

多重代入法におけるバイアス補正推定量に関する研究

富田 裕章

博士（統計科学）

総合研究大学院大学

複合科学研究科

統計科学専攻

平成29（2017）年度

博士論文

多重代入法におけるバイアス補正推定量に関する研究

富田 裕章

目次

第 1 章	序論	5
1.1	欠測データと統計解析	5
1.2	本論文の構成	7
第 2 章	欠測を含むデータの理論と解法	9
2.1	欠測を含むデータの尤度	9
2.2	欠測の機構	11
2.3	欠測のパターン	14
2.4	完全ケース解析	15
2.5	利用可能なデータを用いた解析	16
2.6	代入法	18
第 3 章	多重代入法	21
3.1	多重代入法の概要	21
3.2	多重代入法の理論	23
3.3	多重代入法の計算方法	27
3.4	多重代入法のパッケージ	32
3.5	多項式回帰における多重代入法	33
3.6	本章のまとめ	34
第 4 章	偏り補正を行う多重代入法	35
4.1	最尤推定による統合推定量の性質	36
4.2	重み付き最尤推定法	37
4.3	最小二乗条件付き密度推定	38
4.4	バイアス補正多重代入法	41
第 5 章	BCMI を用いた回帰分析	43

5.1	BCMI による推定量と実行方法	43
5.2	数値実験	44
5.3	実データ分析	47
第 6 章	多重代入法を利用した予測手法	53
6.1	BCMI による推定量	54
6.2	交差検証を利用した新たな予測法	54
6.3	提案手法の手順	55
6.4	数値実験	55
第 7 章	結論	61
付録 A	式 (4.3) の証明	63
付録 B	ベイズ統計学の概略	65
B.1	ベイズ推定	65
B.2	ベイズ統計の計算手法	67

第 1 章

序論

1.1 欠測データと統計解析

一般的な数理統計学の教科書で扱われる統計的推測理論の多くは、標本に欠測が発生せず完全に観測されること（、かつしばしばそれらは独立同一分布に従って発生していること、）を前提としている。しかしながら我々が観測によって得ることができる標本の多くには欠測が含まれうる。機械の作動状況や気象などを観測したデータであれば、観測機器の故障によって欠測は起こりうるし、調査研究におけるパネルデータであれば回答者の無回答によって起こりうる。医療における臨床試験であれば被験者の脱落によって値は欠測となりうる。こうした問題に対する対処方法は古くより様々な分野で検討がなされてきた。例えば調査研究における事後層別法を用いたレギング法 (Deming and Stephan (1940); Ireland and Kullback (1968)) は対応策のひとつである。あるいは臨床試験における経時データにおける脱落に対して、最後の観察された値を代入するという LOCF 法 (Last-Observation-Carried-Forward) も現在から 50 年以上前の 1965 年の論文に利用例が見られる (Woolley et al. (2009))。

他方、ひとつの分野に特化せず理論として欠測に対処する方法に関する研究についても様々な蓄積が存在する。データに欠測値が発生する原因となる変数の分類（欠測メカニズムの分類）や、欠測が標本に対してどのように起こっているかという欠測パターンの分類、あるいは欠測値を含むデータに対してどのように尤度や事後分布を定義するかといった解法の土台となる理論付けは古くから行われてきた。

それとともに、多数の欠測値を含むデータへ対する解法も提案されている。それらのいくつかについては次章以降紹介するが、本論文ではとりわけ代入法について焦点をあてる。代入法は欠測した値になんらかの値を補完を行って解析を行う手法である。代入法の最大の利点は、ひとたび欠測した値になんらかの値を代入すれば、全ての標本の値が揃うため通常の解析手法を適用可能であることである。従って、統計のテキストのみならず、一般的なデータ解析のハンドブックには、欠測値への対処法としてなんらかの代入法が記載されている。しかしなが

ら、このとき問題となるのは代入値の生成方法である。例えば $[0, 1]$ の範囲しかとりえない欠測値に対して 100 といった明らかに誤った数値を代入した場合、標本平均に一致性がなくなることは容易に理解ができる。しかしながら、統計的推定を考えると、一見問題がなく、ハンドブック等にも記載されている「観測された標本の平均値を欠測値に代入する」方法でも結果として一致性のない推定量や推定量の分散が得られてしまうことがある。

一般に推定量の分散まで正しく代入法で評価しようとした場合に用いられる手法として Rubin (1987) が提案した多重代入法 (Multiple Imputation) がある。この方法は補完、解析、統合の 3 つのステップで説明できる (Rubin (1996)) もので、補完を繰り返し行って得られる多数の疑似完全データから推定量を多数回計算し、それを統合することで、一定の条件の下で推定量およびその分散について正しく評価ができるものである。この手法は 2000 年代には医療統計では広く用いられている一般的な手法となっている (Mackinnon (2010))。

一方で、現在の多重代入法の利用状況を分析した Rezvan et al. (2015) によれば、機械的に多重代入法を適用して行った解析や、具体的にどのように多重代入法を適用したかを記していない論文もあると報告されている。多重代入法が適切な推定量を導くためには一定の条件が満たされることが必要であり、そのような状況でなければ偏りを含んだ望ましくない推定値となりうるということが知られている。多重代入法には様々な拡張があるが、推定量に一致性を担保するための条件として、「補完値の生成に使うモデル (代入モデル) が正しく同定できているか」ということが 1 つの条件となる。前述した Rezvan et al. (2015) においても、代入モデルを正しく同定できない場合には望ましくない推定値となりうると述べられている。しかしながら、真の構造が複雑であるほど、妥当な同定は難しくなることが知られている。例えば、変数同士が非線形な関係を含む場合や、連続変数と離散変数が混在する場合の同定は難しい問題である (Carpenter and Kenward (2013))。

これまでの研究では回帰モデルの推定における研究 (Bartlett et al. (2015)) や、代入モデルの同定を誤ったときの分散について考察した研究 (Robins and Wang (2000)) 等を除き、基本的に代入モデルは真の構造を含みうることが前提とされていた。本論文では、代入モデルが真の構造を含まなかったとしても、推定段階で補正を施して漸近的に一致性のある推定量を導出し、またその推定量を実際の統計的推定に利用するための方法 (バイアス補正多重代入法; BCMI) を提示している。この手法の利点は、代入モデルの同定が難しいような状況や、同定できたとしても多重代入法で必要とされる多数の独立した補完値の発生が難しい状況下でも、簡単に補完値を発生できる手法を用いて代入することで、多重代入法を適用して、一致性がある推定量を得ることが可能になることがあげられる。

次章以降で欠測値を含むデータに対する統計理論について整理を行い、バイアス補正多重代入法で利用する推定量の導出および評価を行う。

1.2 本論文の構成

本論文の構成は以下の通りである。第2章では多重代入法への導入として、欠測を含むデータを扱うための統計理論や、解析手法について論じる。欠測値を含む場合の尤度の定義や、欠測の機構や欠測の種類に関する分類について述べたのち代表的な欠測データを解析する手法について概観する。第3章では多重代入法について詳細に述べる。理論的正当化や一致性のある推定量を得るために必要な条件の検討、様々な多重代入法の計算アルゴリズムなどについて触れる。また、5章で行う数値実験の設定として利用した多項式回帰の問題に対する既存手法についても述べる。第4章以降においてバイアス補正を行う推定量について提案を行う。第4章では重みつき最尤法を用いたバイアス補正推定量およびその性質について述べる。第5章では特に回帰モデルの説明変数が欠測した際の解析への応用について述べる。第6章ではこの手法を利用した新たな予測方法について述べる。

第7章では第4, 5章で述べたバイアス補正推定量について考察を行うとともに、将来の課題について論じる。

第 2 章

欠測を含むデータの理論と解法

本章では欠測の基礎理論と解法について概観する。なお、基本的に頻度論に基づいて議論を行うが、多重代入法は本来ベイズ推定に基づく手法のため、必要に応じてベイズ推定に基づく解釈についても記述する。なお、ベイズ推定に関する基本的な知識は付録 B にまとめた。

2.1 欠測を含むデータの尤度

はじめに、欠測を含むデータをどのように記述するかについて、記号を導入しながら述べる。なお、本論文で扱う変数は、特別に記載がない限りすべて連続であるとする。

欠測を含むデータのモデリングについて考える。ここでは、Little and Rubin (2002) に従い、標本中の欠測値は本来実現値として存在するが、ある確率的な機構によって欠測が発生し観測されないと考える。まず $X = (X_1, \dots, X_p)$ を p 次元確率変数とする。 X の各成分が実際に観測されるか否かを p 次元二値確率変数 $R = (R_1, \dots, R_p)$ を用いて表現する。すなわち $R_i = 1$ のとき X_i は観測され、 $R_i = 0$ のとき X_i は欠測すると定義する。 R は観測指標 (response indicator)、もしくは欠測指標 (missing-data indicator) と呼ばれる。本論文では一貫して観測指標と呼ぶことにする。

この定義から理解されるように、標本に欠測値が含まれているか否かによらず、確率変数 X には常に全ての実現値が存在しており、観測指標 R の実現値に応じて X の実現値が観測されるか、欠測値となるかが決定されることになる。なお、欠測がないデータのことを完全データという。また、実際には欠測を含むデータ x に対して完全データを \tilde{x} として表記する。 \tilde{x}, x, r の関係を表 2.1 にまとめた。なお簡単のため、表中では X は一次元の確率変数とした。さらにここで記法を導入する。ある標本 X_j について観測された変数すべてを $x_{\cdot,j}^o$ と表すことにし、欠測した変数全てを $x_{\cdot,j}^m$ と表す。これは標本 X_j と観測指標の実現値 r_j を同時に表したものと考えることができる。また、ある変数 X_i が欠測した標本すべてを含む集合 M_i に属する標本 x_j を取り出し、 $x_{i,j}$ について独立に操作を行うことを簡略的に $x_{i,\cdot}^m$ と表現することが

表 2.1 完全データと観測指標，観測される標本の関係．完全データが観測指標にしたがって標本において実際に得られるか，欠測となるかが示されている．

r.v. \ j	1	2	...	$n-1$	n
\tilde{x}	2	3	...	1	2
r	1	0	...	1	0
			↓		
x	2	?	...	1	?

ある．また，各標本 x_j について， X_i を除く変数のことを， $x_{-i,j}$ と書く．また， X_i が観測された標本の集合を \mathcal{O}_i と書くことがある．さらに， x^o と書いた場合は観測された標本すべてをさすことにする．

続いて， x の尤度について考える．まず，完全データ \tilde{x} についてパラメトリックモデル $f(X; \theta)$ を用いて最尤推定を行うことを考えると，最尤推定量 $\hat{\theta}$ は

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} \log f(\tilde{x}; \theta) \\ &= \arg \max_{\theta} \sum_{j=1}^n \log f(\tilde{x}_j; \theta)\end{aligned}$$

と表される．この最尤推定量は， f が真の構造を含んでいるとき，一致性や漸近正規性を持つ．しかしながら，欠測を含む変数の場合にはこの尤度はそのまま計算することができない．

そこで，観測指標 R と付随するパラメータ φ を用いて，欠測を含んでモデリングすることを考える． φ は欠測の構造を表すパラメータと意味づけすることができる． X と R の同時分布 $f(x, r; \theta, \varphi)$ を考えると，この同時分布から R を消去すると

$$f(x; \theta, \varphi) = \sum_{r \in \Gamma} f(x, r; \theta, \varphi) \quad (2.1)$$

と表される．ただし Γ は R がとりうるすべてのベクトル値を含む集合である．なお，この同時分布は欠測が起こらない通常の同時分布 $f(x; \theta)$ を拡張したものである．これは欠測が存在しないとき Γ は $\Gamma = \{(1, 1, \dots, 1)^T\}$ という一点分布であることを考えれば容易に理解される．式 (2.1) を用いて欠測値を含んだ尤度を定義する．

単純に標本 (x_1, \dots, x_n) のうち，欠測した値 $x_{i,\cdot}^m$ をすべて積分消去することを考える．すなわち，尤度としては

$$\prod_{j=1}^n \left\{ \int_{i:r_{i,j}=0} f(x_{i,j}, r_{i,j}; \theta, \varphi) dx_{i,j} \right\} \quad (2.2)$$

となる。これを簡単に $f(x_r^o, r; \theta, \varphi)$ と表す。式 (2.2) を完全尤度 (full likelihood) とよぶ。明らかにこれは R の実現値によって異なる形となり、また標本 j ごとに欠測成分を積分消去するため、 j 毎に関数の定義域の次元は異なっている。また、Seaman et al. (2013) によれば、完全尤度は完全データ \tilde{x} と観測指標 r との関数で表されるといえる。

式 (2.2) の値を最大化するような $\hat{\theta}$ を求めるのが完全尤度からの最尤推定法である。また、この推定量 $\hat{\theta}$ には一貫性や漸近正規性があることを仮定して議論を進めることも多い (Wang and Robins (1998))。ただし、 θ と φ は分離不可能な場合があることに注意する。

ベイズ推定の場合は事後分布に着目する。事後分布は事前分布と尤度の積に比例することから、完全データの事後分布は

$$p(\theta|\tilde{x}) \propto f(\tilde{x}; \theta)p(\theta)$$

となる。ただし、 $p(\theta)$ は事前分布である。同様に欠測を含む同時分布から記載できる事後確率は

$$p(\theta, \varphi|x, r) \propto f(x, r; \theta, \varphi)p(\theta, \varphi)$$

である。欠測した値を積分消去すると、

$$p(\theta, \varphi|x^o, r) \propto \prod_{j=1}^n \int f(x_{:,j}^m, x_{:,j}^o, r; \theta, \varphi)p(\theta, \varphi)dx_{:,j}^m$$

となる。

2.2 欠測の機構

尤度を定義したので、次に欠測を含むデータを解析する上で必要となる欠測の機構について述べる。

先ほど定義した X と R の同時分布を $f(x, r; \theta, \varphi)$ であるが、一般にこのままでは推定すべきパラメータの数が多すぎることが知られている。多くの場合 X の母集団の推定に関心があるため、 θ の推定さえ正しければよく、 φ については推定を行わなくて良い場合が多いので、このまま推定を行う必要はないといえる。したがって、 θ の推定を容易にするために、 $f(x, r; \theta, \varphi)$ は条件付き分布の積に分解することを考える。分解の方法は大きく 2 通りに分けられる。一つ目の方法は選択モデル (selection model) と呼ばれる分解方法で、

$$f(x, r; \theta, \varphi) = f(x; \theta)f(r|X = x; \varphi) \quad (2.3)$$

と表現される。このモデルでは X に関する分布 $f(x; \theta)$ と、 X の実現値によって条件付けされた R の条件付き分布の積として同時分布が表現される。よって、 θ は X に欠測がないとき (す

すべての標本が完全に観測されたとき)の母集団分布のパラメータであり、 ϕ は条件付き分布を特徴付けるパラメータである。もう1つの分解方法はパターン混合モデル (pattern mixture model) とよばれる。この分解ではパラメータの意味が異なるため、同時分布を $f(x, r; \psi, \xi)$ と書き直すことにする。このとき

$$f(x, r; \psi, \xi) = f(x|r; \psi)f(r; \xi) \quad (2.4)$$

と分解するものである。このとき ψ は観測指標 R ごとの x の分布を特徴付けるパラメータで、 ξ は R の実現値 r の出現頻度を表す分布のパラメータであるといえる。こちらは、主として経時データ解析において用いられる分解方法である。それぞれの分解に応じて欠測のメカニズムが分類されるが、本論文では選択モデルを用いた考察を行うため、選択モデルの場合のみ記す。

選択モデルでは、Rubin (1976) が提唱した概念を用いて、完全にランダムな欠測 (missing completely at random, MCAR), ランダムな欠測 (missing at random, MAR), ランダムでない欠測 (missing not at random, MNAR) の3つに分類される。なお、MNARはMMAR (not missing at random) ということもある (Graham (2012))。

以下にそれぞれの定義を述べる。

Definition 1 (MCAR). 欠測メカニズムが完全にランダムな欠測であるとは、

$$f(r|X = x, \varphi) = f(r; \varphi) \quad (2.5)$$

がすべての r に対して成立することである。

Definition 2 (MAR). 欠測メカニズムがランダムな欠測であるとは、

$$f(r|X = x; \varphi) = f(r|X = x_r^o; \varphi) \quad (2.6)$$

がすべての r に対して成立することである。

上の二つの条件に当てはまらない場合は欠測メカニズムはランダムでない欠測 (MNAR) と分類される。すなわち、観測指標に関する条件付確率関数 $f(r|x; \varphi)$ が欠測している変数の値そのものに依存するような状況を MNAR と分類する。

Example 1. MAR なデータの例 (星野 (2009))

学校において、入学試験の妥当性を検証するために、入学試験の成績 y_1 と入学後の成績 y_2 の関係を調べたいとする。この場合、入学試験に不合格であった受験生を含めて解析を行いたいが、 Y_2 は Y_1 の値によって欠測が起こる。ここでは、ある定数 C に対して $y_1 \leq C$ となるとき y_2 が欠測することにする。

y_1 と y_2 の関係を線形回帰モデル

$$y_2 = \beta_0 + \beta_1 y_1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

で表す。このとき y_1 を条件付けたもとで欠測メカニズムを含めた尤度を書き下すと、

$$f(y_2, R|y_1, \beta, \sigma^2, C) = f(y_2|y_1, \beta, \sigma^2)f(R|y_1, C)$$

となるので、これは MAR の条件を満たす。よって観測されたデータから得られる尤度を最大化するような最尤推定量は、欠測の構造を無視して推定することができる。すなわち、 β_0 , β_1 , σ^2 の推定は y_1 , y_2 がともに観測されたデータのみ、すなわち、入学者の成績だけで行ってよいことになる（欠測を完全に無視してよい）。

さて、欠測を含んだデータの推定を検討する際に重要な概念として欠測の無視可能性がある。この概念は次のように定義される。

Definition 3 (ignorable). 欠測メカニズムが無視可能 (ignorable) であるとは、

- 欠測メカニズムが MAR の条件を満たすこと。
- パラメータ θ, φ が分離可能である。すなわち (θ, φ) のパラメータ空間がそれぞれの空間の直積であること。

が成り立つことである。

なお、前述の定義より MCAR は MAR の特別な場合と考えることができるので、欠測メカニズムが MCAR の場合でもパラメータの分離可能性を満たせば、無視可能な欠測であるといえる。また、ベイズ推定における定義では、分離可能性は事前分布が独立であることが定義となる。すなわち、事前分布を $p(\cdot)$ と表すことにすると、 $p(\theta, \varphi) = p(\theta)p(\varphi)$ が成り立てば、パラメータの分離可能性を満たす。

欠測の無視可能性が成り立つときは、完全尤度に対して重要な性質が成立する。同時分布から考えると、 R の実現値 r の 1 つ 1 つに対して

$$\begin{aligned} f(x, R = r; \theta, \varphi) &= f(x; \theta)f(R = r|x, \varphi) \\ &= f(x; \theta)f(R = r|x_r^o, \varphi) \end{aligned}$$

が成り立つ。完全尤度は、

$$\begin{aligned}
 & \prod_{j=1}^n \int_{i:r_{i,j}=0} f(x_{\cdot,j}; \theta) f(r|x_{\cdot,j}, \varphi) dx_{i,j} \\
 &= \prod_{j=1}^n \int_{i:r_{i,j}=0} f(x_{i,j}^m, x_{\cdot,j}^o; \theta) f(r|x_{\cdot,j}^o, \varphi) dx_{i,j} \\
 &= \prod_{j=1}^n \int_{i:r_{i,j}=0} f(x_{i,j}^m, x_{\cdot,j}^o; \theta) dx_{i,j} f(r|x_{\cdot,j}^o, \varphi) \\
 &= \prod_{j=1}^n f(x_{\cdot,j}^o; \theta) f(r|x_{\cdot,j}^o, \varphi)
 \end{aligned}$$

となる。したがって、完全尤度を最大にするような θ を考えるときは、欠測メカニズムについて考慮せずに、完全データに対する尤度 $f(\tilde{x}; \theta)$ から欠測した値を積分消去することで得られる尤度 $f(x^o; \theta)$ の最大化を行うことで得ることができる。

ベイズ推定の場合も、事後分布について同様の性質が成り立つ。事後分布 $p(\theta, \varphi|x^o, r)$ は、事前分布 $p(\theta, \varphi)$ を用いて

$$p(\theta, \varphi|x^o, r) \propto \prod_{j=1}^n \int_{i:r_{i,j}=0} f(x_{\cdot,j}; \theta) f(r|x_{\cdot,j}, \varphi) p(\theta, \varphi) dx_{i,j}$$

