FYS5429/9429 lecture
April 24, 2025

FYS5429/9429 April 24, 2025

Quick reminder on BM and VAEs

$$X = \{x_0\ x_1\ -\ \cdot\ \cdot\ x_{n-1}\}$$

$x_i \sim$ i.i.d.

$$p(X) = \prod_{x_i \in X} p(x_i)$$

we have a model with hidden layers – h – $h_j \in H$

$$p(x_i, h_j; \theta) = \frac{f(x_i, h_j; \theta)}{Z(\theta)}$$

Discrete case

$$Z(\theta) = \sum_{x_i \in X} \sum_{h_j \in H} f(x_i, h_j; \theta)$$

marginal distribution

$$p(x_i; \theta) = \sum_{h_j \in H} \frac{f(x_i, h_j; \theta)}{Z(\theta)}$$

$$p(h_j; \theta) = \sum_{x_i \in X} f(x_i, h_j; \theta) \Big/ Z(\theta)$$

$x_i, h_j = \{-1, 1\}$

if we have n-values of x

— " ———— m- values of h $\Rightarrow$

number of configurations for

x is 2^n

h is 2^m

$$P(X;\theta) = \prod_{x_i \in X} P(x_i;\theta)$$

$$= \frac{1}{Z(\theta)} \prod_{x_i \in X} \left[\sum_{h_j \in H} f(x_i, h_j; \theta) \right]$$

