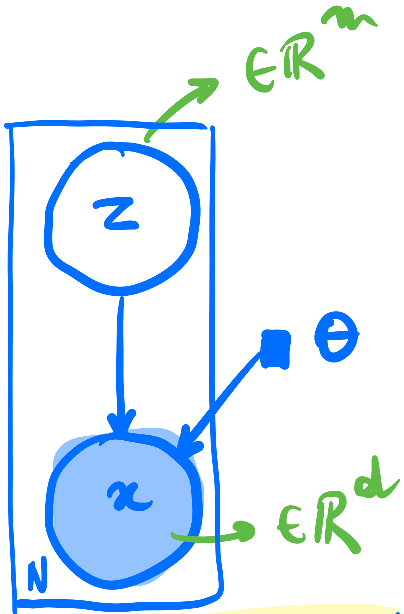
LEC05 Approximate inference, Part I

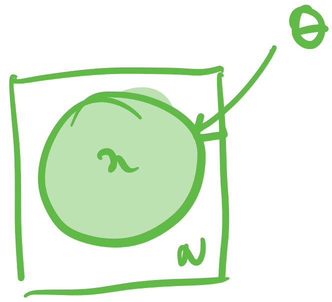


+ GALTON!

$p(z)$

$p_\theta(x|z)$

Start with



How to estimate θ?

How to estimate θ?

$$\max p(x_{1:n}) = \prod p(x_i) \iff \sum \log p(x_i)$$

$$\max_\theta \; p_\theta(x) = \int p_\theta(x, z)\, dz$$

$$= \int p(z)\, p_\theta(x|z)\, dz$$

$$= \mathbb{E}_{p(z)}\left[p_\theta(x|z) \right]$$

$$\approx \frac{1}{B} \sum_b p_\theta(x|z_b)$$

MC approximation

Poor approximation when m is large due to the curse of dimensionality.

I. Expectation - Maximization

$$p_\theta(x) = \mathbb{E}_{p(z)}\left[p_\theta(x|z)\right]$$

IS

$$= \mathbb{E}_{q(z)}\left[\frac{p(z)}{q(z)}\, p_\theta(x|z)\right]$$

if we hold N x_n $\Rightarrow \prod p(x_n) \to \sum \log p(x_n)$

$$\log p_\theta(x) = \log \mathbb{E}_{q(z)}\left[\frac{p(z)}{q(z)}\, p_\theta(x|z)\right]$$

Jensen

$$\geq \mathbb{E}_{q(z)}\left[\log \frac{p(z)}{q(z)}\, p_\theta(x|z)\right]$$

① ELBO

$$= \mathbb{E}_{q(z)}\left[\log \frac{p_\theta(x,z)}{q(z)}\right]$$

$$= \mathbb{E}_{q(z)}\left[\log p_\theta(x|z)\right] - KL\left(q(z)\,\|\,p(z)\right)$$

ELBO(q,θ)

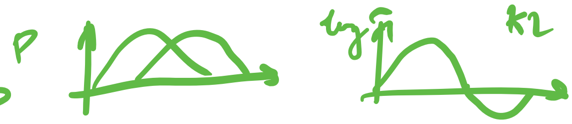
or negative of the variational free energy

②

$$= \mathbb{E}_{q(z)}\left[\log \frac{p(z)}{q(z)}\, p_\theta(x|z)\, \frac{p_\theta(x)}{p_\theta(x)}\right]$$

$$= \mathbb{E}_{q(z)}\left[\log \frac{p_\theta(z|x)}{q(z)}\, p_\theta(x)\right]$$

