So far we talked about

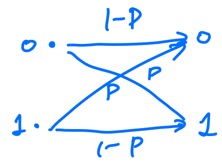
① P_X — single distribution. information source data compression

$$\boxed{P_X^n} \xrightarrow[\text{encoder}]{f} \underset{\text{Alice}}{\{0,1\}^{nR}} \xrightarrow{Id} \underset{\text{Bob}}{\{0,1\}^{nR}} \xrightarrow{\text{decoder}} \boxed{P_{\hat{X}}^n}$$

Rate: $\min R$ s.t. $\lim_{n\to\infty} P(X^n \neq \hat{X}^n) \to 0$

② A channel: $N = P_{Y|X} : \underline{X} \to \underline{Y}$

N_p



p - error probility $0<p<1$ noisy channel

$p=0$ or 1 perfect channel

$$\underset{\text{original message}}{W} \xrightarrow{f} \underset{\text{channel input}}{X} \xrightarrow{N = P_{Y|X}} \underset{\text{channel output}}{Y} \xrightarrow{g} \underset{\text{decoded message}}{\hat{W}}$$

f encoder g decoder

Example: English text $\xrightarrow{f}$ morse code (telegraph seation) $\xrightarrow{\text{telegraph}}$ morsecode (telegraph seation) $\xrightarrow{g}$ English text

Picture $\xrightarrow{f}$ binary bits (smart phone) $\xrightarrow{\text{wireless communication}}$ binary bits (computer) $\xrightarrow{g}$ picture

Of course, we want to minimize error. $\varepsilon = P(W \neq \hat{W})$

but as the size of message $\nearrow$, $\varepsilon \nearrow$.

So what is the $\boxed{\text{largese size of file}}$ that can be sent through N faithfully?

Channel capacity

Mathematical Model

(Deterministic encoder decoder):

$$W \xrightarrow{f} X \xrightarrow{N = P_{Y|X}} Y \xrightarrow{g} \hat{W}$$

$$m \longrightarrow f(m) \longrightarrow P_{Y|X=f(m)} \longrightarrow g(Y)|X=f(m)$$

$$P_{\hat{W}|W}(\hat{m}|m) = \sum_{g(y)=\hat{m}} P_{Y|X}(y|f(m))$$

$$P_W(m) = \frac{1}{M} \qquad P_{\hat{W}W} = P_{\hat{W}W}(\hat{m}, m) = \frac{1}{M} \sum_{g(y)=\hat{m},} P_{Y|X}(y|f(m))$$

(f,g) encoder-decoder $\quad P_e(f,g) = P(\hat{W} \neq W) \quad \varepsilon^*(M) = \inf_{(f,g)} P(\hat{W} \neq W)$

(Probablistic Coding):

$$W \xrightarrow{E = P_{X|W}} X \xrightarrow{N = P_{Y|X}} Y \xrightarrow{D = P_{\hat{W}|Y}} \hat{W}$$

E D classical channel

$$P_{\hat{W}W}(\hat{m}, m) = \frac{1}{M} \sum_{x,y} P_{\hat{W}|Y}(\hat{m}|y) P_{Y|X}(y|x) P_{X|W}(x|m)$$

It turns out probalistic coding is not doing better than deterministic

$$\varepsilon^*(M) = \inf_{(f,g)} P(\hat{W} \neq W) = \inf_{(E,D)} P(\hat{W} \neq W)$$

why choose $m \sim$ uniform on W.?

By source coding. $P_X^n \sim$ uniform $\{0,1\}^{nH(X)}$, so asymptotically it suffices to consider uniform distribution, whose amount information is $\log|W| = \log M$.

$$P_e \simeq \| Id_W - E \circ N \circ D \|$$

Definition: (1) An M-code for $N = P_{Y|X}$ is an encoder/decoder pair (f,g) such that

① $f: \underset{\{1, \dots, M\}}{\underset{\shortparallel}{[M]}} \longrightarrow X$

dectable error

② $g: Y \to [M]$ (or $[M] \cup \{e\}$)

② We say $(f.g)$ is an $(M.\varepsilon)$ code for $N = P_{Y|X}$ if $(f.g)$ is an M-code with

$$P_e = P(W \neq \hat{W}) \leq \varepsilon .$$

We interesed in $M^*(P_{Y|X}, \varepsilon) = \max \{ M : \exists \ (M.\varepsilon)\text{-code} \}$

$\log_2 M^*(\varepsilon)$ largese # of bits can be send through N with error $\leq \varepsilon$.

