Conditional Relative Entropy

$$D(P_{X|Y} \| Q_{X|Y} | P_Y) = \mathbb{E}_{P_Y P_{X|Y}}\left[\log \frac{P_{X|Y}(X|Y)}{Q_{X|Y}(X|Y)}\right]$$

Prop

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z)$$

Chain Rules:

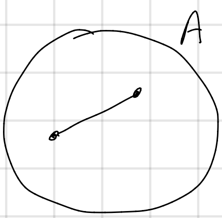
$$H(X_1,\dots,X_n) = \sum_{i=1}^{n} H(X_i|X_1 \cdots X_{i-1})$$

$$I(X_1,\dots,X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_1,\dots,X_{i-1})$$

$$D(P_{XY} \| Q_{XY}) = D(P_X \| Q_X) + D(P_{Y|X} \| Q_{Y|X} | P_X)$$

09/22/2016
Thursday

Convex Set:



$x_1 \in A$, $x_2 \in A$ $\quad \lambda \in [0,1] \quad \lambda x_1 + (1-\lambda) x_2 \in A$

→ convex combination

Convex function:

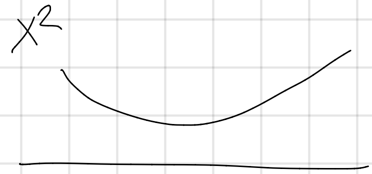
$f : D \to$

$\uparrow$ convex

f is convex if $\forall x_1, x_2 \in D$, $\lambda \in [0,1]$ $\quad f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$

" " strictly convex if equality only occurs when $\lambda = 0$ or 1 or $x_1 = x_2$.

Examples:



If $f''(x) \geq 0 \ \forall x \Rightarrow$ convex

$f''(x) > 0 \ \forall x \Rightarrow$ strictly convex.

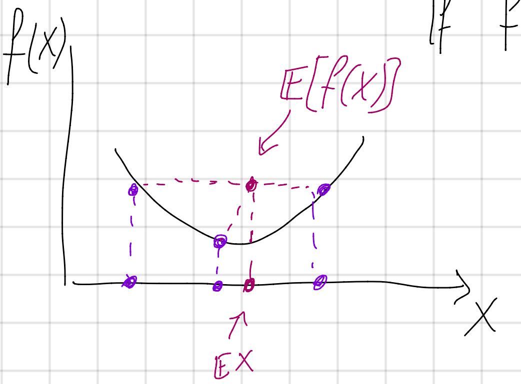
$x \log x$ is convex

Concave function: $-f$ is convex

Jensen's Inequality If f is convex, $\mathbb{E} f(X) \geq f(\mathbb{E}X)$

If f is strictly convex, equality iff X is deterministic

If f is concave



Theorem:

$D(P||Q) \geq 0$ with equality iff $P = Q$

proof:

$$-D(P||Q) = -\mathbb{E}\left[\log \frac{P(X)}{Q(X)}\right]$$

$$= \mathbb{E}\left[\log \frac{Q(X)}{P(X)}\right]$$

$$\leq \log \mathbb{E}\left[\frac{Q(X)}{P(X)}\right]$$

$$= \log \sum_{\substack{x \in \mathcal{X} \\ \mathcal{X}:\, supp(P)}} Q(x) \leq \log 1 = 0$$

Equality iff $\frac{Q(X)}{P(X)}$ is deterministic i.e. $Q(x) = P(x)\ \forall x \in \mathcal{X}$

Theorem $I(X;Y) \geq 0$ w/ equality iff $X \perp\!\!\!\perp Y$

$$I(X;Y) = D(P_{XY} \| P_X P_Y) \geq 0 \quad \text{w/ "=" iff } P_{XY} = P_X P_Y$$
$$\text{iff } X \perp\!\!\!\perp Y.$$

Also

$I(X;Y|Z) \geq 0$ w/ equality iff $X \perp\!\!\!\perp Y | Z$
(conditionally indep given Z)

Theorem $H(X) \leq \log|\mathcal{X}|$ w/ equality iff X is uniformly distributed on $\mathcal{X}$.

proof: Let U be the uniform distribution on $\mathcal{X}$.

$U(x) = \frac{1}{|\mathcal{X}|}\ \forall x \in \mathcal{X}$. Let $X \sim P$

$$D(P\|U) = \mathbb{E}_P\left[\log \frac{P(X)}{U(X)}\right] = \log|\mathcal{X}| - \mathbb{E}\left[\log \frac{1}{P(X)}\right] = \log|\mathcal{X}| - H(X)$$
$$\geq 0 \quad \text{QED.}$$

Theorem

$H(X|Y) \leq H(X)$ w/ equality iff $X \perp\!\!\!\perp Y$

proof $I(X;Y) = H(X) - H(X|Y) \geq 0$

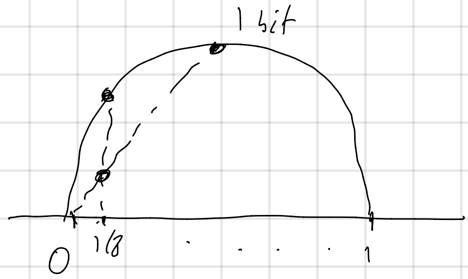
So, "conditioning reduces uncertainty" on average!

Example:

Y \ X	0	1
0	0	3/4
1	1/8	1/8

$P_X(0) = 1/8 \quad P_X(1) = 7/8$

$H(X) = H(1/8)$



$\to H(P)$ concave in P.

