Relative entropy

Surprise $(x) := \log \frac{1}{p(x)}$

Huffman coding $E(x)$ has length $\lceil \log \frac{1}{p(x)} \rceil$ possible with prefix-free encoding

e.g.

x	$p(x)$	$E(x)$
a	1/2	0
b	1/4	10
c	1/8	110
d	1/8	111

$\lceil \cdot \rceil \Rightarrow$ some inefficiency which $\to 0$ if we encode blocks of letters

$$H(p) = \mathbb{E}_x [\text{surprise}(x)] = \sum_x p(x) \log \frac{1}{p(x)}$$

wrong codebook

$$\mathbb{E}_{x\sim p} |\mathrm{Enc}_q(x)| = \sum_x p(x) \log \frac{1}{q(x)} \geq \sum_x p(x) \log \frac{1}{p(x)}$$

$$\text{excess} = D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$ relative entropy

claim: $D(p\|q) \geq 0$

fact $1+z \leq e^z \quad \forall z \in \mathbb{R}$ Pf $f(z) = e^z - 1 - z$ is convex, $f(0) = f'(0) = 0$

$\log y \leq y - 1$

$\log \frac{1}{y} \geq 1 - y$

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$\geq \sum_x p(x) \left(1 - \frac{q(x)}{p(x)}\right) = \sum_x p(x) - q(x) = 0$$

$D = 0 \Leftrightarrow \frac{p(x)}{q(x)} = 1$ always $\Leftrightarrow p = q$ (later we'll make this robust)

not a metric, not symmetric, no Δ ineq, etc

Cor $H(p) \leq \log d$

Pf $u = (1/d, \ldots, 1/d)$ $\quad 0 \leq D(p\|u) = \sum_x p(x) (\log p(x) + \log d)$

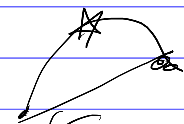
$= \log d - H(p)$

Cor $\quad I(X;Y) \geq 0$

$$D(p_{XY} \| p_X \otimes p_Y) = \sum_{x,y} p_{XY}(x,y)\left(\log p(x,y) - \log p_X(x) - \log p_Y(y)\right)$$

$$= -H(XY) + H(X) + H(Y)$$

$$= I(X;Y) \geq 0$$



Application $\quad$ concavity

$$\sum_x \pi_x H(p_x) \leq H\left(\sum_x \pi_x p_x\right)$$

$$p(x,y) = \pi_x p_x(y) \qquad \overset{\shortparallel}{H(Y|X)} \leq \overset{\shortparallel}{H(Y)}$$

$$H(Y) - H(Y|X) = I(X;Y) \geq 0$$

Hypothesis testing

$x \sim p$ or q

$\alpha = \Pr[\text{guess } q \mid x \sim p] \qquad$ type 1

$\beta = \Pr[\text{guess } p \mid x \sim q] \qquad$ type 2

symmetric $\qquad \min \frac{\alpha+\beta}{2} = \frac{1}{2}\|p-q\|_1$

Bayesian $\qquad \min \pi\alpha + (1-\pi)\beta = \|\alpha p - (1-\pi) q\|_1$

asymmetric $\qquad \beta_\epsilon = \min\{\beta : \alpha \leq \epsilon\}$

we will study $\quad \beta_\epsilon^n = \beta_\epsilon$ for p^n vs q^n $\sim \exp(-nR)$

Chernoff - Stein

$$\lim_{n\to\infty} \frac{-1}{n} \log \beta_\epsilon^n = D(p\|q) \qquad \forall \epsilon \in (0,1)$$

Mostly proved on ps 2

Examples

1) $\quad p = q \Leftrightarrow D(p\|q) = 0 \quad$ i.e. exp. small error always possible

2) $\quad q = U \qquad D(p\|U) = \log d - H(p)$

$\qquad T^n_{p,\delta}$ guess $p \qquad \overline{T^n_{p,\delta}}$ guess U

$$\alpha = p^n\left(\overline{T^n_{p,\delta}}\right) \to 0 \qquad \text{as} \quad n\to\infty$$

$$\beta = u^n\left(T^n_{p,\delta}\right) = \frac{|T^n_{p,\delta}|}{d^n} \le \exp\left(nH(p) + n\delta - n\log d\right)$$
$$= \exp(-n(D(p\|u) - \delta))$$

3) $D(p\|q) = \infty \iff \exists x$ s.t. $p(x) > 0$, $q(x) = 0$

$\iff \operatorname{supp} p \not\subseteq \operatorname{supp} q$

If we see $x \in \operatorname{supp} p - \operatorname{supp} q$, guess p, otherwise q

$$\alpha = p(\operatorname{supp} q)^n \to 0 \qquad \beta = 0$$