2.3 キャリブレーション推定法

キャリブレーション推定法は、補助変数の標本における値、及び母集団における周辺分布の情報に基づいて「ウェイト」の調整を行う、一般的な推定方法である。次節のIPW 法も、ウェイトの調整によって欠測バイアスを緩和する手法であり、直感的な説明はキャリブレーション推定法と共通する部分が大きい。両者で異なる点として、第1に、補助変数に関する母集団特性の情報を、キャリブレーション推定法では用いるが、IPW 法では用いない。第2に、欠測バイアス以外のバイアス(後述)も、キャリブレーション推定法では補正しているが、IPW 法は欠測バイアスのみを補正する。本節ではまず、「ウェイト」について説明したうえで、キャリブレーション推定法の概要を示す。また、キャリブレーション推定の特殊形である事後層化推定について、ウェイト調整によって標本の偏りを補正する考え方の直感的な理解を目指す。

一般的に標本調査におけるウェイトとは、標本に含まれた調査客体のそれぞれが、目標母集団の要素何単位分を代表しているか、を表す尺度である。ウェイトは、目標母集団の要素のそれぞれが標本に含まれる確率(「包含確率」と呼ばれる)の逆数に等しい。1%の確率で標本に含まれる調査客体は、目標母集団の要素 100 単位分(すなわち当該調査客体と他の 99 の調査対象候補)を代表し、20%の確率で標本に含まれる調査客体は、目標母集団の要素 5 単位分(すなわち当該調査客体と他の 4 の調査対象候補)を代表している(標本における母集団代表性の尺度としての包含確率の逆数の妥当性は、Horvitz-Thompson 推定量の不偏性と関連している。)。単純無作為抽出では、すべての調査対象候補が等確率で抽出されるので、すべての調査客体でウェイトの値は等しい。特に非復元単純無作為抽出の場合、すべての調査対象候補について包含確率の値はn/N(標本サイズ/母集団サイズ)、従ってウェイトの値はN/n(母集団サイズ/標本サイズ)である。(復元単純無作為抽出であれば、包含確率は$1-((\mathrm{N}-1)/\mathrm{N})^{\mathrm{n}}$、従ってウェイトの値は$\{1-((\mathrm{N}-1)/\mathrm{N})^{\mathrm{n}}\}^{-1}$である。)このように、標本抽出デザインが決まれば包含確率も決まるので、ウェイトも決まる。ウェイトとしての包含確率の逆数を、特に「抽出ウェイト」と呼ぶ。回答率 100%の標本に対しては、抽出ウェイトを用いることで偏りのない推定が可能である(Horvitz-Thompson 推定量の不偏性に関する次節の説明参照)。標本調査で欠測が生じた場合は、いわば母集団から標本へと選出された代表に欠員が生じているので、回答標本に対して抽出ウェイトを用いた推定は、母集団をあまねく反映していないことになる(都議会で都議に欠員が生じると、当該選挙区の意見が都政に反映されなくなるイメージ)。そこで、キャリブレーション推定法及び IPW 法では、回答標本におけるウェイトを調整することで、回答標本の偏りを補正することが意図される。ここで注意すべきは、ウェイト調整による補正では、欠測となった調査客体が代表している母集団の要素に類似した他の要素を代表する調査客体が、回答標本の中になお残されていなければならないという点である(「サポ

ート問題（support problem）」と呼ばれる）。これがウェイト調整を行う手法の前提となる。

なお、ここでは欠測バイアスを緩和する統計的処理法として、キャリブレーション推定法をとりあげているが、キャリブレーション推定法は、欠測バイアスに限らず、より一般的な標本の偏りを補正する手法である。本書では、「標本の偏り」としては、欠測が生じた場合に分析の対象となる回答標本の偏り、すなわち欠測バイアスのみを考えているが、一般的に、欠測が生じない条件下でも標本の偏りは生じる。それは、「運の悪い標本抽出結果」という形で事後的に生じるものである。単純無作為抽出法は、事前の意味では母集団の縮図となる標本を抽出するが、偏った標本が抽出される確率は0ではない（※確率比例抽出、層化抽出、多段抽出などは、このような悪運を抑制する標本抽出デザインといえる）。キャリブレーション推定法は、欠測バイアスへの対応としても利用できるが、欠測が生じない場合にも利用され、その場合は、事後的に（運悪く）標本が偏ることへの対応となっている。