と書くことができる。よって、無視可能な欠測の下では、

$$\prod_{j=1}^n f(x_{\cdot,j}^o; \theta) p(\theta) f(r|x_{\cdot,j}^o, \varphi) p(\varphi)$$

と分解できる。よって、 θ の推測にあたっては観測データから得られる事後分布 $p(\theta|x^o)$ のみを考慮してもとめてもよいことがいえる。

2.3 欠測のパターン

ここでは欠測のパターンについて議論する。欠測のパターンとは、欠測を含んだ標本を並べたとき、どのように欠測が起こっているかを形式的に分類するものである。一般の欠測、すなわちどの変数も欠測メカニズムに従って確率的に欠測しうる状況下では、確率変数 X と同じ次元を持つ p 次元確率変数として表現された。このとき、 R の実現値である $p \times n$ 次元行列 r_{ij} は、 $r_{i,j} = 1$ のとき、標本 x_j の第 i 成分 $x_{i,j}$ が観測され、 $r_{i,j} = 0$ のときは欠測している状況を表していた。

さて、一部の欠測の問題においては観測指標を整理することができる。代表的な例が単調な欠測 (monotone missingness) とよばれる状況である。単調な欠測はパネル調査などでみられる状況である。

単調な欠測とは、変数や標本の添え字を適当に並び替えることで、確率変数 $X = (X_1, \dots, X_p)$ について、 X_i が欠測するならば、必ず X_{i+1} も欠測するという条件を満たす欠測である。すなわち、 X_1 が欠測するときは必ず X_2, \dots, X_p も欠測し、 X_1 が観測されていて X_2 が欠測するときは必ず X_3, \dots, X_p が欠測する、といった状況であり、図で表すと階段状の欠測となる。このとき、観測指標は $R = 0, \dots, p$ なる $p + 1$ 個の整数で表すことができる。すなわち、 $R = 0$ のとき全ての変数が欠測し、 $R = i$ ($i = 1, \dots, p$) のとき、 X_1, \dots, X_i が観測され X_{i+1}, \dots, X_p は欠測する状況を表すものである。よって、標本 x_1, \dots, x_n が得られたときには、観測指標は n 次元ベクトル r_j となり、 r_j は標本 x_j の欠測パターンを表していることになる。特に一変数のみが欠測する場合は観測指標は $R = 0, 1$ という二値変数で表すことが可能である。すなわち、欠測しうる変数を X_i としたとき、 $R = 0$ ならば X_i は欠測し、そうでなければ X_i は観測される。本論文で提案する手法は欠測しうる変数が高々 1 変数の状況を仮定する。したがって、第 4 章以降では断りがない限り R は 0, 1 の 2 値変数である。

2.4 完全ケース解析

本節からは欠測のあるデータの解析手法について述べる。解析手法のうちもっとも簡単と思われるものとして完全ケース解析 (Complete Cases analysis) を扱う。本手法はリストワイズ削除法 (List-Wise Deletion) と呼ばれる。完全ケース解析とはすべての変数が完全に観測されたデータのみを利用した解析のことである。この手法では、1 つでも欠測した変数を含む標本は解析には全く利用されない。この解析は非常に簡便であり、容易に実現可能である。欠測が MCAR の状況下では推定量に一致性が存在することも知られている。しかしながら、一般に利用できるデータが少なくなるため、推定量の分散は過大評価されることが知られている。また、MCAR の状況でない場合には、平均値の推定量や回帰係数の推定値に大きな偏りを含み得る。Little and Rubin (2002) では欠測の割合が大きくなるにつれて真の値とのずれが大きくなることを示しており、また、Schafer and Graham (2002) では、シミュレーションによって欠測メカニズムが MCAR, MAR, NMAR のそれぞれに従う状況でのパラメータの推定値や信頼区間の精度について検討をしており、MCAR でない場合は精度が悪化するを示している。よって、臨床試験などでは完全ケース解析は適切な手法ではないともされている (Council (2010))。しかしながら、完全ケース解析でも MCAR の場合や、MAR の場合でも目的変数が欠測した場合の回帰係数の推定では一致性を担保することが知られており、Little (1992))、そういった状況下では妥当でない代入をした場合よりもよい推定値を得ることができる。

2.5 利用可能なデータを用いた解析

利用可能なデータに基づく解析では、欠測した標本についても解析に取り込むものである。例えば、欠測したデータを積分消去した尤度に基づく最尤推定法や欠測する確率の逆数を用いてデータに重み付けを行う逆確率重み付け法 (Inverse Probability Weighted method: IPW 法) があげられる。特に IPW 法は二重にロバストな推定 (Doubly Robust Estimation) が構成できることが知られている。これらについての詳細は例えば、Seaman and White (2013) に記載されている。なお、次節で述べる代入法も広い意味では利用可能なデータを用いた解析である。

2.5.1 EM アルゴリズム

利用可能なデータを用いる解析において、完全尤度に基づいて最尤推定を行う手法があると述べたが、一般にこれを直接計算することは難しい。そこで、Dempster et al. (1977) は、欠測値を含むデータから最尤推定を行うアルゴリズムとして EM アルゴリズムを提案した。EM アルゴリズムは E-step と M-Step と呼ばれる 2 つのステップを繰り返し行うことで、推定値を最尤解に収束させるものである。EM アルゴリズムは潜在変数モデルなどその他の問題にも使われているが、ここでは触れない。

Dempster et al. (1977) に従い EM アルゴリズムの実行方法を述べる。欠測メカニズムは MAR を仮定し、完全データに対する対数尤度 $LL_C(\theta) = f(\tilde{x}|\theta)$ 、観測されたデータに対する尤度を $LL(\theta) = f(x^o|\theta)$ とする。

EM アルゴリズムは 2 段階のステップを交互に行う。

E Step

観測されたデータ x^o と t 回目の θ の推定値 $\theta^{(t)}$ が与えられた状態で、

$$Q(\theta; \theta^{(t)}) = E \left[LL_C(\theta | x^o, \theta^{(t)}) \right]$$

で、 LL_C の条件付き期待値を計算する。

M Step

E Step で求めた条件付き期待値 $Q(\theta; \theta^{(t)})$ を最大化するような $\theta^{(t+1)}$ を求める。すなわち、

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta; \theta^{(t)})$$

である。

反復を $\theta^{(t)}$ や $LL(\theta^{(t)})$ の変化量が少なくなるまで交互に繰り返すことで最尤推定値 $\hat{\theta}$ を求められる。

完全データが多変量正規分布に従う際、EM アルゴリズムを利用してパラメータの最尤推定量を求めることができる。

Example 2. 多変量正規分布の場合 (渡辺・山口 (2000))

p 次元確率変数 X が p 次元正規分布 $N_p(\boldsymbol{\mu}, \Sigma)$ に従い、欠測が発生している状況を考える。各 x_j について観測成分と欠測成分をそれぞれ $x_{:,j}^m, x_{:,j}^o$ と表記し、それぞれの成分に対応するパラメータ $\boldsymbol{\mu}$ を $\boldsymbol{\mu}^o, \boldsymbol{\mu}^m, \Sigma$ を

$$\Sigma = \begin{pmatrix} \Sigma^o & \Sigma^{om} \\ \Sigma^{mo} & \Sigma^m \end{pmatrix}$$

と表す。EM アルゴリズムを用いて欠測値を代入する方法は次の通りである。

E Step

十分統計量 $(x_j, x_j x_j^T)$ の条件付き期待値を計算する。

$$\begin{aligned} \mathbb{E} \left[x_{:,j}^m | x_{:,j}^o; \boldsymbol{\mu}^{(t)}, \Sigma^{(t)} \right] &= \boldsymbol{\mu}^{m(t)} + \Sigma^{mo(t)} \Sigma^{o(t)-1} x_{:,j}^o \\ &=: x_{:,j}^{m(t)} \\ \mathbb{E} \left[x_{:,j}^m x_{:,j}^{oT} | x_{:,j}^o; \boldsymbol{\mu}^{(t)}, \Sigma^{(t)} \right] &= x_{:,j}^{m(t)} x_{:,j}^{oT} \\ \mathbb{E} \left[x_{:,j}^m x_{:,j}^{mT} | x_{:,j}^o; \boldsymbol{\mu}^{(t)}, \Sigma^{(t)} \right] &= x_{:,j}^{m(t)} x_{:,j}^{o(t)T} + \Sigma_m^{(t)} - \Sigma^{mo(t)} \Sigma^{o(t)-1} \Sigma^{om(t)} \\ &=: \Sigma^{m(t)} \end{aligned}$$

となる。

M Step

上記 E ステップで求めた条件付き期待値を利用してパラメータの推定値を更新する。表記として、

$$\begin{aligned} x_j^{(t)} &= (x_{:,j}^o, x_{:,j}^{m(t)}) \\ (x_j x_j^T)^{(t)} &= \begin{pmatrix} \Sigma_o & x_{:,j}^o x_{:,j}^{m(t)T} \\ x_{:,j}^{m(t)} x_{:,j}^{oT} & \Sigma^{m(t)} \end{pmatrix} \end{aligned}$$

を導入して

$$\begin{aligned} \boldsymbol{\mu}^{(t+1)} &= \frac{1}{n} \sum_{j=1}^n x_j^{(t)} \\ \Sigma^{(t+1)} &= \frac{1}{n} \sum_{j=1}^n (x_j x_j^T)^{(t)} - \boldsymbol{\mu}^{(t+1)} \boldsymbol{\mu}^{(t+1)T} \end{aligned}$$

としてパラメータを更新する。

なお、Amelia II という R で多重代入法を計算するパッケージでは、EM アルゴリズムを推定に利用している。

2.6 代入法

代入法とは、欠測値に何らかの値を補完する方法である。補完する方法や補完する回数は様々である。多重代入法も代入法の 1 つであるが、これは次章で詳述し、本節では補完する回数が 1 回の単一代入法 (Single Imputation) について述べる。一度しか補完を行わないので後述する多重代入法と異なり、補完方法に確率的構造をもつことや、容易に多数の補完の候補値を発生させられることは必要とはされない。

欠測値を補完する代入法は多数存在するため、いくつかについて簡単に紹介する。

一番簡単な方法は観測された標本から求めた平均値の代入である。すなわち変数 X_i が欠測した標本 $x_{i,\cdot}^m$ に対して、 X_i が観測されたすべての標本 $x_{i,\cdot}^o$ から得られる平均値

$$x_{i,\cdot}^* = \frac{\sum_{j \in \mathcal{O}_i} x_{i,j}}{|\mathcal{O}_i|}$$

を代入するものである。この方法は簡単であり、なおかつ観測値の平均が代入値の平均となるので、MCAR の状況下では平均の推定量 \bar{x} は一致性をもつことになる。ただし、すべての欠測値に同一の平均値を入れるため、通常の完全データに対する解析手法をそのまま用いて分散やパーセンタイルといった推定量を一致推定することはできない (Little and Rubin (2002))。

また、回帰モデルを用いた代入法 (回帰代入法) も用いられる。これは変数 X_i に対して、

$$X_i = f(x_{-i}) + \epsilon_i$$

なるモデルを仮定して、代入を行うものである。ただし、 ϵ_i は平均 0、分散 σ^2 の乱数であり、一般に正規乱数であるとする。モデル f にはパラメトリック、ノンパラメトリックのどちらも仮定することが可能である。パラメトリックモデルの場合のパラメータ、ノンパラメトリックモデルの場合のチューニングパラメータについては観測された標本 $x_{i,j} \in \mathcal{O}_j$ を用いて推定したものをを用いる。なお、補完値の生成に当たっては

$$x_{i,j}^* = \hat{f}(x_{-i,j}^m)$$

とするものと

$$x_{i,j}^* = \hat{f}(x_{-i,j}^m) + \hat{\epsilon}_i$$

とするものの 2 通りが存在する．後者を特に確率的回帰代入法 (Stochastic Regression imputation) ということがある．

回帰代入法では，回帰モデルの設定（モデルの種類，説明変数として含む変数の選択），や推定の手法については解析者が目的に応じて選ぶことになる．古典的手法では，MCAR かつ同時分布が多変量正規分布に従うという仮定の下で，一般の欠測のパターンに対して，平均の推定量が一致性をもつ Buck の方法 (Buck (1960)) が存在する．また理論的には，回帰モデルの同定が正しければ MAR の状況下で平均は一致性をもつことが知られている (van Buuren (2012)).

ノンパラメトリックな回帰代入法は，近傍法を用いた代入 (Cheng (1984, 1994)) やカーネル密度推定で提案された Nadaraya-Watson 推定量 (Nadaraya (1964); Watson (1964)) を用いたカーネル回帰代入法 (Cheng (1994)) がある．カーネル回帰代入法では， $f(x_{-i})$ として，

$$f(x) = \frac{\sum_{j \in \mathcal{O}_j} K_h(x - x_{-i,j}) x_{i,j}}{\sum_{j \in \mathcal{O}_j} K_h(x - x_{-i,j})}$$

を利用する．ただし， K_h はカーネル関数 K を用いて

$$K_h(x) = \prod_{\substack{k=1 \\ k \neq i}}^p \frac{1}{h_k} K\left(\frac{x - x_k}{h_k}\right)$$

として表される関数である (Scott (2015)). K としてはガウシアン型のカーネル

$$K(x) = \exp(-x^2)$$

が良く用いられる．カーネル回帰を用いる場合は，次元が大きい場合には精度が悪化する「次元の呪い」には注意する必要がある．

その他の補完値の発生手法としては，何らかのマッチングを用いて他の標本と比較し，もっとも近そうな標本の値を代入する Hot Deck 法や，外部の知識を用いて何らかの定数を代入する Cold Deck 法があげられる．臨床試験などでよく用いられる LOCF 法は Cold Deck の一種と考えることができる．

単一代入法の問題点は，推定量の分散について一致推定が難しいことである．平均の推定量に一致性が保たれる場合でも，観測された標本と欠測値に対して観測された標本から発生させた補完値を区別せずに扱くと，完全データとは異なり分散が過小評価されてしまう．多重代入法ではこの問題に対してベイズ推定の枠組みからアプローチし，適切な分散を導出している．

最後に完全ケース，利用可能なケース，代入法の違いを述べた簡単な例題を挙げておく．

Example 3. $(x_1, y_1), \dots, (x_r, y_r), (x_{r+1}), \dots, (x_n)$ というデータが得られたとする．すなわち， (x, y) というデータの組において， $i = 1, \dots, r$ においては x, y ともに観測されているが， $i = r + 1, \dots, n$ において y は欠測していると考ええる．

このとき、推定量 $\boldsymbol{\mu} = (\mu_x, \mu_y)$ として (x, y) の期待値を考える。
完全ケース解析では

$$\hat{\mu}_x = \frac{1}{r} \sum_{i=1}^r x_i$$
$$\hat{\mu}_y = \frac{1}{r} \sum_{i=1}^r y_i$$

利用可能なケースでは

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\hat{\mu}_y = \frac{1}{r} \sum_{i=1}^r y_i$$

代入法では、欠測した y_{r+1}, \dots, y_n について、なんらかの手法によりそれぞれ補完する値を与える。補完した値を y_{r+1}^*, \dots, y_n^* と表せば、

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\hat{\mu}_y = \frac{1}{n} \left\{ \sum_{i=1}^r y_i + \sum_{i=r+1}^n y_i^* \right\}$$

である。

第 3 章

多重代入法

本章では、Rubin が提唱した多重代入法 (Multiple Imputation; MI) について説明する。まずはじめに多重代入法の概要についてまとめる。次節では Rubin による多重代入法の理論的な定式化について述べ、それを実現させる手法について説明する。多重代入法の拡張について述べる。特に、van Buuren らによる連鎖式を用いた多重代入法 (Multiple Imputation with Chained Equation; MICE) や、Bartlett らによる主要なモデルに互換的な完全条件付き尤度指定法 (Substantive model compatible fully conditional specification; SMC-FCS) など最近の展開についても触れる。最後に現在 R 等の統計解析ソフトウェアで利用可能な多重代入法の計算パッケージのうちいくつかについて、それらの概要について述べる。なお、本節で用いる訳語は高井他 (2016) を参考にした。

3.1 多重代入法の概要

多重代入法は、欠測値への補完値の代入を繰り返し行うことで得られる多数の推定値を統合して推定を行う手法である。歴史的な経緯を紐解くと「欠測値を独立に複数回補完する」という発想は既に 1970 年代後半頃には Rubin にあったようだが (Scheuren (2005); Rubin (2004, 1996)), Rubin (1987) において、ベイズ推定における事後期待推定量の近似値を、生成した多数のデータセットごとに求めた推定量から導出する方法として提唱されたものが、現在用いられている多重代入法の起源といえる。

多重代入法のアルゴリズムを簡単に記載すると、次の通りである。第一段階として、欠測値の補完を行う。この時に欠測値それぞれに対して、多数の補完値を確率的に発生させられる方法で補完を行い、1 つの欠測値に対して D 個の補完値を生成する。従って補完の結果、 D 個の擬似的な欠測のないデータセットができる。第二段階として、解析を行う。この際 D 個のデータセットそれぞれに対して推定を行い、推定値 $\hat{\theta}^1, \dots, \hat{\theta}^D$ を得る。第三段階として、各推定値 $\hat{\theta}^1, \dots, \hat{\theta}^D$ を統合して、最終的な統合推定値を得る。推定量の統合方法は Rubin のルー

ル (Rubin's rule) が良く用いられる。これは統合推定量 $\hat{\theta}_{MI}$ およびその分散 $\text{Var} [\hat{\theta}_{MI}]$ を与えるもので、それぞれ

$$\hat{\theta}_{MI} = \frac{1}{M} \sum_{d=1}^D \hat{\theta}^d \quad (3.1)$$

$$\begin{aligned} \text{Var} [\hat{\theta}_{MI}] &= \frac{1}{D} \sum_{d=1}^D \text{Var} [\hat{\theta}^d] + \left(1 + \frac{1}{D}\right) \frac{1}{D-1} \sum_{d=1}^D (\hat{\theta}^d - \hat{\theta}_{MI}) \\ &= \frac{1}{D} W_D + \left(1 + \frac{1}{D}\right) B_D \end{aligned} \quad (3.2)$$

と表現される。特に式 (3.2) で与えられる分散の公式については、各擬似完全データセット内での推定量の分散（推定量に付随する共分散行列） W_D と統合推定量との分散 B_D との和で表されることがわかる。

多重代入法を使う利点をいくつかあげる。第一の利点として、疑似完全データをひとたび作成した後は、統計解析において「欠測」が存在することを考えなくて良いことがあげられる。

第二の利点として、推定量の分散の推定が可能なが挙げられる。単一代入法ではしばしば推定量の漸近分散を過小評価することが問題となっているが、多重代入法では補完を複数回行った結果を統合することによって、漸近分散についても式 (3.2) によって評価することが可能である。

3.1.1 MI の例

簡単な例を示す。

Example 4. 平均を求める数値例 (Carpenter and Kenward (2013))

$X_j = j$ ($j = 1, \dots, 10$) とし、 Y を、

$$Y_j = 1 + X_j + e_j$$

として生成する。ただし、 e_j は標準正規分布に従う乱数である。

Y について欠測が発生し、その欠測メカニズムは $X > 5$ のときに限り 50% の確率で欠測が発生するものとする。 Y の平均の推定量 \bar{Y} を求めたい。

代入モデルは $Y_j^* = \beta_0 + \beta_1 X_j + \epsilon_j$ とする。ただし、 ϵ_j は平均 0、分散 σ^2 に従う乱数とする。 $\beta_0, \beta_1, \sigma^2$ はベイズ推測によって選ぶ。

ベイズ推測では、事前分布をうまく選択することで、最小二乗法によって求めた推定値

表 3.1 完全データ，観測値，疑似完全データセットの関係 (Carpenter and Kenward (2013))

	\tilde{x}	x	代入値			
			1	2	3	4
	2.23	2.23	2.23	2.23	2.23	2.23
	3.72	3.72	3.72	3.72	3.72	3.72
	3.54	3.54	3.54	3.54	3.54	3.54
	6.09	6.09	6.09	6.09	6.09	6.09
	6.53	6.53	6.53	6.53	6.53	6.53
	5.63	?	7.08	8.13	6.98	7.08
	8.28	8.28	8.28	8.28	8.28	8.28
	10.64	?	8.68	9.25	9.68	10.49
	10.74	10.74	10.74	10.74	10.74	10.74
	10.58	?	10.87	10.86	11.74	13.96
平均	6.80	5.88	6.78	6.94	6.95	7.27
標準偏差	1.00	1.13	0.94	0.96	1.01	1.16

