ELBO Extension to Diffusion Loss

$$-L_{VAE} = +\log p_\theta(x) - D_{KL}(q_\phi(z|x) \| p_\theta(z|x)) \le \log p_\theta(x)$$

Lower Bound!

$$L_{VAE}: \; -\log p_\theta(x) \le -\log p_\theta(x) + D_{KL}(q_\phi(z|x) \| p_\theta(z|x))$$

$$= D_{KL}(q_\phi(z|x) \| q_\theta(z)) - \mathbb{E}_{z \sim q_\phi(z|x)} \log p_\theta(x|z)$$

Extend to Diffusion *

$$-\log(p_\theta(x_0)) \le -\log p_\theta(x_0) + D_{KL}(q(x_{1:T}|x_0) \| p_\theta(x_{1:T}|x_0))$$

Note:

$$D_{KL}(p(x) | q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx$$

$$= \int q(x_{1:T}|x_0) \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} dx_{1:T}$$

$$p_\theta(x_{1:T}|x_0) = \frac{p_\theta(x_0|x_{1:T})\, p_\theta(x_{1:T})}{p_\theta(x_0)} = \frac{p_\theta(x_0, x_{1:T})}{p_\theta(x_0)} = \frac{p_\theta(x_{0:T})}{p_\theta(x_0)}$$

$$= \int q(x_{1:T}|x_0) \log \frac{q(x_{1:T}|x_0)\, p_\theta(x_0)}{p_\theta(x_{0:T})} dx_{1:T}$$

$$= \int q(x_{1:T}|x_0) \left\{ \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + \log p_\theta(x_0) \right\} dx_{1:T}$$

$$= \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + \log p_\theta(x_0)$$

$$-\log(p_\theta(x_0)) \le -\log p_\theta(x_0) + \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + \log p_\theta(x_0)$$

$$-\log(p_\theta(x_0)) \le \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

Variational Lower Bound for Diffusion!

$$-\log(p_\theta(x_0)) \le \log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

Note:

$$p_\theta(x_{0:T}) = p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \qquad q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$$

$$-\log(p_\theta(x_0)) \le \log\left\{\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}\right\}$$

$$= -\log p(x_T) + \log\left\{\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}\right\}$$

$$= -\log p(x_T) + \sum_{t=1}^{T}\log\left\{\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right\}$$

$$= -\log p(x_T) + \sum_{t=2}^{T}\log\left\{\frac{\boxed{q(x_t|x_{t-1})}}{p_\theta(x_{t-1}|x_t)}\right\} + \log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

(= too uncertain to predict x_{t-1} from x_t)

** too high variance!!

$$q(x_t|x_{t-1}) = \frac{\overset{①}{q(x_{t-1}|x_t)}\,\overset{②}{q(x_t)}}{\underset{③}{q(x_{t-1})}}$$

Can make better prediction of x_{t-1} from x_t given x_0 as well!

$$\frac{q(x_{t-1}|x_t,x_0)\,q(x_t|x_0)}{q(x_{t-1}|x_0)}$$

$$= -\log p(x_T) + \sum_{t=2}^{T}\log\left\{\frac{q(x_{t-1}|x_t,x_0)\,q(x_t|x_0)}{p_\theta(x_{t-1}|x_t)\,q(x_{t-1}|x_0)}\right\} + \log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

$$= -\log p(x_T) + \sum_{t=2}^{T}\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} + \boxed{\sum_{t=2}^{T}\log\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}} + \log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

ex) T=4: $\log\left\{\frac{q(\cancel{x_2}|x_0)}{q(x_1|x_0)} \times \frac{q(\cancel{x_3}|x_0)}{q(\cancel{x_2}|x_0)} \times \frac{q(x_4|x_0)}{q(\cancel{x_3}|x_0)}\right\}$

$$= \log\frac{q(x_T|x_0)}{q(x_1|x_0)}$$

$$= -\log p(x_T) + \sum_{t=2}^{T}\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} + \log\frac{q(x_T|x_0)}{q(x_1|x_0)} + \log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}$$

$$= \log\{q(x_T|x_0) - q(x_1|x_0) + q(x_1|x_0) - p_\theta(x_0|x_1)\}$$

$$= \log q(x_T|x_0) - \log p_\theta(x_0|x_1)$$

$$= \log q(x_T|x_0) - \log p(x_T) + \sum_{t=2}^{T}\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} - \log p_\theta(x_0|x_1)$$

$$= \log\left\{\frac{q(x_T|x_0)}{p(x_T)}\right\} + \sum_{t=2}^{T}\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} - \log p_\theta(x_0|x_1)$$

$$= \underbrace{\log\left\{\frac{q(x_T|x_0)}{p(x_T)}\right\}}_{D_{KL}(q(x_T|x_0)\,\|\,p(x_T))} + \sum_{t=2}^{T} \underbrace{\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)}}_{D_{KL}(q(x_{t-1}|x_t,x_0)\,\|\,p_\theta(x_{t-1}|x_t))} - \log p_\theta(x_0|x_1)$$

$$= D_{KL}(q(x_T|x_0)\,\|\,p(x_T)) + \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,x_0)\,\|\,p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_1)$$

ignore
bc no learnable params

$$\sum_{t=2}^{T} D_{KL}(\underbrace{q(x_{t-1}|x_t,x_0)}_{\sim\mathcal{N}(x_{t-1}|\tilde{\mu}_t(x_t,x_0),\,\tilde{\beta}I)}\,\|\,\underbrace{p_\theta(x_{t-1}|x_t)}_{\sim\mathcal{N}(x_{t-1};\mu_\theta(x_t,t),\,\beta_I)}) - \log p_\theta(x_0|x_1)$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right)$$

$$\mathcal{L}_t = \frac{1}{2\sigma_t^2}\left\|\tilde{\mu}_t(x_t,x_0) - \mu_\theta(x_t,t)\right\|^2$$

$$= \frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right)\right\|^2$$

$$= \boxed{\frac{\beta^2}{2\sigma_t^2\alpha_t(1-\hat{\alpha}_t)}}\left\|\epsilon - \epsilon_\theta(x_t,t)\right\|^2$$

To be ignored!

$$\Rightarrow \left\|\epsilon - \epsilon_\theta(x_t,t)\right\|^2$$

Normalized Image pixels range: $[-1,1]$ D: # of pixels in the image

$$p_\theta(x_0|x_1) = \prod_{i=1}^{D}\boxed{\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)}\mathcal{N}(x;\mu_\theta^i(x_1,1),\beta_1)\,dx}$$

if integral predicts mean close to that of true pixels : $\int\uparrow$
" far away from " : $\int\downarrow$

$$\delta_+(x) = \begin{cases}\infty & \text{if } x=1\\ x+\frac{1}{255} & \text{if } x<1\end{cases} \qquad \delta_-(x) = \begin{cases}-\infty & \text{if } x=-1\\ x-\frac{1}{255} & \text{if } x>-1\end{cases}$$

Border (ex. integrate from one pixel below to one pixel above)

However, also ignored!

[Sampling Time]

- when $t > 1$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sqrt{\beta_t}\,\epsilon$$

- when $t = 1$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right)$$ No noise!

$$\therefore \mathcal{L}oss = \mathbb{E}_{t, x_0, \epsilon}\left\| \epsilon - \epsilon_\theta\left(\underbrace{\boxed{\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon}}_{= x_t}, t\right)\right\|^2$$

$$\Updownarrow$$

$$\mathcal{L}oss = \mathbb{E}_{t, x_0, \epsilon}\left\| \epsilon - \epsilon_\theta(x_t, t)\right\|^2$$