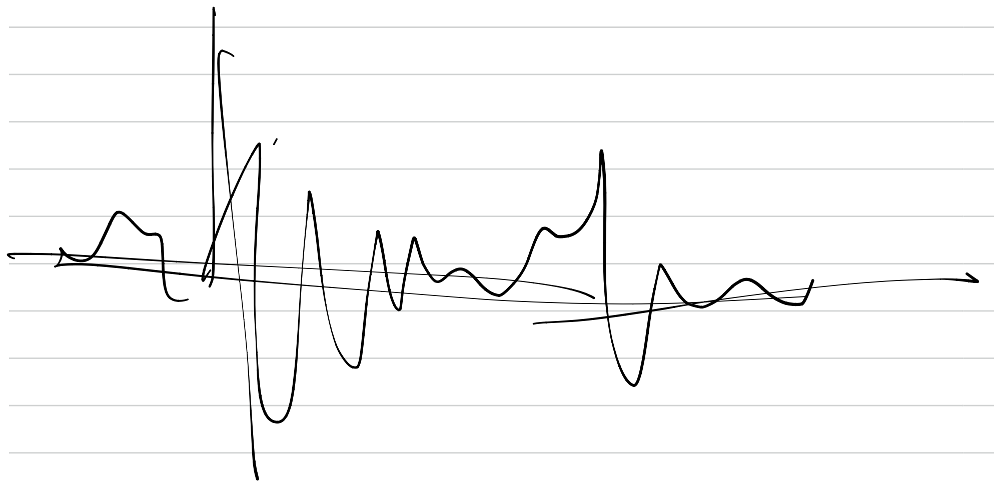
Information Theory & Coding

Sept. 14th 2020

 $\rightsquigarrow$ Xmittable form

ABD... $\longrightarrow$ — . . — . . — ∘ . . .

$\boxed{\text{Source}} \longrightarrow U_1 U_2 U_3 \ldots$

$U_i \in \mathcal{U}$, $|\mathcal{U}| < \infty$

e.g. $\mathcal{U} = \{a, b, c, d, e, f\}$

Letters $U_1, U_2, U_3, \ldots$ are Random Variables.

(i.e., source $\equiv$ stochastic process).

(Question: given a stochastic process, what is the smallest amount of space required to represent it? (bits/letter)) channel

$$\underset{\text{pt } A}{X} \longrightarrow \underset{\text{noisy}}{\bigcirc} \longrightarrow \underset{\text{pt } B}{Y}$$

$$p(y \mid x) = \Pr(Y = y \mid X = x)$$

- probability kernel
- transition probability.
- ...

Source Coding / Data Compression

alphabet $\mathcal{U}$, $u \in \mathcal{U}$, $|\mathcal{U}| < \infty$

letter

U is a random variable taking values in $\mathcal{U}$.

a source code is a mapping

$c : \mathcal{U} \to \{0,1\}^* = \{\text{null}, 0, 1, 00, 01, 10, 11, 000, 001, \ldots\}$

= set of all finite (binary) sequences.

$c_n : \mathcal{U}^n \longrightarrow \{0,1\}^*$

---

ex: $\mathcal{U} = \{a, b, c, d\}$

$c(a) = 0$

$c(b) = 0$

$c(c) = 1$

$c(d) = 0$

singular = non-injective

= bad.

= stupid

Def: a code is called injective (non-singular) if $u \neq v \Rightarrow c(u) \neq c(v)$

Ex: $\mathcal{U} = \{a, b, c, d\}$

$c(a) = 0$

$c(b) = 00$

$c(c) = 1$

$c(d) = 10$

$c^*(aa) = c^*(b)$

not good.

$c(d) = c(c)c(a)$

Def:

given $c : \mathcal{U} \to \{0,1\}^*$ let us define

$c^* : \mathcal{U}^* \to \{0,1\}^*$ as

$c^*(u_1 \ldots u_n) = \underbrace{c(u_1)\, c(u_2)\, c(u_3) \ldots c(u_n)}_{\text{Concatenation}}$

$c^n : \mathcal{U}^n \to \{0,1\}^*$ as

$c^n(u_1 \ldots u_n) =$

Def: c is called uniquely-decodable if c^* is injective.

