# *Note of CD*

$$\boxed{p_i(x \mid \theta) \overset{def}{=} p(X_i = x \mid \theta) = \frac{1}{Z_i(\theta)} e^{\phi_i(x,\theta)},}$$

==>

$$\nabla_\theta p_i(x) \log p_j(x) = p_i(x) \nabla_\theta \log p_j(x) + \log p_j(x) p_i(x) \nabla_\theta \log p_i(x),$$

and

$$\boxed{\nabla_\theta \log p_i(x) = \Psi_i(x) - \mathbb{E}[\Psi_i(X_i)],}$$

==>

$$\begin{aligned}\nabla_\theta p_i(x) \log p_j(x) = p_i(x)(\Psi_j(x) - \mathbb{E}[\Psi_j(X_j)] \\ + \log p_j(x) (\Psi_i(x) - \mathbb{E}[\Psi_i(X_i)])).\end{aligned}$$

==>

$$\nabla_\theta \int p_i(x) \log p_j(x) dx \quad \text{H(Xi,Xj)}$$

$$\begin{aligned}
&= \int p_i(x) \Psi_j(x) dx - \mathbb{E}[\Psi_j(X_j)] + \int p_i(x) \log p_j(x) \Psi_i(x) dx \\
&\quad - \int p_i(x) \log p_j(x) dx \, \mathbb{E}[\Psi_i(X_i)] \\
&= \mathbb{E}[\Psi_j(X_i)] - \mathbb{E}[\Psi_j(X_j)] \\
&\quad + \mathbb{E}[\log p_j(X_i) \Psi_i(X_i)] - \mathbb{E}[\log p_j(X_i)] \mathbb{E}[\Psi_i(X_i)] \\
&= \mathbb{E}[\Psi_j(X_i)] - \mathbb{E}[\Psi_j(X_j)] + \mathbb{C}\mathrm{ov}[\log p_j(X_i), \Psi_i(X_i)] \\
&= \mathbb{E}[\Psi_j(X_i)] - \mathbb{E}[\Psi_j(X_j)] + \mathbb{C}\mathrm{ov}[\phi_j(X_i), \Psi_i(X_i)];
\end{aligned}$$

==>

$$\begin{aligned}\nabla_\theta \mathbb{D}(X_i, X_j \mid \theta) = \mathbb{E}[\Psi_j(X_j)] - \mathbb{E}[\Psi_j(X_i)] \\ + \mathbb{C}ov[\phi_i(X_i) - \phi_j(X_i), \Psi_i(X_i)],\end{aligned}$$

*Hinton's CD statistic*

$$H_t \stackrel{def}{=} \mathbb{E}[\Psi_\infty(\xi)] - \mathbb{E}[\Psi_\infty(X_t)],$$

**Theorem 2.** the $p(\xi \mid \theta)$ is constant wrt $\theta$. $\Psi_\xi(u) = \nabla_\theta \phi(u, \theta) = 0$.
*Let $\{X_t : t \in \mathcal{R}+\}$ be a collection of rvs with distribution*

$$p_t(x_t \mid \theta) \propto e^{\phi_t(x,\theta)}.$$

*Then*

$$\boxed{\nabla_\theta \left(\mathbb{D}(\xi, X_\infty)\right) - \mathbb{D}(X_t, X_\infty)) = -(H_t + \tilde{R}_t),}$$

*and the residual $\tilde{R}_t$ is a covariance statistic,*

$$\tilde{R}_t \stackrel{def}{=} \mathbb{C}ov[\phi_t(X_t) - \phi_\infty(X_t)\ ,\ \Psi_t(X_t)].$$