第3章　最尤推定値の計算と EM アルゴリズム

ここでは最尤推定値を数値計算で求める方法を 3 つ紹介する．節 3.1 では Newton-Raphson 法による最尤推定値の計算法を説明する．次節では, Newton-Raphson 法を修正した Fisher のスコア法を説明する．節 3.3 では, 欠損値データに対して最尤推定値を求めるときに有効な EM アルゴリズムについて説明する．

3.1　Newton-Raphson 法

sec:3-1

いま $g : \mathbb{R} \to \mathbb{R}$ を 2 回微分可能な関数とし, 方程式 $g(x) = 0$ をみたす解 $x = c$ をみつけたいとする．そのために c に近い x ~~に対して,~~ のまわりで, 関数 g を Taylor 展開をする．

$$0 = g(c) \approx g(x) + \dot{g}(x)(c - x)$$

を得る．ただし $\dot{g}(x) = dg/dx$ である．また「≈」は「近い」と漠然と理解することにする．

$\dot{g}(x) \neq 0$ のとき $g(x) + \dot{g}(x)(x - c) \approx 0$ を c について解けば

$$c \approx x - \frac{g(x)}{\dot{g}(x)}$$

を得る．初期値 x_0 を取り点列 $\{x_n\}_{n=1}^{\infty}$ を逐次的に

$$x_{n+1} = x_n - \frac{g(x_n)}{\dot{g}(x_n)}, \qquad n = 0, 1, \ldots \tag{3.1}$$

eq:newton1

で定義する．そして $|g(x_n)/\dot{g}(x_n)|$ が十分小さくなるまで操作を繰り返すとする．

区間 $I = [a, b]$ 上で $\ddot{g}(x) > 0$ かつ $g(a)g(b) < 0$ のとき, $g(x_0) > 0$ となる $x_0 \in (a, b)$ をひとつ見つければ, (3.1) で得られる点列 $\{x_n\}_{n=1}^{\infty}$ は I 上における零点 c に収束することが知られている[1]．Newton-Raphson 法

[1] 杉浦「解析入門 I (東京大学出版会)」p.105 を参照.

を用いて尤度方程式の解として定義される最尤推定値を求めることができる.

いま $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ は同時 p.d.f. または p.m.f. $p(\boldsymbol{x}|\,\theta)$ を持つとする. ここで, p はほとんどいたるところの $\boldsymbol{x}$ で θ に関して 2 回連続微分可能とする. また $\theta \in \Theta \subset \mathbb{R}$ で $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である. $\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの尤度関数を $\mathsf{lik}_n(\theta|\,\boldsymbol{x}) = \mathsf{p}(\boldsymbol{x}|\,\theta)$ と書くことにする. したがって, 最尤推定値 $\widehat{\theta}_n$ は関数

$$\Theta \ni \theta \mapsto \mathsf{lik}_n(\theta|\,\boldsymbol{x})$$

の θ に関する最大値である. したがって, 最尤推定値 $\widehat{\theta}_n$ は尤度方程式

$$S_n(\theta|\,\boldsymbol{x}) = \frac{d}{d\theta}\log \mathsf{lik}_n(\theta|\,\boldsymbol{x}) = 0 \tag{3.2}$$

eq:lequation

の解となる.

方程式 (3.2) の解である最尤推定値を Newton-Raphson 法で求めよう. そのために, $m\ (m \in \mathbb{N})$ 回操作を行ったのちの θ の推定値を $\widehat{\theta}_n^{(m)}(\boldsymbol{x})$ とする. (3.1) から $H_n(\widehat{\theta}_n^{(m)}|\,\boldsymbol{x}) \neq 0$ のとき

$$\widehat{\theta}_n^{(m+1)}(\boldsymbol{x}) = \widehat{\theta}_n^{(m)}(\boldsymbol{x}) + \frac{S_n(\widehat{\theta}_n^{(m)}(\boldsymbol{x})|\,\boldsymbol{x})}{H_n(\widehat{\theta}_n^{(m)}(\boldsymbol{x})|\,\boldsymbol{x})}$$

と書ける[2]. ただし

$$H_n(\theta|\,\boldsymbol{x}) = -\frac{d^2}{d\theta^2}\log \mathsf{lik}_n(\theta|\,\boldsymbol{x})$$

である. この操作を $\widehat{\theta}_n^{(m+1)}(\boldsymbol{x})$ と $\widehat{\theta}_n^{(m)}(\boldsymbol{x})$ との差が十分小さくなるまで繰り返す.

Newton-Raphson 法を用いるために初期推定値 $\widehat{\theta}_n^{(0)}(\boldsymbol{x})$ が必要である. 初期推定値に何を用いるかによってアルゴリズムは収束したりしなかったりする. また尤度方程式 $S_n(\theta|\,\boldsymbol{x}) = 0$ が複数の解を持つ場合には, 尤度方程式 (3.2) の解は尤度関数の極小点, 極大点, 鞍馬点 (saddle point) に対応するので, $\widehat{\theta}_n^{(m)}(\boldsymbol{x})$ は最尤推定値とは異なる点に収束する可能性がある. 収束先が最尤推定値と異なるかどうかが不明な場合には, 複数の初期値で試すとよい. また初期値として, 別の推定値 (モーメント推定量などの別の推定量の実現値) を用いることもできる.

たとえば $\widehat{\theta}_n^{(0)}(\boldsymbol{X})$ が θ の十分よい推定量ならば一段階推定量

$$\widehat{\theta}_n^{(1)}(\boldsymbol{X}) = \widehat{\theta}_n^{(0)}(\boldsymbol{X}) + \frac{S_n(\widehat{\theta}_n^{(0)}|\,(\boldsymbol{X}))}{H_n(\widehat{\theta}_n^{(0)}|\,(\boldsymbol{X}))}$$

[2] (3.1) の右辺の第 2 項の「-」は H に入れていることに注意せよ.