Def: given a code $c : \mathcal{U} \to \{0,1\}^*$, define

$$\underline{\text{KraftSum}}(c) = \underline{\sum_{u \in \mathcal{U}} 2^{-\text{length}(c(u))}}$$

for the code in the Cart example $KS = 3/2$.

Thm: if $c : \mathcal{U} \to \{0,1\}^*$ is injective then

$$\underline{\text{KraftSum}(c) \le \log_2(1 + |\mathcal{U}|)}$$

Pf: Without loss of generality we can assume that if $\underline{k = \text{length}(c(u))}$, then for all binary sequences b with $\text{length}(b) < k$ there is a letter v s.t. $c(v) = b$.

(otherwise if b is not a codeword then we can set $c(u) = b$, without violating injectivity, and this modification $\nearrow$ KS.) Consequently we can assume that, with $\underline{k = \text{length of the longest codeword}}$, the set of codewords, is $\bigcup_{i=0}^{k-1} \{0,1\}^i \cup (\text{subset of } \{0,1\}^k)$

of cardinality $= r$

$1 \leq r \leq 2^k$

0, 00, 10, 11, ~~0000~~

$\rightarrow$ 0, 00, 10, ~~11~~, (null) $\rightarrow$ ~~0~~, 1,

null, 0, 1, 00, 10

$|\mathcal{U}| = \underbrace{1 + 2 + \ldots + 2^{k-1}} + r$

$= 2^k - 1 + r$

$K.S = \sum_{i=0}^{k-1} 2^i \cdot 2^{-i} + r 2^{-k}$

$= k + r 2^{-k} = x$

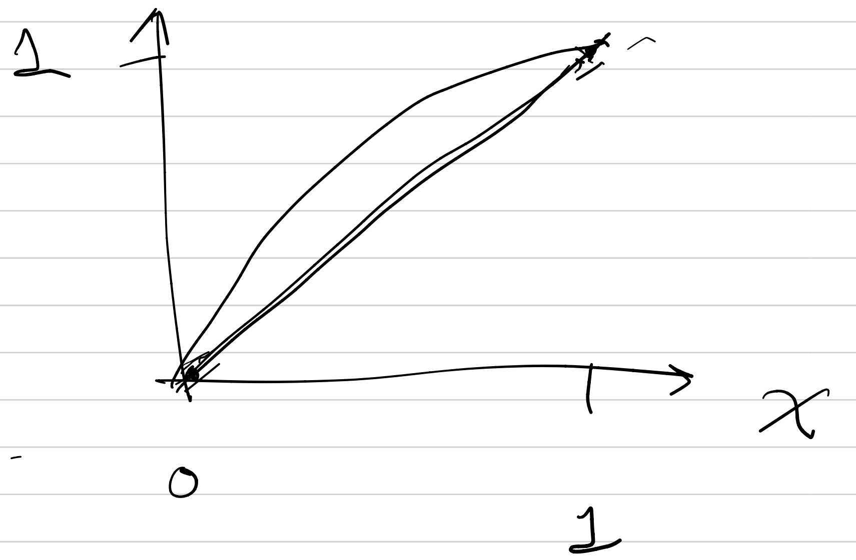
we are supposed to show $k + x \leq \log(1 + |\mathcal{U}|)$

$$\cancel{k} + x \le \log_2(2^k - 1 + 1 + r)$$

$$= \log_2(2^k + r)$$

$$= \log_2\left(2^k\left(1 + \underbrace{r2^{-k}}_{x}\right)\right)$$

$$= \cancel{k} + \log_2(1+x) \qquad 0 < x \le 1$$



$x \to \log_2(1+x) - x$

is concave

& $= 0$ at $x = 0, 1$

$\Rightarrow$ must be ≥ 0 in between. //

Suppose now U is a RV taking values in $\mathcal{U}$, & we have a code $c : \mathcal{U} \to \{0,1\}^*$