$\hat{\beta}_0, \hat{\beta}_1, \hat{e}_i^2$ を用いて事後分布をそれぞれ，

$$p(\sigma^2|X, Y) = \frac{\sum_{j=1}^n \hat{e}_j^2}{(n-2)X^2}, \quad (3.3)$$

$$p(\beta_0, \beta_1|\sigma^2|X, Y) \sim N_2 \left[\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \sigma^2 \begin{pmatrix} n & \sum X_j \\ \sum X_j & \sum X_j^2 \end{pmatrix} \right] \quad (3.4)$$

と書き下すことができる。ただし， X^2 は自由度 $n-2$ の χ^2 分布に従う乱数である。

よって，事後予測分布で用いるパラメータとして $\hat{\sigma}^2$ を，式 (3.3) に基づいて補完ごとに M 回独立に発生させる。また， $\tilde{\beta}_0, \tilde{\beta}_1$ は $\hat{\sigma}^2$ を式 (3.4) に代入したうえで発生させる。この結果を用いて $Y_i^* = \tilde{\beta}_0 + \tilde{\beta}_1 X_i + \epsilon$, $\epsilon \sim N(0, \hat{\sigma}^2)$ から，補完値を発生させる。

完全データ \tilde{y} と観測値 y ，および 4 回の代入例を表 3.1 に示す。このとき，完全ケース解析の平均値は 5.88 であり，多重代入法により平均値は 6.99 であり，より完全データの平均に近い。また標準偏差はそれぞれ 1.13 と 1.05 であり，こちらも多重代入法のほうが完全データのものに近いといえる。

3.2 多重代入法の理論

本節では多重代入法の理論的な展開を述べる。

3.2.1 多重代入法の公式

ここでは MI の公式について考察を行う．多重代入法によって D 個の疑似完全データセットが得られた場合について考える．なお，補完値の生成方法は考察中に述べる．

欠測が無視可能であると仮定する．標本 x_1, \dots, x_n を X^o と書く．また，確率変数 X は $f(x; \theta)$ に従うとする．ここではパラメータ θ のベイズ推測を考える．すなわち， θ の事後期待推定量 $\hat{\theta}$ を求めたい．この $\hat{\theta}$ を観測データの事後分布 $p(\theta|X^o)$ から計算したいが， θ に関する期待値の計算が煩雑であると仮定する．

推定量を求めるために，完全データの尤度を考えると，

$$\begin{aligned} p(\theta|X^o) &= \prod_{j=1}^n \int p(\theta, X_j^m | X_j^o) dX_j^m \\ &= \prod_{j=1}^n \int p(\theta|X_j^o, X_j^m) p(X_j^m | X_j^o) dX_j^m \end{aligned}$$

と結びつけることができる．ただし， $p(X_j^m | X_j^o)$ は標本 X_j で欠測した成分に関して，観測された成分で条件付けした事後予測分布である．

これを用いて θ について考えると，

$$\begin{aligned} E[\theta] &= \int_{\theta} \theta \prod_{j=1}^n \int p(\theta|X_j^o, X_j^m) p(X_j^m | X_j^o) dX_j^m d\theta \\ &= \int \int \prod_{j=1}^n \theta p(\theta|X_j^o, X_j^m) d\theta p(X_j^m | X_j^o) dX_j^m \end{aligned} \quad (3.5)$$

$$\begin{aligned} V[\theta] &= \int \int_{\theta} (\theta - E[\theta])^2 \prod_{j=1}^n p(\theta|X_j^o, X_j^m) d\theta p(X_j^m | X_j^o) dX_j^m \\ &= \int \int_{\theta} (\theta - E[\theta|X_1^m, \dots, X_n^m])^2 \prod_{j=1}^n p(\theta|X_j^o, X_j^m) d\theta p(X_j^m | X_j^o) dX_j^m \\ &\quad + \int \int_{\theta} (E[\theta|X_1^m, \dots, X_n^m] - E[\theta])^2 \prod_{j=1}^n p(\theta|X_j^o, X_j^m) d\theta p(X_j^m | X_j^o) dX_j^m \end{aligned} \quad (3.6)$$

となる．

さて，各標本に対して補完値を事後予測分布 $p(X_j^m | X_j^o)$ から発生させ， X_j^d を D 個疑似完全データセットのうち d 番目の標本とする．また，事後分布として擬似完全データ $(X_1^o, X_1^d, \dots, X_n^o, X_n^d)$ を用いた $p(\theta|X_1^o, \dots, X_n^o, X_1^d, \dots, X_n^d)$ から計算される事後期待推定

量を $\hat{\theta}^d$ とする。また、 $\hat{\theta}^d$ の事後分散を V_d とする。すなわち、

$$\begin{aligned}\hat{\theta}^d &= \text{E} [\theta | X_1^d, \dots, X_n^d, X_1^o, \dots, X_n^o] \\ V_d &= \text{V} [\theta | X_1^d, \dots, X_n^d, X_1^o, \dots, X_n^o]\end{aligned}$$

である。 D が大きいとき、

$$\begin{aligned}\text{E} [\theta] &= \int \int_{\theta} \theta p(\theta | X_1^o, \dots, X_n^o, X_1^d, \dots, X_n^d) d\theta \prod p(X_j^m | X_j^o) dX_j^m \\ &\approx \frac{1}{D} \sum_{d=1}^D \int_{\theta} \theta p(\theta | X_1^o, \dots, X_n^o, X_1^d, \dots, X_n^d) d\theta \\ &= \frac{1}{D} \sum_{d=1}^D \hat{\theta}^d = \hat{\theta}_{MI}\end{aligned}$$

であり、

$$\text{V} [\theta] \approx \frac{1}{D} \sum_{d=1}^D V_d + \frac{1}{D-1} \sum_{d=1}^D (\hat{\theta}^d - \hat{\theta}_{MI})^2 \quad (3.7)$$

とできる。Rubin のルールでは式 (3.7) よりも良い近似として、

$$\text{V} [\theta] \approx \frac{1}{D} \sum_{d=1}^D V_d + \left(1 + \frac{1}{D}\right) \frac{1}{D-1} \sum_{d=1}^D (\hat{\theta}^d - \hat{\theta}_{MI})^2 \quad (3.8)$$

としている (Rubin (1987); Little and Rubin (2002))。

3.2.2 適切な多重代入法

MI によって得られる統合推定量の性質を考える上で Rubin が提唱した概念として、適切 (proper) という概念がある。定義を以下に述べる。

Definition 4 (適切な多重代入法). 多重代入法が適切 (proper) であるためには、以下の 3 条件を満たさなければならない。

1. $\hat{\theta}_{MI}$ は $M \rightarrow \infty$ で $N(\hat{\theta}, T_D)$ に収束する。ただし、 $\hat{\theta}$ は完全データで計算される推定値である。
2. W_D は $M \rightarrow \infty$ で $\text{Var} [\hat{\theta}]$ に収束する。
3. B_D は $M \rightarrow \infty$ で $\text{E} [B_D]$ に収束する。

この条件を満たさない多重代入法を improper な多重代入法と呼ぶことがある。しかしながら、多重代入法が適切であることが推定量に一致性が存在することの必要条件ではないため本論文ではあえて improper には適切でないという訳語を当てないことにする。

一般にベイズ推測に基づく予測分布に従って代入を行い、解析をすれば適切であるとされるが、状況によっては推定量の分散が条件を満たさないなど、適切でないこともありうる。このことは Nielsen (2003) などで議論がなされている。また、一般にこの条件を満たしているか確認するのは難しいとされる (van Buuren (2012))。

3.2.3 improper な多重代入法

すでに述べたように、多重代入法の適切さを満たすためには、基本的には欠測値に対する補完値は観測値を条件付けした事後予測分布から生成すればよい。しかしながら、事後予測分布の定義から、パラメータに関する積分が必要になり計算が困難である場合もある。そのため、事後予測分布の近似として例えば、

$$p(X_j^m | X_j^o, \tilde{\theta}) = f(X_j^m | \tilde{\theta}, X_j^o)$$

を用いる方法がある。 $\tilde{\theta}$ としてはベイズ推測に基づく推定量のほか色々なものが候補として考えられるが、ここでは最尤推定量を用いることにする。このとき得られる推定量の性質については、Wang and Robins (1998) において議論されている。それによれば、補完値の生成を完全ケースからの最尤推定によって求めた推定値から行い、Rubin のルールを適用した結果得られる統合推定量 (Type-B estimator と呼んでいる) には、一定の正則条件の下で推定量の漸近一致性が成り立つことを示している。

3.2.4 多重代入法の融和性

ここでは多重代入法が妥当な統合推定量を得られる条件を検討する上で重要となる補完値を生成する代入モデルと、推定量を求める解析モデルとの関係 (融和性) について考える。多重代入法では、代入モデルと解析モデルが同一である必要はなく、また用いる変数も同一である必要は無い。例えば、調査観察データにおいて調査者と解析者が別である場合、調査者が解析者には秘匿するデータも用いて多重代入法の第一ステップである補完を行ったほうが、解析者が補完から始めるよりも分散が小さい推定量が得られうる (超効率性, Rubin (1996))。

Meng (1994) では、代入モデルと解析モデルがあるとき、解析モデル f から作成されるような欠測値に対する事後予測分布と、代入モデルにおける事後予測分布が一致するとき、2つのモデルが融和的である、と定義している。また、融和性の観点から統合推定量が一致性を持つための条件が、高井他 (2016) において考察されている。それによれば、モデルの融和性と推定量の一致性の関係は表 3.2 となる。これより、一致性のある統合推定量を得るためには、基本的に代入モデルが真の構造を含む必要があることがわかる。本論文では代入モデルが真の構造を含むものを「妥当なモデル」と称し、真の構造からなる条件付き分布から代入できたことを「妥当に同定した、妥当に代入した」と定義する。

表 3.2 モデルの融和性と推定量の一致性の関係（高井他 (2016), 表 4.1 より）

モデルの関係	一致するための条件
解析モデルが代入モデルを含む 融和性がある	代入モデルが真のモデルを含む 代入モデルが真のモデルを含む
代入モデルが解析モデルを含む	解析モデルが真のモデルを含むか、解析モデルは代入モデルの周辺分布となる
上記の場合でない	代入モデルが真のモデルを含み、解析モデルが完全データで一致推定可能

3.2.5 多重代入法の代入回数

多重代入法の実現にあたっては、擬似完全データセットの個数、すなわち代入を何回行えば得られる統合推定量が妥当な近似となるかも重要である。

Rubin (1987, 1996) によれば、代入回数はせいぜい 5 回程度で十分とされていた。その後、数値実験に基づいていくつかの報告がなされている。Graham et al. (2007) によれば、検出力の観点から最低 20 回の代入を推奨している。この研究の代入回数は 100 回まで比較を行っている。また、White et al. (2011) の研究では、欠測割合に 100 を乗じた回数を提唱し、データ解析においては 100 から 500 回を利用している。

本論文で提案した手法は、代入回数を無限大とした時の漸近一致性を前提としたものである。よって、数値実験においては代入回数に起因した推定精度の変化がないようなるべく多い回数で施行することとし、既存研究を参考に代入回数を 100 回と定めた。

3.3 多重代入法の計算方法

本節では多重代入法を実現させる計算手法について述べる。代表的なものとして、逐次回帰代入法、同時分布モデリング、そして連鎖式による多重代入法についてまとめる。他の手法についても簡単に触れる。

3.3.1 逐次回帰代入法

逐次回帰代入法とは欠測が単調であるときに、逐次的に代入を行う方法である。欠測しうる変数を X_1, \dots, X_p とし、 X_i が欠測するときは必ず X_{i-1} までは欠測すると考える。また欠測が発生しない変数を Z とする。変数 X_1, \dots, X_p それぞれに対して、補完式 $h(X_i|x_1, \dots, x_{i-1}, z, \phi_i)$ と事前分布 $p(\phi_i)$ を仮定する。 i が小さいものから順番に補完を行う

手法である.

Example 5. 単調欠測データ (Carpenter, p.78)

データ $Y_{i,j}$ を標本 j の第 i 成分とする. 標本サイズは n , 次元は p とする. 単調に欠測しているため, 各 $i = 2, \dots, p$ について, Y_i を Y_1, \dots, Y_{i-1} で条件つけた回帰モデルを組むことができる.

$$Y_{i,j} = \beta_{0,j} + \sum_{k=1}^j \beta_{k,j} Y_{i,k} + e_{i,j}$$

$$e_{i,j} \sim N(0, \sigma_j^2)$$

とする. (Y_1, \dots, Y_i) が完全となるデータ (n_i 個あったとする) を用いて最小二乗推定量 $\hat{\beta}_{k,j}, \hat{\Sigma}_j^2$ を求める. 続いて, z を自由度 $n_i - i$ の χ^2 分布 $\chi_{n_i - i}^2$ からとり,

$$\tilde{\sigma}_i^2 = \frac{\hat{\sigma}_i^2(n_i - i)}{z}$$

と定める. また,

$$\tilde{\beta}_j \sim N(\hat{\beta}_j, \tilde{\sigma}_j^2 A_j)$$

$$A_j = \left(\sum_{i=1}^{n_j} (1, y_{i1}, \dots, y_{ij-1})(1, y_{i1}, \dots, y_{ij-1})^T \right)^{-1}$$

として, $i = n_j + 1, \dots, n$ について, 回帰式から欠測値を生成する.

3.3.2 同時分布モデリング

同時分布モデリングとは, 完全データの同時分布 $f(x; \theta)$ と事前分布 $p(\theta)$ を仮定し, それに対応する条件付き分布と条件付き分布のパラメータに対応する事前分布を用いて補完を行う手法である. 基本的に完全データの同時分布が多変量正規分布の場合に用いられる. Schafer (1997) に詳細な実行方法が記載されている.

Example 6. コレステロール値に関する推測 (Schafer (1997))

やけどをした患者の血清コレステロールの値に関する経時データを考える. データは $Y = (Y_1, Y_2, Y_3)$ という3変数からなり, それぞれ受傷後 2, 4, 14 日経過後の値である. 標本サイズ $n = 28$ で Y_3 にのみ欠測値を含む. 関心のある推定量は

- μ_3 : Y_3 の平均
- $\sigma_{13} = \mu_1 - \mu_3$: 変動の平均
- $\tau_{13} = 100 \times \sigma_{13} / \mu_1$: 変化率の平均

である。

データ拡大法を用いて補完を行う。Y は多変量正規分布 $N_3(\mu, \Sigma)$ に従うとする。そのため
にまず、事前分布、および計算を始めるための初期推定量を与える。ここでは、事前分布とし
て無情報事前分布 $p(\theta) \propto |\Sigma|^{-2}$ を用いる。また、初期推定量は最尤推定によって与えること
にする。

補完値を生成するに当たっては、 m 回独立にデータ拡大法を行ってそれぞれの収束値を利用
する。

3.3.3 連鎖式による多重代入法

本節では Buuren et al. (2006) による連鎖式による多重代入法 (MICE) について述べる。
なお、これは完全条件付き分布指定法 (FCS) とも呼ばれる。この方法は前述した同時分布モ
デリングとは異なり、同時分布 $f(x; \theta)$ についてのモデリングは行わず、したがって同時分布
について直接的な仮定を行わない。その代わりに欠測が生じた変数 1 つ 1 つに対して他の変
数からの回帰モデルと、それに付随する事前分布を仮定して、Gibbs Sampling 同様の計算を
行うことで、推定量や補完値を求める計算手法である。

よって、MICE では例えば欠測が生じる変数に離散変数と連続変数が混在する場合など、同
時分布によるモデリングが難しい場合にも補完値の生成や推定を行うことができる。反面、発
生させた補完値や計算された統合推定量が妥当なものかの検証は理論的には難しい。

MICE の概略

$X = (X_1, \dots, X_p)$ という p 次元確率変数とする。標本の数を n とする。表記を簡単にする
ために、 $X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)$ とする。また、 X_i が欠測となっている標本を
表す添え字として X_i^m を導入する。この添え字が現れているときの操作は、「 X_i が欠測した
すべての標本に対して、標本ごとに独立に同じ操作を行う」ことを表す。同様に X_i が観測さ
れたすべての標本を表す添え字として X_i^o を導入する。また、 $L(X_i^o | X_{-i}, \varphi_i)$ を「 X_i が観測
されたすべての標本に基づいて計算される条件付き尤度とする。

まず、 X_i ($i = 1, \dots, p$) が欠測したときに補完するモデルを $h_i(X_i^* | x_{-i}, \varphi_i)$ とする。同時
に、各 φ_i に対して事前分布 $p(\varphi_i)$ をそれぞれ定める。続いて、 X_i^m の初期値を、観測された
標本 X_i^o の中からランダムに選ぶ。

以下の操作を繰り返す。反復回数を t とする。 $X_{-i}^t = (X_1^t, \dots, X_{i-1}^t, X_{i+1}^{t-1}, \dots, X_p^{t-1})$ と
する。すなわち、 $i - 1$ 番目までの変数について更新が終わっている状況である。

1. φ_i^t を事後分布 $p(\varphi_i) L(X_i^o | X_{-i}^t, \varphi_i)$ より導く
2. X_i^m を条件付き分布 $h_i(X_i | X_{-i}^t, \varphi_i^t)$ より導く。

収束した X_i^m を補完値とする。

この手法は補完に用いる条件付き分布を与えていることを除けば、ギブスサンプリング法と形式的によく似ているものである。van Buuren (2012) では、ギブスサンプリングにおいて、発生する乱数の分布が定常分布に収束するための条件が MICE において成り立つかどうかについて考察を行っている。

Liu et al. (2014) では、上記の繰り返しによって Gibbs Sampling での定常分布にあたるものが存在した場合に、適切な事後分布に収束するための十分条件 (互換性; compatibility) を述べている。Hughes et al. (2014) によれば、次の条件を満たすとき、この代入が同時モデリングと同等になる。

- $h(x_i|x_{-i}, \varphi_i)$ が同時分布 $f(x; \theta)$ と互換的であること
- $i = 1, \dots, p$ に対して、 $f(x; \theta) = h_i(x_i|x_{-i}, \varphi_i)f(x_{-i}; \theta_{-i})$ と書け、パラメータが独立になること。また、事前分布について $p(\theta) = p(\varphi_i)p(\theta_{-i})$ と分解できること

なお、Liu et al. (2014) では、代入モデルが互換的でなくても、統合推定量が漸近一致性を持つための条件として準互換性 (semi-compatibility) という条件も提示している。

3.3.4 SMC-FCS

SMC-FCS は Bartlett et al. (2015) によって提案された手法である。この手法では回帰分析における説明変数が欠測したという状況について考える。一般に MICE では代入モデルとして目的変数を含めた他のすべての変数を用いたものを用いるが、この場合、回帰モデルが複雑になるとモデルが融和的であるか検討することが難しい。そこで、ベイズの定理を用いることで説明変数と外変数について融和的であれば、目的変数について条件付けしても同様に融和的になりうるという性質を用いて、MICE の改良を行ったものである。

一般に MICE では代入モデル $h(x_i|x_{-i}, y, z; \psi)$ について融和性があることを必要とする。しかしながら、解析モデル (主要なモデルが) が回帰モデルであることを利用し、回帰モデルが真の構造を含んでいるとすると仮定する。さて、ベイズの定理より、

$$h_i(x_i|x_{-i}, y, z) \propto h(y|x_i, x_{-i}, z)h_i(x_i|x_{-i}, z)$$

と表せることから、 $h_i(y|x_i, X_{-i}, z)$ には解析しようとする回帰モデル $f(y|x_i, x_{-i}, z; \theta)$ を当てはめることにする。これにより、 $h_i(x_i|x_{-i}, y, z)$ の同定の問題を $h_i(x_i|x_{-i}, z)$ の同定の問題に置き換えることができる。すなわち、同定した $h_i(x_i|x_{-i}, z)$ を用いて $f(y|x_i, X_{-i}, z; \theta)h_i(x_i|x_{-i}, z)$ に従う X_i を発生させれば、妥当な補完値を生成可能である。

しかし、 $f(x|x_i, x_{-i}, z; \theta)h_i(x_i|x_{-i}, z)$ は一般に通常のパラメトリックモデルでは書き下すことができない。よって棄却サンプリング法を用いて $h_i(x_i|x_{-i}, z)$ から

$f(y|x_i, x_{-i}, z; \theta)h_i(X_i|x_{-i}, z)$ に従う補完値 X_i^* を発生させることにする. 具体的には, 区間 $[0, 1]$ 上の一様乱数 U と, $h_i(x_i|x_{-i}, z)$ から発生させた補完値 X_i^* が

$$U \geq \frac{f(y|X_i = X_i^*, x_{-i}, z)}{c(y, x_{-i}, z)}$$

を満たすとき X_i^* を補完値として採択する. ただし, $c(y, x_{-i}, z)$ は $f(y|x, z; \theta)$ の X_i に対する上限値である.

SMC-FCS が妥当な推定量を導くためには, $h_i(x_i|x_{-i}, z)$ に対して以下の 2 条件が満たされなければならない.

