Kullback-Leibler Divergence

\[ D(p \| q) = \mathbb{E}_p \left[ \log \frac{f(x)}{f_0(x)} \right] \quad \text{if} \; p \; \text{is absolutely continuous w.r.t.} \; q. \]

Chain rule:

\[ D(p(x) \| q(x)) = D(p(x) \| q(x_0)) + \int p(x) D(p(y|x) \| q(y|x)) \, dx. \]

For any prob. measures \( p \) and \( q \).

\[ D(p \| q) = - \int q(x) \log \frac{p(x)}{q(x)} \, dx. \]

Random. M.H. derivative.

As a special case, if \( p \) and \( q \) are a.e. w.r.t. a-finite measure \( \mu \) (Lebesgue measure or counting measure), then

\[ D(p \| q) = \int \frac{p(x)}{q(x)} - 1 \, d\mu. \]

\( \star \) divergence.

Let \( f : [0, \infty) \rightarrow \mathbb{R}^+ \), and \( f(0) = 0 \).

Then, for \( p(x) \) and \( q(x) \), the \( f \)-divergence is defined as

\[ D^f(p \| q) = \mathbb{E}_p \left[ f \left( \frac{p(x)}{q(x)} \right) \right]. \]

Eg: \( f(x) = \log p \), then \( D^f = D \).

Mutual Information

\[ I(X; Y) = D(p(x,y) \| p(x)p(y)) \]

\[ = \frac{1}{p(y)} \int p(x) D(p(y|x) \| p(y)) \, dx. \]

\( I(X; Y) \) is convex in \( p(x,y) \) for a fixed \( p(x|y) \).

\( I(X; Y) \) is concave in \( p(y|x) \) for a fixed \( p(y) \).

Variational characterization

\[ I(X; Y) = \min_{q(y|x)} D(p(y|x) \| q(y|x)) \]

\[ = D(p(y|x) \| q(y|x)) \leq D(p(y|x) \| p(y)) \quad \text{with} \; *= \quad \text{if} \; p = q. \]
Information Capacity

Suppose that $p_{XY}$ is fixed, we are interested in \[
\max_p \min_{q}\ I(X; Y)
\]
\[
\max_p \min_{q}\ I(X; Y) = \max_p \min_{q} \sum_x \sum_y \mathbb{P}(x) \mathbb{P}(y|x) \log \frac{\mathbb{P}(y|x)}{q(y)}
\]

* Min-max theorem:

\begin{align*}
\text{convex in } q(y) & \quad \text{If } A, B \text{ are compact.} \\
\text{linear in } p(x) & \quad \text{if } f \text{ is concave in } x \text{ and convex in } p. \\
\min_{q(y)} \max_p & \quad \text{then } \max_p \min_{q(y)} f(p,Y) = \min_{q(y)} \max_p f(p,Y)
\end{align*}

\[
= \min_{q(y)} \max_p \mathbb{P}(Y) \mathbb{P}(Y|X) \log \frac{\mathbb{P}(Y|X)}{\mathbb{P}_Y(Y)}
\]