

Siddhant Mishra-Sharma (MIT/AI FI) Summer School

166

19

Variational inference

$$D_{\text{KL}}\left(q_{\varphi}(z) \parallel p(z \mid x)\right)=\log p(x)-\left\langle\log p_{\vartheta}(x, z)-\log q_{\varphi}(z)\right\rangle_{q_{\varphi}(z)}$$



Evidence Lower Bound (ELBO)



Evidence





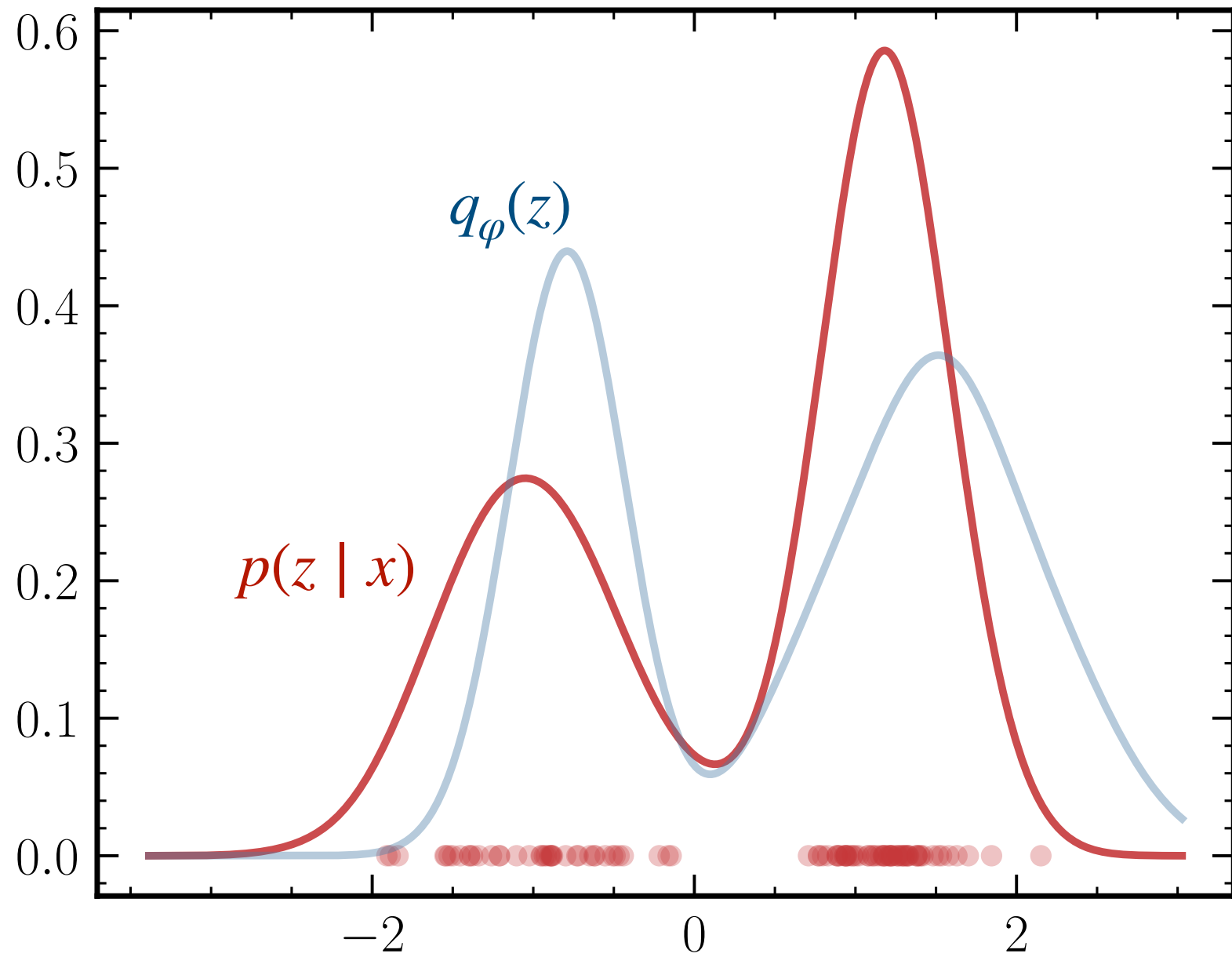
\geq

0

Ageneral-purpose technique for posterior estimation

$$\begin{aligned}\text{ELBO} &= \left\langle \log p_{\vartheta}(x, z) - \log q_{\varphi}(z \mid x) \right\rangle_{q_{\varphi}} \\&= \left\langle \log p_{\vartheta}(x \mid z) + \log p(z) - \log q_{\varphi}(z \mid x) \right\rangle_{q_{\varphi}} \\&= \underbrace{\left\langle \log p_{\vartheta}(x \mid z) \right\rangle_{q_{\varphi}}}_{\text{“Reconstruction”}} - \underbrace{D_{\text{KL}} \left(q_{\varphi}(z \mid x) \parallel p(z) \right)}_{\text{“Regularization”}}\end{aligned}$$

