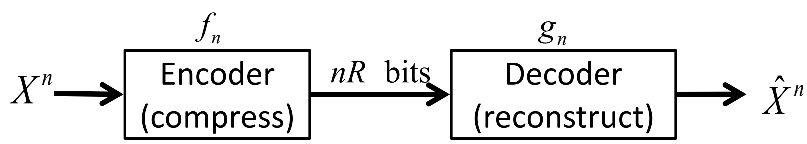
EE231A Information Theory
Lecture 13:
The Rate Distortion Function

A. Rate vs Distortion for Lossy Compression
B. Calculation of the Rate vs. Distortion Function
C. $R(D)$ for a Gaussian source

Part 13A:
Rate vs. Distortion
for
Lossy Compression

Lossy Compression

- $X_1, X_2, \ldots, X_n$ are i.i.d. from alphabet $\mathcal{X}$,



$$\boxed{\hat{X}^n = g_n\left(f_n\left(X^n\right)\right)}$$

- If $R < H(X)$, such a scheme cannot guarantee that $\hat{X}^n = X^n$. There is a loss in fidelity.

Measuring Distortion

- Distortion Metrics $\quad d: \mathcal{X} \times \hat{\mathcal{X}} \rightarrow R^+$
- Maximum Distortion $\quad d_{\max} = \max_{x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x})$
- Bounded distortion $\quad d_{\max} < \infty$

Hamming Distortion

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

- For Hamming distortion $E[d(x,\hat{x})] = P(\hat{X} \neq X)$

Squared-Error Distortion

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

Distortion between sequences

- Distortion between sequences

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

- Other definitions are possible…

Distortion of block-based lossy compression

$$E\left[d\left(X^n, g_n\left(f_n(X^n)\right)\right)\right] = \sum_{x^n} p(x^n) d\left(X^n, g_n\left(f_n(X^n)\right)\right)$$

Rate Distortion Pair

- A rate, distortion pair (R, D) is *achievable* if there is a sequence of $(2^{nR}, n)$ lossy compression codes (f_n, g_n) with

$$\lim_{n\to\infty} E\left[d\left(X^n, g_n\left(f_n(X^n)\right)\right)\right] \le D \quad .$$

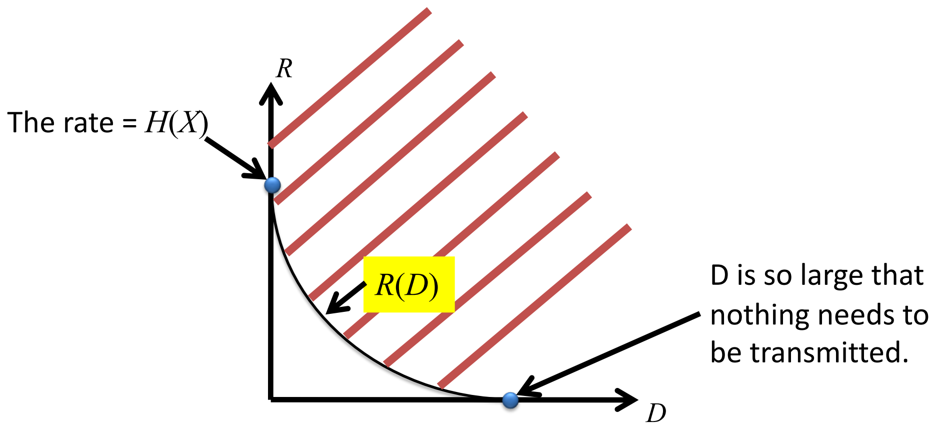
Rate-Distortion function

- For X i.i.d. $\sim p(x)$ and $d(x,\hat{x})$ bounded

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

Achievable Pairs R,D

- Rate distortion function bounds achievable pairs R, D



Part 13B: Calculation of the Rate Distortion Function

Rate-Distortion function

- For X i.i.d. $\sim p(x)$ and $d(x,\hat{x})$ bounded

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

$$E[d] = \sum_{x,\hat{x}} p(x,\hat{x}) d(x,\hat{x})$$

$$= \sum_{x,\hat{x}} p(x) \underbrace{p(\hat{x} \mid x)}_{\text{This is what we control.}} d(x,\hat{x})$$

Proof will come later

- Proof that *R(D)* is achievable is in a future lecture.
- Today, we focus on how to compute *R(D)*.

