ゴール

- 拡散モデルの数理を流し読みして、量子コンピューティング (量子アニーリング) と関係ありそうだと思った
  - ので、これの数理的な関連をざっくりと理解したい
  - 年内いっぱいくらいで「わかった」という気分になりたい
  - なんでかというと、今から量子コンピューティングの勉強にリソース割いておくかどうかを見定めたい

関連しそうな項目

- 拡散モデル
  - 確率過程
- シミュレーティッドアニーリング
- 量子アニーリング
- 代数的確率空間
  - 量子力学においては、行列(=作用素)の trace が期待値
  - この辺を抽象化して、代数の元に「確率的」な情報が含まれているとみなしてゴニョゴニョするのが確率的代数空間 (イメージ)
  - これを理解できると「拡散モデルのマルコフ過程は代数的確率空間的にはこう」みたいな話から「じゃあこんな種類の代数構造から考えたら拡散モデルに当たるものは何?」みたいな議論ができて面白いのでは、という仮説

課題感

- 大規模 LLM は大手に独占されている
  - GPT 4 はパラメーター数が 170 兆
  - 圧倒的データ量

教科書

- **確率論:** https://www.amazon.co.jp/gp/product/4254116004/ref=ox_sc_act_image_1?smid=AN1VRQENFRJN5&psc=1
- **確率微分方程式とその応用:** https://www.amazon.co.jp/gp/product/4627077815/ref=ox_sc_act_image_2?smid=AN1VRQENFRJN5&psc=1

論文まとめ

Deep Unsupervised Learning using Nonequilibrium Thermodynamics
非平衡熱力学を用いた教師なしディープラーニング

一般定理

ガウス積分

$$\int_{-\infty}^{\infty} \exp\left(-a(x-b)^2\right) \mathrm{d}x = \sqrt{\frac{\pi}{a}} \quad \Re\{a\} > 0, b \in \mathbb{C}$$

用語

可測空間

- 組み (Ω, F)
  - Ω : set
  - F : 完全加法族
    - 完全加法性を満たす、Ωの冪集合の部分集合

可測写像

- $(\Omega, F), (\Psi, G)$: 可測空間
  - $f : (\Omega, F) \to (\Psi, G)$ で以下を満たす
  - $\forall A \in G \; f^{-1}(A) \in F$
  - ** $f : (\Omega, F) \to (\Psi, G)$ こう書いたときの実態は $f : F \to G$

測度空間

- 組み(Ω, F, μ)
  - $\mu : F \to \mathbb{R}_{\geq 0}$
    - 可算加法性を満たす
      - $\forall i, j \in \mathbb{N}, A_i, A_j \in F \; A_i \cup A_j = \text{empty} \Rightarrow \mu(\cup_i A_i) = \sum_i \mu(A_i)$

実数 (測度空間の具体例)

- $(\mathbb{R}, B(\mathbb{R}), \mu)$
  - $B(\mathbb{R})$: ボレル集合族
    - RR 上の距離から定まる位相を含む最小の完全加法族
  - μ : ルベーグ測度
    - 距離から自然に定まる「長さ」
    - $\mu([0,1]) = 1$
    - $\mu([a,b]) = b - a$

確率空間

- 測度空間 (Ω, F, P)で以下を満たすもの
  - $P(\Omega) = 1$
- Ω を標本空間という
  - Ωの元を標本という
- F を事象の全体という
  - Fの元を事象という
- P を確率測度という

確率変数 (random variety)

- $X : \Omega \to \mathbb{R}$ で、以下を満たすもの
  - 可測である。つまり、$\forall \omega \in \mathfrak{B}(\mathbb{R}), X^{-1}(\omega) \in F$

Xに由来する確率測度

確率空間(Ω, F, P)と確率変数Xについて、

$P_X : \mathfrak{B}(\mathbb{R}) \to [0, 1],$

$A \to P\left(X^{-1}(A)\right)$

が定まり、$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$は確率空間になっている

d 変数確率変数 (random variety)

$X_1, ..., X_d$:1 変数確率変数

$$(X_1, ..., X_d) : (\Omega, F) \to \left(\mathbb{R}^d, \mathfrak{B}\left(\mathbb{R}^d\right)\right),$$
$$\omega \to (X_1(\omega), ..., X_d(\omega))$$

$(X_1, ..., X_d)$に由来する確率測度

確率空間(Ω, F, P)と d 変数確率変数$(X_1, ..., X_d)$について、

$P_{(X_1, ..., X_d)} : \mathfrak{B}\left(\mathbb{R}^d\right) \to [0, 1],$

$A \to P\left((X_1, ..., X_d)^{-1}(A)\right)$

が定まり、$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$は確率空間になっている

Lebesgue の分解定理

確率空間(Ω, F, P)と確率変数Xについて、

$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$: 確率空間 が定まる。

$\mu : \mathfrak{B}(\mathbb{R}) \to \mathbb{R}$ を Lebegue Measure とする.

$$P_X = P_{X_{\text{abs}}} + P_{X_{\text{sing}}} + P_{X_{\text{dis}}}$$
$$s.t.$$

- $P_{X_{\text{abs}}}$ は μ について、 absolutely conti.
  - すなわち、 $\mu(A) = 0$ ならば $P_{X_{\text{abs}}}(A) = 0 (\forall A \in B(\mathbb{R}))$
- $P_{X_{\text{sing}}}$ は μ について、conti. かつ、 singular.
  - すなわち、 $P_{X_{\text{sing}}}(a) = 0 (\forall a \in \mathbb{R})$ かつ、 $\exists A \in A(\mathbb{R})$ s.t. $\mu(A) = 0$ and $P_{X_{\text{sing}}}(A^c) = 0$
- $P_{X_{\text{dis}}}$ は μ について、 discrete.
  - すなわち、 $\exists \{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ s.t. $P_{X_{\text{dis}}} = \sum_{i=1}^{\infty} a_i \chi_{P_i}(A)$ ただし、$(\chi_{P_i} : B(\mathbb{R}) \to \mathbb{R}, \chi_{P_i}(A) := \left\{ \begin{cases} 0 (p_i \neg\in A) \\ 1 (p_i \in A) \end{cases} \right\})$

と一意に分解できる.

