

次回(8/10)

- 定理番号(counter 関数)を調べる
- 順次定義を直す
 - 一旦連続のみで

Claim 0.1 (テスト): これはテストです

1. 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

1.1. 関連しそうな項目

- 拡散モデル
 - 確率過程
- シミュレーテッドアニーリング
- 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

2. 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

3. 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

4. 論文まとめ

4.1. 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

5. 一般定理

5.1. ガウス積分

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

6. 用語

6.1. 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

6.2. 可測写像

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

6.3. 測度空間

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

6.4. 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
 - \mathbb{R} 上の距離から定まる位相を含む最小の完全加法族
 - μ : ルベーク測度
 - 距離から自然に定まる「長さ」
 - $\mu([0, 1]) = 1$
 - $\mu([a, b]) = b - a$

6.5. 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

6.6. 確率変数 (random variety)

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

6.7. 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、

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

$$A \rightarrow P(X^{-1}(A))$$

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

6.8. d 変数確率変数 (random variety)

$X_1, ..., X_d$: 1 変数確率変数

$$\begin{aligned} (X_1, ..., X_d): (\Omega, F) &\rightarrow (\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d)), \\ \omega &\rightarrow (X_1(\omega), ..., X_d(\omega)) \end{aligned}$$

6.9. $(X_1, ..., X_d)$ 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、

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

$$A \rightarrow P\big((X_1, ..., X_d)^{-1}(A)\big)$$

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

Theorem 6.9.1 (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。

$\mu : \mathfrak{B}(\mathbb{R}) \rightarrow \mathbb{R}$ 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、 $\exists \{a_i\}_{i=1}^\infty \subset \mathbb{R}, \{p_i\}_{i=1}^\infty \subset \mathbb{R}$ s.t. $P_{X_{\text{dis}}}(A) = \sum_{i=1}^\infty a_i \chi_{p_i}(A)$ ただし、 $\left(\chi_{p_i} : B(\mathbb{R}) \rightarrow \mathbb{R}, \chi_{p_i}(A) := \begin{cases} 0(p_i \notin A) \\ 1(p_i \in A) \end{cases}\right)$

と一意に分解できる。

- 拡散モデルの数理を読む上では P_{abs} だけ考えるで OK

6.9.1. 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、

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

$\mu : \mathfrak{B}(\mathbb{R}^d) \rightarrow \mathbb{R}$ を Lebegue Measure とする。

以下は、1 変数の場合と同様。

6.10. Radon-Nikodym の定理

Theorem 6.10.1 (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 $p_{X_{\text{abs}}} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ 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 × 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

6.10.1. 多変数の場合

Theorem 6.10.1.1 (Radon-Nikodym の定理: 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、

$(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), P_{(X_1, ..., X_d)})$: 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 $p_{(X_1, ..., X_d)_{\text{abs}}}^{\text{joint}} : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ が存在する。

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

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

この $p_{(X_1, ..., X_d)_{\text{abs}}}^{\text{joint}}$ 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、

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

Theorem 6.10.1.2 (Radon-Nikodym の定理の逆っぽい定理): $d \in \mathbb{Z}_{\geq 1}$ と、実数値関数 $p : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ について、

p が、

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

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、

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

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

を満たす。

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

Proof: 確率空間を構成する。

- $P_{(X_1, ..., X_d)_{\text{abs}}}(X_1 \leq x_1 \wedge \cdots \wedge X_d \leq x_d) := \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} p(x'_1, ..., x'_d) \, \mathrm{d}x'_1 \cdots \mathrm{d}x'_d$
- $\prod_i (-\infty, x_i]$ 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

□

6.11. 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

6.12. 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)

$$p_{x_{\text{dis}}} : \mathbb{R} \rightarrow [0, 1]$$

$$x \rightarrow P_{x_3}(\{x\})$$

6.13. d 変数確率密度関数

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

—

↓一つの確率測度(確率分布) P について述べている

Definition 6.13.1:

$$\begin{array}{ccccc} p_{\text{abs}}^d & : & \mathbb{R}^d & \rightarrow & \mathbb{R} \text{ 可測関数} \\ & & \cup & & \cup \\ & & (x_1, \dots, x_d) & \mapsto & y \end{array}$$

