III.3. Non-perfect data compression

Thm.: If $C: \mathcal{X}^N \to \{0,1\}^+$ is a code for which $H(X^N) \geq L(C)$, then

$$p_e \geq h^{-1}\left(\frac{H(X^N) - L(C)}{N}\right)$$

where h^{-1} is the inverse of the binary entropy function h on $[0, \frac{1}{2})$, and p_e the average bit error rate after decoding.

proof: define a random variable $Y := C(X^N)$ with range $\mathcal{Y}$ and a code $C': \mathcal{Y} \to \{0,1\}^+$ by $C' = id$. Then

$$L(C) = L(C') \geq H(Y) = H\left(C(X^N)\right) \quad (*)$$

$$H\left(X^N \mid C(X^N)\right) = H\left(X^N, C(X^N)\right) - H\left(C(X^N)\right)$$

$$= H(X^N) - H\left(C(X^N)\right)$$

$$H\left(C(X^N) \mid X^N\right) = 0 \quad \text{and} \quad \geq H(X^N) - L(C) \quad (*)$$

Fano's inequality $\Rightarrow$

$$N h(p_e) \geq H\left(X^N \mid C(X^N)\right) \geq H(X^N) - L(C)$$

□

III.4. Asymptotic equipartition property & typicality

Def.: A sequence of random variables $\{X_i\}_{i \in \mathbb{N}}$ converges to X "in probability" if $\forall \varepsilon > 0 \quad p\{|X_n - X| > \varepsilon\} \to 0$ for $n \to \infty$.

Thm.: (weak law of large numbers)

Let $\{X_i\}_{i \in \mathbb{N}}$ be i.i.d. random variables with mean $E(X_i) = \mu$.

Then $\boxed{\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{n \to \infty} \mu}$ in probability.

Thm.: (asymptotic equipartition property / AEP) Let $\{X_i\}_{i\in\mathbb{N}}$ be i.i.d. random variables with distribution $p(x)$. Then

$$-\frac{1}{n}\log\left(p(X_1,\dots,X_n)\right) \to H(X) \quad \text{in probability.}$$

remark: $p(X)$ means a random variable defined as follows: Let Ω be the sample space, $X:\Omega\to\mathcal{X}$ a r.v. and $p:\mathcal{X}\to[0,1]$, $x\mapsto p(X=x)$. Then $p(X):\Omega\to[0,1]$ is defined as $p(X)=p\circ X$. Hence if μ is the probability measure on Ω, then $p(X):\omega\mapsto\mu\left(\{\omega'\in\Omega \mid X(\omega)=X(\omega')\}\right)$.

proof: $\{X_i\}$ i.i.d. $\Rightarrow$ $\{\log p(X_i)\}$ i.i.d.

Law of large numbers $\Rightarrow$ $\frac{1}{n}\sum_{i=1}^{n}\log(p(X_i)) \to -H(X)$

$$\| \qquad \frac{1}{n}\log\left(p(X_1,\dots,X_n)\right)$$

remark: this can be extended to ergodic stationary stochastic processes. The "Shannon-McMillan-Breiman thm." states that then

$$-\frac{1}{n}\log\left(p(X_1,\dots,X_n)\right) \to H(\{X_i\})$$

Def.: (typical set)

A "typical set" $A_\varepsilon^{(n)} \subseteq \mathcal{X}^n$ w.r.t. to a set of i.i.d. random variables $\{X_i\}_{i\in\mathbb{N}}$ contains all $x\in\mathcal{X}^n$ for which

$$2^{-n(H(X)+\varepsilon)} \le p(x) \le 2^{-n(H(X)-\varepsilon)}$$

Motivation: take a random string $x:=(x_1,\dots,x_n)\in\{1,\dots,k\}^n$

Then $p(x) = \prod_{j=1}^{k} p(X=j)^{n_j} \approx 2^{-nH(X)}$

$\uparrow$

$n_j \approx np(X=j)$

$\to$ we expect a random string to have probability around $2^{-nH(X)}$

Thm.: (properties of sets of typical strings)