キャリブレーション推定法は、補助変数について推定値が母集団特性値の真の値に等しくなるようなウェイトを用いた推定法である。ただし、補助変数について推定値が母集団特性値の真の値に等しくなるようなウェイトは一意ではない。キャリブレーション推定法では、通常無数に存在するウェイトの候補のなかで、抽出ウェイトに最も近いものを採用する。補助変数に関して、あるウェイトを用いた推定値がその母集団特性値に等しいことを表す条件式を、当該ウェイトの「キャリブレーション方程式（calibration equation）」と呼び、キャリブレーション方程式を満たしかつ抽出ウェイトからの距離を最小化するウェイトを、「キャリブレーションウェイト（calibration weight）」と呼ぶ。キャリブレーションウェイトを用いた推定が、キャリブレーション推定である。キャリブレーションウェイトを算出するためには、用いる補助変数の母集団特性値が知られていなければならない。

キャリブレーション推定法の要点を理解するために、キャリブレーション推定法の特殊形のひとつである「層サイズによる事後層化推定」の考え方を、図2－4－1に示す。目的となる変数Yに欠測が生じ、補助変数Xの値はすべての調査客体で観測されている。図2－4－1のXY平面上の散布図は、仮に目的となる変数Yの値がすべての調査客体で観測される（すなわち完全データが観測される）場合に得られるものであり、観測データのレコードを記号○、欠測データのレコードを記号△で表す（記号○についてはX座標とY座標が両方とも知られているが、記号△についてはX座標のみが知られている）。図2－4－1では、完全データにおいて目的となる変数Yと補助変数Xの間に高い正の相関がある状況を考える。

図2－4－1（イ）に示す3つのパネルそれぞれの上方側のグラフは、目的となる変数Yについて、真の分布及び欠測によってゆがめられた分布を示したものである。目的となる変数Yの分布で、灰色の領域は、変数Yが観測されないレコードの、変数Yの値ごとの頻度を示す（図1－1～1－3のヒストグラム参照）。このグラフによると、欠測の起こ

りやすさが欠測する変数Yの値に依存しているので MNAR である(変数Yとの相関が高い補助変数Xが利用可能でなければ)。特に、変数Yの値が大きいほど欠測が起こりやすくなっている。このことは、散布図からも確認できる。すなわち、右側の領域ほど記号△で表される欠測レコードの割合が高くなっている。

図2－4－1(イ)に示す3つのパネルそれぞれの左側方のグラフは、補助変数Xの真の分布、及び目的となる変数Yが観測されているという条件による補助変数Xの条件付分布、を示したものである。補助変数X自体は欠測が生じない変数なので、上側方の目的となる変数Yに関する分布のグラフとの相違点に注意を要する。左側方の補助変数Xに関するグラフでは、灰色の領域は、(補助変数Xではなく)目的となる変数Yが観測されないレコードの、補助変数Xの値ごとの相対的頻度を示す。

ここで、補助変数Xの値に基づいて不完全データの標本を層化することができる。層の数をKとする。図2－4－1(イ)及び(ロ)のそれぞれに示す3つのパネルは左から順に第 1 層、第k層$(k = 2, 3, \cdots, K-1)$及び最後の第K層に注目した場合を示したものである。補助変数Xはすべて観測されているので、任意の第k層において(つまりどの層においても)、観測レコードと欠測レコードの構成比を知ることができる。たとえば、図の第k層では、観測レコード7件(7個の記号○)と欠測レコード6件(6個の記号△)から成っている(第k層の回答者数$n_k^R = 7$及び無回答者数$n_k^M = 6$)。この層では、観測レコード 1 件を(7+6)/7 倍に膨らませることで、(観測レコードのみを用いて)補助変数Xの分布を完全データのものに一致させることができる。さらに、母集団について任意の第k層のサイズN_kが知られていれば、第k層の観測レコード 1 件を(母集団第k層のサイズ/母集団サイズ)/(回答標本第k層のサイズ/標本サイズ)の倍率で膨らませることで、(観測レコードのみを用いて)補助変数Xの分布を母集団のものに一致させることができる。この場合のキャリブレーション方程式は、$\sum_{i \in S_k^R} w_i^C = N_k$である(ただし$S_k^R$は回答標本の第$k$層である)。

もっとも補助変数Xはすべて観測されているので、補助変数Xに関する推定が目的であればことさらに上記の処理をする必要はない。上記の処理を目的となる変数Yについて実行できればよいが、それは不可能である。そこで、目的となる変数Yのかわりに補助変数Xについて実行するのである。欠測は、目的となる変数Yの回答標本における分布をゆがませるが、それと連動して補助変数Xの回答標本における分布もゆがませる。この連動性に着目すると、補助変数Xの回答標本における分布のゆがみを補正すれば、それに連動して変数Yの回答標本における分布のゆがみも補正されていることが期待される。これが、キャリブレーション方程式を制約条件とすることの動機となっている。