で、

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x') \, \mathrm{d}x' = 1$$

7. 平均情報量(エントロピー)

$$H_{\text{entropy}}(p_{\text{abs}}) := - \int_X \left(p_{\text{abs}}(x) \log(p_{\text{abs}}(x)) \right) \, \mathrm{d}\mu(x)$$

7.1.1.1. エントロピーの値の範囲

$$0 \leq H_{\text{entropy}}(p_{\text{abs}}) \text{ (誤り)}$$

8. 条件付き確率 確率密度関数

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

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

8.1. 確率空間複数バージョン

$$\text{確率空間 } (\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), P_d), (\mathbb{R}^{d-1}, \mathfrak{B}(\mathbb{R}^{d-1}), P_{d-1})$$

$$\text{確率変数 } X_d : \mathbb{R}^d \rightarrow \mathbb{R}, X_{d-1} : \mathbb{R}^{d-1} \rightarrow \mathbb{R} \text{ について、}$$

$$A_d \in \mathfrak{B}(\mathbb{R}^d), B_{d-1} \in \mathfrak{B}(\mathbb{R}^{d-1}) \text{ について、}$$

$$P_{\text{condi}}^{P_d, P_{d-1}} : \mathfrak{B}(\mathbb{R}^d) \times \mathfrak{B}(\mathbb{R}^{d-1}) \rightarrow [0, 1]$$

$$P_{\text{condi}}^{P_d, P_{d-1}}(A_d \mid B_{d-1}) := \frac{P_{d(A_d \cap (B_{d-1} \times \mathbb{R}))}}{P_{d-1}(B_{d-1})}$$

9. 条件つき確率密度関数

$$\text{確率空間 } (\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$$

$$(Y_1, Y_2) \text{: 1 変数確率変数 } Y_1, Y_2 \text{ によって定まる 2 変数確率変数}$$

$$p_{(Y_1, Y_2)_{\text{abs}}}^{\text{condi}}(\cdot \mid \cdot) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$$

$$p_{(Y_1, Y_2)_{\text{abs}}}^{\text{condi}}(y_1 \mid y_2) := \frac{p_{(Y_1, Y_2)_{\text{abs}}}^{\text{joint}}(y_1, y_2)}{p_{Y_2}(y_2)_{\text{abs}}}$$

9.1. 確率空間複数バージョン

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

$$p_{\text{condi}}^{P_d \mid P_{d-1}} : \mathbb{R}^d \times \mathbb{R}^{d-1} \rightarrow \mathbb{R}_{\geq 0}$$

を、条件付き確率密度関数という

10. 結合確率

$$\text{確率空間 } (\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), P_d), (\mathbb{R}^{d-1}, \mathfrak{B}(\mathbb{R}^{d-1}), P_{d-1})$$

$$\text{確率変数 } X_d : \mathbb{R}^d \rightarrow \mathbb{R}, X_{d-1} : \mathbb{R}^{d-1} \rightarrow \mathbb{R} \text{ について、}$$

$$A_d \in \mathfrak{B}(\mathbb{R}^d), B_{d-1} \in \mathfrak{B}(\mathbb{R}^{d-1}) \text{ について、}$$

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

11. 結合エントロピー (joint entropy)

11.1. 確率空間ひとつバージョン

$$\text{確率空間 } (\mathbb{R}, \mathfrak{B}(\mathbb{R}), P) \text{ 確率変数 } X, Y \text{ について、}$$

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

定義から直ちに、

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

がいえる。

11.2. 確率空間複数バージョン

12. 条件付きエントロピー (conditional entropy)

$$\text{確率空間 } (\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$$

$$\text{確率変数 } X, Y \text{ について、}$$

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

Theorem 12.1 (Claim): 以下が成り立つ

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

Proof:

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

□

12.1. 交差エントロピー (cross entropy)

X : 可測空間 (Ω, \mathcal{F}) 上の確率変数

$(\Omega, \mathcal{F}, P), (\Omega, \mathcal{F}, Q)$: 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、

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

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

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

ともかく、

12.2. KL-ダイバージェンス

X : 可測空間 (Ω, \mathcal{F}) 上の確率変数

$(\Omega, \mathcal{F}, P), (\Omega, \mathcal{F}, Q)$: 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、

