August 16, 2017

# Lecture 4: Consequence of Jensen's Inequality

**Thm:** (J.I) If $f(\cdot)$ is a convex function and $X$ is a RV,
$$f(\mathbb{E}(X)) \le \mathbb{E}[f(X)]$$
Moreover, if $f(\cdot)$ is strictly convex, equality implies that $X = E(X)$ i.e., $X$ is a constant.

**Thm:** $D(p\|q) \geq 0$

**Pf:**
$$D(p(x)\,\|\,q(x)) = \sum_{x\in\mathcal{X}_0} p(x)\log\frac{p(x)}{q(x)}$$

adopt convention as part of definition that
$$0\log\frac{0}{0} = 0 \qquad 0\log\frac{0}{q(x)} = 0 \qquad p(x)\log\frac{p(x)}{0} = \infty$$

Let $A = \{x \in \mathcal{X}_0 \mid p(x) > 0\}$

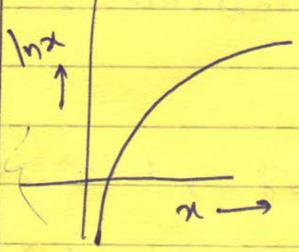
$$\Rightarrow D(p\|q) = \sum_{x\in A} p(x)\log\frac{p(x)}{q(x)}$$

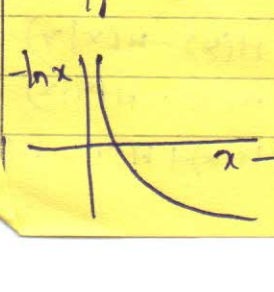
Case ①: Suppose $\exists x \in A$ s.t. $q(x) = 0$
$D(p\|q) = \infty$ (nothing to prove)

Case ②: Assume $q(x) > 0$ for $x \in A$.
Consider the variable $Y$ ~~$= \frac{q(X)}{p(X)}$~~ takes on values ~~with $y =$~~ $\frac{q(x)}{p(x)}$ with prob $p(x)$, $x \in A$.

then
$$\begin{aligned}
D(p\|q) &= \mathbb{E}[-\log Y] \\
&\geq -\log E[Y] \to ① \\
&= -\log \sum_{x\in A} p(x)\frac{q(x)}{p(x)} \\
&= -\log \sum_{x\in A} q(x) = \log\frac{1}{\sum_{x\in A} q(x)} \\
&\geq 0 \quad \left[\text{as } \sum_{x\in A} q(x) \le 1\right]
\end{aligned}$$

Since $-\log(\cdot)$ is strictly convex equality ① holds iff $Y$ = constant i.e., $Y = c$.
$$\Rightarrow \frac{p(x)}{q(x)} = c, \quad x \in A$$
$$\Rightarrow \sum_{x\in A} p(x) = c\sum_{x\in A} q(x) = 1 \;\; (\hookrightarrow \text{by } ②)$$
$\Rightarrow c = 1$ and $p(x) = q(x)$ whenever $p(x) > 0$




$\therefore$ if $D(p\|q) = 0$ then $p(x) = q(x) \;\; \forall x \in \mathcal{X}_0$

**TODAY**

- $D(p\|q) > 0$ → log sum inequality
- $I(X;Y) \geq 0$ → consequences
- $D(P(y|x) \| q(y|x)) \geq 0$
- $I(X;Y|Z) \geq 0$
- $H(X) \leq \log|\mathcal{X}|$
- $H(X) \geq H(X|Y)$
- $H(X_1, \dots X_n) \leq \sum_{i=1}^{n} H(X_i)$ → Excercise: Read!!

$D(p\|q)$ is convex.

$H(p)$ is concave

$I(X;Y)$: concave in $p(x)$, convex in $p(y|x)$ ~~Convex in p(~~ ~~Concave in p(~~

Thm: $I(X;Y) \geq 0$.
with equality iff
X and Y are independant

Pf: $I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

$= D(p(x,y) \| p(x)p(y))$

$\geq 0$

Equality holds iff

$p(x,y) = p(x)p(y)$ i.e., X and Y are independant

Cor: $D_{p(x,y),\, q(x,y)}(p(y|x) \| q(y|x)) \geq 0$

Pf: $\triangleq \sum_x p(x) \sum_y D_{p(y|x),\, q(y|x)}(p(y|x) \| q(y|x)) = \sum_x p(x) \underbrace{\sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)}}_{\geq 0}$

$\geq 0$.

Note: Equality holds when?

$$D_{p(y|x),\, q(y|x)}(p(y|x) \| q(y|x)) = 0 \quad \forall x \in A$$

$$p(y|x) = q(y|x), \quad \forall y, \quad \forall x \in A$$

$p(x) > 0$.

Cor: $I(X;Y|Z) \geq 0$, Equality holds iff X, Y are independant given Z.

Can write

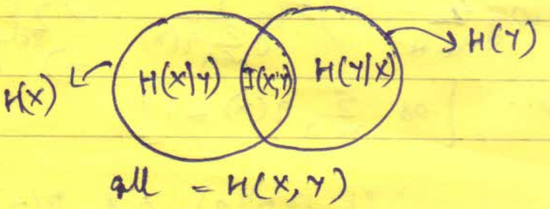
$$I(X;Y|Z) = \sum_z p(z)\, I(X;Y|Z=z)$$

$$= \sum_z p(z)\, D(p(x,y|z) \| p(x|z)p(y|z))$$

$$\geq 0.$$

with equality iff $p(x,y|z) = p(x|z)\,p(y|z)$, $p(z) > 0$

(i.e., given z, X, Y are independant)




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

**Thm:** $H(X) \le \log|\mathcal{X}_0|$ where $X$ takes values from alphabet $\mathcal{X}_0$
Equality holds iff ~~the~~ distribution $P(x)$ is uniform over $\mathcal{X}_0$.

