セミナーの内容

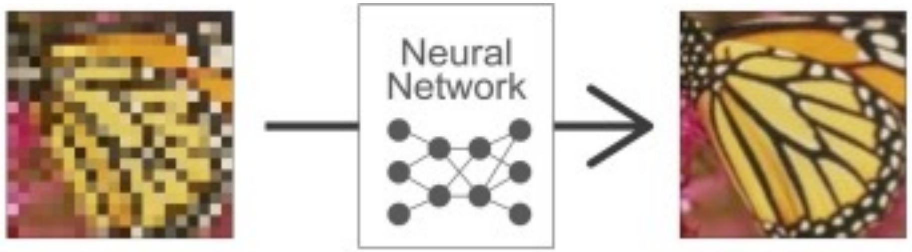
- 拡散モデルを利用する動機と確率過程を学ぶ動機
- 確率過程の説明
- 順過程から逆過程への変形
- 拡散モデルの学習と生成
- 条件付き生成問題への応用

* 理論ノートで引用している文献は角括弧で番号のみを記す (例: [1])

データ生成と微分方程式

■ データ生成

- 画像や文章などを ML モデルで作成
- 現代の AI 技術の基礎
- 既知データに変換を施す
- 複雑なデータの生成は変換も複雑



■ 微分方程式の応用

- 微小変換の積み重ねでデータを生成
- 微分方程式の数値積分と類似
  - 正規化流 [9] や Neural ODE [10]

■ 確率微分方程式 (SDE) の応用

- 拡散モデルは SDE で生成を行う[3]

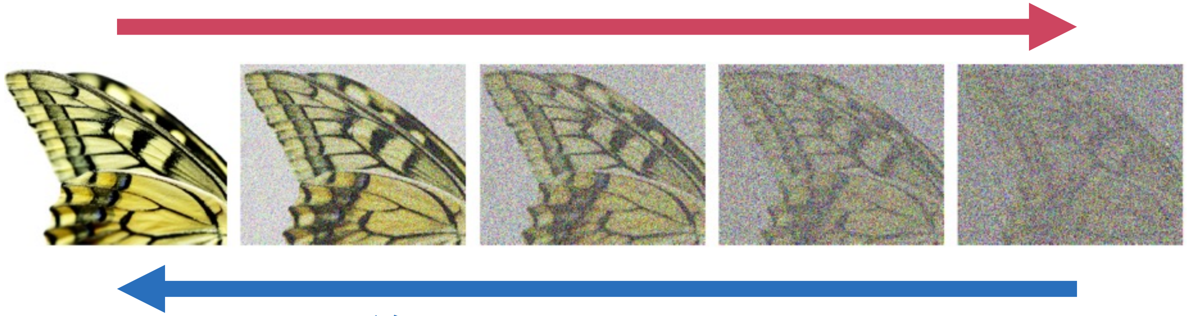
$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, t)$$

$$\boldsymbol{x}_{t+\Delta t} = \boldsymbol{x}_t + \boldsymbol{f}(\boldsymbol{x}_t, t)\,\Delta t$$

拡散モデルのデータ生成の概要

- 順過程と逆過程の存在[3,12]
  - 順過程: 拡散過程
    - データをノイズに変換
  - 逆過程: 順過程の逆回し
    - ノイズからデータを生成

順過程 ($t = 0 \to t = 1$)

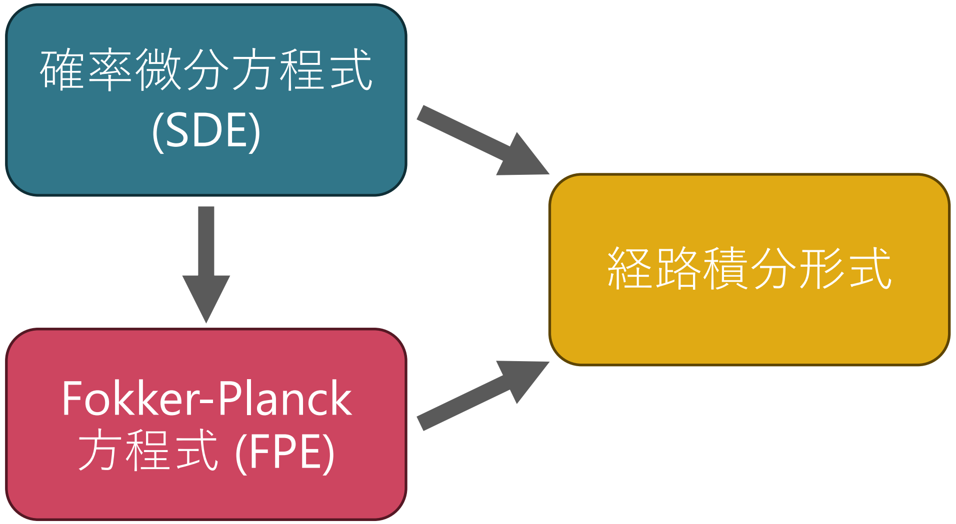


逆過程 ($t = 1 \to t = 0$)