1) $x \in A_\varepsilon^{(n)} \iff H(X) - \varepsilon \leq -\frac{1}{n} \log p(x) \leq H(X) + \varepsilon$

2) $p(A_\varepsilon^{(n)}) := p\{X^n \in A_\varepsilon^{(n)}\} > 1 - \varepsilon$ for suff. large n

3) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$

4) $|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$ for suff. large n

proof: 1) from definition

2) $p(A_\varepsilon^{(n)}) = p\left\{\left| -\frac{1}{n} \log p(X) - H(X) \right| \leq \varepsilon\right\}$

$$= 1 - p\left\{\left| -\frac{1}{n} \log p(X) - H(X) \right| > \varepsilon\right\}$$

AEP $\Rightarrow \exists N \in \mathbb{N}\ \forall n > N:\ p\left\{\left| -\frac{1}{n} \log p(X) - H(X) \right| > \varepsilon\right\} < \varepsilon$

3) $1 \geq \sum_{x \in A_\varepsilon^{(n)}} p(x) \geq \sum_{x \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)+\varepsilon)}$

4) from 2) $\Rightarrow \exists N \in \mathbb{N}\ \forall n > N:\ p(A_\varepsilon^{(n)}) > 1 - \varepsilon$

$$\sum_{x \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} \overset{!!}{=} |A_\varepsilon^{(n)}|\, 2^{-n(H(X)-\varepsilon)}$$

□

loosely speaking:

- sequences are typically typical ones
- there are $\sim 2^{nH(X)}$ typical sequences of length n
- each of them occurs with probability $\sim 2^{-nH(X)}$

III.5. Data compression based on AEP

$$\mathcal{X}^n = A_\varepsilon^{(n)} \cup \overline{A_\varepsilon^{(n)}}$$

Define an injective map $C: \mathcal{X}^n \to \{0,1\}^+$ such that

- $x \in A_\varepsilon^{(n)} \Rightarrow C(x) = 0y$ where $y \in \{0,1\}^{\lceil n(H(X)+\varepsilon)\rceil}$
  - remember that $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$
    $\to$ C can be chosen injective on $A_\varepsilon^{(n)}$
  - the prefix "0" indicates that $x \in A_\varepsilon^{(n)}$
- $x \notin A_\varepsilon^{(n)} \Rightarrow C(x) = 1\tilde{x}$ where $\tilde{x} \in \{0,1\}^{\lceil n \log|\mathcal{X}|\rceil}$
  - $\tilde{x} = x$ if $\mathcal{X} = \{0,1\}$
  - the prefix "1" encodes $x \notin A_\varepsilon^{(n)}$

We obtain for the average codeword length:

$$\begin{aligned} L(C) &= \sum_{x \in A_\varepsilon^{(n)}} p(x)\, l(x) + \sum_{x \notin A_\varepsilon^{(n)}} p(x)\, l(x) \\ &\leq \underbrace{p(A_\varepsilon^{(n)})}_{\leq 1} \left(n(H(X)+\varepsilon)+2\right) + \underbrace{\left(1-p(A_\varepsilon^{(n)})\right)}_{\leq \varepsilon} \left(n \log|\mathcal{X}|+2\right) \\ &\leq n\left(H(X)+\varepsilon\right) + 2 + \varepsilon\left(n \log|\mathcal{X}|+2\right) \end{aligned}$$

$\Rightarrow$ **Thm.:** (Shannon's source coding theorem) Let $\{X_i\}_{i\in\mathbb{N}}$ be i.i.d. random variables with range $\mathcal{X}$. $\forall \delta > 0\ \exists n \in \mathbb{N}\ \exists C: \mathcal{X}^n \to \{0,1\}^+$ uniquely decodable:

$$\boxed{\frac{1}{n} L(C) \leq H(X) + \delta}$$