In source coding, what to do when rate required is less than entropy of source? Converse of source coding thm shows error probability $\to 1$ as $n \to \infty$.

Allow distortion, but control its nature so fidelity is reasonable.

Quantize source : Lloyd-Max iterative algorithm
Sharma (1978) dynamic programming

functional source coding: certain parts of source alphabet less important than others, so not as fine quantization.

Analysis in high-resolution regime: fine-grained quantization cells

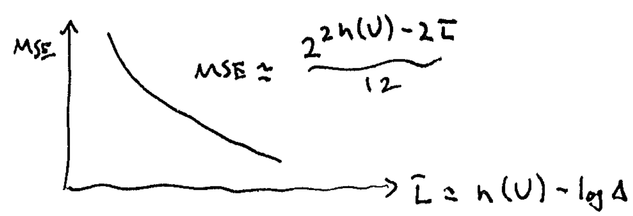
Differential entropy of analog r.v. U with pdf $f_U(u)$ is

$$h(U) = \int_{-\infty}^{\infty} -f_U(u) \log f_U(u)\, du .$$

Quantize finely, Δ, so mean-square error is approximately

$$\int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta} u^2 du = \frac{\Delta^2}{12} .$$

$$\text{entropy} \approx \int_{-\infty}^{\infty} -f_U(u) \log [f_U(u) \Delta] du = h[U] - \log \Delta .$$



Can extend to vector quantization, etc.

Rather than rate asymptotics, consider block length asymptotics.

Define a distortion measure between each source seq and each reproduction seq.

↳ try to design rate-distortion code that, w.h.p., reproduces source sequence with distortion within a tolerance level.

Let $\{X_k\}_{k=1}^{\infty}$ be an iid source from $\mathcal{X}$ according to $p(x)$.

where x_1^n is a source sequence reproduced as $\hat{x}_1^n$ when $\hat{x}_i \in \hat{\mathcal{X}}$

and both $\mathcal{X}, \hat{\mathcal{X}}$ finite sets.

define single-letter distortion measure, average distortion.

<u>def</u>: $d: \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$

) abuse of notation in extending to n letters

<u>def</u>: average distortion for sequences

$$d(x_1^n, \hat{x}_1^n) = \frac{1}{n} \sum_{k=1}^{n} d(x_k, \hat{x}_k)$$

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

<u>def</u>: Let $\hat{x}^*$ minimize $E[d(X, \hat{x})]$ over all $\hat{x} \in \hat{\mathcal{X}}$ and define

$D_{max} = E[d(X, \hat{x}^*)]$, here $\hat{x}^*$ is best estimate of X without any knowledge def: An (n, M) rate-distortion code has encoding/decoding

$$f : \mathcal{X}^n \to \{1, 2, \dots, M\}$$

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

where $g(f(1)), g(f(2)), \dots, g(f(M))$ are codewords.

def: rate of (n, M) code is $\frac{1}{n} \log M$.

def: a pair (R, D) is asymptotically achievable if for any $\varepsilon > 0$, there exists for large n, an (n, M) code s.t.

$$\frac{1}{n} \log M \le R + \varepsilon$$

$$\Pr\left[d(X, \hat{X}) > D + \varepsilon \right] \le \varepsilon$$

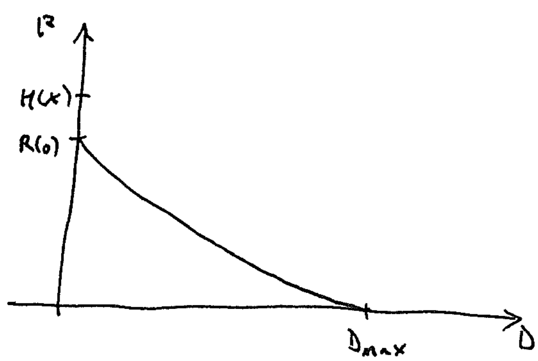
where $\hat{X} = g(f(X))$.

Clearly if (R, D) achievable, then (R', D) and (R, D') also achievable for all $R' \ge R$, $D' \ge D$.

def: rate-distortion region is subset of $\mathbb{R}^2$ containing all achievable pairs (R, D)

thm: rate-distortion region is closed and convex.

~~def.~~ thm Properties



① $R(D)$ is non-decreasing in D

② $R(D)$ is convex

③ $R(D) = 0$ for $D \ge D_{max}$

④ $R(0) \le H(X)$.

def: informational rate-distortion function

for $D \geq 0$

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

↑ optimize over $\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \}$

↑ forward test channel.