- $h_i(x_i|x_{-i}, z, \varphi_i)$ が X, Z の同時分布 $f(x, z)$ と互換的であること
- $i = 1, \dots, p$ に対して, $f(x|z, \phi) = f(x_i|z, \phi_i)f(x_{-i}|z; \phi_{-i})$ と書け, パラメータが独立になること. また, 事前分布について $p(\phi) = p(\phi_i)p(\phi_{-i})$ と分解できること

また, 本手法では $c(y, X_{-i}, z)$ について評価が必要である. これが可能な場合として Bartlett らは解析モデル $f(y|x, z; \theta)$ が, (1) 誤差項が正規分布で表される回帰モデル, (2) Y が離散変数となるモデル, (3) 比例ハザードモデルの 3 つをあげている.

3.3.5 その他の多重代入法

ここまでに 3 つの多重代入法の実装方法について述べた. これらはすべて補完に用いる代入モデルとしてパラメトリックモデルを仮定してきた. もちろん, 補完方法にノンパラメトリックなモデルを用いて補完値を生成する方法も存在する. それらについて簡単に紹介する.

ノンパラメトリックな手法として代表的なものに Efron が提案したブートストラップ法 (Bootstrap) を用いて補完値を生成する手法がある (Rubin and Schenker (1986); Efron (1994)). この方法では補完値を生成するのは簡単であるが, 作成される値は観測された値に限られてしまう. したがって, 観測された標本の数が少ない場合は補完値として生成されうる値が限られてしまうという欠点がある.

このほか Titterington and Sedransk (1989) によるカーネル回帰とブートストラップを利用したノンパラメトリック代入法や, 単一代入法で触れた Cheng (1994) による Nadaraya-Watson 推定量を用いたカーネル回帰代入法を繰り返し適用する代入法がある. これらの方法には, 同じモデルから多数の補完値を独立に発生させることが難しいという問題がある. また, IPW 法との融合により二重にロバストな性質を持たせた多重代入法 (Seaman et al. (2012a)) といったものも存在する.

3.4 多重代入法のパッケージ

本節では代表的な計算パッケージについてまとめる。

3.4.1 Norm

Shafer が主として開発した R のパッケージで、おそらく多重代入法を実現させる最も古いものである。このパッケージでは、同時分布モデリング法を同時分布が多変量正規分布である場合に関して計算するものである。

3.4.2 Amelia

Amelia は Honaker らによって開発されている R のパッケージである (Honaker et al. (2011))。現在のパッケージは Amelia II という名称である。

利用するための前提条件は

- 完全データは多変量正規分布に従う。
- 欠測のメカニズムは MAR。

である。

補完アルゴリズムは EMB アルゴリズムを利用してパラメータ推定を行い、それに基づいて補完を行う。

EMB アルゴリズム

EMB(Expectation-Maximization Bootstrap) は Honaker and King (2010) によると、ブートストラップと EM アルゴリズムを組み合わせたパラメータの事後分布の推定方法である。

まず、欠測を含んだ n 個の標本を含むデータセットからブートストラップを行い、 D 個の新たなデータセットを作成する。生成した各データセットに対して EM アルゴリズムを用いて欠測値の補完とパラメータ推定を行う。推定値を Rubin のルールで統合して統合推定量とする。

3.4.3 mice

MICE を提案した、van Buuren らによって R で提供されているパッケージである (van Buuren and Groothuis-Oudshoorn (2011))。

mice の特徴である回帰式を自在に設定できることを有効に使えるよう、補完値の生成方法

は次の中から選ぶことができる。

- ベイズ線形回帰
- ベイズロジスティック回帰
- norm
- 平均値代入
- 予測平均マッチング
- 回帰による分類木
- 線形回帰（頻度論的）
- 多値回帰
- 二次多項式回帰

3.4.4 proc mi

SAS を用いた場合に、補完値を生成するメソッドである。詳細は Yuan (2011) を参照せよ。連続変数に対してはこれまでにのべた回帰式による代入法や、同時分布モデリング、mice のうち回帰、ロジスティック回帰、予測平均マッチング法などを用いることができる。

3.5 多項式回帰における多重代入法

本節では、数値実験にて扱う解析モデルが多項式回帰、すなわち目的変数 Y と説明変数 X の間に非線形関係が存在する場合における、多重代入法の適用について述べる。

二次多項式モデル $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$ (ϵ は正規乱数) が真の関係であり、解析モデルとして用いる状況下で説明変数 X が欠している場合の計算手法は Vink and van Buuren (2013) が提示した多項式結合法 (Polynomial Combination Method) がある。この方法では、まず $Z = \beta_1 X + \beta_2 X^2$ という新しい変数を作り、 X の欠測を Z の欠測と置き換えて補完を行った後、 X と X^2 の項に補完値を分解するという手段を用いる。この手法は MICE において実行可能である。

一般の多項式回帰モデルには以下の 4 手法がある。

第 1 の手法として、予測平均マッチング (Predictive Mean Matching; PMM) (Rubin (1986); Little (1988)) を利用する方法である。PMM は回帰代入法とマッチングを組み合わせた手法であり、回帰代入法によって生成される補完値

$$x_{i,\cdot}^* = \hat{f}(x_{-i,\cdot}^m)$$

と値が近い M 個の観測された標本の値を補完値とする。Seaman et al. (2012b) によればこの方法は推奨されないとされている。

2番目の方法として Von Hippel (2009) が提案した「別の変数として扱う」(Just Another Variable) 法がある。これは例えば X と X^2 を別の変数として取り扱い、それぞれ代入を行う方法である。ただし、Seaman et al. (2012b) において、欠測メカニズムが MCAR でない限り、この手法を用いると大きなバイアスが生じうると報告されている。

3番目の方法として、同時分布モデリングの利用があげられる。このとき、同時分布を通常のパラメトリックモデルによって記述することはできない (Carpenter and Kenward (2013)) ので、解析モデル $f(y|x)$ と説明変数に関する周辺分布 $g(x)$ の積 $f(y|x)g(x)$ としてモデリングする。これから代入モデルを導く。このとき、代入モデルと解析モデルの融和性は成立しているため、周辺分布 g を正しく同定できれば正しい代入となる。計算においてはメトロポリス・ヘイスティング法もしくは棄却サンプリングを用いる。本手法は代入モデルが多変量正規分布に従うなどの仮定が計算上必要などの条件がある。詳細は Carpenter and Kenward (2013) を参照されたい。

最後に Buuren et al. (2006) らによる MICE の利用がある。特に Bartlett et al. (2015) らによる SMC-FCS では、多項式モデルについても数値実験を行い、有用性を示している。

3.6 本章のまとめ

本章では多重代入法について概観した。最後に、補完値の生成方法の妥当性という観点から各手法を振り返りたい。

まず、代表的な多重代入法による計算手法として、逐次回帰法、同時分布モデリング法、連鎖式による多重代入法について述べた。逐次回帰法は単調な欠測である場合に有効な手法であるが、回帰モデルが正しいものでなければならない。同時分布モデリング法は Rubin の提唱した適切さの条件を満たすが、同時分布について正しく同定し、パラメトリックモデルによって書き下すことができなければならない。連鎖式による多重代入法は、同時分布に対する仮定は不要だが、各欠測値に対する代入モデルを真の同時分布と(準)互換性があるように記述しなければ、統合推定量には漸近一致性が担保できない。理論的観点からは、連鎖式による多重代入法を利用する上では検討すべき課題は多いといえる (van Buuren (2012))。

ノンパラメトリック法からの導出も簡単に触れた。ノンパラメトリック法では真の同時分布と融和的な代入モデルを同定する必要はないが、これら手法では多数の独立した補完値を生成することが難しい場合が多いと考えられる。

第 4 章

偏り補正を行う多重代入法

本章では偏り補正を行う多重代入法 (Bias corrected multiple imputation; BCMI) の導出と提案を行う。これは, Tomita et al. (2017); 富田他 (2017) において示した導出を一般の形にしたものである。

詳細な導出を行う前に, 本論文で提案する推定量について簡潔に説明を行う。提案する推定量は, 補完によって生成されたすべての擬似完全データから計算される尤度関数を統合して求められるものである。これは Rubin のルールによる統合推定量, 多重代入法で一般的に用いられている, 擬似データ毎に推定量を計算して求める推定量とは異なる。しかしながら, 提案する推定量は最尤推定法の枠組みで導出されるため, 推定量の一致性などの漸近的性質を検討することが容易である。よって, 代入による偏りを補正するような統合推定量の導出法として, すべての擬似完全データから計算される尤度関数を統合する方法を検討する。

なお, 過去にも Rubin のルールに基づかない統合推定量を用いた研究は存在する。例えば多重代入法を利用した対数尤度比検定を提案した Meng and Rubin (1992) では, 対数尤度比に基づく統計量の統合方法として, 単純な Rubin のルールとは異なる統合を行っている。

以後の節において提案推定量の導出を行う。まずはじめに代入モデルの同定が成功した場合に, 単純な最尤推定法に基づいて推定を行った場合でも推定量に漸近一致性が担保されることを示す。続いて, 代入モデルの同定が必ずしも成功していない場合にも, ある種の重みを補正として用いた重みつき最尤推定法によって推定量に漸近一致性が担保されることを示す。この重みは真の分布を利用するため, 直接計算することはできないが, 機械学習の手法を用いて分布を推定することで重みの推定を行うことが可能である。その手法について述べる。最後に提案する手法の手順についてまとめる。

本章で用いる記号の定義について述べる。 X は欠測が発生する変数であり, Z は欠測が発生しない変数である。 X の欠測メカニズムは MAR であるとする。標本の添え字を並び替えることで, $i = 1, \dots, n_1$ においては X は観測され, $i = n_1 + 1, \dots, n$ で X は欠測したものとする。欠測した X を補完する代入モデルを $h(X|z; \psi)$ とし, これは線形回帰モデル等の乱数

が発生させやすいモデルとする。補完を M 回繰り返すことにし、 X_i ($i = n_1 + 1, \dots, n$) の m ($m = 1, \dots, M$) 番目の補完値を X_i^m と表す。解析を行う主要なモデルを $f(x, z; \theta)$ とし、 θ の推定量 $\hat{\theta}$ について議論を行う。

4.1 最尤推定による統合推定量の性質

本節では多重代入法を用いたときの最尤推定量 $\hat{\theta}_{M,n}$ として、

$$\hat{\theta}_{M,n} = \arg \max_{\theta} L_{M,n}(\theta) \quad (4.1)$$

$$L_{M,n}(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i, Z_i; \theta) + \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n \log f(X_i^m, Z_i; \theta) \right\} \quad (4.2)$$

を考える。この推定量が代入モデル $h(x|z, \psi)$ が正しく特定されている場合に $M, n \rightarrow \infty$ とすると漸近一致性を持つことを示す。

はじめに $M \rightarrow \infty$ とすると、式 (4.2) は、

$$L_{M,n}(\theta) \xrightarrow{P} L_n(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i, Z_i; \theta) + \sum_{i=n_1+1}^n \int h(x|Z_i; \hat{\phi}) \log f(x, Z_i; \theta) dx \right\}$$

となる。続いて、 $n \rightarrow \infty$ とすると、

$$\begin{aligned} L_n(\theta) \xrightarrow{P} L(\theta) &= c \int g(x, z) \log f(x, z; \theta) dx dz \\ &\quad + (1 - c) \int g(z) \int h(x|z; \hat{\phi}) \log f(x, z; \theta) dx dz \end{aligned}$$

となる。ただし、 c は $n \rightarrow \infty$ としたときの n_1/n の極限であり、 $g(x, z)$ 、 $g(z)$ はそれぞれ X, Z の同時分布、 Z の周辺分布を表す。上式から正則条件 (van der Vaart (1998)) の下で

$$\hat{\theta}_{M,n} = \arg \max_{\theta} L_{M,n}(\theta) \xrightarrow{P} \arg \max_{\theta} L(\theta) = \tilde{\theta}$$

が成り立つことがわかる。代入モデルが正しく同定された時、すなわち、 $h(x|z; \hat{\psi}) = g(x|z, R = 0)$ となる場合を考えているので、式 (2.6) で表される MAR の仮定より、 z が与えられたもとでの x の条件付き分布やモーメントは R の実現値によらず等しくなること、すなわち

$$\begin{aligned} g(x|z, R = 0) &= g(x|z) \\ &= g(x|z, R = 1) \end{aligned} \quad (4.3)$$

が成立することに注意する。 $L(\theta)$ は

$$L(\theta) = \int g(x, z) \log f(x, z; \theta) dx dz$$

となる。よって、 $f(x, z; \theta^*) = g(x, z)$ となるような θ が存在するならば、 $\tilde{\theta} = \theta^*$ となり、漸近的に一致性を持つ。

さて、第 5 章では回帰モデルの推定問題を扱うが、この場合も同様である。解析モデルを $f(x|z; \theta)$ とすると、同様の变形で $M, n \rightarrow \infty$ のとき、 $\hat{\theta}_{M,n}$ は何らかの値 $\tilde{\theta}$ に収束する。条件付き分布の場合でも一致性を持つことを示すために条件付き分布に対するカルバック・ライブラー情報量を用いて考察を続ける。一般に $p_{X|Z}(x|z)$, $q_{X|Z}(x|z)$ をそれぞれ $Z = z$ で条件付けた X の条件付き密度関数とし、 $p_Z(z)$ を Z の周辺密度関数とする。 $p_Z(z)$ の下で 2 つの条件付き密度関数 $p_{X|Z}(x|z)$, $q_{X|Z}(x|z)$ のカルバック・ライブラー情報量 $D(p_{X|Z}, q_{X|Z}; p_Z)$ は

$$\begin{aligned} D(p_{X|Z}, q_{X|Z}; p_Z) &= E_{p_Z} \left[E_{p_{X|Z}} \left[\log \frac{p_{X|Z}(x|z)}{q_{X|Z}(x|z)} \right] \right] \\ &= \int p_Z(z) \int p_{X|Z}(x|z) \log \frac{p_{X|Z}(x|z)}{q_{X|Z}(x|z)} dx dz \end{aligned}$$

となる。

ここで、カルバック・ライブラー情報量は次の 2 条件を満たす。

- (i) $D(p_{X|Z}, q_{X|Z}; p_Z) \geq 0$.
- (ii) $D(p_{X|Z}, q_{X|Z}; p_Z) = 0 \Leftrightarrow p_{X|Z} = q_{X|Z}$ (a.e.)

このとき、 $\tilde{\theta}$ は、

$$\begin{aligned} \tilde{\theta} &= \operatorname{argmax}_{\theta} \int g(z) g(x|z) \log f(x|z; \theta) dx dz \\ &= \operatorname{argmax}_{\theta} \int p_X(x) p_{X|Z}(x|z) \log f(x|z; \theta) dx dz \\ &= \operatorname{argmax}_{\theta} -D(p_{X|Z}, f_{\theta}; p_Z) + \int g(z) \int g(x|z) \log g(x|z) dz dx \end{aligned}$$

となる。ただし、上式においてそれぞれ、 $f_{\theta}(x) = f(x|z; \theta)$, $p_Z(z) = g(z)$, $p_{X|Z}(x|z) = g(x|z)$ とする。なお、式 (4.3) の証明は付録 A に与えた。

ここで仮定より、 $p_{X|Z}(x|z) = f(x|z; \theta^*)$ となる θ^* が存在するから、

$$\tilde{\theta} = \operatorname{argmin}_{\theta} D(p_{X|Z}, f_{\theta}; p_Z) = \theta^*$$

が成り立つ。従って、 $\hat{\theta}_{M,n}$ は θ^* に確率収束する。

以上より、多重代入法を用いた解析において最尤推定法の枠組みで導かれる推定量を利用したとしても、漸近的に一致性を持つことが示された。

4.2 重み付き最尤推定法

前項では代入モデル $h(x|z; \phi)$ が正しく同定された場合について考察を行った。続いて、 $h(x|z; \phi)$ が正しく同定されていない場合について考える。この場合は、通常 of 多重代入法に

よる推定量 $\hat{\theta}_{MI}$ は θ^* に収束するとは保証されない (Rubin (1996), Clayton et al. (1998), Robins and Wang (2000)). また, 最尤推定法による推定量 $\hat{\theta}_{M,n}$ についても漸近バイアス $\tilde{\theta} - \theta^*$ が 0 でないため, やはり θ^* に収束することは保証できない. この問題を解決するために, 最尤推定に重みを加えた重み付き最尤推定法を利用する.

提案する重み付き最尤推定量 $\hat{\theta}$ は,

$$\hat{\theta} = \arg \max_{\theta} U_{M,n}(\theta) \quad (4.4)$$

$$U_{M,n}(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i|Z_i; \theta) + \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n w(X_i^m, Z_i, \hat{\phi}) \log f(X_i^m|Z_i; \theta) \right\}$$

である. ただし, $w(x, z, \phi)$ は重み関数で

$$w(x, z, \phi) = \frac{g(x|z)}{h(x|z; \phi)}$$

によって定義される.

前項同様に $M \rightarrow \infty$ とすると $U_{M,n}(\theta) \xrightarrow{P} U_n(\theta)$ となる. ここで $U_n(\theta)$ は

$$\begin{aligned} U_n(\theta) &= \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i, Z_i; \theta) + \sum_{i=n_1+1}^n \int h(y|X_i; \hat{\phi}) w(x, Z_i, \hat{\phi}) \log f(x, Z_i; \theta) dx \right\} \\ &= \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i, Z_i; \theta) + \sum_{i=n_1+1}^n \int h(x|Z_i; \hat{\phi}) \frac{g(x|Z_i)}{h(x|Z_i; \hat{\phi})} \log f(x, Z_i; \theta) dx \right\} \\ &= \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(X_i, Z_i; \theta) + \sum_{i=n_1+1}^n \int g(x|Z_i) \log f(x, Z_i; \theta) dx \right\} \end{aligned}$$

と表現される関数である. 従って $n \rightarrow \infty$ としたとき $U_n(\theta)$ は $L(\theta)$ に確率収束する. 前項の議論を用いれば, 提案した推定量 $\hat{\theta}$ は $M, n \rightarrow \infty$ とすれば, 適当な正則条件のもとで θ^* に確率収束する.

4.3 最小二乗条件付き密度推定

前項で提案した重み付き推定量は重み関数 $w(x, z, \phi)$ として,

$$w(x, z, \phi) = \frac{g(x|z)}{h(x|z; \phi)}$$

を用いたが, 一般に真の条件付き分布 $g(x|z)$ は未知であるため, 重み関数 $w(x, z, \phi)$ の値を求めることは不可能である. よって, 完全に観測されたデータ $\mathcal{Z} = \{(x_1, z_1), \dots, (x_{n_1}, z_{n_1})\}$ から条件付き密度推定で $g(x|z)$ を推定し, その結果を $\hat{g}(x|z)$ とし, 重み関数に代入した関数 $\hat{w}(x, z, \phi)$ の値を利用する.

条件付き分布の推定方法として, Sugiyama et al. (2010) による最小二乗条件付き密度推定 (Least Square Conditional Density Estimation; LSCDE) を利用する. LSCDE は条件付き分布を, 同時分布 $g(x, z)$ と z に関する周辺分布 $g(z)$ の密度比としてとらえ, Kanamori et al. (2009) によって提案された拘束無し最小二乗重要度適応法 (unconstrained Least-Squares Importance fitting; uLSIF) を用いて推定するものである. LSCDE によって推定される分布 $\hat{g}(x|z)$ は,

$$\hat{g}(x|z) = \sum_{b=1}^B \nu_b \psi_b(x, z)$$

と表される. ここで $\psi_b(x, z)$ は基底関数であり, ガウシアンカーネル

$$\psi_b(x, z) = \exp\left(-\frac{\|x - u_{xb}\|^2}{2\sigma_x^2}\right) \exp\left(-\frac{\|z - u_{zb}\|^2}{2\sigma_z^2}\right) \quad (4.5)$$

を利用する. また, (u_{xb}, u_{zb}) ($b = 1, \dots, B$) はカーネルの中心点であり, 完全に観測されたデータから B 個ランダムに選択する.