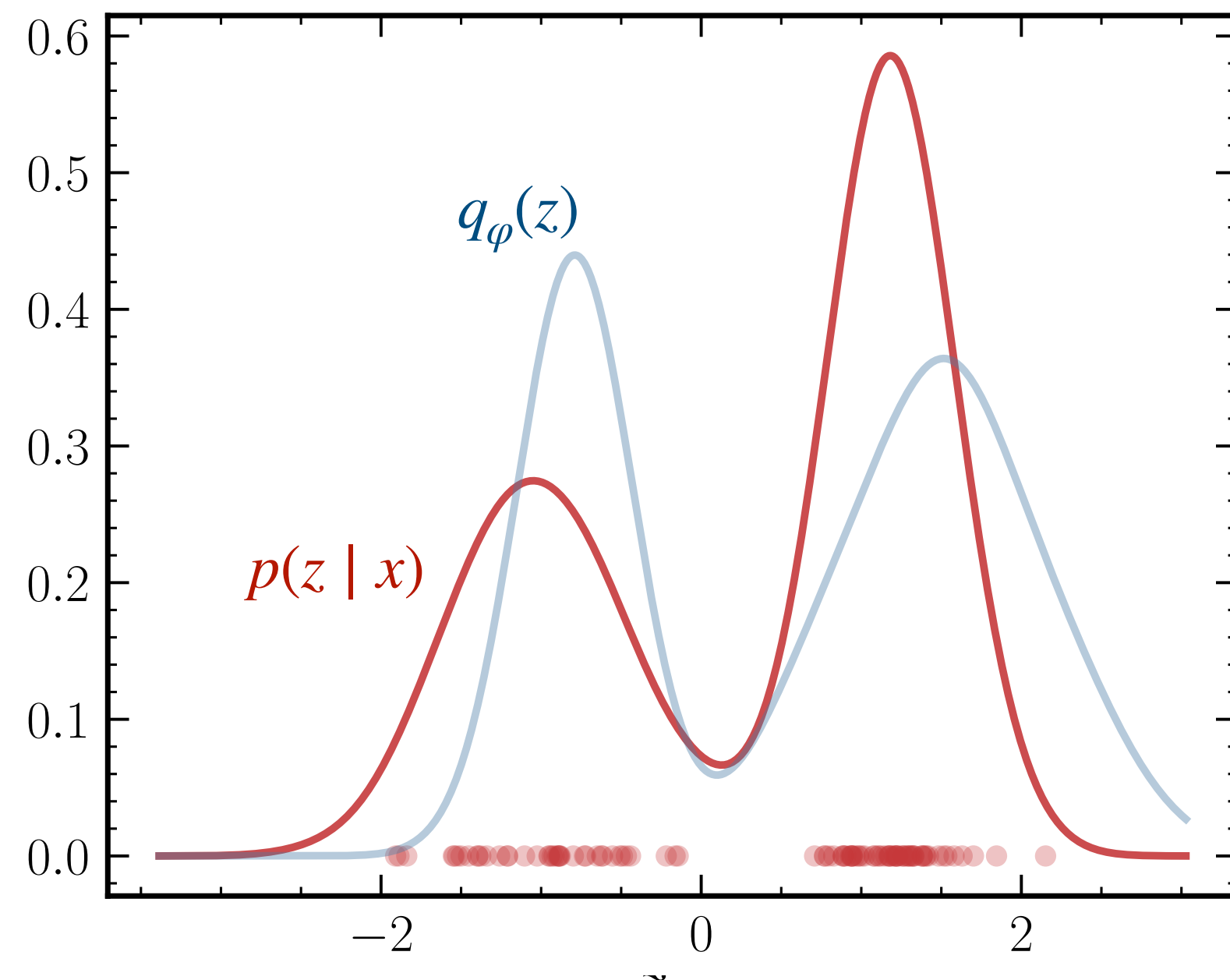




Variational inference

A general-purpose technique for posterior estimation

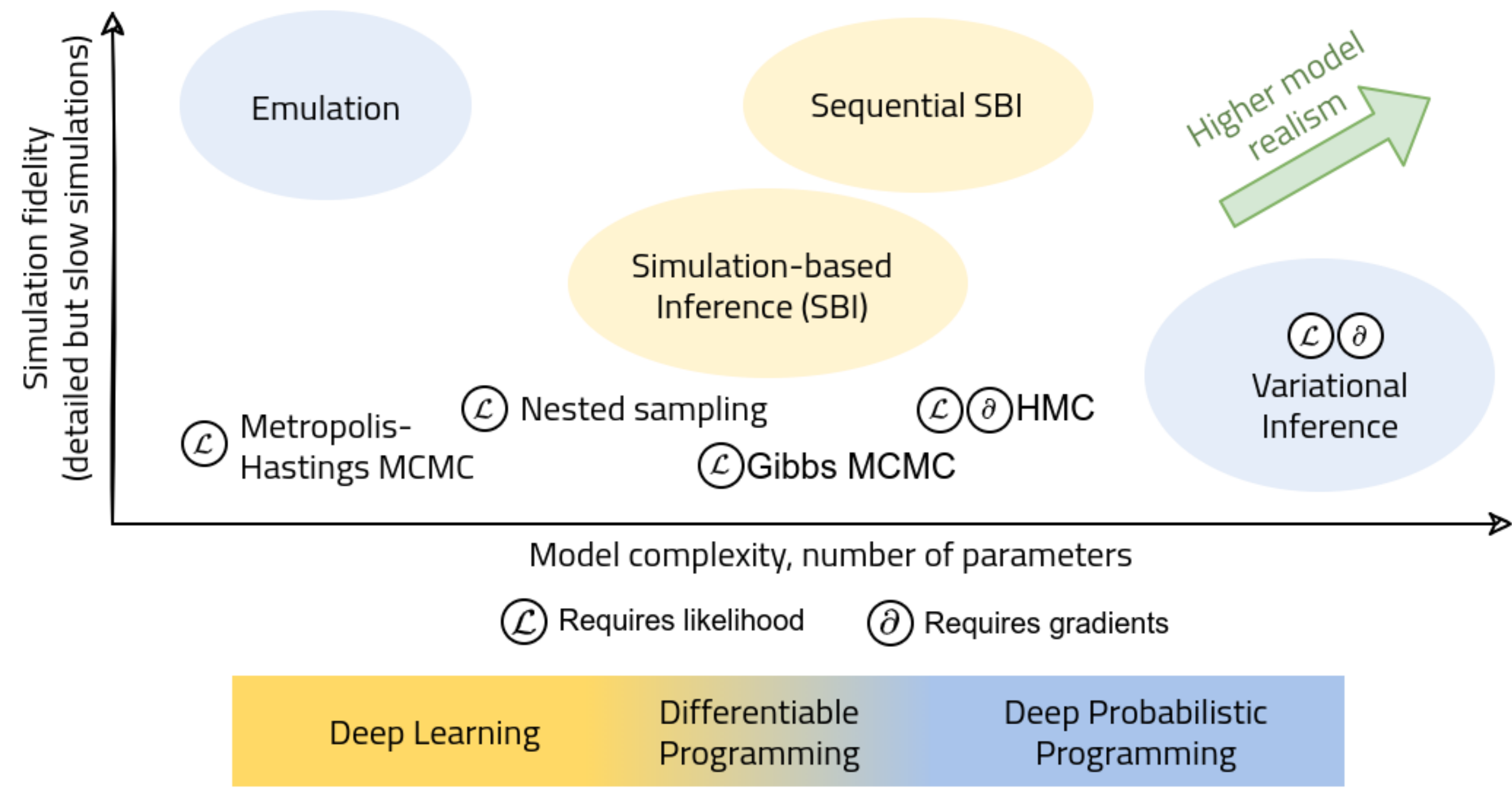
$$\begin{array}{c} \geq 0 \qquad \text{Evidence} - \text{Evidence Lower BOund (ELBO)} \\ \hline D_{\text{KL}} \left(q_{\phi}(z) \parallel p(z \mid x) \right) = \log p(x) - \left\langle \log p_{\theta}(x, z) - \log q_{\phi}(z) \right\rangle_{q_{\phi}(z)} \end{array}$$



$$\begin{aligned} \text{ELBO} &= \left\langle \log p_{\theta}(x, z) - \log q_{\phi}(z \mid x) \right\rangle_{q_{\phi}} \\ &= \left\langle \log p_{\theta}(x \mid z) + \log p(z) - \log q_{\phi}(z \mid x) \right\rangle_{q_{\phi}} \\ &= \underbrace{\left\langle \log p_{\theta}(x \mid z) \right\rangle_{q_{\phi}}}_{\text{"Reconstruction"}} - \underbrace{D_{\text{KL}} \left(q_{\phi}(z \mid x) \parallel p(z) \right)}_{\text{"Regularization"}} \end{aligned}$$

Variational inference

A general-purpose technique for posterior estimation: *optimization* instead of *sampling*



[EuCAPT White Paper 2021]