は最尤推定量と同じ性質を漸近的には同じ性質をもつことが知られている．ただし, $\widehat{\theta}_n^{(0)}(\boldsymbol{X})$, $S(\widehat{\theta}_n^{(0)}|\,\boldsymbol{X})$ と $H_n(\widehat{\theta}_n^{(0)}|\,\boldsymbol{X})$ は $\widehat{\theta}_n^{(0)}(\boldsymbol{x})$, $S_n(\widehat{\theta}_n^{(0)}|\,\boldsymbol{x})$, $H_n(\widehat{\theta}_n^{(0)}|\,\boldsymbol{x})$ の $\boldsymbol{x}$ に $\boldsymbol{X}$ を形式的に代入したものとする． 正確に言えば $\sqrt{n}\{\widehat{\theta}_n^{(0)}(\boldsymbol{X}) - \theta\}$ は正規分布に分布収束するならば $\sqrt{n}\{\widehat{\theta}_n^{(1)}(\boldsymbol{X}) - \widehat{\theta}_n^{(0)}(\boldsymbol{X})\} \xrightarrow{P} 0\,(n \to \infty)$ となる．ただし $\widehat{\theta}_n(\boldsymbol{X})$ は θ の最尤推定量である．

auchy-newton **例 3.1.** この例では推定値のみを扱うので $\widehat{\theta}_n(\boldsymbol{x})$ を $\widehat{\theta}_n$ と書くことにする. $S(\theta|\,\boldsymbol{x})$, $H(\theta|\,\boldsymbol{x})$ も $S(\theta)$, $H(\theta)$ と書く. $\theta > 0$ とする．つぎの p.d.f. を持つ Cauchy 分布からの大きさ n のランダム標本を $X_1, X_2, \ldots, X_n$ とする.

$$\mathsf{p}(x|\,\theta) = \frac{1}{\pi\{1 + (x-\theta)^2\}} \qquad (-\infty < x < \infty).$$

このとき実現値 $x_1, x_2, \ldots, x_n$ に対する対数尤度関数は

$$\log \mathsf{lik}_n(\theta) = -\sum_{i=1}^{n} \log\{1 + (x_i - \theta)^2\} - n\log\pi$$

となる．最尤推定値 $\widehat{\theta}_n$ は尤度方程式

$$S_n(\widehat{\theta}_n) = \sum_{i=1}^{n} \frac{2(x_i - \widehat{\theta}_n)}{1 + (x_i - \widehat{\theta}_n)^2} = 0$$

の解である．$S_n(\theta)$ は θ の単調関数でないので, 与えられた $(x_1, x_2, \ldots, x_n)$ に対して尤度方程式は複数の解を持つ可能性がある．したがって適切な初期値 $\widehat{\theta}^{(0)}$ を選ぶことが重要である．Cauchy 分布は $\mathsf{E}[X_1]$ が定義されないので, 初期値として標本平均を用いるのは適当ではない．X_1 の分布は θ に関して対称であることに注目して, 標本中央値を初期値 $\widehat{\theta}^{(0)}$ として用いる．これを用いて逐次的に $\widehat{\theta}^{(m)}\,(m = 1, 2, \ldots)$ を

$$\widehat{\theta}_n^{(m)} = \widehat{\theta}_n^{(m-1)} + \frac{S_n(\widehat{\theta}^{(m-1)})}{H_n(\widehat{\theta}^{(m-1)})}$$

で定める．ただし

$$H_n(\theta) = 2\sum_{i=1}^{n} \frac{1 - (x_i - \theta)^2}{\{1 + (x_i - \theta)^2\}^2}$$

である．$\theta = 10$ の Cauchy 分布から標本の大きさ $n = 100$ のランダム標本に基づいて最尤推定値を求めた例が次である．

m	$\widehat{\theta}_{100}^{(m)}$	$\log \mathsf{lik}_{100}(\widehat{\theta}_{100}^{(m)}) + 100\log(\pi)$
0	9.932387	11.95144
1	9.98055	11.9517
2	9.980323	11.9517
3	9.980323	11.9517

□

つぎに母数の次元が $k\ (k \in \mathbb{N},\ k \geq 2)$ の場合を考える．母数 $\boldsymbol{\theta} = (\theta_1,\ \theta_2,\ \ldots,\ \theta_k)^\top$ が k 次元のとき，$\boldsymbol{\theta}$ の最尤推定値 $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ は尤度方程式

$$\boldsymbol{S}_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \boldsymbol{0}$$

の解として定義される．ただし

$$\boldsymbol{S}_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \left(\frac{\partial}{\partial\theta_1}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x}),\ \frac{\partial}{\partial\theta_2}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x}),\ \ldots,\ \frac{\partial}{\partial\theta_k}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})\right)^\top$$
$$\boldsymbol{0} = \underbrace{(0,\ 0,\ \ldots,\ 0)}_{k\ 個}{}^\top$$

である．さらに $\mathbf{H}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を $p \times p$ の行列で，$i,\ j = 1,\ 2,\ \ldots,\ k$ について $\mathbf{H}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ の $(i,\ j)$ 成分は

$$H_{n,\,ij}(\boldsymbol{\theta}|\,\boldsymbol{x}) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathsf{lik}_n(\boldsymbol{\theta})$$

