1. Making $P_\theta(x,z)$ tractable with Monte-carlo

2. Making $P_\theta(x,z)$ tractable with importance sampling

$$\log \sum_z P_\theta(x,z) = \log \sum_z q_\phi(z) \frac{P_\theta(x,z)}{q_\phi(z)}$$

$$= \log E_{z\sim q_\phi(z)}\left[\frac{P_\theta(x,z)}{q_\phi(z)}\right]$$

$$\simeq \log \frac{1}{K} \sum_{k=1}^{K} \frac{P_\theta(x,z_k)}{q_\phi(z_k)},$$ where z_k are sampled from $q_\phi(z_k)$

Lower variance compared to Monte carlo.

But what is a good $q_\phi(z)$??

→ Learn $q_\phi(z)$ from data

Big picture

Our learning objective is to maximise the log probability

$$\log E_{z\sim q_\phi(z)}\left[\frac{P_\theta(x,z)}{q_\phi(z)}\right] \simeq \log \frac{1}{K} \sum_{k=1}^{K} \frac{P_\theta(x,z_k)}{q_\phi(z_k)}$$