Information Theory & Coding

Oct. 19th 2020.

<u>So far:</u>

- Source Coding -
- Prefix-free, Uniq. Decodable; Kraft Sum, etc.
- Huffman Codes
- Entropy as a lower bound to the # of bits per letter
- LZ algo as a "universal" method

- $H(U)$, $H(UV)$, $H(U_1 \dots U_n)$
  $H(U|V)$, chain rules, entropy rate
  $$\mathcal{H}(\{U_i\}) = \lim_{n \to \infty} \frac{1}{n} H(U_1 \dots U_n)$$

- mutual inf. $I(U;V) = H(U) - H(U|V)$
  conditional versions, chain rules, etc..

- Data processing theorem
  $$U - V - W \Rightarrow I(U;V) \geq I(U;W)$$

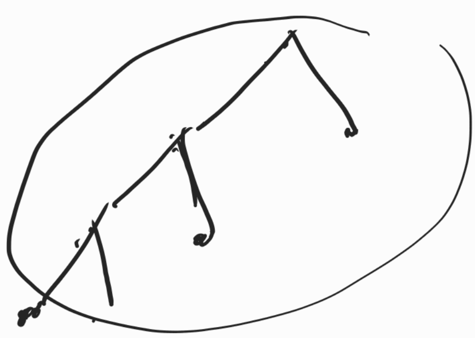
- Fano's Ineq. $u, v \in \mathcal{U}$

$$H(U|V) \le h_2(p) + p \log_2(|\mathcal{U}|-1)$$

$$p = \Pr(U \ne V), \quad h_2(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$

<u>Recall</u>: $H(U) \sim$ guessing effort to learn the value of the RV U

[Analogy between prefix-free codes & the game of 20 questions]



$$H(U|V) := \sum_{v} H(U|V=v) p_V(v)$$

$$\underbrace{H(U|V)}\le H(U).$$

$$I(U;V) = H(U) - H(U|V)$$

$$I(U;V) = D(p_{UV} \| p_U p_V)$$

Given P_{UV}, let $(\tilde{U}_i, \tilde{V}_i)$ be iid RVs

with distribution $\tilde{P}_{UV} = P_U \cdot P_V$.

($P_{UV}(u,v)$, let $P_U(u) = \sum_v P_{UV}(u,v)$

$P_V(v) = \sum_u P_{UV}(u,v)$

$\tilde{P}_{UV}(u,v) = P_U(u) P_V(v)$ is also a distribution)

$$I(U;V) = D(P_{UV} \| \tilde{P}_{UV})$$

$$\Pr\left(\left((\tilde{U}_1, \tilde{V}_1), (\tilde{U}_2, \tilde{V}_2), \ldots (\tilde{U}_n, \tilde{V}_n)\right) \in T(n, \varepsilon, P_{UV})\right)$$

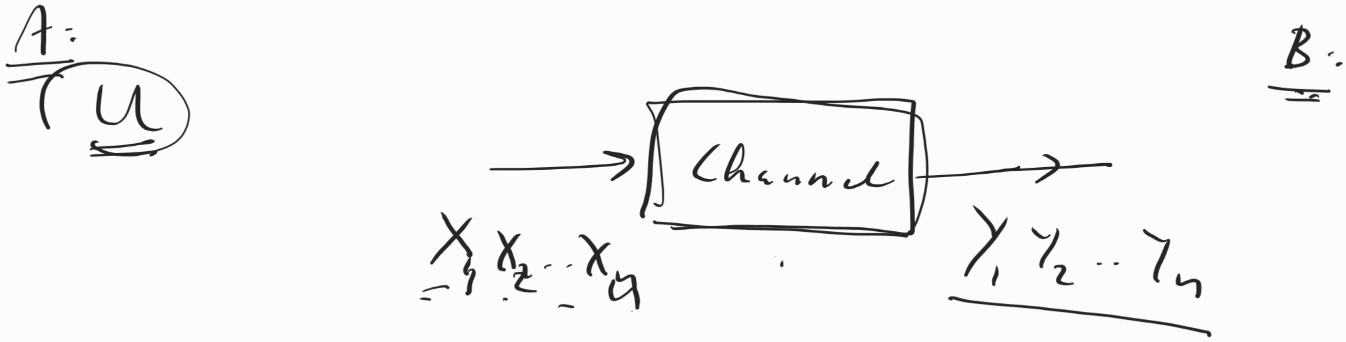
$$\approx 2^{-n(D(P \| \tilde{P}) + o(\varepsilon))}$$

$$= 2^{-n I(U;V)}$$

Data Transmission:

Setup:

A random variable U.