で定義されるものとする．$\mathbf{H}_n(\widehat{\boldsymbol{\theta}}_n^{(m-1)}|\,\boldsymbol{x})$ が正則のとき，$m\ (m \in \mathbb{N})$ 回目の逐次解 $\widehat{\boldsymbol{\theta}}_n^{(m)}(\boldsymbol{x})$ は

$$\widehat{\boldsymbol{\theta}}_n^{(m)}(\boldsymbol{x}) = \widehat{\boldsymbol{\theta}}_n^{(m-1)}(\boldsymbol{x}) + \left\{\mathbf{H}_n(\widehat{\boldsymbol{\theta}}_n^{(m-1)}|\,\boldsymbol{x})\right\}^{-1} S_n(\widehat{\boldsymbol{\theta}}_n^{(m-1)}|\,\boldsymbol{x}) \tag{3.3}$$

eq:newton2

で定義される．

3.2 Fisher のスコア法

sec:3-2

Newton-Raphson アルゴリズムの簡単な修正として Fisher のスコアアルゴリズムがある．Fisher のスコアアルゴリズムは (3.3) の中の $\boldsymbol{H}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ の代わりに

eq:newton2

$$\mathbf{H}_n^*(\boldsymbol{\theta}) = \mathsf{E}_\theta[\mathbf{H}_n(\boldsymbol{\theta}|\,\boldsymbol{X})] = -\mathsf{E}_\theta\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{X})\right]$$

である．ただし

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{X})$$

の $(i,\ j)$ 成分は

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{X}) \qquad (i,\ j = 1,\ 2,\ \ldots,\ k)$$

である．

したがって $m\,(m \in \mathbb{N})$ 回目の逐次解 $\widehat{\boldsymbol{\theta}}_n^{(m-1)}(\boldsymbol{x})$ は

$$\widehat{\boldsymbol{\theta}}_n^{(m)}(\boldsymbol{x}) = \widehat{\boldsymbol{\theta}}_n^{(m-1)}(\boldsymbol{x}) + \left[\mathbf{H}_n^*(\widehat{\boldsymbol{\theta}}^{(m-1)})\right]^{-1} \boldsymbol{S}_n\left(\widehat{\boldsymbol{\theta}}_n^{(m-1)} \,\middle|\, \boldsymbol{x}\right) \tag{3.4}$$ eq:fisher-1

で定義される．

例 3.2. (例 3.1 の続き) ex:cauchy-newton

$$H_n(\theta) = 2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{\{1+(x_i-\theta)^2\}^2}$$

から

$$H_n^*(\theta) = \frac{2n}{\pi}\int_{-\infty}^{\infty} \frac{1-(x-\theta)^2}{\{1+(x-\theta)^2\}^3}\,dx = \frac{n}{2} \tag{3.5}$$ eq:keisan

から Fisher のスコアアルゴリズムは

$$\widehat{\theta}_n^{(m)} = \widehat{\theta}_n^{(m-1)} + \frac{4}{n}\sum_{i=1}^{n} \frac{x_i - \widehat{\theta}^{(m-1)}}{1+(x_i-\widehat{\theta}^{(m-1)})^2} \qquad (m = 1,\, 2,\, \ldots)$$

となる．最後に (3.5) eq:keisan の計算をする．$z = x - \theta$ とおく．さらに $y = \tan\gamma$ とおけば

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1-y^2}{(1+y^2)^3}\,dy &= \int_{-\infty}^{\infty} \frac{2}{(1+y^2)^3}\,dy - \int_{-\infty}^{\infty} \frac{1}{(1+y^2)^2}\,dy \\
&= 2\int_{-\pi/2}^{\pi/2} \frac{1}{(1+\tan\gamma^2)^3}\frac{1}{\cos\gamma^2}\,d\gamma - \int_{-\pi/2}^{\pi/2} \frac{1}{(1+\tan\gamma^2)^2}\frac{1}{\cos\gamma^2}\,d\gamma \\
&= 2\int_{-\pi/2}^{\pi/2} \cos^4\gamma\,d\gamma - \int_{-\pi/2}^{\pi/2} \cos^2\gamma\,d\gamma, \\
&= 2\left[\frac{\cos 4\gamma + 1}{8} + \cos 2\gamma + \frac{1}{4}\right]_{-\pi/2}^{\pi/2} - \left[\frac{\cos 2\gamma + 1}{2}\right]_{-\pi/2}^{\pi/2} \\
&= \frac{\pi}{4}
\end{aligned}$$

となることがわかる．

3.3 EM アルゴリズム

sec:3-3

$(\mathbb{X},\, \mathcal{B},\, \mu)$ を測度空間とする．ここで μ は σ 有限な測度[3]とする．$\mathbb{X}$ 値確率変数 $\boldsymbol{X}$ は母数 $\theta\,(\theta \in \Theta)$ の確率測度 P_θ を持つ．さらに P_θ は μ に

[3] ある部分集合の列 $\{G_n\}_{n=1}^{\infty}$, $G_n \in \mathcal{A}$ で $\cup_{n=1}^{\infty} G_n = \mathcal{X}$ かつ各 n に対して $\mu(G_n) < \infty$ なるものが存在することである．

関する p.d.f. $\mathsf{p}(\cdot\,|\,\theta)$ をもつとする．すなわち

$$\mathsf{P}_\theta(B) = \int_B \mathsf{p}(\boldsymbol{x}|\,\theta)\, d\mu(\boldsymbol{x}) \qquad (B \in \mathcal{B}).$$

