データ分析プロセス
「3.2.9 完全情報最尤推定法」に関する補足

2015 年 10 月 1 日

1　完全情報最尤推定法についての補足

完全情報最尤推定法は拙著で取り上げている他の手法と異なり，欠損値を削除したり，代入することに重きが置かれているのではなく，平均値や分散などの統計量をバイアスを極力減らして推定するための手法です．そのため，本書でも平均値 $\boldsymbol{\mu}$ と分散共分散行列 $\boldsymbol{\Sigma}$ を求めているに留めています．

p.45 の図 3.8「欠損値への対応のフロー」では,「欠損値の特定」の後に,「欠損値の削除」,「最尤推定法」,「欠損値の代入」とあり，最尤推定法だけが特別扱いされているのは，欠損値を削除する対応でも，補完した値を代入する対応でもないことを表しています．この図は，Kabacoff[2] で取り上げられているものですが，この図からも完全情報最尤推定法が，欠損値を削除したり値を補完する手法とは異なることがわかります．

また，p.46 の表 3.1「欠損値への対応手法」には，完全情報最尤推定法の概要として,「サンプルごとに欠損パターンに応じた尤度関数を仮定して最尤推定を実施して得られる多変量正規分布を用いて平均値や分散共分散行列を推定」と記載しました．

2　サンプルデータを用いた完全情報最尤推定法に関する補足

従業員の IQ と業務のパフォーマンスの関係を表すデータに対して，p.58–p.59 では mvnmle パッケージの mlest 関数を用いて完全情報最尤推定法を実行しています．本書では紙面の関係上，省略しましたが，この実行例の背後で完全情報最尤推定法がどのように実行されるかについて補足説明します．

対象データには，20 個のサンプルがあります．

```
> employee.IQ.JP <- read.csv("data/missdata/employee_IQ_JP.csv", row.names=NULL,
+   colClasses=c(rep("integer", 3), "factor", "integer", "factor", "integer",
+                "factor"))
> employee.IQ.JP

   IQ JobPerformance MCAR MCAR.is.missing MAR MAR.is.missing MNAR
1  78              9   NA               1  NA              1    9
2  84             13   13               0  NA              1   13
```

```
3   84             10   NA              1  NA              1   10
4   85              8    8              0  NA              1   NA
5   87              7    7              0  NA              1   NA
6   91              7    7              0   7              0   NA
7   92              9    9              0   9              0    9
8   94              9    9              0   9              0    9
9   94             11   11              0  11              0   11
10  96              7   NA              1   7              0   NA
11  99              7    7              0   7              0   NA
12 105             10   10              0  10              0   10
13 105             11   11              0  11              0   11
14 106             15   15              0  15              0   15
15 108             10   10              0  10              0   10
16 112             10   10              0  10              0   10
17 113             12   12              0  12              0   12
18 115             14   14              0  14              0   14
19 118             16   16              0  16              0   16
20 134             12   NA              1  12              0   12
   MNAR.is.missing
1                0
2                0
3                0
4                1
5                1
6                1
7                0
8                0
9                0
10               1
11               1
12               0
13               0
14               0
15               0
16               0
17               0
18               0
19               0
20               0
```

平均値 $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$ が与えられている前提で，それぞれのサンプルが生起する確率密

度関数は，次式

$$f(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right) \tag{1}$$

で与えられます．ここで，d は $\boldsymbol{x}$ の次元です．

ここでは，IQ と JobPerformance の平均値 $\boldsymbol{\mu}$ が次式

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{\mathrm{IQ}} \\ \mu_{\mathrm{JP}} \end{pmatrix} \tag{2}$$

で表されているとします．ここで，μ_{IQ} は IQ の平均値，μ_{JP} は JobPerformance の平均値を表します．また，分散共分散行列 $\boldsymbol{\Sigma}$ は，次式

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{\mathrm{IQ}} & \sigma_{\mathrm{IQ,JP}} \\ \sigma_{\mathrm{JP,IQ}} & \sigma^2_{\mathrm{JP}} \end{pmatrix} \tag{3}$$

