Compression (Transform coding)

Question

for simplicity $\mathbb{R}$, $\mathbb{C}$ analogous.

$f : \Omega \to \mathbb{R}$

- How many bits do we need to encode the function f?
- Example $\Omega = [0, T]$ or $\Omega = [0,1] \times [0,1]$

(1) Need a function class $\mathcal{F}$.

(2) An encoder / decoder

$$E : \mathcal{F} \to \{0,1\}^R$$

$$D : \{0,1\}^R \to \mathcal{F}$$

(3) $$d(R, f) := \| f - D \circ E(f) \|_{L^2}^2$$

$$d(R, \mathcal{F}) := \sup_{f \in \mathcal{F}} \| f - D \circ E(f) \|_{L^2}^2$$

Background: compression of discrete sequences.

$\mathcal{X} = \{1, \dots k\}$. finite alphabet

p_* prob distribution over $\mathcal{X}$

$$T_n(p_*, \epsilon) = \left\{ x \in \mathcal{X}^n : \|\hat{p}_x - p_*\|_{TV} \leq \epsilon \right\}$$

$$\hat{p}_x(s) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i = s}.$$

$$\lim_{n\to\infty} \frac{1}{n} \log |T_n(p_*, \epsilon)| = H(p_*) + \Delta(\epsilon)$$

$$H(q) = -\sum_x q(x) \log q(x)$$

$\forall t > h_*(p)$

Propo Assume $\epsilon_n \downarrow 0$. Then there exist $\varphi : T_n(p_*, \epsilon_n) \to \{0,1\}^{nr}$

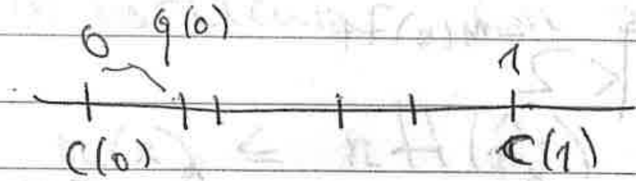
st $\cancel{\varphi^{-1} \circ \varphi(x) = x}$ $\quad \psi : \{0,1\}^{nr} \to \mathcal{X}^n$

$$\psi \circ \varphi(x) = x \quad \forall x \in T_n(p_*, \epsilon_n)$$

Arithmetic coding

- Prob distribution q on $\mathcal{X}$ $\quad (q(x))_{x \in \mathcal{X}}$
- Define $c(z) = \sum_{z'=1}^{z} q(z')$ $\quad c(0) = 0$

$c(k) = 1$



Construct recursively an interval

$I_n(\underline{x}) = (a_n(\underline{x}), b_n(\underline{x})]$ for $x \in \mathcal{X}^n$

* $I_0(\underline{x}) := (0, 1]$

* Given $I_k(\underline{x}) = (a_k(\underline{x}), b_k(\underline{x})]$

$$I_{k+1}(x) = a_k(x) + (b_k(x) - a_k(x)) \cdot \cancel{c(x_{k+1})} - c(x_k)$$

J_{x_k}

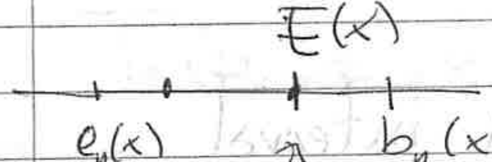
$\cdot J_{x_{k+1}}$

where $J_k = (c(x-1), c(x)]$

$\underline{\text{Remark}}$ $|I_n(\underline{x})| = q(\underline{x}) := \prod_{i=1}^{n} q(x_i)$

Let $m(x) := \min\{m : 2^{-m} \le |I_n(x)|\}$

$$2^{-m(x)} \le |I_n(x)| < 2^{-m(x)+1}$$



$m(x) \ge \log_2 |I|^{-1} \ge m(x)-1$

either one or two points
with $m(x)$ ~~di~~ binary digits

$E(x) :=$ rightmost such point

$\cancel{\mathbb{E}\,\text{length}(x)} = \text{length } \cancel{\text{ne}}\, E(x) = m(x)$

$$\log \frac{1}{q(x)} \le m(x) \le \log \frac{1}{q(x)} + 1$$

$$\log \frac{1}{q(x)} = -n \sum_{z} \hat{p}_x(z) \log q(z)$$

$$= \cancel{H(} \, n\left[H(\hat{p}_x) + D(\hat{p}_x \| q)\right]$$

In particular, can achieve

$$m(x) \le nH(\hat{p}_x) + 1$$

if can store $\hat{p}_x$ separately and use $q = \hat{p}_x$

Or

$$m(x) \le nH(q) + O(\sqrt{n})$$

if $\hat{p}_x(x)$ concentrates around q

(eg iid or MC or _)