をみたす非負関数 p が存在するとする．

いま $(\mathbb{X}, \mathcal{B})$ は隠れた空間で $\boldsymbol{X}$ のすべてを観測できないとする．実際にはある可測空間 $(\mathbb{Y}, \mathcal{C})$ と可測関数 $T : \mathbb{X} \mapsto \mathbb{Y}$ が存在して $\boldsymbol{Y} = T(\boldsymbol{X})$ のみが観測できるとする．T は $(\mathbb{Y}, \mathcal{C})$ 上の測度を誘導する．

$$\mathsf{Q}_\theta(C) = \mathsf{P}_\theta \circ T^{-1}(C) = \mathsf{P}_\theta(T^{-1}(C)) \qquad (C \in \mathcal{C})$$

とする．さらに ν を $(\mathbb{Y}, \mathcal{C})$ 上の σ 有限な測度としたとき，確率測度 Q_θ は測度 ν に関して p.d.f. $\mathsf{q}(\cdot\,|\,\theta)$ を持つとする．

$$\mathsf{Q}_\theta(C) = \int_C \mathsf{q}(\boldsymbol{y}|\,\theta)\, d\nu(\boldsymbol{y}) \qquad (C \in \mathcal{C})$$

である．

EM アルゴリズムは観測 $\boldsymbol{Y} = \boldsymbol{y}$ が与えられたとき，θ の関数として $\mathsf{q}(\boldsymbol{y}|\,\theta)$ を最大化することで θ の最尤推定値を求める方法である．

EM アルゴリズムは以下のように行う．$\boldsymbol{y}$ が与えられたとき θ の初期推定値の $\widehat{\theta}^{(0)}(\boldsymbol{y})$ から始める．$\widehat{\theta}^{(0)}(\boldsymbol{y})$ より $\mathsf{P}_{\widehat{\theta}^{(0)}}$ と $\mathsf{Q}_{\widehat{\theta}^{(0)}} = \mathsf{P}_{\widehat{\theta}^{(0)}} \circ T^{-1}$ が初期の推定された確率測度となる．

E 段階 (E-Step): 各 $\theta \in \Theta$ に対し，条件付き期待値

$$\phi_1(\theta|\,\boldsymbol{y}) = \mathsf{E}_{\widehat{\theta}^{(0)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\theta)\,\middle|\, T(\boldsymbol{X}) = \boldsymbol{y}\right] \tag{3.6}$$

eq:emalgo-1

を求める．ただし $\mathsf{E}_{\widehat{\theta}^{(0)}}[\cdot]$ は $\mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(0)})$ に関して期待値を取ったものである．同様に $\mathsf{P}_{\widehat{\theta}^{(0)}}$ は $\mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(0)})$ のもとでの確率分布である．$\phi_1(\theta)$ は $\boldsymbol{y}$ の関数になることに注意する．$\mathsf{P}_{\widehat{\theta}^{(0)}}\big(T(\boldsymbol{X}) = \boldsymbol{y}\big) > 0$ [4]の場合には，$T(\boldsymbol{X}) = \boldsymbol{y}$ が与えられたときの $\boldsymbol{X}$ の条件付き分布を求め，それに関して関数 $\boldsymbol{x} \mapsto \log \mathsf{p}(\boldsymbol{x}|\,\theta)$ の期待値を求めればよい．$\mathsf{P}_{\widehat{\theta}^{(0)}}\big(T(\boldsymbol{X}) = \boldsymbol{y}\big) = 0$ のときは注意が必要であるが (3.6) の正当化は可能であることが知られている．以後は簡単のために $\mathsf{P}_{\widehat{\theta}^{(0)}}\big(T(\boldsymbol{X}) = \boldsymbol{y}\big) > 0$ の場合を考える．

eq:emalgo-1

みればよいの
記すること．
/30 記．

M 段階 (M-Step): θ に関して

$$\phi_1(\theta|\,\boldsymbol{y})$$

を最大化する．最大を与える点 (存在すれば) を $\widehat{\theta}^{(1)}(\boldsymbol{y})$ とおく．つぎに $\mathsf{P}_{\widehat{\theta}^{(0)}}$ の代わりに $\mathsf{P}_{\widehat{\theta}^{(1)}}$ を用いる．

[4] $\mathsf{P}_{\widehat{\theta}^{(0)}}\big(T(\boldsymbol{X}) = y\big) := \mathsf{P}_{\widehat{\theta}^{(0)}}\big(\{\boldsymbol{x} \in \mathcal{X};\, T(\boldsymbol{x}) = y\}\big)$ の意味である．

E 段階 (E-Step): 各 $\theta \in \Theta$ に対して

$$\phi_2(\theta|\,\boldsymbol{y}) = \mathsf{E}_{\widehat{\theta}^{(1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\theta)\middle|\, T(\boldsymbol{X}) = \boldsymbol{y}\right]$$

を求める.

M 段階 (M-Step): θ に関して

$$\phi_2(\theta|\,\boldsymbol{y})$$

を最大化する. 最大を与える点 (存在すれば) を $\widehat{\theta}^{(2)}(\boldsymbol{y})$ とおく.

一般には m 段階 $(m = 1, 2, \ldots)$ は以下のようになる.

E 段階 (E-Step): 各 $\theta \in \Theta$ に対し, 条件付き期待値

$$\phi_m(\theta|\,\boldsymbol{y}) = \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\theta)\middle|\, T(\boldsymbol{X}) = \boldsymbol{y}\right] \tag{3.7}$$

eq:emalgo-2

を求める.

M 段階 (M-Step): θ に関して

$$\phi_m(\theta|\,\boldsymbol{y})$$

を最大化する. 最大を与える点 (存在すれば) を $\widehat{\theta}^{(m)}(\boldsymbol{y})$ とおく.

