LEC06

# Approximate inference, part I

$$p(z|x) = \frac{p(x|z)\,p(z)}{p(x)}$$

$$= \frac{p(x,z)}{p(x)}$$

$$= \frac{p(x,z)}{\int p(x,z')\,dz'}$$

easy

hard :-c

## I. Variational inference

Approximate $p(z|x)$ with some variational distribution $q_\phi(z)$.

$$KL\left(q_\phi(z)\,||\,p(z|x)\right)$$

Reverse KL

$$= \mathbb{E}_{q_\phi(z)}\left[\log \frac{q_\phi(z)}{p(z|x)}\right]$$

Can't evaluate since we don't know $p(z|x)$!

$$= \mathbb{E}_{q_\phi(z)}\left[\log q_\phi(z)\right] - \mathbb{E}_{q_\phi(z)}\left[\log p(z|x)\right]$$

$$= \mathbb{E}_{q_\phi(z)}\left[\log q_\phi(z)\right] - \mathbb{E}_{q_\phi(z)}\left[\log p(x,z) - \log p(x)\right]$$

$$= \mathbb{E}_{q_\phi(z)}\left[\log q_\phi(z) - \log p(x,z)\right] + \log p(x)$$

$$= \underbrace{- \mathbb{E}_{q_\phi(z)}\left[\log \frac{p(x,z)}{q_\phi(z)}\right]}_{\text{ELBO}(\phi)} + \log p(x)$$

$\Rightarrow \min\limits_{\phi} KL\left(q_\phi(z) \| p(z|x)\right)$

is equivalent to $\max\limits_{\phi} \text{ELBO}(\phi)$

$$\boxed{\begin{aligned} \text{ELBO}(\phi) &= \mathbb{E}_{q_\phi(z)}\left[\log \frac{p(x,z)}{q_\phi(z)}\right] \\ &= \mathbb{E}_{q_\phi(z)}\left[\log p(x,z)\right] + \mathbb{H}\left[q_\phi(z)\right] \end{aligned}}$$

automatic differentiation variational inf.

## II. ADVI

How to choose q ?

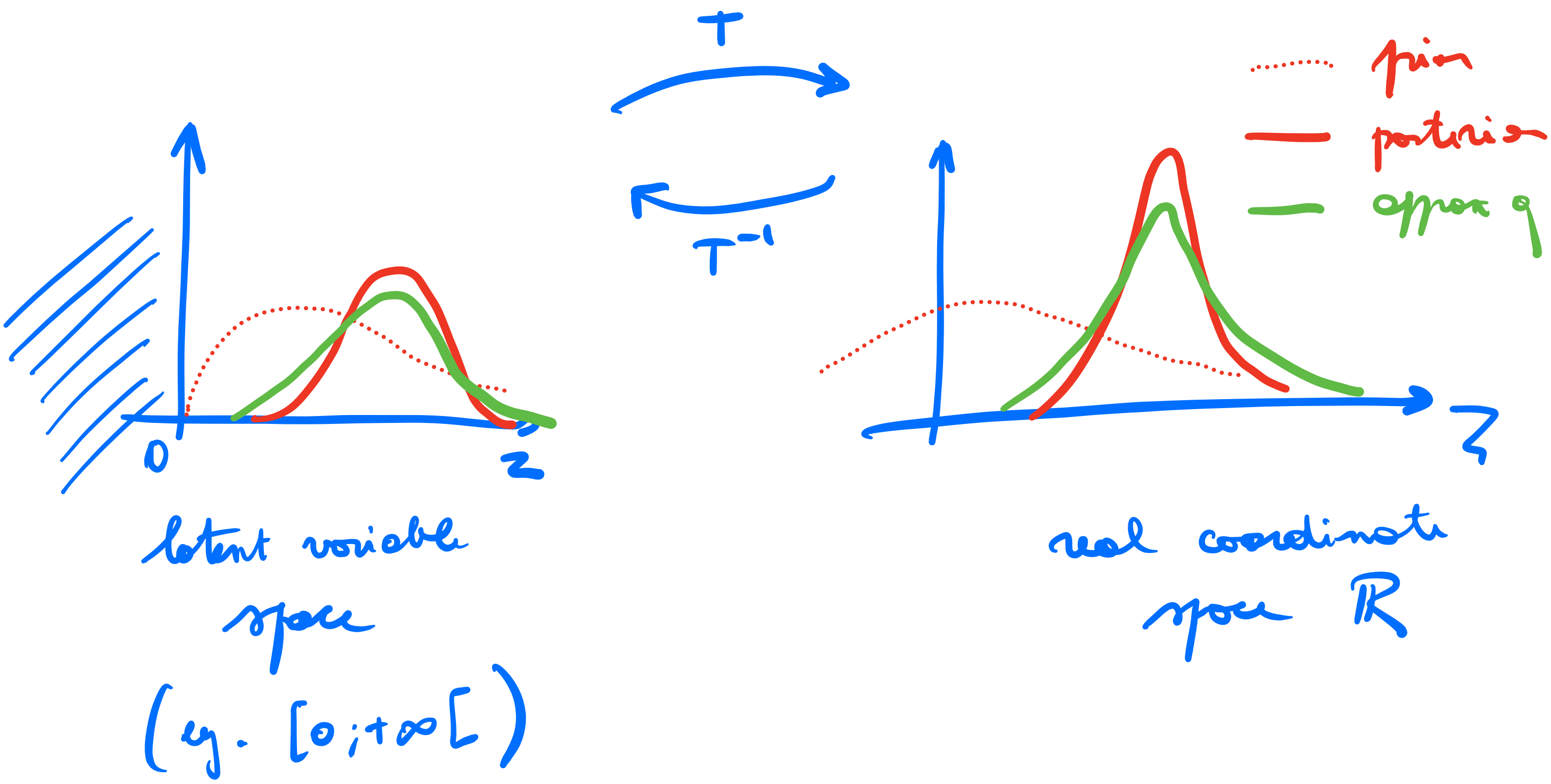
1) Transform the support of the latent variable z such that they live in $\mathbb{R}^K$.