Thm rate-distortion thm

$R(D) = R_I(D)$.

converse proof Let (R,D) be any achievable rate-distortion pair. Then for any $\epsilon > 0$, there exists, for sufficiently large n, an (n,M) code with

$$\frac{1}{n} \log M \leq R + \epsilon \qquad \Pr[d(X,\hat{X}) > D+\epsilon] \leq \epsilon.$$

$$\begin{aligned}
n(R+\epsilon) &\geq \log M \\
&\geq H(f(X)) \\
&\geq H(g(f(X))) \\
&= H(\hat{X}) \\
&= H(\hat{X}) - H(\hat{X}|X) \\
&= I(X;\hat{X}) \\
&= H(X) - H(X|\hat{X}) \\
&= \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k | \hat{X}^n, X_1, X_2, \dots, X_{k-1}) \\
&\geq \sum_{k=1}^{n} H(X_k) - \sum_{k=1}^{n} H(X_k|\hat{X}_k) \quad \text{since conditioning doesn't increase entropy} \\
&= \sum_{k=1}^{n} [H(X_k) - H(X_k|\hat{X}_k)] = \sum_{k=1}^{n} I(X_k;\hat{X}_k)
\end{aligned}$$

$$\geq \sum_{k=1}^{n} R_I\left(E\left[d(X_k, \hat{X}_k)\right]\right) \quad \text{by definition of } R_I(D).$$

$$= n\left[\frac{1}{n}\sum_{k=1}^{n} R_I\left(E[d(X_k,\hat{X}_k)]\right)\right]$$

$$\geq n\, R_I\left(\frac{1}{n}\sum_{k=1}^{n} E[d(X_k,\hat{X}_k)]\right) \quad \text{by convexity of } R_I(D) \text{ and Jensen's ineq.}$$

$$= n\, R_I\left(E[d(X_1^n,\hat{X}_1^n)]\right)$$

a little bit of continuity arguments then imply

$$R(D) \geq R_I(D).$$

---

achievability : want to show $R(D) \leq R_I(D)$.

use random coding argument

① construct an (n, M) codebook $\mathcal{C}$ by randomly generating M codewords in $\hat{\mathcal{X}}^n$ iid $\sim p(\hat{x})^n$. Denote these by $\hat{X}_1^n(1), \hat{X}_1^n(2), \ldots \hat{X}_1^n(M)$.

② reveal codebook $\mathcal{C}$ to both encoder and decoder.

③ Source sequence is generated according to $p(x)^n$.

④ the <u>encoder</u> uses typicality as follows:

encoder encodes source sequence X_1^n into an index $K \in \{1,2,\ldots,M\}$ where K takes value i if

ⓐ $(X_1^n, \hat{X}_1^n(i)) \in T^n_{[X\hat{X}]\delta}$, the strongly jointly typical set.

ⓑ for all $i' \in \{1,2,\ldots,M\}$ if $(X_1^n, \hat{X}_1^n(i')) \in T^n_{[X\hat{X}]\delta}$, then $i' \leq i$

otherwise $K = 1$

⑤ K is sent to decoder, which produces $\hat{X}(K)$ as reproduction.

more description of encoding step

After X_1^n generated, we search through all codewords in $\mathcal{C}$ for those which are jointly typical with X_1^n with respect to $p(x,\hat{x}) = p(x)p(\hat{x}|x)$.

if there is at least one such codeword, we let i be largest index of such codewords and set $K=i$. if no such codeword, $K=1$.

where does distortion come into the picture?

clever choice of δ in strongly typical set definition.

An example due to Erokhin.

Bernoulli(p) source with Hamming distortion

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

rather than minimizing $I(X;\hat{X})$ directly, find lower bound that is achieved.

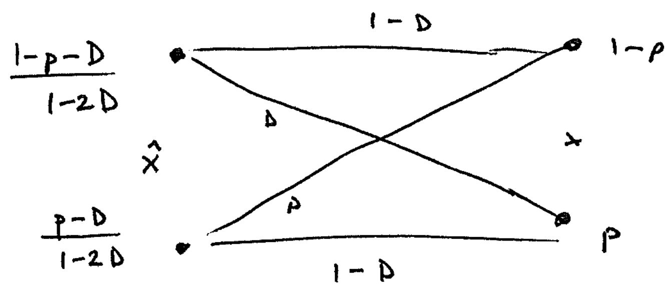
let $\oplus$ denote mod 2 addition so $X \oplus \hat{X} = 1$ is equivalent to $X \neq \hat{X}$.

for any joint distribution satisfying distortion constraint

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_2(p) - H(X \oplus \hat{X} | \hat{X}) \\ &\ge h_2(p) - H(X \oplus \hat{X}) \\ &\ge h_2(p) - h_2(D) \end{aligned}$$

so $R(D) \ge h_2(p) - h_2(D)$

this can be achieved with forward test channel:



that is BSC.

$$\text{so } R(D) = \begin{cases} h_2(p) - h_2(D), & 0 \le D \le \min(p, 1-p) \\ 0, & D > \min(p, 1-p). \end{cases}$$