Random binning

Recall that in proving the source coding thm, we indexed all elements of the typical set and didn't really worry about atypical sequences (or gave them very long codewords).

Here we index everything but reject atypical sequences later.

Procedure: for each sequence X_1^n, draw an index at random from $\{1, 2, \dots, 2^{nR}\}$.

The set of sequences with the same index form a bin since think of throwing X_1^n into bins at random.

To decode the source from the bin index, look for a typical X_1^n sequence in bin. If one and only one typical X^n sequence in bin, declare that as $\hat{X}_1^n$. Otherwise declare error.

This is a source code.

To analyze performance, divide X_1^n sequences into typical and atypical.

1. if source sequence is typical, its bin will contain at least one typical sequence (itself) so error only if more than one typical sequence in bin.

2. if source sequence atypical, always error

(if # bins much larger than # typical sequences, then 1 not so common) formally, let $f(X^n)$ be bin corresponding to X^n and $g(\cdot)$ its decoding.

Probability of error, averaged over random choice of codes is

$$\Pr\left[g(f(X^n)) \neq X^n\right] \leq \Pr\left[X^n \notin A_\varepsilon^{(n)}\right] + \sum_{x^n \in \mathcal{X}^n} \Pr\left[\exists\, x' \neq x : x' \in A_\varepsilon^{(n)},\ f(x') = f(x)\right] \cdot p(x_i^n)$$

$$\leq \varepsilon + \sum_{x} \sum_{\substack{x' \in A_\varepsilon^{(n)} \\ x' \neq x}} \Pr\left[f(x') = f(x)\right] p(x).$$ (notating x_i^n as x, x_i^n as x')

$$\leq \varepsilon + \sum_{x} \sum_{x' \in A_\varepsilon^{(n)}} 2^{-nR} p(x)$$

$$= \varepsilon + \sum_{x' \in A_\varepsilon^{(n)}} 2^{-nR} \sum_{x} p(x)$$

$$\leq \varepsilon + \sum_{x' \in A_\varepsilon^{(n)}} 2^{-nR}$$

$$\leq \varepsilon + 2^{n(H(X)+\varepsilon)} 2^{-nR}$$ and if $R > H(X) + \varepsilon$ and n large

$$\leq 2\varepsilon$$

so if rate of code is greater than entropy, P_e is arbitrarily small.

$\Rightarrow$ a different source coding construction than before, but this also works

Note that the binning scheme doesn't require explicit characterization of typical set at encoder; needed only at decoder.

This property super important to ensure things work in distributed setting

Lemma on typicality for Markov chains.

A triplet of sequences x^n, y^n, z^n is said to be ε-strongly typical if

$$\left| \frac{1}{n} N(a,b,c \mid x^n, y^n, z^n) - p(a,b,c) \right| < \frac{\varepsilon}{|\mathcal{X}||\mathcal{Y}||\mathcal{Z}|}.$$

this implies (x^n, y^n) and (y^n, z^n) are also jointly strongly typical,

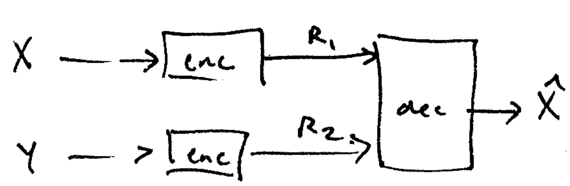
but if $(x^n, y^n) \in A_\varepsilon^{*(n)}(X,Y)$ and $(y^n, z^n) \in A_\varepsilon^{*(n)}(Y,Z)$, doesn't imply

$(x^n, y^n, z^n) \in A_\varepsilon^{*(n)}(X,Y,Z)$, in general.

But if $X \to Y \to Z$, forms a Markov chain, this is true.

Lemma: Let (X,Y,Z) form Markov chain $X \to Y \to Z$, i.e. $p(x,y,z) = p(x,y)\, p(z|y)$. If for a given $(y^n, z^n) \in A_\varepsilon^{*(n)}(Y,Z)$ and if ~~then~~ X^n drawn $\sim \prod_{i=1}^{n} p(x_i|y_i)$, then $Pr\left[(X^n, y^n, z^n) \in A_\varepsilon^{*(n)}(X,Y,Z)\right] > 1-\varepsilon$ for sufficiently large n.

