

Denoising Diffusion Probabilistic Models の 変分推論

正田 備也

masada@rikkyo.ac.jp

Contents

周辺尤度

ELBO

変分事後分布

観測データのモデリング

Appendix

周辺尤度 (marginal likelihood)

- ▶ 結合分布が次のように書ける確率モデルを考える。

$$p_{\theta}(\mathbf{x}_{0:T}) = p_{\theta}(\mathbf{x}_T) \prod_{t=1}^T p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t) \quad (1)$$

- ▶ \mathbf{x}_0 だけは、その値が観測されている確率変数。
- ▶ 他の $\{\mathbf{x}_t : t = 1, \dots, T\}$ は、latent な確率変数。
- ▶ そして、 \mathbf{x}_{t-1} の分布は \mathbf{x}_t だけに条件付けられている。
 - ▶ 階層がいくつもあるベイズモデルだと思っていい。
- ▶ \mathbf{x}_0 の対数周辺尤度は、次のように書ける。

$$\log p_{\theta}(\mathbf{x}_0) = \log \int p_{\theta}(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T} \quad (2)$$

Contents

周辺尤度

ELBO

変分事後分布

観測データのモデリング

Appendix

ELBO

- ▶ Jensen の不等式を使って、対数周辺尤度の下界を得る。

$$\begin{aligned}\log p_{\theta}(\mathbf{x}_0) &= \log \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \frac{p_{\theta}(\mathbf{x}_{0:T})}{q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &\geq \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{0:T})}{q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{0:T})}{q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0)} d\mathbf{x}_{1:T} \equiv L_{\text{VLB}} \quad (3)\end{aligned}$$

- ▶ $q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0)$ が変分事後分布。
 - ▶ 変分事後分布は、VAE 同様、観測データ \mathbf{x}_0 に条件付けられている。
 - ▶ 変分事後分布のパラメータ ψ は、観測データ $\mathcal{X} \equiv \{\mathbf{x}_0^{(1)}, \dots, \mathbf{x}_0^{(N)}\}$ を使って、VAE と同様に、amortized な仕方で training する。

変分事後分布におけるマルコフ性の仮定

- ▶ 条件付き分布の chain rule より

$$q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) = q_\psi(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \dots, \mathbf{x}_1, \mathbf{x}_0) \quad (4)$$

- ▶ ここで、 $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \dots, \mathbf{x}_1, \mathbf{x}_0) = q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$ を、全ての $t = 2, \dots, T$ について仮定し、変分事後分布を単純化する。
- ▶ すると、 $q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0)$ は、次のように分解できる。

$$q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) = q_\psi(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) \quad (5)$$

ELBO の書き直し

- ▶ このとき、式 (3) の変分下界 L_{VLB} は、以下のように書き直せる。

$$\begin{aligned} L_{\text{VLB}} &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{0:T})}{q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_T) \prod_{t=1}^T p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q_{\psi}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_{\theta}(\mathbf{x}_T) d\mathbf{x}_{1:T} + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \sum_{t=2}^T \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_{\psi}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0)} d\mathbf{x}_{1:T} \end{aligned} \tag{6}$$

- ▶ 以下、この ELBO の式をさらに変形していく。
- ▶ が、その前に、 $T = 1$ の場合（つまり、階層が 1 層だけの場合）に ELBO がどうか、見ておく。

$T = 1$ の場合の ELBO

- ▶ $T = 1$ のとき、式 (3) の変分下界 L_{VLB} は、以下のように書き直せる。

$$\begin{aligned} L_{\text{VLB}} &= \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{0:1})}{q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0)} d\mathbf{x}_1 \\ &= \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log p_{\theta}(\mathbf{x}_1) d\mathbf{x}_1 + \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_0 | \mathbf{x}_1)}{q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0)} d\mathbf{x}_1 \\ &= \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log p_{\theta}(\mathbf{x}_0 | \mathbf{x}_1) d\mathbf{x}_1 + \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_1)}{q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0)} d\mathbf{x}_1 \\ &= \int q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \log p_{\theta}(\mathbf{x}_0 | \mathbf{x}_1) d\mathbf{x}_1 - D_{\text{KL}}(q_{\psi}(\mathbf{x}_1 | \mathbf{x}_0) \parallel p_{\theta}(\mathbf{x}_1)) \end{aligned} \quad (7)$$