で表されているとします．ここで，σ^2_{IQ} と σ^2_{JP} はそれぞれ IQ と JobPerformance の分散，$\sigma_{\mathrm{IQ,JP}}$ と $\sigma_{\mathrm{JP,IQ}}$ は IQ と JobPerformance の共分散を表します．なお，$\sigma_{\mathrm{IQ,JP}} = \sigma_{\mathrm{JP,IQ}}$ なので，以下では特に断りがない限り $\sigma_{\mathrm{IQ,JP}}$ のみを用います．

まず，1 番目のサンプルは IQ が 78，JobPerformance が NA(欠損値) のため，

$$\boldsymbol{x}_1 = \begin{pmatrix} 78 \\ \mathrm{NA} \end{pmatrix} \tag{4}$$

と表されます．しかし，この 1 番目のサンプルでは JobPerformance が欠損しているため，JobPerformance の平均値，IQ と JobPerformance の共分散 $\sigma_{\mathrm{IQ,JP}}(= \sigma_{\mathrm{JP,IQ}})$ の推定にこのサンプルを役立てることは困難です．したがって，データは IQ だけの 1 行のベクトルであると考えて，

$$\boldsymbol{x}_1 = (78) \tag{5}$$

と表記します．また，平均値は IQ の平均だけの 1 行のベクトル

$$\boldsymbol{\mu} = (\mu_{\mathrm{IQ}}) \tag{6}$$

であり，分散共分散行列も単に IQ の分散だけの 1 行 1 列の行列

$$\boldsymbol{\Sigma}_1 = (\sigma^2_{\mathrm{IQ}}) \tag{7}$$

となります．このとき，1 番目のサンプルの確率密度関数は次式

$$f(\boldsymbol{x}_1|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}_1|^{1/2}} \exp\left\{-\frac{1}{2}\left((78 - \mu_{\mathrm{IQ}})\frac{1}{\sigma^2_{\mathrm{IQ}}}(78 - \mu_{\mathrm{IQ}})\right)\right\} \tag{8}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{IQ}}} \exp\left\{-\frac{1}{2\sigma^2_{\mathrm{IQ}}}(78 - \mu_{\mathrm{IQ}})^2\right\} \tag{9}$$

で与えられます．5 番目のサンプルまでは JobPerformance が欠損値であるため，1 番目のサンプ

ルと同様にして確率密度関数は，次式

$$f(\boldsymbol{x}_2|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{IQ}}}\exp\left(-\frac{1}{2\sigma_{\mathrm{IQ}}^2}(84-\mu_{\mathrm{IQ}})^2\right) \tag{10}$$

$$f(\boldsymbol{x}_3|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{IQ}}}\exp\left(-\frac{1}{2\sigma_{\mathrm{IQ}}^2}(84-\mu_{\mathrm{IQ}})^2\right) \tag{11}$$

$$f(\boldsymbol{x}_4|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{IQ}}}\exp\left(-\frac{1}{2\sigma_{\mathrm{IQ}}^2}(85-\mu_{\mathrm{IQ}})^2\right) \tag{12}$$

$$f(\boldsymbol{x}_5|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{IQ}}}\exp\left(-\frac{1}{2\sigma_{\mathrm{IQ}}^2}(87-\mu_{\mathrm{IQ}})^2\right) \tag{13}$$

で求められます．

6 番目から 20 番目までのサンプルでは，IQ と JobPerformance はともに欠損がなく値が観測されています．そのため，各サンプルの平均値と分散共分散行列は等しく，次式

$$\boldsymbol{\mu}_6 = \cdots = \boldsymbol{\mu}_{20} = \begin{pmatrix} \mu_{\mathrm{IQ}} \\ \mu_{\mathrm{JP}} \end{pmatrix} \tag{14}$$

$$\boldsymbol{\Sigma}_6 = \cdots = \boldsymbol{\Sigma}_{20} = \begin{pmatrix} \sigma_{\mathrm{IQ}}^2 & \sigma_{\mathrm{IQ,JP}} \\ \sigma_{\mathrm{JP,IQ}} & \sigma_{\mathrm{JP}}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{\mathrm{IQ}}^2 & \sigma_{\mathrm{IQ,JP}} \\ \sigma_{\mathrm{IQ,JP}} & \sigma_{\mathrm{JP}}^2 \end{pmatrix} \tag{15}$$