この操作を $\widehat{\theta}^{(m)}(\boldsymbol{y})$ が $\widehat{\theta}^{(m-1)}(\boldsymbol{y})$ とほとんど変化がなくなるまで繰り返す.

例 3.3. $\theta > 0$ とする. $X_1, X_2, \ldots, X_n$ は独立同一に指数分布

$$\mathsf{p}(x|\,\theta) = \frac{1}{\theta}e^{-x/\theta}\mathbb{1}_{(0,\infty)}(x), \qquad \mathbb{1}_{(0,\infty)}(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x \le 0) \end{cases}$$

に従う[5]とする. しかし各 $X_i\,(i = 1, 2, \ldots, n)$ は直接観測されず, 各 X_i の整数部分のみが観測されるとする. すなわち $Y_i = \lfloor X_i \rfloor$ である. $Y_1, Y_2, \ldots, Y_n$ の観測に基づいて θ の最尤推定値を求めよう.

この場合, $\mathbb{X} = (0, \infty)^n$, $\mathcal{B}$ は $(0, \infty)^n$ 上の Borel 可測集合族[6]であり, P_θ は $\mathbb{R}^n$ 上の Lebesgue 測度[7]に関する p.d.f.

$$\mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \mathsf{p}_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n|\,\theta) = \prod_{i=1}^{n}\mathsf{p}(x_i|\,\theta) \tag{3.8}$$

$$= \theta^{-n}\exp\left(-\sum_{i=1}^{n} x_i/\theta\right) \qquad (\boldsymbol{x} = (x_1, x_2, \ldots, x_n)) \tag{3.9}$$

eq:ex3-1

[5]指数分布の p.d.f. は

$$\mathsf{p}(x|\,\lambda) = \lambda e^{-\lambda x}\mathbb{1}_{(0,\infty)}(x)$$

であるが, 母数を $\lambda = 1/\theta$ としている.

[6]$(0, \infty)^n$ の開集合族を含む最小の σ 加法族のこと.

[7]$d\boldsymbol{x}$ と書くことにする. また $\mathbb{R}$ 上の Lebesgue 測度は dx と書く.

を持つ．また $\mathbb{Y} = \{0, 1, \ldots\}^n$ で，$\mathcal{C}$ は $\mathbb{Y}$ のすべての部分集合の集まりからなる σ 加法族である．さらに関数 $T : \mathbb{X} \mapsto \mathbb{Y}$ は

$$T(\boldsymbol{x}) = (\lfloor x_1 \rfloor, \lfloor x_2 \rfloor, \ldots, \lfloor x_n \rfloor)$$

で定義される．

~~いま~~ $y \in \{0, 1, 2, \ldots\}$ に対して，

$$\mathsf{Pr}(Y_i = y) = \int_y^{y+1} \mathsf{p}(x|\,\theta)\,dx = e^{-y/\theta}(1 - e^{-1/\theta}) \qquad (3.10) \quad \boxed{\texttt{eq:emalgo-2a}}$$

となる．このことより

$$\mathsf{q}(\boldsymbol{y}|\,\theta) = \mathsf{q}(y_1, y_2, \ldots, y_n|\,\theta) = \Pi_{i=1}^n e^{-y_i/\theta}(1 - e^{-1/\theta})$$

となることがわかる．直接 $\mathsf{q}(\boldsymbol{y}|\,\theta)$ を θ に関して最大化して θ の最尤推定値を求めることはできるが，EM アルゴリズムを用いるとどうなるかを観てみよう．

まず $\lfloor X_i \rfloor = y$ が与えられたとき，θ のもとでの X_i の条件付き p.d.f. を求める．

$$\begin{aligned}
\mathsf{k}_\theta(x|\,y) &= \lim_{\Delta x \to 0;\, \Delta x > 0} \frac{1}{\Delta x} \frac{\mathsf{Pr}(x \le X_i \le x + \Delta x,\ \lfloor X_i \rfloor = y)}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \\
&= \frac{1}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \lim_{\Delta x \to 0;\, \Delta x > 0} \frac{1}{\Delta x} \int_x^{x+\Delta x} \frac{1}{\theta} e^{-z/\theta} \mathbb{1}_{[y,\, y+1)}(z)\,dz \\
&= \frac{1}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \lim_{\Delta x \to 0;\, \Delta x > 0} \frac{1}{\Delta x} \Big[-e^{-z/\theta} \Big]_x^{x+\Delta x} \\
&= \frac{1}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \lim_{\Delta x \to 0;\, \Delta x > 0} \frac{-e^{-(x+\Delta x)/\theta} \mathbb{1}_{[y,\, y+1}(x + \Delta x) + e^{-x/\theta} \mathbb{1}_{[y,\, y+1)}(x)}{\Delta x} \\
&= \frac{-e^{-x/\theta}}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \lim_{\Delta x \to 0;\, \Delta x > 0} \frac{-e^{-\Delta x/\theta} \mathbb{1}_{[y,\, y+1}(x + \Delta x) + \mathbb{1}_{[y,\, y+1)}(x)}{\Delta x} \\
&= \frac{-e^{-x/\theta}}{\mathsf{Pr}(\lfloor X_i \rfloor = y)} \left(-\frac{1}{\theta} \mathbb{1}_{[y,\, y+1)}(x) \right) \\
&= \frac{\theta^{-1} e^{-x/\theta} \mathbb{1}_{[y,\, y+1)}(x)}{e^{-y/\theta}(1 - e^{-y/\theta})} \qquad (\because\ (3.10)) \quad \texttt{eq:emalgo-2a}
\end{aligned}$$