$$T : \text{support}(p(z)) \longmapsto \mathbb{R}^K$$

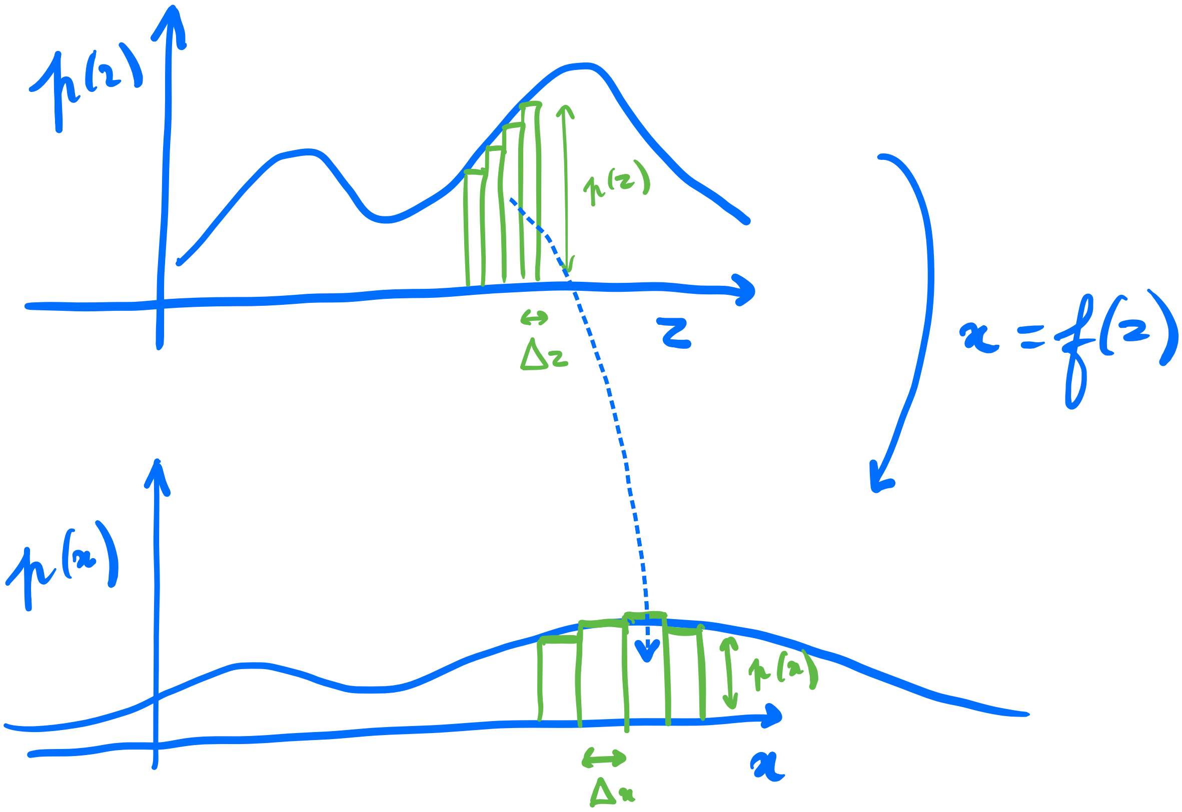
$$\zeta = T(z)$$

$$\Rightarrow p(x, \zeta) = p(x, T^{-1}(\zeta)) \left| \det J_{T^{-1}}(\zeta) \right|$$

change of variable thm



# Change of variable theorem:




Conservation de la densité

$$p(z)\,\Delta z = p(x)\,\Delta x$$

$$\rightarrow p(x) = p(z)\,\frac{\Delta z}{\Delta x}$$

$z = f^{-1}(x)$

$$= p(f^{-1}(x))\,\frac{\Delta f^{-1}}{\Delta x}$$

$\Delta x \to 0$

$$= p(f^{-1}(x))\left|\frac{\partial f^{-1}}{\partial x}\right|$$

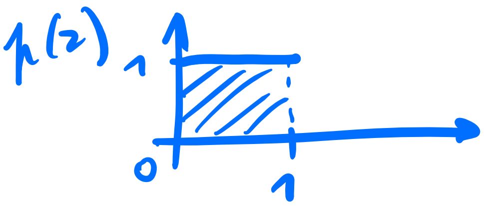
→ $|\cdot|$ because of ratio of two areas

$$\rightarrow p(x) = p(z)\,\frac{\Delta z}{\Delta x}$$

$$p(f(z)) = p(z)\,\frac{\Delta z}{\Delta f} \qquad x = f(z)$$

$$\overset{\Delta z \to 0}{=} p(z)\left|\frac{\partial f}{\partial z}\right|^{-1}$$

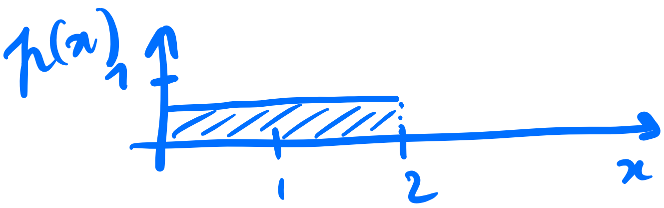
Example:




$$p(z) = \begin{cases} 1 & \text{if } z \in [0;1] \\ 0 & \text{else} \end{cases}$$

$$x = f(z) = 2z \;\longrightarrow\; z = f^{-1}(x) = \frac{1}{2}x$$

$$p(x) = p(f^{-1}(x))\left|\frac{\partial f^{-1}}{\partial x}\right| = p(f^{-1}(x))\,\frac{1}{2}$$




