Information Theory & Coding

Oct 5th 2020

So far:

- codes, ..
  - Expected codeword length

$$H(U) \le E[\hat{L}] \le H(U)+1$$

  - Huffman procedure to design $\hat{c}$
  - Entropy Rate $\frac{1}{n} H(U_1 \ldots U_n)$

$$H(U_n | U_1 \ldots U_{n-1})$$

  - [Chain Rules, mutual information, ...]
  - Typicality & typical sets.

given $\mathcal{U}$, p a distribution on $\mathcal{U}$

"Robust typicality" we call $\boxed{u_1 \ldots u_n}$: ε-typical wrt p if $\frac{1}{n} |\{i : u_i = u\}| = p(u)(1 \pm \varepsilon)$ $\forall u \in \mathcal{U}$

  - $\Pr(\underbrace{U_1 \ldots U_n}_{\text{iid} \sim p} \in \{\text{typical set of sequences}\}) \approx 1$
  - $\boxed{\Pr(\underbrace{U_1 \ldots U_n}_{\text{iid } p} = \underbrace{u_1 \ldots u_n}_{\text{an element of } T(p)}) = 2^{-n H(U)(1 \pm \varepsilon)}}$ $\sim p$
  - $|T(n, \varepsilon, p)| \doteq 2^{n H(U)(1 \pm \varepsilon)}$

Example: $\mathcal{U} \times \mathcal{V}$, $\underline{p_{uv}}$: given.

$$p_u(u) = \sum_v p_{uv}(u,v).$$

$$p_v(v) = \sum_u p_{uv}(u,v)$$

suppose we have $(\underline{u_1}, v_1), (\underline{u_2} v_2) \dots (\underline{u_n}, v_n)$

$\in T(\varepsilon, n, p_{uv})$

$$(1-\varepsilon) \dots \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\underbrace{(u_i v_i) = (uv)}\} \leq p(uv)(1+\varepsilon)$$

Q: is $(u_1 \dots u_n) \in T(\varepsilon, n, p_u)$. ?

A: yes.

$$\underline{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{u_i = u\}} = \frac{1}{n} \sum_{i=1}^{n} \sum_{v \in \mathcal{V}} \mathbb{1}\{u_i = u, v_i = v\}$$

$$= \sum_{v \in \mathcal{V}} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{u_i v_i = uv\}}_{\underline{\leq p_{uv}(uv)(1+\varepsilon)}}$$

$$\leq p_u(u)(1+\varepsilon) \quad .. \quad \text{similar}$$

$$\dots \geq (1-\varepsilon)\, p(u)$$

$\Rightarrow (u_1 \dots u_n) \in T(\varepsilon, n, p_u).$

when $U_1 \dots U_n$ is not matched to $T(p, \varepsilon, n)$

Suppose : $\underbrace{U_1 \dots U_n}_{\text{iid} \sim \underline{\underline{q}}}$ & $\underbrace{u_1 \dots u_n}_{\in \underline{\underline{T}}(p, \varepsilon, n)}$

$$\Pr(U_1 \dots U_n = u_1 \dots u_n)$$

$$= \prod_{i=1}^{n} \Pr(\underline{\underline{U_i = u_i}}) = \prod_{i=1}^{n} \underline{\underline{q(u_i)}}$$

$$= \prod_{u \in \mathcal{U}} q(u)^{\overbrace{\underbrace{\quad\quad\quad\quad\quad}_{= np(u)(1 \pm \varepsilon)}}^{\text{number of times } u \text{ appears in } (u_1 \dots u_n)}}$$

$$\Rightarrow \left(\prod_{u \in \mathcal{U}} q(u)\right)^{np(u)(1+\varepsilon)} \leq \Pr(\quad) \leq \left(\prod_{u \in \mathcal{U}} \underline{\underline{q(u)}}^{np(u)}\right)^{(1-\varepsilon)}$$

$$2^{n(1+\varepsilon)\sum_{u \in \mathcal{U}} p(u) \log q(u)} \leq \dots \leq 2^{n(1-\varepsilon)\sum_{u \in \mathcal{U}} p(u) \log q(u)}$$