- ▶ 途中、 $p_{\theta}(\mathbf{x}_1)$ と $p_{\theta}(\mathbf{x}_0 | \mathbf{x}_1)$ を入れ替えた。
- ▶ 式 (7) は、VAE の ELBO。逆に言うと・・・
- ▶ 式 (6) の、 $\int q_{\psi}(\mathbf{x}_{1:T} | \mathbf{x}_0) \sum_{t=2}^T \log \frac{p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_{\psi}(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0)} d\mathbf{x}_{1:T}$ という項は、VAE の ELBO には無かった。この項の扱いが、やや厄介。
- ▶ だから、式 (6) を、以下、さらに変形していく。

ELBO の式の変形 (1/5)

- ▶ 式 (6) に現れる $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$ についてベイズ則を使うと、次を得る。

$$q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = \frac{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q_\psi(\mathbf{x}_t|\mathbf{x}_0)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_0)} \quad (8)$$

- ▶ 式 (8) を使うと、 $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$ と $q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ の間を、行ったり来たりできる。
- ▶ つまり、 \mathbf{x}_{t-1} と \mathbf{x}_t の並び順を、いつでもひっくり返せる。
- ▶ この式 (8) にもとづいて、式 (6) に現れる $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$ を $\frac{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q_\psi(\mathbf{x}_t|\mathbf{x}_0)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_0)}$ で置き換えると、変分下界 L_{VLB} は以下のように書き換えられる。

$$\begin{aligned} L_{\text{VLB}} = & \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_\theta(\mathbf{x}_T) d\mathbf{x}_{1:T} + \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \sum_{t=2}^T \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\ & + \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \sum_{t=2}^T \log \frac{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q_\psi(\mathbf{x}_t|\mathbf{x}_0)} d\mathbf{x}_{1:T} + \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}{q_\psi(\mathbf{x}_1|\mathbf{x}_0)} d\mathbf{x}_{1:T} \end{aligned} \quad (9)$$

ELBO の式の変形 (2/5)

$$\begin{aligned}
 L_{\text{VLB}} &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_{\theta}(\mathbf{x}_T) d\mathbf{x}_{1:T} + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \sum_{t=2}^T \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_{\psi}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\
 &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0) \cancel{q_{\psi}(\mathbf{x}_2|\mathbf{x}_0)} \cdots \cancel{q_{\psi}(\mathbf{x}_{T-1}|\mathbf{x}_0)}}{\cancel{q_{\psi}(\mathbf{x}_2|\mathbf{x}_0)} \cancel{q_{\psi}(\mathbf{x}_2|\mathbf{x}_0)} \cdots q_{\psi}(\mathbf{x}_T|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\
 &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\
 &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_{\theta}(\mathbf{x}_T) d\mathbf{x}_{1:T} + \sum_{t=2}^T \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_{\psi}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\
 &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{\cancel{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0)}}{q_{\psi}(\mathbf{x}_T|\mathbf{x}_0)} d\mathbf{x}_{1:T} + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{\cancel{q_{\psi}(\mathbf{x}_1|\mathbf{x}_0)}} d\mathbf{x}_{1:T} \\
 &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_T)}{q_{\psi}(\mathbf{x}_T|\mathbf{x}_0)} d\mathbf{x}_{1:T} + \sum_{t=2}^T \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_{\psi}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\
 &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_{\theta}(\mathbf{x}_0|\mathbf{x}_1) d\mathbf{x}_{1:T} \equiv L_T + \sum_{t=2}^T L_{t-1} + L_0
 \end{aligned}$$

ELBO の式の変形 (3/5)

- ▶ ここで、式 (8) より

$$\begin{aligned} q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_{t-2}, \mathbf{x}_0) &= \frac{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q_\psi(\mathbf{x}_t|\mathbf{x}_0)}{\cancel{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_0)}} \cdot \frac{q_\psi(\mathbf{x}_{t-2}|\mathbf{x}_{t-1}, \mathbf{x}_0)\cancel{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_0)}}{q_\psi(\mathbf{x}_{t-2}|\mathbf{x}_0)} \\ &= \frac{q_\psi(\mathbf{x}_t|\mathbf{x}_0)}{q_\psi(\mathbf{x}_{t-2}|\mathbf{x}_0)} q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q_\psi(\mathbf{x}_{t-2}|\mathbf{x}_{t-1}, \mathbf{x}_0) \end{aligned} \quad (11)$$