カーネルの係数 $\nu = (\nu_1, \dots, \nu_B)$ はリッジ罰則を用いた最小二乗推定量 $\tilde{\nu}$ に対し, $\hat{g}(x|z)$ が確率密度関数となるよう制約付けた推定量 $\hat{\nu}$ を利用する. 最小二乗推定量 $\tilde{\nu}$ は

$$\tilde{\nu} = \arg \min_{\nu} \left\{ \frac{1}{2} \iint (\hat{g}(x|z) - g(x|z))^2 g(z) dx dz + \delta \nu^T \nu \right\}$$

と定義されるが, これは ν に対する定数を無視し, 真の分布に基づく期待値を経験分布におきかえると,

$$\tilde{\nu} = \arg \min_{\nu} \left\{ \frac{1}{2} \nu^T \hat{H} \nu - \hat{h} \nu + \delta \nu^T \nu \right\}$$

となる. ただし, \hat{H} , \hat{h} はそれぞれ

$$\begin{aligned} \hat{H} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \int \psi(x, z_i) \psi(x, z_i)^T dx \\ \hat{h} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \psi(x_i, z_i) \end{aligned}$$

と表される関数である. なお, 上式では

$$\psi(x, z) = (\psi_1(x, z), \dots, \psi_B(x, z))^T$$

という表現を利用した. このとき, uLSIF による解は,

$$\tilde{\nu} = \left(\hat{H} + \delta I_B \right)^{-1} \hat{h}$$

と解析的に求められる. ただし I_B は B 次元単位行列である.

さらに、 $\hat{g}(x|z)$ は確率密度関数であるため、 $\hat{\nu}_b = \max\{\hat{\nu}_b, 0\}$ として係数がすべて 0 以上の値をとるようにしたうえで $\sum_{b=1}^B \hat{\nu}_b \psi_b(x, z)$ を正規化する処理を行う。その結果 $\hat{g}(x|z)$ は

$$\frac{\sum_{b=1}^B \hat{\nu}_b \psi_b(x, z)}{\int \sum_{b=1}^B \hat{\nu}_b \psi_b(x, z) dx}$$

として与えられる。なお、正則化項のパラメータ δ 、および基底関数のパラメータ σ_x, σ_z は $g(x|z)$ と $\hat{g}(x|z)$ 間のカルバック・ライブラー情報量に基づく K 分割交差検証法によって値を定める。

K 分割交差検証法では、真の同時分布 $g(x, z)$ と、 z の周辺分布 $g(z)$ と推定した条件付き密度 $\hat{g}(x|z)$ から構成される推定された同時密度とのカルバック・ライブラー情報量

$$KL_0 = \iint g(x, z) \log \frac{g(x, z)}{\hat{g}(x|z)g(z)} dx dz \quad (4.6)$$

に基づき、 KL_0 を最小にするようなパラメータを選択するものを用いる。ここで、

$$KL = - \iint g(x, z) \log \hat{g}(x|z) dx dz,$$

$$C_{KL} = \iint g(x, z) \log g(x|z) dx dz$$

という量を定義すると、式 (4.6) は、

$$KL_0 = KL + C_{KL}$$

と表される。このとき C_{KL} は推定量 $\hat{g}(x|z)$ について定数であることに注意すれば、 KL を最小化するような推定量となるパラメータを選択すれば KL_0 も最小化できることがわかる。

KL を小さくするようなパラメータを選ぶ K 分割交差検証法は以下の通りである。まず、完全に観測されたデータ \mathcal{Z} を K 分割し、 k 番目に属する標本を \mathcal{Z}_k と表す。続いて、すべてのパラメータの候補点 $\delta, \sigma_x, \sigma_z$ についてそれぞれ以下の操作を行い、 KL の推定値 $\hat{KL}_{\delta, \sigma_x, \sigma_z}$ を求める。はじめに $k = 1, \dots, K$ について、条件付き密度の推定量 $\hat{g}_{k, \delta, \sigma_x, \sigma_z}(x|z)$ を k 番目の分割以外に属する標本 $\mathcal{Z} \setminus \mathcal{Z}_k$ より作る。次に、 \mathcal{Z}_k を用いて、

$$\hat{KL}_{k, \delta, \sigma_x, \sigma_z} = \frac{1}{|\mathcal{Z}_k|} \sum_{(x, z) \in \mathcal{Z}_k} \hat{g}_{k, \delta, \sigma_x, \sigma_z}(x|z)$$

を計算する。最後に k について平均を取った値

$$\hat{KL}_{\delta, \sigma_x, \sigma_z} = \frac{1}{K} \sum_{k=1}^K \hat{KL}_{k, \delta, \sigma_x, \sigma_z}$$

をパラメータ $\delta, \sigma_x, \sigma_z$ に対応する KL の推定値とする。 $\hat{KL}_{\delta, \sigma_x, \sigma_z}$ を最小にするようなパラメータの値を採択する。

なお、上記手順で構成される推定量 \hat{KL} はほぼ KL と一致性をもつことが知られている (Schölkopf and Smola (2002)). また、カルバック・ライブラー情報量ではなく最小二乗法に基づいて交差検証を行う方法もある。詳しくは Sugiyama et al. (2012) を参照されたい。

4.4 バイアス補正多重代入法

本章では、重み付き最尤推定法と、その重みを推定するための条件付密度推定法について説明した。これらを利用し「バイアス補正多重代入法」(BCMI) のアルゴリズムを次のように提示する。

1. 代入モデル $h(x|z; \phi)$ を適当に定める。例えば、線形回帰モデルを利用する。完全ケースからの最尤推定によってパラメータを求める。
2. $h(x|z; \hat{\phi})$ から確率的回帰代入によって M 個の疑似完全データセットを生成する。
3. LSCDE によって $\hat{g}(x|z)$ を求める。
4. 各補完値 X_i^m について、重み \hat{w}_i を

$$\hat{w}(X_i^m, Z_i, \hat{\phi}) = \frac{\hat{g}(X_i^m | Z_i)}{h(X_i^m | Z_i; \hat{\phi})}$$

として計算する。ただし、推定を安定させるために $h(X_i^m | Z_i; \hat{\phi})$ の値が極めて小さい場合は重みを 0 とする。

5. 式 (4.4) に重みを代入して、重み付き最尤推定を行い推定量 $\hat{\theta}$ を求める

LSCDE におけるチューニングパラメータやカーネルの数 B 、及び、手順 4 において $w(x, z, \phi) = 0$ とする h の閾値はデータに基づいて適宜定める。

第 5 章

BCMI を用いた回帰分析

前章で提案した BCMI の応用として、欠測を含むデータに対して回帰分析を行うことを考える。本章では説明変数のうち 1 変数が欠測した場合について検討する。

はじめに、本章に渡って通用するモデルと変数の定義を行う。 Y を目的変数とし、 X, Z を説明変数とする。 X は欠測が発生しうる変数であり、ここでは一変数であるとする。 Y, Z には欠測が発生しないものとする。回帰モデルを $f(y|x, z; \theta)$ とし、 θ の推定を行うことを考える。前章同様、 X の欠測メカニズムは MAR とし、標本の添え字を適宜並び替えることで、 $i = 1, \dots, n_1$ ではすべての変数が観測され、 $i = n_1 + 1, \dots, n$ では X が欠測したとする。

欠測した X に対して、観測された変数 Y, Z から補完を行う代入モデルは $h(x|y, z; \psi)$ と表す。実際に補完を行うときには ψ として完全に観測されたデータに基づく推定値 $\hat{\psi}$ を利用する。欠測した標本 x_i について m 番目の補完値を X_i^m と表す。

5.1 BCMI による推定量と実行方法

前章で提案した BCMI の推定量は、本章の設定においては、

$$\hat{\theta} = \arg \max_{\theta} U_{M,n}(\theta), \quad (5.1)$$

$$U_{M,n}(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(Y_i|X_i, Z_i; \theta) + \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n w(X_i^m, Y_i, Z_i; \hat{\phi}) \log f(Y_i|X_i^m, Z_i; \theta) \right\},$$

と表せる。また、重み関数 $w(x, y, z, \phi)$ は

$$w(x, y, z, \phi) = \frac{g(x|y, z)}{h(x|y, z; \phi)}$$

となる。LSCDE によって推定した条件付き密度分布 $\hat{g}(x|y, z)$ は、

$$\hat{g}(x|y, z) = \sum_{b=1}^B \nu_b \psi_b(x, y, z)$$

となり、また、

$$\psi_b(x, y, z) = \exp\left(-\frac{\|x - u_{xb}\|^2}{2\sigma_x^2}\right) \exp\left(-\frac{\|y - u_{yb}\|^2}{2\sigma_y^2}\right) \exp\left(-\frac{\|z - u_{zb}\|^2}{2\sigma_z^2}\right) \quad (5.2)$$

である。ただし、式 (5.2) における (u_{xb}, u_{yb}, u_{zb}) , $(b = 1, \dots, B)$ は、完全に観測されたデータ $(x_1, y_1, z_1), \dots, (x_{n_1}, y_{n_1}, z_{n_1})$ よりランダムに選択する。

よって、第4章で述べた BCMI の実行方法は次のように書き換えられる。

1. 代入モデル $h(x|y, z; \phi)$ を定める。完全ケースからの最尤推定によってパラメータを求める。
2. $h(x|y, z; \hat{\phi})$ から確率的回帰代入によって M 個の疑似完全データセットを生成する。
3. LSCDE によって $\hat{g}(x|y, z)$ を求める。LSCDE のパラメータは $B = 100$ としたほか、チューニングパラメータはデータによって異なるため、数値実験および実データ解析それぞれで設定した値を述べることにする。
4. 各補完値 X_i^m について、重み $\hat{w}(x, y, z, \phi)$ を

$$\hat{w}(X_i^m, Y_i, Z_i, \hat{\phi}) = \frac{\hat{g}(X_i^m|Y_i, Z_i)}{h(X_i^m|Y_i, Z_i; \hat{\phi})}$$

として計算する。ただし、推定を安定させるために $h(X_i^m|Y_i, Z_i; \hat{\phi})$ の値が極めて小さい場合は重みを 0 とする。本章を通じて $h(X_i^m|Y_i, Z_i; \hat{\phi})$ が両側 80% 点を下回った場合は 0 とした。

5. 式 (4.4) に重みを代入して、重み付き最尤推定を行い推定量 $\hat{\theta}$ を求める。

B の値やチューニングパラメータ $\sigma_x, \sigma_y, \sigma_z$ および δ はデータによって最適なものを選ぶことになる。数値実験及び実データ解析においてどのような値を選んだかを記述する。

5.2 数値実験

本節では BCMI の性能を確認するために、数値実験を行って既存手法（補正を行わない通常の多重代入法 (MI), 完全ケース解析 (CC)) と比較した結果を述べる。White and Carlin (2010) は説明変数が欠測する状況下で MI と CC から得られる推定量のバイアスと効率性について比較を行った。その研究によれば、欠測メカニズムが MCAR もしくは MAR であり、かつ代入モデルが妥当であれば、MI によって得られる推定量のバイアスは CC によるものよ

りも小さくなると報告されている。それを踏まえて今回の数値実験では代入モデルを誤特定した場合を含むシミュレーションを行い、得られる推定量のバイアスについて検討を行う。

5.2.1 設定

欠測を含む説明変数 X とすべてが完全に観測される説明変数 Z はいずれも標準正規分布 $N(0, 1)$ から独立に発生し、目的変数 Y は

$$Y = X^d + 0.5Z + \varepsilon, \quad \varepsilon \sim N(0, 0.75^2), \quad (5.3)$$

に従って発生させる。ここで d は定数であり、今回は $d = 1, 2, 3$ の 3 通りについて検討した。欠測メカニズムは MAR であり、 X の観測指標 R は

$$\Pr(R = 1|Y, Z) = \frac{1}{1 + \exp(-a - Y)},$$

に従う。ここで a は定数であり、 X が観測される標本が約 70% となるように経験的に定める。

欠測した X を補完する代入モデル $h(x|y, z; \phi)$ を

$$X = \alpha_0 + \alpha_1 Y + \alpha_2 Z + \epsilon', \quad \epsilon' \sim N(0, \tau'^2), \quad (5.4)$$

と定める。式 (5.3) において $d \neq 1$ であるとき、式 (5.4) で与えられる代入モデルは妥当でないことは自明である。また、補完を行う際にはパラメータ $\alpha_0, \alpha_1, \alpha_2$ および τ'^2 は完全に観測されたデータを用いて最尤推定を行い、推定値 $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ 及び $\hat{\tau}'^2$ を利用する。補完後に X, Z から Y への回帰関係を推定する解析モデルを

$$Y = \beta_0 + \beta_1 X^d + \beta_2 Z + \epsilon, \quad \epsilon \sim N(0, \tau^2)$$

として、BCMI, MI, CC それぞれの方法で推定を行う。なお、このモデルには真の $Y|X, Z$ の関係が含まれていることに注意する。

パラメータ $\theta = (\beta_0, \beta_1, \beta_2, \tau^2)$ の推定量について検討する。BCMI, MI, CC それぞれの方法で得られた推定量を $\hat{\theta}_{BC}$, $\hat{\theta}_{CC}$ と $\hat{\theta}_{MI}$ とする。このとき、 $\hat{\theta}_{CC}$ と $\hat{\theta}_{MI}$ は

$$\begin{aligned} \hat{\theta}_{CC} &= \arg \max_{\theta} \sum_{i=1}^{n_1} \log f(Y_i|X_i, Z_i; \theta), \\ \hat{\theta}_{MI} &= \frac{1}{M} \sum_{m=1}^M \hat{\theta}^m, \\ \hat{\theta}^m &= \arg \max_{\theta} \left\{ \sum_{i=1}^{n_1} \log f(Y_i|X_i, Z_i; \theta) + \sum_{i=n_1+1}^n \log f(Y_i|X_i^m, Z_i; \theta) \right\} \end{aligned}$$

とあらわされる。また、 $\hat{\theta}_{BC}$ は式 (4.4) より得られる。

表 5.1 3 手法に基づくバイアスの比較

		BCMI	MI	CC
$d = 1$	β_1	0.042	0.000	-0.095
	β_2	-0.012	0.000	-0.047
$d = 2$	β_1	0.000	-0.363	-0.051
	β_2	-0.012	0.037	-0.044
$d = 3$	β_1	0.022	-0.537	-0.020
	β_2	0.001	0.016	-0.037

ここで、 $\hat{\theta}_{BC}$ を計算する際に用いた LSCDE のパラメータの設定を述べる.. はじめに、カーネルの数 B について $B = 100$ とした. ただし、カーネルを構成するのに利用できる標本の数が B を下回る時は、標本の数とした. また、 $\sigma_x, \sigma_y, \sigma_z$ は、 σ_x, σ_z は 0.01, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.3 からとり、 σ_y は 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6 から選択を行った. また λ は 0.001 から 10 の間で対数スケールで等間隔に 9 点の中から選んだ.

標本の数は $n = 300$ とし、代入を行う回数を $M = 100$ とした. また 1000 回の数値実験を繰り返し行った. なお、Rubin のルールに基づかず、単純な最尤推定法を用いて得られる MI の推定値もほぼ Rubin のルールに基づく通常の MI で得られるものとほぼ一致していることや、通常の MI のほうが実際の統計解析にあたって広く用いられていることを考慮し、今回の数値実験では通常の MI を用いている.

5.2.2 結果

各手法の評価の指標として、回帰係数 β_1, β_2 に対する推定量のバイアスを用いた. すなわち、

$$\text{Bias}(\hat{\beta}_j) = E \left[\hat{\beta}_j - \beta_j \right] \quad (j = 1, 2)$$

を計算し、比較した. 数値実験より近似的に計算されたバイアスを表 5.1 に示す. 代入モデルを正しく同定できる $d = 1$ の場合では、MI は回帰係数 β_1, β_2 両方に対してほぼバイアス無く推定できている. しかしながら、正しく同定できない $d = 2, 3$ の場合では、MI では大きなバイアスが生じる. BCMI について考えると、 $d = 1$ のときは MI よりもバイアスが大きい、十分 0 に近い値といえる. また MI ではバイアスが生じる $d = 2, 3$ の場合についても MI に比べてバイアスが少ないことがわかる. とりわけ $d = 2$ の場合では、BCMI は他の手法と比較して最も小さいバイアスであった. BCMI と CC を比較すると、 $d = 1, 2$ のときは BCMI のほうがバイアスが小さく、 $d = 3$ の場合、 β_1 については CC のほうがわずかにバイアスが小

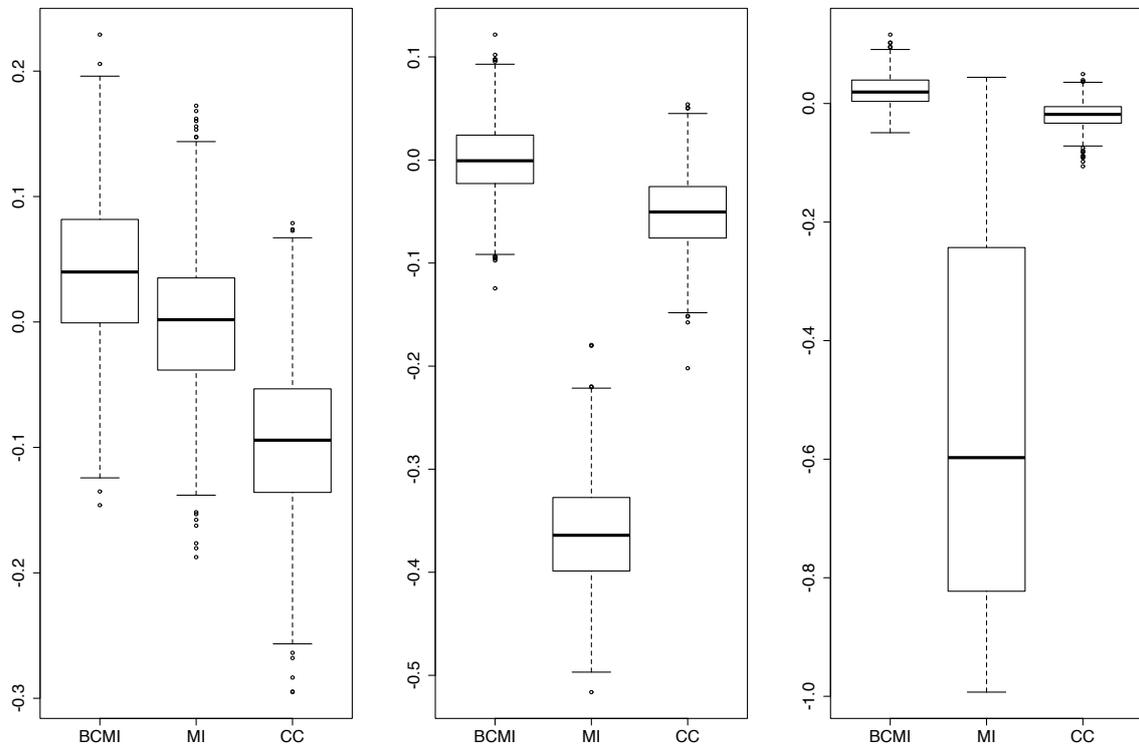


図 5.1 3 手法による $\hat{\beta}_1 - \beta_1$ の比較. 左より $d = 1, d = 2, d = 3$ の場合をそれぞれ示す.

