AEP: (X_n) stochastic process, state space $\mathcal{X}$ Then

$$-\frac{1}{n}\log_2 p_n(X_1,\dots,X_n) \to h \text{ almost surely}$$

h as entropy

p_n joint distribution of $X_1,\dots,X_n$

Theorem 0.1 (Ergodic theorem for finite Markov chains)**.** *$(\mathcal{X}, P)$ where P is a transition matrix $P = (p(y|x))_{x,y\in\mathcal{X}}$ irreducible ν unique stationary (invariant) probability distribution: $\nu = (\nu(x))_{x\in\mathcal{X}}$ row vector; $\nu P = \nu$. Then for any initial distribution $\mu = (\mu(x))_{x\in\mathcal{X}}$ a Markov chain $(X_n)_{n\geq 0}$ for every function $f : \mathcal{X} \to \mathbb{R}$,*

$$\frac{f(X_0)+f(X_1)+\cdots+f(X_{n-1})}{n} \xrightarrow{a.s.} \int_{\mathcal{X}} f d\nu (= \sum_{x\in\mathcal{X}} f(x)\nu(x))$$

Theorem 0.2. *Under the above assumptions $(X_n)_{n\geq 0}$ has the AEP.*

Proof. $p_n(x_0,x_1,\dots,x_n) = \mathbb{P}[X_0 = x_0, X_1 = x_1, \dots, X_n = x_n] = \mu(x_0)p(x_1|x_0)p(x_2|x_1)\dots p(x_n|x_{n-1})(= \mu_{x_0}p_{x_0,x_1}p_{x_1,x_2}\dots p_{x_{n-1},x_n}$ We want to study

$$-\frac{1}{n+1}\log_2[\mu(X_0)p(X_1|X_0)\dots p(X_n|X_{n-1})]$$

$$= \frac{1}{n+1}[-\log_2(\mu(X_0)) + (-\log_2 p(X_1|X_0)) + \cdots + (-\log_2(p(X_n|X_{n-1}))$$

Remark. *we can replace $\frac{1}{n+1}$ by $\frac{n}{n+1}\frac{1}{n}$ since $\frac{n}{n+1}$ tends to 1.*

with $f(x,y) = -\log_2 p(y|x)$, then this above

$$\sim \frac{1}{n}(f(X_0,X_1) + f(X_1,X_2) + \cdots + f(X_{n-1},X_n))$$

Look at new Markov chain: $(X_n, X_{n+1})_{n\geq 0}$

$$\mathbb{P}[(X_n,X_{n-1}) = (x_2,y_2)|(X_{n-1},X_n) = (x_1,y_1)] = q(x_2,y_2|x_1,y_1) = \begin{cases} p(y_2|y_1) & x_2 = y_1 \\ 0 & x_2 \neq y_1 \end{cases}$$

State space: $\mathcal{S} = \{(x,y) \in \mathcal{X}^2 : p(y|x) > 0\}$ = oriented edges of the graph of $(\mathcal{X}, P)$. $(\mathcal{S}, Q)$ is irreducible ? stationary distribution $\hat{\nu}$ Try

add pic

$$\mathbb{P}[(X_0,X_1) = (x,y)] = \nu(x)p(y|x)$$

if X_0 has distribution ν.

?: $\hat{\nu}Q \stackrel{?}{=} \hat{\nu}$, $(x,y) \in \mathcal{S}$

$$\sum_{(u,v)\in\mathcal{S}} \hat{\nu}(u,v) \underbrace{((x,y)|(u,v))}_{p(y|x)\text{ if } v=x, 0\text{else}} = \sum_{u:(u,x)\in\mathcal{S}} \widehat{\nu(u,x)} p(y|x)$$

$$= \sum_u \nu(u)p(x|u)p(y|x)$$

$$= \nu(x)p(y|x)\checkmark$$

Check whether $p(y|x)$ is a probability distribution.