- いくつかの疑問
  1. 任意のデータをノイズに変換可能？
  2. どうやって時間反転する？
  3. SDE をどうやって学習させる？

確率過程の全体像[5,6]

- 確率微分方程式[5,6]
  - ランダム系の時間発展式
  - 例え) サイコロの次に出る目の予測式
  - 予測はランダムで試行毎に異なる
- Fokker-Planck 方程式 (FPE)[5,6]
  - アンサンブルの確率分布の時間発展
  - 例え) サイコロの確率分布の発展式
  - 予測は決定論的でランダム性はない
- 経路積分形式[6]
  - 各パス (SDE の解) の確率を記述
  - セミナーでは扱わない



確率微分方程式 (SDE)

- 拡散モデルの原理における主役
  - $\mathrm{d}x = -ax\,\mathrm{d}t + b\,\mathrm{d}W$
- もしノイズ (第二項) がなければ
  - $x_t = x_0 e^{-at}$
  - 初期条件を 0 へと指数的に減衰させる
  - $b\,\mathrm{d}W$ がある場合，正規ノイズへ緩和させる

任意のデータを
ノイズへと変換する仕組み

- 一変数系でよいのか？
  - x は画像のあるピクセル値に対応
  - 拡散モデルは，各ピクセルを独立に減衰させ，ノイズへ緩和させる[1,2]
    - 相関構造を無視して破壊し，各成分が独立な正規ノイズへ緩和させる
  - 相関構造はスコア関数を通して復元 (生成) される[1,2,3]

ノイズ $\mathrm{d}W$ の統計性

■ 数値的に方程式を解く

- $\mathrm{d}W \to \xi_k$ と置き換えて数値的に解く
- 乱数 ξ_k を得れば，漸化式が解ける

$$\mathrm{d}x = -ax\,\mathrm{d}t + b\,\mathrm{d}W$$

$$\Rightarrow \quad x_{k+1} = x_k + (-ax_k)\Delta t + \xi_k$$

■ 乱数 ξ の性質は？

- 正規分布に従う $\xi \sim \mathcal{N}(0, \nu^2)$
- $\nu^2 = (\Delta t)^2$ or $\nu^2 = \Delta t$?
  - $\Delta t = \frac{1}{N}$
  - 分散の総和が発散しない条件は？
  - $\nu^2 = \Delta t$

$$\mathbb{E}\left[\sum_k \xi_k^2\right] = \sum_k \mathbb{E}\left[\xi_k^2\right] = \sum_k \nu^2 = N\nu^2$$

$$x_{k+1} = x_k + (-ax_k)\Delta t + b\epsilon_k\sqrt{\Delta t}$$

$$\Rightarrow \quad \mathrm{d}x = -ax\,\mathrm{d}t + b\,\mathrm{d}W$$

■ $\mathrm{d}W$ の計算規則[5,6]

- $(\mathrm{d}W_t)^2 = \mathrm{d}t$
- $\mathrm{d}W_{t_1}\mathrm{d}W_{t_2} = 0$

なぜ $(\mathrm{d}W_t)^2$ は確定値 $\mathrm{d}t$ となる？
微小擾乱が足し合わされて大数の法則が働く [6]

[補足] 伊藤の公式

$$\begin{aligned}
\mathrm{d}f &= \frac{\mathrm{d}f}{\mathrm{d}x}\mathrm{d}x + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(\mathrm{d}x)^2 + O((\mathrm{d}x)^3) && (14)\\
&= \frac{\mathrm{d}f}{\mathrm{d}x}\left[-ax\,\mathrm{d}t + b\,\mathrm{d}W\right] + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\left[-ax\,\mathrm{d}t + b\,\mathrm{d}W\right]^2 + O((\mathrm{d}x)^3) && (15)\\
&= \frac{\mathrm{d}f}{\mathrm{d}x}\left[-ax\,\mathrm{d}t + b\,\mathrm{d}W\right] + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\left[(ax)^2(\mathrm{d}t)^2 - 2abx(\mathrm{d}t\,\mathrm{d}W) + b^2\mathrm{d}t\right] + O((\mathrm{d}x)^3) && (16)\\
&= \left[\frac{\mathrm{d}f}{\mathrm{d}x}(-ax) + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}b^2\right]\mathrm{d}t + \frac{\mathrm{d}f}{\mathrm{d}x}(b\,\mathrm{d}W) + o(\mathrm{d}t) && (17)
\end{aligned}$$