$$\begin{array}{cc} & j \\ i & \left[\begin{array}{ccc} \cdots & \frac{\partial f_i}{\partial x_j} & \end{array}\right] \end{array}$$

Multivariate case:

$$p(x) = p(f^{-1}(x))\left|\det \frac{\partial f^{-1}}{\partial x}\right| \rightarrow \text{Jacobian of } f^{-1}$$

$$p(x = f(z)) = p(z)\left|\det \frac{\partial f}{\partial x}\right|^{-1} \rightarrow \text{Jacobian of } f$$

2) Posit a mean-field Gaussian variational approximation:

$$q_\phi(\underline{z}) = \mathcal{N}(\underline{z} \mid \underline{\mu}, \underline{\sigma})$$

$$= \prod_{k=1}^{K} \mathcal{N}(z_k \mid \underbrace{\mu_k, \sigma_k}_{})$$

⚠ Strong assumption
⇒ no posterior covariance

$$\phi = \{\mu_1, \sigma_1, \dots, \mu_K, \sigma_K\}$$

(1) + (2):

$$\mathcal{L}(\underline{\mu}, \underline{\sigma}) = \mathbb{E}_{q(z)}\left[\log p(x, T^{-1}(z)) \left|\det J_{T^{-1}}(z)\right|\right]$$

$$+ \underbrace{\mathbb{H}[q(z)]}_{}$$

$$= \frac{K}{2}(1 + \log 2\pi) + \sum_{k=1}^{K} \log \sigma_k$$

3) Optimization:

$$\underline{\mu}^*, \underline{\sigma}^* = \arg\max_{\underline{\mu}, \underline{\sigma}} \mathcal{L}(\underline{\mu}, \underline{\sigma})$$

$$\text{s.t.} \quad \underline{\sigma} > 0$$

$\Rightarrow$ Stochastic gradient ascent on $\mathcal{L}$.

