$$H(X|Y) = p(Y=0)H(X|Y=0) + p(Y=e)H(X|Y=e) + p(Y=1)H(X|Y=1)$$

since $p(X=0|Y=0)=1$ and $p(X=1|Y=1)=1$

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

So:

$$H(X|Y) = \varepsilon H(X)$$

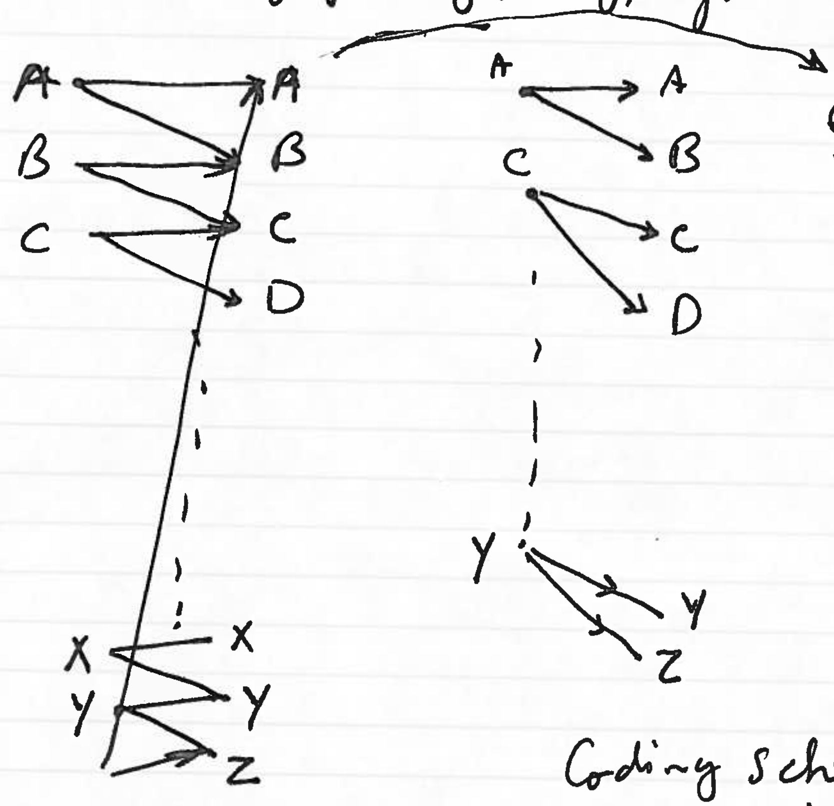
So:

$$C = \max_{p(x)} [1-\varepsilon] H(X) = (1-\varepsilon)\max_{p(x)} H(X) = 1-\varepsilon$$

X Lecture 5, Sept. 30, 2003.

Channel Coding Theorem and its Converse

Example of faulty (noisy) type writer:

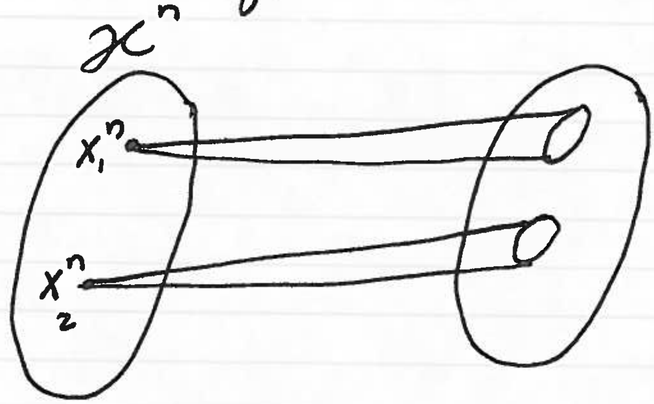


e.g. $p(Y=A|X=A) = \frac{1}{2} = p(Y=B|X=A)$

Coding scheme to achieve the capacity of $\log_2 13$

$$\boxed{C = \max I(X;Y) = \max H(Y) - H(Y|X) = \log_2 26 - 1 = \log_2 13}$$

The whole point (and the basis) of channel coding ~~the~~ is to pick a subset of channel inputs sequences that ~~have~~ result in disjoint sequences at the output. The encoding we described above for the noisy type writer shows one such affort.



channel after n uses.

For large n, for each n-sequence, there are $\approx 2^{nH(Y|X)}$ possible Y sequences. But there are $\approx 2^{nH(Y)}$ Y sequences in total. So, in order for different input n-sequences to be identifiable, we ~~need to~~ can at most have $\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(X;Y)}$

So, we can at most have $\approx 2^{nI(X;Y)}$ sequences at the input.

Next, we try to substantiate this intuitive result.