補助変数Xの次元で行われる補正が、目的となる変数Yの次元でどのような効果をもつかを示したのが、図2－4－1(ロ)である。ただし上述の通り、キャリブレーション推

定法には欠測バイアス以外のバイアス(ここでは「運の悪い標本抽出」によるバイアスのみ)も同時に補正する機能があり、ここでは欠測バイアスを補正する機能のみを示したいので、ウェイト補正を欠測バイアスに対応する部分とそれ以外のバイアスに対応する部分の2つに分解し、前者の効果のみを図示する。単純無作為抽出の抽出ウェイトであれば、母集団について任意の第k層のサイズN_kの値を用いて、第k層の観測レコード 1 件を(母集団第k層のサイズ/母集団サイズ)/(回答標本第k層のサイズ/標本サイズ)の倍率で膨らませるが、この倍率を項(標本第k層のサイズ)/(回答標本第k層のサイズ)と項(標本サイズ/標本第k層のサイズ)×(母集団第k層のサイズ/母集団サイズ)の積としてみると、前者は欠測バイアスの補正、後者は「運の悪い標本抽出」によるバイアスの補正となっている。第k層に属する記号○(すなわち観測レコード)を(7+6)/7 倍に膨らませるのであるから、(ロ)上側方の補正前分布では、矢印で示したとおりの垂直方向の拡張が第k層に属する観測レコードの各点で起こる。これが、事後層化第k層における補正の効果である。同様の補正が他のすべての層においても行われ、目的となる変数Yの分布のゆがみが補正される。第 1 層では、3 件の観測値と 0 件の欠測値があるから、観測値のウェイトは(3+0)/3 倍に調整される。第K層では、1 件の観測値と 3 件の欠測値があるから、観測値のウェイトは(1+3)/1 倍に調整される。

図2－4－1に示した層サイズによる事後層化推定法の例では、補助変数Xの値に基づく任意の第k層において(つまりどの層においても)、観測レコード数(記号○の数)と欠測レコード数(記号△の数)の合計に占める観測レコード数(記号○の数)の割合の逆数を抽出ウェイトに乗じたものをウェイトとして、目的となる変数Yの分布を推定することで、欠測バイアスが除かれる。事後層化推定法は、目的となる変数Yと補助変数Xとの相関が高いほど、目的となる変数Yの分布の推定における欠測バイアスをより多く取り除くことができる。図2－4－1に示した原理は、ウェイト調整に関する一般的な原理であり、IPW 法でも同様にはたらいている。

次に、目的となる変数Yと補助変数Xとの相関が低いときには欠測バイアスを緩和する効果が小さくなることを、図2－4－2により示す。図2－4－2は、目的となる変数Yと補助変数Xとに相関がないときに事後層化推定法を適用した場合である。記号や配色の意味は、図2－4－1と同様である。図2－4－1との違いは、目的となる変数Yと補助変数Xとの相関がないという点だけである。図2－4－1及び図2－4－2に示された点線の楕円形は、変数Yと変数X間の全データの散布図における散らばり方(正確には同時分布の等量線)を表している。欠測レコード(記号○)と観測レコード(記号△)の散布図上の位置は、この点線の楕円形によって確率的に制約される。図2－4－1では楕円形が細長いので、レコードのY軸座標がひとたび決まれば、とり得るX軸座標の範囲はかなり狭まる(逆も然りである)が、図2－4－2では楕円形が幅広いので、レコードのY軸座標が決まっても、とり得るX軸座標の範囲は依然として広い(逆も然りである)。図2－4－1及び図2－4－2のいずれにおいても、等しく目的となる変数Yの値が

大きいほど欠測が起こりやすくなっているので、記号△で表されるレコードは、Y軸の右方により多く集中する。その結果、楕円形が細長い図2－4－1では、必然的にX軸方向の下方により多く集中することになるが、楕円形が幅広い図2－4－2では、X軸方向に関しては均一に分布してしまう。このため図2－4－2では、すべての層において欠測レコードの割合が等しくなっている。図2－4－2では、すべての層で観測レコードのウェイトに等しい値を掛けるので、実質的にウェイトの調整がなされないことになる。

補助変数の目的となる変数に対する相関の有無によるウェイト調整結果の違いは、図2－4－1(ロ)と図2－4－2(ロ)を比較することでも分かる。相関のある図2－4－1(ロ)で分布の左側が拡張されているのと比べて、相関のない図2－4－2(ロ)では分布のやや右寄りが拡張されており、ゆがみが正しく補正されていない。この図ではひとつの層(第 k 層)におけるウェイト調整の効果を示しているが、他の層についても同様の効果がみられる。

