(4) $q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\,x_{t-1}, \beta_t I)$

服从高斯正态分布 $\mu = \sqrt{1-\beta_t}\,x_{t-1}$ $\sigma^2 = \beta_t I$

令 $\alpha_t = 1-\beta_t$ $z_t \sim N(0, I)$ 为标准正态分布

则 $x_t = \sqrt{\alpha_t}\,x_{t-1} + \sqrt{1-\alpha_t}\,z_{t-1}$

由 $z = \dfrac{x-\mu}{\sigma}$ 即正态分布和标准正态分布之间的关系得到

$$x_t = \sqrt{\alpha_t}\,x_{t-1} + \sqrt{1-\alpha_t}\,z_{t-1}$$

$$x_{t-1} = \sqrt{\alpha_{t-1}}\,x_{t-2} + \sqrt{1-\alpha_{t-1}}\,z_{t-2}$$

$$\vdots$$

$$x_1 = \sqrt{\alpha_1}\,x_0 + \sqrt{1-\alpha_1}\,z_0$$

令 $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

则 $x_t = \sqrt{\bar{\alpha}_t}\,x_0 + \underbrace{\dfrac{\sqrt{\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_1}}\sqrt{1-\alpha_1}\,z_0 + \dfrac{\sqrt{\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_2}}\sqrt{1-\alpha_2}\,z_1 + \cdots + \dfrac{\sqrt{\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}\sqrt{1-\alpha_t}\,z_{t-1}}_{\text{设这一部分为随机变量 } \tilde{z}_{t-1}}$

而 $E(z_i) = 0$ 故 $E(\tilde{z}_{t-1}) = 0$

$$Var(\tilde{z}_{t-1}) = \frac{\bar{\alpha}_t}{\bar{\alpha}_1}(1-\alpha_1) + \frac{\bar{\alpha}_t}{\bar{\alpha}_2}(1-\alpha_2) + \cdots + \frac{\bar{\alpha}_t}{\bar{\alpha}_t}(1-\alpha_t)$$

$$= \alpha_2\alpha_3\cdots\alpha_t\beta_1 + \alpha_3\cdots\alpha_t\beta_2 + \alpha_4\cdots\alpha_t\beta_3 + \cdots + \alpha_t\beta_{t-1} + \beta_t$$

而 $\alpha_t\alpha_{t-1}\cdots\alpha_1 + \alpha_t\alpha_{t-1}\cdots\alpha_2\beta_1 + \alpha_t\cdots\alpha_3\beta_2 + \cdots + \alpha_t\beta_{t-1} + \beta_t$

$= \alpha_t\alpha_{t-1}\cdots\alpha_2(\alpha_1+\beta_1) + \alpha_t\cdots\alpha_3\beta_2 + \cdots + \alpha_t\beta_{t-1} + \beta_t$

$= \alpha_t\alpha_{t-1}\cdots\alpha_2 + \alpha_t\cdots\alpha_3\beta_2 + \cdots + \alpha_t\beta_{t-1} + \beta_t$

$= \alpha_t\alpha_{t-1}\cdots\alpha_3(\alpha_2+\beta_2) + \cdots + \alpha_t\beta_{t-1} + \beta_t$

$= \alpha_t\alpha_{t-1}\cdots\alpha_3 + \cdots + \alpha_t\beta_{t-1} + \beta_t$

$= \cdots$

$= \alpha_t + \beta_t$

$= 1$

因此 $Var(\tilde{z}_{t-1}) = 1 - \alpha_t\alpha_{t-1}\cdots\alpha_1 = 1 - \bar{\alpha}_t$

因此 $x_t = \sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\bar{z}_t$ $\bar{z}_t \sim N(0, I)$

即 $q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}\,x_0, (1-\bar{\alpha}_t) I)$ (4)式证毕

(3) 目标：得到尽可能真实的 x_0，即求 θ 使得 $p_\theta(x_0)$ 最大

噪声空间难以积分，所以 $-\log p_\theta(x_0)$ 难以优化，故考虑优化 $\log p_\theta(x_0)$ 的变分下界 $\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$

下证 $-\log p_\theta(x_0) \le E_q\left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$ 和 $E\left[-\log p_\theta(x_0)\right] \le E_q\left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$

极大似然函数 $p_\theta(x_0) = \int_{x_1}\int_{x_2}\cdots\int_{x_T} p_\theta(x_0, x_1, x_2, \cdots, x_T)\, dx_1 dx_2 \cdots dx_T$

$$= \int_{x_1}\int_{x_2}\cdots\int_{x_T} q(x_{1:T}|x_0) \cdot \frac{p_\theta(x_0, x_1, x_2, \cdots, x_T)}{q(x_{1:T}|x_0)} dx_1 dx_2 \cdots dx_T$$

$$= E_{q(x_{1:T}|x_0)}\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

故 $\log p_\theta(x_0) = \log E_{q(x_{1:T}|x_0)}\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right] \ge E_{q(x_{1:T}|x_0)}\left[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$