$P_{X|Y=0}(0) = 0 \qquad H(P_{X|Y=0}) = 0$ bits

$P_{X|Y=0}(1) = 1 \qquad H(P_{X|Y=1}) = 1$ bit

$$H(X|Y) = \mathbb{E}\left[H\left(P_{X|Y}(\cdot|y)\right)\right]$$

$$\leq H\left(\mathbb{E}\left[P_{X|Y}(\cdot|y)\right]\right) = H(P_X) = H(X)$$

Mut. Info. also average over conditioned variables

i.e. $I(X;Y|Z) = \mathbb{E}\left[I\left(P_{XY|Z=z}\right)\right] \qquad z \sim P_Z$

Theorem: $H(X_1, \dots, X_n) \leq \sum_{i=1}^{n} H(X_i)$ w/ equality if $\{X_i\}$ are independent.

proof Chain rule

$$H(X_1, \dots, X_n) = \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1})$$

$$\leq \sum_{i=1}^{n} H(X_i)$$

Theorem $D(P\|Q)$ is convex in pairs (P, Q)

i.e. $D(\lambda P_1 + (1-\lambda)P_2 \,\|\, \lambda Q_1 + (1-\lambda)Q_2) \leq \lambda D(P_1\|Q_1) + (1-\lambda)D(P_2\|Q_2)$

proof Let P_{XY}, Q_{XY} s.t. P_Y, Q_Y both Bern(λ)

$P_{X|Y=1} = P_1 \quad P_{X|Y=0} = P_2 \quad Q_{X|Y=1} = Q_1 \quad Q_{X|Y=0} = Q_2$

$P_X = \lambda P_1 + (1-\lambda)P_2$

$Q_X = \lambda Q_1 + (1-\lambda)Q_2$

$$D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + D(P_{Y|X}\|Q_{Y|X}|P_X)$$
$$= \underset{0 \ P_Y \| Q_Y}{\cancel{D(P_Y\|Q_Y)}} + \underbrace{D(P_{X|Y}\|Q_{X|Y}|P_Y)}_{= \lambda D(P_1\|Q_1) + (1-\lambda)D(P_2\|Q_2)}$$

Corollary: $D(P\|U) = \log|X| - H(X)$ $\quad D(P\|U)$ is convex in P

$\Rightarrow H(X)$ is concave

Theorem: 1) $I(X;Y)$ is a convex function of P_X when $P_{Y|X}$ is fixed

2) $I(X;Y)$ is a convex function of $P_{Y|X}$ when P_X is fixed.

proof

1) $I(X;Y) = H(Y) - H(Y|X)$

$$= \underbrace{\underset{\substack{\uparrow \\ \text{concave function}}}{H(Y)} - \sum P_X(x) \underset{\hookrightarrow \text{linear function}}{H(P_{Y|X=x})}}_{\to \text{ concave funct.}}$$

2.) $I(X;Y) = D(P_{XY} \| P_X P_Y) \Rightarrow \forall P_{XY} \quad Q_{XY} \quad \lambda \in [0,1]$

$$D(\lambda P_{XY} + (1-\lambda) Q_{XY} \| \lambda P_X P_Y + (1-\lambda) Q_X Q_Y)$$
$$\leq \lambda I(P_{XY}) + (1-\lambda) I(Q_{XY})$$

if $P_X = Q_X$ then
$P_X (\lambda P_Y + (1-\lambda) Q_Y)$

Consider $P_{XY} = P_X P_{Y|X}$

$Q_{XY} = P_X Q_{Y|X}$

$$\overline{P}_{XY} = \lambda P_{XY} + (1-\lambda) Q_{XY} = P_X \, (\lambda P_{Y|X} + (1-\lambda) Q_{Y|X})$$

$$\overline{P}_X = P_X$$

$$\overline{P}_Y = \sum_x P_X(x) \left(\lambda P_{Y|X}(y|x) + (1-\lambda) Q_{Y|X}(y|x)\right)$$

$$= \lambda \Big(\sum_x \underbrace{P_X(x) P_{Y|X}(y|x)}_{P_Y}\Big) + (1-\lambda)\Big(\underbrace{\sum_x P_X Q_{Y|X}}_{Q_Y}\Big)$$

Markovity:

$X \to Y \to Z$ Markov chain

$$P_{XYZ} = P_{XY} P_{Z|Y}$$
$$= P_X P_{Y|X} P_{Z|Y}$$
$$= P_Y P_{X|Y} P_{Z|Y}$$
$$= P_Z P_{Y|Z} P_{X|Y}$$

$X \leftarrow Y \rightarrow Z$

$X \leftarrow Y \leftarrow Z$

Markovity means: Conditional independence: $X \perp\!\!\!\perp Z \mid Y$

$X - Y - Z - W$ means $X \perp\!\!\!\perp (ZW) \mid Y$

$(XY) \perp\!\!\!\perp W \mid Z$

Notice that always $X - Y - f(Y)$ where f is deterministic funct.

Data Processing Inequality

If $X - Y - Z$, $I(X;Y) \geq I(X;Z)$ with equality iff $X - Z - Y$

proof

$$I(X;Y,Z) = I(X;Y) + \cancel{I(X,Z|Y)} \to 0$$

$$= I(X;Z) + \underbrace{I(X,Y|Z)}_{\geq 0}$$

w/ equality iff $X - Z - Y$

QED.

Corollary $I(X;Y) \geq I(X;f(Y))$ w/ equality iff

$X - f(Y) - Y$

proof $X - Y - f(Y)$

↑

suff. statistic of Y for X

$I(X;Y) \quad \lesseqgtr ? \quad I(X;Y|Z)$

↳ None!

Thm If $X - Y - Z$, then $I(X;Y) \geq I(X;Y|Z)$

If $X \perp\!\!\!\perp Z$, then $I(X;Y) \leq I(X;Y|Z)$