- ▶ 同様に考えて、

$$\prod_{t=2}^T q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = \frac{q_\psi(\mathbf{x}_T|\mathbf{x}_0)}{q_\psi(\mathbf{x}_1|\mathbf{x}_0)} \prod_{t=2}^T q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \quad (12)$$

- ▶ 両辺に $q_\psi(\mathbf{x}_1|\mathbf{x}_0)$ を掛けて

$$q_\psi(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = q_\psi(\mathbf{x}_T|\mathbf{x}_0) \prod_{t=2}^T q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \quad (13)$$

- ▶ 式 (5) より、これは $q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0)$ に等しい。
- ▶ つまり、 $\mathbf{x}_1, \dots, \mathbf{x}_T$ の並び順をひっくり返しても、 $q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0)$ に等しい。

ELBO の式の変形 (4/5)

- ▶ 式 (13) を使って式 (10) の L_{t-1} を下のように書き換え、 $q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ を分けて扱う。

$$\begin{aligned} L_{t-1} &\equiv \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &= \int \left(q_\psi(\mathbf{x}_T|\mathbf{x}_0) \prod_{t' \neq t} q_\psi(\mathbf{x}_{t'-1}|\mathbf{x}_{t'}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \quad (14) \end{aligned}$$

- ▶ ここで、 $q_\psi(\mathbf{x}_T|\mathbf{x}_0)q_\psi(\mathbf{x}_{T-1}|\mathbf{x}_T, \mathbf{x}_0) = q_\psi(\mathbf{x}_{T-1}, \mathbf{x}_T|\mathbf{x}_0)$ より、

$$L_{t-1} = \int \left(q_\psi(\mathbf{x}_{T-1}, \mathbf{x}_T|\mathbf{x}_0) \prod_{t' \neq t \wedge t' < T} q_\psi(\mathbf{x}_{t'-1}|\mathbf{x}_{t'}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T} \quad (15)$$

- ▶ \mathbf{x}_T を積分消去すると

$$L_{t-1} = \int \left(q_\psi(\mathbf{x}_{T-1}|\mathbf{x}_0) \prod_{t' \neq t \wedge t' < T} q_\psi(\mathbf{x}_{t'-1}|\mathbf{x}_{t'}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:T-1} \quad (16)$$

- ▶ 同様の手順で $\mathbf{x}_{T-1}, \dots, \mathbf{x}_{t+1}$ を積分消去する。(次ページ)

ELBO の式の変形 (5/5)

$$L_{t-1} = \int \left(q_\psi(\mathbf{x}_t | \mathbf{x}_0) \prod_{t'=2}^{t-1} q_\psi(\mathbf{x}_{t'-1} | \mathbf{x}_{t'}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:t} \quad (17)$$

▶ ここで、再び式 (13) を使って、今度は $\mathbf{x}_1, \dots, \mathbf{x}_{t-2}$ を積分消去する。

$$L_{t-1} = \int \left(\frac{q_\psi(\mathbf{x}_t | \mathbf{x}_0)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_0)} q_\psi(\mathbf{x}_1 | \mathbf{x}_0) \prod_{t'=2}^{t-1} q_\psi(\mathbf{x}_{t'} | \mathbf{x}_{t'-1}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{1:t}$$

$$= \int \left(\frac{q_\psi(\mathbf{x}_t | \mathbf{x}_0)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_0)} q_\psi(\mathbf{x}_2 | \mathbf{x}_0) \prod_{t'=3}^{t-1} q_\psi(\mathbf{x}_{t'} | \mathbf{x}_{t'-1}, \mathbf{x}_0) \right) q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{2:t}$$

...