- 任意関数 $f(x)$ の変分[5,6]
  - 通常の微分積分学の計算では $\mathrm{d}f = f'\mathrm{d}t$ となり，一階微分まで現れる
  - 確率過程では $(\mathrm{d}W_t)^2 = \mathrm{d}t$ の特異性により，二階微分まで必要となる
  - 確率分布の「拡散」の源がこの特異性である

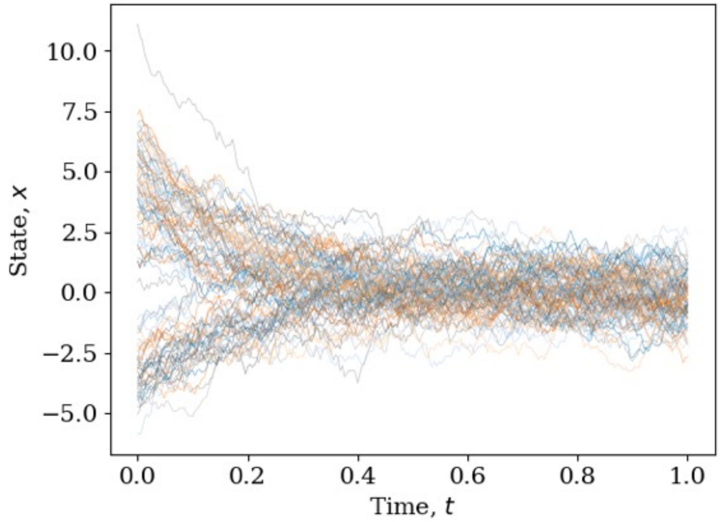
SDE の解析解[5]

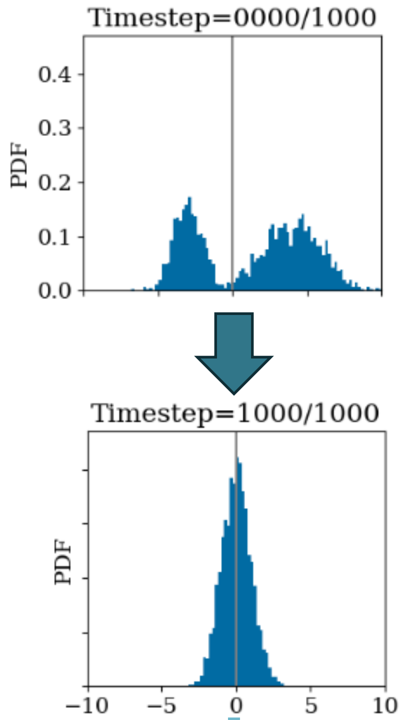
- 拡散モデルの実装で直接利用[2,3]
  - 第一項: 摩擦による初期条件の減衰
  - 第二項: ノイズ $\mathrm{d}W$ の積算値
- 第二項の性質
  - 正規ノイズの和は正規分布に従う (再生性)
  - ノイズの和を標準正規乱数 ϵ で書く
- 疑問とその答え
  - 任意のデータをノイズに変換可能？
  - 任意のデータに対して，摩擦により指数関数的に減衰させ，ノイズを支配的にする

$$\mathrm{d}x = -ax\,\mathrm{d}t + b\,\mathrm{d}W$$

$$\begin{aligned} x_t &= e^{-at}x_0 + \int_0^t be^{-a(t-t')}\mathrm{d}W_{t'} \\ &= e^{-at}x_0 + \sigma_t\epsilon \end{aligned}$$

$$\sigma_t^2 := \frac{b^2}{2a}\left[1 - e^{-2at}\right]$$





アンサンブルの確率分布の時間発展[5,6]

Fokker-Planck 方程式 (FPE) の導入

$$dx = -ax\,dt \boxed{+\,b\,dW}$$

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[\mathrm{axp}(x,t)] + \boxed{\frac{b^2}{2}\frac{\partial^2}{\partial x^2}p(x,t)}$$

- 全ての実現可能性を集めて確率分布を構成
- FPE はその時間発展を記述

ノイズと拡散項の対応関係

↑

この対応関係を壊さないように
時間反転を行う

FPE の解釈

- 第一項: 決定論的な移流
- 第二項: ノイズによる不確実性の増大
  (確率分布の拡散項)

FPE は不可逆な拡散現象を記述するため
方程式の形を変更しないとダメ

データ分布が正規分布へ変換される時間発展を FPE は記述[5]