ここで注目すべき点は、図2－4－2(イ)及び(ロ)左側方のグラフである。目的となる変数Yと補助変数Xとに高い相関がある図2－4－1の場合は、欠測の起こりやすさと補助変数Xとの間に相関があったものの、目的となる変数Yと補助変数Xとに相関がない図2－4－2の場合は、欠測の起こりやすさと補助変数Xとの間に相関がない。これは、もともと欠測の起こりやすさが目的となる変数Yに依存しているので、目的となる変数Yとの相関が高い補助変数は、欠測の起こりやすさとも相関が高く、目的となる変数Yとの相関がない補助変数は、欠測の起こりやすさとも相関がないということの表れである。図2－4－1では変数Yの欠測・観測別分布と変数Xの欠測・観測別分布に連動性があったが、図2－4－2では解消している。回答者に関する変数Xの分布を真の分布へ向けて補正することで、間接的に変数Yの分布を補正するキャリブレーション推定法では、目的となる変数Yと補助変数Xの連動性(相関)が重要である。

このように、補助変数が目的となる変数と相関をもたなければ(より正確には、補助変数が欠測確率に対して説明力をもたなければ)、事後層化推定法は欠測バイアスを緩和できない。この点は、単一代入法、多重代入法、IPW 法などの他の欠測データ処理法についても当てはまる。

キャリブレーション推定法を実行するためには、補助変数の(層別)母集団総計を知っている必要がある。この点が、他の手法と比べてキャリブレーション推定法を実行する上での大きな制約となる。政府統計においては、母集団データベースの活用が期待される所以である。