Note: $$\sum_{u \in \mathcal{U}} p(u) \log q(u) = \sum_{u \in \mathcal{U}} p(u) \left[\log p(u) + \log \frac{q(u)}{p(u)}\right]$$

{ Aside, example: $\mathcal{U} = \{\alpha, \beta\}$

$$p: \quad \frac{1}{2} \quad \frac{1}{2}$$

$$q: \quad 3/4 \quad 1/4$$

p-typical sequence look like $\alpha \ldots \beta$

$\approx \frac{n}{2}$ α's $\frac{n}{2}$ β's

$$\Pr(\underbrace{U_1 \ldots U_n}_{\sim q} \in T(n, p, \varepsilon)) = ?$$

So for: $u_1 \ldots u_n \in T(n, p, \varepsilon)$

$$\Pr(\underbrace{U_1 \ldots U_n}_{\sim q} = u_1 \ldots u_n) = 2^{-n\left(\sum_{u \in \mathcal{U}}' p(u) \log \frac{1}{q(u)}\right)(1 \pm \varepsilon)}$$

$$\sum_{u \in \mathcal{U}}' p(u) \log \frac{1}{q(u)} = \sum_{u \in \mathcal{U}}' p(u) \left[\log \frac{1}{p(u)} + \log \frac{p(u)}{q(u)}\right]$$

$$= H(p) + \underbrace{\sum_{u \in \mathcal{U}}' p(u) \log \frac{p(u)}{q(u)}}_{D(p \| q)}$$

"KL-divergence from q to p"

So: if $\underbrace{U_1 \dots U_n}_{\text{iid } q}$ & $u_1 \dots u_n \in T(p, n, \varepsilon)$

Then:

then:

$$\Pr(U_1 \dots U_n = u_1 \dots u_n)$$

$$= 2^{-n(1 \pm \varepsilon)[H(p) + \underline{D(p \| q)}]}.$$

and $\Pr(\underbrace{U_1 \dots U_n}_{\text{iid} \sim q} \in T(p, \varepsilon, n)) \approx 2^{-n \underline{D(p\|q)}}$

details in the proof.

Pf: $\Pr(U_1 \dots U_n \in T(p, \varepsilon, n))$

$$= \sum_{(u_1 \dots u_n) : \in T(\ )} \underline{\Pr(U_1 \dots U_n = u_1 \dots u_n)}$$

$$\leq \sum_{(u_1 \dots u_n) \in T(p, \varepsilon, n)} \underbrace{2^{-n[D(p\|q) + H(p)](1-\varepsilon)}}$$

$$= \underline{|T(p, \varepsilon, n)|} \cdot 2^{-n(1-\varepsilon)[D(p\|q) + H(p)]}.$$

$$\leq 2^{n(1+\varepsilon)H(p)} \, 2^{-n(1-\varepsilon)(D(p\|q) + H(p))}$$

$$= 2^{-n[(1-\varepsilon) \underline{D(p\|q)} - 2\varepsilon H(p)]}$$

$$\Pr(\underbrace{U_1 \ldots U_n}_{\sim q} \in T(p, \varepsilon, n))$$

$$\leq 2^{-n\left[\underbrace{D(p\|q)} - \varepsilon \underbrace{\text{junk}}_{(2H(p) + D(p\|q))}\right)}$$

Similarly: for n large we know

$$|T(n, p, \varepsilon)| \geq (1-\varepsilon)\, 2^{n H(p)(1-\varepsilon)}$$

So the same computation will give

$$\Pr(U_1 \ldots U_n \in T(\quad))$$

$$\geq (1-\varepsilon)\, 2^{-n\left(D(p\|q) + \varepsilon \,\underline{\text{junk}}\right]}$$

For concreteness:

