#4. $\phi''(x) \geq 0$ then $\phi(x)$ is convex.

pf) $\forall x_1, x_2 \in C$, $x_1 \leq x_2$, $\eta \in (0,1)$

$$\frac{\phi(\eta x_1 + (1-\eta)x_2) - \phi(x_1)}{(1-\eta)(x_2 - x_1)} = \phi'(x_3) \quad \exists x_3 \in (x_1, \eta x_1 + (1-\eta)x_2) \text{ by Mean Value Theorem}$$

$$\frac{\phi(x_2) - \phi(\eta x_1 + (1-\eta)x_2)}{\eta(x_2 - x_1)} = \phi'(x_4) \quad \exists x_4 \in (\eta x_1 + (1-\eta)x_2, x_2) \text{ by M.V.T}$$

$$\frac{\phi'(x_4) - \phi'(x_3)}{x_4 - x_3} = \frac{1}{(x_4 - x_3)} \frac{1}{(x_2 - x_1)} \left(\frac{\phi(x_2) - \phi(\eta x_1 + (1-\eta)x_2)}{\eta} - \frac{\phi(\eta x_1 + (1-\eta)x_2) - \phi(x_1)}{(1-\eta)} \right) = \phi''(c) \geq 0$$

$\exists c \in (x_3, x_4)$ by M.V.T

$$\therefore (1-\eta)\left(\phi(x_2) - \phi(\eta x_1 + (1-\eta)x_2)\right) \geq \eta\left(\phi(\eta x_1 + (1-\eta)x_2) - \phi(x_1)\right)$$

$$\Leftrightarrow \eta\phi(x_1) + (1-\eta)\phi(x_2) \geq \phi(\eta x_1 + (1-\eta)x_2) \quad \blacksquare$$

이제, $D_{KL}(p\|q) \geq 0$ 임을 보이자.

pf) $f(x) = -\ln(x) \Rightarrow f'(x) = -\frac{1}{x} \Rightarrow f''(x) = \frac{1}{x^2} \geq 0 \quad (x > 0)$

$\therefore$ since $f''(x) > 0$, $f(x)$ is convex. Therefore, by Jensen's inequality $(P(I=i) = p_i)$

$$D_{KL}(p\|q) = \mathbb{E}_I\left[\log\left(\frac{p_I}{q_I}\right)\right] = \mathbb{E}_I\left[-\log\left(\frac{q_I}{p_I}\right)\right] \geq -\log\left(\mathbb{E}_I\left[\frac{q_I}{p_I}\right]\right)$$

$$= -\log\left(\sum_{i=1}^{n} p_i \frac{q_i}{p_i}\right) = -\log(1) = 0.$$

$D_{KL}(p\|q) \geq 0 \quad \blacksquare$

#5 $\phi''(x) > 0$ then $\phi(x)$ is strictly convex:

pf) $x_1 < x_2$, $\eta \in (0,1)$

$$\frac{\phi(\eta x_1 + (1-\eta)x_2) - \phi(x_1)}{(1-\eta)(x_2 - x_1)} = \phi'(x_3) \quad \exists x_3 \in (x_1, \eta x_1 + (1-\eta)x_2) \text{ by M.V.T}$$

$$\frac{\phi(x_2) - \phi(\eta x_1 + (1-\eta)x_2)}{\eta(x_2 - x_1)} = \phi'(x_4) \quad \exists x_4 \in (\eta x_1 + (1-\eta)x_2, x_2) \text{ by M.V.T}$$

$$\frac{\phi'(x_4) - \phi'(x_3)}{x_4 - x_3} = \phi''(c) > 0 \quad \exists c \in (x_3, x_4) \text{ by M.V.T}$$

$\Leftrightarrow$ $(1-\eta)(\phi(x_2) - \phi(\eta x_1 + (1-\eta)x_2)) > \eta(\phi(\eta x_1 + (1-\eta)x_2) - \phi(x_1))$

$\Leftrightarrow$ $\eta \phi(x_1) + (1-\eta)\phi(x_2) > \phi(\eta x_1 + (1-\eta)x_2)$: strictly convex! ∎

이제, $D_{KL}(p \| q) > 0$ $(p \neq q)$ 임을 보이자.

$f(x) = -\ln(x) \Rightarrow f''(x) = \frac{1}{x^2} > 0$ $(x > 0)$ 이므로 f는 strictly convex. ∴ By Jensen,

$\phi(E[X]) < E[\phi(X)]$ holds for non constant r.v X.

Consider random variable $\frac{p_I}{q_I}$ where $P(I=i) = p_i$. since $p \neq q$, $\frac{p_I}{q_I}$ is a non constant r.v

$$\therefore D_{KL}(p \| q) = E_I\left[-\log \frac{q_I}{p_I}\right] > -\log\left(E_I\left[\frac{q_I}{p_I}\right]\right) = -\log\left(\sum_{i=1}^{n} p_i \cdot \frac{q_i}{p_i}\right) = -\log 1$$

$$= 0$$

$\therefore D_{KL}(p \| q) > 0$

∎

#6 $f_\theta(x) = \mathbb{u}^T \sigma(\mathbb{a}x+\mathbb{b}) = \sum_{i=1}^{p} u_j \sigma(a_j x + b_j)$

$\dfrac{\partial f_\theta(x)}{\partial u_i} = \sigma(a_j x + b_j)$ 이므로, $\nabla_u f_\theta(x) = (\sigma(a_1 x+b_1), \cdots, \sigma(a_p x+b_p)) = \underline{\sigma(\mathbb{a}x+\mathbb{b})}$

$\dfrac{\partial f_\theta(x)}{\partial b_i} = u_j \sigma'(a_j x+b_j)$ 이므로, $\nabla_b f_\theta(x) = (u_1 \sigma'(a_1 x+b_1), \cdots, u_p \sigma'(a_p x+b_p))$

$$= \sigma'(\mathbb{a}x+\mathbb{b}) \odot \mathbb{u} = \begin{pmatrix} a_1x+b_1 & & & 0 \\ & a_2x+b_2 & & \\ & & \ddots & \\ 0 & & & a_px+b_p \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix} = \mathrm{diag}(\sigma'(\mathbb{a}x+\mathbb{b}))\,\mathbb{u}$$

$\dfrac{\partial f_\theta(x)}{\partial a_i} = u_j \sigma'(a_j x+b_j)\, x$ 이므로, $\nabla_a f_\theta(x) = x(u_1 \sigma'(a_1 x+b_1), \cdots, u_p \sigma'(a_p x+b_p))$

$$= (\sigma'(\mathbb{a}x+\mathbb{b}) \odot \mathbb{u})\, x = x \begin{pmatrix} a_1x+b_1 & & & 0 \\ & a_2x+b_2 & & \\ & & \ddots & \\ 0 & & & a_px+b_p \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix} = (\mathrm{diag}(\sigma'(\mathbb{a}x+\mathbb{b}))\,\mathbb{u})\, x$$