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

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

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

ともかく、

12.2.1. 確率変数が複数ある状況の場合

X, Y : 可測空間 (Ω, \mathcal{F}) 上の確率変数

(Ω, \mathcal{F}, P) : 確率空間

を考えると、

$\text{id}_{\mathbb{R}}$: 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

$\left(\mathbb{R}, \mathfrak{B}(\mathbb{R}), (P_X)_{\text{id}_{\mathbb{R}}}\right), \left(\mathbb{R}, \mathfrak{B}(\mathbb{R}), (Q_X)_{\text{id}_{\mathbb{R}}}\right)$: 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。

12.3. parametric な問題

$\exists p : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (x, \theta) \mapsto p_{\theta(x)},$ 具体的に計算可能,

$$\exists \theta_0 \in \mathbb{R} \text{ s.t. } p_{X_{\text{abs}}} = p_{\theta_0}$$

$p_{X_{\text{abs}}}$ 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)

12.4. likelihood function

$(x_i)_{i=1}^n$: n 回の trial が与えられた時、

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

12.5. log-likelihood function

$(x_i)_{i=1}^n$: n 回の trial が与えられた時、

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

12.6. 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

13. 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

13.1. 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。

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

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

$X \in F_q^n$ は、

$$X=\left\{\left(\ell_i^1\right)_{i=1,...,n},...,\left(\ell_i^j\right)_{i=1,...,n}\right\}$$

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

と書ける

$$r^n: \text{全ての根源事象の数}$$

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

$$P^n:F_q^n\rightarrow \mathbb{R} \text{ を、}$$

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

と定める。

$$\left(k_i\right)_{i=1,...,n} \text{ について、}$$

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

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

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

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。

13.2. Sanov の定理

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

確率測度は以下のように定める

$$P'_{\mathcal{P}_n,q}:\mathcal{P}_n \text{ の一点集合の集合} \rightarrow \mathbb{R}$$

$$\left\{\left(\frac{k_1}{n},..., \frac{k_r}{n}\right)\right\}\mapsto \frac{n!}{k_1!...k_r!}q_1^{k_1}...q_r^{k_r}$$

$$P_{\mathcal{P}_n,q}:2^{\mathcal{P}_n}\rightarrow \mathbb{R}$$

$$\{x_1,...,x_k\}\mapsto \sum_i^k P'_{\mathcal{P}_n,q}(\{x_i\})$$

Theorem 13.2.1 (1): $A\subset \mathcal{P}$ が open である時

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

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、大きくなるかわからない

(右辺)は、大きくなるかわからない

Theorem 13.2.2 (2): $A\subset \mathcal{P}$ について、

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

Theorem 13.2.3 (3): $A\subset \mathcal{P}$ が、内部が閉包に含まれるならば、

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

- 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

→ 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

14. 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

14.1.1.1. マルコフ拡散カーネル (Markov kernel, 積分変換)

$$(\beta \text{は、拡散率})$$

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 $q:\mathbb{R}^{n\cdot d}\rightarrow \mathbb{R}$ に対して、

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

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

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

14.1.1.1.1. 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

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

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

とおく

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

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

あわなかったので、 $T_{\pi(y \mid y'; \beta)} := \mathcal{N}(y, my', \beta)$ で計算して、噛み合う m を見つける。

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

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

とおく

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

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

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 間の拡散率

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

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

14.2. 2.1. Forward Trajectory

次回(7/20)

- q^0 という実数値函数から、 $q^{0...T}$ まで定める
 - q^0 の条件は可測で全域積分すると 1, $q^{0...T}$ も満たしている(はず)
 - > 確率空間($\mathbb{R}^{T+1}, \mathcal{B}(\mathbb{R}^{T+1}), Q^{0...T}$) が定まる
- $q^{0...T}$ を周辺化(Tの軸で微分する)すると、 $q^{0...T-1}$ 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

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

$q_{X_{\text{abs}}}^0: X^0 \rightarrow \mathbb{R}$,

Q^0 から定まる確率密度関数

$q_{X_{\text{abs}}}^{0...i}: X^{0...i} \rightarrow \mathbb{R}$ を

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

$q_{X_{\text{abs}}}^T: X^T \rightarrow \mathbb{R}$,

$q_{X_{\text{abs}}}^T(\boldsymbol{x}^T) := \int \mathrm{d}\boldsymbol{y}^{0...T-1} q_{X_{\text{abs}}}^{0...T}(\boldsymbol{y}^{0...T-1}, \boldsymbol{x}^T)$