$$= \log p_\theta(x) - KL\left(q(z)\,\|\,p_\theta(z|x)\right)$$



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

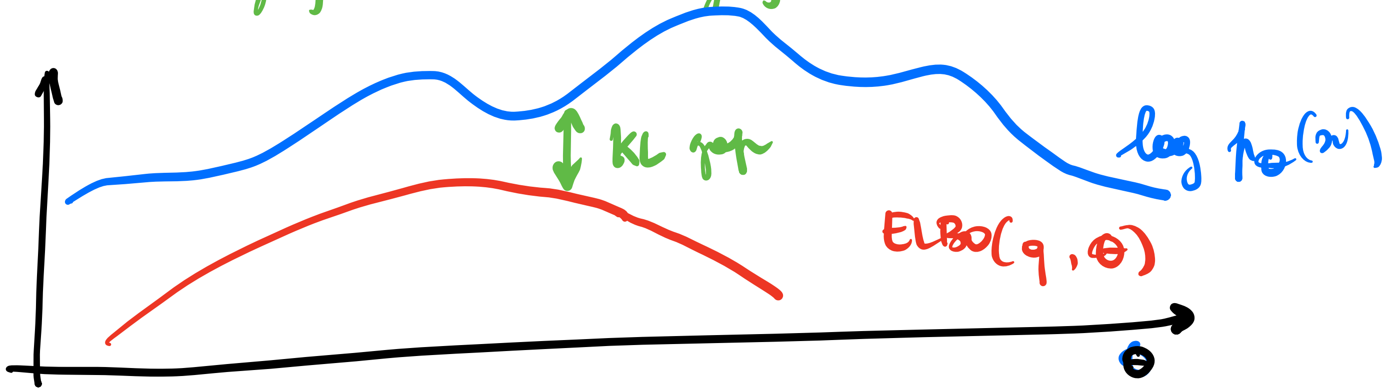
revising one's prior beliefs from Q to P

① + ② $\Rightarrow \log p_\theta(x) = \text{ELBO}(q, \theta) + KL(q(z) \| p_\theta(z|x))$

