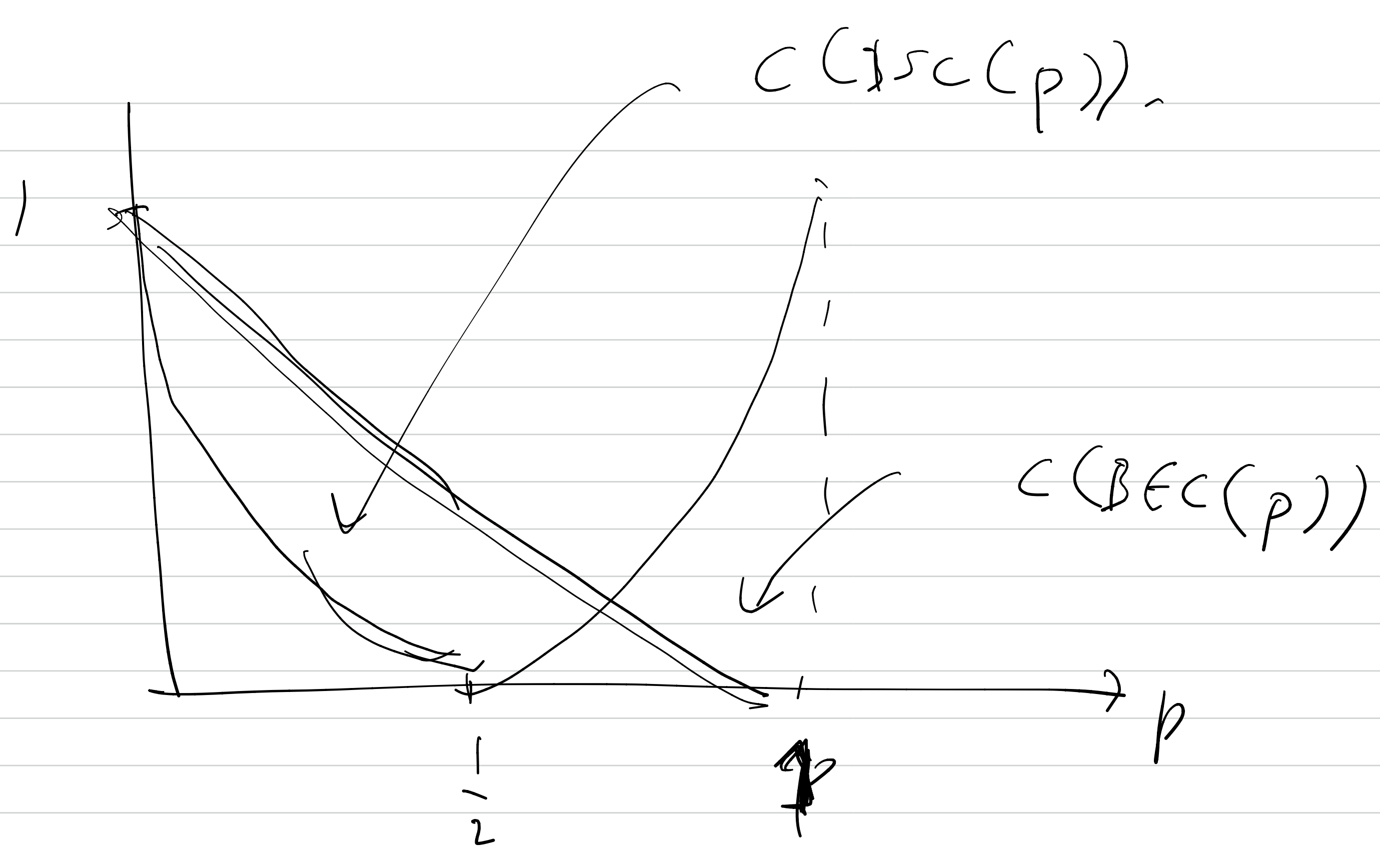
$$= \frac{\Pr(X=0)\Pr(Y=?|X=0)}{p} = \Pr(X=0)$$