で表されます．また，分散共分散行列の行列式は，次式

$$|\boldsymbol{\Sigma}_6| = \cdots = |\boldsymbol{\Sigma}_{20}| = \sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2 \tag{16}$$

で表されます．

6 番目のサンプルの確率密度関数は，次式

$$\begin{aligned}
&f(\boldsymbol{x}_6|\boldsymbol{\mu},\boldsymbol{\Sigma}) \\
&= \frac{1}{2\pi|\boldsymbol{\Sigma}_6|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}_6-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}_6-\boldsymbol{\mu})\right) && (17)\\
&= \frac{1}{2\pi\sqrt{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2}}\times \\
&\quad \exp\left\{-\frac{1}{2}(91-\mu_{\mathrm{IQ}}, 7-\mu_{\mathrm{JP}})\frac{1}{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2}\begin{pmatrix} \sigma_{\mathrm{JP}}^2 & -\sigma_{\mathrm{IQ,JP}} \\ -\sigma_{\mathrm{IQ,JP}} & \sigma_{\mathrm{IQ}}^2 \end{pmatrix}\begin{pmatrix} 91-\mu_{\mathrm{IQ}} \\ 7-\mu_{\mathrm{JP}} \end{pmatrix}\right\} && (18)\\
&= \frac{1}{2\pi\sqrt{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2}}\times \\
&\quad \exp\left\{-\frac{1}{2(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2)}\begin{pmatrix} (91-\mu_{\mathrm{IQ}})\sigma_{\mathrm{JP}}^2-(7-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ,JP}} \\ -(91-\mu_{\mathrm{IQ}})\sigma_{\mathrm{IQ,JP}}+(7-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ}}^2 \end{pmatrix}^\top\begin{pmatrix} 91-\mu_{\mathrm{IQ}} \\ 7-\mu_{\mathrm{JP}} \end{pmatrix}\right\} && (19)\\
&= \frac{1}{2\pi\sqrt{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2}}\times \\
&\quad \exp\left\{-\frac{(91-\mu_{\mathrm{IQ}})^2\sigma_{\mathrm{JP}}^2-2(91-\mu_{\mathrm{IQ}})(7-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ,JP}}+(7-\mu_{\mathrm{JP}})^2\sigma_{\mathrm{IQ}}^2}{2(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2-\sigma_{\mathrm{IQ,JP}}^2)}\right\} && (20)
\end{aligned}$$

で算出されます．以下同様に，7 番目から 20 番目のサンプルの確率密度関数も算出できます．

$$
\begin{aligned}
f(\boldsymbol{x}_7|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = & \frac{1}{2\pi\sqrt{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2}} \times \\
& \exp\left\{ -\frac{(92-\mu_{\mathrm{IQ}})^2\sigma_{\mathrm{JP}}^2 - 2(92-\mu_{\mathrm{IQ}})(9-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ,JP}} + (9-\mu_{\mathrm{JP}})^2\sigma_{\mathrm{IQ}}^2}{2(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{IQ}}^2 - \sigma_{\mathrm{IQ,JP}}^2)} \right\} \qquad (21) \\
& \vdots \\
f(\boldsymbol{x}_{20}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = & \frac{1}{2\pi\sqrt{\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2}} \times \\
& \exp\left\{ -\frac{(134-\mu_{\mathrm{IQ}})^2\sigma_{\mathrm{JP}}^2 - 2(134-\mu_{\mathrm{IQ}})(12-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ,JP}} + (12-\mu_{\mathrm{JP}})^2\sigma_{\mathrm{IQ}}^2}{2(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2)} \right\} \\
& \qquad (22)
\end{aligned}
$$

以上により，サンプルの独立性を仮定すると，$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{20}$ の同時確率密度関数は，次式

$$
\begin{aligned}
f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{20}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= f(\boldsymbol{x}_1|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ldots f(\boldsymbol{x}_{20}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) && (23) \\
&= \prod_{i=1}^{20} f(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) && (24)
\end{aligned}
$$