⚠

- How to enforce $\underline{\sigma} > 0$?

  $\Rightarrow$ Reparametrize $\rightarrow$ $\underline{\omega} = \log \underline{\sigma} \in \mathbb{R}^K$

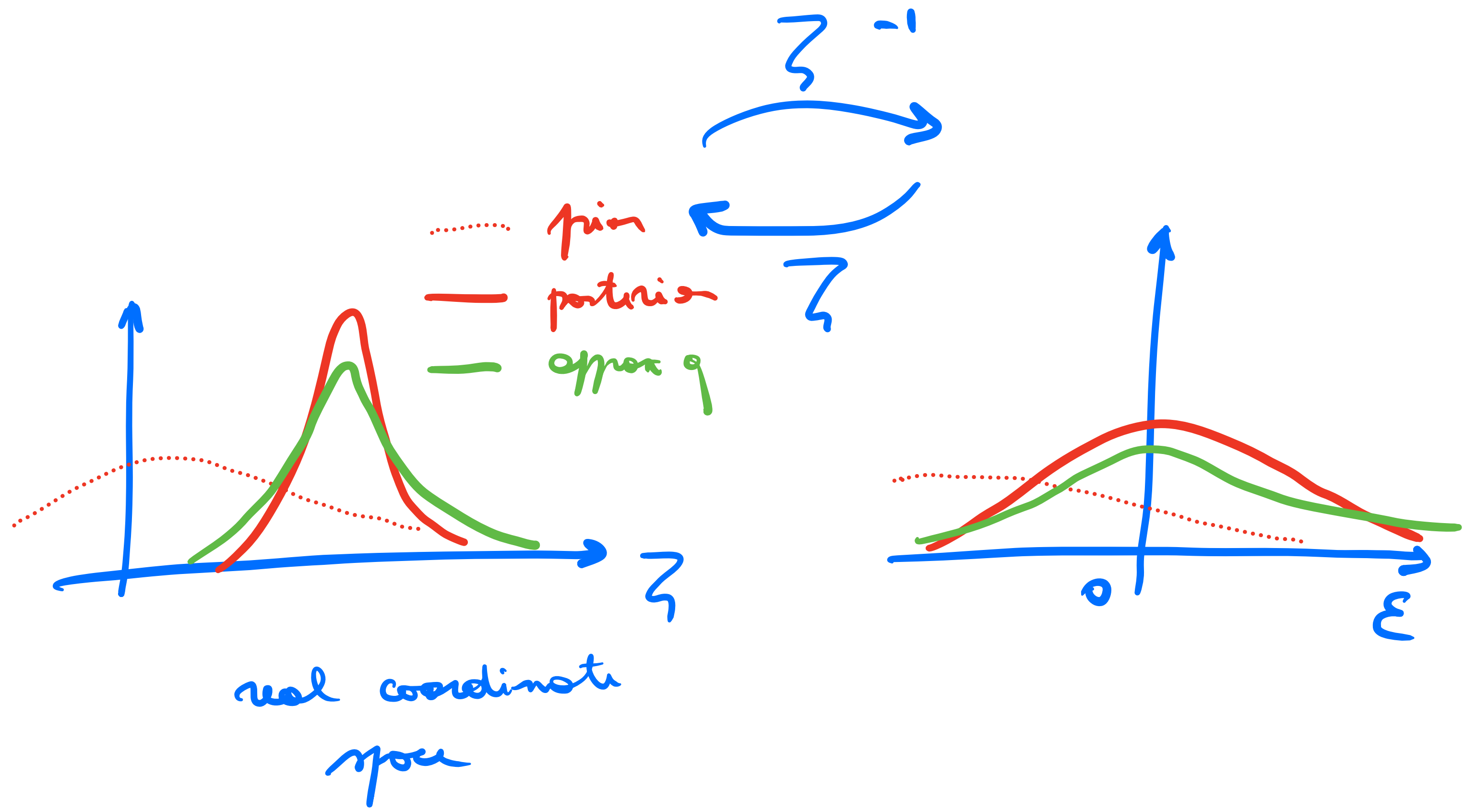
- $\mathbb{E}_q$ depends on $\phi$, hence