理由：$-\log x$ 是凸函数，利用 Jensen 不等式

故有 $-\log p_\theta(x_0) \le E_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right]$

$$-E_{q(x_0)} \log p_\theta(x_0) = -E_{q(x_0)} \log\left(\int p_\theta(x_{1:T}|x_0) p_\theta(x_0)\, dx_{1:T}\right)$$

$$= -E_{q(x_0)} \log\left(\int p_\theta(x_{0:T})\, dx_{1:T}\right)$$

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

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

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

$$= E_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right]$$

(3) 式中的不等式证毕

$$= E_{q(x_{0:T})}\left[\log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=1}^{T} \log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_{q(x_{0:T})}\left[-\log p_\theta(x_T) - \sum_{t=1}^{T} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right]$$

(3) 式全部证毕

(6)、(7) 逆扩散过程

在扩散过程中，我们知道了 $q(x_t|x_{t-1})$ 和 $q(x_t|x_0)$

但考虑逆扩散降噪过程，$q(x_{t-1}|x_t)$ 是不能容易得到的

因此用 $p_\theta(x_{t-1}|x_t)$ 来近似 $q(x_{t-1}|x_t)$，而 $p_\theta(x_{t-1}|x_t)$ 正是我们要训练的

下证 $q(x_{t-1}|x_t,x_0)$ 可以用 $q(x_t|x_0)$ 和 $q(x_t|x_{t-1})$ 来表示，即 $q(x_{t-1}|x_t,x_0)$ 可知

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

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

由于扩散过程是马尔可夫过程，即 t 时的状态只和 t-1 时的状态有关

故 $q(x_t|x_{t-1}x_0)=q(x_t|x_{t-1})$

所以 $q(x_{t-1}|x_t,x_0)=q(x_t|x_{t-1})\cdot\dfrac{q(x_{t-1}|x_0)}{q(x_t|x_0)}$

依次对 $q(x_t|x_{t-1})$，$q(x_t|x_0)$，$q(x_{t-1}|x_0)$ 进行处理

$$q(x_t|x_{t-1})=N(x_t;\sqrt{1-\beta_t}\,x_{t-1},\beta_t I)$$

$$=\frac{1}{\sqrt{2\pi(1-\alpha_t)}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_t}}$$

$$q(x_t|x_0)=\frac{1}{\sqrt{2\pi(1-\bar\alpha_t)}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_t-\sqrt{\bar\alpha_t}x_0)^2}{1-\bar\alpha_t}}$$

$$q(x_{t-1}|x_0)=\frac{1}{\sqrt{2\pi(1-\bar\alpha_{t-1})}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}x_0)^2}{1-\bar\alpha_{t-1}}}$$

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

$$=\frac{\dfrac{1}{\sqrt{2\pi(1-\alpha_t)}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{1-\alpha_t}}\cdot\dfrac{1}{\sqrt{2\pi(1-\bar\alpha_{t-1})}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}x_0)^2}{1-\bar\alpha_{t-1}}}}{\dfrac{1}{\sqrt{2\pi(1-\bar\alpha_{t-1})}}\cdot e^{-\frac{1}{2}\cdot\frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}x_0)^2}{1-\bar\alpha_{t-1}}}}$$

故 $q(x_{t-1}|x_t,x_0)=N\left(x_{t-1};\dfrac{(1-\bar\alpha_{t-1})\sqrt{\alpha_t}\,x_t}{1-\bar\alpha_t}+\dfrac{\beta_t\sqrt{\bar\alpha_{t-1}}\,x_0}{1-\bar\alpha_t},\dfrac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t\right)$

(7) 式证毕

(5) 式证明: $E_{q(x_{0:T})}\left[-\log p_\theta(x_T)+\sum_{t=1}^{T}\log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$

$$=E_{q(x_{0:T})}\left[-\log p_\theta(x_T)+\sum_{t=2}^{T}\log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}+\log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right]$$

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

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

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

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

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

$$=E_q\left[\underbrace{D_{KL}(q(x_T|x_0)\|p(x_T))}_{L_T}+\underbrace{\sum_{t>1}D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t))}_{L_{t-1}}\underbrace{-\log p_\theta(x_0|x_1)}_{L_0}\right]$$

$$L_{t-1}=D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t))$$

$$=D_{KL}(N(x_{t-1};\tilde{\mu}(x_t,x_0),\sigma_t^2 I)\|N(x_{t-1};\mu_\theta(x_t,t),\sigma_t^2 I))$$

$$=\log 1+\frac{\sigma_t^2+\|\tilde{\mu}_t(x_t,x_0)-\mu_\theta(x_t,t)\|^2}{2\sigma_t^2}-\frac{1}{2}$$

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

故 $L_{t-1}=E_q\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t,x_0)-\mu_\theta(x_t,t)\|^2\right]+C$