$$P_0(x) = \frac{1}{Z} e^{-\mathcal{E}(x)}, \quad Z = \int e^{-\mathcal{E}(x)} dx$$

$$\mathcal{E}(x): \mathbb{R}^d \to \mathbb{R}$$

$$x_t = x_0 + \sigma_t \epsilon, \quad \epsilon \sim N(0, I) \qquad P(z|x_0) \sim N(z; x_0, \sigma_t^2 I)$$

$$\beta_\Delta^2 = \sigma_t^2 - \sigma_{t-\Delta}^2$$

#2 $$P_t(z) = \int P_0(x_0)\, p(z|x_0)\, dx_0$$

$$= \int P_0(x_0)\, N(z; x_0, \beta_\Delta^2 I)\, dx_0$$

$$(f * g)(z) := \int f(x_0)\, g(z - x_0)\, dx_0$$

$$P_t(z) = \int P_0(x_0)\, \underbrace{N(z - x_0; 0, \sigma_t^2 I)}_{:= g(z - x_0)}\, dx_0$$

$$\boxed{= P_0 * N(0, \sigma_t^2 I)}$$

#1 反向核

$$p(\underbrace{x_{t-\Delta}}_{:=y} \mid \underbrace{x_t}_{:=x}) = \frac{p(x_t \mid x_{t-\Delta})\, p_{t-\Delta}(x_{t-\Delta})}{\underbrace{p_t(x_t)}_{C.}} \qquad \beta_\Delta^2 := \sigma_t^2 - \sigma_{t-\Delta}^2$$

$$\propto p(x \mid y)\, p_{t-\Delta}(y)$$

$$= N(x; y, \beta_\Delta^2 I)\, p_{t-\Delta}(y)$$

$$\log N(x; y, \beta_\Delta^2 I) = \log\left((2\pi\beta_\Delta^2)^{d/2} \exp\left(\frac{\|x-y\|^2}{2\beta_\Delta^2}\right)\right) \qquad u := y - x$$

$$= \underbrace{\log(2\pi\beta_\Delta^2)^{d/2}}_{C_1} - \frac{\|u\|^2}{2\beta_\Delta^2}$$

$$\log p_{t-\Delta}(y) = \log p_{t-\Delta}(x+u) = \log p_{t-\Delta}(x) + \underbrace{\nabla \log p_{t-\Delta}(x)}_{:=S}\, u + \cdots$$

$$\log p(y \mid x) = C_2 - \frac{\|u\|^2}{2\beta_\Delta^2} + Su + O(\Delta)$$

$$\frac{-\|u-a\|^2}{2\beta_\Delta^2} = \frac{-\|u\|^2}{2\beta_\Delta^2} + \frac{\cancel{2}au}{\cancel{2}\beta_\Delta^2} - \overset{C}{\cancel{\frac{\|a\|^2}{2\beta_\Delta^2}}}$$

$$S = \frac{a}{\beta_\Delta^2} \Rightarrow a = S\beta_\Delta^2 = g^2(t)\Delta$$

$$g^2(t) := \frac{d\sigma_t^2}{dt} \qquad \beta_\Delta^2 = \sigma_t^2 - \sigma_{t-\Delta}^2 \qquad \sigma_{t-\Delta}^2 = \sigma_t^2 - \frac{d\sigma_t^2}{dt}\Delta + O(\Delta)$$

$$\beta_\Delta^2 = \frac{d\sigma_t^2}{dt}\Delta$$

$$\nabla \log p_{t-\Delta}(x) = \nabla \log p_t + O(\Delta)$$

$$\log p(y \mid x) = -\frac{\|u-a\|^2}{2\beta_\Delta^2} + C + O(\Delta)$$

$$\sim N(y; a+x, \beta_\Delta^2 I)$$

写出来 $\nabla \log p_t(x)$

#3

$$p_t(z) = \int p_0(x_0)\ \mathcal{N}(z; x_0, \sigma_t^2 I)\, dx_0$$

$$= \int \frac{p_0(x_0)}{\cancel{\tfrac{1}{Z}}\ C}\ \mathcal{N}(z; x_0, \sigma_t^2 I)\, dx_0$$