$$\nabla_\phi \mathbb{E}_{q_\phi} \neq \mathbb{E}_{q_\phi} \nabla_\phi$$

Reparametrization trick:

$$\underline{\varepsilon} \sim \mathcal{N}(0, I)$$

$$\underline{z} = \underline{\mu} + \mathrm{diag}(\exp(\underline{\omega}))\, \varepsilon$$




$$\underline{\mu}^*, \underline{\omega}^* = \arg\max_{\mu,\omega} \mathcal{L}(\underline{\mu}, \underline{\omega})$$

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

$$= \arg\max_{\underline{\mu}, \underline{\omega}} \mathbb{E}_{\mathcal{N}(\varepsilon; 0, I)} \Big[ \log p(x, T^{-1}(\zeta(\varepsilon))) + \log \left| \det J_{T^{-1}} (\zeta(\varepsilon)) \right| \Big] + \sum_{k=1}^{K} \omega_k$$

$$\nabla_\mu \mathcal{L} = \mathbb{E}_{\mathcal{N}(\varepsilon)}\big[\underbrace{\nabla_\mu(\ldots)}\big]$$

$$\nabla_{w_k} \mathcal{L} = \mathbb{E}_{\mathcal{N}(\varepsilon)}\big[\underbrace{\nabla_{w_k}(\ldots)}\big] + 1$$

with auto-diff

approximate the expectation using M samples

## II. Neural posterior approximation

amortize over all $x$

$$\min_\phi \mathbb{E}_{p(x)}\Big[\mathrm{KL}\big(p(z|x) \,\|\, \underbrace{q_\phi(z|x)}_{\text{neural density estimator}}\big)\Big]$$

forward KL

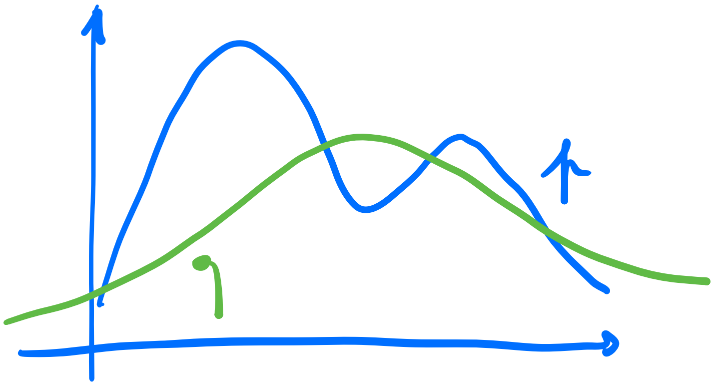
$$= \min_\phi \mathbb{E}_{p(x)\,p(z|x)}\left[\log \frac{p(z|x)}{q_\phi(z|x)}\right]$$

$$= \max_\phi \mathbb{E}_{p(x,z)}\left[\log q_\phi(z|x)\right]$$

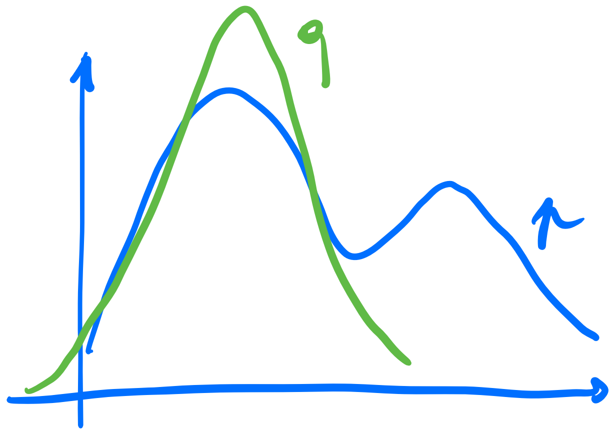
Forward KL vs. Reverse KL

$$KL(p \| q) = \mathbb{E}_p\left[\log \frac{p}{q}\right]$$

$$KL(q \| p) = \mathbb{E}_q\left[\log \frac{q}{p}\right]$$




mean-seeking




mode-seeking