Definition: A discrete memoryless channel (DMC), ~~to~~ denote by $(\mathcal{X}, p(y|x), \mathcal{Y})$, consists of an input alphabet $\mathcal{X}$, and output alphabet $\mathcal{Y}$ and probability mass function $p(y|x)$, one for each $x \in \mathcal{X}$ such that $\forall x, y$, $p(y|x) \geq 0$ and for $\forall x$, $\sum_{x} p(y|x) = 1$.

Definition: The n-th extension of the DMC is the channel $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$, where,

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k), \quad k = 1, 2, \ldots, n.$$

If the channel does not have feedback, i.e., if $p(x_k|x^{k-1}, y^{k-1}) = p(x_k|x^{k-1})$ then

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

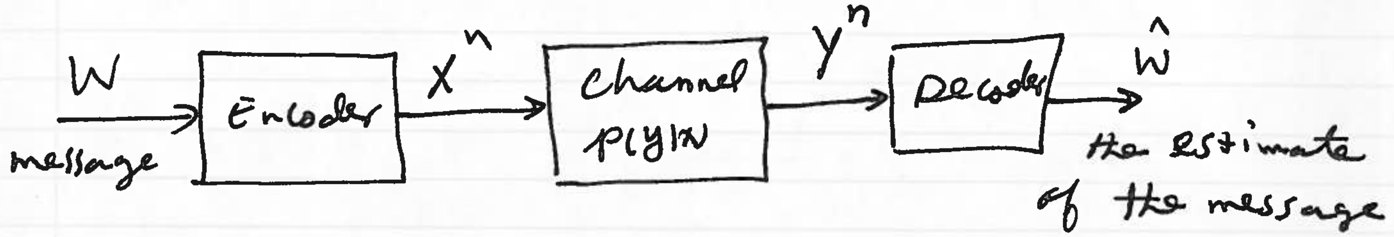
Definition: An (M, n) code ~~for~~ the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of:

1) An index set $\{1, 2, \ldots, M\}$
2) An encoding function $X^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$ yielding codewords $X^n(1), X^n(2), \ldots, X^n(M)$. The set of codewords is called the codebook.

3) A decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \dots, M\}$$

which is ~~the~~ a deterministic rule which assigns a symbol to each possible received vector.



<u>Definition</u>: (Probability of error)

$$\lambda_i = \Pr(g(Y^n) \neq i \mid X^n = x^n(i))$$

$$= \sum_{y^n} p(y^n \mid x^n(i))\, I(g(y^n) \neq i)$$

is the conditional probability of error given the index i was sent. $I(\cdot)$ is the indicator function.

<u>Definition</u>: The <u>maximal probability of error</u> $\lambda^{(n)}$ for an (M, n) code is,

$$\lambda^{(n)} = \max_{i \in \{1, \dots, M\}} \lambda_i$$

Definition: The average probability of error, $P_e^{(n)}$ is,

$$P_e^{(n)} = \frac{1}{M}\sum_{i=1}^{M} \lambda_i = \Pr(I \neq g(Y^n))$$

where I is a random variable uniformly distributed on the set $\{1, 2, \ldots, M\}$.

Definition: The rate R of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bits/transmission}$$

Definition: A rate R is said to be achievable if there ~~is~~ exists a sequence of $(\lceil 2^{nR} \rceil, n)^*$ codes such that the maximal probability of error ~~ton~~ $\lambda^{(n)}$ tends to zero as $n \to \infty$.

~~The Capa~~

Definition: The capacity of a DMC is the supremum of all achievable rates.

---

* In practice, we do not have to use $\lceil . \rceil$ function, and can identify the code as a $(2^{nR}, n)$ code.

Jointly typical sequences

Definition: The set $A_\varepsilon^{(n)}$ of jointly typical sequences $\{(x^n, y^n)\}$ w.r.t. the distribution $p(x,y)$ is the set of n-sequences with empirical entropies ε-close to the true entropies, i.e.,

$$A_\varepsilon^{(n)} = \{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n :$$

$$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \varepsilon,$$

$$\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \varepsilon,$$

$$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| < \varepsilon \}.$$

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

Theorem (Joint AEP): Let (X^n, Y^n) be sequences of length n drawn i.i.d. according to $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$. Then,

1) $\Pr\left((X^n, Y^n) \in A_\varepsilon^{(n)} \right) \to 1$ as $n \to \infty$

2) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$ and $|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)}$

3) If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, i.e., if $\tilde{X}^n$ and $\tilde{Y}^n$ are independent with the same marginals as X^n and Y^n ~~$p(x^n, y^n)$~~), then

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \leq 2^{-n(I(X;Y) - 3\varepsilon)}$$