さいが、ほぼ同程度であり、 β_2 については BCMI のほうが小さいバイアスとなった。

推定の安定性についても検討した。図 5.1 は $\hat{\beta}_1 - \beta_1$ について、図 5.2 は $\hat{\beta}_2 - \beta_2$ について、1000 回の数値実験からそれぞれ計算された値を箱ひげ図にしたものである。BCMI と CC については箱の大きさがほぼ同じであるといえる。しかしながら、バイアスについては全体的に BCMI のほうが 0 に近い。他方 MI については、代入モデルを正しく同定できない状況 ($d = 2, 3$) では、MI の箱の大きさは他の手法よりも大きく、推定として不安定である。特に $d = 3$ の場合で β_1 について比較すると MI は BCMI の約 16 倍の大きさになっている。以上より、BCMI は推定方法としてバイアスの意味でも推定の安定性でもよいことが示唆された。

5.3 実データ分析

本節では、BCMI を用いて実際のデータを解析した結果を示す。用いたデータは Vanderbilt Biostatistics Wiki*¹ に掲載されている、Duke Cardiac Catheterization Coronary Artery Disease Diagnostic Dataset である。このデータセットは Duke 大学が動脈性心疾患の患者に

*¹ URL: <http://biostat.mc.vanderbilt.edu/DataSets>

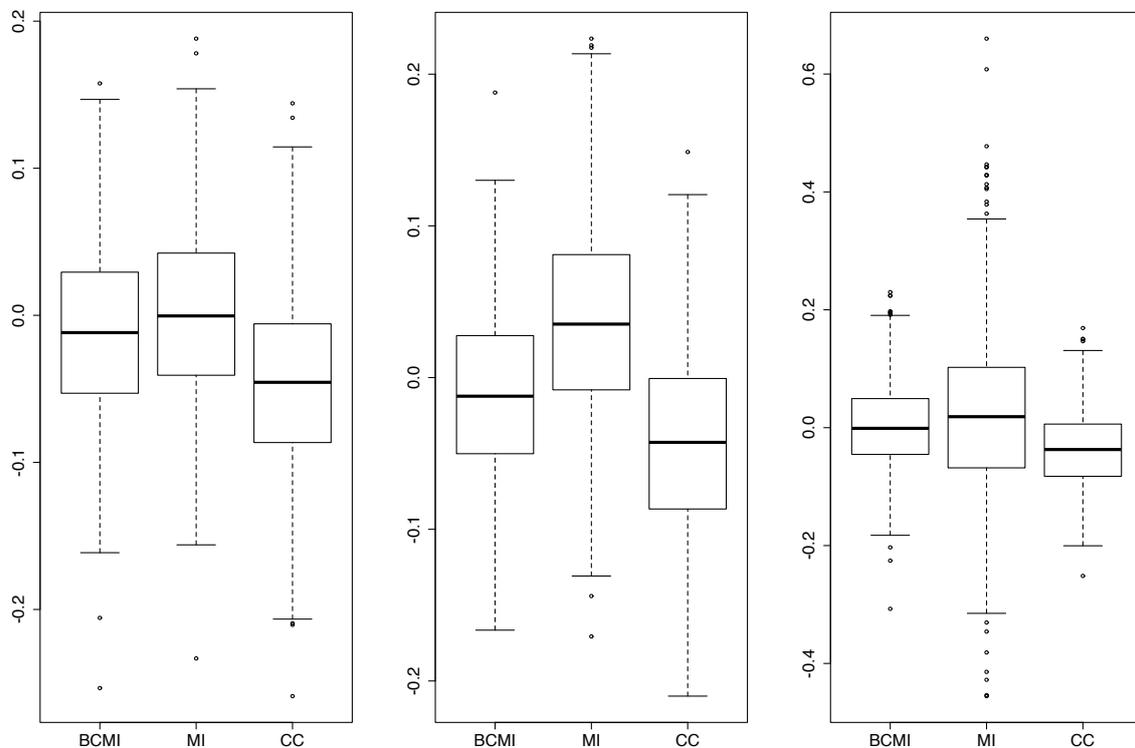


図 5.2 3 手法による $\hat{\beta}_2 - \beta_2$ の比較. 左より $d = 1$, $d = 2$, $d = 3$ の場合をそれぞれ示す.

対して心臓カテーテル法という検査を行った結果についてまとめたものである. 本データセットは, 性別 (sex), 年齢 (age), 心疾患が発生している期間 (cad.dur), コレステロールのレベル (cholesterol) という 4 つの変数と, 心臓カテーテル法によって分類された症状の重さを表す 2 値変数 (sigdz) から構成されている. 標本数は $n = 3504$ である. 欠測が生じている変数はコレステロールのレベルだけであり, 他の変数はすべて観測がなされている. 欠測が生じた標本数は 1246 個である.

5.3.1 分析方法

本研究では性別, 年齢, 症状が発生している期間, コレステロールの 4 変数を説明変数とし, 症状の重さを目的変数とする回帰モデルの推定を行った. 推定にあたっては, はじめに男女で層別を行った. 続いて, すべての連続変数について正規化処理を行った. 推定する回帰モデルは

$$\text{logit} \{ \Pr(\text{sigdz} = 1) \} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{cholesterol} + \beta_3 \log(1 + \text{cad.dur}) \quad (5.5)$$

とした．欠測したコレステロールレベルの値を代入するモデルは，

$$\text{cholesterol} = \alpha_0 + \alpha_1 \text{ age} + \alpha_2 \log(1 + \text{cad.dur}) + \varepsilon, \quad \varepsilon \sim N(0, \tau^2). \quad (5.6)$$

とした．式 (5.6) におけるパラメータ α_i ($i = 0, 1, 2$), τ^2 は完全に観測されたデータを用いて推定する．特に今回は sigdz の値に応じて独立に最尤推定を行い，推定値をそれぞれ， $(\hat{\alpha}_{0,1}, \dots, \hat{\alpha}_{2,1}, \hat{\tau}_1^2)$ ($\hat{\alpha}_{0,0}, \dots, \hat{\alpha}_{2,0}, \hat{\tau}_0^2$) とした．従って，sigdz = 1 であり，コレステロールが欠測している標本については，

$$\text{cholesterol} = \hat{\alpha}_{0,1} + \hat{\alpha}_{1,1} \text{ age} + \hat{\alpha}_{2,1} \log(1 + \text{cad.dur}) + \varepsilon, \quad \varepsilon \sim N(0, \hat{\tau}_1^2)$$

にもとづいて補完値を生成し，sigdz = 0 であり，コレステロールが欠測している標本に対しては，

$$\text{cholesterol} = \hat{\alpha}_{0,0} + \hat{\alpha}_{1,0} \text{ age} + \hat{\alpha}_{2,0} \log(1 + \text{cad.dur}) + \varepsilon, \quad \varepsilon \sim N(0, \hat{\tau}_0^2)$$

にもとづいて補完値を生成する．今回の解析では $M = 100$ 個の補完値を独立に生成し，100 個の擬似完全データセットを作成した．式 (5.5) で定めたロジスティック回帰モデルのパラメータ $\beta_0, \beta_1, \beta_2, \beta_3$ について，BCMI の定義式 (4.4) に従って重みつき最尤推定を行った．

BCMI のパラメータは $B = 100$ とした．各ガウスカーネルのパラメータ σ については男性のデータでは， $[0.717, 0.71725, 0.71750, 0.7175, 0.718, 0.71825, 0.7185, 0.71875, 0.719, 0.7195, 0.5920, 0.5921, 0.5923, 0.5924, 0.5925, 0.5926, 0.5927, 0.5928, 0.5929, 0.5930]$ を，女性は， $[0.8846, 0.8847, 0.8848, 0.8849, 0.885, 0.8851, 0.8852, 0.8853, 0.8854, 0.8855, 0.6155, 0.6156, 0.6157, 0.6158, 0.6159, 0.6160, 0.6161, 0.6162, 0.6163]$ をそれぞれ候補点とした． δ は区間 $[0, 10]$ から等間隔に 100 点発生させたものを候補とし， σ と δ は 5 分割交差検証法で選んだ．

比較に用いた既存手法での解析手法を簡単に述べると次の通りである．完全データ解析 (CC) では，欠測がないデータのみを用いて $\beta_0, \beta_1, \beta_2, \beta_3$ を最尤推定した．通常の多重代入法 (MI) では，100 個の擬似完全データセットそれぞれにおいて， $\beta_0, \beta_1, \beta_2, \beta_3$ を最尤推定し，その後 Rubin のルールに従って統合推定量を求めた．

今回の推定では 95% 信頼区間をブートストラップに基づくパーセンタイル法を用いて近似的に算出した．計算手法を以下に述べる．はじめに，欠測値を含む標本の割合を固定した状態で， $B = 1000$ 回再サンプリングを行う．すなわち，欠測値を含む標本と含まない標本に分け，それぞれ独立に再サンプリングを実行した．これによって， B 個の推定値 $\hat{\beta}_j^b$ ($j = 0, 1, 2, 3$; $b = 1, \dots, B$) が得られる．続いて各 j 毎に B 個の推定値を昇順に並び替える．すなわち， j 毎に $\hat{\beta}_j^1 \leq \dots \leq \hat{\beta}_j^{1000}$ という関係が成立するように並び替えを行う．パーセンタイル法では β_j の近似的な 95% 信頼区間は，

$$\left[\hat{\beta}_j^{26}, \hat{\beta}_j^{975} \right]$$

として与えられる (下平 (2012))．この操作を BCMI, MI, CC それぞれに対して行うことで，手法ごとに近似的な 95% 信頼区間を得ることができる．

表 5.2 ロジスティック回帰で得られた推定値と 95% 信頼区間 (男性 $n = 2405$, 欠測割合:34.8%)

	BCMI		MI		CC	
	推定値	95%CI	推定値	95%CI	推定値	95%CI
const	1.534	(1.466, 1.623)	1.489	(1.433, 1.562)	1.550	(1.475, 1.648)
age	0.850	(0.727, 0.979)	0.845	(0.731, 0.966)	0.824	(0.682, 0.980)
cholesterol	0.583	(0.421, 0.757)	0.464	(0.337, 0.608)	0.499	(0.367, 0.645)
cad.dur	0.064	(-0.051, 0.191)	0.085	(-0.017, 0.205)	0.057	(-0.092, 0.223)

表 5.3 ロジスティック回帰で得られた推定値と 95% 信頼区間 (女性 $n = 1099$, 欠測割合:37.9%)

	BCMI		MI		CC	
	推定値	95%CI	推定値	95%CI	推定値	95%CI
const	-0.364	(-0.416, -0.272)	-0.376	(-0.435, -0.327)	-0.375	(-0.435, -0.321)
age	0.607	(0.445, 0.794)	0.651	(0.503, 0.814)	0.483	(0.287, 0.695)
cholesterol	0.541	(0.327, 0.785)	0.397	(0.239, 0.587)	0.466	(0.305, 0.661)
cad.dur	-0.133	(-0.256, 0.013)	-0.072	(-0.196, 0.062)	-0.149	(-0.319, 0.030)

5.3.2 結果と考察

各手法によって計算された回帰係数の推定値と近似的な 95% 信頼区間をまとめると表 5.2, 5.3 となる.

得られた推定値について考察を行う. はじめに、女性における心疾患が発生している期間 (cad.dur) に関して検討すると、通常の MI は完全ケース解析で得られる推定値の約半分となっている。このとき BCMI では完全ケース解析で得られた推定量とほぼ同じ推定量が得られている。本解析においては完全ケース解析が必ずしも正しい推定量を与えているとは限らず、信頼区間を用いて検討すると BCMI, CC ともに cad.dur については有意ではないが、BCMI は MI の推定量を適切に補正しているといえる。

信頼区間の幅について検討を行うと、BCMI によって与えられる信頼区間は MI や CC とほぼ同程度の幅である。このことから、ノンパラメトリック法による条件付き密度推定を用いた BCMI でも実用に足る信頼区間を構成できているといえる。

男性の定数項および年齢に対する回帰係数の推定値は推定方法によらず女性の推定値と大きく異なっている。とりわけ定数項については正負が逆転していることが特徴的である。このことから、本解析で行った性別に基づく層別化には意味があるといえる。

最後にこのデータセットを用いて過去に行われたロジスティック回帰モデルを用いた解析 Harrell (2001); Jin and Wang (2010) との比較を行う。過去の研究ではいずれも sigdz を予測するモデルについて年齢とコレステロールのレベルの 2 変数のみを説明変数として用いてい

た. このモデルは今回の解析結果によれば, 男性のデータでは十分なモデルと考えられるが, 女性のデータについては BCMI や CC における β_3 の変数が十分小さくないため十分なモデルではないと考えることができる.

第 6 章

多重代入法を利用した予測手法

本章では、回帰分析の目的変数が欠測した場合について多重代入法を利用した予測手法を検討する。なお、目的変数以外の変数はすべて観測されるものとし、欠測は MAR を仮定する。この状況下では欠測値を補完しないで行う完全ケース解析でも、観測変数をすべて説明変数として用いる回帰分析では一致性をもつことが知られている。Little (1992); von Hippel (2007) にあるように、 $Y|X$ の回帰モデルを同定する目的において、 Y の補完を行ったとしても、新たに得られる情報は特に存在せず、したがって積極的に補完することへの意味づけが難しいという問題がある。

一方で、機械学習の理論によれば、 n 個の学習データ $(x_1, y_1), \dots, (x_n, y_n)$ よって得られた $Y|X$ のモデルにおける期待予測二乗誤差は、定数項を除くと $O(1/n)$ で抑えられることが知られている (Hastie et al. (2009))。したがって、特に欠測の割合が大きい場合や標本の数が小さい場合には、欠測値に対して補完を行うことによって利用可能な標本の数を増やすことで、予測誤差を小さくすることができるかもしれない。回帰分析においては、予測モデルの構築が目的になることも多いため、予測に注目して検討することは有用と考えられる。本章では、予測の観点から目的変数が欠測したときに多重代入法を利用した予測手法を検討した結果を述べる。特に第 4 章で提示した BCMI を利用した上で、完全ケース解析による一致性を活用した、新たな予測手法を提案する (富田他 (2017))。この手法は、代入の誤り具合に応じて適切な解析方法をバランスよく選ぶものである。

本章での設定は以下の通りである。 Y を目的変数とし、 X を説明変数とする。 Y は欠測が発生しうる変数であり、 X には欠測が発生しない。回帰モデルを $f(y|x; \theta)$ とし、 θ の推定を行うことを考える。欠測した Y に対して、観測された変数 X から補完を行う代入モデルは $h(y|x; \phi)$ と表す。 ϕ は完全に観測されたデータに基づく最尤推定値 $\hat{\phi}$ を利用する。6.1 節において、前章で提案した BCMI を用いた θ の推定量 $\hat{\theta}$ について本章で利用する形で書き直す。6.2 節では目的変数が MAR に基づいて欠測する場合に成り立つ完全ケース解析の一致性を利用した新しい予測方法を提示する。6.3 節で提案手法の手順についてまとめ、6.4 節で数値実

験を行った結果を述べる.

6.1 BCMI による推定量

BCMI による推定量 $\hat{\theta}$ は, 前章までの記法を利用すると

$$\hat{\theta} = \arg \max_{\theta} U_{M,n}(\theta) \quad (6.1)$$

$$U_{M,n}(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(Y_i|X_i; \theta) + \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n w(X_i, Y_i^m, \hat{\phi}) \log f(Y_i^m|X_i; \theta) \right\}$$

と表される. ただし, $w(x, y, \phi)$ は重み関数で

$$w(x, y, \phi) = \frac{g(y|x)}{h(y|x; \phi)}$$

と表される.

6.2 交差検証を利用した新たな予測法

本節では多重代入法を利用した新しい予測法について議論する. 先ほど述べたように, 回帰分析において目的変数のみが欠測する場合には, 完全ケース解析による推定量に一致性が存在することが知られている. よって, 完全ケースを基準として用いた交差検証法を利用することで新しい推定量を提案することができる. 具体的な手法は以下の通りである. 前節で示した推定量 (6.1) にパラメータ α, λ , ($0 \leq \alpha \leq 1, 0 \leq \lambda \leq 1$) を導入して,

$$\hat{\theta} = \arg \max_{\theta} \tilde{U}_{M,n,\lambda,\alpha}(\theta) \quad (6.2)$$

$$\tilde{U}_{M,n,\lambda,\alpha}(\theta) = \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \log f(Y_i|X_i; \theta) + \lambda \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n w(X_i, Y_i^m; \hat{\phi})^\alpha \log f(Y_i^m|X_i; \theta) \right\}$$

とする. ここでパラメータ α は条件付き密度推定を利用して求めた重みの重視度合い, λ は補完したデータの重視度合いと解釈することができる. すなわち, さて, パラメータ α, λ の選択にあたっては, 次の交差検証法を用いる. まず, 完全に観測されたデータ $(X_1, Y_1), \dots, (X_{n_1}, Y_{n_1})$ を 5 分割し, 各分割に属する標本の添え字集合を S_k ($k = 1, \dots, 5$) と表す. また, 添え字集合 $\{1, \dots, n_1\} \setminus S_k$ を S_{-k} と定義する. 推定量 $\hat{\theta}_{k,\lambda,\alpha}$ を

$$\hat{\theta}_{k,\lambda,\alpha} = \arg \max_{\theta} \tilde{U}_{M,n,k}(\theta, \lambda, \alpha)$$

$$\tilde{U}_{M,n,k}(\theta, \lambda, \alpha) = \frac{1}{n} \left\{ \sum_{i \in S_{-k}} \log f(Y_i|X_i; \theta) + \lambda \frac{1}{M} \sum_{m=1}^M \sum_{i=n_1+1}^n w(X_i, Y_i^m; \hat{\phi})^\alpha \log f(Y_i^m|X_i; \theta) \right\}$$

によって求める.

各 $\hat{\theta}_{k,\lambda,\alpha}$ を利用して, 添え字集合 S_k に対応する標本 $\{(X_i, Y_i) | i \in S_k\}$ に関して負の対数尤度

$$e(k, \lambda, \alpha) = \frac{1}{|S_k|} \sum_{i \in S_k} -\log f(Y_i | X_i; \hat{\theta}_{k,\lambda,\alpha}) \quad (6.3)$$

を計算する. $e(k, \lambda, \alpha)$ を k に関して和を取った $e(\lambda, \alpha) = \sum_{k=1}^5 e(k, \lambda, \alpha)$ を最小とするような (λ, α) を選択する. なお, $\lambda = 0$ となるときは α を不定とする. 以下本章では本手法のことを提案手法と呼ぶ.

6.3 提案手法の手順

本節では提案手法の手順についてまとめて示す.

Step1. M 個の独立な擬似完全データセットを代入モデル $h(y|x, \hat{\phi})$ に基づいて生成する.

Step2. 完全に観測されたデータから条件付き密度推定を行い $\hat{g}(y|x)$ を求める.

Step3. 代入した標本 (X_i, Y_i^m) について, 重み $\hat{w}(X_i, Y_i^m; \hat{\psi})$ を

$$\hat{w}(X_i, Y_i^m; \hat{\phi}) = \frac{\hat{g}(Y_i^m | X_i)}{h(Y_i^m | X_i; \hat{\phi})}$$

で計算する. 以後の推定における計算を安定させるために, 代入モデルの密度 $h(Y_i^m | X_i; \hat{\phi})$ が極めて小さい値については $\hat{w}(X_i, Y_i^m; \hat{\phi}) = 0$ とした. 今回はその閾値を $h(y|x; \hat{\phi})$ の両側 80% 点と定めた.

Step4. λ, α を交差検証法で決定する.

Step5. 式 (6.2) に基づいて重み付き最尤推定法を行い, $\hat{\theta}$ を計算する.

6.4 数値実験

提案手法の性能を数値実験で検証した. 比較する既存手法は通常 of 多重代入法 (MI) と完全ケース解析 (CC) である. また, BCMI に追加した交差検証の挙動を検証するために, BCMI とも比較した. ただし, ここでの多重代入法は前章同様事前分布を仮定しない, 頻度論的な多重代入法である.

6.4.1 方法

$X = (X_1, X_2)$ を説明変数とし, X_1, X_2 を区間 $[-1, 1]$ 上の一様分布からそれぞれ独立に $n = 1000$ 個発生させた. 目的変数 Y は X_1 の 3 次多項式に X_2 の項を加えたモデル

$$Y = X_1 + dX_1^3 + X_2 + \epsilon$$

によって生成した. ただし, ϵ は正規分布 $N(0, 0.3^2)$ に従う乱数である. 今回は d として $d = 0, 0.5$ を考察した. 欠測メカニズムは, Y の観測指標 R について

$$\Pr(R = 1 | X_1 = x_1, X_2 = x_2) = \frac{0.5}{1 + \exp(-5x_1^2)}$$

で定める. ただし, 右辺の値が 1 以上となる場合は $R = 1$ とする. Y が観測される割合はおよそ 40% である. 観測確率は $X_1 = 0$ において最小となる.

LSCDE によって条件付き確率密度の推定量 $\hat{g}(x|y, z)$ を求める際に用いたハイパーパラメータ σ_x, σ_y の候補は 0.05 から 0.5 まで 0.025 刻みでとり, δ の候補は $[10^{-3}, 10]$ の間から対数スケールで等間隔に選んだ 9 点である.

欠測した標本 Y_i を X_{1i}, X_{2i} から補完する代入モデルは, 線形モデル

$$Y^* = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \epsilon'$$

とした. ただし, ϵ' は正規分布 $N(0, \tau'^2)$ に従う乱数である. 補完を行う際には各パラメータ $\alpha_j (j = 0, \dots, 2)$ と τ'^2 には, 完全に観測された標本を用いて最尤推定することで得られた推定値を利用した.

補完後に解析するメカニズムは

$$Y = m(X_1, X_2; \beta) + \epsilon,$$

$$m(X_1, X_2; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_1^4 + \beta_5 X_2,$$

とした. ただし, ϵ は正規分布 $N(0, \tau^2)$ に従うものとする.