図2－4－1　層サイズによる事後層化推定の考え方

Y：目的となる変数、X：補助変数、○：観測レコード、△：欠測レコード

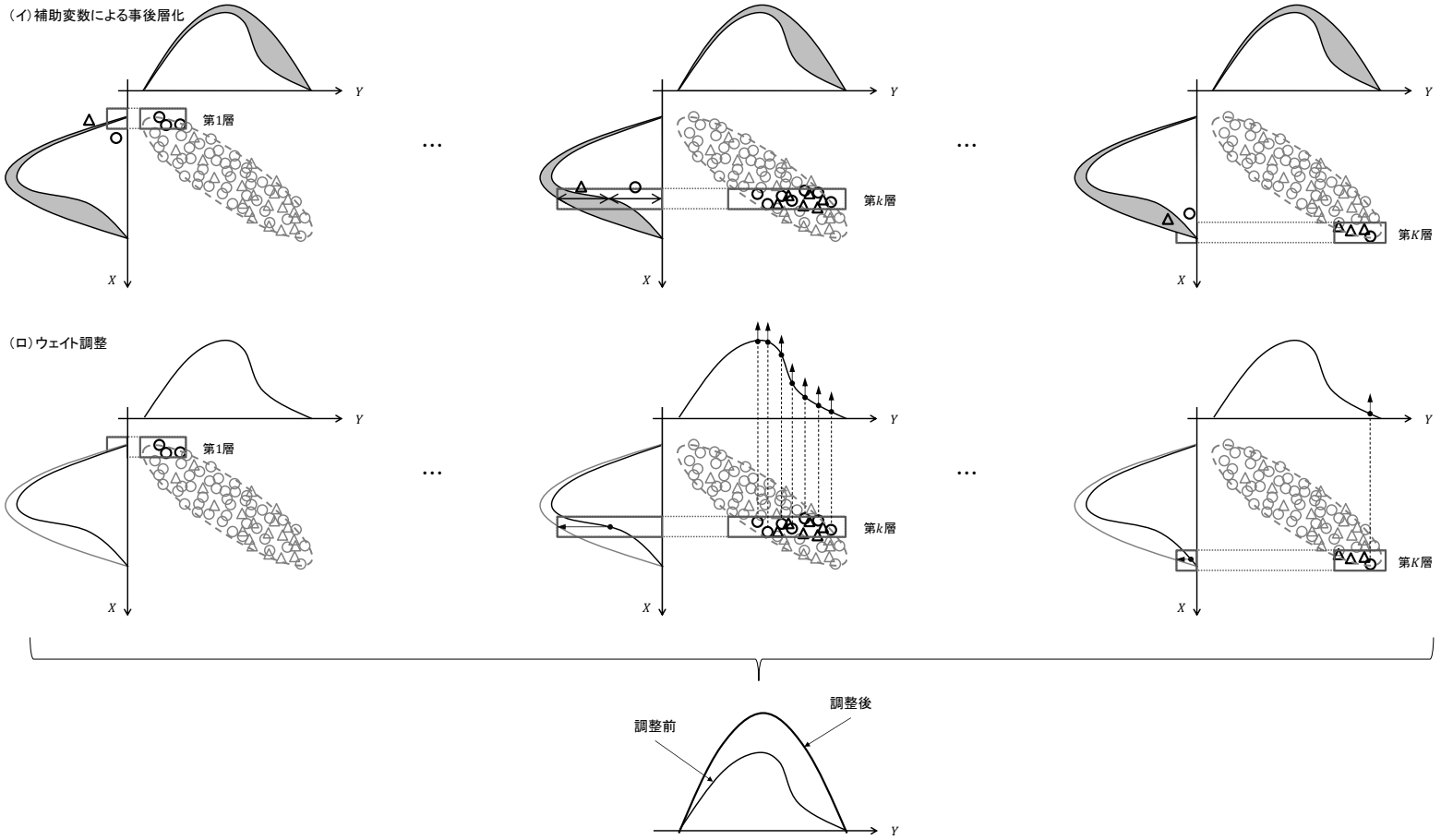
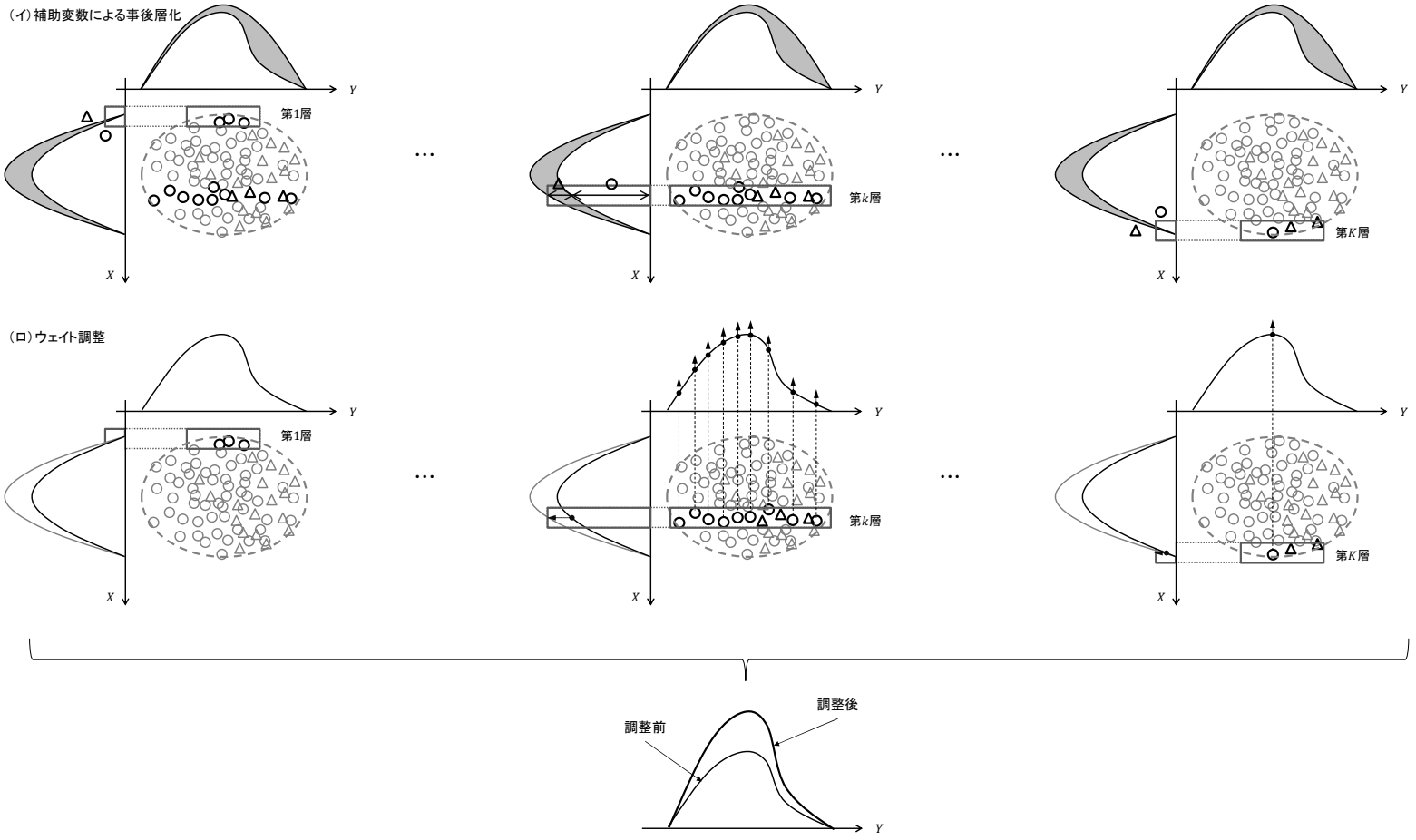