順過程から逆過程へ

- 式変形の要請
  - 順過程の FPE の解を逆回ししたものが逆過程の方程式の解である
  - 逆過程の方程式にも拡散項が正の符号を持って存在している
- 何が問題か？
  - 素朴に時間反転をするとマイナス倍されて拡散項に負の符号がつく
- 式変形の説明
  - $p^R(x, t^R)$: FPE の解を逆回ししたもの
  - 見かけの形が異なるが $p(x,t)$ と等価
  - 両関数共に順 FPE の解である

$$\mathrm{d}x = -ax\,\mathrm{d}t \boxed{+\,b\,\mathrm{d}W}$$

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[\mathrm{axp}(x,t)] + \boxed{\frac{b^2}{2}\frac{\partial^2}{\partial x^2}p(x,t)}$$

$$\frac{\partial}{\partial t^{\mathrm{R}}}p = \frac{\partial t}{\partial t^{\mathrm{R}}}\frac{\partial}{\partial t}p = -\frac{\partial}{\partial t}p$$

$$\boxed{t^{\mathrm{R}} := (T - t)}$$

$$\begin{aligned}
\frac{\partial}{\partial t^{\mathrm{R}}}p^{\mathrm{R}}(x,t^{\mathrm{R}}) &= -\frac{\partial}{\partial t}p(x,t) \\
&= -\frac{\partial}{\partial x}[axp(x,t)] - \frac{b^2}{2}\frac{\partial^2}{\partial x^2}p(x,t) \\
&= -\frac{\partial}{\partial x}[axp^{\mathrm{R}}(x,t^{\mathrm{R}})] - \frac{b^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}(x,t^{\mathrm{R}})
\end{aligned}$$

$$\boxed{p^{\mathrm{R}}(x,t^{\mathrm{R}}) := p(x,t)|_{t=T-t^{\mathrm{R}}}}$$

スコア関数の導入

■ 0 を足し算して拡散項をつくる[8]

- 恒等的な式変形であり，p^R は順 FPE の解のまま
- 最終的な結果が逆過程の FPE となる

$$\frac{\partial}{\partial t^{\mathrm{R}}} p^{\mathrm{R}} = -\frac{\partial}{\partial x}[axp^{\mathrm{R}}] - \frac{b^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}} + \underbrace{\left[-\frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}} + \frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}\right]}_{=0}$$

$$= \frac{\partial}{\partial x}\left[-axp^{\mathrm{R}} - \left(\frac{b^2}{2} + \frac{c^2}{2}\right)\frac{\partial}{\partial x}p^{\mathrm{R}}\right] + \frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}$$

$$= \frac{\partial}{\partial x}\left[p^{\mathrm{R}}\left\{-ax - \frac{b^2+c^2}{2p^{\mathrm{R}}}\frac{\partial}{\partial x}p^{\mathrm{R}}\right\}\right] + \frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}$$

$$= -\frac{\partial}{\partial x}\left[p^{\mathrm{R}}\left\{ax + \frac{b^2+c^2}{2}\frac{\partial}{\partial x}\ln p^{\mathrm{R}}\right\}\right] + \frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}$$

■ 逆過程の FPE と SDE

- SDE は，FPE との対応関係から導出
- スコア関数: $\partial_x \ln p(x,t)$
- 拡散項の前の符号は正である
- p^R は順 FPE と逆 FPE の解であり
  順/逆過程の記述する確率分布は同じ
- 逆過程は正規分布をデータ分布へと時間発展させる

$$\mathrm{d}x = \left[ax + \left(\frac{b^2+c^2}{2}\right)\frac{\partial}{\partial x}\ln p^{\mathrm{R}}\right]\mathrm{d}t^{\mathrm{R}} + c\,\mathrm{d}W^{\mathrm{R}}$$

$$\frac{\partial p^{\mathrm{R}}}{\partial t^{\mathrm{R}}} = -\frac{\partial}{\partial x}\left[p^{\mathrm{R}}\left\{ax + \left(\frac{b^2+c^2}{2}\right)\frac{\partial}{\partial x}\ln p^{\mathrm{R}}\right\}\right] + \frac{c^2}{2}\frac{\partial^2}{\partial x^2}p^{\mathrm{R}}$$