I.I.d Setting. $\dfrac{\log_2 M^*(P_{Y|X}^n, \varepsilon)}{n}$ (largest # of bits per use of channel - - - - - error $\leq \varepsilon$

<u>Definition</u> (Channel Capacity)

The Shannon capacity $C(N)$ of $N = P_{Y|X}$ is

$$C_\varepsilon(N) := \liminf_{n \to \infty} \frac{1}{n} \log M^*(n.\varepsilon) \qquad C(N) = \lim_{\varepsilon \to 0^+}$$

Theorem (Shannon's Noisy Channel Coding , 1948)

$$C = \sup_{P_X} I(X:Y) \qquad (P_{XY} = P_{Y|X} P_X)$$

Alternative angle : $\varepsilon^*(M.N) = \inf_{\substack{(f.g)\text{-}M\text{ code} \\ \text{of } N}} P(W \neq \hat{W})$

Optimal error, when sending $\log_2 M$ bits

I.I.d Setting. $\varepsilon^*(2^{nR}, N^n)$ optimal error sending nR bits over n use of N.

Theorem : $\lim_{n \to \infty} \varepsilon^*(2^{nR}. N^n) = \begin{cases} 1 & \text{if } R > I(X:Y) \quad \text{(Strong converse)} \\ 0 & \text{if } R < I(X:Y) \quad \text{(Direct coding} \end{cases}$

Recall $I(X:Y) = D(P_{XY} \| P_X \times P_Y)$ $\quad I(X:Y|Z) = I(XZ:Y) - I(Y:Z)$

$$D(P_{XYZ} \| P_{XZ} \times P_Y) - D(P_{YZ} \| P_Y \times P_Z)$$

Lemma: ① If $Y \xrightarrow{N} Z$, $\quad I(X:Y) \geq I(X:Z)$

② $I(X:Z|Y) \geq 0$, equality iff $X \to Y \to Z$

Pf: $P_{XYZ} = P_{Z|Y} P_{Y|X} P_X$ $\quad P_{XYZ}(x,y,z) = \sum_{Y}\sum_{X} P_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x)$

$$P_{XZ|Y} = \frac{\sum_X P_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x)}{P_Y(y)} = P_{Z|Y} P_{X|Y}$$

③ If $X \to Y \to Z$, $\quad I(X:Z) \leq I(Y:Z)$

$$I(XY:Z) = I(Y:Z) + \boxed{I(X:Z|Y)} \; {=0}$$

$$= I(X:Z) + I(X:Y|Z)$$

③ If $X \to Y \to Z \to W$

$$I(Y:Z) \geq I(X:Z) \geq I(X:W)$$

④ Chain Rule: $\quad I(X_1 \cdots X_n : Y) = \sum_{k=1}^{n} I(X_k : Y | X_1 \cdots X_{k-1})$

Pf: $I(X_1 \cdots X_n : Z) = I(X_n : Y | X_1 \cdots X_{n-1}) + I(X_1 \cdots X_{n-1} : Y)$

$$= \cdots\cdots \sum_{k=1}^{n} I(X_k : Y | X_1 \cdots X_{k-1})$$

Lemma: If $X_1 \to Y_1$ & $X_2 \to Y_2$,

$$I(X_1X_2 : Y_1Y_2) \leq I(X_1:Y_1) + I(X_2:Y_2)$$

Pf: $I(X_1X_2 : Y_1Y_2) - I(X_1:Y_1) - I(X_2:Y_2)$

$= H(X_1X_2) + H(Y_1Y_2) - H(X_1X_2Y_1Y_2) - H(X_1) - H(Y_1) + H(X_1Y_1)$

$\quad - H(X_2) - H(Y_2) + H(X_2Y_2)$ $\qquad X_1 \to X_1Y_1 \to X_1$

$= \boxed{I(X_2Y_2 : X_1Y_1) - I(X_1:X_2)} - I(Y_1:Y_2) = -I(Y_1:Y_2) \leq 0$

Theorem (Weak Converse by DPI)

Any M-code satisfies

$$\log M \le \frac{\sup_X I(X;Y) + h(P_e)}{1-P_e}$$

Pf: $W \xrightarrow{E} X \xrightarrow{N} Y \xrightarrow{D} \hat{W}$

$$\sup_{P_X} I(X;Y) \ge I(X;Y) \ge I(W;\hat{W}) \ge d(P(W \ne \hat{W}) \| \tfrac{1}{M})$$

$h: W\hat{W} \to \{0,1\}$ $\quad\quad \ge -h(P_e) + (1-P_e)\log M$

$$h(m,\hat{m}) = \begin{cases} 1 & m=\hat{m} \\ 0 & \text{otherwise} \end{cases}$$

Lemma: $\sup_{P_{X^n}} I(X^n;Y^n) = n \sup_{P_X} I(X;Y)$

Now Apply it for N^n

If $R > I(X;Y)$, $\exists\ \varepsilon > 0$ s.t. $(1-\varepsilon)R > \sup_{P_X} I(X;Y)$

$M = 2^{nR}$ $\quad \log 2^{nR} \le \dfrac{\sup I(X^n;Y^n) + h(\varepsilon)}{1-\varepsilon}$

$$nR \le \frac{n \sup_{P_X} I(X;Y) + h(\varepsilon)}{1-\varepsilon}$$

$n \to \infty$ $\quad\quad R \le \dfrac{\sup_{P_X} I(X;Y)}{1-\varepsilon}$ $\quad$ contradiction.