~~Also~~ Also for sufficiently large n,

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \geq (1-\varepsilon)\, 2^{-n(I(X;Y) + 3\varepsilon)}$$

<u>Proof</u>: By the weak law of large numbers

$$-\frac{1}{n} \log p(X^n) \to -E[\log p(X)] = H(X) \text{ in probability}$$

So, given $\varepsilon > 0$, there exists n_1, such that $\forall n > n_1$

$$\Pr\left(\left|-\frac{1}{n} \log p(X^n) - H(X)\right| > \varepsilon\right) < \frac{\varepsilon}{3}$$

Similarly,

$$-\frac{1}{n} \log p(Y^n) \to -E[\log p(Y)] = H(Y) \text{ in probability}$$

and

$$-\frac{1}{n} \log p(X^n, Y^n) \to -E[\log p(X,Y)] = H(X,Y) \text{ in prob.}$$

and there exists n_2 and n_3 such that $\forall n \geq n_2$

$$\Pr\left(\left|-\frac{1}{n} \log p(Y^n) - H(Y)\right| > \varepsilon\right) < \frac{\varepsilon}{3}$$

and $\forall n \geq n_3$

$$\Pr\left(\left|-\frac{1}{n} \log p(X^n, Y^n) - H(X,Y)\right| > \varepsilon\right) < \frac{\varepsilon}{3}$$

for $n > \max(n_1, n_2, n_3)$

$$\Pr((X^n, Y^n) \in A_\varepsilon^{(n)}) \leq \Pr(|-\frac{1}{n}\log p(x^n) - H(X)|)$$
$$+ \Pr(|-\frac{1}{n}\log p(y^n) - H(Y)|)$$
$$+ \Pr(|-\frac{1}{n}\log p(x^n, y^n) - H(X,Y)|)$$
$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

Part 2:

$$1 = \sum p(x^n, y^n) \geq \sum_{A_\varepsilon^{(n)}} p(x^n, y^n)$$
$$\geq |A_\varepsilon^{(n)}| \, 2^{-n(H(X,Y)+\varepsilon)}$$

So,

$$A_\varepsilon^{(n)} \leq 2^{n(H(X,Y)+\varepsilon)}$$

For sufficiently large n,

$$1 - \varepsilon \leq \sum_{(x^n, y^n) \in A_\varepsilon^n} p(x^n, y^n) \leq |A_\varepsilon^n| \, 2^{-n(H(X,Y)-\varepsilon)}$$

So:

$$|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)}$$

Part 3:

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) = \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n) p(y^n)$$

$$\leq 2^{n(H(X,Y)+\varepsilon)} \cdot 2^{-n(H(X)-\varepsilon)} \, 2^{-n(H(Y)-\varepsilon)}$$

$$= 2^{-n(I(X;Y)-3\varepsilon)}$$

Since $I(X;Y) = H(X) + H(Y) - H(X,Y)$.

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^n) = \sum_{A_\varepsilon^{(n)}} p(x^n) p(y^n)$$

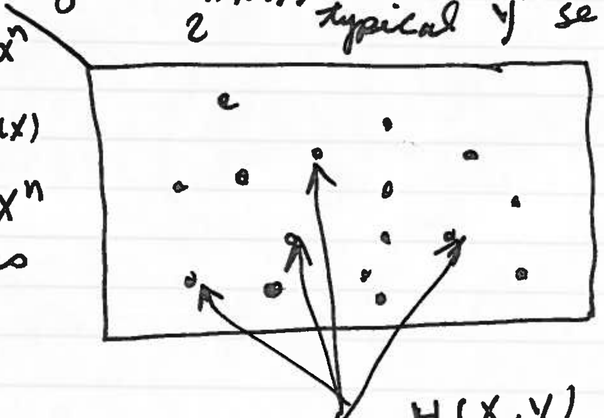
$$\geq (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)} \, 2^{-n(H(X)+\varepsilon)} \cdot 2^{-n(H(X)+\varepsilon}$$

$$= (1-\varepsilon) 2^{-n(I(X;Y)+3\varepsilon)}$$

y^n $\quad 2^{nH(Y)}$ typical y^n sequences

x^n

$2^{nH(X)}$ typical x^n sequences



Chance of randomly picking a jointly typical pair is

$$\frac{2^{nH(X,Y)}}{2^{nH(X)} \cdot 2^{nH(Y)}} = 2^{-nI(X;Y)}$$

$2^{nH(X,Y)}$ jointly typical (x^n, y^n) sequences

So, there are

$$\frac{2^{nH(Y)} \times 2^{nH(X)}}{2^{nH(X,Y)}} = 2^{nI(X;Y)}$$

distinguishable X^n's.