<u>Def</u>: given two distributions p, q on the set $\mathcal{U}$, we define

$$D(p\|q) = \sum_{u \in \mathcal{U}} p(u) \log \frac{p(u)}{q(u)},$$

__Thm__: $D(p\|q) \geq 0$, equality iff $p = q$.

__Pf__: $$D(p\|q) = -\sum_{u \in \mathcal{U}}' p(u) \log \frac{q(u)}{p(u)}$$

$$\leq -\log \sum_{u \in \mathcal{U}}' \cancel{p(u)} \frac{q(u)}{\cancel{p(u)}}$$

$$= -\log \sum_{u}' q(u)$$

$$= -\log 1$$

$$= 0. \quad //$$

Observe:

$$I(U;V) = \sum_{u,v}' p(uv) \log \frac{p(uv)}{p(u)\,p(v)}$$

$$= D(p_{uv} \| p_u p_v)$$

$(p_u p_v)(u,v) = p_u(u)\, p_v(v).$

More places we encounter $D(\ )$.

Suppose that we are given a distribution q on an alphabet $\mathcal{U}$ and we design a code for $\mathcal{U}$ with the belief that the distrib. is indeed q. The ideal codeword lengths would be $\log_2 \frac{1}{q(u)}$, and the expected codeword length would be

$$H(q) = \sum_{u\in\mathcal{U}}' q(u) \log \frac{1}{q(u)},$$

If the true distribution is p (not q).

then

$$E[\text{length}] = \sum_{u \in \mathcal{U}}' p(u) \log \frac{1}{q(u)}$$

$$= \sum_{u \in \mathcal{U}}' p(u) \left(\log \frac{1}{p(u)} + \log \frac{p(u)}{q(u)} \right)$$

$$= \underline{H(p)} + \underbrace{D(p \| q)}_{\text{penalty for designing for } q \text{ when truth} = p}.$$

Also note: when we design a code (u.d.)

$c : \mathcal{U} \to \{0,1\}^*$, we have a

$$\underline{q(u) \triangleq 2^{-\text{length}(c(u))}},$$ which satisfies

$$\sum_{u \in \mathcal{U}}' q(u) \leq 1.$$

if $\stackrel{(1)}{=}$ then q is a distribution

$<$ then pick a fictitious letter not in $\mathcal{U}$ u_0

with $q(u_0) = 1 - \sum_{u \in \mathcal{U}}' q(u)$.

Then $\underset{u \sim p}{E}[\text{length } c(U)] = \sum' p(u) \log \frac{1}{q(u)}$

$$= H(p) + D(p\|q).$$

Logic: code $\longrightarrow$ q $\rightsquigarrow$ $D(p\|q)$

Question: can we do code design ($\equiv$ find a distribution q) without knowing the true distribution p?

Example: Suppose we know that $U_1 U_2 U_3 \ldots$ is iid on $\{0,1\}$, i.e.,

$$\Pr(U_i = 0) = 1 - \underbrace{\Pr(U_i = 1)}_{= \theta} \qquad 0 \leq \theta \leq 1 \text{ unknown}$$

& $U_1 \ldots U_n$ are independent.

Suggestion for a code:

$$C : \{0,1\}^{n} \longrightarrow \{0,1\}^* \quad \text{(prefix-free)}.$$

given $(u_1 \ldots u_n)$ count the # of 1's in it.

$$k \in \{0, 1, \ldots, n\}$$

describe this count using $\lceil \log_2 (n+1) \rceil$ bits.

At this moment all I need to specify is

which one among the binary sequences of k 1's & $(n-k)$ 0's

is $(u_1 \dots u_n)$. There are

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$ such sequences. So

$\left\lceil \log_2 \binom{n}{k} \right\rceil$ bits suffices to describe $u_1 \dots u_n$

over all:

$$\boxed{\text{length}\,(c(u_1 \dots u_n)) = \lceil \log (n+1) \rceil + \left\lceil \log_2 \binom{n}{k} \right\rceil}$$