So If $R > I(X;Y)$, $\quad \lim_{n\to\infty} \varepsilon^*(2^{nR}, N^n) \ne 0$

Shannon's Achievability bound.

Theorem: Given $N = P_{Y|X}$, $\forall P_X$, $\forall \tau > 0$ $\exists (M, \varepsilon)$-code s.t.

$$\varepsilon \le P\left[\log \frac{P_{XY}}{P_X \times P_Y} \le \log M + \tau \right] + e^{-\tau}$$

where $P_{XY} = P_{Y|X} P_X$.

Denote $i(X;Y) = \log \frac{P_{XY}}{P_X \times P_Y} = \log \frac{P_{X|Y}}{P_X}$

$$\boxed{M} \xrightarrow{f} \boxed{X} \xrightarrow{N} \boxed{Y} \xrightarrow{g} \boxed{\hat{M}}$$

Pf: Define $C_m = f(m)$, code word for $m \in [M] = \{1 \cdots M\}$

We need to find good C_m and decoder g.

Define $g(y) = \begin{cases} m, & \exists!\, C_m \text{ s.t. } i(C_m; y) \geq \log M + \tau \\ e, & \text{o.w.} \end{cases}$

Interpretation: $i(C_m; y) \geq \log M + \tau \Leftrightarrow P_{X|Y}(C_m|Y) \geq M e^{\tau} P_X(C_m)$

there is a unique m s.t. the probability m being sent given y received is above certain threshold.

Given a code book $\{C_1 \cdots C_M\}$,

$$P(W = \hat{W} | W = m) = P\left(\{ i(C_m, Y) \geq \log M + \tau \} \cap \{ \nexists\, \bar{m} \neq m, \text{ s.t. } i(C_{\bar{m}}, Y) > \log M + \tau \} \right)$$

$$P_e(C_1 \cdots C_M) = P(W \neq \hat{W}) = \frac{1}{M} \sum_{m=1}^{M} P\left(\{ i(C_m, Y) < \log M + \tau \} \cup \{ \exists\, \bar{m} \neq m \text{ s.t. } i(C_{\bar{m}}, Y) > \log M + \tau \} \right)$$

$\downarrow$

W is uniform on $[M]$

Random coding: we choose $C_m \sim P_X$, i.i.d. By symmetry

$$\mathbb{E}_{C_m \sim P_X}[P_e(C_1 \cdots C_M)]$$
$$= \mathbb{E}[P_e(C_1 \cdots C_M) \mid W=1]$$
$$= P[\{i(C_1 : Y) \le \log M + \tau\} \cup \{\exists\, \bar{m} \ne 1,\ i(C_{\bar{m}}, Y) > \log M + \tau\} \mid W=1]$$
$$\le P[\{i(C_1 : Y) \le \log M + \tau \mid W=1\} + \sum_{m=2}^{M} P(i(C_m, Y) > \log M + \tau \mid W=1\}$$
$$\le P[i(X:Y) \le \log M + \tau] + (M-1)\, \mathbb{P}[i(\bar{X} : Y) > \log M + \tau]$$
$$\le P[i(X:Y) \le \log M + \tau] + (M-1)\exp(-\log M - \tau)$$
$$\le P[\ -\ —\ —\ ] + e^{-\tau}$$

where we used $\forall x$,

$$\mathbb{P}[i(x, Y) > \tau] = \mathbb{P}[\log \frac{P_{Y|X=x}}{P_Y} > \tau] \le e^{-\tau}$$

Indeed, $Q[\log \frac{P}{Q} \ge t] = \sum_{\log \frac{P(x)}{Q(x)} \ge t} Q(x) \le \sum e^{-t} P(x) \le e^{-t}$.

Since for $C_1 \cdots C_m \sim P_X$ i.i.d $\exists\, (M, \varepsilon)$ code as desired, there exists some deterministic code make this happen.

Proof of Shannon's Theorem, achievability. For $M_n = 2^{nR}$, $R < I$ $\exists\, \delta$ s.t. $R + 2\delta < I$, Choose $\tau = \delta n$

$$\varepsilon_n^*(M) \le P[i(X^n, Y^n) \le \log M_n + \tau_n] + e^{-n\delta}$$
$$= P[i(X^n : Y^n) \le nR + n\delta] + e^{-n\delta}$$
$$= P[\log \frac{P_{Y^n|X^n}}{P_{Y^n}} \le nR + n\delta]$$
$$= P[\sum_{j=1}^{n} \log \frac{P_{Y_j|X_j}}{P_{Y_j}} \le nR + n\delta] + e^{-n\delta}$$

$$= P\left[\sum_{k=1}^{n} i(X_k ; Y_k) \leq n I(X;Y) - \delta n\right] + \exp(-\delta n) \to 0$$

by W.L.L.N.

Note that the above argument holds for $\forall P_X$. $\forall \varepsilon > 0$

so $C_\varepsilon = \lim_{n\to\infty} \frac{1}{n} \log M^*(n, \varepsilon) \geq \sup_{P_X} I(X;Y) - 2\delta \quad \forall \delta > 0$