逆過程は何をやっているのか？

■ 逆過程の SDE は閉じていない

- $p(x,t)$ を知らないとダメ
- 不可逆系を無理やり時間反転した代償

$$p(x,t) = \frac{1}{Z(t)} \exp[-U(x,t)]$$

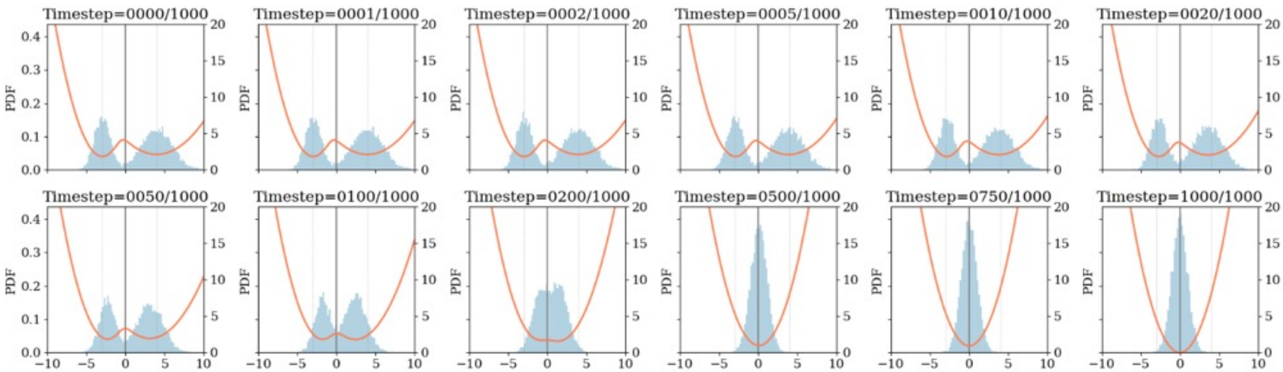
$$Z(t) = \int \mathrm{d}x \exp[-U(x,t)]$$

■ 逆 SDE のダイナミクス

- ポテンシャルの導入
- 逆過程はポテンシャルの低い向きへ引っ張る勾配系 (Gradient System, [16])
- ポテンシャルは段々複雑になっていく
- 画像の場合:
  - 順過程: 画像の構造を破壊しつつ正規ノイズへ緩和
  - 逆過程: スコア関数を通して画像の構造を復元

$$\mathrm{d}x = \left[x + \frac{\partial}{\partial x} \ln p^{\mathrm{R}}\left(x, t^{\mathrm{R}}\right)\right] \mathrm{d}t^{\mathrm{R}} + \mathrm{d}W^{\mathrm{R}}$$

$$\mathrm{d}x = \left[x - \frac{\partial}{\partial x} U^{\mathrm{R}}\left(x, t^{\mathrm{R}}\right)\right] \mathrm{d}t^{\mathrm{R}} + \mathrm{d}W^{\mathrm{R}}$$



確率過程から機械学習へ

損失関数の素朴な変形 (1/2)

何がやりたいか？

- スコア関数を答えとして損失関数を構成 (別名，目的関数)
- 損失関数による最適化でスコア関数を推定したい
- しかし，答えとなるスコア関数がそもそも求められない・・・

スコア関数の推定

- ここから機械学習の話題
  - スコア関数の推定
- スコア関数を解析的に求めてみる
  - 条件付き分布 $p(x_t|x_0)$ を使って書く
  - $p(x_t|x_0)$ は正規分布となる
    - 初期条件を固定している点に注意

$$\left.\begin{aligned} \mathrm{d}x_t &= -ax_t\,\mathrm{d}t + b\,\mathrm{d}W_t \\ x_t &= e^{-at}x_0 + \sigma_t\epsilon \\ \sigma_t^2 &:= \frac{b^2}{2a}\left[1 - e^{-2at}\right] \\ \epsilon &\sim \mathcal{N}(0,1) \end{aligned}\right\}$$

x_0 の固定により
x_t は正規変数に

$$p(x,t \mid x_0) = \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{1}{2}\frac{\left(x - e^{-at}x_0\right)^2}{\sigma_t^2}\right]$$

$$\begin{aligned} \frac{\partial}{\partial x}\ln p(x,t) &= \frac{\partial_x p(x,t)}{p(x,t)} \\ &= \frac{\int\left[\partial_x p(x,t \mid x_0)\right]p(x_0)\,\mathrm{d}x_0}{p(x,t)} \\ &= \int \frac{\partial_x p(x,t \mid x_0)}{p(x,t \mid x_0)}\underbrace{\frac{p(x,t \mid x_0)p(x_0)}{p(x,t)}}_{=p(x_0|x_t)}\mathrm{d}x_0 \\ &= \int\left[\partial_{x_t}\ln p(x_t \mid x_0)\right]p(x_0|x_t)\,\mathrm{d}x_0 \\ &= \int \mathrm{d}x_0\,p(x_0|x_t)\left[\partial_{x_t}\ln p(x_t|x_0)\right] \\ &= \int \mathrm{d}x_0\,p(x_0|x_t)\left[-\frac{x_t - e^{-at}x_0}{\sigma_t^2}\right] \end{aligned}$$