図2－4－2　目的となる変数Yと補助変数Xに相関がない場合の事後層化推定

Y:目的となる変数、X:補助変数、○:観測レコード、△:欠測レコード



2.4 IPW 法

標本調査において欠測が生じない場合は、包含確率（母集団の各要素が標本に含まれる確率）の逆数を調査客体ごとのウェイトとすることで、偏りのない推定ができる（Horvitz-Thompson 推定量の不偏性）。この原理を欠測が生じる場合へ拡張した推定手法が、IPW（inverse probability weighting）法である。

拡張は次のように考えることで容易となる。欠測が生じない場合は標本抽出という試行によって分析対象となる標本が得られ、欠測が生じる場合は標本抽出並びに「回答の成否」という2つの試行が合成された試行によって分析対象となる回答標本が得られる。つまり、欠測が生じない場合における試行としての標本抽出を、上述の合成試行に置換えて考えればよい。このことから、欠測が生じる場合は、包含確率の代わりに「母集団の各要素が標本に含まれかつ回答する確率」の逆数を調査客体ごとのウェイトとすることで偏りのない推定となる。

ここで、包含確率の値は標本抽出デザインに応じて決まるためその真の値が知られているのに対して、一般的に母集団の各要素について「標本に含まれかつ回答する確率」の真の値を知ることはできずデータから推定しなければならないということが問題として現れる。IPW 法の適性は、調査客体ごとの「標本に含まれかつ回答する確率」の真の値を正しく推定できるか否かにかかっている。「標本に含まれかつ回答する確率」は非常に緩やかな条件の下で包含確率と「回答する確率」に分解でき、「回答する確率」は別の2つの条件の下で不完全データから正しく推定できる。IPW 法でウェイトに用いる確率の分解及び回答確率の推定における仮定を順にみていく。

まず、一般的に、任意の調査客体の「標本に含まれかつ回答する確率」は当該調査客体の「標本に含まれた場合に回答する確率」と当該調査客体の包含確率との積に等しい。また、標本抽出と回答成否の事象が互いに母集団の要素の属性による条件付独立でありかつ母集団の要素ごとの属性が固定されている場合、任意の調査客体の「標本に含まれた場合に回答する確率」は、単に当該調査客体の「回答する確率」に等しい。まとめると、標本抽出と回答成否の事象が互いに母集団の要素の属性による条件付独立でありかつ母集団の要素ごとの属性が固定されているという条件の下では、任意の調査客体の「標本に含まれかつ回答する確率」は、当該調査客体の「回答する確率」と当該調査客体の包含確率との積に等しい（「数式を使った説明」参照）。

2つの条件（1）標本抽出と回答の有無が互いに条件付独立であること及び（2）母集団の要素ごとの属性が固定されていることは、厳しいものではない。特に、無作為抽出による測定誤差のない標本調査の場合はこれらの条件が成立している。IPW 法ではこれら2つの緩い条件を前提として任意の調査客体の「回答する確率」の値を推定し、それに標本抽出デザインによって決まる包含確率の値を乗じることで、調査客体ごとの「標本に含まれかつ回答する確率」を求める。

次に、不完全データから調査客体ごとの「回答する確率」の値を正しく推定できるためには、さらに別の条件が成立していなければならない。第1に、IPW 法においては、任意の調査客体の「回答する確率」を当該調査客体の属性の関数としてモデル化し、そのモデルのパラメータを推定することで、任意の調査客体の「回答する確率」を推定する。つまり、IPW 法では、回答確率（観測確率）のモデルが正しく特定化されていなければならない。モデルの特定化に誤りがあると、調査客体ごとの「回答する確率」の推定値に誤設定バイアスが伴うからである。第2に、回答確率（観測確率）のモデルは不完全データから推定できなければならない。すなわち、IPW 法では欠測データメカニズムは MAR でなければならない（欠測データメカニズムが MNAR だと、回答確率が観測されない値に依存するため、回答確率モデルは推定できない）。これらの2つの条件（1）観測確率モデルが正しく特定化されていること、及び（2）欠測データメカニズムが MAR であることは、IPW 法における重要な仮定である。