Channel is described by a probability kernel

$$\underset{\geq 0}{p(y|x)}, \quad \text{for each } x \in \mathcal{X}, \quad \sum_{y \in \mathcal{Y}} p(y|x) = 1$$

$p(u, x_1, \ldots x_n, y_1 \ldots y_n)$

$$= \Pr(U=u) \Pr(X_1 = x_1 | U=u) \Pr(Y_1 = y_1 | X_1 = x_1, U=u)$$
$$\Pr(X_2 = x_2 | U=u, X_1 = x_1, Y_1 = y_1) \Pr(Y_2 = y_2 | X_2 = x_2, X_1 = x_1, Y_1 = y_1, U=u)$$

$$\ldots$$

$$= P(U=u) \prod_{i=1}^{n} P(X_i = x_i | U=u, (X_1 \ldots X_{i-1}) = (x_1 \ldots x_{i-1}), (Y_1 \ldots Y_{i-1}) = (y_1 \ldots y_{i-1}))$$
$$\Pr(Y_i = y_i | \ldots\ldots, X_i = x_i)$$

always true

Def:

- A channel is called memoryless & stationary if

$$\left[\Pr\big(Y_i = y_i \mid U=u, (X_1 \ldots X_{i-1}) = x_1 \ldots x_{i-1}, (Y_1 \ldots Y_{i-1}) = y_1 \ldots y_{i-1}, X_i = x_i\big) = p(y_i|x_i). \right]$$

Def.

- A communication system is called "without feedback" if

$$\Pr\big(X_i = x_i \mid U=u, (X_1 \ldots X_{i-1}) = (x_1 \ldots x_{i-1}), (Y_1 \ldots Y_{i-1}) = (y_1 \ldots y_{i-1})\big)$$
$$= \Pr\big(X_i = x_i \mid (U, X_1 \ldots X_{i-1}) = (u, x_1 \ldots x_{i-1})\big)$$

So: if we have [a stationary, memoryless channel & system w/o feedback,] then

$$\Pr(U X^n Y^n = u x^n y^n) = p(u) \prod_{i=1}^{n} p(x_i|x^{i-1}, u)\, p(y_i|x_i)$$

$[a^n = a_1 \ldots a_n]$

$$= p(u)\, p(x^n|u)\, p(y^n|x^n)$$

where $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$

[We see that in such a system

$(U - X^n - Y^n)$ &

$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$.]

$$U \to X_1 \ldots X_n \to \square - Y_1 \ldots Y_n$$

no hidden connections $\equiv U - X^n - Y^n$

In particular, in such a system

$$H(Y^n|X^n) = E \log \frac{1}{p(Y^n|X^n)}$$

$$= E\left(\log \frac{1}{\prod_{i=1}^{n} p(Y_i|X_i)}\right)$$

$$= \sum_{i=1}^{n} E\left(\log \frac{1}{p(Y_i|X_i)}\right) = \sum_{i=1}^{n} H(Y_i|X_i)$$

Example: $\mathcal{U} = \{1, 2, 3\}$

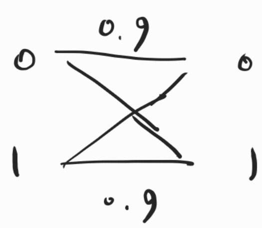
$p_u \quad \frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}$

when $u = 1$ $\quad X_1 X_2 = \begin{cases} 0\,0 & \frac{1}{2} \\ 0\,1 & \frac{1}{2} \end{cases}$

$u = 2 \quad X_1 X_2 = 0\,0$

$u = 3 \quad X_1 X_2 = 1\,1$

Channel



"BSC(0.1)".

$Pr(U=1, X^2 = 00, Y^2 = 01) = \frac{1}{3}\frac{1}{2}(0.9)(0.1)$

$Pr(U=2, X^2 = 00, Y^2 = 01) = \frac{1}{3} \cdot 1 (0.9)(0.1)$

$Pr(U=3, X^2 = 11, Y^2 = 01) =$

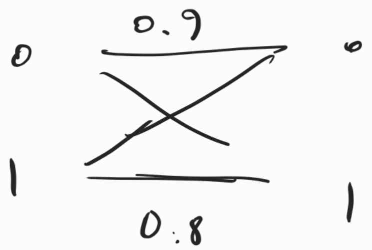
$\frac{1}{3} \cdot 1 \cdot (0.9)(0.1)$

$H(Y_1|X_1) = ?$

$$\underbrace{\sum_{x_1}^{1} \Pr(X_1 = x_1)}\ \underbrace{\sum_{y_1}^{1} \Pr(Y_1 = y_1 | X_1 = x_1) \log \frac{1}{P(Y_1 = y_1 | X_1 = x_1)}}_{h_2(0.1)}$$

$= h_2(0.1)$

If we keep U, X^2 same & change



$$H(Y_1|X_1) = \sum_{x_1}^{1} \Pr(X_1 = x_1) \underbrace{\sum_{y_1}^{1} P(Y_1 = y_1 | X_1 = x_1) \log \frac{1}{P(Y_1 = y_1 | X_1 = x_1)}}_{\begin{cases} h_2(0.1) & \text{if } x_1 = 0 \\ h_2(0.2) & \text{if } x_1 = 2 \end{cases}}$$

$$= \underbrace{\Pr(X_1 = 0)}_{2/3} h_2(0.1) + \underbrace{\Pr(X_1 = 1)}_{1/3} h_2(0.2)$$