また、Radon-Nikodym の逆より、 $q_{X_{\text{abs}}}^{0...i}, q_{X_{\text{abs}}}^T$ からそれぞれ、

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

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

が定まる。

14.2.1.1.1. 論文との対応

- $H_{q(\boldsymbol{x}^t)} = H_{\text{entropy}}\left(q_{X_{\text{abs}}}^{0...t}\right)$
- $H_{q(\boldsymbol{x}^{t-1} \mid \boldsymbol{x}^t)} = - \int_{\boldsymbol{x}^{0...t-1} \in \prod_{j=0}^{t-1} X^j, \boldsymbol{x}^{0...t} \in \prod_{j=0}^t X^j} \mathrm{d}\boldsymbol{x}^{0...i} \mathrm{d}y p_{(Y, X)_{\text{abs}}}^{\text{joint}}(y, \boldsymbol{x}^{0...i}) \log\left(p_{(Y, X)_{\text{abs}}}^{\text{condi}}(y \mid \boldsymbol{x}^{0...i})\right)$

- $q(\boldsymbol{x}^j \mid \boldsymbol{x}^{j-1}) = T_{\pi(\boldsymbol{x}^j \mid \boldsymbol{x}^{j-1}; \beta_j)}$

14.2.1.2. 公式 1 (ガウス積分)

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

14.2.1.3. 公式 2

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

14.2.1.3.1.1. 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 14.2.1.3.1.1.1 (Claim $H_{\text{entropy}}\Big(q_{X_{\text{abs}}}^{(0\dots t)}\Big) \geq H_{\text{entropy}}\Big(q_{X_{\text{abs}}}^{(0\dots t-1)}\Big)$ (**論文**(28))):

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

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

$$H_{\text{entropy}}\Big(q_{X_{\text{abs}}}^{(0\dots t)}\Big) \geq H_{\text{entropy}}\Big(q_{X_{\text{abs}}}^{(0\dots t-1)}\Big)$$

が成り立つ。

Proof:

$$q_{X_{\text{abs}}}^{(0\dots t)}\Big(x^{(0\dots t)}\Big) = q_{X_{\text{abs}}}^{(0\dots t-1)}\Big(x^{(0\dots t-1)}\Big) \cdot T_{\pi(x^{(t)} \mid x^{(t-1)}; \beta_t)}$$

であるから、

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

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

－ 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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot T_{\pi(x^{(t)} \mid x^{(t-1)}; \beta_t)} \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(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)} \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(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{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \right) \cdot \underbrace{\int_{x^{(t)} \in X^{(t)}} \mathrm{d}x^{(t)} \left(\frac{1}{\underbrace{\sqrt{2\pi\beta_t^2}}_{\alpha}} \exp \left(-\frac{1}{\underbrace{2\beta_t^2}_{\gamma}} \cdot \left(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)} \right)^2 \right) \right)}_1 \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \log \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \right) \right) \\
&= \text{右辺} \\
(*) \text{ 第2項} &= - \int_{x^{(0\dots t)} \in \prod_{j=0}^t X^{(j)}} \mathrm{d}x^{(0\dots t)} \left(q_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \int_{x^{(t)} \in X^{(t)}} \mathrm{d}x^{(t)} \left(\frac{1}{\sqrt{2\pi\beta_t^2}} \exp \left(-\frac{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^2}{2\beta_t^2} \right) \cdot \left(\log \left(\frac{1}{\sqrt{2\pi\beta_t^2}} \exp \left(-\frac{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \cdot \left(\int_{x^{(t)} \in X^{(t)}} \mathrm{d}x^{(t)} \left(\frac{1}{\sqrt{2\pi\beta_t^2}} \exp \left(-\frac{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^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{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^2}{2\beta_t^2} \right) \cdot \left(-\frac{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^2}{2\beta_t^2} \right) \right)}_{\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{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^2}{2\beta_t^2} \right) \cdot \left(-\frac{(x^{(t)} - \sqrt{1-\beta_t^2}x^{(t-1)})^2}{2\beta_t^2} \right) \right)}_{-\frac{1}{2} \cdot \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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_{X_{\text{abs}}}^{(0\dots t-1)}(x^{(0\dots t-1)}) \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 □

