X Lecture 10, ~~Oct.~~ Nov. 4, 2003

Data compression subject to a fidelity criterion

~~(Rate Distortion Theory)~~

Quantization:

- Scalar Quantization
- Vector Quantization

Definition: A fidelity criterion or a distortion measure is a function $d(\cdot,\cdot)$ from $\mathcal{X}\times\hat{\mathcal{X}}$ to R^+, where $\mathcal{X}$ and $\hat{\mathcal{X}}$ are the source and reproduction alphabets respectively.

examples: Hamming (Prob. of error) distance

$$d(x,\hat{x}) = \begin{cases} 0 & x=\hat{x} \\ 1 & x\neq x' \end{cases}$$

$$E[d(x,\hat{x})] = P(X\neq\hat{X}) = P_r(\mathcal{E})$$

- squared error distortion

$$d(x,\hat{x}) = (x-\hat{x})^2$$

- absolute error distortion

$$d(x,\hat{x}) = |x-\hat{x}|$$

<u>Definition</u>: A single letter fidelity criterion is a distortion measure from $\mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}^+$ such that:

$$d(X^n, \hat{X}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)$$

example of a fidelity criterion that is not single letter is ℓ^∞ distortion measure:

$$d(X^n, \hat{X}^n) = \max |X_i - \hat{X}_i|$$

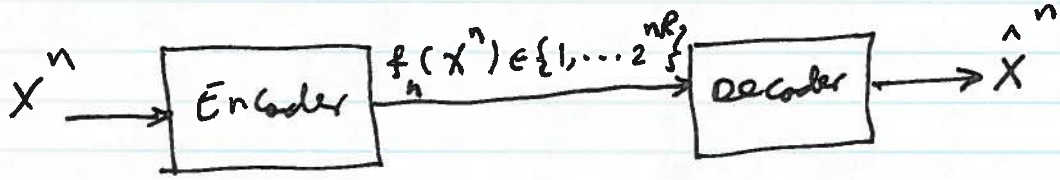
A source encoding (data Compression) system:
(a $(2^{nR}, n)$ Code)
Consists of:

1) Encoder

$$f_n(X^n) : X^n \longrightarrow \{1, 2, \ldots, 2^{nR}\}$$

2) Decoder

$$g_n(\cdot) : \{1, 2, \ldots, 2^{nR}\} \to \hat{\mathcal{X}}^n$$



The distortion associated with the $(2^{nR}, n)$ Code is

$$D = E\, d(X^n, g_n(f_n(X^n)))$$

where

$$D = \sum_{x^n} p(x^n)\, d(x^n, g_n(f_n(x^n)))$$

Definition: A rate distortion pair (R, D) is said to be achievable if there exists a ~~code~~ sequence $(2^{nR}, n)$ of data compression Codes (f_n, g_n) with $\lim E\, d(X^n, g_n(X^n, f_n(X^n))) \leq D$.

Definition: The Rate-Distortion, $R(D)$, is the infimum of rates R such that (R, D) is achievable for a given D.

Definition: The information rate-distortion function $R^I(D)$ for a source X with distortion measure $d(x, \hat{x})$ is:

$$R^I(D) = \min_{p(\hat{x}|x) : \sum_{x,\hat{x}} p(x) p(\hat{x}|x) d(x,\hat{x}) \leq D} I(X; \hat{X})$$

The source coding Theorem and its converse shows ~~that the rate-Distortion function $R(D)$ and~~ the information rate-distortion function $R^I(D)$ are the same. So, we will use the term rate-distortion function for both and denote them both by $R(D)$. Before trying to prove source coding theorem, we discuss ~~the way~~ how to find $R(D)$.

<u>Example:</u> Memoryless Binary Source.

Assume that the output of a source is a ~~binary~~ Bernoulli (p) process. The rate-distortion function for ~~a~~ the Hamming distortion measure is

$$R(D) = \begin{cases} H(p) - H(D) & 0 \leq D \leq \min(p, 1-p) \\ 0 & D > \min(p, 1-p) \end{cases}$$

$$R(D) = \min_{p(\hat{x}|x):\, E[d(x,\hat{x})] \leq D} I(X;\hat{X})$$

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(p) - H(X \oplus \hat{X} | \hat{X}) \\ &\geq H(p) - H(X \oplus \hat{X}) \\ &\geq H(p) - H(D) \end{aligned}$$

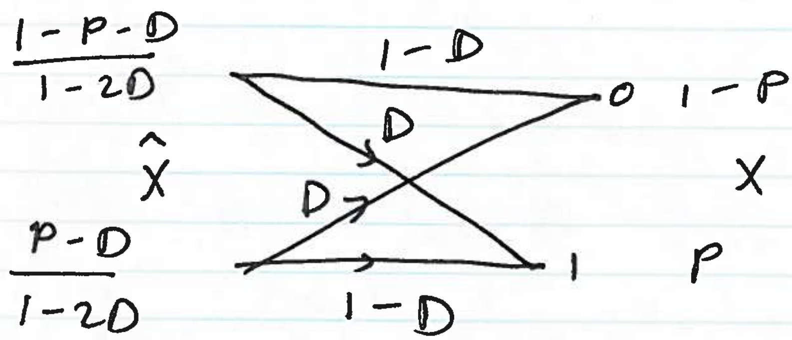
The last inequality is due to the fact that $\Pr(X \neq \hat{X}) \leq D$ and $H(D)$ increases with D for $D \leq 1/2$. So:

