

[59] をもとに (6.18) について再度考える。

(6.14) を考慮すると以下ようになる。

$$\begin{aligned} p(Y_U, Z_A, Z_U | Y_A, X_A, X_U) &= \frac{p(Y_A, Y_U, Z_A, Z_U, X_A, X_U)}{p(Y_A, X_A, X_U)} = \frac{p(X_A | Y_A, Z_A) p(Y_A) p(Z_A) p(X_U | Y_U, Z_U) p(Y_U) p(Z_U)}{p(Y_A, X_A, X_U)} \\ &= p(X_A | Y_A, Z_A) p(Z_A) p(X_U | Y_U, Z_U) p(Y_U) p(Z_U) \frac{p(Y_A)}{p(Y_A, X_A, X_U)} = p(X_A | Y_A, Z_A) p(Z_A) p(X_U | Y_U, Z_U) p(Y_U) p(Z_U) \exp\{c\} \end{aligned} \quad (1)$$

3 つめの等号のあとの分数の部分がデータとしてわかっているので定数。

また、 q を平均場近似を用いて、以下のようにする。

$$\begin{aligned} q &\equiv q(Y_U, Z_A, Z_U, \psi_2; X_A, X_U, Y_A, \psi_{1a}, \psi_{1b}, \psi_3) \approx q(Z_A; X_A, X_U, Y_A, \psi_{1a}, \psi_{1b}, \psi_3) q(Z_U; X_A, X_U, Y_A, \psi_{1a}, \psi_{1b}, \psi_3) \\ &\quad q(Y_A; X_A, X_U, Y_A, \psi_{1a}, \psi_{1b}, \psi_2, \psi_3) q(Y_U, \psi_2; X_A, X_U, Y_A, \psi_{1a}, \psi_{1b}, \psi_3) \\ &\approx q(Z_A; X_A, Y_A, \psi_{1a}) q(Z_U; X_U, \psi_{1b}) q(Y_U; X_U, \psi_2) q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3) \end{aligned} \quad (2)$$

本と異なり、 ψ_2 がパラメータ ψ_3 によって決まる確率変数としている。

これらは (6.6), (6.15)-(6.17) のように定数や、得られたデータを用いて、表される。また、対称ディリクレ分布を仮定して、 $q(\psi_2 | \psi_3 = (\alpha)) = \text{Dir}(\psi_2 | (\alpha)) = C_D((\alpha)) \prod_k \psi_{2k}^{\alpha-1}$

(3.10) の KL ダイバージェンスの定義を考慮すると、

$$\begin{aligned} D_{KL}(q(Y_U, Z_A, Z_U, \psi_2; X_A, Y_A, X_U, \psi_{1a}, \psi_{1b}, \psi_3) || p(Y_U, Z_A, Z_U | X_A, X_U, Y_A)) &= -\mathbb{E}_{q(Z_A; X_A, Y_A, \psi_{1a})} [\ln p(X_A | Y_A, Z_A)] \\ &\quad -\mathbb{E}_{q(Z_A; X_A, Y_A, \psi_{1a})} [\ln p(Z_A)] + \mathbb{E}_{q(Z_A; X_A, Y_A, \psi_{1a})} [\ln q(Z_A; X_A, Y_A, \psi_{1a})] \\ &\quad -\mathbb{E}_{q(Y_U; X_U, \psi_2) q(\psi_2; \psi_3) q(Z_U; X_U, \psi_{1b})} [\ln p(X_U | Y_U, Z_U)] - \mathbb{E}_{q(Y_U; X_U, \psi_2) q(\psi_2; \psi_3)} [\ln p(Y_U)] - \mathbb{E}_{q(Z_U; X_U, \psi_{1b})} [\ln p(Z_U)] \\ &\quad + \mathbb{E}_{q(Y_U; X_U, \psi_2) q(\psi_2; \psi_3)} [\ln q(Y_U; X_U, \psi_2)] + \mathbb{E}_{q(Z_U; X_U, \psi_{1b})} [\ln q(Z_U; X_U, \psi_{1b})] \\ &\quad + \mathbb{E}_{q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3)} [\ln q(Y_A; X_A, \psi_2)] + \mathbb{E}_{q(\psi_2; \psi_3)} [\ln q(\psi_2; \psi_3)] + c \end{aligned} \quad (3)$$

これで (6.18) に相当する式が求まった。

(6.18) との違いは、期待値を取る確率分布 q の若干の違いもあるが、本質的には以下のみ異なる。

$$\begin{aligned} &\mathbb{E}_{q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3)} [\ln q(Y_A; X_A, \psi_2)] + \mathbb{E}_{q(\psi_2; \psi_3)} [\ln q(\psi_2; \psi_3)] = \\ &\mathbb{E}_{q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3)} [\ln q(Y_A; X_A, \psi_2)] + \mathbb{E}_{q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3)} [\ln q(\psi_2; \psi_3)] \\ &= \mathbb{E}_{q(Y_A; X_A, \psi_2) q(\psi_2; \psi_3)} [\ln q(Y_A; X_A, \psi_2) + \ln q(\psi_2; \psi_3)] \end{aligned} \quad (4)$$

$q(\psi_2; \psi_3)$ に対称ディリクレ分布を仮定しているため、 ψ_2 の各要素は、ほぼ同一だと想定され、 $q(Y_A; X_A, \psi_2)$ は以下のように仮定する。

$$q(Y_A; X_A, \psi_2) \equiv q(Y_A; \psi_2) = \text{Cat}(Y_A | \psi_2) = \prod_a \prod_k \psi_{2k}^{y_{ak}} \quad (5)$$

そうした場合、

$$\ln q(Y_A; X_A, \psi_2) + \ln q(\psi_2; \psi_3) = \left(\sum_k \left(\sum_a y_{ak} + \alpha - 1 \right) \ln \psi_{2k} \right) + c \quad (6)$$

ψ_2 の要素がほぼ同一だと $\sum_a y_{ak}$ が、 k によらず、ほぼ同一になり、 $(\sum_a y_{ak} + \alpha - 1)$ が定数とみなせる。よって、

$$\ln q(Y_A; X_A, \psi_2) + \ln q(\psi_2; \psi_3) = \left(\sum_k \left(\sum_a y_{ak} + \alpha - 1 \right) \ln \psi_{2k} \right) + c \approx \beta \ln \psi_2 \propto \ln q(Y_A; X_A, \psi_2) \quad (7)$$

よって、(本の方の符号は違う気がするが、)(6.22) のような式に妥当性があるように見える。