$$= \frac{1}{Z(\theta)} \prod_{x_i \in X} f(x_i;\theta)$$

$$\underset{\theta \in \mathbb{R}^P}{\operatorname{argmax}} \log p(X;\theta)$$

$$\nabla_\theta \log p(X;\theta) =$$

$$\nabla_\theta \left[\sum_{x_i \in X} \log f(x_i;\theta) \right]$$

$$- \mathbb{E}\left[\log f(x_i;\theta) \right] = 0$$

VAEs

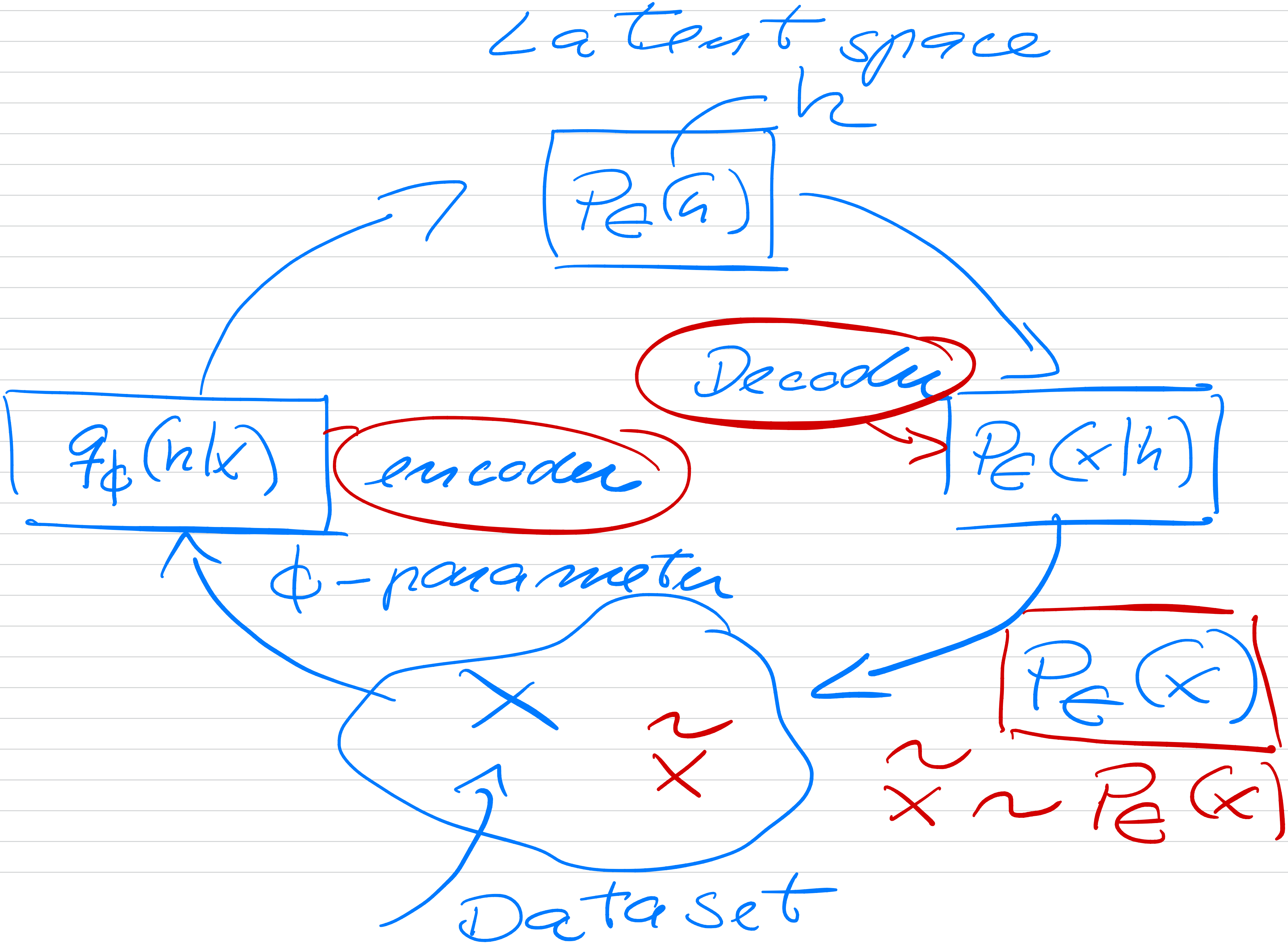
$x_i \to x \; ; \; h_i \to h$

$p(x;\theta) \to p_\theta(x)$

$p_\theta(x) = p_\theta(x,h) / p_\theta(h|x)$

$$p_\theta(x) = \int_{h \in H} dh \, p_\theta(x,h)$$

$$= \int dh \, p_\theta(x|h) \, p_\theta(h)$$



Latent space
h
$P_\theta(h)$
Decoder
$P_\theta(x|h)$
$q_\phi(h|x)$
encoder
ϕ-parameter
x
$\tilde{x}$
$P_\theta(x)$
$\tilde{x} \sim P_\theta(x)$
Dataset

Derivation of VAE equations

$\log P_{\Theta}(x)$ we want gradient

$$\int q_{\phi}(h|x)\,dh = 1$$

$$\log p(x) = \log p(x) \int q_{\phi}(h|x)\,dh$$

$$= \int q_{\phi}(h|x) \log p(x)\,dh$$

$$= \mathbb{E}_{h \sim q_{\phi}}\left[\log p(x)\right]$$

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

$$1 = \frac{q_\phi(h|x)}{q_\phi(h|x)}$$

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

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

$$\underbrace{\mathbb{E}_{q_\phi(h|x)}\left[\log \frac{q_\phi(h|x)}{p(h|x)}\right]}_{\text{Kullback-Leibler divergence}}$$

$$D_{KL}\left(q_\phi(h|x) \,\|\, p(h|x)\right) \geq 0$$

$$\Rightarrow$$

$$\log p(x) = \mathbb{E}_{q_\phi(h|x)}\left[\log \frac{p(x,h)}{q_\phi(h|x)}\right] + D_{KL}\left(q_\phi \,\|\, p(h|x)\right)$$

$$= \overbrace{\boxed{\mathbb{E}_{q_\phi(h|x)}\left[\frac{\log p(x,h)}{q_\phi(h|x)}\right]}}^{\text{ELBO}} \geq 0$$

ELBO = evidence lower bound, and leads to variational bound

we don't know $p(x,h)$

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

introduce θ

$$p_\theta(x,h) = \left(p(x,h;\theta)\right)$$

$$p_\theta(x|h)\,p_\theta(h)$$

recall that $p_\theta(x) = \int dh\, p_\theta(x|h)\,p_\theta(h)$

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

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

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

$$+ \quad \mathbb{E}_{q_\phi(z|x)}\left[\log \frac{p_\theta(z)}{q_\phi(z|x)}\right]$$

$z \sim q_\phi(z|x)$

$$= \overbrace{\mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]}^{\text{Reconstruction term}}$$

$$- \underbrace{D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right)}_{\text{Prior matching term}}$$

$q_\phi(z|x)$ is treated as an encoder

$$q_\phi(z|x) \sim N(z;\, \mu_\phi(x);\, \sigma_\phi^2)$$

multivariate Gaussian

$$P_\theta(h) = p(h) \sim N(h; 0, 1)$$

we are learning a deterministic function $\boxed{P_\theta(x|h)}$

Decoder

$$\arg\max_{\theta, \phi} \left\{ \mathbb{E}_{q_\phi(h|x)} \left[\log P_\theta(x|h) \right] - D_{KL}(q_\phi(h|x) \,||\, p(h)) \right\}$$

$$\simeq \arg\max_{\theta, \phi} \left\{ \frac{1}{M} \sum_{l=1}^{M} \log P_\theta(x|h_l) - D_{KL}(q_\phi \,||\, p(h)) \right\}$$

Diffusion models & Markov chains:

$P_i(t)$ = probability of being in a state $-i-$ at time $-t-$

Markov-chain

$$P_i(t) = \sum_j T_{ij}\, P_j(t')$$

Transition probability (unknown)

$$\sum_j T_{ij} = 1$$

element T_{ij} of stochastic matrix

$$T$$

$$T_{ij} = \underbrace{T(j \to i)}_{\text{can be time-dep or time-independent}}$$

$$\sum_j P_j = 1$$

$$p(t) = \begin{bmatrix} P_0(t) \\ P_1(t) \\ \vdots \\ P_{n-1}(t) \end{bmatrix}$$