- 拡散モデルの数理を読む上では Pabs だけ考えるで OK

多変数の場合

確率空間(Ω, F, P)と確率変数$(X_1, ..., X_d)$について、

$\left(\mathbb{R}^d, \mathfrak{B}\left(\mathbb{R}^d\right), P_{(X_1, ..., X_d)}\right)$: 確率空間 が定まる。

$\mu : \mathfrak{B}\left(\mathbb{R}^d\right) \to \mathbb{R}$ を Lebegue Measure とする.

以下は、1 変数の場合と同様。

Radon-Nikodym の定理

X: 確率変数

Lebegue の $P_{X_{\text{abs}}}$ について、以下を満たす Integrable な関数 $p_{X_{\text{abs}}} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ が存在する。

$\forall x \in \mathbb{R}$

$$\int_{-\infty}^{x} p_{X_{\text{abs}}}(x') \, \mathrm{d}x' = P_{X_{\text{abs}}}((-\infty, x])$$

この $P_{X_{\text{abs}}}$ を、確率密度関数 (Probabilty Density Function) と呼ぶ。

- 気温が 20 度から 30 度になる確率を求めたい時に、
  - 「気温 x 度」を実数 x にマップする ← 確率変数という
  - 実数直線上で 20-30 の区間で積分する

多変数の場合

確率空間(Ω, F, P)と$(X_1, ..., X_d)$: d 変数確率変数について、

$\left(\mathbb{R}^d, \mathfrak{B}\left(\mathbb{R}^d\right), P_{(X_1, ..., X_d)}\right)$: 確率空間 が定まる。

Lebegue の $P_{(X_1, ..., X_d)_{\text{abs}}}$ について、以下を満たす Integrable な関数 $p^{\text{joint}}_{(X_1, ..., X_d)_{\text{abs}}} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ が存在する。

$\forall x_1, ..., x_1 \in \mathbb{R}$

$$\int_{-\infty}^{x_1} ... \int_{-\infty}^{x_d} p^{\text{joint}}_{(X_1, ..., X_d)_{\text{abs}}}(x') \, \mathrm{d}x' = P_{(X_1, ..., X_d)_{\text{abs}}}((-\infty, x_1] \times ... \times (-\infty, x_d])$$

この $p^{\text{joint}}_{(X_1, ..., X_d)_{\text{abs}}}$ を、結合確率密度関数 (Joint Probabilty Density Function) と呼ぶ。

　フビニの定理により、$\mathbb{R}^d$についての積分は d 回の$\mathbb{R}$についての積分と等しくなる。

$X_1, ..., X_d$ は互いに入れ替え可能である。 特に、$d = 2$の場合、

$$p^{\text{joint}}_{(X_1, X_2)_{\text{abs}}}(x_1, x_2) = p^{\text{joint}}_{(X_2, X_1)_{\text{abs}}}(x_2, x_1)$$

Radon-Nikodym の定理の逆っぽい定理

Theorem (Claim). $d \in \mathbb{Z}_{\geq 1}$と、実数値関数 $p : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ について、

p が、
- pは可測関数
- $\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p(x') \, \mathrm{d}x' = 1$

を満たすとき、

確率空間$(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), P)$がただ一つ存在して、

$\boldsymbol{x} \in \mathbb{R}^d$

$$\int_{-\infty}^{\mathrm{pr}_1(\boldsymbol{x})} ... \int_{-\infty}^{\mathrm{pr}_d(\boldsymbol{x})} p(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' = P((-\infty, \mathrm{pr}_1(\boldsymbol{x})] \times ... \times (-\infty, \mathrm{pr}_d(\boldsymbol{x})])$$

を満たす。

ただし、$1 \leq i \leq d$ に対して、 $X_i := \mathrm{pr}_i : \mathbb{R}^d \to \mathbb{R}$ (射影)

確率空間を構成する。

- $P_{(X_1, ..., X_d)_{\text{abs}}}(X_1 \leq x_1 \wedge ... \wedge X_d \leq x_d) := \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_d} p(x'_1, ..., x'_d) \, \mathrm{d}x'_1 ... \mathrm{d}x'_d$
- $\prod_i (-\infty, x_i]$ は $\mathfrak{B}(\mathbb{R}^d)$ の生成元
- よって、$P_{(X_1, ..., X_d)_{\text{abs}}}$ は $\mathfrak{B}(\mathbb{R}^d)$ 上の確率測度

□

累積分布関数

- 確率密度関数を$-\infty$からxまで積分したもの

確率質量関数 (= **Px3** の時に確率分布って適当に書かれてたらこれ)

$$p_{x_{\text{dis}}} : \mathbb{R} \to [0, 1]$$
$$x \to P_{x_3}(\{x\})$$

d 変数確率密度関数

Radon-Nikodym の定理の多変数版から定まる $p_{(X_1, ..., X_d)_{\text{abs}}} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$

$-$ ↓一つの確率測度(確率分布)Pについて述べている

平均情報量(エントロピー)

(Ω, F, P): 確率空間

X: 確率変数

$($`bb`$\{R\},$ `frak`$\{B\}($`bb`$\{R\}), P_X)$: 上記から定まる確率空間

$$H^P_{\text{entropy}}(X) := H_{\text{entropy}}\left(p_{X_{\text{abs}}}\right) := -\int_{x \in X} \mathrm{d}x p_{X_{\text{abs}}}(x) \log p_{X_{\text{abs}}}(x)$$

エントロピーの値の範囲

$$0 \leq H_{\text{entropy}}\left(p_{X_{\text{abs}}}\right)$$

(誤り)