を得る．これより

$$\begin{aligned}
\phi_1(\theta) &= \mathsf{E}_{\widehat{\theta}^{(0)}} \left[\log p(\boldsymbol{X}|\,\theta) \middle| T(\boldsymbol{X}) = \boldsymbol{y} \right] \\
&= \mathsf{E}_{\widehat{\theta}^{(0)}} \left[-\frac{1}{\theta} \sum_{i=1}^n X_i - n \log \theta \middle| T(\boldsymbol{X}) = \boldsymbol{y} \right] \qquad ((3.9)\ を代入) \quad \texttt{eq:ex3-1} \\
&= -\frac{1}{\theta} \mathsf{E}_{\widehat{\theta}^{(0)}} \left[\sum_{i=1}^n X_i \middle| T(\boldsymbol{X}) = \boldsymbol{y} \right] - n \log \theta
\end{aligned}$$

と

$$
\begin{aligned}
\mathsf{E}_{P_{\widehat{\theta}^{(0)}}}\left[X_i|T(\boldsymbol{X})=\boldsymbol{y}\right] &= \frac{1}{\widehat{\theta}^{(0)}e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\int_y^{y+1} xe^{-x/\widehat{\theta}^{(0)}}\,dx \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{\left[-xe^{-x/\widehat{\theta}^{(0)}}\right]_y^{y+1}+\int_y^{y+1} e^{-x/\widehat{\theta}^{(0)}}\,dx\right\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\Bigg\{-(y+1)e^{-(y+1)/\widehat{\theta}^{(0)}}+ye^{-y/\widehat{\theta}^{(0)}} \\
&\qquad -\widehat{\theta}^{(0)}e^{-(y+1)/\widehat{\theta}^{(0)}}+\widehat{\theta}^{(0)}e^{-y/\widehat{\theta}^{(0)}}\Bigg\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{-(y+\widehat{\theta}^{(0)})e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})-e^{-(y+1)/\widehat{\theta}^{(0)}}\right\} \\
&= -\left(y-\frac{1}{e^{1/\widehat{\theta}^{(0)}}-1}+\widehat{\theta}^{(0)}\right)
\end{aligned}
$$

となる．よって E 段階は

$$
\phi_1(\theta|\,\boldsymbol{y}) = n\left(-\log\theta+\frac{1}{\theta(e^{1/\widehat{\theta}^{(0)}}-1)}-\frac{\bar{y}_n+\widehat{\theta}^{(0)}}{\theta}\right)
$$

となる．ただし $\bar{y}_n=(1/n)\sum_{i=1}^n y_i$ である．次に M 段階は上の式を θ に関して最大化する：

$$
\widehat{\theta}^{(1)}(\boldsymbol{y}) = \arg\max_\theta \phi_1(\theta|\,\boldsymbol{y}) = \widehat{\theta}^{(0)}(\boldsymbol{y})+\bar{y}_n-\frac{1}{e^{1/\widehat{\theta}^{(0)}(\boldsymbol{y})}-1}
$$

となる[8]．したがって EM アルゴリズムは

$$
\widehat{\theta}^{(m)}(\boldsymbol{y}) = \arg\max_\theta \phi_m(\theta|\,\boldsymbol{y}) = \widehat{\theta}^{(m-1)}(\boldsymbol{y})+\bar{y}_n-\frac{1}{e^{1/\widehat{\theta}^{(m-1)}(\boldsymbol{y})}-1}
$$

で与えられる． □

つぎに EM アルゴリズムがどうしてうまく働くかを観る．$\boldsymbol{Y}=\boldsymbol{y}$ が与えられたとき，$\widehat{\theta}$ を θ の最尤推定値とする． ~~とし~~さらに，Θ の内部上で関数

$$
\Theta \ni \theta \mapsto q(\boldsymbol{y}|\,\theta) \in \mathbb{R}
$$

は微分可能と仮定する． ~~する．~~ このとき $\widehat{\theta}$ は関数 $\theta \mapsto \mathsf{q}(\boldsymbol{y}|\,\theta)$ を最大にする点なので

$$
\left.\frac{\partial}{\partial\theta}\mathsf{q}(\boldsymbol{y}|\,\theta)\right|_{\theta=\widehat{\theta}} = 0
$$

[8] $\arg\max_\theta\phi(\theta)$ は関数 $\theta\mapsto\phi_\theta$ を最大にする点の集合を表す．したがって厳密には $\widehat{\theta}^{(1)}\in\arg\max_\theta\phi(\theta)$ と書くべきである．たとえば $\arg\max_{0\leq\theta4\pi}\sin\theta=\{\pi/2,\,5\pi/2\}$ となる．

である．ただし x が変数の関数 $f(x)$ に x_0 を代入することを $f(x)\big|_{x=x_0}$ と書いている．

いま $\widehat{\theta}^{(\infty)}(\boldsymbol{y})$ は~~を~~ Θ の内点とし，EM アルゴリズムの収束先とする．このとき，$\widehat{\theta}^{(\infty)}(\boldsymbol{y})$ は関数

$$\theta \mapsto \mathsf{E}_{\widehat{\theta}^{(\infty)}}\left[\log \mathsf{p}(\boldsymbol{X}\,\middle|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\right] \tag{3.11}$$

$\boxed{\texttt{eq:2.11}}$

を最大化する．eq:2.11 ~~(3.11) は~~ Θ の内部で微分可能とし，期待値と微分記号の交換が可能とすれば，$\theta=\widehat{\theta}^{(\infty)}(\boldsymbol{y})$ において