$$= \int \left(\frac{q_\psi(\mathbf{x}_t | \mathbf{x}_0)}{\cancel{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_0)}} \cancel{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_0)} \right) q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{t-1:t}$$

$$= \int q_\psi(\mathbf{x}_t | \mathbf{x}_0) \left(\int q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)}{q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)} d\mathbf{x}_{t-1} \right) d\mathbf{x}_t$$

$$\equiv -\mathbb{E}_{q_\psi(\mathbf{x}_t | \mathbf{x}_0)} [D_{\text{KL}}(q_\psi(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t))]$$

(18)

変形後の ELBO

- 以上の結果をまとめると

$$\begin{aligned} L_{\text{VLB}} &= \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_{\theta}(\mathbf{x}_T)}{q_{\psi}(\mathbf{x}_T|\mathbf{x}_0)} d\mathbf{x}_{1:T} \\ &\quad - \sum_{t=2}^T \mathbb{E}_{q_{\psi}(\mathbf{x}_t|\mathbf{x}_0)} \left[D_{\text{KL}}(q_{\psi}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)) \right] \\ &\quad + \int q_{\psi}(\mathbf{x}_{1:T}|\mathbf{x}_0) \log p_{\theta}(\mathbf{x}_0|\mathbf{x}_1) d\mathbf{x}_{1:T} \\ &\equiv L_T + \sum_{t=2}^T L_{t-1} + L_0 \end{aligned} \tag{19}$$

Contents

周辺尤度

ELBO

変分事後分布

観測データのモデリング

Appendix

変分事後分布の一つの設定方法

- ▶ 変分事後分布 $q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) = q_\psi(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$ は、 $\psi \equiv \{\alpha_1, \dots, \alpha_T\}$ をパラメータとする以下のような多変量正規分布の積として構成されていると仮定する。

$$q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1 - \alpha_t)\mathbf{I}) \quad \text{for } t = 2, \dots, T \quad (20)$$

- ▶ すると、 $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0) = q_\psi(\mathbf{x}_t|\mathbf{x}_{t-1})$ となり、 \mathbf{x}_1 以外は \mathbf{x}_0 に依存しなくなる。
- ▶ ところで、Appendix の式 (37) より、 $q_\psi(\mathbf{x}_t|\mathbf{x}_{t-2})$ は、下のように書き換えられる。

$$q_\psi(\mathbf{x}_t|\mathbf{x}_{t-2}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2}, (1 - \alpha_t\alpha_{t-1})\mathbf{I}) \quad (21)$$

- ▶ 同じ議論を繰り返すと、 $q_\psi(\mathbf{x}_t|\mathbf{x}_0)$ は、下のように書き換えられる。

$$q_\psi(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad \text{where } \bar{\alpha}_t = \prod_{s=1}^t \alpha_s \quad (22)$$

- ▶ この $q_\psi(\mathbf{x}_t|\mathbf{x}_0)$ からは、簡単にサンプルを得られる (式 (27) を参照)。
- ▶ よって、式 (18) の期待値は、モンテカルロ近似できる。

変分事後分布の導出(1/3)

- ▶ ψ を自由パラメータとみなし、これ以降、 ψ を notations から脱落させることにする。
- ▶ 式 (20) と式 (22) より、式 (18) の $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ は、以下のように書き換えられる。

$$\begin{aligned} q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) &= \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \quad (\text{cf. Eq. (8)}) \\ &\propto \exp \left(-\frac{1}{2} \left(\frac{(\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_{t-1})^2}{1 - \alpha_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}} \mathbf{x}_0)^2}{1 - \bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0)^2}{1 - \bar{\alpha}_t} \right) \right) \\ &\propto \exp \left(-\frac{1}{2} \left(\left(\frac{\alpha_t}{1 - \alpha_t} + \frac{1}{1 - \bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1}^2 - 2 \left(\frac{\sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}} \mathbf{x}_0 \right) \mathbf{x}_{t-1} \right) \right) \end{aligned} \quad (23)$$

- ▶ つまり、 $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ は正規分布であることが分かる。
- ▶ そこで、その平均を $\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0)$ 、分散を $\tilde{\beta}_t$ と書くことにする。つまり

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \equiv \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t) \quad (24)$$

と設定する。

変分事後分布の導出(2/3)