$\Rightarrow$ $H(X|Y=?) = H(X)$. so

$$I(X;Y) = (1-p)\,H(X).$$

$$C(BEC(p)) = \max_{p_X} (1-p)\,H(X) = (1-p)$$

with p_X uniform on $\{0,1\}$



On the nature of how to compute $C(P)$.

$$C(P) = \max_{P_X} \underbrace{I(X;Y)}_{\text{funct}(P_X,\, P)}$$

$I(X;Y)$ as a function of P_X.

$$= H(Y) - \underbrace{H(Y|X)}_{\underbrace{\sum_x P_X(x) \underbrace{H(Y|X=x)}_{\text{funct}(P)}}_{\text{linear function of } P_X}}$$

$$\underbrace{\Pr(Y=y)}_{} = \sum_x \underbrace{P(y|x)}_{} P_X(x)$$

$\Rightarrow$ P_Y is a linear function of P_X.

$*$ $H(Y)$ is a concave function of P_Y.

Proof that $H(U)$ is a concave function of P_U.

Let $H(p_U)$ denote the entropy of a random variable U

$$\underbrace{H(\lambda p + (1-\lambda) q)}_{\text{dist on } \mathcal{U}} \geq \underline{\lambda H(p) + (1-\lambda) H(q)} \quad 0 \leq \lambda \leq 1$$

Let V be binary $\{1, 2\}$
w.p. $\lambda, 1-\lambda$

Let $\left.\begin{array}{l} \Pr(U=u \mid V=1) = p(u) \\ \Pr(U=u \mid V=2) = q(u). \end{array}\right\}$ $\Pr(U=u) = \lambda p(u) + (1-\lambda) q(u)$

$H(\lambda p + (1-\lambda) q) = H(U)$

$\left.\begin{array}{l} H(p) = H(U \mid V=1) \\ H(q) = H(U \mid V=2) \end{array}\right\}$ $\lambda H(p) + (1-\lambda) H(q) = H(U \mid V)$

Since $H(U) \geq H(U \mid V)$ we find that

$H(\lambda p + (1-\lambda) q) \geq \lambda H(p) + (1-\lambda) H(q)$

then $p \mapsto H(p)$ is $\cap$ concave.

$$I(X;Y) = H(Y) - \underbrace{H(Y \mid X)}_{\text{linear in } P_X}$$

concave funct
of a lin. funct of P_X

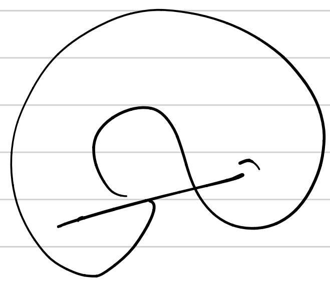
is a concave function of P_X

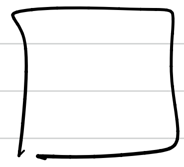
Diversion to concave / convex functions:

Recall: a set $S \subseteq \mathbb{R}^n$ is called convex if

$\forall x, y \in S$ & $0 \leq \lambda \leq 1$ we have

$$\lambda x + (1-\lambda) y \in S.$$





A function $f : S \mapsto \mathbb{R}$ is said to be convex

if $$\underline{f(\lambda x + (1-\lambda) y)} \leq \underline{\lambda f(x) + (1-\lambda) f(y)}$$
$$0 \leq \lambda \leq 1.$$

f is concave if $(-f)$ is convex.

Lemma: if $f : (a,b) \to \mathbb{R}$ is twice cont. differentiable,

and if $f'' \geq 0$ then f is convex.

given $x, y,$ $0 \leq \lambda \leq 1$ let

$$z = \lambda x + (1-\lambda) y.$$

$$f(x) = \underline{f(z) + f'(z)(x-z)} + \frac{1}{2} \overbrace{f''(\xi)} \overbrace{(x-z)^2}$$
$$f(y) = \underline{f(z) + f'(z)(y-z)} + \frac{1}{2} f''(\xi)(y-z)^2$$

$$\Rightarrow \boxed{\begin{matrix} \lambda f(x) \\ + (1-\lambda) f(y) \end{matrix} \geq f(z)} + f'(z) \underbrace{\left(\lambda x - \lambda z + (1-\lambda) y - (1-\lambda) z \right)}_{=0}$$

//