Theorem 14.2.1.3.1.1.2 (付録 A (30)):

$$H_{q(\boldsymbol{X}^{t-1} \mid \boldsymbol{X}^t)} \leq H_{q(\boldsymbol{X}^t \mid \boldsymbol{X}^{t-1})}$$

$$H_{\text{condi}} = - \int (x \in X, y \in Y) \, \mathrm{d}x \, \mathrm{d}y p_{(Y, X)_{\text{abs}}}^{\text{joint}}(y, x) \log \left(p_{(Y, X)_{\text{abs}}}^{\text{condi}}(y \mid x) \right)$$

(次回 6/29)

- ↑を $q^{0\dots i}$ を使って書く
- この時、 $q^{0\dots i}$ は joint としてしまえることがわかったので、 $p^{\wedge}(\text{joint})$ のところに $q^{\wedge}(0\dots i)$ をそのまま入れる
- 型は、 $p_{((X^{t-1}, \boldsymbol{X}^t))_{\text{abs}}}^{\text{condi}}(\cdot \mid \cdot) : \mathbb{R}^t \times \mathbb{R}^{t-1} \rightarrow \mathbb{R}_{\geq 0}$ みたいになるはず。(ちゃんとやる)

$\beta_t \geq \sqrt{\frac{1}{2\pi e}}$ のとき、

(次回 6/22)

14.2.2. メモ

14.2.2.1.1. 大きな方針

- とりあえずこの本の理論は最後まで追ってみる
 - 有限と捉えてちゃんと書いてみた時に、自然と $\sqrt{1-\beta}$ が出てきてくれないか
 - 代数的確率空間的に書いてみるとどうなるか
- 各ステップのエントロピー差が有界であることを示している
 - (30)で上界がわかる
 - ゴールは (36)

14.3. 2.2. Reverse Trajectory

$$p_{X_{\text{abs}}}^T : X^T \rightarrow \mathbb{R}, p_{X_{\text{abs}}}^T(\boldsymbol{x}^T) := \pi(\boldsymbol{x}^T)$$

$$p_{X_{\text{abs}}}^{i\dots T} : \prod_{t=i}^T X^t \rightarrow \mathbb{R}$$

$$p_{X_{\text{abs}}}^{i\dots T}(\boldsymbol{x}^{i\dots T}) := p_{X_{\text{abs}}}^T(\boldsymbol{x}^T) \prod_{t=1}^T T_{\pi(\boldsymbol{x}^{t-1} \mid \boldsymbol{x}^t; \beta_t)}$$

14.3.1.1.1. 論文との対応

- $p(x^{j-1} \mid x^j) = T_{\pi(\boldsymbol{x}^{j-1} \mid \boldsymbol{x}^j; \beta_j)}$

14.4. 2.3. Model Probability

> これは、統計物理学における準静的過程の場合に相当する

- [非平衡科学](https://sosuke110.com/noneq-physics.pdf)

・[機械学習のための確率過程](https://www.amazon.co.jp/%E6%A9%9F%E6%A2%B0%E5%AD%A6%E7%BF%92%E3%81%AE%E3%81%9F%E3%82%81%E3%81%AE%E7%A2%BA%E7%8E%87%E9%81%8E%E7%A8%8B%E5%85%A5%E9%96%80-%E2%80%95%E7%A2%BA%E7%8E%87%E5%BE%AE%E5%88%86%E6%96%B9%E7%A8%8B%E5%BC%8F%E3%81%8B%E3%82%89%E3%83%99%E3%82%A4%E3%82%BA%E3%83%A2%E3%83%87%E3%83%AB%E3%81%BE%E3%81%A7%E2%80%95-%E5%86%85%E5%B1%B1%E7%A5%90%E4%BB%8B-ebook/dp/B0CK176SH5/ref=tmm_kin_swatch_0?_encoding=UTF8&qid=&sr=)

$$\begin{aligned} p_{X_{\text{abs}}}^0 : X^0 &\rightarrow \mathbb{R}, p_{X_{\text{abs}}}^0(\boldsymbol{x}^0) := \int \mathrm{d}\boldsymbol{y}^{1\dots T} p_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T}) \\ &= \int \mathrm{d}\boldsymbol{y}^{1\dots T} p_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T}) \frac{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})} \\ &= \int \mathrm{d}\boldsymbol{y}^{1\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T}) \frac{p_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})} \\ &= \int \mathrm{d}\boldsymbol{y}^{1\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T}) \cdot \frac{p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)}{q_{X_{\text{abs}}}^0(\boldsymbol{x}^0)} \cdot \frac{T_{\pi(\boldsymbol{x}^0 \mid \boldsymbol{y}^1; \beta_1)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{x}^0; \beta_1)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \end{aligned}$$