このとき, 式 (6.2) に基づく重み付き尤度は

$$\begin{aligned} \tilde{U}_{M,n,\lambda,\alpha}(\beta, \tau^2) = & \frac{1}{n} \left[-\frac{n}{2} \log(2\pi\tau^2) + \sum_{i=1}^{n_1} \left\{ -\frac{(Y_i - m(X_{1i}, X_{2i}; \beta))^2}{2\tau^2} \right\} \right. \\ & \left. + \lambda \frac{1}{M} \sum_{i=n_1+1}^n w(X_{1i}, X_{2i}, Y_i^m, \hat{\phi})^\alpha \left\{ -\frac{(Y_i^m - m(X_{1i}, X_{2i}; \beta))^2}{2\tau^2} \right\} \right] \end{aligned}$$

と表される. これは β の 2 次形式で表され, 最小とする $\hat{\beta}$ は

$$\hat{\beta} = (A^T A)^{-1} A^T b$$

である。ただし、

$$A = \begin{pmatrix} \sqrt{M} & \sqrt{M}X_{11} & \sqrt{M}X_{11}^3 & \sqrt{M}X_{21} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{M} & \sqrt{M}X_{1n_1} & \sqrt{M}X_{1n_1}^3 & \sqrt{M}X_{2n_1} \\ \sqrt{\lambda w^\alpha} & \sqrt{\lambda w^\alpha}X_{1n_1+1} & \sqrt{\lambda w^\alpha}X_{1n_1+1}^3 & \sqrt{\lambda w^\alpha}X_{2n_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\lambda w^\alpha} & \sqrt{\lambda w^\alpha}X_{1n} & \sqrt{\lambda w^\alpha}X_{1n}^3 & \sqrt{\lambda w^\alpha}X_{2n} \end{pmatrix},$$

$$b = \begin{pmatrix} \sqrt{M}Y_1 \\ \vdots \\ \sqrt{M}Y_{n_1} \\ \sqrt{\lambda w^\alpha}Y_{n_1+1}^* \\ \vdots \\ \sqrt{\lambda w^\alpha}Y_n^* \end{pmatrix},$$

である。

提案手法では、重み w に関する交差検証を行った。この交差検証では回帰パラメータ β_0, \dots, β_5 に関してのみ行い、 λ, α は 0, 0.25, 0.5, 0.75, 1.0 の値から選択した。なお、このとき ε は正規分布に従うことから、式 (6.3) によって定義される $e(k, \lambda, \alpha)$ はテストデータに対する二乗誤差平均

$$\text{err}(k, \lambda, \alpha) = \frac{1}{|S_k|} \sum_{i \in S_k} \|Y_i - m(X_1, X_2; \hat{\beta}_{\lambda, \alpha})\|^2$$

と等価であることに注意する。本実験では二乗誤差平均を最小にする λ, α を交差検証の結果として採用した。

本論文では回帰関数そのものをターゲットとしたときに利用される二乗誤差

$$\mathbb{E} \left[\left\{ (X_1 + dX_1^3 + X_2) - m(X_1, X_2; \hat{\beta}) \right\}^2 \right] \quad (6.4)$$

を考えた。これは予測二乗誤差 (PSE)

$$\mathbb{E} \left[\left\{ Y - m(X_1, X_2; \hat{\beta}) \right\}^2 \right]$$

から誤差項の分散を差し引いた誤差でもある。ここではシミュレーションデータとは独立に $N = 50000$ 個の標本を生成して式 (6.4) の近似を行って評価した。

6.4.2 結果

推定の結果を表 6.1 に示す。代入モデルを正しく同定できる場合 ($d = 0$)、MI による予測が誤差を最小にする。提案手法は MI にわずかに及ばないものの、CC に比べると精度よく予

表 6.1 $\{(X_1 + dX_1^3 + X_2) - m(X_1, X_2; \hat{\beta})\}^2$ の結果

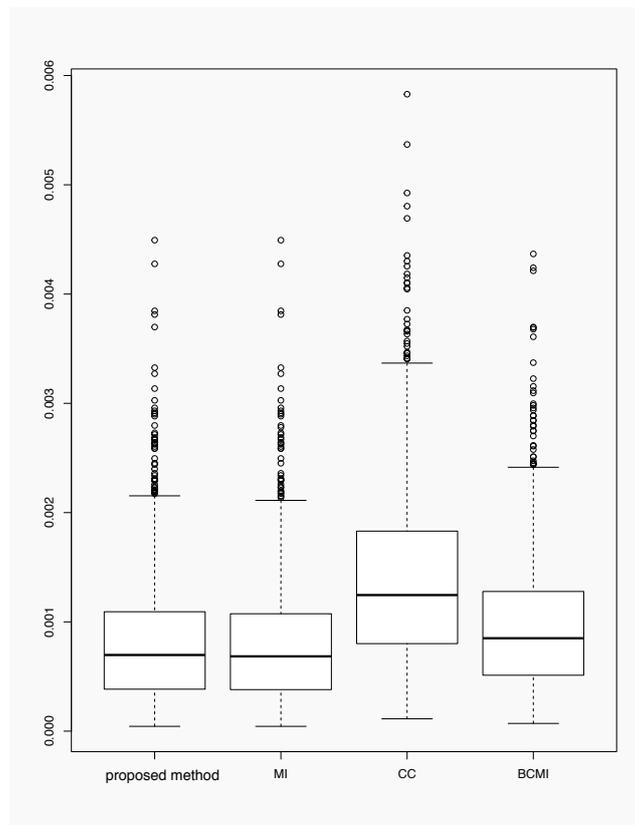
	$d = 0$				$d = 0.5$			
	提案手法	MI	CC	BCMI	提案手法	MI	CC	BCMI
min	0.00004	0.00004	0.00011	0.00007	0.00008	0.00088	0.00009	0.00019
1st-Qu.	0.00038	0.00038	0.00080	0.00051	0.00091	0.00230	0.00081	0.00121
median	0.00070	0.00068	0.00125	0.00085	0.00145	0.00280	0.00128	0.00174
mean	0.00084	0.00082	0.00141	0.00099	0.00166	0.00292	0.00144	0.00187
3rd-Qu.	0.00109	0.00107	0.00183	0.00128	0.00217	0.00346	0.00186	0.00244
max	0.00449	0.00449	0.00583	0.00437	0.00620	0.00834	0.00593	0.00721

測ができることがわかる。代入モデルの特定を誤った場合 ($d = 0.5$)、MI による予測の誤差は他の手法に比べて大きくなる。提案手法は重み付け最尤推定によって予測誤差を小さくしている。提案手法は CC に比べて予測誤差がわずかに大きくなっている。

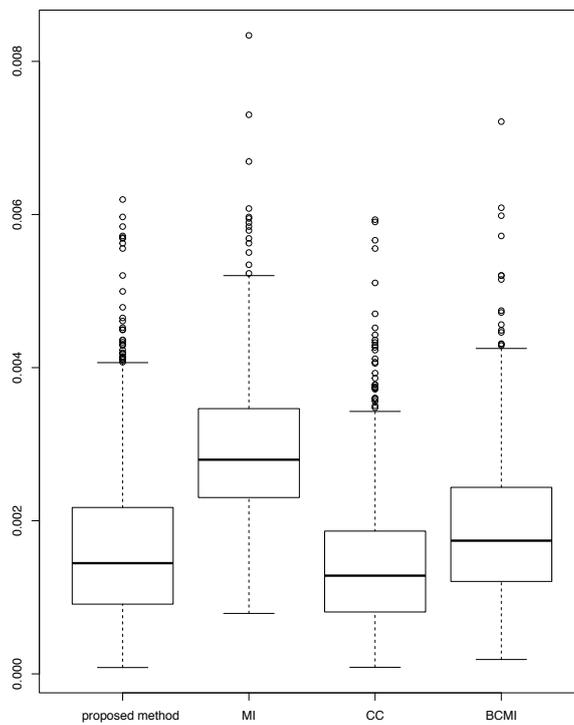
続いて、重み w に関する交差検証によって選択されたチューニングパラメータ λ, α の値の頻度を表 6.2 に示す。代入モデルの同定が正しく行われているとき、95% 以上の頻度で $(\lambda, \alpha) = (1, 0)$ が選択される。すなわち、ほとんどの場合で提案手法は最尤法に基づく MI を行っていることがわかる。これは代入モデルの同定が正しく行われている状況では妥当な選択であるといえる。代入モデルの同定を誤っている時には、最尤法に基づく MI が採択される割合は減少する。推定された重みに関するチューニングパラメータ α は 1 となるものは多いので、BCMI によって推定された重みを活かす推定が採択されることになる。他方、欠測データの重要度と解釈されるチューニングパラメータ λ については、完全ケース解析に近い $\lambda = 0.25, 0.5$ の採択率が上昇する。

以上の結果より、この交差検証は、代入が正しい時には最尤法に基づく通常の MI を行い、代入の誤りが大きくなると重みを利用したり、代入を行っていない完全ケースのデータを重要視することで、式 (6.4) によって定義される最終的な予測誤差が小さくなるようにチューニングパラメータを選択する傾向であることがわかる。すなわち、多重代入法を利用した上で、完全ケース解析などの情報も利用しながら、回帰関数の予測誤差を小さくするような予測手法であるといえる。

また、 N 回のシミュレーション結果を箱ひげ図を用いて示したものが図 6.1 である。代入モデルの誤特定が起こらない場合は、MI の箱は他の手法に比べてわずかに小さくなる。このとき、提案手法は密度推定に基づく重み付けを行っているため、箱のサイズは MI に比べて大きい、CC よりははるかに小さい。代入モデルを誤特定した場合には、どの手法でも箱のサイ



(a) $d = 0$



(b) $d = 0.5$

図 6.1 $\{(X_1 + dX_1^3 + X_2) - m(X_1, X_2; \hat{\beta})\}^2$ の箱ひげ図

表 6.2 λ, α の選択結果

(a) $d = 0$							(b) $d = 0.5$						
α	λ					合計	α	λ					合計
	0	0.25	0.5	0.75	1			0	0.25	0.5	0.75	1	
不定	0	-	-	-	-	0	不定	17	-	-	-	-	17
0	-	0	0	2	965	967	0	-	47	18	12	39	116
0.25	-	0	2	0	3	5	0.25	-	50	17	6	10	83
0.5	-	0	0	0	2	2	0.5	-	21	4	1	3	29
0.75	-	0	0	0	4	4	0.75	-	21	9	3	5	38
1	-	0	1	1	20	22	1	-	292	229	119	77	717
合計	0	0	3	3	994	1000	合計	17	431	277	141	134	1000

ズはほぼ同じである。従って、提案手法は、既存手法と比べて予測のばらつきを大きく増加させることはないと言える。

第7章

結論

本論文では，多重代入法において代入時に必ずしも適切なモデルから補完値を生成できなかったとしても，推定量を求める際に補正して漸近的に一致性のある推定量を導出する方法を提案した．これは欠測が無視可能であるという仮定の下で，欠測した変数に関する条件付き分布について成立する性質を利用した重み付き最尤推定法である．すなわち，欠測変数が一変数 X で，他のすべて観測された変数が Z であるときに， $X|Z$ の条件付き分布は X が欠測するか否かによらず同じものになることを用いて，欠測を補完した標本に対応する尤度に，真の条件付き密度と補完に用いた代入モデルの密度の比として表される重みを乗じて最尤推定を行うものである．加えて，一般には $X|Z$ の真の条件付密度 $g(x|z)$ は未知であることから，条件付き密度推定法を用いて推定した密度 $\hat{g}(x|z)$ を用いて計算する方法（バイアス補正多重代入法：BCMI）を提示した．

本論文では BCMI を回帰分析の説明変数が欠測する場合の推測に適用した．数値シミュレーションと実データ解析によって，バイアス補正多重代入法が推定量の偏りを補正していることを示した．また多重代入法，BCMI を利用した回帰分析の目的変数の予測をする方法についても提案した．

代入モデルに対して条件を課さず，推定時に補正を行うことで推定量の偏りを取り除く手法は，代入モデルを簡易なものにすることで多数の補完値を発生させやすいことから多重代入法と親和性が高く，また一般に妥当な推定量を導出するために代入モデルに課される融和性の条件を検証することは難しいことから，本手法は有用性が高いといえる．

課題

本論文で提案した BCMI には未解決である課題が存在する．特に重要と考えるものを挙げる．第一の課題は，推定量の漸近分散を導出することがあげられる．多重代入法を用いる上での利点の一つとして，推定量の漸近分散 $\text{Var}[\hat{\theta}]$ を見積もることが可能である．適切な多重代

入法では、Rubin のルールによって推定量の漸近分散が、式 (3.2) として与えられている。しかしながら、BCMI は重みつき最尤推定法を用いて推定量を導出しているため、Rubin の提案した手法に従っていないため、漸近的な分散についても別途導出する必要がある。

第二の課題として、BCMI の重み w の閾値、条件付き密度推定におけるパラメータ σ, δ や第6章における新たな予測手法で導入したチューニングパラメータ α, λ の定め方である。まず、重み w については、分母である代入モデルの密度 h が非常に小さい値のとき、重み w が非常に大きくなるため、推定が不安定になる。本論文では重み w を安定させるため、 h の値が定められた閾値より小さいとき重みを 0 とした。この閾値を理論的に定める方法を検討する必要がある。また、条件付き密度推定におけるパラメータや最尤推定における重要度のチューニングパラメータをすべて交差検証法で求めている。交差検証ではチューニングパラメータの候補となる点をあらかじめ与える必要があるが、与え方については未知な事柄が多い。今回は均等に分割する形で候補点を与えたが、効率よく推定を行うためには、候補点の与え方に対する指針が必要となると考えられる。

第三の課題として、多変量の欠測問題への BCMI の拡張である。本論文で扱った BCMI は1変数のみが欠測する状況下のものであるが、現実のデータにおいて1変数のみが欠測することはまれである。また、欠測のパターンも複数存在する（すなわち、 R が2値変数ではない）ことが多い。よって、複数の欠測パターンが存在する場合への BCMI の拡張を考える必要があるといえる。

付録 A

式 (4.3) の証明

本章では，式 (4.3) について証明を行う．

Propostion 1. X, Z を確率変数とする． X のみ無視可能な欠測が生じうるとし，観測指標を R と表す．つまり， X の欠測メカニズムが MAR に従うとする．このとき， X について他の変数を条件つけた真の条件付き密度を $g(x|z, r)$ について，

$$g(x|z, R = 0) = g(x|z, R = 1)$$

となる．すなわち， X について Z を条件付けした密度は R によらない．

Proof. X は無視可能な欠測なので，

$$\Pr(R = 0|X, Z) = \Pr(R = 0|X, Z)$$

が成り立つ．ここでベイズの定理を用いると， $g(x|z, R = r)$, ($r = 0, 1$) は

$$\begin{aligned} g(x|z, R = r) &= \frac{g(x, R = r, z)}{g(R = r, z)} \\ &= \frac{g(x, z)g(R = r|x, z)}{g(z)g(R = r|z)} \end{aligned}$$

と展開できる．さて，無視可能性の定義を利用することで

$$\begin{aligned} g(R = 0|x, z) &= g(R = 0|z), \\ g(R = 1|x, z) &= 1 - g(R = 0|x, z) \\ &= 1 - g(R = 0|z) \\ &= g(R = 1|z) \end{aligned}$$

がいえるので,

$$\begin{aligned}g(x|z, R = r) &= \frac{g(x, z)g(R = r|z)}{g(z)g(R = r|z)} \\ &= \frac{g(x, z)}{g(z)} \\ &= g(x|z)\end{aligned}$$

となる。したがって,

$$\begin{aligned}g(x|z, R = 0) &= g(x|z, R = 1) \\ &= g(x|z)\end{aligned}$$

がいえた。

□

付録 B

ベイズ統計学の概略

多重代入法はベイズ統計学の理論に基づく展開がなされているため、ベイズ推定における用語の定義を述べる。

B.1 ベイズ推定

確率変数を Y ，パラメータを θ とする。頻度論と異なり，パラメータ θ とデータ Y の両方が確率変数と考え， $f(Y; \theta)$ をデータとパラメータの同時密度関数とする。このとき，条件付き分布の定義やベイズの定理を用いると，

$$\begin{aligned} f(\theta|Y) &= \frac{f(\theta : Y)}{f(Y)} \\ &= \frac{f(\theta : Y) f(\theta)}{f(\theta) f(Y)} \\ &= \frac{f(Y|\theta)f(\theta)}{f(Y)} \\ &= \frac{f(Y|\theta)f(\theta)}{\int f(Y|\theta)f(\theta)d\theta} \end{aligned}$$

などと書けることを利用する。

ベイズ推定では $f(Y; \theta)$ をパラメータ θ とデータ Y の両方が確率変数であり，同時分布として考える，このとき，確率 $p(\theta)$ を事前分布とすると，尤度 $L(Y|\theta) = f(Y|\theta)$ を用いて，事後分布 $p(\theta|Y)$ は

$$p(\theta|Y) = \alpha p(\theta)L(Y|\theta)$$

ただし， α は正規化の定数であり，陽にかくと $\int_{\theta} p(\theta)L(Y|\theta)d\theta$ である。

事後分布からパラメータ θ を推定するとき、代表的な推定方法は 3 つある。1 つは事後期待値平均 (expected a posteriori : EAP) であり、

$$\hat{\theta}_{EAP} = E[\theta|\mathbf{y}] = \int_{\theta} \theta p(\theta|\mathbf{y}) d\theta$$

として求められる。以下においても断りがなければ、パラメータの点推定としては $\hat{\theta}_{EAP}$ を用いる。他の方法は事後確率最大化 (maximized a posteriori : MAP) であり、

$$\hat{\theta}_{MAP} = \arg \max_{\theta} p(\theta|\mathbf{y})$$

として求める方法と、事後中央値 (posterior median : MED) で、

$$\hat{\theta}_{MED} = \arg \text{solve} \int_{\theta}^{\hat{\theta}_{MED}} p(\theta|\mathbf{y}) d\theta = \frac{1}{2}$$

とする方法がある。

また、点推定量の事後分散は

$$\begin{aligned} V[\theta|\mathbf{y}] &= E\left[\left(\theta - \hat{\theta}_{EAP}\right)^2 \mid \mathbf{y}\right] \\ &= \int_{\theta} \left(\theta - \hat{\theta}_{EAP}\right)^2 p(\theta|\mathbf{y}) d\theta \end{aligned}$$

として求められる。

さらに、事後予測分布 (posterior predictive distribution) という分布を次のように定義する。 $f(Y^*|\theta)$ をモデルの分布とし、 $p(\theta|\mathbf{y})$ を事後分布として、事後予測分布 $p(Y^*|\mathbf{y})$ を

$$p(Y^*|\mathbf{y}) := \int_{\theta} f(Y^*|\theta) p(\theta|\mathbf{y}) d\theta$$

と定義する。なお、正確に書くならば、

$$p(Y^*|\mathbf{y}) := \int_{\theta} f(Y^*|\mathbf{y}, \theta) p(\theta|\mathbf{y}) d\theta$$

だが、 Y, Y^* は θ が与えられたもとで独立、すなわち、 $f(Y^*, Y|\theta) = f(Y^*|\theta) f(Y|\theta)$ であるから

$$\begin{aligned} f(Y^*|\mathbf{y}, \theta) &= \frac{f(Y^*, Y|\theta)}{f(Y|\theta)} \\ &= \frac{f(Y^*|\theta) f(Y|\theta)}{f(Y|\theta)} \\ &= f(Y^*|\theta) \end{aligned}$$

と整理できる。

B.1.1 事前分布の種類

ベイズ推定では事後分布が事前分布と尤度によって決定されるため、最尤法と異なり事前分布の選び方によって推定結果が変化する。ここでは、簡単に事前分布についてまとめる。比較的良く使われるものとして、無情報事前分布 (noninformative prior) や共役事前分布 (conjunctive prior) がある。無情報事前分布は、事後分布が事前分布からの影響を与えないような分布である。事前分布として局所的に一様なものを用いることが多い。無情報事前分布としては例えば Jeffery's prior と呼ばれる事前分布が存在する。これは事前分布 $p(\theta)$ が尤度関数のフィッシャー情報量 $I(\theta)$ との間に

$$p(\theta) \propto \sqrt{I(\theta)}$$

なる関係がある事前分布である。

共役事前分布は、事後分布と事前分布が同じ分布となるような事前分布である。この事前分布を用いることによって、事後分布や事後予測分布を解析的に求めることができる。Schlaifer and Raiffa (1961) によって提案され、尤度関数の分布にしたがって共役事前分布が定義される。例えば、尤度関数がベルヌーイ分布であるならば、共役事前分布はベータ分布である。

Schafer and Olsen (1998) によれば、一般のデータ解析であれば無情報事前分布を用いれば、多くの場合問題なく解析できるとしている。

B.2 ベイズ統計の計算手法

多重代入法の実現においては、ベイズ統計学における計算手法を用いている。これらについて簡潔に述べる。

B.2.1 モンテカルロ積分

モンテカルロ積分とは、直接積分をするのが難しい関数に対して、数値計算を用いて積分の値を求める手法である。 X の分布を $f(x)$ とし、 $g(x)$ の期待値 $E_f[g(X)] = \int f(x)g(x)dx$ を求めることを考える。この積分を直接実行することは難しいが、 $f(x)$ から標本をサンプリングすることは可能であると仮定する。