条件付き確率

(Ω, F, P): 確率空間, $A, B \in F$ について、

$P_{\text{condi}}(\cdot \mid \cdot) : F \times F \to [0, 1]$

$$P_{\mathrm{condi}}(A|B) := \frac{P(A \cap B)}{P(B)}$$

確率空間複数バージョン

確率空間 $\left(\mathtt{bb}\{R\}^{d}, \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d}\right), P_{d}\right), \left(\mathtt{bb}\{R\}^{d-1}, \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d-1}\right), P_{d-1}\right)$

確率変数 $X_{d} : \mathtt{bb}\{R\}^{d} \to \mathtt{bb}\{R\}, X_{d-1} : \mathtt{bb}\{R\}^{d-1} \to \mathtt{bb}\{R\}$ について、

$A_{d} \in \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d}\right), B_{d-1} \in \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d-1}\right)$ について、

$P_{\mathrm{condi}}^{P_{d}, P_{d-1}} : \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d}\right) \times \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d-1}\right) \to [0,1]$

$$P_{\mathrm{condi}}^{P_{d}, P_{d-1}}(A_{d} \mid B_{d-1}) := \frac{P_{d\left(A_{d} \cap \left(B_{d-1} \times \mathtt{bb}\{R\}\right)\right)}}{P_{d-1}(B_{d-1})}$$

条件つき確率密度関数

確率空間 $(\mathtt{bb}\{R\}, \mathtt{frak}\{B\}(\mathtt{bb}\{R\}), P)$

(Y_{1}, Y_{2}): 1 変数確率変数 Y_{1}, Y_{2} によって定まる 2 変数確率変数

$p_{(Y_{1},Y_{2})_{\mathrm{abs}}}^{\mathrm{condi}}(\cdot \mid \cdot) : \mathtt{bb}\{R\} \times \mathtt{bb}\{R\} \to \mathtt{bb}\{R\}_{\geq 0}$

$$p_{(Y_{1},Y_{2})_{\mathrm{abs}}}^{\mathrm{condi}}(y_{1} \mid y_{2}) := \frac{p_{(Y_{1},Y_{2})_{\mathrm{abs}}}^{\mathrm{joint}}(y_{1}, y_{2})}{p_{\underset{\mathrm{abs}}{Y_{2}}}(y_{2})}$$

確率空間複数バージョン

確率測度、$P_{\mathrm{condi}}^{P_{d}, P_{d-1}}$ から Randon-Nikodym の定理により定まる、

$$p_{\mathrm{condi}}^{P_{d} \mid P_{d-1}} : \mathtt{bb}\{R\}^{d} \times \mathtt{bb}\{R\}^{d-1} \to \mathtt{bb}\{R\}_{\geq 0}$$

を、条件付き確率密度関数という

結合確率

確率空間 $\left(\mathtt{bb}\{R\}^{d}, \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d}\right), P_{d}\right), \left(\mathtt{bb}\{R\}^{d-1}, \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d-1}\right), P_{d-1}\right)$

確率変数 $X_{d} : \mathtt{bb}\{R\}^{d} \to \mathtt{bb}\{R\}, X_{d-1} : \mathtt{bb}\{R\}^{d-1} \to \mathtt{bb}\{R\}$ について、

$A_{d} \in \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d}\right), B_{d-1} \in \mathtt{frak}\{B\}\left(\mathtt{bb}\{R\}^{d-1}\right)$ について、

$$P_{\mathrm{joint}}(A_{d}, B_{d-1}) := P_{d(A_{d})} P_{\mathrm{condi}}(A_{d} \mid B_{d-1})$$

結合エントロピー (joint entropy)

確率空間ひとつバージョン

確率空間 $(\mathtt{bb}\{R\}, \mathtt{frak}\{B\}(\mathtt{bb}\{R\}), P)$ 確率変数 X, Y について、

$$H_{\mathrm{joint}}^{P}(X,Y) := H_{\mathrm{joint}}\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}\right) := -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y\, p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)\right)$$

定義から直ちに、

$$H_{\mathrm{joint}}(Y,X) = H_{\mathrm{joint}}(X,Y) \qquad \text{論文}(25)$$

がいえる。

確率空間複数バージョン

条件付きエントロピー (conditional entropy)

確率空間 $(\mathtt{bb}\{R\}, \mathtt{frak}\{B\}(\mathtt{bb}\{R\}), P)$

確率変数 X, Y について、

$$\begin{aligned} H_{\mathrm{condi}}^{P}(Y|X) &:= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y\, p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{joint}}(y,x) \log\left(\frac{p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{joint}}(y,x)}{p_{X_{\mathrm{abs}}}(x)}\right) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y\, p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{joint}}(y,x) \log\left(p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{condi}}(y \mid x)\right) \end{aligned}$$

Theorem (Claim). 以下が成り立つ

$$H_{\mathrm{condi}}^{P}(Y \mid X) = H_{\mathrm{condi}}^{P}(X \mid Y) + H_{\mathrm{entropy}}^{P}(Y) - H_{\mathrm{entropy}}^{P}(X)$$