MAR の下では、IPW 法の回答確率モデルにおいて、任意の調査客体の「回答する確率」は、適当な補助変数の値で条件付けた回答確率、すなわち回答の傾向スコアに他ならない。つまり、IPW 法は、MAR の仮定の下で、回答の傾向スコアを推定し、その推定値と包含確率の積の逆数をウェイトとして推定を行う。MAR の下で、傾向スコアの確率モデルが正しければ、IPW 法は欠測バイアスを緩和することができる。

図2－5は、IPW 法によって欠測バイアスが緩和される原理を示したものである。与えられた不完全データに対応する完全データについて、目的となる変数Yと補助変数Xの散布図を最下部に示す。目的となる変数Yの値が観測されている調査客体を記号○、観測されていない調査客体を記号△で表している。つまり、記号○はX座標とY座標の両方が知られているが、記号△はX座標しか知られていない。散布図の上には、完全データにおける変数Yのヒストグラムを示す。灰色部分は欠測値、白色部分は観測値に対応する。ヒストグラムの上には、データから推定される変数Yの分布の形状を示す。2つの曲線のうち、上は仮に完全データが観測された場合の推定分布であり、下は回答標本から推定される分布である。ヒストグラムとの対応を分かりやすくするために、2つの分布の縮尺はそろえていない。2つの曲線に挟まれる領域の垂直距離によって、変数Yの値ごとの欠測率が表される。図2－5のデータ例では、変数Yの値が大きいほど欠測率が高くなる。これだけだと、欠測確率が欠測する変数の値に依存する、すなわち MNAR となるが、同時に図2－5のデータ例では、目的となる変数Yと補助変数Xの相関が大きい。そこで、欠測率は補助変数にも依存し、特に、補助変数だけで説明できれば MAR となる。ここでは、MAR であるとする。

図2－5左下のグラフは、散布図に示されるデータに基づいて傾向スコアを推定したときの結果を表したものである。通常傾向スコアは、観測指標Rの補助変数Xによる2項回帰モデル（ロジットモデルやプロビットモデル）によって推定される。点線が推定さ

れた傾向スコアを表す。この例では、補助変数Xの値が大きいほど傾向スコアは小さくなる。

既に述べたとおり IPW 法は、MAR の仮定の下で推定された傾向スコアの逆数を、抽出ウェイト（包含確率の逆数）に乗じることでウェイトの調整を行う。欠測によって変数Yの分布に生じたゆがみが、傾向スコアによるウェイト調整で補正される効果をみるために、補助変数Xの値ごとのウェイト調整を分けて示す。図では、補助変数Xの値が標本における最も小さい値である場合（$X = \underline{x}$）、最も大きい値の場合（$X = \overline{x}$）及び中間値の場合（$X = x$）、の3通りについて示している。このため、同じ散布図、ヒストグラム及び分布が3つ並んでいる。左から順に、最小値、中間値、最大値の場合である。

左端の散布図において強調表示された記号〇で表された調査客体は、補助変数Xの値が最も小さい。当該調査客体の傾向スコアの推定値は、左端のグラフによると、9/10 である。当該調査客体は、事前には 90%の確率で変数Yの値が観測されるという性質をもっている。変数Yに関する IPW 法の推定において、当該調査客体のウェイトは、推定された傾向スコアの逆数である 10/9 倍に調整される。このように調整されたウェイトによると、当該調査客体は、回答標本において、自身と回答標本に含まれなかった他の要素 1/9 単位分を代表するものとして扱われる。散布図の上のヒストグラムにある上向きの矢印は、この調整によって当該調査客体のウェイトが 10/9 倍に増加することを表している。

中央の散布図において強調表示された記号〇で表された調査客体は、補助変数Xが値xをとる。これら調査客体の傾向スコアの推定値は、左端のグラフによると、1/2 である。当該調査客体は、事前には 50%の確率で変数Yの値が観測されるという性質をもっており、実際に当該調査客体の変数の値Yは観測された。変数Yに関する IPW 法の推定において、当該調査客体のウェイトは、推定された傾向スコアの逆数である 2 倍に調整される。このように調整されたウェイトによると、当該調査客体は、回答標本において、自身と回答標本に含まれなかった他の要素 1 単位分を代表するものとして扱われる。中央のヒストグラムにある上向きの矢印は、この調整によって、当該調査客体のウェイトが 2 倍に増加することを表している。

