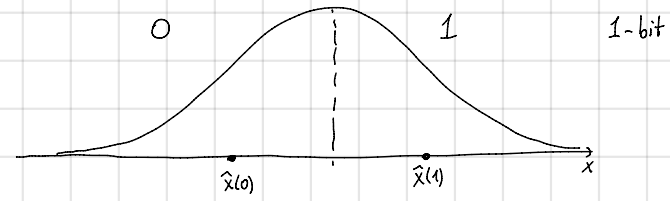
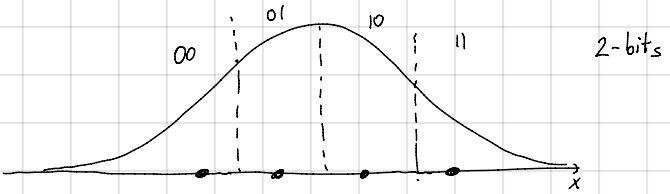
10/06/2016
Thursday

## Rate distortion theory: (Lossy Compression)

How do you encode with $R < H(X)$?

### Quantization:



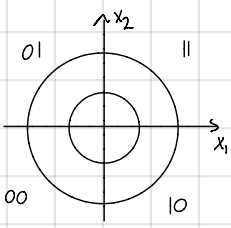

### Necessary Conditions:

1. Given encoding regions; $\hat{x}$ is conditional mean

2. Given reconstruction points; encode to nearest point (Voronoi region)

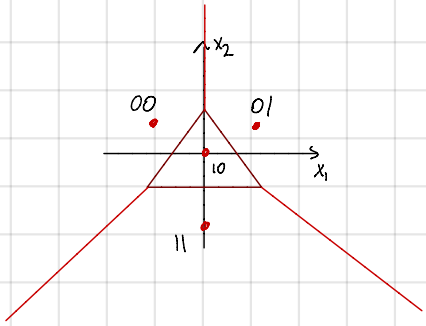
   Lloyd's algorithm: Iterate

2. IID Gaussian $X_1, X_2$



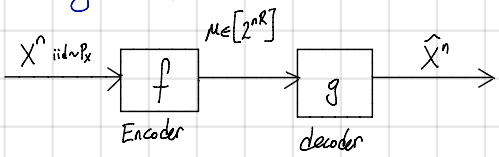

1-bit scalar quant

Vector (2-bit) Quantization




→ Better MSE even if $X_1 \perp\!\!\!\perp X_2$

# Lossy Compression:




Distortion function $d(x, \hat{x})$

ex: squared error: $d(x,\hat{x}) = (x-\hat{x})^2$

Hamming: $d(x,\hat{x}) = 1_{\{x \neq \hat{x}\}}$

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

Achievability: A rate distortion pair $(R, D)$ is achievable if $\exists\, n, f, g$ at rate $R$ such that

$$\mathbb{E}[d(X^n, \hat{X}^n)] \leq D$$

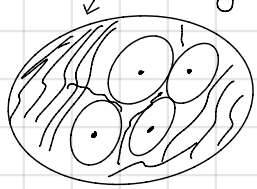
Rate-distortion function: $R(D) = \inf\{R : (R,D) \text{ is achievable}\}$

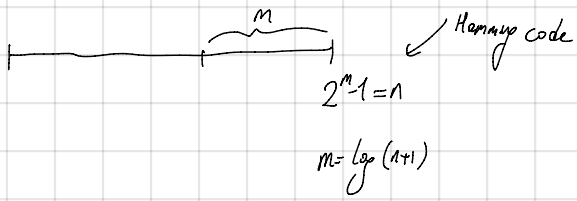
Distortion-Rate function: $D(R) = \inf\{D : (R,D) \text{ is achievable}\}$

Puzzle: $n$ random bits

At most 1 error allows

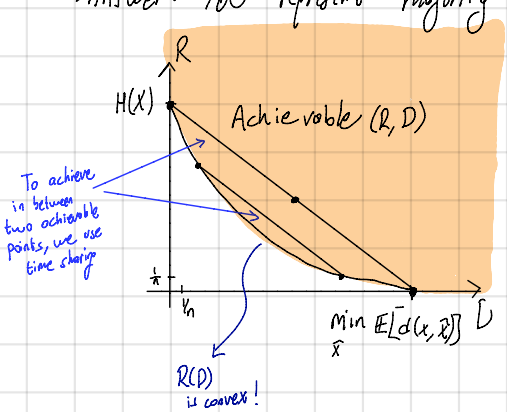
How many bits can be transmitted? Answer: $n - \log(n+1)$






Another Puzzle: Use 1-bit, what is smallest Hamming distance? $\frac{n}{2} - c\sqrt{n}$

→ Answer: You represent majority



## Rate-distortion Thm

We assume $|\mathcal{X}|<\infty$, $|\hat{\mathcal{X}}|<\infty$ and $d(x,\hat{x})<\infty$ $\forall x,\hat{x}$ but we don't need these assumptions.

