df Relative Entropy (KL-divergence)

$$D(p\|q) = \mathbb{E}_p \log \frac{1}{q} - \mathbb{E}_p \log \frac{1}{p}$$

$\frac{1}{q}$ ← bits used to encode $p(X)$ as $q(X)$

$$= \mathbb{E}_p \log \frac{p}{q}$$

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

$$= \sum_x q(x) \left(\frac{p(x)}{q(x)} \log \frac{p(x)}{q(x)} \right) = D_\phi(p\|q)$$

w/ $\phi = t \log t$

note Suppose that we are estimating $p(X)$ w/ $q(X)$

then MLE is

$$\max_{q \in Q} q(X) = \min_{q \in Q} \overbrace{-\log q(X)}^{\text{loss}}$$

$$\text{Risk}(q) = \mathbb{E}_p \log \frac{1}{q(X)} = \mathbb{E}_p \log \frac{p(X)}{q(X)} + \mathbb{E}_p \log \frac{1}{p(X)}$$

$$= D_{KL}(p\|q) + \text{Risk}(p)$$

thm (Pinsker inequality)

$$D_{TV}(p\|q) \leq \sqrt{\frac{1}{2} D_{KL}(p\|q)}$$

$\Rightarrow D_{KL}(f_+ \| f_-) \to 0$ then $R^* \to \frac{1}{2}$

in [Class]