Information theory example

optimal entropy of a prefix code

$$A = \{a_1, \ldots, a_K\}$$

$$\phi : A \longrightarrow X = \bigcup_{n=1,\ldots,N} \{0,1\}^n \qquad \text{binary coding}$$

A has distribution $\{p(a)\}$

$a_1, \ldots, a_{v_r}$ word

$$\phi(a_1, \ldots, a_{v_r}) = \phi(a_1, \ldots, a_{v_{r-1}})\,\phi(a_v)$$

prefix = instantaneously decodable code

$$A = \{a_1, a_2, a_3\}$$

$$\phi(a_1) = 0, \quad \phi(a_2) = 10, \quad \phi(a_3) = 11 \qquad \text{prefix}$$

$$\underline{1}111 = \phi(a_3\, a_3)$$

$$\phi(a_1) = 1, \quad \phi(a_2) = 10, \quad \phi(a_3) = 11 \qquad \text{not prefix}$$

$$\underline{1}111 = \phi(\quad) \qquad \text{decode by context}$$

$$n(a) = \#\text{ of bits in } \phi(a) = |\phi(a)|$$

$$Z = \sum 2^{-n(a)} \leq 1 \qquad \text{Kraft Inequality}$$

relative entropy

$$0 \leq \sum_A p(a) \log_2 \frac{p(a)}{\frac{2^{-n(a)}}{Z}}$$

$$\leq \sum_A n(a)\,p(a) + \sum p(a) \log_2 p(a) \cancel{+ \log_2 Z} \leq 0 \qquad \text{Kraft}$$

equality: $n(a) = -\log_2 p(a)$

Shannon-Wiener information

Markov chain

$X = \{x_1, \dots, x_r\}$

$p^*(x)$, $p(x)$ $\qquad p(x,y)$ joint $\qquad p(y|x) \sim P$

$q^*(x)$, $q(x)$ $\qquad q(x,y)$ joint $\qquad q(y|x) \sim Q$

conditional relative entropy

$$\Phi\left(p(y|x) \| q(y|x)\right) = \sum_{x \in X} p^*(x) \sum_{y \in Y} p(y|x) \log \frac{p(y|x)}{q(y|x)} \geq 0$$

chain rule

$$\Phi\left(p(x,y) \| q(x,y)\right) = \Phi\left(p^* \| q^*\right) + \Phi\left(p(y|x) \| q(y|x)\right)$$

proof

$$\Phi\left(p(x,y) \| q(x,y)\right) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{q(x,y)}$$

$$= \sum p(x,y) \log \frac{p^*(x)\, p(y|x)}{q^*(x)\, q(y|x)}$$

$$= \sum p(x,y) \log \frac{p^*(x)}{q^*(x)} + \sum p(x,y) \log \frac{p(y|x)}{q(y|x)}$$

$$= \Phi\left(p^* \| q^*\right) + \Phi\left(p(y|x) \| q(y|x)\right)$$

Apply this to Markov chain

at step n $\qquad P = \left(p(y|x)\right) = \left(p_{ij}\right)$

$p^{(n)}$, $q^{(n)}$

$$p\left(x^{(n)}, x^{(n+1)}\right) = p^{(n)} P$$

$$q\left(x^{(n)}, x^{(n+1)}\right) = q^{(n)} P$$

Apply chain rule:

$$\Phi\left(p(x^{(n)}, x^{(n+1)}) \| q(x^{(n)}, x^{(n+1)})\right) =$$

$$= \Phi\left(p(x^{(n)}) \| q(x^{(n)})\right) + \Phi\left(p(x^{(n+1)} | x^{(n)}) \| q(x^{(n+1)} | x^{(n)})\right) \quad 0$$

$$= \Phi\left(p(x^{(n+1)}) \| q(x^{(n+1)})\right) + \Phi\left(p(x^{(n)} | x^{(n+1)}) \| q(x^{(n)} | x^{(n+1)})\right)$$

$$\Phi\left(p(x^{(n+1)} | x^{(n)}) \| q(x^{(n+1)} | x^{(n)})\right) = \sum p(x,y) \log \frac{p(y|x)}{p(y|x)} = 0$$

$$\Phi\left(p(x^{(n+1)}) \| q(x^{(n+1)})\right) \leq \Phi\left(p(x^{(n)}) \| q(x^{(n)})\right)$$

or

$$\Phi\left(p^{(n+1)} \| q^{(n+1)}\right) \leq \Phi\left(p^{(n)} \| q^{(n)}\right)$$

relative entropy decreases as move along a Markov chain

Cover + Thomas

$$q^{(n)} = p^{eq}$$

$\Phi\left(p^{(n)} \| p^{eq}\right)$ decreasing with $n \to \infty$

? ergodic $\left(p(x,y) > 0 \Rightarrow p^{eq} \text{ unique}\right)$

decreasing exponentially

large deviation statement   look in Ellis

D + Z

Dupuis + Ellis

$$\boxed{\begin{array}{l}\text{Laplace method} \\ -\log p(\omega) = \frac{1}{n} \text{mm} \\ |\omega| = n\end{array}}$$