Chapter 6: Noisy channel coding

6.1 Channel mutual information

In the literature it is often called 'channel capacity'.

Def: Let $\mathcal{X}$ be a finite alphabet of input symbols.
$\mathcal{Y}$ " of output symbols.

Let ω be a conditional probability distribution, or channel, $\omega_{Y|X}(y|x)$. Then, the channel mutual information of ω is defined as

$$I(\omega) := \max_{P_X} I(X:Y)_P$$

where $P_{XY}(x,y) = P_X(x)\,\omega_{Y|X}(y|x)$.

Lemma: For fixed ω, $I(X:Y)_P$ is a concave function of P_X.

Proof:

$$I(X:Y)_P = H(Y)_P - H(Y|X)_P$$
$$= \underbrace{H(Y)_P}_{} - \underbrace{\sum_x P_X(x)\,H(Y|X=x)_\omega}_{\text{linear in } P_X}$$

concave in P_Y, thus also P_X

$$\underbrace{P_Y(y) = \sum_x P_X(x)\,\omega(y|x)}_{\text{linear in } P_X}$$

$\square$

Lemma: Consider ω such that for every permutation $\pi \in S_{|\mathcal{X}|}$ there exist $\tilde{\pi} \in S_{|\mathcal{Y}|}$ such that
$\omega(y|x) = \omega(\tilde{\pi}(y)\,|\,\pi(x))$ for all x, y.

Then, the $I(\omega)$ is achieved by the uniform distribution on inputs.

For such channel:

$$X \longleftrightarrow \pi(X) \overset{\omega}{\longleftrightarrow} \hat{\pi}^{-1}(Y) \longleftrightarrow Y \qquad (*)$$

$$\Updownarrow$$

$$X \overset{\omega}{\longleftrightarrow} Y$$

Proof: Let P_X be a maximizer. Assume $|\mathcal{X}| = d$. Then $I(X:Y)_P = I(X:Y)_{P^\pi}$ for every $\pi \in S_n$, $P_X^\pi(x) = P_X(\pi^{-1}(x))$ due to the data-processing inequality and $(*)$.

$$\text{Thus } I(\omega) = I(X:Y)_P = \sum_{\pi \in S_d} \frac{1}{|S_d|} I(X:Y)_{P^\pi}$$

$$\leq I(X:Y)_Q$$

$$\text{where } Q(x) = \sum_{\pi \in S_d} \frac{1}{|S_d|} P_X^\pi(x)$$

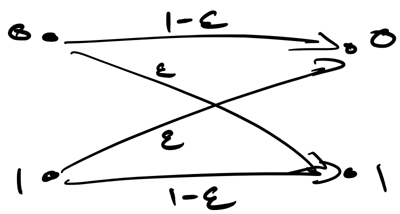
$$= \sum_{\pi \in S_d} \frac{1}{d!} P_X(\pi^{-1}(x))$$

$$= \sum_{y \in \mathcal{X}} \frac{(d-1)!}{d!} P_X(y) = \frac{1}{d}$$

$\square$

Examples

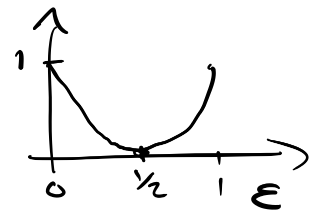
1) Binary symmetric channel (BSC)



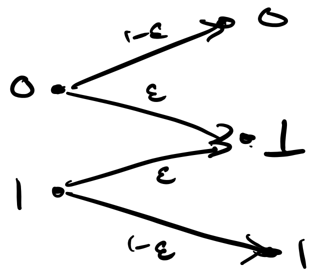
$\omega(1|0) = \varepsilon = \omega(0|1)$
$\omega(0|0) = 1-\varepsilon = \omega(1|1)$

$$I(\omega) = I(X:Y)_Q = H(Y) - \sum_{x\in\{0,1\}} \frac{1}{2} H(Y|X=x)$$

$$= 1 - h(\varepsilon)$$



2. Binary erasure channel (BEC)



$\Rightarrow$ symmetric

$I(\omega) =$

$$I(X:Y)_Q = H(X) - H(X|Y)$$

$$= 1 - \sum_y P_Y(y)\, H(X|Y=y)$$

$$= 1 - P_Y(\perp) = 1-\varepsilon$$

<u>Prop</u>: $I(\omega) = \min_{Q\in\mathcal{P}(Y)} \max_x D(\omega(\cdot|x)\|Q)$