$$\begin{aligned} (\text{右辺}) &= H_{\mathrm{condi}}^{P}(X \mid Y) + H_{\mathrm{entropy}}^{P}(Y) - H_{\mathrm{entropy}}^{P}(X) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y\, p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{condi}}(x \mid y)\right) - \int_{y \in Y} \mathrm{d}y\, p_{Y_{\mathrm{abs}}}(y) \log p_{Y_{\mathrm{abs}}}(y) + \int_{x \in X} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log p_{X_{\mathrm{abs}}}(x) \qquad (\because \mathrm{def.}) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y\, p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \log\left(\frac{p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)}{p_{Y_{\mathrm{abs}}}(y)}\right) - \int_{y \in Y} \mathrm{d}y\, p_{Y_{\mathrm{abs}}}(y) \log p_{Y_{\mathrm{abs}}}(y) + \int_{x \in X} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log p_{X_{\mathrm{abs}}}(x) \qquad (\because p_{\mathrm{joint}} \text{ の def.}) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)\right) - \log\left(p_{Y_{\mathrm{abs}}}(y)\right)\right)\right) - \int_{y \in Y} \mathrm{d}y\, p_{Y_{\mathrm{abs}}}(y) \log p_{Y_{\mathrm{abs}}}(y) + \int_{x \in X} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log p_{X_{\mathrm{abs}}}(x) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)\right)\right)\right) + \cancel{\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{Y_{\mathrm{abs}}}(y)\right)\right)\right) - \int_{y \in Y} \mathrm{d}y\, p_{Y_{\mathrm{abs}}}(y) \log p_{Y_{\mathrm{abs}}}(y)} + \int_{x \in X} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log p_{X_{\mathrm{abs}}}(x) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)\right)\right)\right) + \int_{x \in X} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log p_{X_{\mathrm{abs}}}(x) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)\right)\right)\right) + \int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(p_{X_{\mathrm{abs}}}(x)\right)\right)\right) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y) \cdot \left(\log\left(\frac{p_{(X,Y)_{\mathrm{abs}}}^{\mathrm{joint}}(x,y)}{p_{X_{\mathrm{abs}}}(x)}\right)\right)\right) \\ &= -\int_{x \in X, y \in Y} \mathrm{d}x\,\mathrm{d}y \left(p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{joint}}(y,x) \cdot \left(\log\left(\frac{p_{(Y,X)_{\mathrm{abs}}}^{\mathrm{joint}}(y,x)}{p_{X_{\mathrm{abs}}}(x)}\right)\right)\right) \\ &= H_{\mathrm{condi}}^{P}(Y \mid X) \qquad (\text{論文}(27)) \end{aligned}$$

□

交差エントロピー (cross entropy)

X: 可測空間(Ω, F)上の確率変数

$(\Omega, F, P), (\Omega, F, Q)$: 確率空間

$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_{X}), (\mathbb{R}, \mathfrak{B}(\mathbb{R}), Q_{X})$: 上記から定まる確率空間

について、

$$H_{\mathrm{cross}}(P_{X}, Q_{X}) := -\sum_{i=1}^{\infty} p_{X_{\mathrm{dis}}}(x_{i}) \log\left(q_{X_{\mathrm{dis}}}(x_{i})\right) - \int_{x \in \mathbb{R}} \mathrm{d}x\, p_{X_{\mathrm{abs}}}(x) \log\left(q_{X_{\mathrm{abs}}}(x)\right)$$

$p_{X_{\mathrm{dis}}} = q_{X_{\mathrm{dis}}} = 0$のとき、

$$H_{\mathrm{cross}}\left(p_{X_{\mathrm{abs}}}, q_{X_{\mathrm{abs}}}\right) := H_{\mathrm{cross}}(P_X, Q_X)$$

ともかく。

KL-ダイバージェンス

X: 可測空間(Ω, F)上の確率変数

$(\Omega, F, P), (\Omega, F, Q)$: 確率空間

$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X), (\mathbb{R}, \mathfrak{B}(\mathbb{R}), Q_X)$: 上記から定まる確率空間

について、

$$D_{KL}(P_X \parallel Q_X) := \sum_{i=1}^{\infty} p_{X_{\mathrm{dis}}}(x_i) \log\left(\frac{p_{X_{\mathrm{dis}}}(x_i)}{q_{X_{\mathrm{dis}}}(x_i)}\right) + \int_{\mathbb{R}} p_{X_{\mathrm{abs}}}(x) \log\left(\frac{p_{X_{\mathrm{abs}}}(x)}{q_{X_{\mathrm{abs}}}(x)}\right) \mathrm{d}x$$

$p_{X_{\mathrm{dis}}} = q_{X_{\mathrm{dis}}} = 0$のとき、

$$D_{KL}\left(p_{X_{\mathrm{abs}}} \parallel q_{X_{\mathrm{abs}}}\right) := D_{KL}(P_X \parallel Q_X)$$

ともかく。

確率変数が複数ある状況の場合

X, Y: 可測空間(Ω, F)上の確率変数

(Ω, F, P): 確率空間

を考えると、

$\mathrm{id}_{\mathbb{R}}$: 可測空間$(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$上の確率変数 (恒等写像)

$(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X), (\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_Y)$: 確率空間

$\left(\mathbb{R}, \mathfrak{B}(\mathbb{R}), (P_X)_{\mathrm{id}_{\mathbb{R}}}\right), \left(\mathbb{R}, \mathfrak{B}(\mathbb{R}), (Q_X)_{\mathrm{id}_{\mathbb{R}}}\right)$: 確率空間

として、KL-ダイバージェンス を考えることができる。

parametric な問題

$$\exists p : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (x, \theta) \mapsto p_{\theta(x)}, \text{具体的に計算可能},$$
$$\exists \theta_0 \in \mathbb{R} \text{ s.t. } p_{X_{\mathrm{abs}}} = p_{\theta_0}$$

$p_{X_{\mathrm{abs}}}$を求める問題を parametric な問題という

(p_{x_3}についてもだいたい同じ)

likelihood function

$(x_i)_{i=1}^n$: n 回の trial が与えられた時、

$$L : \mathbb{R} \to \mathbb{R}, \quad \theta \mapsto \prod_{i=1}^n p(x_i, \theta)$$

log-likelihood function

$(x_i)_{i=1}^n$: n 回の trial が与えられた時、

$$L_{\log} : \mathbb{R} \to \mathbb{R}, \quad \theta \mapsto \sum_{i=1}^n \log(p(x_i, \theta))$$

likelihood estimation

$L, L_{\log}$の最大値、もしくは極値を求める

コイン投げの場合、尤度関数は上に凸なので、極値が求まれば良い

Sanov の定理

参考) https://genkuroki.github.io/documents/mathtodon/2017-04-27%20Kullback-Leibler 情報量とは何か？.pdf

経験分布

r 面サイコロを振るということを考える

$$\mathcal{P} = \{p = (p_1, ..., p_r) \in \mathbb{R}^r \mid p_1, ..., p_r \geq 0, p_1 + ... + p_r = 1\}$$