Background: Quantization on $\mathbb{R}$

$X \sim p$ probability density function on $\mathbb{R}$

~~$\emptyset$ $S \subset \mathbb{R}$ discrete~~

~~$Q(x)$~~

$\mathbb{R} = \biguplus_{k \in \mathbb{Z}} J_k$ disjoint intervals

$J_k = (b_k, b_{k+1}]$

$q_k \in J_k$

$$Q(x) = \sum_{k \in \mathbb{Z}} q_k \mathbb{1}_{J_k}(x)$$

distortion $\mathbb{E}\{(X - Q(X))^2\}$

for small intervals $p(x) \sim$ unif over the interval

$$\Delta_k = (b_{k+1} - b_k)$$

$$\cancel{p_k \approx p(J_k)} \quad p_k \approx p(x_k)\, \Delta_k$$

$$E\{(X - Q(X))^2\} \approx \sum_{k \in \mathbb{Z}} p_k \cdot \frac{\Delta_k^2}{12}$$

$$\int_{-1/2}^{1/2} x^2\, dx = \frac{x^3}{3}\Big|_{-1/2}^{1/2} = \frac{1}{12}$$

$$\begin{cases} \min & H(p) \\ \text{st} & \frac{1}{12} \sum_{k \in \mathbb{Z}} p_k \Delta_k^2 \le \delta \end{cases}$$

$$H(p) = -\sum_k p(x_k)\, \Delta_k \log(p(x_k)\, \Delta_k)$$

$$\approx -\int \log[p(x) \cdot \Delta(x)]\, p(x)\, dx$$

$$= h(p) - \int \log \Delta(x)\, p(x)\, dx$$

$$\frac{1}{12} \sum_k p_k \Delta_k^2 \approx \frac{1}{12} \int p(x)\, \Delta(x)^2\, dx$$

$$\mathcal{L}(\Delta) = -\int \log \Delta(x)\, p(x)\, dx + \frac{\lambda}{2}\int \Delta(x)^2\, p(x)\, dx$$

$$\Delta(x)^{-1} + \lambda \Delta(x) = 0$$

$$\Delta(x) = \text{const}$$

$$d(R,p) = \frac{1}{12}\Delta^2$$

$$R = h(p) + \log\frac{1}{\Delta}$$

$$\Delta = e^{h(p)-R}$$

$$\boxed{d(R,p) \simeq \frac{1}{12} e^{2h(p)-2R}}$$

~~use R bits~~ Correct as $R \to \infty$

use R bits to ~~$(R = H(p)$~~ encode ~~symbol~~ x $\quad x \sim p$

Transform coding

$$f : \Omega \to \mathbb{R} \qquad |\Omega| = N$$

eg $\Omega = \{0, \frac{T}{N}, \dots \frac{T(N-1)}{N}\}$

$\Omega = \{0, \dots \sqrt{N}\} \times \{0, \dots \sqrt{N}\}$.

$$f = \sum_{n=1}^{N} a_n \varphi_n \qquad a_n \in \mathbb{R}$$

$$E(f) = A([Q(a_1), \dots, Q(a_N)])$$

arithmetic coding

quantize

~~Assume empirical distribution~~.

~~Assume probabilistic model~~

~~Distribution of $\langle f, \varphi_i \rangle$~~

Assume empirical distribution of $a_1 \dots a_N$ $\hat{p}_a \Rightarrow p$ continuous distr.

$$d(R,p) \approx N e^{2h(p) - 2R/N}$$

<u>Which basis should I choose?</u>

Assume prob model

$a_i \sim p_i$ $i \leq N$ not nec independent

use quantiz Δ_i

$$R = \sum_{i=1}^{N} \left[h(p_i) - \log \Delta_i \right]$$

$$d \approx \sum_{i=1}^{N} \Delta_i^2$$

$$\begin{cases} \min \sum_{i=1}^{N} \log 1/\Delta_i \\ \text{st} \sum_{i=1}^{N} \Delta_i^2 \le d \end{cases}$$

optimal choice $\Delta_i = \Delta$

$$d \simeq N\Delta^2 \qquad \frac{R}{N} = \frac{1}{N}\sum_i h(p_i) - \log$$

$\cancel{d \simeq}$

$$\boxed{d \simeq N \exp\left\{ \frac{1}{N}\sum_{i=1}^{N} h(p_i) - \frac{R}{N} \right\}.}$$

Need minimize $\sum_{i=1}^{N} h(p_i)$

IT Classical ineq

$$h(p_{x_1 \ldots x_N}) \le \sum_{i=1}^{N} h(p_{x_i})$$

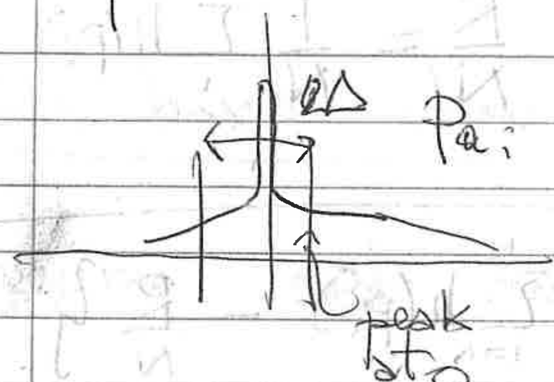
with equality iff indep

$\Rightarrow$ KL transform.

