$$p(x|\theta)$$

$$\mathbb{E}_{p(x|\theta)} \underbrace{\frac{\partial \log p(x \mid \theta)}{\partial \theta}}_{\text{score function}} = \int p(x \mid \theta) \frac{\partial \log p(x \mid \theta)}{\partial \theta} dx =$$

$$= \int p(x \mid \theta) \frac{1}{p(x \mid \theta)} \frac{\partial p(x \mid \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int p(x \mid \theta) dx = 0$$

$$\frac{\partial^2 \log p(x \mid \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} \log p(x \mid \theta) =$$

$$= \frac{\partial}{\partial \theta} \frac{1}{p(x \mid \theta)} \frac{\partial p(x \mid \theta)}{\partial \theta} = -\frac{1}{p(x \mid \theta)^2} \left(\frac{\partial p(x \mid \theta)}{\partial \theta} \right)^2 + \frac{1}{p(x \mid \theta)} \frac{\partial^2}{\partial \theta^2} \log p =$$

$$= -\left(\frac{\partial \log p(x \mid \theta)}{\partial \theta} \right)^2 + \underbrace{\frac{1}{p(x \mid \theta)} \frac{\partial^2}{\partial \theta^2} \log p(x \mid \theta)}_{\mathbb{E}_p = 0}$$

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

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

$$\frac{\partial^2 \log p(x \mid \theta)}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2} (\log f(x) + \theta^T u(x) - \log g(\theta)) =$$

$$= -\frac{\partial^2 \log g(\theta)}{\partial \theta^2} = -F(\theta)$$

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

$$p(x, z, \theta) = p(x, z \mid \theta) p(\theta) =$$

$$= \underbrace{p(x \mid z, \theta)} \underbrace{p(z \mid \theta) p(\theta)}$$

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

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

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

- сопр. на θ при изв. z_i

- сопр. на z при изв. θ

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

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

$$\log q(z) = \mathbb{E}_{q(\theta)} \log p(x, z, \theta) + \text{const} =$$
$$= \mathbb{E}_{q(\theta)}\left(\sum_{i=1} \log f(\theta) + \log p(x_i, z_i) + \theta^T h(x_0, z_i) + \nu_0 \log f(\theta) + \theta^T \eta_0\right) + \text{const} =$$
$$= \sum \log p(x_i, z_i) + \mathbb{E}\theta^T \log p(x_i, z_i) + \text{const} = \sum q_i(z)$$

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

$$\nu = \nu_0 + n$$
$$\eta = \eta_0 + \sum_i \mathbb{E}h(x_i, z_i)$$

$$x = (x_1, \ldots, x_n)$$
$$x >> 1 \Rightarrow \text{stop bayes, use ml}$$

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

SVI to the battle!

$$\mathcal{L}(\eta) = \int q(\theta \mid \eta) q(z)(\log p(x, z \mid \theta) + \log p(\theta) - \log q(\theta \mid \eta) - \cancelto{\text{const}}{\log q(z)})d\theta dz =$$

$$= \int q(\theta \mid \eta) q(z)[[\sum_{i=1}^{n} \cancel{f(\theta)} + \cancelto{\text{const}}{\log p(x_i, z_i)} + \theta^T h(x_i, z_i)] +$$

$$\cancel{\nu_0 \log f(\theta)} + \theta^T \eta_0 - \cancelto{\text{const}}{\log g(\nu_0, \eta_0} - \cancel{(\nu_0 + n) \log f(\theta)} -$$

$$-\theta^T n + \log g(\nu_0 + n, \eta)]dz d\theta =$$

$$= \int q(\theta \mid \eta)[\theta^T(\eta_0 - \eta + \sum_{i=1} \mathbb{E}_z h(x_i, z_i))]d\theta + \log g(\nu_0 + n, \eta) + \text{const} =$$

$$= \underbrace{(\mathbb{E}_{\theta \mid \eta} \theta)}_{= \frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta)} [\theta^T(\eta_0 - \eta + \sum_{i=1} \mathbb{E}_z h(x_i, z_i))] + \log g(\nu_0 + n, \eta)$$

$$\frac{\partial}{\partial \eta} = \cancel{\frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta)} + (\eta_0 - \eta + \sum_{i=1}^{n} \mathbb{E} \log p(x_i, z_i)) \cdot$$

$$\cdot \underbrace{\frac{\partial^2}{\partial \eta^2} \log g(\nu_0 + n, \eta)}_{F(\eta)} - \cancel{\frac{\partial}{\partial \eta} \log g(\nu_0 + n, \eta)} = \frac{\partial \mathcal{L}}{\partial \eta}$$

$$\text{nat grad} \mathcal{L} = \cancel{F^{-1}} \cancel{F} (\eta_0 - \eta + \sum_{i=1}^{n} \mathbb{E}_z h(x_i, z_i)) \approx$$

$$\approx \eta_0 - \eta + n \mathbb{E}_{z_j} h(x_j, z_j)$$

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

стох. вар. вывод // Блей, 2011 @ LDA