このとき、 x_1, \dots, x_n を $f(x)$ から独立にとった標本とすれば、

$$\begin{aligned} E_f[g(X)] &= \int f(x)g(x)dx \\ &\approx \frac{1}{n} \sum_{j=1}^n g(x_j) \end{aligned}$$

として近似する．これをモンテカルロ積分という．

Example 7. 事後予測分布の計算

事後予測分布 $p(X^*|x)$ は,

$$\int f(X^*|x, \theta)p(\theta|x)d\theta = E[f(X^*|x, \theta)]$$

として定義される．よって，事後分布から θ の標本を多数サンプリングして，各標本 θ^i について $f(X^*|x, \theta^i)$ を求めて平均することで事後予測分布を得ることができる．

多重代入法では，パラメータの事後期待推定量をモンテカルロ積分の考え方をういて求めている．

B.2.2 ギブスサンプリング

ギブスサンプリング (Gibbs Sampling) (Geman and Geman (1984)) とはマルコフ連鎖モンテカルロ法を用いた乱数発生方法のひとつである．乱数の発生のほか，ある事後分布に従うパラメータのサンプリングにも利用することができる．

p 次元確率分布 $f(x_1, \dots, x_p)$ からの直接のサンプリングは難しいが，各 $i = 1, \dots, p$ について $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ によって条件つけた分布 $f(x_i|x_{-i})$ による乱数発生が行える状況を考える．このとき，次の手順で $f(x)$ に従う乱数を求めることができる．

Step 1

初期値 $x^0 = (x_1^0, \dots, x_p^0)$ を適当に定める．

Step 2

各 i について，逐次的に，

$$\begin{aligned} x_i^t &\sim p(x_i|x_{-i}^t), \\ x_{-i}^t &:= (x_1^t, \dots, x_{i-1}^t, x_{i+1}^{t-1}, \dots, x_p^{t-1}) \end{aligned}$$

として標本を更新する．

Step3

t が十分大きい数になるまで繰り返す．最終的に得られた標本を $f(x)$ からの標本とする．

パラメータについても同様で，パラメータ θ が p 次元である場合に， $p(\theta_i|\theta_{-i}, x)$ という条件付き事後分布が正確にわかるのであれば，ギブスサンプリングを用いて事後分布 $p(\theta|x)$ に従う θ を発生させることができる．

ギブスサンプリングが正当化できるためには、マルコフ連鎖による平衡分布が同時分布になる必要がある。このための条件は例えば Gilks et al. (1995) を参照のこと。

B.2.3 データ拡大法

データ拡大法、もしくはデータ増幅法 (Data Augmentation) とは Tanner and Wong (1987) が提案した手法で、ギブスサンプリングを用いて、欠測値の補完と補完を行った事後分布からの推定量のサンプリングを繰り返すことで、適切な事後期待推定量と事後予測分布に基づく補完値を導く手法である。

この手法は、I-Step, P-Step の 2 段階の繰り返しで構成される。それぞれのステップで、欠測値 X^m の補完とパラメータ θ の推定を行い、それらを反復する。すなわち、これは同時分布 $f(x, \theta)$ が陽に与えられていないときのギブスサンプリング法であるといえ、前節に記載した標本の発生とパラメータの発生を同時に行っているものとなっている。

この手法は、完全データの同時分布 X は多変量正規分布に従い、欠測メカニズムはベイズ推定の意味で無視可能なときに用いられる。完全データに対する分布の条件は、共役事前分布、事後分布や条件付き事後分布がすべて解析的に求められることがこの手法を用いる上で必要であることに由来する。

データ拡大法では次の 2 ステップを繰り返し行う。反復回数の添え字を t とする。

I-Step $X^{m(t+1)}$ を $f(X^m | X^o, \theta^{(t)})$ より導く。

P-Step $\theta^{(t+1)}$ を $p(\theta | X^o, X^{m(t+1)})$ より導く。

なお、初期推定量 $\theta^{(0)}$ は θ の事後分布に近い分布から得られたものとする。

これを十分繰り返したときに得られる標本 x^m を補完値とし、パラメータ θ を補完値に対応する事後推定量とする。

Example 8. 二次元正規分布におけるデータ拡大法 (Little & Rubin(2002) Ex10.3)

2 変数 Y_1, Y_2 は、二次元正規分布 $N_2(\mu, \Sigma)$ に従うとする。このうち欠測は Y_2 にのみ含まれるとする。 $Y_2 | Y_1$ は正規分布 $N(\beta_{20.1} + \beta_{21.1} Y_1, \sigma_{22.1})$ に従う。よって、I-Step では、その時点で求められているパラメータ $(\beta_{20.1}^{(t)}, \beta_{21.1}^{(t)}, \sigma_{22.1}^{(t)})$ を用いて、 Y_2 が欠測した i について独立に、

$$y_{i2}^{(t)} \sim N(\beta_{20.1}^{(t)} + \beta_{21.1}^{(t)} y_{i1}, \sigma_{22.1}^{(t)})$$

として欠測値を補完する。

P-Step では、データ $(Y_{obs}, Y_{mis}^{(t)})$ が実際に二次元正規分布 $N_2(\mu, \Sigma)$ に従って観測されたデータと考える。事前分布として無情報事前分布のひとつである Jeffrey's prior

$$p(\mu, \Sigma) \propto |\sigma|^{-\frac{3}{2}}$$

を用いると、パラメータ μ, Σ について事後分布をそれぞれ求めることができる。事後分布は逆ウィッシュャート分布や正規分布で書き表されることが知られている (Gelman et al. (2013))。

$$\begin{aligned}\Sigma|Y &\sim W^{-1}(S, n-1) \\ \mu|\Sigma, Y &\sim N_2(\bar{y}, \Sigma/n)\end{aligned}$$

ただし、 S はスケール行列 (Scale Matrix) とよばれる行列である。実際に $\mu^{(t)}, \Sigma^{(t)}$ を求める時には、 S^{-1} のコレスキー分解やカイ二乗分布などを用いて計算を行うことになる。パラメータ $(\beta_{20.1}^{(t)}, \beta_{21.1}^{(t)}, \sigma_{22.1}^{(t)})$ は $\mu^{(t)}, \Sigma^{(t)}$ より計算可能である。

謝辞

博士課程において、主任指導教員として支えてくださった統計数理研究所の逸見昌之准教授に御礼申し上げます。振り返るとご面倒なお願いとご心配ばかりおかけしてきましたが、暖かくご支援下さったことには感謝の言葉のほかございません。また、副指導教員および主査をお引き受けいただいた統計数理研究所の藤澤洋徳教授には、研究を進めるにあたって様々なお助言、ご指摘をいただいたほか、様々なお相談にのってくださいました。深く御礼申し上げます。学位審査においては統計数理研究所の野間久史准教授、大阪大学の服部聡教授、東京工業大学の金森敬文教授にはお忙しい中副査を引き受けていただき、多数の有益なお指摘をいただきました。感謝申し上げます。

この研究は様々な方のご支援の賜物と認識しております。総合研究大学院大学および統計数理研究所に在籍、在職した3年6ヶ月の間に、自由な研究環境を提供してくださった統計数理研究所の全ての教員、職員の皆様に感謝申し上げます。とくに最後の半年間お世話になりました統計思考院の皆様、および様々な事務手続きを行っていただきましたデータ系支援室および大学院担当の皆様には特に御礼申し上げます。また、総合研究大学院大学の学生の皆様とは、研究やそれ以外の由無き雑談を色々とさせていただきました。これらは学生生活を送る上での支えとなりました。さらには、総合研究大学院大学に入学する前には、研究生として籍をおかせていただき、また入学後も様々な形でご支援くださった東京大学の美馬秀樹准教授、増田勝也特任助教にも感謝申し上げます。

最後になりますが、常に暖かく励まし、見守ってくれた両親、家族に心から感謝をいたします。

参考文献

- Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., and for the Alzheimer's Disease Neuroimaging Initiative* (2015) "Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model," *Statistical Methods in Medical Research*, Vol. 24, No. 4, pp. 462-487, URL: <https://doi.org/10.1177/0962280214521348>, DOI: <http://dx.doi.org/10.1177/0962280214521348>, PMID: 24525487.
- Buck, S. F. (1960) "A Method of Estimation of Missing Values in Multivariate Data Suitable for use with an Electronic Computer," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 22, No. 2, pp. 302-306.
- van Buuren, S. (2012) *Flexible Imputation of Missing Data (Chapman & Hall/CRC Interdisciplinary Statistics)*: Chapman and Hall/CRC.
- van Buuren, S. and Groothuis-Oudshoorn, K. (2011) "mice: Multivariate imputation by chained equations in R," *Journal of statistical software*, Vol. 45, No. 3, pp. 1-67.
- van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C., and Rubin, D. B. (2006) "Fully conditional specification in multivariate imputation," *Journal of statistical computation and simulation*, Vol. 76, No. 12, pp. 1049-1064.
- Carpenter, J. and Kenward, M. (2013) *Multiple imputation and its application*: John Wiley & Sons.
- Cheng, P. (1994) "Nonparametric estimation of mean functional with data missing at random," *Journal of the American Statistical Association*, Vol. 89, No. 425, pp. 81-87.
- Cheng, P. E. (1984) "Strong consistency of nearest neighbor regression function estimators," *Journal of Multivariate Analysis*, Vol. 15, No. 1, pp. 63 - 72, URL: <http://www.sciencedirect.com/science/article/pii/0047259X84900678>, DOI: [http://dx.doi.org/https://doi.org/10.1016/0047-259X\(84\)90067-8](http://dx.doi.org/https://doi.org/10.1016/0047-259X(84)90067-8).
- Clayton, D., Spiegelhalter, D., Dunn, G., and Pickles, A. (1998) "Analysis of longitudinal binary data from multiphase sampling," *Journal of the Royal Statistical Society: Series*

- B (Statistical Methodology)*, Vol. 60, No. 1, pp. 71–87.
- Council, N. R. (2010) *The Prevention and Treatment of Missing Data in Clinical Trials*, Washington, DC: The National Academies Press, URL: <https://www.nap.edu/catalog/12955/the-prevention-and-treatment-of-missing-data-in-clinical-trials>, DOI: <http://dx.doi.org/10.17226/12955>.
- Deming, W. E. and Stephan, F. F. (1940) “On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known,” *The Annals of Mathematical Statistics*, Vol. 11, No. 4, pp. 427–444.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.
- Efron, B. (1994) “Missing data, imputation, and the bootstrap,” *Journal of the American Statistical Association*, Vol. 89, No. 426, pp. 463–475.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013) *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science: Taylor & Francis.
- Geman, S. and Geman, D. (1984) “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 6, No. 6, pp. 721–741.
- Gilks, W., Richardson, S., and Spiegelhalter, D. (1995) *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics: Taylor & Francis, URL: https://books.google.co.jp/books?id=TRXrMWY_i2IC.
- Graham, J. W., Olchowski, A. E., and Gilreath, T. D. (2007) “How many imputations are really needed? Some practical clarifications of multiple imputation theory,” *Prevention science*, Vol. 8, No. 3, pp. 206–213.
- Graham, J. (2012) *Missing Data: Analysis and Design*, Statistics for Social and Behavioral Sciences: Springer New York.
- Harrell, F. (2001) *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*: Springer.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, Springer Series in Statistics: Springer New York.
- von Hippel, P. T. (2007) “Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data,” *Sociological Methodology*, Vol. 37, pp. 83–117, URL:

- <http://www.jstor.org/stable/20451132>.
- Honaker, J. and King, G. (2010) “What to do about missing values in time-series cross-section data,” *American Journal of Political Science*, Vol. 54, No. 2, pp. 561–581.
- Honaker, J., King, G., and Blackwell, M. (2011) “Amelia II: A Program for Missing Data,” *Journal of Statistical Software, Articles*, Vol. 45, No. 7, pp. 1–47, URL: <https://www.jstatsoft.org/v045/i07>, DOI: <http://dx.doi.org/10.18637/jss.v045.i07>.
- 星野崇宏 (2009) 『調査観察データの統計科学 因果推論・選択バイアス・データ融合』, 確率と情報の科学, 岩波書店.
- Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., and Sterne, J. A. (2014) “Joint modelling rationale for chained equations,” *BMC Medical Research Methodology*, Vol. 14, No. 1, p. 28, Feb, URL: <https://doi.org/10.1186/1471-2288-14-28>, DOI: <http://dx.doi.org/10.1186/1471-2288-14-28>.
- Ireland, C. T. and Kullback, S. (1968) “Contingency Tables with Given Marginals,” *Biometrika*, Vol. 55, No. 1, pp. 179–188.
- Jin, L. and Wang, S. (2010) “A Model Validation Procedure when Covariate Data are Missing at Random,” *Scandinavian Journal of Statistics*, Vol. 37, No. 3, pp. 403–421.
- Kanamori, T., Hido, S., and Sugiyama, M. (2009) “A least-squares approach to direct importance estimation,” *Journal of Machine Learning Research*, Vol. 10, No. Jul, pp. 1391–1445.
- Little, R. J. A. (1988) “Missing-Data Adjustments in Large Surveys,” *Journal of Business & Economic Statistics*, Vol. 6, No. 3, pp. 287–296.
- Little, R. J. A. (1992) “Regression With Missing X’s: A Review,” *Journal of the American Statistical Association*, Vol. 87, No. 420, pp. 1227–1237.
- Little, R. J. A. and Rubin, D. B. (2002) *Statistical analysis with missing data*: John Wiley & Sons.
- Liu, J., Gelman, A., Hill, J., Su, Y.-S., and Kropko, J. (2014) “On the stationary distribution of iterative imputations,” *Biometrika*, Vol. 101, No. 1, pp. 155–173.
- Mackinnon, A. (2010) “The use and reporting of multiple imputation in medical research a review,” *Journal of Internal Medicine*, Vol. 268, No. 6, pp. 586–593.
- Meng, X.-L. (1994) “Multiple-Imputation Inferences with Uncongenial Sources of Input,” *Statistical Science*, Vol. 9, No. 4, pp. 538–558, 11.
- Meng, X.-L. and Rubin, D. B. (1992) “Performing likelihood ratio tests with multiply-imputed data sets,” *Biometrika*, Vol. 79, No. 1, pp. 103–111.
- Nadaraya, E. A. (1964) “On estimating regression,” *Theory of Probability & Its Applications*, Vol. 9, No. 1, pp. 141–142.

- Nielsen, S. F. (2003) “Proper and Improper Multiple Imputation,” *International Statistical Review / Revue Internationale de Statistique*, Vol. 71, No. 3, pp. 593–607.
- Rezvan, P. H., Lee, K. J., and Simpson, J. A. (2015) “The rise of multiple imputation: a review of the reporting and implementation of the method in medical research,” *BMC Medical Research Methodology*, Vol. 15, No. 1, p. 30.
- Robins, J. M. and Wang, N. (2000) “Inference for Imputation Estimators,” *Biometrika*, Vol. 87, No. 1, pp. 113–124.
- Rubin, D. B. (1976) “Inference and missing data,” *Biometrika*, Vol. 63, No. 3, pp. 581–592.
- Rubin, D. B. (1986) “Statistical Matching Using File Concatenation With Adjusted Weights and Multiple Imputations,” *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp. 87–94, DOI: <http://dx.doi.org/10.1080/07350015.1986.10509497>.
- Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*: Wiley, pp.258.
- Rubin, D. B. (1996) “Multiple Imputation After 18+ Years,” *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 473–489.
- Rubin, D. B. (2004) “The Design of a General and Flexible System for Handling Nonresponse in Sample Surveys,” *The American Statistician*, Vol. 58, No. 4, pp. 298–302, URL: <http://dx.doi.org/10.1198/000313004X6355>, DOI: <http://dx.doi.org/10.1198/000313004X6355>.
- Rubin, D. B. and Schenker, N. (1986) “Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse,” *Journal of the American Statistical Association*, Vol. 81, No. 394, pp. 366–374.
- Schafer, J. L. (1997) *Analysis of incomplete multivariate data*: CRC press.
- Schafer, J. L. and Graham, J. W. (2002) “Missing data: our view of the state of the art.,” *Psychological methods*, Vol. 7, No. 2, p. 147.
- Schafer, J. L. and Olsen, M. K. (1998) “Multiple imputation for multivariate missing-data problems: A data analyst’s perspective,” *Multivariate behavioral research*, Vol. 33, No. 4, pp. 545–571.
- Scheuren, F. (2005) “Multiple Imputation: How It Began and Continues,” *The American Statistician*, Vol. 59, pp. 315–319, 02.
- Schlaifer, R. and Raiffa, H. (1961) *Applied statistical decision theory*: MIT Press.
- Schölkopf, B. and Smola, A. J. (2002) *Learning with kernels: support vector machines, regularization, optimization, and beyond*: MIT press.
- Scott, D. (2015) *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics: Wiley.

- Seaman, S. R. and White, I. R. (2013) “Review of inverse probability weighting for dealing with missing data,” *Statistical Methods in Medical Research*, Vol. 22, No. 3, pp. 278-295, DOI: <http://dx.doi.org/10.1177/0962280210395740>.
- Seaman, S. R., White, I. R., Copas, A. J., and Li, L. (2012a) “Combining multiple imputation and inverse-probability weighting,” *Biometrics*, Vol. 68, No. 1, pp. 129–137.
- Seaman, S. R., Bartlett, J. W., and White, I. R. (2012b) “Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods,” *BMC medical research methodology*, Vol. 12, No. 1, p. 46.
- Seaman, S. R., Galati, J., Jackson, D., and Carlin, J. (2013) “What Is Meant by” Missing at Random”?” *Statistical Science*, pp. 257–268.
- 下平英寿 (2012) 「ブートストラップ」, 国友直人・山本拓 (編) 『21 世紀の統計科学 Vol. III 数理・計算の統計科学』, 東京大学出版会, 第 8 章, 193–221 頁.
- Sugiyama, M., Takeuchi, I., Suzuki, T., Kanamori, T., Hachiya, H., and Okanohara, D. (2010) “Least-squares conditional density estimation,” *IEICE Transactions on Information and Systems*, Vol. 93, No. 3, pp. 583–594.
- Sugiyama, M., Suzuki, T., and Kanamori, T. (2012) *Density ratio estimation in machine learning*: Cambridge University Press.
- 高井啓二・星野崇宏・野間久史 (2016) 『欠測データの統計科学——医学と社会科学への応用』, 第 1 巻, 調査観察データ解析の実際, 岩波書店.
- Tanner, M. A. and Wong, W. H. (1987) “The calculation of posterior distributions by data augmentation,” *Journal of the American statistical Association*, Vol. 82, No. 398, pp. 528–540.
- Titterington, D. M. and Sedransk, J. (1989) “Imputation of missing values using density estimation,” *Statistics & Probability Letters*, Vol. 8, No. 5, pp. 411–418.
- 富田裕章・藤澤洋徳・逸見昌之 (2017) 「欠測データに対する多重代入法から得られる推定量のバイアス補正について」, 応用統計学. 掲載確定..
- Tomita, H., Fujisawa, H., and Henmi, M. (2017) “A bias-adjusted estimator in multiple imputation for missing data,” in submission.
- van der Vaart, A. W. (1998) *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics: Cambridge University Press.
- Vink, G. and van Buuren, S. (2013) “Multiple imputation of squared terms,” *Sociological Methods & Research*, Vol. 42, No. 4, pp. 598–607.
- Von Hippel, P. T. (2009) “How to impute interactions, squares, and other transformed variables,” *Sociological Methodology*, Vol. 39, No. 1, pp. 265–291.

- Wang, N. and Robins, J. M. (1998) “Large-sample theory for parametric multiple imputation procedures,” *Biometrika*, Vol. 85, No. 4, pp. 935–948.
- 渡辺美智子・山口和範 (2000) 『EM アルゴリズムと不完全データの諸問題』, 多賀出版.
- Watson, G. S. (1964) “Smooth regression analysis,” *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372.
- White, I. R. and Carlin, J. B. (2010) “Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values,” *Statistics in Medicine*, Vol. 29, No. 28, pp. 2920–2931.
- White, I. R., Royston, P., and Wood, A. M. (2011) “Multiple imputation using chained equations: Issues and guidance for practice,” *Statistics in Medicine*, Vol. 30, No. 4, pp. 377–399.
- Woolley, S. B., Cardoni, A. A., and Goethe, J. W. (2009) “Last-Observation-Carried-Forward Imputation Method in Clinical Efficacy Trials: Review of 352 Antidepressant Studies,” *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, Vol. 29, No. 12, pp. 1408–1416.
- Yuan, Y. (2011) “Multiple Imputation Using SAS Software,” *Journal of Statistical Software, Articles*, Vol. 45, No. 6, pp. 1–25, DOI: <http://dx.doi.org/10.18637/jss.v045.i06>.