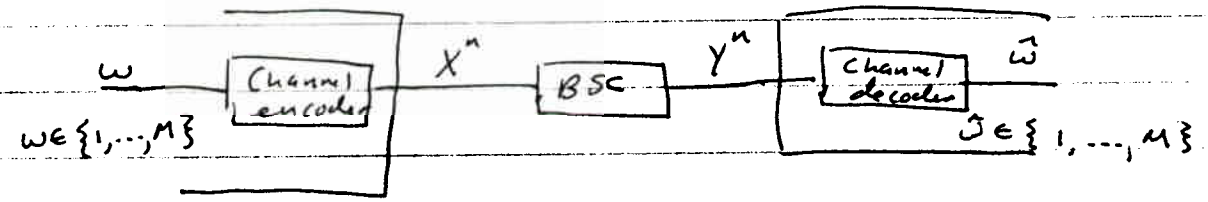
Noisy Channel Coding Thm

Theorem: Given a discrete memoryless channel w/ capacity C (in bits/channel use) a channel code can be found such that communications across the channel with $R < C$ is possible with an arbitrarily small error probability, namely $P(w \neq \hat{w}) < \varepsilon$ for any $\varepsilon > 0$

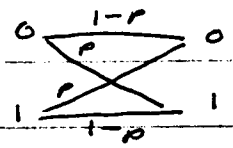
Proof: proof will be for BSC only.

Scenario:



$X^n = (X_1, \ldots, X_n)$ $\quad Y^n = (Y_1 \cdots Y_n)$

BSC:



Note: Since channel is BSC, then X^n and Y^n are binary strings of length n. That is, $X_i \in \{0,1\}$ and $Y_i \in \{0,1\}$

$$R = \frac{\log_2 M}{n} \quad \left(\frac{\text{information bits}}{\text{channel use}}\right)$$

$$\Rightarrow \quad M = 2^{nR}$$

Channel Encoder: The channel encoder is a look up table. There are M possible inputs to channel encoder and thus there are M possible channel codewords $X_1^n, X_2^n, \ldots, X_M^n$

w	X^n
1	$X_1^n = (X_{11}, X_{12}, \ldots, X_{1n})$
2	$X_2^n = (X_{21}, X_{22}, \ldots, X_{2n})$
3	↖ each X_{ij} is 0 or 1
⋮	
M	$X_M^n = (X_{M1}, X_{M2}, \ldots, X_{Mn})$

Depending on what w is, the corresponding codeword $X_i^n = (X_{i1}, \ldots X_{in})$ is transmitted across the channel.

Begin the proof: proof proceeds in 3 steps

Preview: 1) Choose a set of $M = 2^{nR}$ codewords of length n and find an upper bound on $P_e = P_r(w \neq \hat{w})$

2) (Known as Random Coding step). Compute an upper bound on average error probability when M codewords are chosen randomly with equal probability from the set of 2^n possible sequences. This random coding is done with replacement. We see that $\bar{P}_E$ can be made small provided $R < C$

3) Observe that P_E for at least one set of codewords must be less than $\bar{P}_E$.

(*) Step 1) Determine an upperbound on $P_e = Pr(w \neq \hat{w})$.

The set of M binary codewords $\{X_1^n, X_2^n, \ldots, X_M^n\}$ each of length n is used over a BSC with crossover probability p. Let z be a random variable denoting the number of bit errors made by transmitting n bits across the BSC.

z is a Bernoulli random variable (like # heads in a coin flip)