$$H(Y_2|X_2) = \underbrace{\Pr(X_2 = 0)}_{1/2} h_2(0.1) + \underbrace{\Pr(X_2 = 1)}_{1/2} h_2(0.2)$$

$$H(Y^2|X^2) = H(Y_1|X_1) + H(Y_2|X_2) = \ldots\ldots$$

Suppose we have a communication system of the type we have above

$$\mathcal{U} = (U_1 \dots U_k)$$

$$U \longrightarrow X^n \text{ — } \boxed{\text{Channel}} \text{ — } Y^n \longrightarrow V = (V_1 \dots V_k)$$

memoryless, stationary

may

We want $\Pr(\mathcal{U} \neq V)$ small $\longrightarrow E[\max_i \mathbb{1}\{U_i \neq V_i\}]$

or we may want $\frac{1}{K}\sum_{i=1}^{k} \Pr(U_i \neq V_i) = E\left[\frac{\text{\# of wrong } i\text{'s}}{k}\right]$

$$\mathbb{1}\{U_i \neq V_i\}$$

$$\frac{1}{K}\sum_{i=1}^{k} \mathbb{1}\{U_i \neq V_i\}$$

Fano's Inequality revisited.

- $H(\mathcal{U}|V) \leq h_2(p) + p\log(|\mathcal{U}|-1)$.

we want

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

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

all u_i, V_i belong to $\mathcal{U}$.

Proof:

$$H(\mathcal{U}^k|V^k) = \sum_{i=1}^{k} H(\mathcal{U}_i|\mathcal{U}^{i-1}V^k) \quad \text{chain rule}$$

$$\le \sum_{i=1}^{k} H(\mathcal{U}_i|V_i) \quad \text{cond. red. ent.}$$

$$\le h_2(p_i) + p_i \log(|\mathcal{U}|-1) \quad p_i = \Pr(\mathcal{U}_i \neq V_i) \text{ Fano.}$$

So $\frac{1}{k} H(\mathcal{U}^k|V^k) \le \frac{1}{k}\sum_i \left[\qquad \right]$

$$= \underbrace{\frac{1}{k}\sum_{i=1}^{k} h_2(p_i)} + \bar{p}\log_2(|\mathcal{U}|-1)$$

$$\le \underbrace{h_2(\bar{p})} + \bar{p}\log_2(|\mathcal{U}|-1)$$

Concavity of $h_2(\,)$.

$h_2(p) = -p\log p - (1-p)\log(1-p)$ is concave:

$$\frac{\partial h_2(p)}{\partial p} = -\log p - \cancel{1} + \log(1-p) \cancel{+1}$$

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

$$\frac{\partial^2}{\partial p^2} h_2(p) = -\frac{1}{1-p} - \frac{1}{p} = -\frac{1}{p(1-p)} \le 0. \quad \cap \quad //$$

Given a memoryless, stationary channel $p(y|x)$

Compute: $\underline{C} = \max_{P_X} I(X;Y)$ Then in

$U^k - X^n - \boxed{Ch} \to Y^n - V^k$

$\bar{p} = \frac{1}{k} \sum^{l} Pr(U_i \neq V_i)$ satisfies

$$h_2(\bar{p}) + \bar{p} \log_2(|U|-1) \geq \left[\frac{1}{k} H(U^k) - \frac{n}{k} C\right].$$

Pf: By the Revisited Fano

$$h_2(\bar{p}) + \bar{p} \log_2(|U|-1) \geq \frac{1}{k} H(U^k|V^k)$$

$$= \frac{1}{k}\left[H(U^k) - I(U^k;V^k)\right]$$

$$\geq \frac{1}{k}\left(H(U^k) - I(X^n;Y^n)\right)$$ Data proc. $U^k - X^n - Y^n - V^k$

$$\geq \frac{1}{k}\left[H(U^k) - nC\right]$$

because $I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n)$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i)$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \quad ''$$

$$= \sum_{i=1}^{n} \underbrace{I(X_i;Y_i)}_{\leq C} \leq nC. \qquad //$$

Consequently:

if $U_1 U_2 \dots$ is a stationary source we also have

$$\underbrace{\frac{1}{k} H(U_1 \dots U_k)} = \frac{1}{k} \sum_{i=1}^{k} H(U_i | U^{i-1})$$

$$= \frac{1}{k} \left[H(U_1) + H(\underbrace{U_2 | U_1}) \dots + H(U_k | U^{k-1}) \right]$$

$$= \frac{1}{k} \left(\underbrace{H(U_k)} + \underbrace{H(U_k | U_{k-1})} + \dots + H(U_k | U^{k-1}) \right)$$

$$\geq \frac{1}{\cancel{k}} \cdot \cancel{k}\, H(U_k | U^{k-1}) \searrow$$

$$\geq \mathcal{H}(\{U_i\}).$$

$\Rightarrow$ $\bar{p}$ satisfies $\underbrace{h_2(\bar{p}) + \bar{p} \log_2(|\mathcal{U}| - 1)} \geq \underline{\mathcal{H}(\{\}) - \frac{n}{k} C}$

$\frac{n}{k}$: # of times the channel is used per transmitted source letter.