$f(x)$	stoch. grad
$\sum_{i=1}^{n} f_i(x)$	$n\frac{\partial f_j(x)}{\partial x} \quad j \sim U\{1,\dots,n\}$
$\int p(y)g(x,y)dy$	$\frac{\partial}{\partial x} g(x,\hat{y}) \quad \hat{y} \sim p(y)$
$\int p(y)\sum_{i=1}^{n} g_i(x,y)dy$	$n\frac{\partial g_j(x,\hat{y})}{\partial x}\hat{y} \quad j \sim U\{1,\dots,n\}$
$\int p(y \mid x)g(x,y)dx$	reinforce: $\left(\frac{\partial}{\partial x} + \frac{\partial \log p(\hat{y}\mid x)}{\partial x}\right) g(x,\hat{y})$

$$\theta \in \mathbb{R}^d, n \gg 1, d \gg 1$$

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

$$x = (x_1,\dots,x_n), p(\theta \mid x) = q(\theta \mid \phi) =$$

$$= \arg\min_{\phi} \mathrm{KL}(q(\theta \mid \phi) \| p(\theta \mid X)) =$$

$$= \arg\max_{\phi} \mathcal{L}(\phi)$$

$$\mathcal{L}(\phi) = \int q(\theta\|\phi) \log \frac{p(x,\theta)}{q(\theta \mid \phi)} d\theta$$

$$\frac{\partial}{\partial\phi}\mathcal{L}(\phi) = \frac{\partial}{\partial\phi}\int q(\theta\|\phi)[\log p(x \mid \theta) + \log p(\theta) - \log q(\theta \mid \phi)] =$$

$$= \int q(\theta \mid \phi)\frac{\partial \log q(\theta \mid \phi)}{\partial\phi}[\sum_{i=1}^{n}\log p(x_i \mid \theta) + \log\frac{p(\theta)}{q(\theta \mid \phi)}]d\theta -$$

$$- \cancelto{0}{\int q(\theta \mid \phi)\frac{\partial \log q(\theta \mid \phi)}{\partial\phi} d\theta} \quad \approx$$

$$\approx \frac{\partial \log q(\theta \mid \phi)}{\partial\phi}[\sum_{i=1}^{n}\log p(x_i \mid \hat{\theta}) + \log\frac{p(\hat{\theta})}{q(\hat{\theta} \mid \phi)}] \quad (\text{where } \hat{\theta} \sim q(\theta \mid \phi)) \approx$$

$$\approx \frac{\partial \log q(\hat{\theta} \mid \phi)}{\partial\phi}\left[n\log p(x_j \mid \hat{\theta}) + \log\frac{p(\hat{\theta})}{q(\hat{\theta} \mid \phi)}\right]; \quad j \sim U\{1,\dots,n\}$$

var.red. (control variates)

$$\frac{\partial \log q(\hat{\theta} \mid \phi)}{\partial\phi}\left(n\log p(x_j \mid \hat{\theta}) + \log\frac{p(\hat{\theta})}{q(\hat{\theta} \mid \phi)} + b\right)$$

$$b, b(x_j), b(x_j,\phi)$$

$$p(t, \theta \mid x, \Lambda) = p(t \mid x, \theta, \Lambda) p(\theta \mid \Lambda) = \frac{\prod \mathcal{N}(\theta_j \mid 0, \lambda_j^2)}{1 + \exp(-t\theta^T x)}$$

$$t \in \{-1, +1\} \quad x \in \mathbb{R}^d \quad \theta \in \mathbb{R}^d$$

$$\Lambda^* = \arg\max_{\Lambda} p(T \mid X, \Lambda) =$$

$$= \arg\max_{\Lambda} \int \prod_{i=1}^{n} p(t_i \mid x_i, \theta) p(\theta \mid \Lambda) d\theta$$

$$\log p(T \mid X, \Lambda) \geq \int q(\theta \mid \phi) \log \frac{p(T \mid X, \theta) p(\theta \mid \Lambda)}{q(\theta \mid \phi)} d\theta \to \max_{\Lambda, \phi}$$

$$\int q(\theta \mid \phi) \log p(T \mid X, \theta) d\theta - \underbrace{\int q(\theta \mid \phi) \log \frac{q(\theta \mid \phi)}{p(\theta || \Lambda)} d\theta}_{\mathrm{KL}(q(\theta|\phi) \| p(\theta||\Lambda))} =$$

$$= \sum_{i=1}^{n} \int q(\theta \mid \phi) \log p(t_i \mid x_i, \theta) d\theta - \sum_{j=1}^{d} \mathrm{KL}(\mathcal{N}(\theta_j \mid \mu_j, \sigma_j^2) \| \mathcal{N}(\theta_j \mid 0, \lambda_j^2)) \to \max_{\mu, \sigma, \Lambda}$$

$$\mathrm{KL}(q \| p) = \log \frac{\lambda_j}{\sigma_j} + \frac{\sigma_j^2 + \mu_j^2}{2\lambda_j^2}$$