と定めると、$\mathcal{P}$は、有限集合 $\{1, ..., r\}$上の確率質量関数の集合と同一視できる。

$q = (q_1, ..., q_r) \in \mathcal{P}$ を固定する。これはあるサイコロ q を固定するということ。

$\Omega_q := \{$サイコロ q を投げた時に i が出る $\mid i = 1, ..., r\}$

確率空間: $\left((\Omega_q)^n, F_q^n = 2^{(\Omega_q)^n}, P^n\right)$

$X \in F_q^n$は、

$$X = \left\{\left(\ell_i^1\right)_{i=1,...,n}, ..., \left(\ell_i^j\right)_{i=1,...,n}\right\}$$
$$\left(j = 0, ..., r^n, 1 \leq \ell_i^j \leq r, \text{ただし j=0 の時は X= empty}\right)$$

と書ける

$$r^n : \text{全ての根源事象の数}$$
$$\left(\ell_i^j\right)_{i=1,...,n} : \text{ある事象 X の中で、i 回目に}\ell_i^j\text{が出るような j 番目の根源事象}$$

$P^n : F_q^n \to \mathbb{R}$ を、

$$P^{n\left(\left\{\left(\ell_i^1\right)_{i=1,...,n}, ..., \left(\ell_i^j\right)_{i=1,...,n}\right\}\right)} := \left(q_{\ell_1^1} q_{\ell_2^1} ... q_{\ell_n^1}\right) + ... + \left(q_{\ell_1^j} q_{\ell_2^j} ... q_{\ell_n^j}\right)$$

と定める。

$(k_i)_{i=1,...,n}$について、

$$X_{(k_w)_{w=1,...,r}} := \left\{\left(\ell_i^j\right)_{i=1,...,n} \mid^{\forall} w\left(\left(\ell_i^j = w\text{となる個数}\right) = k_w\right)\right\}$$

なる事象の起こる確率は、$\# X_{((k_w))_{w=1,...,r}} = \frac{n!}{(k_1!...k_r!)}$なので、

$$P^{n\left(X_{(k_w)_{w=1,...,r}}\right)} = \frac{n!}{k_1!...k_r!} q_1^{k_1} ... q_r^{k_r}$$

となる。

ここで、集合$\mathcal{P}_n$を、

$$\mathcal{P}_n = \left\{\left(\frac{k_1}{n}, ..., \frac{k_r}{n}\right) \mid k_i = 0, 1, ..., n, k_1 + ... + k_r = n\right\}$$

と定めると、

$$\mathcal{P}_n \subset \mathcal{P}$$

であり、$\# \mathcal{P}_n \leq (n+1)^r$ である。

このような$\mathcal{P}_n$の元を、

$$X_{(k_i)_{i=1,...,n}} \text{に対しての経験分布}$$

という。

Sanov の定理

$q = (q_1, ..., q_r) \in \mathcal{P}$と、上記の記号の元、確率空間$\left(\mathcal{P}_n, 2^{\mathcal{P}_n}, P_{\mathcal{P}_n,q}\right)$を考える

確率測度は以下のように定める

$$P'_{\mathcal{P}_n,q} : \mathcal{P}_n \text{ の一点集合の集合} \to \mathbb{R}$$
$$\left\{\left(\frac{k_1}{n}, ..., \frac{k_r}{n}\right)\right\} \mapsto \frac{n!}{k_1!...k_r!} q_1^{k_1}...q_r^{k_r}$$
$$P_{\mathcal{P}_n,q} : 2^{\mathcal{P}_n} \to \mathbb{R}$$
$$\{x_1, ..., x_k\} \mapsto \sum_i^k P'_{\mathcal{P}_n,q}(\{x_i\})$$

Theorem (1). $A \subset \mathcal{P}$ が open である時

$$\liminf_{n\to\infty} \frac{1}{n} \log P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n) \geq -\inf_{p\in A} D_{KL}(p \parallel q)$$

A を大きくすると、

((左辺)の$\frac{1}{n} \log P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n)$)は、大きくなるか変わらない

(右辺)は、大きくなるか変わらない

Theorem (2). $A \subset \mathcal{P}$について、

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n) \leq -\inf_{p\in A} D_{KL}(p \parallel q)$$

Theorem (3). $A \subset \mathcal{P}$が、内部が閉包に含まれるならば、

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n) = -\inf_{p\in A} D_{KL}(p \parallel q)$$

- 仮説 A をとる
  - A は絞られてるほうが、実用上「いい仮説」と言える
    - が、Sanov の定理ではそこは興味ない
  - $P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n)$: n 回試行時の経験分布が A に入る確率

  $\to$ n 回振って得られた経験分布を何個も取得して、$P_{\mathcal{P}_n,q}(A \cap \mathcal{P}_n)$を定める
  - 左辺の n をどんどん大きくして、収束する値によって、A の「正解っぽさ」がわかる
    - もし 0 に収束しているのならば、A に真の分布 q が含まれていることがわかる

2. Algorithm

- 順拡散過程と逆拡散過程の定義
- 逆拡散過程の学習方法
- 逆拡散過程の entropy bounds(エントロピー下界？)を導出する

マルコフ拡散カーネル (Markov kernel, 積分変換)

$$T_\pi : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}_{\geq 0}, (\boldsymbol{y}, \boldsymbol{y}', \beta) \mapsto (\boldsymbol{y} \mid \boldsymbol{y}'; \beta)$$

(βは、拡散率)

T_πは、$d \in \mathbb{Z}_{\geq 1}$として、任意の$n \cdot d$変数確率密度関数 $q : \mathbb{R}^{n\cdot d} \to \mathbb{R}$ に対して、

$$q(x_1, ..., x_d) \cdot T_{\pi(x_{d+1} \mid x_d; \beta)} \text{ は } \mathrm{n} \cdot (d+1) \text{変数確率密度関数}$$