条件付き期待値

$$\frac{\partial}{\partial x}\ln p(x,t) = -\frac{x_t - e^{-at}\boxed{\mathbb{E}\left[x_0 \mid x_t\right]}}{\sigma_t^2}$$

簡単にはスコア関数は推定できなさそう

損失関数の素朴な変形 (2/2)

- とりあえず，愚直に損失関数を設定

$$\mathcal{L}_{\boldsymbol{\theta}} = \frac{1}{T}\mathbb{E}\left[\left\|\hat{s}_{\boldsymbol{\theta}}(x,t) - \frac{\partial}{\partial x}\ln p(x,t)\right\|^2\right]$$

$\hat{s}_\theta(s,t)$: ニューラルネットワーク

$$\mathcal{L}_{\boldsymbol{\theta}} = \int\left[\hat{s}_{\boldsymbol{\theta}}^2 - 2\hat{s}_{\boldsymbol{\theta}}\frac{\partial}{\partial x}\ln p(x,t) + \left(\frac{\partial}{\partial x}\ln p(x,t)\right)^2\right]p(x,t)dx$$

  - 困難になるのは第二項だけ

- 第二項だけ変形してみる[2,7,13]
  - 条件付き分布を使って式変形

$$p\left(x,t \mid x_0\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{1}{2}\frac{\left(x - e^{-at}x_0\right)^2}{\sigma_t^2}\right]$$

$$\begin{aligned}
\int \mathrm{d}x\, p(x,t)\left[\hat{s}_{\boldsymbol{\theta}}\frac{\partial}{\partial x}\ln p(x,t)\right] &= \int \mathrm{d}x\, p(x,t)\left[\hat{s}_{\boldsymbol{\theta}}\frac{1}{p(x,t)}\frac{\partial}{\partial x}p(x,t)\right] \\
&= \int \mathrm{d}x\left[\hat{s}_{\boldsymbol{\theta}}\frac{\partial}{\partial x}\int \mathrm{d}x_0\, p(x,t|x_0)p(x_0)\right] \\
&= \int \mathrm{d}x\mathrm{d}x_0\, p(x_0)\left[\hat{s}_{\boldsymbol{\theta}}\frac{\partial}{\partial x}p(x,t|x_0)\right] \\
&= \int \mathrm{d}x\mathrm{d}x_0\, p(x_0)\left[\hat{s}_{\boldsymbol{\theta}}\frac{p(x,t|x_0)}{p(x,t|x_0)}\frac{\partial}{\partial x}p(x,t|x_0)\right] \\
&= \int \mathrm{d}x\mathrm{d}x_0\, p(x,t|x_0)p(x_0)\left[\hat{s}_{\boldsymbol{\theta}}\frac{1}{p(x,t|x_0)}\frac{\partial}{\partial x}p(x,t|x_0)\right] \\
&= \int \mathrm{d}x\mathrm{d}x_0\, p(x,t|x_0)p(x_0)\left[\hat{s}_{\boldsymbol{\theta}}\frac{\partial}{\partial x}\ln p(x,t|x_0)\right]
\end{aligned}$$

- 条件付きスコア関数の推定問題に

$$\frac{1}{T}\int\left\|\hat{s}_{\boldsymbol{\theta}} - \frac{\partial}{\partial x}\ln p(x,t|x_0)\right\|^2 p(x,t|x_0)p(x_0)\,\mathrm{d}x\mathrm{d}x_0\mathrm{d}t$$

デノイジングスコアマッチング (DSM)

- ノイズの推定問題に[2,7,13]
  - 条件付き分布と SDE の解を代入
  - ノイズ推定の損失関数が得られる
  - 期待値の評価: モンテカルロ法の利用
- なぜノイズからスコアが分かる？
  - $p(x,t)$: 正規分布 $p(x,t|x_0)$ の重ね合わせ
  - 正規分布に対して，スコア関数は摂動 (ノイズ) そのものになる
  - ノイズの平均的な傾向の推定がスコア関数の推定となる

$$p\left(x, t \mid x_0\right)=\frac{1}{\sqrt{2 \pi \sigma_t^2}} \exp \left[-\frac{1}{2} \frac{\left(x-e^{-a t} x_0\right)^2}{\sigma_t^2}\right]$$

$$x-e^{-at}x_0=\sigma_t\epsilon$$

対応する条件付き確率

SDE の解は，減衰する初期条件 + 正規ノイズ

$$\frac{1}{T}\int\left\|\hat{s}_{\boldsymbol{\theta}}-\frac{\partial}{\partial x}\ln p(x,t|x_0)\right\|^2 p(x,t|x_0)p(x_0)\,\mathrm{d}x\mathrm{d}x_0\mathrm{d}t$$

$$\mathcal{L}_{\boldsymbol{\theta}}^{\mathrm{DSM}}=\frac{1}{N}\sum_i\left\|\hat{s}_{\boldsymbol{\theta}}(x^{(i)},t^{(i)})+\frac{\epsilon^{(i)}}{\sigma_{t^{(i)}}}\right\|^2$$