- ▶ すると、以下を得る。

$$\tilde{\beta}_t = 1 / \left(\frac{\alpha_t}{1 - \alpha_t} + \frac{1}{1 - \bar{\alpha}_{t-1}} \right) = \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{\alpha_t - \alpha_t \bar{\alpha}_{t-1} + 1 - \alpha_t} = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t) \quad (25)$$

$$\begin{aligned} \tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) &= \left(\frac{\sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}} \mathbf{x}_0 \right) / \left(\frac{\alpha_t}{1 - \alpha_t} + \frac{1}{1 - \bar{\alpha}_{t-1}} \right) \\ &= \left(\frac{\sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}} \mathbf{x}_0 \right) \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t) \\ &= \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \bar{\alpha}_t} \mathbf{x}_0 \end{aligned} \quad (26)$$

- ▶ 式 (22) にもとづくと、 \mathbf{x}_t は、次のように reparameterize できる。

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon} \quad \text{for } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (27)$$

- ▶ よって

$$\mathbf{x}_0 = \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \boldsymbol{\epsilon} \quad \text{for } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (28)$$

変分事後分布の導出(3/3)

- ▶ 式 (28) を式 (26) に代入することで、 $\tilde{\boldsymbol{\mu}}$ を以下のように書き換えることができる。

$$\begin{aligned}\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \boldsymbol{\epsilon}) &= \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \bar{\alpha}_t} \left(\frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \boldsymbol{\epsilon} \right) \\&= \left(\frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} + \frac{1 - \alpha_t}{(1 - \bar{\alpha}_t)\sqrt{\alpha_t}} \right) \mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}\sqrt{\alpha_t}} \boldsymbol{\epsilon} \\&= \frac{1}{\sqrt{\alpha_t}} \left(\left(\frac{\alpha_t(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} + \frac{1 - \alpha_t}{1 - \bar{\alpha}_t} \right) \mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon} \right) \\&= \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon} \right)\end{aligned}\tag{29}$$

- ▶ 変分事後分布 $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)$ の平均 $\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0)$ は、 \mathbf{x}_t と $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ とを使えば、計算できることがわかる。

Contents

周辺尤度

ELBO

変分事後分布

観測データのモデリング

Appendix

観測データのモデリング

- ▶ ここで初めて、データ・モデルの詳細を以下のように指定する。

$$p_{\theta}(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I}) \quad (30)$$

$$p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t)) \quad (31)$$

- ▶ ただし、 $\boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I}$ と仮定する ([1] を参照)。
- ▶ 式 (24) と式 (31) より、式 (18) にある KL 情報量は、一つの正規分布から別の正規分布への KL 情報量であると分かる。
- ▶ したがって、式 (18) の L_{t-1} は、以下のように書き直せる。¹

$$L_{t-1} = -\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \left[\frac{1}{2\sigma_t^2} \|\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \epsilon) - \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t)\|^2 \right] + const. \quad (32)$$

¹https://scoste.fr/posts/dkl_gaussian/

ELBO revisited (1/3)

- ▶ 式 (29) を使うと、 L_{t-1} は、さらに、以下のように書き直せる。

$$L_{t-1} = -\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \left[\frac{1}{2\sigma_t^2} \left\| \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon} \right) - \boldsymbol{\mu}_\theta(\mathbf{x}_t, t) \right\|^2 \right] + const. \quad (33)$$

- ▶ ここで、 $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ を、関数 $\boldsymbol{\epsilon}_\theta$ を使って次のように parameterize する ([1] を参照)。

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) \quad (34)$$

- ▶ この式 (34) と、式 (27) の \mathbf{x}_t を使うと、 L_{t-1} は、以下のように書き換えられる。

$$L_{t-1} = -\mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\frac{(1-\alpha_t)^2}{2\sigma_t^2(1-\bar{\alpha}_t)} \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2 \right] + const. \quad (35)$$

- ▶ 関数 $\boldsymbol{\epsilon}_\theta$ は、 $\boldsymbol{\epsilon}$ を予測する役割を果たすと言える。
- ▶ これにともなって、外側の期待値も、標準正規分布に関する期待値に書き換えてある。
- ▶ [1] の Algorithm 1 の 4 行目は、上式の期待値をモンテカルロ近似するサンプリング。