Since $KL \geq 0$,

$$\log p_\theta(x) \geq \text{ELBO}(q, \theta)$$

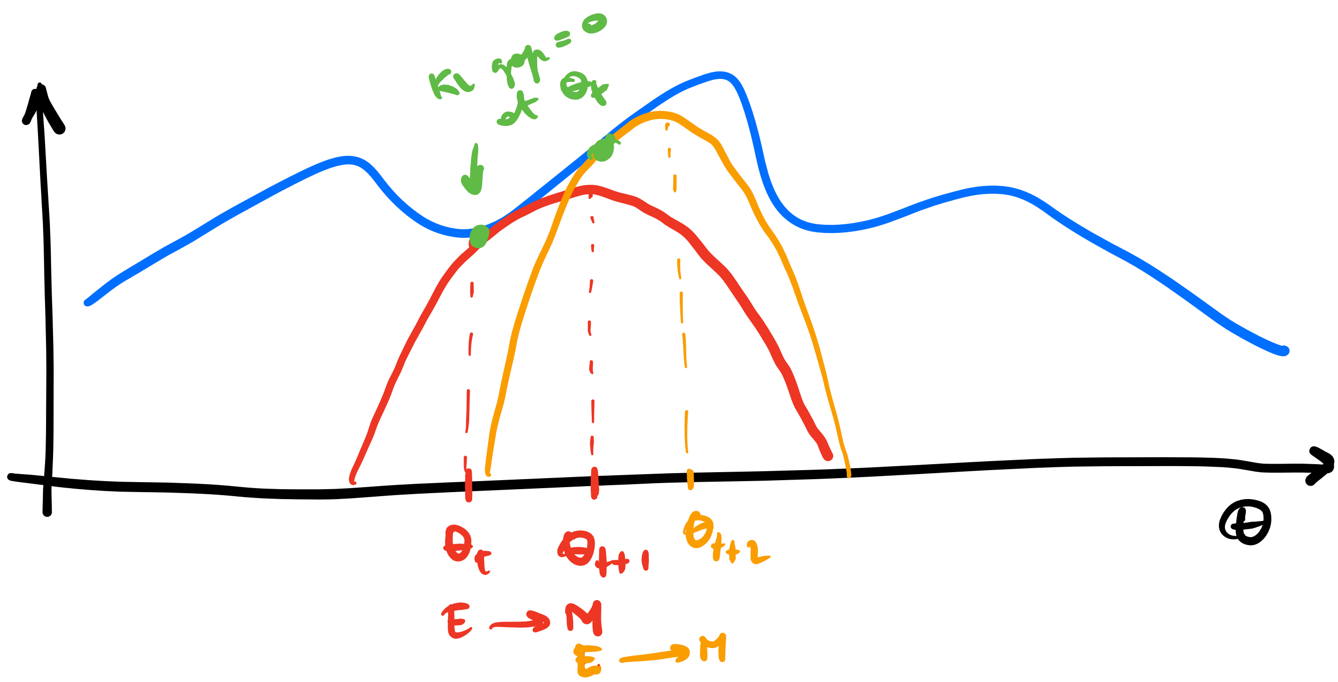
maximizing $\Leftrightarrow$ maximizing



How to pick q? Assume a candidate θ_t, make the bound tight and set

$$q(z) := p_{\theta_t}(z|x)$$

Assuming the posterior can be derived analytically.



Hence,

$$\mathrm{ELBO}(q,\theta) = \mathbb{E}_{q(z)}\left[\log \frac{p_\theta(x,z)}{q(z)}\right]$$

$$= \mathbb{E}_{p_{\theta_t}(z|x)}\left[\log p_\theta(x,z)\right]$$

$$- \mathbb{E}_{p_{\theta_t}(z|x)}\left[\log p_{\theta_t}(z|x)\right]$$

$$= Q(\theta,\theta_t) - \mathbb{H}\left[p_{\theta_t}(z|x)\right]$$

EM:

$\theta_0 = \mathrm{rand}()$

1. E-step : Define $Q(\theta|\theta_t) = \mathbb{E}_{p_{\theta_t}(z|x)}\left[\log p_\theta(x,z)\right]$

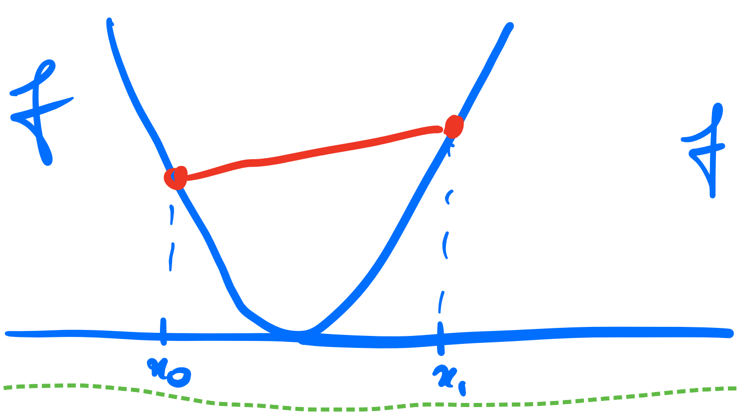
2. M-step : $\theta_{t+1} := \arg\max_\theta Q(\theta,\theta_t)$

(+ $\log p(\theta)$) for MAP estimation)

Jensen's inequality

if f is convex then $f(\mathbb{E}[x]) \leq \mathbb{E}[f(x)]$

concave $f(\mathbb{E}[x]) \geq \mathbb{E}[f(x)]$



$$f\left(\sum \lambda_i x_i\right) \leq \sum \lambda_i f(x_i)$$

$\mathbb{E}_{p(x)}[\cdot] \Leftrightarrow \sum_{x_i} p(x_i)$

II. Variational inference

What if $p_\theta(z|x)$ is not tractable?

Replace q with a variational family $q_\phi(z)$ and solve

VI :

$$\underbrace{\theta^*, \phi^*}_{\text{both together!}} = \arg\max_{\theta,\phi} \mathrm{ELBO}(q_\phi, \theta) = \arg\max_{\theta,\phi} \mathbb{E}_{q_\phi(z)}\left[\log \frac{p_\theta(x,z)}{q_\phi(z)}\right]$$

→ Coordinate ascent VI
Stochastic VI

EM and VI provide algorithms for both fitting θ (to $p_\theta(x)$) and estimating the posterior $p_{\theta^*}(z|x)$.