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathsf{E}_{\widehat{\theta}^{(\infty)}}\left[\log \mathsf{p}(\boldsymbol{X}\,\middle|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\right]\Bigg|_{\theta=\widehat{\theta}^{(\infty)}} &= \mathsf{E}_{\widehat{\theta}^{(\infty)}}\left[\frac{\partial}{\partial\theta}\log \mathsf{p}(\boldsymbol{X}|\,\theta)\,\middle|\,T(\boldsymbol{X})=\boldsymbol{y}\right]\Bigg|_{\theta=\widehat{\theta}^{(\infty)}} \\ &= 0 \end{aligned} \tag{3.12}$$

$\boxed{\texttt{eq:emalgo-3}}$

となる．ここで

$$S(\theta|\,\boldsymbol{x}) = \frac{\partial}{\partial\theta}\log \mathsf{p}(\boldsymbol{x}|\,\theta)$$

とおけば，適当な仮定のもとで

$$\begin{aligned}\mathsf{E}_\theta\left[S(\theta|\,\boldsymbol{X})\middle|T(\boldsymbol{X})=\boldsymbol{y}\right] &= \int_{T(\boldsymbol{x})=\boldsymbol{y}} S(\theta|\,\boldsymbol{x})\mathsf{p}(\boldsymbol{x}|\,\theta)\,d\mu(\boldsymbol{x}) \\ &= \int_{T(\boldsymbol{x})=\boldsymbol{y}}\left(\frac{\partial}{\partial\theta}\mathsf{p}(\boldsymbol{x}|\,\theta)\right)\,d\mu(\boldsymbol{x}) \\ &= \frac{\partial}{\partial\theta}\int_{T(\boldsymbol{x})=\boldsymbol{y}}\mathsf{p}(\boldsymbol{x}|\,\theta)\mu(d\boldsymbol{x}) \\ &= \frac{\partial}{\partial\theta}\mathsf{q}(\boldsymbol{y}|\,\theta)\end{aligned}$$

となる．すなわち，EM アルゴリズムの収束先 $\widehat{\theta}^{(\infty)}$ は，$\boldsymbol{Y}=\boldsymbol{y}$ を与えたときの θ の最尤推定値 $\widehat{\theta}$ に一致することがわかる．

よって，$\mathsf{q}(\boldsymbol{y}|\,\theta)>0$ となる $\boldsymbol{y}$ に対して，eq:emalgo-3 ~~(3.12) は~~

$$\frac{\partial}{\partial\theta}\log \mathsf{q}(\boldsymbol{y}|\,\theta)\Bigg|_{\theta=\widehat{\theta}^{(\infty)}} = 0 \tag{3.13}$$

$\boxed{\texttt{eq:emalgo-4}}$

を意味する．したがって，最尤推定値が一意に存在するならば，$\widehat{\theta}=\widehat{\theta}^{(\infty)}$ となる．

$\theta=\widehat{\theta}^{(\infty)}$ なる解をもつ方程式

$$\mathsf{E}_\theta\{S(\theta|\,\boldsymbol{X})|T(\boldsymbol{X})=\boldsymbol{y}\}=0$$

を**自己一致方程式** (self-consistency equation) という．

以上の議論から，つぎの場合には EM アルゴリズムはうまく機能されるかは保証されていないことがわかる．

1. 最大値を与える点が Θ の内点に含まれない.
2. 尤度関数がその最大を取る点で微分可能ではない.
3. スコア方程式 (3.13) が複数の解を持ち，そのいくつかは尤度関数を最大にしない.

最後に EM アルゴリズムの各段階で，対数尤度関数 $\log \mathsf{q}(\boldsymbol{y}|\,\theta)$ は非減少であることを示す：すなわち

$$\log \mathsf{q}(\boldsymbol{y}|\,\widehat{\theta}^{(m)}(\boldsymbol{y})) \geq \log \mathsf{q}(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y})), \qquad m = 0, 1, \ldots \tag{3.14}$$

eq:emalgo-5

を示す．まず $T(\boldsymbol{X}) = \boldsymbol{y}$ を与えたとき，$\boldsymbol{X}$ の条件付 p.d.f. $\mathsf{k}_\theta(\boldsymbol{x}|\,\boldsymbol{y})$ は~~次で与えられる：~~

$$\mathsf{k}_\theta(\boldsymbol{x}|\,\boldsymbol{y}) = \frac{\mathsf{p}(\boldsymbol{x}|\,\theta)}{\mathsf{q}(\boldsymbol{y}|\,\theta)} \mathbb{1}_{T^{-1}(\boldsymbol{y})}(\boldsymbol{x})$$

となることに注意する．上式から，$T(\boldsymbol{x}) = \boldsymbol{y}$, $\mathsf{p}(\boldsymbol{x}|\,\theta) > 0$, $\mathsf{q}(\boldsymbol{y}|\,\theta) > 0$ の場合，上式から

$$\log \mathsf{q}(\boldsymbol{y}|\,\theta) = \log \mathsf{p}(\boldsymbol{x}|\,\theta) - \log \mathsf{k}_\theta(\boldsymbol{x}|\,\boldsymbol{y})$$

を得る．~~となる.~~

以下では最尤推定値の候補は $\mathsf{q}(\boldsymbol{y}|\,\theta) > 0$ をみたしていなければいけないので，$\mathsf{q}(\boldsymbol{y}|\,\theta) > 0$ を仮定して議論を進める．なぜならば，~~これは，~~ そうでなければ，$\log \mathsf{q}(\boldsymbol{y}|\,\theta) > 0$ を最大にしないことからこのことはわかる．