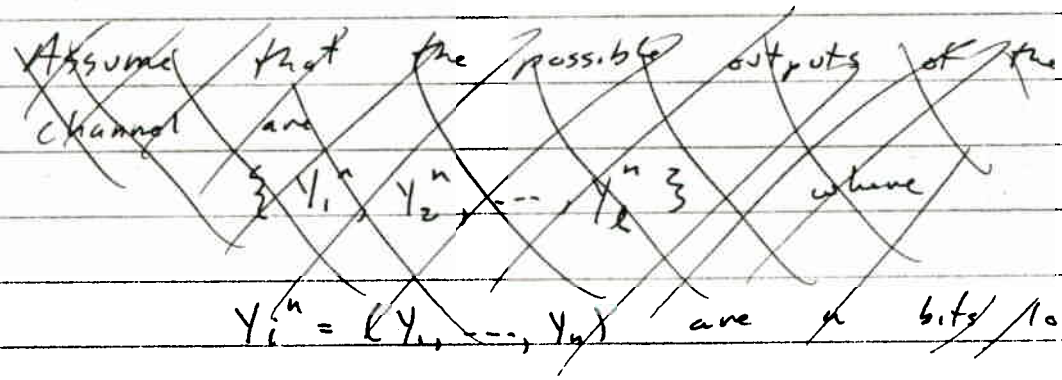
If z is # bit errors, then

$$P(z=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

and:

$$E[z] = np$$

$$Var(z) = np(1-p)$$



Then by Weak Law of Large Numbers

(Recall for a R.V. $X = \sum_{i=1}^{n} X_i$ with mean μ σ_X^2

$$Pr\left(\left|\frac{X}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma_X^2}{n^2 \varepsilon^2}$$)

we get:

$$P\left[\left|\frac{z}{n} - p\right| \geq \varepsilon\right] \leq \frac{\sigma_z^2}{n^2 \varepsilon^2} \quad \text{where } \operatorname{var} z = np(1-p)$$

$$\Rightarrow P\left[\frac{1}{n}\left|z - np\right| \geq \varepsilon\right] \leq \frac{p(1-p)}{n^2 \varepsilon^2}$$

$$\Rightarrow P\left[\left|z - np\right| \geq n\varepsilon\right] \leq \frac{p(1-p)}{n^2 \varepsilon^2}$$

if we let $\delta = \frac{p(1-p)}{n^2 \varepsilon^2}$ and rearrange we get:

$$P\left[z \geq n(p+\varepsilon)\right] \leq \delta \quad (A)$$

Alternate expression which is not used

$$P\left[z \leq n(p-\varepsilon)\right] \leq \delta$$

Probability that the # of bit errors in n channel uses is $\geq n(p+\varepsilon)$

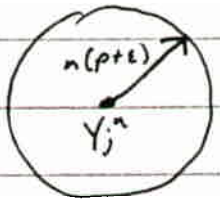
Next, consider the following decoding rule:

X_i — [BSC] — Y_j^n

Define T_j as the set of all binary sequences of length n whose Hamming distance (# of places where 2 binary sequences differ) from a received codeword Y_j^n is less than or equal to $n(p+\varepsilon)$

pictorially, for each received codeword Y_j^n we form a sphere:

T_j:



T_j = set of all sequences with a Hamming distance $n(p+\varepsilon)$ of Y_j^n

Aside

Example of T_j: let $p = .25$ and $\varepsilon = .00001$

$n = 4$ (i.e. codewords are of length 4)

and assume channel output $Y_j^4 = 0000$

$n(p+\varepsilon) = 4(.25 + .00001) = 1 + .00004$

What is T_j? Those binary sequences that have Hamming distance ≤ 1.00004 from $Y_j^4 = 0000$

$\therefore T_j =$ 0000, 0001, 0010, 0100, 1000

Decoding rule for Y_j^n: Given that Y_j^n is received, we decide that codeword X_k^n was sent if and only if it is the only codeword in T_j. If there are no valid codewords in T_j or more than 1 codeword in T_j, then we declare an error (i.e. $w \neq \hat{w}$).

Using this decoding rule, we next find $P(\text{error} \mid w_i \text{ sent}) = P[E \mid w_i]$.

It follows that given message w_i was sent (i.e. codeword X_i^n was transmitted) then a error occurs (namely $w \neq \hat{w}$) if and only if either

(1) X_i^n does not lie in test set T_j

or

(2) some other codeword X_ℓ^n lies in T_j

$$\therefore \quad P[E|w_i] = \underbrace{P[z \geq n(\rho+\varepsilon)]}_{\text{case (1) above}} + \underbrace{\sum_{\substack{\ell=1 \\ i\neq\ell}}^{m} P[X_\ell^n \in T_j]}_{\text{case (2) above}}$$

using result (A)

$$P[E|w_i] \leq \delta + \sum_{\substack{\ell=1 \\ \ell\neq i}}^{m} P[X_\ell^n \in T_j] \qquad (C)$$

This is an upper bound on $P[E|w_i]$.