$x_0 := z + \sigma_t \epsilon$ $\qquad$ $x_0 = z + \sigma_t \epsilon$

$\qquad\qquad\qquad\qquad$ $dx_0 = \sigma_t^d\, d\epsilon$

$$\propto \int e^{-\mathcal{E}(z + \sigma_t \epsilon)} \left(2\pi \cancel{\sigma_t^2}\right)^{d/2} \exp\left(-\frac{\|\cancel{\sigma_t}\epsilon\|^2}{2\cancel{\sigma_t^2}}\right) \cancel{\sigma_t^d}\, d\epsilon$$

$$\propto \int e^{-\mathcal{E}(z + \sigma_t \epsilon)}\ \mathcal{N}(0, I)\, d\epsilon$$

$$\propto \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)}\left[e^{-\mathcal{E}(z + \sigma_t \epsilon)}\right]$$

#4 $\nabla \log p_t = \frac{\nabla p_t}{p_t}$

$$\frac{d}{dx} p_t(x) = \frac{d}{dx} \mathbb{E}_{\epsilon \sim N(0,I)}\left[e^{-\mathcal{E}(x+\sigma_t \epsilon)}\right]$$

$$= \frac{d}{dx} \int f(x,\epsilon)\, \pi(\epsilon)\, d\epsilon$$

$$= \int \pi(\epsilon)\, \nabla_x f(x,\epsilon)\, d\epsilon$$

$$\nabla_x f(x,\epsilon) = \nabla_x e^{-\mathcal{E}(x+\sigma_t \epsilon)} = e^{-\mathcal{E}(x+\sigma_t \epsilon)} \cdot \underbrace{\nabla\left[-\mathcal{E}(x+\sigma_t \epsilon)\right]}_{-\nabla_x \mathcal{E}(x+\sigma_t \epsilon)}$$

$$\nabla_x \log p_t = \frac{\mathbb{E}_\epsilon\left[-e^{-\mathcal{E}(x+\sigma_t \epsilon)} \nabla_x \mathcal{E}(x+\sigma_t \epsilon)\right]}{\mathbb{E}_\epsilon\left[e^{-\mathcal{E}(x+\sigma_t \epsilon)}\right]}$$

#5

$S(x,t) := \nabla_x \log P_t(x)$

$\hat{S}_K(x,t)$: at x, t, with K samples

$$\hat{S}_K(x,t) = \frac{\sum_i^K \left[-e^{-\mathcal{E}(x+\sigma_t \epsilon_i)} \nabla \mathcal{E}(x+\sigma_t\epsilon_i) \right]}{\sum_i^K e^{-\mathcal{E}(x+\sigma_t\epsilon_i)}}$$

$\epsilon \overset{iid}{\sim} N(0, I)$

$w_i := e^{-\mathcal{E}(x+\sigma_t\epsilon_i)}$

$$\hat{S}_K(x,t) = \frac{\sum -w_i \nabla \mathcal{E}(x_i+\sigma_t\epsilon_i)]}{\sum w_i}$$

bias $O(1/K)$

$$\hat{S}_K = \frac{A}{B} \qquad \mathbb{E}\left[\frac{A}{B}\right] \neq \mathbb{E}[A]/\mathbb{E}[B]$$

$$\frac{A}{B} = \frac{\alpha + \delta_a}{\beta + \delta_B} = \frac{a + \delta_A}{\beta} \cdot \left(1 + \frac{\delta_B}{\beta}\right)^{-1}$$

$$(1+u)^{-1} = 1 - u + u^2 + \cdots$$

$$\text{bias} \approx \frac{\alpha}{\beta^3} Var(\beta) - \frac{1}{\beta^2} Cov(A,B)$$

$$Var,\ cov \propto \frac{1}{K}$$

ESS

$$ESS := \frac{(\Sigma_i w_i^2)}{\Sigma_i w_i} \in [1, K]$$

$$w_i \sim N(m, s^2),\ \mathbb{E}[w] = e^{m + s^2/2},\ \mathbb{E}[w^2] = e^{2m + 2s^2}$$

$$ESS \approx \frac{K^2 e^{2m+s^2}}{K e^{2m+2s^2}} = K e^{-s^2}$$

loss

$$\mathcal{L} = \mathbb{E}_t\left[\lambda(t) \| s_\theta(x,t) - \hat{S}_K(x,t) \|^2\right]$$

cost

n_train_batch × buffer × K