$$\boxed{E[\text{length } c(U)]}$$

$$= \sum_u p(u) \underbrace{\text{length}(c(u))}$$

$$= \sum_u p(u) \log_2 \frac{p(u)}{2^{-\text{length}(c(u))}\, p(u)}$$

$$= \sum_u p(u) \log \frac{1}{p(u)} - \sum_u p(u) \log_2 \frac{2^{-\text{length } c(u)}}{p(u)}$$

$$\geq \sum_u \quad '' \quad - \log_2 \sum_u \cancel{p(u)} \frac{2^{-\text{length } c(u)}}{\cancel{p(u)}}$$

concavity of log

$$= \underbrace{\sum_u p(u) \log \frac{1}{p(u)}}_{\text{intrinsic to } U} - \log_2(\text{KraftSum}(c))$$

Summary:

$$E[\text{length } c(U)] \geq H(U) - \log_2(\underbrace{\text{KraftSum}(c)}_{})$$

entropy of the source

$$\triangleq \sum_u p(u) \log_2 \frac{1}{p(u)}$$

Thm: if c is injective then,

$$E[\text{length } c(U)] \geq \underline{H(U)} - \underbrace{\log_2 \underbrace{\log_2(1+|\mathcal{U}|)}_{}}_{}$$

Thm: if c is uniquely decodable then

$$E[\text{length}(c(U))] \geq H(U).$$

Pf: if c is uniquely decodable then

$c^n: \mathcal{U}^n \to \{0,1\}^*$ defined as

$c^n(u_1 \ldots u_n) = c(u_1) \ldots c(u_n)$ is injective.

$$\underline{E[\text{length } c^n(U_1 \ldots U_n)]} \geq \underline{H(U_1 \ldots U_n)} - \log\log(1+|\mathcal{U}|^n)$$

note: $\text{length } c^n(u_1 \ldots u_n)$

$$= \text{length } c(u_1) + \text{length } c(u_2) + \ldots \text{length } c(u_n)$$

$$E(\text{length}(c^n(U_1 \ldots U_n)))$$

$$= E(\text{length } c(U_1)) + \ldots + E[\text{length } c(U_n)]$$

$$= n\, E(\text{length } c(U)).$$

$$H(U_1 \ldots U_n) = \sum_{u_1 \ldots u_n}' p(u_1 \ldots u_n) \log_2 \frac{1}{p(u_1 \ldots u_n)}$$

$p(u_1 \ldots u_n) = p(u_1)\, p(u_2) \ldots p(u_n)$ so

$$\log \frac{1}{p(u_1 \ldots u_n)} = \underline{\log \frac{1}{p(u_1)}} + \log \frac{1}{p(u_2)} + \ldots \log \frac{1}{p(u_n)}$$

$$H(U_1 \ldots U_n) = \underbrace{\sum_{u_1 u_2 \ldots u_n}' p(u_1)\, p(u_2) \ldots \underline{p(u_n)} \log \frac{1}{p(u_1)}}_{} + \ldots$$

$$= \sum_{u_1}' p(u_1) \log_2 \frac{1}{p(u_1)} + \ldots$$

$$= H(U) + H(U) + \ldots + H(U) = n H(U)$$

So we haven

$$n\,E(\text{length } c(U)) \geq n H(U) - \log_2 \log_2 (1+|U|^n)$$

$$\Rightarrow \underline{E(\text{length } c(U))} \geq \underline{H(U)} - \frac{1}{n}\log\log(1+|U^n|)$$

By taking lim as $n \nearrow \infty$ we get $O\left(\frac{1}{n}\log n\right)$

$$\underline{E(\text{length } c(U))} \geq \underline{\underline{H(U)}}.$$

//