Note: $0 \le t \le 1$

$$1 = 1^n = (1-t+t)^n = \sum_{i=0}^{n} \binom{n}{i} (1-t)^{n-i} t^i$$

$$\geq \binom{n}{k} (1-t)^{n-k} t^k$$

$$\Rightarrow \binom{n}{k} \le \frac{1}{(1-t)^{n-k}\, t^k} \qquad \forall\, t \in [0,1)$$

$$\Rightarrow \binom{n}{k} \le \frac{1}{\left(1-\frac{k}{n}\right)^{n-k} \left(\frac{k}{n}\right)^k}$$

$$\log_2 \binom{n}{k} \le (n-k)\log \frac{1}{1-\frac{k}{n}} + k \log \frac{1}{(k/n)}$$

$$\frac{1}{n}\log_2 \binom{n}{k} \le \underbrace{(1-t)\log \frac{1}{1-t} + t \log \frac{1}{t}}_{h(t)} \qquad t = k/n.$$

Consequences

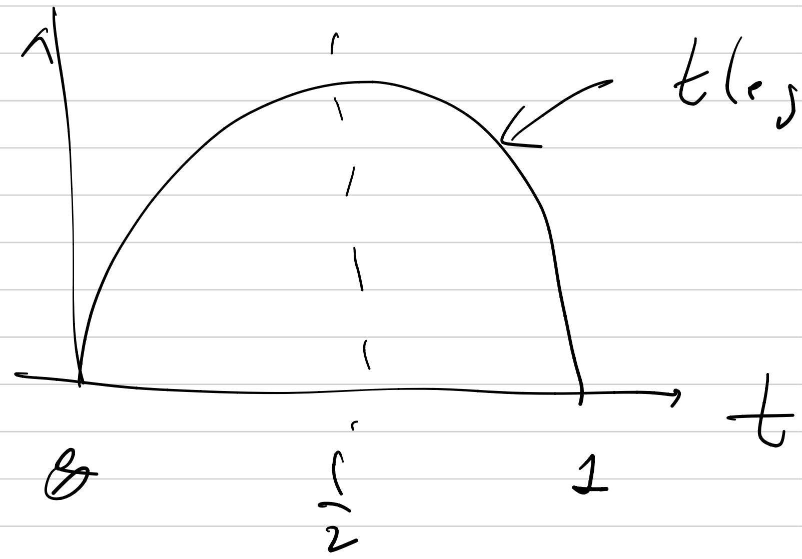
$$\frac{1}{n} \text{ length } c(\underline{u_1 .. u_n}) \le \underbrace{\frac{2}{n} + \frac{\log(1+n)}{n}} + \qquad t = k/n.$$

$$k = k(u_1 .. u_n) = \# \text{ of } 1's \text{ in } u_1 .. u_n.$$

$$E\left[\frac{1}{n} \text{ length } c(U_1 .. U_n)\right] \le \frac{2}{n} + \frac{\log(1+n)}{n}$$

$$\underline{+ E\left[h\left(\frac{k(U_1 .. U_n)}{n}\right)\right]}.$$

$h(t)$



$t \log \frac{1}{t} + (1-t)\log \frac{1}{1-t}$

= binary entropy function.

is a concave function of t.

$$\Rightarrow E[h(\quad)] \le h(E[\_\_])$$

So:

$$\frac{1}{n} E\left[\text{length } c(U_1 .. U_n)\right] \le \frac{2}{n} + \frac{\log(1+n)}{n} + h\left(\frac{E\left(k(U_1 .. U_n)\right)}{n}\right)$$

if $U_i = \begin{cases} 0 & \text{w.p. } 1-\theta \\ 1 & \text{w.p. } \theta \end{cases}$

$$\frac{1}{n} E\left(k(U_1 .. U_n)\right) = \cancel{\theta}\ \theta$$

$$= \boxed{\frac{1}{n} E(\text{length}(\ ))} \le \underbrace{\frac{2}{n} + \frac{\log(1+n)}{n}}_{\approx 0} + \underbrace{h(\theta)}_{H(U_1) = \mathcal{H}\{U\}}$$

$k(U_1 .. U_n) =$ # of 1's in $U_1 .. U_n$

$$= \sum_{i=1}^{n} U_i$$

$$E\left(k(U_1 .. U_n)\right) = \sum_{i=1}^{n} E(U_i) = n\, E(U_1)$$

$$= n\,\theta$$