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

Equivalently: Closure of set of achievable pairs $(R,D)$ is

$$\{(R,D) : \exists P_{\hat{X}|X} \text{ s.t. } R \geq I(X;\hat{X}),\ D \geq \mathbb{E}[d(X,\hat{X})]\}$$

Sanity check: If $R=0$, $\Rightarrow I'(X;\hat{X})=0$ $\quad D \geq \min_{P_{\hat{X}}} \mathbb{E}[d(X,\hat{X})] = \min_{\hat{x}} \mathbb{E}[d(X,\hat{x})]$

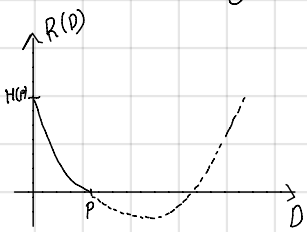
If $D=0$, $\Rightarrow \mathbb{E}[d(X,\hat{X})]=0$

If $d$ is such that $d(x,\hat{x})=0$ iff $x=\hat{x}$ $\Rightarrow X=\hat{X}$ w.p. 1

$\Rightarrow R \geq I(X;\hat{X}) = I(X;X) = H(X)$

**Example:** $X \sim \mathrm{Bern}(p)$ (Hamming distortion)

Claim: $R(D) = \begin{cases} H(p)-H(D) & D<p \text{ and } D<1-p \\ 0 & \text{else} \end{cases}$

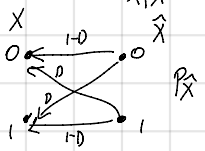



proof of claim: WLOG $p<1/2$

$D>p$ $\quad \hat{X}=0$ ✓

$D<p$ $\quad$ Construct $P_{\hat{X}|X}$ from $P_{X\hat{X}}$

we first prove achievability.




like "reverse" channel

$P_{\hat{X}}$ s.t.

$P_X \sim \mathrm{Ber}(p)$

$\mathbb{P}[\hat{X}=1]=\hat{p}$

$\mathbb{P}[X=1] = \hat{p}(1-D) + (1-\hat{p})D \overset{\text{need}}{=} p$

$\Rightarrow \hat{p} = \frac{p-D}{1-2D}$

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

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

$= H(p) - H(D)$

$$R \geq I(X;\hat{X})$$

$$= H(X) - H(X|\hat{X})$$

$$= H(X) - H(X \oplus \hat{X} | \hat{X})$$

$$\geq H(X) - H(X \oplus \hat{X})$$

$$= H(X) - H(E) \qquad E \text{ is binary}, \ \mathbb{E}[E] \leq D$$

$$\geq H(X) - H(D)$$

Example:

Let $X \sim \mathcal{N}(0,P)$ squared error $R(D) = \begin{cases} \frac{1}{2}\log\frac{P}{D} & D \leq P \\ 0 & D > P \end{cases}$

Empirical Distribution: $\pi(x|x^n) = \frac{1}{n}\sum_{i=1}^{n} 1\{x_i = x\}$

$\uparrow$ frequency of letter $x$ in $x^n$

Sequence Average Property:

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{x} 1(x_i = x)\right) f(x_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{x} 1(x_i = x) f(x) \qquad \text{"Sifting" property}$$

$$= \sum_{x} f(x)\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} 1(x_i = x)\right)}_{\pi(x|x^n)}$$

$$= \mathbb{E}_{\pi(x|x^n)} f(X)$$

Let $T \sim \text{unif}\{1,\dots,n\}$ and $T \perp\!\!\!\perp X^n$

$X_T$ is a R.V. that is a function of $X^n$ and $T$.

Properties:

1. $X_T \sim \mathbb{E}[\pi(x|x^n)] = \frac{1}{n}\sum_{i=1}^{n} P_{X_i}(x)$

2. If $X^n$ is i.i.d. $\sim P_X$, then $X_T \perp\!\!\!\perp T$ and $X_T \sim P_X$

Strong Typical Set: Given $P_X$

$$\mathcal{T}_\varepsilon^{(n)} = \left\{ x^n : |\pi(x|x^n) - P_X(x)| \leq \varepsilon \frac{P_X(x)}{H(X)} \right\}$$

convenience

Note: Jointly typical set defined in the same way.

Properties:

1. If $X^n \sim$ i.i.d. $\sim P_X$ $\quad \mathbb{P}(T_\varepsilon^{(n)}) \longrightarrow 1 \quad \forall \varepsilon > 0$ as $n \to \infty$

2. $\mathcal{T}_\varepsilon^{(n)} \subseteq A_\varepsilon^{(n)}$ (inherits all properties of $A_\varepsilon^{(n)}$)

<u>proof</u>: Let $x^n \in \mathcal{T}_\varepsilon^{(n)} \Rightarrow \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} = \frac{1}{n}\sum_{i=1}^{n}\log\frac{1}{P_X(x_i)} = \mathbb{E}_{\pi(x|x^n)}\log\frac{1}{P_X(x)}$

$$= \sum_x \pi(x|x^n)\log\frac{1}{P_X(x)}$$

$$\leq \sum_y P_X(x)\left(1+\frac{\varepsilon}{H(X)}\right)\log\frac{1}{P_X(x)}$$

$$= \left(1+\frac{\varepsilon}{H(X)}\right)H(X)$$

$$= H(X)+\varepsilon$$