

Data $x^{(1)}, \dots, x^{(n)}$



(1) aggregation

Observed moments $\mathbb{E}[t(x)]$



(2) factorization

Latent moments $\mathbb{E}[t(x) \mid h]$



(3) optimization

Parameters θ