Step 2) Random Coding Step.

Next we compute $\overline{P[E|w_i]}$, the average error probability over all possible codeword sets of length n. The random selection is done with replacement.

There are $(2^n)^m$ possible code word sets, each equally likely

Therefore, given that X_i^n is transmitted and Y_j^n is received, the probability that X_ℓ^n ($\ell \neq i$ belongs to T_j (test set for Y_j^n)

Therefore, given that X_i^n is transmitted and Y_j^n is received, the probability that X_ℓ^n ($\ell \neq i$) belongs to T_j (test set for Y_j^n) is:

$$P_r\left[X_\ell^n \in T_j\right] = \frac{\#\text{ sequences in } T_j}{\text{total } \#\text{ of sequences of length } n}$$

(namely, it's the chance another codeword was picked by chance to be in T_j).

$$= \frac{\#\text{ possible ways of choosing a codeword within hamming distance} \leq n(p+\varepsilon) \text{ from } Y_i}{\text{total } \#\text{ of sequences of length } n}$$

$$= \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n} \qquad (D)$$

Combining (C) & (D), the error probability given w_i was sent (using random coding)

$$P\left[E \mid w_i\right] \leq \delta + \sum_{\substack{\ell=1 \\ \ell \neq i}}^{m} P_r\left[X_\ell^n \in T_j\right]$$

$$= \delta + \sum_{\ell=1}^{m} \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n}$$

$$= \delta + \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n} \sum_{\substack{\ell=1 \\ \ell \neq i}}^{m} 1$$

$$= \delta + (m-1) \sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}$$

$$= \delta + (M-1) \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n}$$

$$P(E|w_i) \leq \delta + M \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n}$$

$$\therefore \bar{P}_E = \sum_{i=1}^{M} P[E|w_i] P(w_i)$$

$$\leq \sum_{i=1}^{M} \left(\delta + \frac{M \sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n} \right) P(w_i)$$

$$= \delta + \frac{M \sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n} \sum_{i=1}^{M} P(w_i)$$

$$\Rightarrow \bar{P}_E \leq \delta + M \frac{\sum_{k=0}^{n(p+\varepsilon)} \binom{n}{k}}{2^n} \quad (E)$$

Using a well-known inequality (see me for reference if you're interested)

$$\sum_{k=0}^{N\alpha} \binom{N}{k} < 2^{N H(\alpha)} \quad \text{of } \alpha < \tfrac{1}{2}$$

$\therefore$ (E) becomes

$$\bar{P}_E \leq \delta + M \, 2^{-n[1-H(p+\varepsilon)]}$$

$$= \delta + 2^{nR} \, 2^{-n[1-H(p+\varepsilon)]}$$

$$\therefore P_e \leq \delta + 2^{-n[1 - H(p+\varepsilon) - R]} \quad (F)$$

Recall $C = 1 - H_{(p)}$

$$\delta = \frac{p(1-p)}{n\varepsilon^2}$$

and ε very small

Then $\bar{P}_e$ can be made arbitrarily small by letting $n \to \infty$ provided $R < C$

(F) becomes:

$$\bar{P}_e \leq \delta^{\to 0} + \underbrace{2^{-n[C-R]}}_{\text{get small provided } R<C \text{ as } n \to \infty}$$

Step 3) Since $\bar{P}_e$ can be made arbitrarily small using random coding, then there must exist one of those codes such that $P_e < \bar{P}_e$ // QED

Therefore, we have shown for a BSC w/ crossover probability p, that if $R < C$ then there exists a code with rate R s.t. $P_E = P(w \neq \hat{w}) < \varepsilon$ for any $\varepsilon > 0$.