$$\sum_{x,y} \nu(x)p(y|x) = \sum_{y} \overbrace{\sum_{x} \nu(x)p(y|x)}^{=\nu(y)} = \sum_{y} \nu(y) = 1$$

Thus,

$$\begin{aligned}\frac{1}{n}(f(X_0, X_1) + f(X_1, X_2) + \cdots + f(X_{n-1}, X_n)) &\to \sum_{(x,y)\in\mathcal{S}} f(x,y)\hat{\nu}(x,y) \\ &= -\sum_{x,y} \nu(x)p(y|x)\log_2 p(y|x) = \sum_{x} \nu(x)H(p(\cdot|x)) = h\end{aligned}$$

$\square$

1 Codes

Data in $\mathcal{X}$ have to be encoded efficiently (and transmitted and decoded). Data comes in with certain probabilities. $p(x) = \mathbb{P}[X = x]$, X random variable, random input. to be encoded by words (strings) over some finite "alphabet" (set) Σ (typically $\Sigma = \{0, 1\}$).

Definition. Source code*: $C : \mathcal{X} \to \Sigma^+$ where $\Sigma^+ = \{a_1 \ldots a_n | n \in \mathbb{N}, a_i \in \Sigma\}$ denotes the set of all non-empty words. $C(x)$ is the codeword of $x \in \mathcal{X}$. $w \in \Sigma^*$: $l(w)$ length (number of letters)*

Expected code length*:*

$$L_C = \sum_{x \in \mathcal{X}} l(C(x))p(x) = \mathbb{E}(l(C(x)))$$

Example. *a)*
$\mathcal{X} = \{1, 2, 3, 4\}$*, and* $\Sigma = \{0, 1\}$.

$$p(1) = \frac{1}{2}$$
$$p(2) = \frac{1}{4}$$
$$p(3) = p(4) = \frac{1}{8}$$

Then

$$C(1) = 0$$
$$C(2) = 10$$
$$C(3) = 110$$
$$C(4) = 111$$

$L_C = \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = \frac{7}{4}$ $(= H(X) = H(p))$

b)
$\mathcal{X} = \{1, 2, 3\}$ $p(\cdot) = \frac{1}{3}$

$$C(1) = 0$$
$$C(2) = 10$$
$$C(3) = 11$$

$L(C) = \frac{5}{3} = 1,66 > 1,58 \approx H(X) = \log_2(3)$

Definition. Extension of C to $\mathcal{X}^+$*:* $\mathcal{X}^+ \to \Sigma^+$ *by*

$$C(x_1 \ldots x_n) = C(x_1)C(x_2) \ldots C(x_n)$$

("concatenation") (remark for me: Semigroup homomorphism, free groups, $\Rightarrow$ unique extension; with $C(\epsilon_{\mathcal{X}} = \epsilon_\Sigma$ we have a monoid-homomorphism)

Definition (Properties of codes)**.** *1. $C : \mathcal{X} \to \Sigma^+$ is called* non-singular *if it is injective*

2. uniquely decodable *if the extension* $C : \mathcal{X}^+ \to \Sigma^+$ *is injective*

3. instantaneous (prefix free; prefix code) *if no codeword* $C(x)$, $x \in \mathcal{X}$ *is a prefix of another codeword [*$\forall x, y \in \mathcal{X} : C(x)$ *is not prefix of* $C(y)$*]*

Example. $\mathcal{X} = \{a, b, c, d\}$ *and* $\Sigma = \{0, 1\}$.

1.
$$\begin{aligned} C(a) &= 0 \\ C(b) &= 010 \\ C(c) &= 01 \\ C(d) &= 10 \end{aligned}$$

   is non-singular, but $C(ad) = C(b)$ *and thus not uniquely decodable*

2.
$$\begin{aligned} C(a) &= 10 \\ C(b) &= 00 \\ C(c) &= 11 \\ C(d) &= 110 \end{aligned}$$

   not instantaneous, but uniquely decodable. Why?

$$C(cbda) = 110011010$$

   Think about it (exercise)