$x_0 \sim q(x_0)$

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

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, \epsilon$$

$\beta_t \in [0,1]$: 方差

variance/noise schedule $\{\beta_1 = \cdots, \beta_2 = \cdots \quad \cdots, \beta_T = \cdots\}$

$$\beta_1 < \beta_2 < \cdots \beta_T$$

$x_t = x_{t-1} + \text{noise} \Rightarrow$ 与 $x_{t-2}, x_{t-3} \cdots$ 无关 $\Rightarrow$ markov

② 令 $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

则 $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$

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

③ 反向去噪:

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

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$p(x_T) = N(0, I)$$

④ $q(x_{t-1}|x_t)$ 无法得到, 但 $q(x_{t-1}|x_t, x_0)$ 可得到:

$$q(x_{t-1}|x_t, x_0) = N(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$$

其中 $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t$

$$\tilde{\mu}(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\, \beta_t}{1-\bar{\alpha}_t} x_0$$

why?

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

而 $q(x_t) = \int q(x_t|x_0)\, q(x_0)\, dx_0$

$q(x_t|x_0)$: 有公式 ✓; $q(x_0)$: 未知

$\therefore q(x_t), q(x_{t-1})$ 推不出来!

终极目标是得到好的 $p_\theta(x_{t-1}|x_t)$

⑤ KL, MLE, VLB/ELBO:

想要 minimize $KL(q(x)\|P_\theta(x))$

而 $KL(q\|P_\theta) = E_{q(x)}[\log q(x) - \log P_\theta(x)]$

$\log q(x)$ 与θ无关!

等价于 minimize $E_{q(x)}[-\log P_\theta(x)]$

或 maximize $E_{q(x)}[\log P_\theta(x)]$

也可以从 MLE 角度理解: $\log \text{Likelihood} = E_{q(x)}[\log P_\theta(x)]$

而 $\log P_\theta(x_0) = \log \int P_\theta(x_{0:T}) dx_{1:T}$

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

Jensen 不等式

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

VLB/ELBO

则 maximize $E_{q(x)}[\log P_\theta(x)] \Rightarrow$ maximize ELBO

$\Rightarrow \text{Loss} = -\text{ELBO} = E_{q(x_{1:T}|x_0)}\left[\log \frac{P_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$

经过推导, $\text{Loss} \approx \sum_{t=2}^{T} L_{t-1} = \sum_{t=2}^{T} E_{q(x_t|x_0)}\left[D_{KL}(q(x_{t-1}|x_t, x_0)\|P_\theta(x_{t-1}|x_t))\right]$

minimize Loss 等价于 minimize $E_{q(x_t|x_0)}\left[D_{KL}(q(x_{t-1}|x_t, x_0)\|P_\theta(x_{t-1}|x_t))\right]$

这也解释了为什么前面要假设 $P_\theta(x_{t-1}|x_t)$ 服从 Gaussian

因为 $q(x_{t-1}|x_t, x_0)$ 就是 Gaussian!

⑥ $P_\theta(x_{t-1}|x_t)$ 的方差, DDPM 做了简化, 为定值 $\sigma_t^2 I$ (σ_t^2 可取 $\beta_t / \tilde{\beta}_t$, 实际上对应上/下限,

由于 $q(x_{t-1}|x_t, x_0)$ 的方差是 $\tilde{\beta}_t$, 这里取 $\sigma_t^2 = \tilde{\beta}_t$ 见 IDDPM)

⑦ 推导得 $L_{t-1} = E_{q(x_t|x_0)}\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2\right]$

代入 $x_t(x_0,\epsilon)=\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\,\epsilon$:

$$L_{t-1}=E_{x_0,\epsilon\sim N(0,I)}\left[\frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(x_t(x_0,\epsilon)-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right)-\mu_\theta(x_t(x_0,\epsilon),t)\right\|^2\right]$$

为了匹配第一项的形状，对 μ_θ 重参数化，让神经网络预测 ϵ：

$$\mu_\theta(x_t(x_0,\epsilon),t)=\frac{1}{\sqrt{\alpha_t}}\left(x_t(x_0,\epsilon)-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t(x_0,\epsilon),t)\right)$$

所以
$$L_{t-1}=E_{x_0,\epsilon\sim N(0,I)}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\left\|\epsilon-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\,\epsilon,t)\right\|^2\right]$$

简化为
$$L_{t-1}=E_{x_0,\epsilon\sim N(0,I)}\left[\left\|\epsilon-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\,\epsilon,t)\right\|^2\right]$$

⑧ 训练：通过 minimize 上述 L_{t-1} 优化网络

推理：每一步都从 $p_\theta(x_{t-1}|x_t)$ 采样得到 x_{t-1}

其中均值，方差