

$$\begin{aligned}
& p(x|\theta) \\
& \mathbb{E}_{p(x|\theta)} \underbrace{\frac{\partial \log p(x|\theta)}{\partial \theta}}_{\text{score function}} = \int p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta} dx = \\
& = \int p(x|\theta) \frac{1}{p(x|\theta)} \frac{\partial p(x|\theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int p(x|\theta) dx = 0 \\
& \frac{\partial^2 \log p(x|\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} \log p(x|\theta) = \\
& = \frac{\partial}{\partial \theta} \frac{1}{p(x|\theta)} \frac{\partial p(x|\theta)}{\partial \theta} = -\frac{1}{p(x|\theta)^2} \left(\frac{\partial p(x|\theta)}{\partial \theta} \right)^2 + \frac{1}{p(x|\theta)} \frac{\partial^2}{\partial \theta^2} \log p = \\
& = -\left(\frac{\partial \log p(x|\theta)}{\partial \theta} \right)^2 + \underbrace{\frac{1}{p(x|\theta)} \frac{\partial^2}{\partial \theta^2} \log p(x|\theta)}_{\mathbb{E}_p=0}
\end{aligned}$$

$$\mathbb{E} \frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} = -\mathbb{E} \left(\frac{\partial \log p(x|\theta)}{\partial \theta} \right)^2 = -F(\theta)$$

$$p(x|\theta) = \frac{f(x)}{g(\theta)} \exp(\theta^T u(x))$$

$$\begin{aligned}
\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} &= \frac{\partial^2}{\partial \theta^2} (\log f(x) + \theta^T u(x) - \log g(\theta)) = \\
&= -\frac{\partial^2 \log g(\theta)}{\partial \theta^2} = -F(\theta)
\end{aligned}$$

$$\text{grad} f(x) \propto \arg \max_{\Delta x} f(x + \Delta x) \text{ subj. } \|\Delta x\|_E < \varepsilon$$

$$\begin{aligned}
p(x, z, \theta) &= p(x, z|\theta) p(\theta) = \\
&= \underbrace{p(x|z, \theta)} \underbrace{p(z|\theta) p(\theta)}
\end{aligned}$$

$$p(x, z, \theta) = \prod_{i=1}^n p(x_i, z_i|\theta) p(\theta)$$

$$x = (x_1, \dots, x_n)$$

$$p(z, \theta | x)$$

- сопр. на θ при изв. z_i

- сопр. на z при изв. θ

$$p(z, \theta \mid x) \approx q(\theta)q(z) = \arg \min_q \text{KL}(q(\theta)q(z) \parallel p(\theta, z \mid x))$$

$$\begin{aligned} \text{let } p(x, z \mid \theta) &= f(\theta)p(x, z) \exp(\theta^T h(x, z)) \\ p(\theta) &= (f(\theta))^{\nu_0} \exp(\theta^T \eta_0) / g(\nu_0, \eta_0) \end{aligned}$$

$$\begin{aligned} \log q(z) &= \mathbb{E}_{q(\theta)} \log p(x, z, \theta) + \text{const} = \\ &= \mathbb{E}_{q(\theta)} \left(\sum_{i=1} \log f(\theta) + \log p(x_i, z_i) + \theta^T h(x_i, z_i) + \nu_0 \log f(\theta) + \theta^T \eta_0 \right) + \text{const} = \\ &= \sum \log p(x_i, z_i) + \mathbb{E} \theta^T \log p(x_i, z_i) + \text{const} = \sum q_i(z) \end{aligned}$$

$$\begin{aligned} \log q(\theta) &= \mathbb{E}_{q(z)} \log p(x, z, \theta) + \text{const} = \\ &= \mathbb{E}_{q(z)} \left(\sum \log f(\theta) + \log p(x_i, z_i) + \theta^T h(x_i, z_i) + \nu_0 \log f(\theta) + \theta^T \eta_0 \right) + \text{const} = \\ &= n \log f(\theta) + \theta^T \sum_i \mathbb{E} h(x_i, z_i) + \nu_0 \log f(\theta) + \theta^T \eta_0 + \text{const} = \\ &= (n + \nu_0) \log f(\theta) + \theta^T \left(\eta_0 + \sum_{i=1}^n \mathbb{E} h(x_i, z_i) \right) + \text{const} \end{aligned}$$

$$\nu = \nu_0 + n$$

$$\eta = \eta_0 + \sum_i \mathbb{E} h(x_i, z_i)$$

$$\begin{aligned} x &= (x_1, \dots, x_n) \\ x \gg 1 &\Rightarrow \text{stop bayes, use ml} \end{aligned}$$

$$n \gg 1, d \gg 1 \Rightarrow \text{still bayes, ok}$$

SVI to the battle!

$$\begin{aligned}
\mathcal{L}(\eta) &= \int q(\theta | \eta) q(z) (\log p(x, z | \theta) + \log p(\theta) - \log q(\theta | \eta) - \log q(z) \xrightarrow{\text{const}}) d\theta dz = \\
&= \int q(\theta | \eta) q(z) \left[\left[\sum_{i=1}^n \cancel{f(\theta)} + \log p(x_i, z_i) \right] \xrightarrow{\text{const}} + \theta^T h(x_i, z_i) \right] + \\
&\quad \cancel{\nu_0 \log f(\theta)} + \theta^T \eta_0 - \log g(\nu_0, \eta_0) \xrightarrow{\text{const}} - (\nu_0 + n) \log f(\theta) - \\
&\quad - \theta^T n + \log g(\nu_0 + n, \eta)] dz d\theta = \\
&= \int q(\theta | \eta) [\theta^T (\eta_0 - \eta + \sum_{i=1}^n \mathbb{E}_z h(x_i, z_i))] d\theta + \log g(\nu_0 + n, \eta) + \text{const} = \\
&= \underbrace{(\mathbb{E}_{\theta|\eta} \theta)}_{= \frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta)} [\theta^T (\eta_0 - \eta + \sum_{i=1}^n \mathbb{E}_z h(x_i, z_i))] + \log g(\nu_0 + n, \eta)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \eta} &= \frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta) + (\eta_0 - \eta + \sum_{i=1}^n \mathbb{E} \log p(x_i, z_i)) \cdot \\
&\quad \cdot \underbrace{\frac{\partial^2}{\partial \eta^2} \log g(\nu_0 + n, \eta)}_{F(\eta)} - \frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta) = \frac{\partial \mathcal{L}}{\partial \eta} \\
\text{nat grad } \mathcal{L} &= \cancel{F'} (\eta_0 - \eta + \sum_{i=1}^n \mathbb{E}_z h(x_i, z_i)) \approx \\
&\approx \eta_0 - \eta + n \mathbb{E}_{z_j} h(x_j, z_j)
\end{aligned}$$

- $j \sim U\{1, \dots, n\}$
- $q_j(z_j)$
- $\eta_{t+1} = \eta_t + \varkappa_t (\eta_0 - \eta + n \mathbb{E}_{z_j} h(x_j, z_j))$

стох. вар. вывод // Блей, 2011 @ LDA