by definition

cdf is integral of pdf

cdf non invertible

$\sigma^2 << 1$, $X^{(n-1)} \perp\!\!\!\perp Z^{(n)}$

large n

$O(\sigma^2)$ taylor expansion centered at Y

$$\underset{\displaystyle X^{n}}{\overset{\displaystyle Y}{\shortparallel}} \,|\, \underset{\displaystyle X^{(n-1)}}{\overset{\displaystyle X}{\shortparallel}} \sim N(X^{(n-1)}, \sigma^2)$$

$$X^{(n-1)} | X^{n} \sim N(\mu_n(X^{(n)}), \sigma^2)$$

↳ also can be $\mathbb{E}[X^{(n-1)} | X^{(n)}]$

① train $\mu_n(X^{(n)})$ w/ NN

gaussian

from above $P_X(x)$

not depend on x

$= 0$

approx $e^x \approx 1+x$

newly added
isotropic

HW

HW

↳ diff is big, σ is not assume small, so cannot use $P_x(y)$

$S_T(X^{(n)}, n)$

directly compare $X^{(n)}$ to $X^{(n)}$

$\sim$ gaussian

do same

$= \exp(t^2/2)$

$\|$

MGF of

$U \sim N(0,1)$

$(1 - B_i)^{\text{large \#}} = 0$

$q \sim x^{(i)}$

tower property

markov property

$q(x^{(0:N)})$

standard gaussian

n=1 term

↳ bayes rule multiple times to prove

still want to min this

will cancel

some term cancel out

∇ w.r.t. $q(x^{(0)})$

goal of training match these 2 dists.

$D_{KL}(N(\mu_n(x^{(n)}|x^{(0)}), \beta_n)$

min $\rightarrow$

HW

HW

plug

$u_n(X^{(n)}|X^{(n)})$

$\underbrace{\quad}_{s\theta}$

$\underbrace{\qquad\qquad}_{u_\theta(X^{(n)}, n)}$

$$\sqrt{1-\beta} \approx 1 - \frac{1}{2}\beta$$

$$\beta_n \approx \beta_t \cdot dt$$

Unknown

take deriv

$(\partial P_t = (\partial_t \log P_t) \cdot P_t$

score fnc

bayes rule

p_t is unknown

$p_{t|0}$ is known

as $n \to \infty$

$\sim N(x_0, \frac{\sigma^2}{2\beta})$

can expand
session

① pick $t \sim \text{Unif}[0, T]$

② $X_t = \gamma_t X_0 - \sigma_t \varepsilon$

③ $\frac{\lambda(t)}{\sigma_t^2} \| \varepsilon_\theta(X_t, t) - \varepsilon \|^2$

④ update θ to minimize it

if t small, $\sigma_t \downarrow 0$

so only train $\varepsilon(X_t, t)$ up to $t \geq \delta$

$X_\delta \longleftrightarrow X_t$, as

$\varepsilon_t(X_t, t)$ for $t < \delta$ inaccurate.

$X^{(n)} \mid X^{(n-1)} \sim N(X^{(n-1)}, \sigma^2 I) \rightarrow \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \sigma^2 \end{bmatrix}$

same formulation as gradient case but gradients instead of derivitive

$\| X^{(0)} - \hat{X}_\theta(X^{(n)}, n) \|^2 \downarrow$

$\underbrace{\qquad\qquad}_{S_\theta(X^{(n)}, n)}$

can train model to estimate f_θ or S_θ

$$\mathbb{X}^{(n)} \underset{\theta}{\sim} N(0, I)$$

$$\varepsilon_\theta(\mathbb{X}^{(n)}, n) = \sqrt{1-\bar{\alpha}_n}\, s_\theta(\mathbb{X}^{(n)}, n)$$

$\mathbb{R}^d$

$\mathbb{R}^d$ $\mathbb{R}^{d\times d}$ $\mathbb{R}^d$

diagonal

?, we don't know

$$\int_0^T \lambda(t) \| s_\theta(x_t, t) - \nabla_t \log p_t(x_t) \|^2$$

→ deterministic generation

control amount of noise added in reverse process

Generate $x^{(T)}, x^{(n)}$ first

Then generate $x^{(n-1)}, \dots x^{(1)}$

↘ from DDPM training

$\hat{x}^{(0)}$

↳ noise

sampling

① $x_N \sim N(0, I)$

② $x_n \sim q(x_n \mid x_{n+1}, x_0 = \hat{x}_0)$

additional NN $\rightarrow$

often hard to train $\rightarrow$ classifie x_t $0 \le t \le N$

$* \; P(z_t, \Omega(X_t) \mid \Omega(\bar{X}_0))$

marginalization

approx

$$\underbrace{\nabla \log p_t(z_t, \Omega(X_t))}_{s_\theta((z_t, \Omega(X_t)), t)}$$

<u>unconditional</u>

1. $X^{(0)}$
2. $X^{(0)} \rightarrow X^{(t)}$ (w/ ε_t)
3. $s_\theta(X^{(t)}, t)$ $\leftrightarrow$ $\varepsilon_\theta(X^{(t)}, t)$
4. $\| \varepsilon_\theta(X^{(t)}, t) - \varepsilon_t \| \downarrow$

conditional

① $x^{(0)}, y$

② $x^{(0)} \to x^{(t)}$ (w/ ε_t)

③ $S_\theta(x_t, y, t)$

$\varepsilon_\theta(x_t, y, t)$

④ $\|\varepsilon_\theta(x_t, y, t) - \varepsilon_t\|^2 \downarrow$

NN to estimate

learn this

want $\to 0$

don't care

Goal: find θ

hyperparameter $\to 0$

$\to 0$

minimized $S_\theta(x_t, t)_y = \frac{p_t(y)}{p_t(x_t)}$

need to sample x_t, $S_\theta(x_t, t)$

$= [\ \ ] \leftarrow S_\theta(x_t, t)_y$

con control

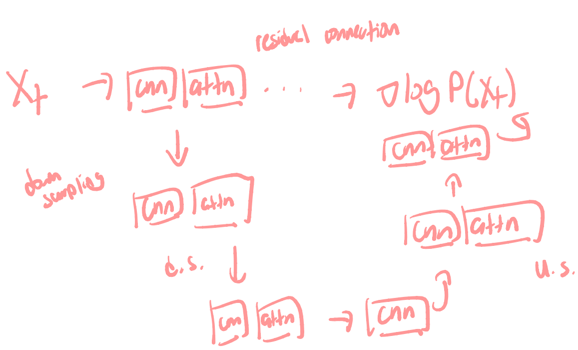
$S_\theta(y, t) = [\ \ ] \rightarrow$ x_t th component $S_\theta(y, t)_{x_t}$

infeasible if $k \gg$

replace w/ conditional similar to continuous case

$\| S_\theta(x_t, t) - \nabla \log p_t(x_t) \|^2$

$\Rightarrow \| S_\theta(x_t, t) - \nabla \log p_{t|0}(x_t|x_0) \|^2$


residual connection
X_t
cnn
attn
$\nabla \log P(X_t)$
down sampling
d.s.
u.s.

$S_\theta(X_t, y, t)$