geometrical interpretation as radius of channel image

Proof: $I(W) = \max_{P_X} D(P_{XY} \| P_X \times P_Y)$

$$= \max_{P_X} \min_{Q_Y} D(P_{XY} \| P_X \times Q_Y)$$

$$= \max_{P_X} \min_{Q_Y} \sum_x P_X(x) D(W(\cdot|x) \| Q_Y)$$

Sion's minimax
$$= \min_{Q_Y} \max_{P_X} \sum_x P_X(x) D(W(\cdot|x) \| Q_Y)$$

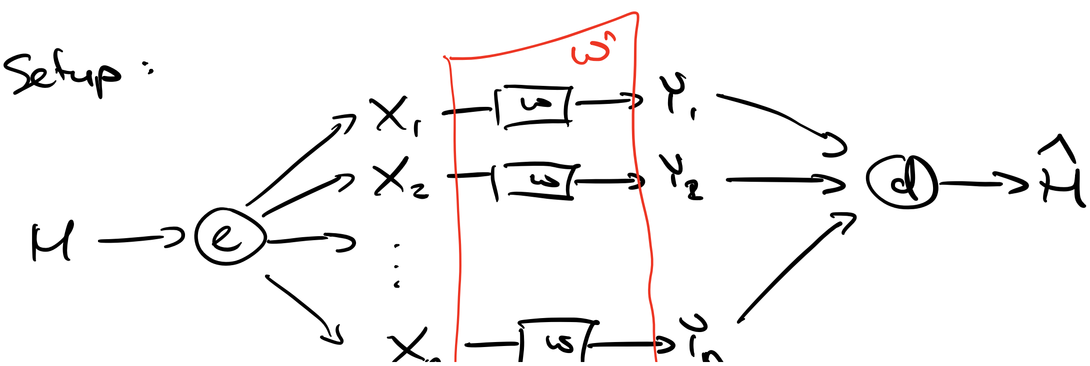
$$= \min_{Q_Y} \max_{x} D(W(\cdot|x) \| Q_Y).$$

6.2 Channel coding theorem

Def.: A discrete memoryless channel (DMC) W is fully characterized by a conditional probability distribution $W(y|x)$. For all $n \in \mathbb{N}$

$$P[Y^n = y^n | X^n = x^n] = \boxed{W^n(y^n|x^n)}$$
$$= \prod_{i=1}^{n} W(y_i|x_i)$$

Setup:



Def. An $(\varepsilon, |M|, n)$-code for W is comprised of an encoder $e : [|M|] \to \mathcal{X}^n$ and a decoder $d : \mathcal{Y}^n \to [|M|]$ so that the Markov chain

$$M \xrightarrow{e} X^n \xrightarrow{W^n} Y^n \xrightarrow{d} \hat{M} \quad \text{satisfies}$$

$$\Pr[M \neq \hat{M}] \leq \varepsilon \text{ when } M \text{ is uniform.}$$

n : block length
ε : allowed probability of error
$|M|$: number of different messages

Def. We say that a rate R is achievable on a DMC W if there exists a sequence of $(\varepsilon_n, 2^{\lceil nR \rceil}, n)$-codes for all $n \in \mathbb{N}$ such that

$$\lim_{n \to \infty} \varepsilon_n = 0.$$

The capacity of W, denoted $C(W)$, is the supremum over all achievable rates.

Theorem: For any DMC W with conditional distribution W, we have $C(W) = I(W)$.

Proof in two steps ① $C(W) \leq I(W)$ converse

② $C(W) \geq I(W)$ achievability

6.2.1 The meta-converse

$$M \longrightarrow (e) \longrightarrow X \xrightarrow{W} Y \longrightarrow (d) \longrightarrow \hat{M}$$

Prop. For $(\varepsilon, |M|, 1)$-code for W, we have

$$|M| \leq \max_{P_X} \min_{Q_Y} \frac{1}{\beta^*_\varepsilon(P_{XY} \| P_X \times Q_Y)},$$

where $P_{XY}(x,y) = P_X(x) W_{Y|X}(y|x)$ and β^*_ε is the minimal error of the second kind when the error of the first kind is bounded by ε for the hypothesis test

$$H_0 : P_{XY}$$
$$H_1 : P_X \times Q_Y$$

Proof: (Assume that e, d are deterministic and that e is injective)

Assume we have an $(\varepsilon, |M|, 1)$ code that induces a distribution P_X on the channel input. We have $P[M \neq \hat{M}] \leq \varepsilon$. Consider the test

$$A = \{(x,y) : x \neq e(d(y))\}$$

$$\alpha(A) = P_{XY}(A) = P[X \neq e(d(Y))]$$
$$= P[e(M) \neq e(\hat{M})]$$

$$\leq \mathbb{P}[\hat{M} \neq M] \leq \varepsilon$$

$$\begin{aligned}
\beta(A) &= P_X \times Q_Y(A^c) \\
&= \sum_{x,y} P_X(x) Q_Y(y) \mathbb{1}\{x = e(d(y))\} \\
&= \sum_m \frac{1}{|M|} \sum_y Q_Y(y) \mathbb{1}\{e(m) = e(d(y))\} \\
&= \frac{1}{|M|} \sum_y Q_Y(y) \underbrace{\sum_m \mathbb{1}\{e(m) = e(d(y))\}}_{\leq 1} \\
&\leq \frac{1}{|M|}
\end{aligned}$$

$$\Rightarrow \beta_\varepsilon^*(P_{XY} \| P_X \times Q_Y) \leq \frac{1}{|M|}$$

$$\Rightarrow |M| \leq \min_{Q_Y} \frac{1}{\beta_\varepsilon^*(P_{XY} \| P_X \times Q_Y)} \qquad \square$$

