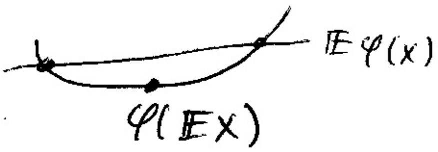
thm (Jensen's Inequality) Let φ be convex, X a R.V.,

$$\varphi(\mathbb{E}X) \le \mathbb{E}\,\varphi(X)$$



thm (Gibbs Information Inequality)

$$^{*}\underline{D_\varphi(p\|q)} = {}^{*}\int \varphi\left(\frac{\partial p}{\partial q}(x)\right) \partial q(x)$$

$$= \mathbb{E}_q\, \varphi\left(\frac{\partial p}{\partial q}(X)\right) \underline{\ge} \varphi\left(\mathbb{E}_q \frac{\partial p}{\partial q}(X)\right)$$

$$= \varphi\left(\int \frac{\partial p}{\partial q}(x)\, \partial q(x)\right) = \varphi(1) = \underline{0}$$

(✱) <u>Source coding</u> is a mapping of random variable $X \in \mathcal{X}$ ($\mathcal{X}$ is finite) to $\{1, \dots, |\mathcal{X}|\}$ w/ binary encoding. Call it $\sigma(X)$ and $\log_2 \sigma(X)$ is length of code.

thm (Shannon Source Coding Thm)

N iid R.V. w/ entropy $H(X)$ can be compressed into more than $NH(X)$ w/ vanishing loss as $N \to \infty$, and compressing into fewer bits will lead to non-negligible information loss.