この積分は、複数のサンプル $q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})$ の平均を取ることで、素早く評価できる。

β_t が無限小のとき、 $\frac{T_{\pi(\boldsymbol{x}^0 \mid \boldsymbol{y}^1; \beta_t)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{x}^0; \beta_t)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) = 1$ となる。

- つまり、 $1 \leq t \leq T$ について、 $T_{\pi(\boldsymbol{x}^{t-1} \mid \boldsymbol{y}^t; \beta_t)} = T_{\pi(\boldsymbol{x}^t \mid \boldsymbol{y}^{t-1}; \beta_t)} \leftarrow ?$
- このとき、上記の積分を評価するのに要求されるのは、単一のサンプル $q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{x}^0, \boldsymbol{y}^{1\dots T})$ のみである

> $q_{x_i}^0$: 与えられた真のデータの分布 > $p_{x_i}^0$: Reverse trajectory を用いて計算された、 $q_{x_i}^0$ の近似

以下で、 $p_{x_i}^0, q_{x_i}^0$ の Cross Entropy を最小化する

14.5. 2.4. Training

$H(p_{X_{\text{abs}}}^0, q_{X_{\text{abs}}}^0)$: Cross Entropy

$$\begin{aligned} H(p_{X_{\text{abs}}}^0, q_{X_{\text{abs}}}^0) &= - \int_{X^0} \mathrm{d}\boldsymbol{y}^0 q_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \cdot \log p_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \\ &= - \int_{X^0} \mathrm{d}\boldsymbol{y}^0 q_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \cdot \log \left[\int \mathrm{d}\boldsymbol{y}^{1\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \frac{p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)}{q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)} \cdot \frac{T_{\pi(\boldsymbol{y}^0 \mid \boldsymbol{y}^1; \beta_1)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{y}^0; \beta_1)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \right] \\ &= - \int_{X^0} \mathrm{d}\boldsymbol{y}^0 q_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \cdot \log \left[\int \mathrm{d}\boldsymbol{y}^{1\dots T} \frac{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)} \cdot p_{X_{\text{abs}}}^T(\boldsymbol{y}^T) \cdot \frac{T_{\pi(\boldsymbol{y}^0 \mid \boldsymbol{y}^1; \beta_1)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{y}^0; \beta_1)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \right] \\ &\geq - \int_{X^0} \mathrm{d}\boldsymbol{y}^0 q_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \cdot \left(\int_{X^{1\dots T}} \mathrm{d}\boldsymbol{y}^{1\dots T} \cdot \frac{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)} \cdot \log \left[p_{X_{\text{abs}}}^T(\boldsymbol{y}^T) \cdot \frac{T_{\pi(\boldsymbol{y}^0 \mid \boldsymbol{y}^1; \beta_1)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{y}^0; \beta_1)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \right] \right) \end{aligned}$$

(\because [Jensen の不等式](https://ja.wikipedia.org/wiki/%E3%82%A4%E3%82%A7%E3%83%B3%E3%82%BB%E3%83%B3%E3%81%AE%E4%B8%8D%E7%AD%89%E5%BC%8F), $\int_{X^{1\dots T}} \frac{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)} = 1$ (\because 周辺分布の定義))

$$= - \int_{X^0} \mathrm{d}\boldsymbol{y}^0 \left(\int_{X^{1\dots T}} \mathrm{d}\boldsymbol{y}^{1\dots T} q_{X_{\text{abs}}}^0(\boldsymbol{y}^0) \cdot \frac{q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T})}{q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)} \cdot \log \left[p_{X_{\text{abs}}}^T(\boldsymbol{y}^T) \cdot \frac{T_{\pi(\boldsymbol{y}^0 \mid \boldsymbol{y}^1; \beta_1)}}{T_{\pi(\boldsymbol{y}^1 \mid \boldsymbol{y}^0; \beta_1)}} \cdot \left(\prod_{t=2}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \right] \right)$$