各 m に対し

$$\begin{aligned}\log \mathsf{q}(\boldsymbol{y}|\,\theta) &= \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{q}(\boldsymbol{y}|\,\theta)\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] \\ &= \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\theta) - \log \mathsf{k}_\theta(\boldsymbol{X}|\,\boldsymbol{y})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] \\ &= \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\theta)\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_\theta(\boldsymbol{X}|\,\boldsymbol{y})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right]\end{aligned} \tag{3.15}$$

eq:emalgo-6

となる．上式の最右辺の各項を別々に評価していく．

まず，$\widehat{\theta}^{(m)}(\boldsymbol{y})$ の定義から

$$\mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\widehat{\theta}^{(m)})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}(\boldsymbol{X}|\,\widehat{\theta}^{(m-1)})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] \geq 0 \tag{3.16}$$

eq:emalgo-7

がわかる．

次に，(3.15) の最右辺の 2 項目を評価する：

$$\begin{aligned}&\mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] \\ &\qquad = \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right] \\ &\qquad \leq \log \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\,\middle|\,T(\boldsymbol{X}) = \boldsymbol{y}\right]\end{aligned} \tag{3.17}$$

eq:emalgo-6a

となる．最後の不等号は Jensen(注意 (A.55)) の不等式よりわかる．
いま

$$\mathsf{g}(\boldsymbol{y}) = \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left(\frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|\, T(\boldsymbol{X}) = \boldsymbol{y}\right) \tag{3.18}$$

eq:emalgo-8b

とおく．条件付き期待値の定義から任意の $C \in \mathcal{C}$ に対し

$$\int_C g(\boldsymbol{y})\, d\mathsf{Q}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y}) = \int_{T^{-1}(C)} \frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{x}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x}|\,\boldsymbol{y})}\, d\mathsf{P}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x}) \tag{3.19}$$

eq:emalgo-8

となる．しかし

$$\frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{x}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x}|\,\boldsymbol{y})} = \frac{\mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))}{\mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))} \cdot \frac{\mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))}{\mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))} \tag{3.20}$$

eq:emalgo-8a

に注意すれば (3.19) の右辺は

$$\begin{aligned}
&\int_{T^{-1}(C)} \frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\, d\mathsf{P}_{\widehat{\theta}^{(m-1)}} \\
&\quad = \int_{T^{-1}(C)} \frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})} \mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))\, d\mu(\boldsymbol{x}) \\
&\quad = \int_{T^{-1}(C)} \frac{\mathsf{p}(\boldsymbol{x}|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))}{\mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))} \mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))\, d\mu(\boldsymbol{x}) \qquad (\because (3.20)) \\
&\quad = \int_{T^{-1}(C)} \frac{\mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))}{\mathsf{q}(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))}\, d\mathsf{P}_{\widehat{\theta}^{(m)}}(\boldsymbol{x}) \\
&\quad = \int_C \frac{\mathsf{q}(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)}(\boldsymbol{y}))}{\mathsf{q}(\boldsymbol{y}|\,\widehat{\theta}^{(m)}(\boldsymbol{y}))}\, d\mathsf{Q}_{\widehat{\theta}^{(m)}}(\boldsymbol{y}) \qquad (\because \text{条件付き期待値の定義}) \\
&\quad = \int_C 1 \cdot d\mathsf{Q}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y})
\end{aligned}$$

となる．~~これより~~上式と (3.19) から，任意の $C \in \mathcal{C}$ に対して

$$\int_C g(\boldsymbol{y})\, d\mathsf{Q}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y}) = 1$$

となので

$$g(\boldsymbol{y}) = 1 \qquad (\mathsf{Q}_{\widehat{\theta}^{(m-1)}}\text{-a.e.})$$

がわかる．よって，(3.18) から

$$\mathsf{g}(\boldsymbol{y}) = \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\frac{\mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{\mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|\, T(\boldsymbol{X}) = \boldsymbol{y}\right] = 1, \qquad a.e. \quad \mathsf{Q}_{\widehat{\theta}^{(m-1)}}$$

となる．この式と (3.17) から

$$\mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})\middle|\, T(\boldsymbol{X})=\boldsymbol{y}\right] - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\right] \le 0 \tag{3.21}$$

がわかる．よって (3.16) と (3.21) から

$$\begin{aligned}
\log \mathsf{q}\big(\boldsymbol{y}|\,\widehat{\theta}^{(m)}\big) &= \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}\big(\boldsymbol{X}|\,\widehat{\theta}^{(m)}\big)|\,T(\boldsymbol{X})=\boldsymbol{y}\right] \\
&\quad - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m)}}\big(\boldsymbol{X}\big|\,\boldsymbol{y}\big)\big|T(\boldsymbol{X})=\boldsymbol{y}\right] \\
&\le \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{p}\big(\boldsymbol{X}|\,\widehat{\theta}^{(m-1)}\big)|\,T(\boldsymbol{X})=\boldsymbol{y}\right] \\
&\quad - \mathsf{E}_{\widehat{\theta}^{(m-1)}}\left[\log \mathsf{k}_{\widehat{\theta}^{(m-1)}}\big(\boldsymbol{X}\big|\,\boldsymbol{y}\big)\big|T(\boldsymbol{X})=\boldsymbol{y}\right] \\
&= \log \mathsf{q}\big(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)}\big)
\end{aligned}$$

となるので (3.14) は示せた.

注意 3.4. $\{\log \mathsf{q}(\boldsymbol{y}|\,\widehat{\theta}^{(m)})\}_{m=1}^{\infty}$ は非減少である．よって，この列が有界のとき，大域的に収束することがわかる．このことが EM アルゴリズムの安定性を保証することになる．[13, 17] を参照のこと．

黒田の 2 章の内容をきちんと書きたい. 2023/09/11 記.