$\pi : \mathbb{R}^n \to \mathbb{R}$ は、以下の第二種フレドホルム積分方程式の解

$$\pi(\boldsymbol{y}) = \int \mathrm{d}\boldsymbol{y}' T_{\pi(\boldsymbol{y} \mid \boldsymbol{y}'; \beta)} \pi(\boldsymbol{y}')$$

正規分布を使ったπとT_πがフレドホルム積分方程式の解であること
一次元でチェックする

$\pi(y) := \mathcal{N}(y, 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$

$$T_{\pi(y \mid y'; \beta)} := \mathcal{N}\left(y, y'\sqrt{1-\beta}, \beta\right) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{\left(y - y'\sqrt{1-\beta}\right)^2}{2\beta^2}\right)$$

とおく

$\pi(y) = \int \mathrm{d}y' T_{\pi(y \mid y'; \beta)} \pi(y')$ が、上記に対して成り立つことを確認する

$$
\begin{aligned}
(\text{右辺}) &= \int \mathrm{d}y' \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{\left(y - y'\sqrt{1-\beta}\right)^2}{2\beta^2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{{y'}^2}{2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{\left(y - y'\sqrt{1-\beta}\right)^2}{2\beta^2} - \frac{{y'}^2}{2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{\left(y - y'\sqrt{1-\beta}\right)^2}{2\beta^2} - {y'}^2\frac{\beta^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\left({y'}^2\frac{\beta^2}{2\beta^2} + \frac{\left(-y'\sqrt{1-\beta} + y\right)^2}{2\beta^2}\right)\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{{y'}^2\beta^2 + \left(-y'\sqrt{1-\beta} + y\right)^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{{y'}^2\beta^2 + (1-\beta){y'}^2 - 2\sqrt{1-\beta}yy' + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{{y'}^2(1-\beta+\beta^2) - 2\sqrt{1-\beta}yy' + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left({y'}^2 - \frac{2\sqrt{1-\beta}yy'}{1-\beta+\beta^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left({y'}^2 - \frac{2\sqrt{1-\beta}yy'}{1-\beta+\beta^2} + \frac{(1-\beta)y^2}{\left(1-\beta+\beta^2\right)^2} - \frac{(1-\beta)y^2}{\left(1-\beta+\beta^2\right)^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2 - \frac{(1-\beta)y^2}{\left(1-\beta+\beta^2\right)^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2 - (1-\beta+\beta^2)\frac{(1-\beta)y^2}{\left(1-\beta+\beta^2\right)^2} + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2 - \frac{(1-\beta)y^2}{1-\beta+\beta^2} + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2 - \frac{(1-\beta)-(1-\beta+\beta^2)}{1-\beta+\beta^2}y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2 + \frac{\beta^2}{1-\beta+\beta^2}y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \exp\left(-\frac{\frac{\beta^2}{1-\beta+\beta^2}y^2}{2\beta^2}\right) \int \mathrm{d}y' \exp\left(-\frac{(1-\beta+\beta^2)\left(y' - \frac{\sqrt{1-\beta}y}{1-\beta+\beta^2}\right)^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \exp\left(-\frac{\frac{\beta^2}{1-\beta+\beta^2}y^2}{2\beta^2}\right) \sqrt{\frac{\pi}{\frac{1-\beta+\beta^2}{2\beta^2}}} \\
&= \frac{1}{2\pi\beta} \sqrt{\frac{2\pi\beta^2}{1-\beta+\beta^2}} \exp\left(-\frac{y^2}{2(1-\beta+\beta^2)}\right) \\
&= \frac{1}{\sqrt{2\pi(1-\beta+\beta^2)}} \exp\left(-\frac{y^2}{2(1-\beta+\beta^2)}\right) \\
&= \mathcal{N}\left(y, 0, \sqrt{1-\beta+\beta^2}\right)
\end{aligned}
$$

あわんかったので、$T_{\pi(y \mid y';\beta)} \coloneqq \mathcal{N}(y, my', \beta)$で計算して、噛み合う$m$を見つける。
$\pi(y) \coloneqq \mathcal{N}(y, 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$

$$
T_{\pi(y \mid y';\beta)} \coloneqq \mathcal{N}(y, my', \beta) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(y - my')^2}{2\beta^2}\right)
$$

とおく

$$
\begin{aligned}
(\text{右辺}) &= \int \mathrm{d}y' \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(y - my')^2}{2\beta^2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{{y'}^2}{2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{{y'}^2\beta^2 + m^2{y'}^2 - 2myy' + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2){y'}^2 - 2myy' + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2)\left({y'}^2 - \frac{2myy'}{\beta^2+m^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2)\left({y'}^2 - \frac{2myy'}{\beta^2+m^2} + \frac{m^2y^2}{\left(\beta^2+m^2\right)^2} - \frac{m^2y^2}{\left(\beta^2+m^2\right)^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2)\left(\left(y' - \frac{my}{\beta^2+m^2}\right)^2 - \frac{m^2y^2}{\left(\beta^2+m^2\right)^2}\right) + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2)\left(y' - \frac{my}{\beta^2+m^2}\right)^2 - \frac{m^2y^2}{\beta^2+m^2} + y^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \exp\left(-\frac{\frac{m^2y^2}{\beta^2+m^2} - y^2}{2\beta^2}\right) \int \mathrm{d}y' \exp\left(-\frac{(\beta^2 + m^2)\left(y' - \frac{my}{\beta^2+m^2}\right)^2}{2\beta^2}\right) \\
&= \frac{1}{2\pi\beta} \exp\left(-\frac{\frac{m^2y^2-(\beta^2+m^2)y^2}{\beta^2+m^2}}{2\beta^2}\right) \sqrt{\frac{\pi}{\frac{\beta^2+m^2}{2\beta^2}}} \\
&= \frac{1}{2\pi\beta} \sqrt{\frac{2\pi\beta^2}{\beta^2+m^2}} \exp\left(-\frac{\frac{\beta^2y^2}{\beta^2+m^2}}{2\beta^2}\right) \\
&= \frac{1}{\sqrt{2\pi(\beta^2+m^2)}} \exp\left(-\frac{y^2}{2(\beta^2+m^2)}\right) \\
&= \mathcal{N}\left(y, 0, \sqrt{\beta^2+m^2}\right)(*)
\end{aligned}
$$