Computing $R(D)$

- *First example:* $R(D)$ for a binary source with Hamming distortion: $P(x=1)=p\le\frac{1}{2}$
- One way to find $R(D)$ is to find a lower bound on $I(X;\hat{X})$ and then achieve it.
- We have the constraint $E[d]=P(\hat{X}\ne X)\le D$

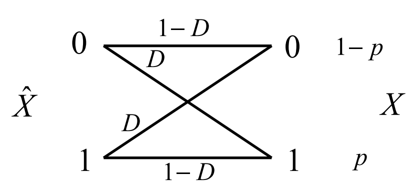
Lower bound on $I(X;\hat{X})$

$$\begin{aligned}
I(X;\hat{X}) &= H(X)-H(X\mid\hat{X})\\
&= H(p)-H(X\oplus\hat{X}\mid\hat{X})\\
&\ge H(p)-H(X\oplus\hat{X})\\
&= H(p)-H(P(X\ne\hat{X}))\\
&\ge H(p)-H(D) \qquad \text{for } D\le\frac{1}{2}
\end{aligned}$$

Achievability of the lower bound

- So $R(D) \geq H(p) - H(D)$
- Can we achieve $I(X;\hat{X}) = H(p) - H(D)$ with $E(d) \leq D$?
- We need to find a $p(x,\hat{x})$ that does that.

The Test Channel



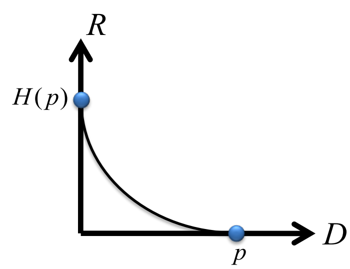
- $H(p) - H(D)$ is the $I(X;\hat{X})$ for a BSC with transition probability D and output distribution $p, 1\text{-}p$.
- Can we find an input distribution $p(\hat{x})$ to make it work?

Achievability of the lower bound (cont.)

$$P(X=0)=(1-D)P(\hat{X}=0)+DP(\hat{X}=1)$$
$$=1-p$$
$$P(X=1)=(1-D)P(\hat{X}=1)+DP(\hat{X}=0)$$
$$=p$$

$$\boxed{\begin{aligned} p(\hat{x}=0)&=\frac{1-p-D}{1-2D} \\ p(\hat{x}=1)&=\frac{p-D}{1-2D}\end{aligned}}$$

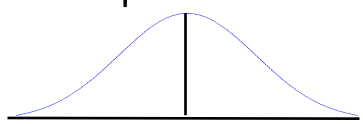
- So $R(D)=H(p)-H(D)$.



Part 13C:
R(D) for a Gaussian source

One bit quantization of Gaussian

- $R(D) = \min_{p(\hat{x}|x), E[d] \le D} I(X;\hat{X})$
- Consider one bit quantization of a Gaussian.



- Optimal 1-bit quantization

(Conditional mean)

$$S_1 = \{x : x > 0\} \qquad \hat{x}_1 = \sqrt{2/\pi}\sigma$$
$$S_2 = \{x : x \le 0\} \qquad \hat{x}_2 = -\sqrt{2/\pi}\sigma$$
$$E[(x - \hat{x}_i(x))^2] = \frac{\pi - 2}{\pi}\sigma^2 = 0.36\sigma^2$$

$R(D)$ for Gaussian source

- $R(D)$ for $X_1, X_2, \ldots$ i.i.d. $N(0,\sigma^2)$
- Find a lower bound on $I(X;\hat{X})$ and achieve it.

$$\begin{aligned}
I(X;\hat{X}) &= h(X) - h(X \mid \hat{X}) \\
&= \frac{1}{2}\log(2\pi e\sigma^2) - h(X - \hat{X} \mid \hat{X}) \\
&\ge \frac{1}{2}\log(2\pi e\sigma^2) - h(X - \hat{X}) \\
&\ge \frac{1}{2}\log(2\pi e\sigma^2) - h(\mathcal{N}(0, E(X-\hat{X})^2)) \\
&\ge \frac{1}{2}\log(2\pi e\sigma^2) - h(\mathcal{N}(0, D))
\end{aligned}$$

Gaussian test channel

$$I(X;\hat{X}) \geq \frac{1}{2}\log(2\pi e\sigma^2) - h(\mathcal{N}(0,D))$$
$$= \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right)$$
$$= \frac{1}{2}\log\left(1+\frac{\sigma^2 - D}{D}\right)$$

$$\boxed{R(D) = \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right)}$$

$$Z \sim \mathcal{N}(0,D)$$

$$\hat{X} \sim \mathcal{N}(0,\sigma^2 - D) \longrightarrow \oplus \longrightarrow X \sim \mathcal{N}(0,\sigma^2)$$

$R(D)$ for Gaussian channel (cont.)

- So we can achieve the lower bound.

$$R(D)\Big|_{D=0.36\sigma^2} = \frac{1}{2}\log\left(\frac{1}{0.36}\right)$$
$$= 0.737 \quad \text{bits}$$

or $R(0.25\sigma^2) = 1$ bit

So Information Theory indicates we can do a better job than optimal one-bit quantization by considering multiple symbols at a time.