INFORMATION THEORY PILLS - ENTROPY

$p(x)$

$-\log p(x)$ - SELF INFORMATION

$$H[p] = \mathbb{E}_p[-\log p(x)] = -\int p(x) \log p(x)\, dx$$

max is Normal R.V.

$$H[p] = -\sum_i p(x_i) \log p(x_i)$$

max for unif. distr. $\log K$.

KULLBACK LEIBLER DIVERGENCE

RELATIVE ENTROPY $\quad p, q$

$$KL[q \| p] = \int q(x) \log \frac{q(x)}{p(x)} dx = -H[q] - \mathbb{E}_q[\log p]$$

$$KL[q \| p] = 0 \text{ iff } q = p$$

KL is a convex functional and $KL \geq 0$

$\rightsquigarrow$ p is fixed but unknown $\quad q = q_\theta$ can vary

what is the best q_θ that approx p.

$$\theta^* = \underset{\theta}{\arg\min}\ KL[q_\theta \| p] \quad \Big\} \text{ on VARIATIONAL INFERENCE}$$

$$I[x, y] = KL[p(x,y) \| p(x)p(y)] \quad \text{MUTUAL INFORMATION}$$

$\underline{x} : x_1, \dots, x_N$

EMPIRICAL DISTRIBUTION $\quad p_{emp}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(x = x_i)$

q

$$KL[p_{emp} \| q] = \underbrace{\mathbb{E}_{p_{emp}}[-\log q(x)]} - H[p_{emp}]$$

$$\underbrace{-\frac{1}{N} \sum \log q(x_i)}$$ CROSS ENTROPY

If $q = q_\theta$ $\qquad -\frac{1}{N} L(\theta) \rightsquigarrow$ MAXIMIZING $L(\theta)$

$\Updownarrow$

MINIMIZING K.L. divergence between p_{emp} and q_θ.