Recall that, for $\delta < 1-\varepsilon$

$$\log \frac{1}{\beta_\varepsilon^*(P_{XY} \| P_X \times Q_Y)} \leq D_s^{\varepsilon+\delta}(P_{XY} \| P_X \times Q_Y) + \log \frac{1}{\delta}$$

Lemma: $D_s^{\mu}(P_{XY} \| P_X \times Q_Y) \leq \max_x D_s^{\varepsilon}(W(\cdot|x) \| Q_Y)$

Proof skipped

Prop. Let W be a DMC with W and let $\varepsilon \in (0,1)$ then for any sequence of $(\varepsilon_n, 2^{\lceil nR \rceil}, n)$-codes with $\lim_{n \to \infty} \varepsilon_n < \varepsilon$ we must have $R \leq I(W)$.

(Strong converse)

Proof: for sufficiently large n, we have $\epsilon_n < \epsilon$

$$nR \le \log|M| \le \max_{P_{X^n}} \min_{Q_{Y^n}} \log \frac{1}{\beta^*(P_{X^nY^n} \| P_{X^n} \times Q_{Y^n})}$$

choose Q_{Y^n} i.i.d. with Q_Y the minimizer for

$$I(W) = \min_{Q_Y} \max_x D(W(\cdot|x) \| Q_Y)$$

$$\le \max_{P_{X^n}} D_s^{\epsilon+\delta}(P_{X^nY^n} \| P_{X^n} \times Q_Y^n) + \log\frac{1}{\delta}$$

$$\le \max_{x^n} D_s^{\epsilon+\delta}(W^n(\cdot|x^n) \| Q_Y^n) + \log\frac{1}{\delta}$$

$$= \sup\left\{R : W^n_{Y^n|X^n=x^n}\left[\log \frac{W_{Y^n|X^n=x^n}(Y^n)}{Q_Y^n(Y^n)} \le R\right] < \epsilon+\delta\right\}$$

$$E\left[\log \frac{W_{Y^n|X^n=x^n}(Y^n)}{Q_Y^n(Y^n)}\right]$$

$$= E\left[\underline{\sum_{i=1}^n \log \frac{W_{Y|X}(Y_i|x_i)}{Q_Y(Y_i)}}\right]$$

n independent random variables

$$= \sum_{i=1}^n D(W(\cdot|x_i) \| Q_Y)$$

$$= n \boxed{\sum_{x \in X} P_X^{x^n}(x) D(W(\cdot|x) \| Q_Y)} = K$$

$$\le n \max_x D\left(W(\cdot|x) \,\|\, Q_Y\right) = n\, I(W)$$

$$\mathrm{Var}\left[\log \frac{W_{Y^n|X^n=x^n}(Y^n)}{Q_Y^n(Y^n)}\right]$$

$$= \sum_{i=1}^n \underbrace{\mathrm{Var}\left[\log \frac{W_{Y|X}(Y_i|x_i)}{Q_Y(Y_i)}\right]}_{\le \sigma^2} \le n\sigma^2$$

Set $\hat{R} = n(K + \nu)$ for some $\nu > 0$.

By Chebyshev's inequality:

$$W_{Y^n|X^n=x^n}\left[\sum_{i=1}^n \log \frac{W_{Y|X}(Y_i|x_i)}{Q_Y(Y_i)} \ge \hat{R}\right] \le \frac{n\sigma^2}{\nu^2 n^2} = \frac{\sigma^2}{\nu^2 n}$$

$$W_{Y^n|X^n=x^n}\left[\sum_{i=1}^n \log \frac{W_{Y|X}(Y_i|x_i)}{Q_Y(Y_i)} < \hat{R}\right] \ge 1 - \frac{\sigma^2}{\nu^2 n} \ge \varepsilon + \delta$$

for n large enough

$$\Longrightarrow$$

$$\sup\left\{R : W_{Y^n|X^n=x^n}\left[\log \frac{W_{Y^n|X^n=x^n}(Y^n)}{Q_Y^n(Y^n)} \le R\right] < \varepsilon + \delta\right\}$$

$$\le \hat{R} = n(K+\nu) \le n\left(I(W) + \nu\right)$$

$$\Rightarrow nR \leq n(I(\omega) + \upsilon) + \log \frac{1}{\delta}$$

$$\Rightarrow R \leq I(\omega) + \upsilon + \frac{1}{n} \log \frac{1}{\delta}$$

$$\Rightarrow R \leq I(\omega)$$ since υ is arbitrarily small and $\lim_{n \to \infty} \frac{1}{n} \log \frac{1}{\delta} = 0$.