よくある書き換え $\hat{s}_{\boldsymbol{\theta}}(x,t)=-\frac{\hat{\epsilon}(x,t)}{\sigma_t}$

$$\mathcal{L}_{\boldsymbol{\theta}}^{\mathrm{DSM}}=\frac{1}{N}\sum_i\frac{\left\|\hat{\epsilon}(x^{(i)},t^{(i)})-\epsilon^{(i)}\right\|^2}{\sigma_{t^{(i)}}^2}$$

拡散モデルの学習とデータ生成

拡散モデルの学習[2,3]

- 順 SDE に従いデータサンプルをにノイズ ϵ を加える
- 加えたノイズ ϵ の推定問題を解きスコア関数を推定

$$\mathrm{d}x_t = -ax_t\,\mathrm{d}t + b\,\mathrm{d}W_t$$

$$x_t = e^{-at}x_0 + \sigma_t\epsilon$$

$$\sigma_t^2 := \frac{b^2}{2a}\left[1 - e^{-2at}\right]$$

$$\epsilon \sim \mathcal{N}(0,1)$$

拡散モデルのデータ生成[2,3]

- 逆 SDE を積分して正規ノイズをデータに変換する
- 逆 SDE の中のスコア関数がニューラルネットワークによる推定となる
- 逆過程のダイナミクスはポテンシャルを使った勾配系として理解できる

$$\mathrm{d}x = \left[ax - \left(\frac{b^2+c^2}{2}\right)\frac{\hat{\epsilon}(x,t)}{\sigma_t}\right]\mathrm{d}t^{\mathrm{R}} + c\,\mathrm{d}W^{\mathrm{R}}$$

$$\hat{s}_{\boldsymbol{\theta}}(x,t) = -\frac{\hat{\epsilon}(x,t)}{\sigma_t}$$

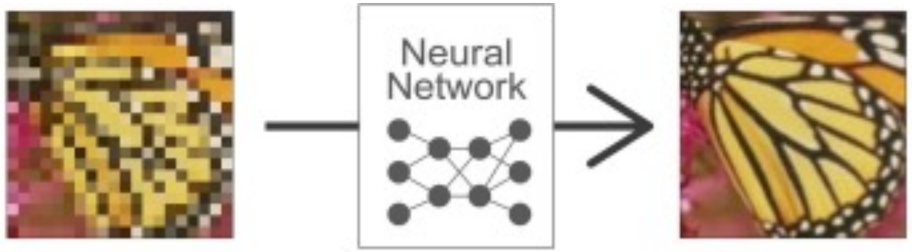
疑問への答え

- 任意のデータを正規ノイズへ時間発展させるには SDE をどう設定するか？
  - 線形摩擦を含む SDE を設定し，任意のデータを指数関数的に減衰させる
- この SDE を時間について逆回しする方法はなにか？
  - スコア関数を導入することで，順過程と同じ確率分布を解とする逆過程を構成する
- そもそも SDE をどうやってデータから決定（または学習）するか？
  - デノイジングスコアマッチングによりノイズの推定問題を通してスコア関数の近似を得る

条件付き生成問題への応用

条件付き生成問題とは

- 事後分布 $p(x|y)$ の推定問題[3,14]
  - x: 予測変数
  - y: 条件 (固定する変数)
    - 低解像度画像，観測値など
  - 複数の x があり得る → $p(x|y)$
- Bayes の定理の応用[3,14]
  - 尤度 $p(y|x)$ を変換
    - $p(y|x)$ は陽に得られることが多い
- 無条件スコア関数を利用できないか
  - あくまで条件付き生成を行うための方法の一つ[3,14]を説明



$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(y) := \int \mathrm{d}x \; p(y|x)p(x)$$

アイディア: 拡散モデルで無条件スコアを推定[3,14]

学習時

- y を考慮せずに無条件スコアを推定
- 今まで通りに拡散モデルを学習

推論時

- 無条件スコアに対数尤度の微分を足す
- 事後分布に従うデータ生成ができる？

尤度が正規分布の例

- x を y に緩和させる項が出てくる
- y の情報を反映した x が生成できそう
- 生成された x は事後分布に従う？

$$\frac{\partial}{\partial x}\ln p(x|y) = \frac{\partial}{\partial x}\ln p(y|x) + \frac{\partial}{\partial x}\ln p(x)$$

$$\mathrm{d}x = \left[x + \frac{\partial}{\partial x}\boxed{\left\{\ln p^{\mathrm{R}}(x,t) + \ln p(y|x)\right\}}\right]\mathrm{d}t^{\mathrm{R}} + \mathrm{d}W^{\mathrm{R}}$$

$$p(y|x) = \frac{1}{\sqrt{2\pi\lambda^2}}\exp\left[-\frac{1}{2}\frac{(y-x)^2}{\lambda^2}\right]$$

$$\frac{\partial}{\partial x}\ln p(y|x) = -\frac{x-y}{\lambda^2}$$

理論の修正が必要[3,14]

