Information Theory

def φ-divergence (convex) $\quad D_\varphi(p\|q) = \int \varphi\left(\frac{\partial p}{\partial q}\right)\partial q \quad \varphi(1)=0$

note Common measures of distance btw two densities are D_φ, but are not always distance

ex $\varphi(t) = \frac{1}{2}|t-1|$ then $\int \frac{1}{2}\left|\frac{f_1}{f_0} - 1\right| f_0 = \frac{1}{2}\int |f_1 - f_0|$
$=$ total variation divergence $= D_{TV}(f_1\|f_0)$

recall in [Class] $R^* = \frac{1}{2} - \frac{1}{2} D_{TV}(f_+\|f_-)$ $\quad f_+ =$ density of $X|Y=+1$

$\Rightarrow D_{TV}(f_+\|f_-) \to 0$ then $R^* \to \frac{1}{2}$ $\quad f_- =$ " $X|Y=-1$

def for discrete RV, X, w/ pmf p

Entropy $\quad H(X) = \sum_x p(x)\log_2 \frac{1}{p(x)} = -\mathbb{E}_p[\log p(x)]$

Joint Entropy of X, Z $\quad$ (see ★)

$$H(Z|X) = \sum_x p(x)\, H(Z|X=x)$$

thm discrete RV can be encoded into $O(H(X))$ bits w/ diminishing loss

$$= -\sum_x p(x) \sum_z p(z|x)\log_2 p(z|x)$$

$$= -\sum_x \sum_z p(x,z)\log_2 p(z|x)$$

$$= -\mathbb{E}\log\frac{1}{p(z|x)}$$