**Pf:** Let $u(x)$ be the uniform distribution:

$$u(x) = \frac{1}{|\mathcal{X}_0|} \qquad x \in \mathcal{X}_0$$

Consider
$$D(p\|u) = \sum_{x\in\mathcal{X}_0} p(x)\log\frac{p(x)}{u(x)} = \sum_{x\in\mathcal{X}_0} p(x)\log p(x) - \sum_{x\in\mathcal{X}_0} p(x)\log\frac{1}{|\mathcal{X}_0|}$$
$$= -H(X) + \sum_{x\in\mathcal{X}_0} p(x)\log|\mathcal{X}_0| = -H(X) + \log|\mathcal{X}_0|$$
$$\ge 0$$
$$\Rightarrow \quad H(X) \le \log|\mathcal{X}_0|$$

Equality holds iff $p(x) = u(x)$, $x \in \mathcal{X}_0$

**Corollary:** $H(X|Y) \le H(X)$
$I(X;Y) = H(X) - H(X|Y) \ge 0$.

**Eg:**

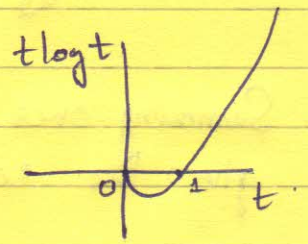
| $Y \backslash X$ | 1 | 2 |
|---|---|---|
| 1 | 0 | 3/4 |
| 2 | 1/8 | 1/8 |

$$H(X,Y) = \frac{3}{4}\log\frac{4}{3} + \frac{2}{8}\log 8$$

$$H(X) = \frac{1}{8}\log 8 + \frac{7}{8}\log\frac{8}{7} = 0.544$$

$H(X|Y=1) = 0 \quad \Rightarrow H(X|Y) = \frac{1}{4} = 0.25$
$H(X|Y=2) = 1$

$$H(X|Y) < H(X)$$




**Thm:** (Log Sum Inequality)
for non-negative numbers $(a_1, a_2, \ldots a_n)$, $b_1, \ldots b_n$

$$\sum_{i=1}^{n} a_i \log\left(\frac{a_i}{b_i}\right) \ge \sigma_a \log\frac{\sigma_a}{\sigma_b} \qquad \text{where } \sigma_a = \sum_{i=1}^{n} a_i,\ \sigma_b = \sum_{i=1}^{n} b_i$$

Equality holds iff $\frac{a_i}{b_i} = c$ a constant for all $i$

**Pf:** Will use the fact that fn $t\log t$ is convex

$$\sum_{i=1}^{n} a_i\log\left(\frac{a_i}{b_i}\right) = \sigma_b \sum_{i=1}^{n}\left(\frac{b_i}{\sigma_b}\right)\left(\frac{a_i}{b_i}\right)\log\left(\frac{a_i}{b_i}\right) \triangleq \sigma_b E[f(Y)] \ge \sigma_b f(E[Y])$$

by Jensens inequality

where $f(t) = t\log t$
and $Y = \frac{a_i}{b_i}$ with prob $\frac{b_i}{\sigma_b}$ $\forall i \in [n]$

$= \sigma_b E[Y] \log E[Y]$ $\quad\quad E[Y] = \sum_{i=1}^{n} \frac{b_i}{\sigma_b}\left(\frac{a_i}{b_i}\right) = \frac{\sum a_i}{\sigma_b} = \frac{\sigma_a}{\sigma_b}$

$= \sigma_b \frac{\sigma_a}{\sigma_b} \log \frac{\sigma_a}{\sigma_b} = \sigma_a \log \frac{\sigma_a}{\sigma_b}$

Equality holds iff $Y$ is a constant (as $f(\cdot) = t\log t$ is strictly convex).

i.e., $\frac{a_i}{b_i} = c \quad \forall i \in [n]$ (c a constant)

**Theorem:** (Convexity of relative entropy)
$D(p\|q)$ is convex in the pair $(p, q)$ i.e., $(p_1, q_1), (p_2, q_2)$ are two pairs of probability mass functions, then:

$$D(\lambda p_1 + (1-\lambda) p_2 \;\|\; \lambda q_1 + (1-\lambda) q_2) \leq \lambda D(p_1\|q_1) + (1-\lambda) D(p_2\|q_2).$$

**Pf:**

$$(\lambda p_1(x) + (1-\lambda) p_2(x)) \log \frac{\lambda p_1(x) + (1-\lambda) p_2(x)}{\lambda q_1(x) + (1-\lambda) q_2(x)}$$

$$\leq \lambda p_1(x) \log \frac{p_1(x)}{q_1(x)} + (1-\lambda) p_2(x) \log \frac{p_2(x)}{q_2(x)}.$$

this follows from log-sum inequality.

where $a_1 = \lambda p_1(x)$, $a_2 = (1-\lambda) p_2(x)$ | $b_1 = \lambda q_1(x)$, $b_2 = (1-\lambda) q_2(x)$

Summing over $x \in \mathcal{X}_0$ on both sides gives the result.

Let $X$ be a r.v that takes values $\frac{p_1(x)}{q_1(x)}, \frac{p_2(x)}{q_2(x)}$ w.p $\frac{q_1(x)\lambda}{\lambda q_1(x) + (1-\lambda) q_2(x)}$ and $\frac{(1-\lambda) q_2(x)}{(1-\lambda) q_2(x) + \lambda q_1(x)}$ respectively.

$E[f(X)] \geq f[E[X]]$

where $f = t\log t$

(Alternately)