で与えられます．(24) 式の両辺に対数をとると，

$$
\begin{aligned}
& \log f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{20}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
& = \sum_{i=1}^{20} \log f(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad (25) \\
& = \underbrace{-\frac{1}{2}\log 2\pi - \log\sigma_{\mathrm{IQ}} - \frac{(78-\mu_{\mathrm{IQ}})^2}{2\sigma_{\mathrm{IQ}}^2}}_{\text{1 番目のサンプルの対数尤度}} \underbrace{-\frac{1}{2}\log 2\pi - \log\sigma_{\mathrm{IQ}} - \frac{(84-\mu_{\mathrm{IQ}})^2}{2\sigma_{\mathrm{IQ}}^2}}_{\text{2 番目のサンプルの対数尤度}} \ldots \\
& \quad \underbrace{-\log 2\pi - \frac{1}{2}\log\left(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2\right) - \frac{(134-\mu_{\mathrm{IQ}})^2\sigma_{\mathrm{JP}}^2 - 2(134-\mu_{\mathrm{IQ}})(12-\mu_{\mathrm{JP}})\sigma_{\mathrm{IQ,JP}} + (12-\mu_{\mathrm{JP}})^2\sigma_{\mathrm{IQ}}^2}{2(\sigma_{\mathrm{IQ}}^2\sigma_{\mathrm{JP}}^2 - \sigma_{\mathrm{IQ,JP}}^2)}}_{\text{20 番目のサンプルの対数尤度}} \\
& \qquad (26)
\end{aligned}
$$

(26) 式が最大となるように，EM アルゴリズムを適用して，$\mu_{\mathrm{IQ}}, \mu_{\mathrm{JP}}$，および $\sigma_{\mathrm{IQ}}^2, \sigma_{\mathrm{IQ,JP}}(=\sigma_{\mathrm{JP,IQ}}), \sigma_{\mathrm{JP}}^2$ を求めます．本文中で `mvnmle` パッケージの `mlest` 関数を用いて完全情報最尤推定法を実行した結果，平均値 $\boldsymbol{\mu}$ と分散共分散行列 $\boldsymbol{\Sigma}$ は，次式

$$
\boldsymbol{\mu} = \begin{pmatrix} 99.99989 \\ 9.84867 \end{pmatrix} \qquad (27)
$$

$$
\boldsymbol{\Sigma} = \begin{pmatrix} 189.60050 & 28.369839 \\ 28.36984 & 8.617752 \end{pmatrix} \qquad (28)
$$

のように求まります．この結果を用いて，確率密度関数は次式

$$f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{29}$$

$$= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\begin{pmatrix} 99.99989 \\ 9.84867 \end{pmatrix}\right)^\top \begin{pmatrix} 189.60050 & 28.369839 \\ 28.36984 & 8.617752 \end{pmatrix}^{-1} \left(\boldsymbol{x}-\begin{pmatrix} 99.99989 \\ 9.84867 \end{pmatrix}\right)\right\} \tag{30}$$

$$= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\begin{pmatrix} 99.99989 \\ 9.84867 \end{pmatrix}\right)^\top \begin{pmatrix} 0.01039433 & -0.03421836 \\ -0.03421836 & 0.22868718 \end{pmatrix} \left(\boldsymbol{x}-\begin{pmatrix} 99.99989 \\ 9.84867 \end{pmatrix}\right)\right\} \tag{31}$$

で表されることがわかります．このようにして，平均値 $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$ を求めると，データの確率密度関数は求まりますが，各サンプルの補完すべき値まではわかりません．

以上の詳細については，本文中でも引用している Enders[1] を参照すると良いでしょう．また，本文中では引用していませんでしたが，Enders[1] の内容をコンパクトにまとめた村山 [3] は説明がわかりやすく，この資料から読むと理解が深まるかもしれません．

参考文献

[1] C.K.Enders. *Applied missing data analysis.* Guilford Press, 2010.

[2] R.Kabacoff. *R in Action: Data Analysis and Graphics with R.* Mannig Pubns Co, 2011.

[3] 村山航. 欠損データ分析 (missing data analysis) –完全情報最尤推定法と多重代入法–. `http://koumurayama.com/koujapanese/missing_data.pdf`, 2011.