$$R(D) \geq H(p) - H(D).$$

Now, we choose $\hat{x}$ such that $p(\hat{x}|x)$ results in

$\Pr(X \neq \hat{X}) = D$ and $I(X;\hat{X}) = H(p) - H(D)$

Let a reverse channel connect $\hat{X}$ to X in the following way



Then,

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) = H(p) - H(D)$$

and

$$\Pr(X \neq \hat{X}) = D$$

If $D \geq p$, we can achieve $R(D) = 0$ by letting $\hat{X} = 0$ with probability 1. This way, we have $I(X;\hat{X}) = 0$ and $D = p$. Similarly we can get

$R(D) = 0$ by letting $\hat{X} = 1$ and having $D = 1 - p$.

$R(D)$ for a Gaussian Source

Theorem: The rate-distortion function for a $\mathcal{N}(0, \sigma^2)$ source with squared-error distortion is

$$R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2 \\ 0 & D \geq \sigma^2 \end{cases}$$

Proof:

by definition

$$R(D) = \min_{f(\hat{x}|x) : E[(X-\hat{X})^2] \leq D} I(X; \hat{X})$$

$$I(X;\hat{X}) = h(X) - h(X|\hat{X}) = \frac{1}{2}\log(2\pi e\sigma^2) - h(X - \hat{X} | \hat{X})$$

$$\geq \frac{1}{2}\log(2\pi\sigma^2) - h(X - \hat{X})$$

$$= \frac{1}{2}\log(2\pi\sigma^2) - h(Z)$$

where $Z = X - \hat{X}$

where Z_g is a normal zero-mean r.v. with variance $E[(X-\hat{X})^2]$

$$\geq \frac{1}{2}\log(2\pi\sigma^2) - h(Z_g)$$

$$= \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log 2\pi E[(X-\hat{X})^2]$$

$$\geq \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log 2\pi D$$

$$= \frac{1}{2}\log\frac{\sigma^2}{D}$$

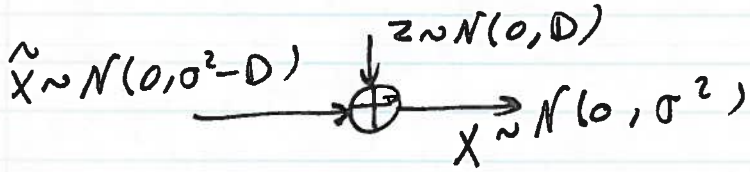
So,

$$R(D) \geq \frac{1}{2} \log \frac{\sigma^2}{D}$$

Now, we show that this ~~of~~ lower bound can be achieved by using a <u>test channel</u>, relating $\hat{X}$ to X so that $I(X;\hat{X}) = \frac{1}{2}\log\frac{\sigma^2}{D}$

Let $X = \hat{X} + Z$ where $\hat{X} \sim \mathcal{N}(0, \sigma^2 - D)$, $Z \sim \mathcal{N}(0, D$

i.e., we take $\hat{X}$ as a normal zero-mean r.v. with variance $\sigma^2 - D$.



This is possible if $\sigma^2 - D \geq 0$ or $D \leq \sigma^2$

$$\begin{aligned} I(X;\hat{X}) &= h(X) - h(X|\hat{X}) = \frac{1}{2}\log\frac{\sigma^2}{D} \\ &= h(X) - h(\hat{X}+Z|\hat{X}) \\ &= h(X) - h(Z|\hat{X}) \\ &= h(X) - h(Z) = \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

also $E[(X - \hat{X})^2] = D$.

So

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2 \\ 0 & D > \sigma^2 \end{cases} = \left(\frac{1}{2}\log\frac{\sigma^2}{D}\right)^+$$

This can be written as

$$\frac{\sigma^2}{D} = 2^{+2R}$$

where $\frac{\sigma^2}{D}$ is the Signal-To-Quantization-Noise Ratio

Changin into decibel, we get

$$SQNR = 10 \log \frac{\sigma^2}{D} = 10 \log 2^{+2R} = +20R \log 2$$

$$SQNR = +6R \quad dB$$

This means that for each 1 bit increase in resolution, we gain 6 dB (4 times improvement in SQNR.

X 11th lecture, Nov. 11, 2003

Simultaneous Quantization (Description) of Independent Gaussian Random variables:

Take normal sources $X_1, X_2, \ldots, X_m$ where $X_i \sim N(0, \sigma_i^2)$. We would like to find

$$R(D) = \min_{f(\hat{x}^m | x^m) : Ed[X^m, \hat{x}^m) \leq D} I(X^m; \hat{X}^m)$$

We have:

$$I(X^m; \hat{X}^m) = h(X^m) - h(X^m | \hat{X}_m)$$

$$= \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i | X^{i-1}, \hat{X}^m)$$

$$(a) \quad \geq \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i | \hat{X}_i)$$