$$p([a, a+\Delta]) \underset{\sim}{\sim} \Delta p(a) \quad \cancel{p(x)\sim} \quad \cancel{p([a_i, b_i}$$

- In reality if rate is not very high the high-rate quantiz assumption does not hold

- In particular in wavelet basis



$$\cancel{p[(-\frac{\Delta}{2}, \frac{\Delta}{2})] \neq \Delta \cdot p(0)} \gg O(\Delta)$$

$$p([-\frac{\Delta}{2}, \frac{\Delta}{2}]) \gg \Delta$$

~~encode~~

$R = R_0 + R_1$ encode the "values of nonzeros" at quant Δ

bits used to encode positions of "non zeros"

Assume # nonzeros $\leq M$

$$R_0 = \log \sum_{\ell=0}^{M} \binom{N}{\ell} \approx \cancel{M \log \frac{N}{M}} \quad \text{for } N \gg M$$

$$= M \log \frac{Ne}{M} + O\left(\frac{M}{N}^2\right)$$

For R_1 will assume a simple model

$|a|_{(k)} \sim C k^{-\alpha}$ ("smoothness α")

$\# \{ k : \ell\Delta < |a_{(k)}| \leq (\ell+1)\Delta \}$

$= \# \{ k : \ell\Delta < C k^{-\alpha} \leq (\ell+1)\Delta$

$$\left(\frac{(\ell+1)\Delta}{C}\right)^{-\frac{1}{\alpha}} < k \leq \left(\frac{\ell\Delta}{C}\right)^{-1/\alpha}$$

$$\approx \frac{1}{\Delta^{1/\alpha}} \left[\frac{1}{\ell^{1/\alpha}} - \frac{1}{(\ell+1)^{1/\alpha}} \right] = \ell^{-1-1/\alpha} \Delta^{-1/\alpha}$$

$$\hat{p}_a(\ell) = \frac{\ell^{-1-1/\alpha}}{C(\alpha)}$$

$$H(\hat{p}_a) = \left(1+\frac{1}{\alpha}\right) \sum_\ell \frac{\ell^{-1+1/\alpha}}{C(\alpha)} \log \ell = C$$

$$R_1 \simeq M$$

$$\boxed{R = M \log \frac{N}{M} + O(M)}$$

dominated by positions of nonzeros!

$$d(R,f) \lesssim \Delta^2 \cdot M + \sum_{k>M} |a_k|^2 = \Delta^2 M + M^{1-2\alpha}$$

Connecting Δ to M

$$M = \#\{k : |a_{(k)}| > \tfrac{\Delta}{2}\}$$

$$\quad = \max\{k : Ck^{-\alpha} > \tfrac{\Delta}{2}\}$$

$$k < C'\Delta^{-1/\alpha}$$

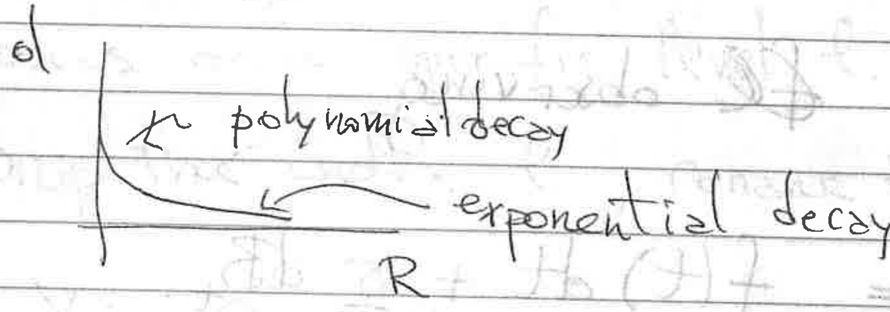
$$M = \Delta^{-1/\alpha} \qquad \Delta \simeq M^{-\alpha} \Rightarrow d(R,f) \simeq M^{1-2\alpha}$$

$$M \simeq \frac{R}{\log \frac{N}{R}}$$

$$d(R,f) \simeq R^{1-2\alpha} \left(\log \frac{N}{R}\right)^{2\alpha-1}$$

[to be compared with

$$\cancel{R}\, d(R,f) \simeq N e^{2h(p) - \frac{R}{N}}$$



at low rate want to minimize $\cancel{R}$ M

=> maximize sparsity] .