- なぜか？
  - SDE では t ごとに x の「汚れ具合」が異なる
- t に応じて尤度 $p(y|x_t)$ を修正する
  - クリーンなデータ x_0 に対して定義された尤度 $p(y|x_0)$ を任意の t 二適用可能にする
  - 事後分布 $p(x_t|y)$ の変形により平均尤度 $l(y|x_t)$ を使えばよいことが分かる

$$l(y \mid x_t) = \int p(y \mid x_0) p(x_0 \mid x_t) \mathrm{d}x_0$$

$$\begin{aligned}
p(x_t \mid y) &= \int p(x_t \mid x_0) p(x_0 \mid y) \mathrm{d}x_0 \\
&= \int p(x_t \mid x_0) \frac{p(y|x_0)p(x_0)}{p(y)} \mathrm{d}x_0 \\
&= \frac{1}{p(y)} \int p(y \mid x_0) p(x_t|x_0) p(x_0) \frac{p(x_t)}{p(x_t)} \mathrm{d}x_0 \\
&= \frac{p(x_t)}{p(y)} \int p(y \mid x_0) \frac{p(x_t \mid x_0) p(x_0)}{p(x_t)} \mathrm{d}x_0 \\
&= \frac{p(x_t)}{p(y)} \int p(y \mid x_0) p(x_0 \mid x_t) \mathrm{d}x_0 \\
&= \frac{1}{p(y)} l(y \mid x_t) p(x_t)
\end{aligned}$$

近似の必要性[14]

- 今までの議論
  - 上のすべての議論は非正規分布をもつデータに対して適用可能
  - 正規性の近似は一切用いていない
- 平均尤度を正規近似から求める[14]

$$p(y \mid x_0) = \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left[-\frac{1}{2}\frac{(y-x_0)^2}{\lambda^2}\right] = \mathcal{N}\left(y; x_0, \lambda^2\right)$$

$$p(x_0 \mid x_t) = \mathcal{N}\left(x_0; \mathbb{E}[x_0 \mid x_t], \mathrm{Var}[x_0 \mid x_t]\right)$$

  - 平均尤度も正規分布になる

$$l(y|x_t) = \mathcal{N}\left(y; \mathbb{E}[x_0 \mid x_t], \lambda^2 + \mathrm{Var}[x_0 \mid x_t]\right)$$

$l(y|x_t)$ を生成に用いれば，無条件スコアの学習のみで追加の訓練なしに事後分布 $p(x_0|y)$ に従うデータを生成

$$\mathrm{d}x_t = \left[a x_t + \left(\frac{b^2+c^2}{2}\right)\frac{\partial}{\partial x_t}\left\{\ln l(y \mid x_t) + \ln p^{\mathrm{R}}(x_t)\right\}\right]\mathrm{d}t^{\mathrm{R}} + c\,\mathrm{d}W_t^{\mathrm{R}}$$

$$\begin{aligned}
l(y \mid x_t) &= \int p(y \mid x_0) p(x_0 \mid x_t) \mathrm{d}x_0 \\
&= \int \mathcal{N}(y; x_0, \lambda^2) \mathcal{N}(x_0; \mathbb{E}[x_0|x_t], \mathrm{Var}[x_0|x_t]) \mathrm{d}x_0 \\
&= \frac{1}{2\pi\sqrt{\lambda^2 \mathrm{Var}[x_0|x_t]}} \int \mathrm{d}x_0 \exp\left[-\frac{1}{2}\frac{(x_0-y)^2}{\lambda^2}\right] \exp\left[-\frac{1}{2}\frac{(x_0-\mathbb{E}[x_0|x_t])^2}{\mathrm{Var}[x_0|x_t]}\right] \\
&= \frac{1}{\sqrt{2\pi(\lambda^2+\mathrm{Var}[x_0|x_t])}} \exp\left[-\frac{1}{2}\frac{(y-\mathbb{E}[x_0|x_t])^2}{\lambda^2+\mathrm{Var}[x_0|x_t]}\right] \\
&= N(y; \mathbb{E}[x_0|x_t], \lambda^2 + \mathrm{Var}[x_0|x_t])
\end{aligned}$$