ELBO revisited (2/3)

- ▶ 次に、式 (19) の L_T を考える。

$$L_T \equiv \int q_\psi(\mathbf{x}_{1:T}|\mathbf{x}_0) \log \frac{p_\theta(\mathbf{x}_T)}{q_\psi(\mathbf{x}_T|\mathbf{x}_0)} d\mathbf{x}_{1:T} \quad (36)$$

- ▶ ノイズ分布 $p_\theta(\mathbf{x}_T)$ は trainable なパラメータを持たない。
- ▶ 近似事後分布 $q_\psi(\mathbf{x}_T|\mathbf{x}_0)$ についても、 ψ は自由パラメータ。
- ▶ したがって、 L_T は定数と見なせる。
 - ▶ なお、自由パラメータ $\psi \equiv \{\alpha_1, \dots, \alpha_T\}$ の値の設定は、いわゆる noise scheduling に相当する [2]。

ELBO revisited (3/3)

- ▶ 最後に、式 (19) の L_0 を考える。
- ▶ L_0 をどのように最大化するかは、 $p_\theta(\mathbf{x}_0|\mathbf{x}_1)$ をどのように指定するかに依存する。
- ▶ そして、この $p_\theta(\mathbf{x}_0|\mathbf{x}_1)$ は、直接的に観測データをモデル化する分布である。
 - ▶ 例えば、[1] の Section 3.3 や [4] を参照されたい。

注意 ここでは、denoising diffusion probabilistic models の変分推論だけを議論している。このモデルがどこから来たのかについては、議論していない。(この点については [3] を参照。)

Contents

周辺尤度

ELBO

変分事後分布

観測データのモデリング

Appendix

Appendix

$$\begin{aligned} & \int \exp \left(-\frac{(x-ay)^2}{2s^2} - \frac{(y-bz)^2}{2t^2} \right) dy = \int \exp \left(-\frac{t^2(x-ay)^2 + s^2(y-bz)^2}{2s^2t^2} \right) dy \\ &= \int \exp \left(-\frac{(s^2 + t^2a^2)y^2 - 2(s^2bz + t^2ax)y + t^2x^2 + s^2b^2z^2}{2s^2t^2} \right) dy \\ &= \exp \left(-\frac{t^2x^2 + s^2b^2z^2}{2s^2t^2} \right) \int \exp \left(-\frac{s^2 + t^2a^2}{2s^2t^2} \left(y^2 - \frac{2(s^2bz + t^2ax)}{s^2 + t^2a^2} y \right) \right) dy \\ &= \exp \left(-\frac{t^2x^2 + s^2b^2z^2}{2s^2t^2} + \frac{(s^2bz + t^2ax)^2}{2s^2t^2(s^2 + t^2a^2)} \right) \int \exp \left(-\frac{s^2 + t^2a^2}{2s^2t^2} \left(y - \frac{s^2bz + t^2ax}{s^2 + t^2a^2} \right)^2 \right) dy \\ &\propto \exp \left(-\frac{s^2t^2x^2 + s^4b^2z^2 + t^4a^2x^2 + s^2t^2a^2b^2z^2 - t^4a^2x^2 - 2s^2t^2abzx - s^4b^2z^2}{2s^2t^2(s^2 + t^2a^2)} \right) \\ &= \exp \left(-\frac{x^2 - 2abzx + a^2b^2z^2}{2(s^2 + t^2a^2)} \right) = \exp \left(-\frac{(x-abz)^2}{2(s^2 + t^2a^2)} \right) \end{aligned} \tag{37}$$

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020. <https://arxiv.org/abs/2006.11239>.
- [2] Nikolas Kirschstein. Noise schedules for diffusion probabilistic models. 2022. <https://www.kirschstein.io/portfolio/2022-oxford-tdl/>.
- [3] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. *CoRR*, abs/1503.03585, 2015. <https://arxiv.org/abs/1503.03585>.
- [4] Wei Yi. Understanding the denoising diffusion probabilistic model (DDPMs), the Socratic way. 2023. <https://towardsdatascience.com/understanding-the-denoising-diffusion-probabilistic-model-the-socratic-way-445c1bdc5756>.