$\sqrt{\beta^2+m^2} = 1$より、$m = \sqrt{1-\beta^2}$

よって、

$$T_{\pi(y \mid y';\beta)} := \mathcal{N}\left(y, \sqrt{1-\beta^2}y', \beta\right) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{\left(y-\sqrt{1-\beta^2}y'\right)^2}{2\beta^2}\right)$$

次回(6/1)
- 連続だと思って計算しているのが、本当は離散じゃないと成り立たないみたいな状況を定量化できると面白いのでは?
  - 松尾研で質問してみる @浅香先生 (6/10)
- Appendix A の各定理をまとめる。(28) - (36)
  - (28)を↑の積分チェックして完了
  - T_πになんか条件があるかも (Tpi 自体確率密度関数みたいな) ($p(0) = \pi$ とか)
- (次回:Appendix B. で、上記の計算を進める)
  - $H_{q(X^T)}$ を、定義に合わせる
  - すると、Appendix B. (43) に着地
  - Appendix A.を読む
    - ↑L の下限が 2.4 (14) の形にすることで示せる
- [いつか] 離散で書き直してみる
- 関数解析に踏み込むとしたら
  - 形状最適化問題
- vscode で Tex で書く方法に移植する
- ↑β が小さい時に forward と reverse が等しくなる、はよくわからないので保留
- モデルの実装(on github)も見てみる https://github.com/Sohl-Dickstein/Diffusion-Probabilistic-Models/blob/master/model.py
- 一旦読み終えてみてから、参考文献見てみる?
  - ガウス過程云々

—

def.

$0 \leq i \leq T$に対して $X^i = \mathbb{R}^n$

$1 \leq i \leq T$に対して $\beta_i \in \mathbb{R}$: t-1 と t の間の拡散率

$X^{i\ldots j} := \prod_{t=i}^{j} X^t$

$\boldsymbol{x}^{i\ldots j} := (\boldsymbol{x}^i, \ldots, \boldsymbol{x}^j) \in X^{i\ldots j}$ と書く

2.1. Forward Trajectory

次回(7/20)
- q^0という実数値函数から、$q^{0\ldots T}$まで定める
  - q^0の条件は可測で全域積分すると 1, $q^{0\ldots T}$も満たしている(はず)
  - -> 確率空間$(\mathbb{R}^{T+1}, \mathcal{B}(\mathbb{R}^{T+1}), Q^{0\ldots T})$ が定まる
- $q^{0\ldots T}$を周辺化(T の軸で微分する)すると、$q^{0\ldots T-1}$になることを確かめる
- 結合確率・条件付き確率・エントロピーなど、一通り、確率密度関数を使って定義し直す

$Q^0 : \mathcal{B}(X^0) \to \mathbb{R}$: 確率測度

$q^0_{X_{\text{abs}}} : X^0 \to \mathbb{R}$,

Q^0 から定まる確率密度関数

$q^{0\ldots i}_{X_{\text{abs}}} : X^{0\ldots i} \to \mathbb{R}$を

$$q^{0\ldots i}_{X_{\text{abs}}}(\boldsymbol{x}^{0\ldots i}) := q^0_{X_{\text{abs}}}(\boldsymbol{x}^0) \prod_{j=1}^{i}\left(T_{\pi(\boldsymbol{x}^j \mid \boldsymbol{x}^{j-1};\beta_j)}\right)$$

$q^T_{X_{\text{abs}}} : X^T \to \mathbb{R}$,

$q^T_{X_{\text{abs}}}(\boldsymbol{x}^T) := \int \mathrm{d}\boldsymbol{y}^{0\ldots T-1} q^{0\ldots T}_{X_{\text{abs}}}\left(\boldsymbol{y}^{0\ldots T-1}, \boldsymbol{x}^T\right)$

また、Radon-Nikodym の逆より、$q^{0\ldots i}_{X_{\text{abs}}}, q^T_{X_{\text{abs}}}$ からそれぞれ、

$Q^{0\ldots i} : \mathcal{B}(X^{0\ldots i}) \to \mathbb{R}$: 確率測度

$Q^T : \mathcal{B}(X^T) \to \mathbb{R}$: 確率測度

が定まる。

論文との対応
- $H_{q(\boldsymbol{x}^t)} = H_{\text{entropy}}\left(q^{0\ldots t}_{X_{\text{abs}}}\right)$
- $H_{q(\boldsymbol{x}^{t-1} \mid \boldsymbol{x}^t)} = -\int_{\boldsymbol{x}^{0\ldots t-1} \in \prod_{j=0}^{t-1} X^j, \boldsymbol{x}^{0\ldots t} \in \prod_{j=0}^{t} X^j} \mathrm{d}\boldsymbol{x}^{0\ldots i}\, \mathrm{d}y p^{\text{joint}}_{(Y,X)_{\text{abs}}}(y, \boldsymbol{x}^{0\ldots i}) \log\left(p^{\text{condi}}_{(Y,X)_{\text{abs}}}(y \mid \boldsymbol{x}^{0\ldots i})\right)$
- $q(\boldsymbol{x}^j \mid \boldsymbol{x}^{j-1}) = T_{\pi(\boldsymbol{x}^j \mid \boldsymbol{x}^{j-1};\beta_j)}$

公式 1 (ガウス積分)

$$\int_x \left(\alpha \cdot \exp\left(-\gamma \cdot (x-\beta)^2\right)\right) \mathrm{d}x = \alpha \cdot \sqrt{\frac{\pi}{\gamma}}$$