$$p(t) = T\,p(t')$$

initial time

$$p(t_m) = T^m\,p(t_0)$$

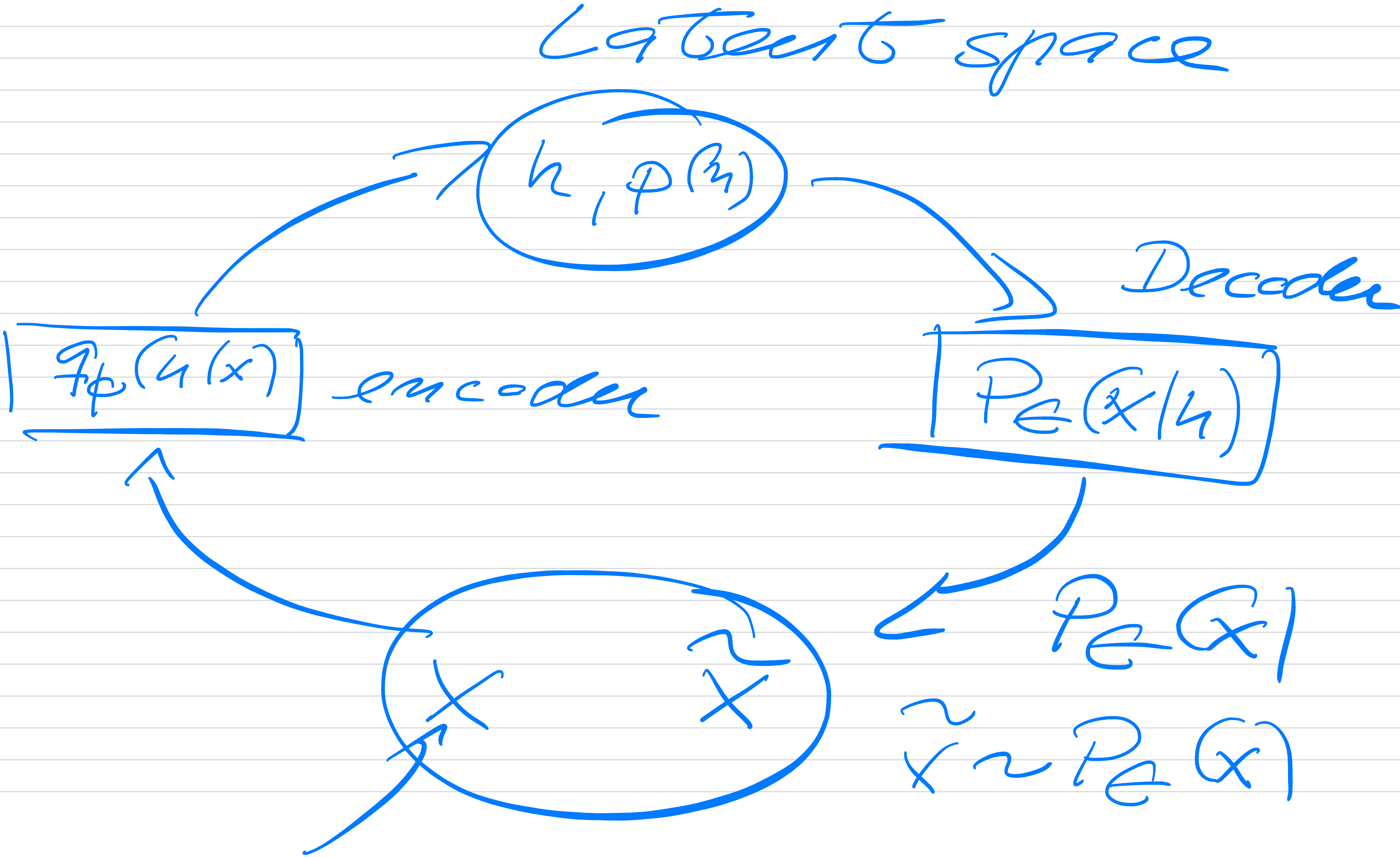
$$\lim_{t_m \to \infty} \quad p(t_m) \Rightarrow T^m\,p(t_m)$$

$$p(t_m) = T^m\,p(t_m)$$

$p(t_0)$ $\quad$ $p(t_1)$ $\quad$ $p(t_2)$ $\quad$ $p(t_m)$

$$\begin{tikzcd}
t_0 \arrow[r, bend left, "T(t_0 \to t_1)"'] & t_1 \arrow[r, bend left, "T(t_1 \to t_2)"'] & t_2 \arrow[r, dash] & \cdots \arrow[r, bend right, "T(t_{m-1} \to t_m)"'] & t_m
\end{tikzcd}$$

VAE



$$p_\theta(x) = \int dh\, p_\theta(x|h)\, p_\theta(h)$$

Diffusion models

$$\begin{tikzcd}
x_0 \arrow[r, bend right, "q(h_1|x_0)"'] & h_1 \arrow[l, bend right, "p(x_0|h_1)"'] \arrow[r, bend right, "q(h_2|h_1)"'] & h_2 \arrow[l, bend right, "p(h_1|h_2)"'] \arrow[r] & \cdots \arrow[r, bend right, "q(h_m|h_{m-1})"'] & h_m \arrow[l, bend right, "p(h_{m-1}|h_m)"']
\end{tikzcd}$$

$$\tilde{x}_0 \sim p(x_0)$$