(\because $q_{X_{\text{abs}}}^0(\boldsymbol{y}^0)$ は $\int_{X^{1\dots T}}$ において定数)

$$= - \int_{X^{0\dots T}} \mathrm{d}\boldsymbol{y}^{0\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^{0\dots T}) \cdot \log \left[p_{X_{\text{abs}}}^T(\boldsymbol{y}^T) \cdot \left(\prod_{t=1}^T \frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right) \right]$$

(\because 約分 & 重積分はまとめられる & 細かい略記) (Appendix B. (38) と一致)

$$= - \int_{X^{0\dots T}} \mathrm{d}\boldsymbol{y}^{0\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \left(\log [p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)] + \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) \right)$$

(1) (\because log を和に分解)

$$= - \int_{X^{0\dots T}} \mathrm{d}\boldsymbol{y}^{0\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \left(\sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) + \log [p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)] \right)$$

(1.5) (\because 括弧内の和の順序入れ替え)

$$= - \int_{X^{0\dots T}} \left(q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) \right) \mathrm{d}\boldsymbol{y}^{0\dots T} - \int_{X^{0\dots T}} \left(q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \log [p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)] \right) \mathrm{d}\boldsymbol{y}^{0\dots T}$$

(2) (\because 積分を和で分解)

$$= - \int_{X^{0\dots T}} \left(q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) \right) \mathrm{d}\boldsymbol{y}^{0\dots T} - \int_{X^T} \int_{X^{0\dots T-1}} \left(q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \log [p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)] \right) \mathrm{d}\boldsymbol{y}^{0\dots T}$$

(2.5) (\because 第二項の積分を被積分変数で分解)

$$= - \int_{X^{0\dots T}} \left(q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) \right) \mathrm{d}\boldsymbol{y}^{0\dots T} - \int_{X^T} \left(q_{X_{\text{abs}}}^T(\boldsymbol{y}^T) \cdot \log [p_{X_{\text{abs}}}^T(\boldsymbol{y}^T)] \right) \mathrm{d}\boldsymbol{y}^T$$

(3) (\because 第二項の内側の積分実行)

$$= - \int_{X^{0\dots T}} \mathrm{d}\boldsymbol{y}^{0\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) - H_{\text{cross}}(q_{X_{\text{abs}}}^T, p_{X_{\text{abs}}}^T)$$

(4)

$$= - \int_{X^{0\dots T}} \mathrm{d}\boldsymbol{y}^{0\dots T} q_{X_{\text{abs}}}^{0\dots T}(\boldsymbol{y}^0, \boldsymbol{y}^{1\dots T}) \cdot \sum_{t=1}^T \left(\log \left[\frac{T_{\pi(\boldsymbol{y}^{t-1} \mid \boldsymbol{y}^t; \beta_t)}}{T_{\pi(\boldsymbol{y}^t \mid \boldsymbol{y}^{t-1}; \beta_t)}} \right] \right) - H_{\text{cross}}(q_{X_{\text{abs}}}^T, p_{X_{\text{abs}}}^T)$$

(5) \leftarrow (論文 14 と対応させたい)

14.6. 2.5. Multiplying Distributions, and Computing Posteriors

14.7. 2.6. Entropy of Reverse Process

15. 3. Experiments

15.1. 3.1. Toy Problem

15.2. 3.2. Images

16. 4. Conclusion

17. メモ

- 小林さん MTG めも

17.1. 小林さん MTG メモ

18. ざっくりした物理的イメージ

- 生成モデルって？
 - 何か \rightarrow 似たやつを作る
 - \rightarrow = 確率分布
- 確率分布 = 分配関数
 - 分配関数がわかれば確率分布がわかる
 - エネルギーが定義できる状況ならなんでも分配関数が使える

- 情報では？
 - 尤度 = エネルギー
 - 学習したデータの尤度を高める = その点のポテンシャルを最小化する
- エネルギー(尤度)を最適させるのに適切な関数は何？
 - 機械学習で言えば損失関数
 -
- 量子コンピュータの種類
 - 量子アニーリング
 - 量子ゲート