Now back to achievability of WAK theorem.



$R_1 \geq H(X|U)$

$R_2 \geq I(Y;U)$ for some $p(x,y)\, p(u|y)$.

i.e. $X \to Y \to U$.

① Fix $p(u|y)$ for a specific auxilliary random variable U.

calculate $p(u) = \sum_y p(y) p(u|y)$.

② generate codebooks :

generate 2^{nR_2} independent codewords of length n

according to $\prod_{i=1}^{n} p(u_i)$ that are mapped as $U(w_2)$, $w_2 \in \{1, 2, \dots, 2^{nR_2}\}$.

randomly bin all X^n sequences into 2^{nR_1} bins by independently

generating an index b distributed uniformly on $\{1, 2, \dots, 2^{nR_1}\}$

for each X^n. Let $B(i)$ denote set of X^n sequences allotted to bin i.

③ encoding

The X sender sends the index i of the bin X^n falls into.

The Y sender looks for index s such that $(Y^n, U^n(s))$

$\in A_\varepsilon^{*(n)}(Y, U)$. $\begin{cases} \text{If more than one such } s, \text{ send the smallest.} \\ \text{If no such } U^n(s) \text{ in codebook, send } s = 1 \end{cases}$

④ decoding

The decoder looks for a unique $X^n \in B(i)$ such that

$(X^n, U^n(s)) \in A_\varepsilon^{*(n)}(X, U)$. If nothing typical or more than one typical, declare error.

Analyze 4 ~~sources~~ causes of error.

① Pair (X^n, Y^n) from the source is atypical

small probability for sufficiently large n.

② Sequence Y^n is typical but there is no $U^n(s)$ in the codebook that is jointly typical with it. From channel coding arguments or rate-distortion arguments, if there are enough codewords

$$R_2 > I(Y;U)$$

then this is quite unlikely for large n.

③ the codeword $U^n(s)$ is jointly typical with y^n but not with x^n

By the Markov chain $X \to Y \to U$ and the lemma, this has small probability for large n

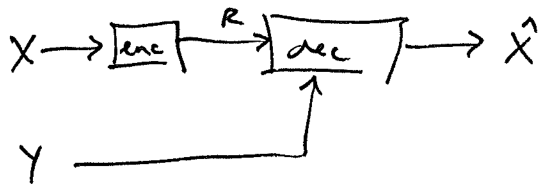
④ there is another typical $X^n \in B(i)$ which is jointly typical with $U^n(s)$. Probability that any other X^n is JT with $U^n(s)$ is less than $2^{-n(I(U;X)-3\epsilon)}$ and so

$$\left| B(i) \cap A_\epsilon^{*(n)}(X) \right| 2^{-n(I(X;U)-3\epsilon)} \leq 2^{n(H(X)+\epsilon)} 2^{-nR_1} 2^{-n(I(X;U)-3\epsilon)}$$

which goes to zero if $R_1 > H(X|U)$

so we have controlled all the error events. QED.

Wyner-Ziv problem



$$E[d(X,\hat{X})] = D.$$

Rate-distortion function with side information $R_Y(D)$ is minimum rate needed to achieve distortion D if side information Y available at decoder.

Thm: Let (X,Y) be drawn iid $\sim p(x,y)$ and let $d(x^n,\hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i)$ be given. The rate-distortion with side information

$$R_Y(D) = \min_{p(w|x)} \min_{f} \left(I(X;W) - I(Y;W) \right)$$

where minimization is over all functions $f: \mathcal{Y} \times \mathcal{W} \to \hat{\mathcal{X}}$ and conditional pmf $p(w|x)$, $|\mathcal{W}| \le |\mathcal{X}|+1$ s.t.

$$\sum_{x}\sum_{w}\sum_{y} p(x,y)\,p(w|x)\,d(x,f(y,w)) \le D.$$