公式 2

$$\int_x \left(\alpha \cdot \left(-\gamma \cdot (x-\beta)^2\right) \cdot \exp\left(-\gamma \cdot (x-\beta)^2\right)\right) \mathrm{d}x = -\frac{\alpha}{2}\sqrt{\frac{\pi}{\gamma}}$$

導出

$y := \sqrt{\gamma}(x-\beta)$ とおくと、$\mathrm{d}x = \frac{1}{\sqrt{\gamma}}\,\mathrm{d}y$ より

$$\begin{aligned}
&\int_y (\alpha \cdot \exp(-y^2) \cdot (-y^2)) \frac{1}{\sqrt{\gamma}}\,\mathrm{d}y \\
&= -\frac{1}{\sqrt{\gamma}}\alpha \int_y y^2 \cdot \exp(-y^2)\,\mathrm{d}y \\
&= -\frac{1}{\sqrt{\gamma}}\alpha \left(\frac{\sqrt{\pi}}{2}\right) \\
&= -\frac{\alpha}{2}\sqrt{\frac{\pi}{\gamma}}
\end{aligned}$$

Theorem (Claim $H_{\text{entropy}}\left(q^{(0\ldots t)}_{X_{\text{abs}}}\right) \geq H_{\text{entropy}}\left(q^{(0\ldots t-1)}_{X_{\text{abs}}}\right)$ (論文(28))).

$$T_{\pi(x^{(t)} \mid x^{(t-1)};\beta_t)} := \mathcal{N}\left(x^{(t)}, \sqrt{1-\beta_t^2}x^{(t-1)}, \beta_t\right) = \frac{1}{\sqrt{2\pi\beta_t^2}} \exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)$$

とするとき、$\beta_t \geq \sqrt{\frac{1}{2\pi e}}$ ならば、

$$H_{\text{entropy}}\left(q^{(0\ldots t)}_{X_{\text{abs}}}\right) \geq H_{\text{entropy}}\left(q^{(0\ldots t-1)}_{X_{\text{abs}}}\right)$$

が成り立つ。

$$q^{(0\ldots t)}_{X_{\text{abs}}}\left(x^{(0\ldots t)}\right) = q^{(0\ldots t-1)}_{X_{\text{abs}}}\left(x^{(0\ldots t-1)}\right) \cdot T_{\pi(x^{(t)} \mid x^{(t-1)};\beta_t)}$$

であるから、

$$
\text{右辺} = H_{\mathrm{entropy}}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\right) = -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)} q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)
$$

$$
\begin{aligned}
\text{左辺} = H_{\mathrm{entropy}}\left(q^{(0\dots t)}_{X_{\mathrm{abs}}}\right) &= -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t)}\right)\cdot\log\left(q^{(0\dots t)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t)}\right)\right)\right) \\
&= -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right) \\
&= -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)+\log\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right)\right) \\
&= -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\right)\right) \\
&\quad -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right)\right) \quad \dots \quad (*)
\end{aligned}
$$

− 項を分ける

$$
\begin{aligned}
(*)\ \text{第 1 項} &= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\cdot\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\cdot\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\cdot\underbrace{\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\underbrace{\frac{1}{\sqrt{2\pi\beta_t^2}}}_{\alpha}\exp\left(-\underbrace{\frac{1}{2\beta_t^2}}_{\gamma}\cdot\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2\right)\right)}_{1}\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\log\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)\right) \\
&= \text{右辺}
\end{aligned}
$$

$$
\begin{aligned}
(*)\ \text{第 2 項} &= -\int_{x^{(0\dots t)}\in\prod_{j=0}^{t} X^{(j)}} \mathrm{d}x^{(0\dots t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\cdot\left(\log\left(T_{\pi(x^{(t)}\mid x^{(t-1)};\beta_t)}\right)\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\cdot\left(\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\right)\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\left(\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\cdot\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\right)\right)+\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\cdot\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\right)\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\left(\frac{1}{\cancel{\sqrt{2\pi\beta_t^2}}}\cdot\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\right)\cdot\underbrace{\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\right)}_{\cancel{\sqrt{2\pi\beta_t^2}}\quad(\because\text{ガウス積分})}+\frac{1}{\cancel{\sqrt{2\pi\beta_t^2}}}\cdot\underbrace{\int_{x^{(t)}\in X^{(t)}}\mathrm{d}x^{(t)}\left(\exp\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\cdot\left(-\frac{\left(x^{(t)}-\sqrt{1-\beta_t^2}x^{(t-1)}\right)^2}{2\beta_t^2}\right)\right)}_{-\frac{1}{2}\cdot\cancel{\sqrt{2\pi\beta_t^2}}\quad(\because\text{公式 2})}\right)\right) \\
&= -\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\cdot\left(\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\right)-\frac{1}{2}\right)\right) \\
&= -\left(\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\right)-\frac{1}{2}\right)\cdot\underbrace{\int_{x^{(0\dots t-1)}\in\prod_{j=0}^{t-1} X^{(j)}} \mathrm{d}x^{(0\dots t-1)}\left(q^{(0\dots t-1)}_{X_{\mathrm{abs}}}\left(x^{(0\dots t-1)}\right)\right)}_{1} \\
&= -\left(\log\left(\frac{1}{\sqrt{2\pi\beta_t^2}}\right)-\frac{1}{2}\right) \\
&= \frac{1}{2}+\frac{1}{2}\cdot\log(2\pi\beta_t^2) \\
&= \frac{1}{2}\cdot(1+\log(2\pi\beta_t^2))
\end{aligned}
$$

よって、

$$
\begin{aligned}
\text{左辺} - \text{右辺} &= \left(\text{右辺} - \frac{1}{2}(1+\log(2\pi\beta_t^2))\right) - \text{右辺} \\
&= \frac{1}{2}(1+\log(2\pi\beta_t^2))
\end{aligned}
$$

ゆえに、$\beta_t \geq \sqrt{\frac{1}{2\pi e}}$ の時、不等式が成り立つ

□

めも(次回): ここから typst 変換続ける