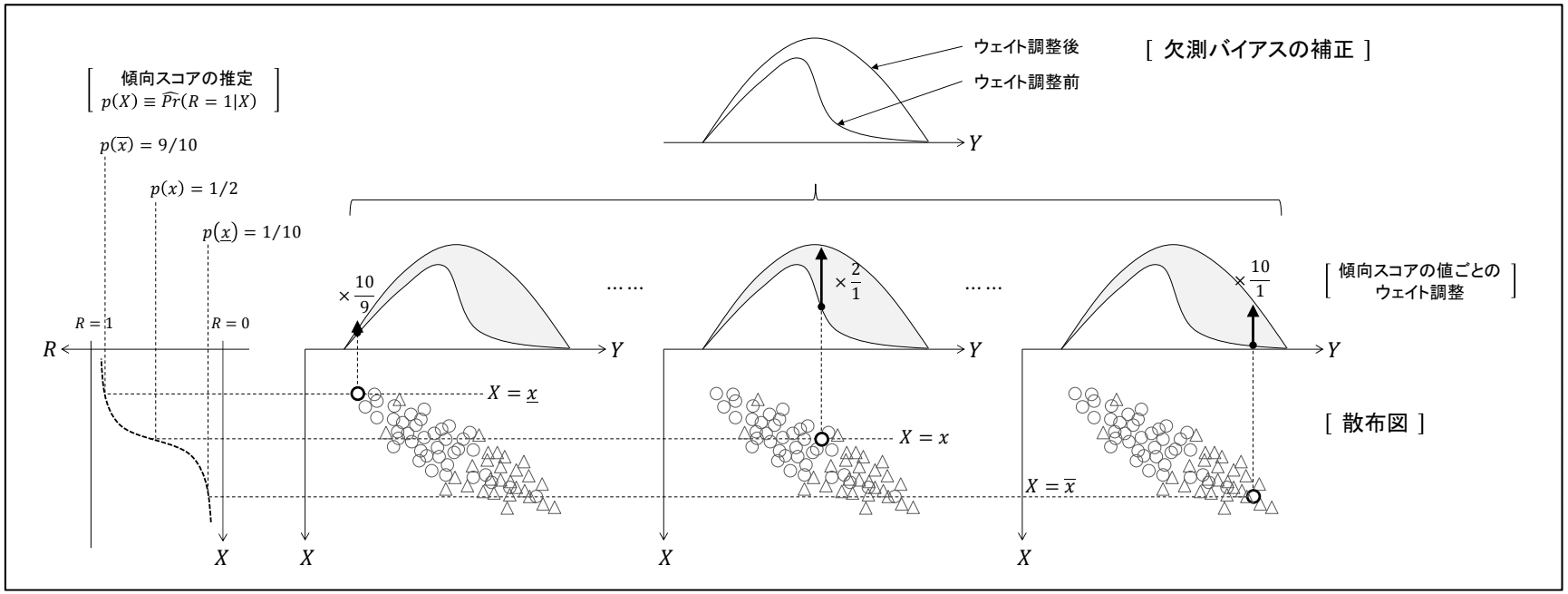
右端の散布図において、強調表示された記号〇で表された調査客体は、補助変数Xの値が最も大きい。当該調査客体の傾向スコアの推定値は、左端のグラフによると、1/10 である。当該調査客体は、事前には 10%の確率で変数Yの値が観測されるという性質をもっている。変数Yに関する IPW 法の推定において、当該調査客体のウェイトは、推定された傾向スコアの逆数である 10 倍に調整される。このように調整されたウェイトによると、当該調査客体は、回答標本において、自身と回答標本に含まれなかった他の要素 9 単位分を代表するものとして扱われる。右端のヒストグラムにある上向きの矢印は、この調整によって当該調査客体のウェイトが 10 倍に増加することを表している。

回答標本のすべての調査客体について推定された傾向スコアによりウェイトを調整した結果をまとめると、図2－5最上部に示す分布のように、推定の欠測バイアスが緩和される。緩和の程度は、目的となる変数Yと補助変数Xとの（正負を問わない）相関の強さに依存する。相関係数の絶対値が1であれば、欠測バイアスは完全に除去される。相関係数の値が0であれば欠測バイアスは全く緩和されない。このことは、図2－5中央の散布図及びヒストグラムから理解することができる。相関が強ければ、散布図で水平に並んだ記号○及び△の相互の距離が小さくなる。記号△と当該記号△を代理する記号○との水平距離が小さいと、変数Yに関する代理の妥当性は強くなる。逆に相関が弱ければ、記号△と当該記号△を代理する記号○との水平距離が大きくかい離し、変数Yに関して互いに大きく異なる記号△を記号○が代理することになる。

IPW 法は、MAR の仮定の下で、傾向スコアによりウェイトを調整することで、欠測バイアスを緩和する。MAR の仮定は、回答の傾向スコアを不完全データから正しく推定するための条件である。このほか IPW 法では、傾向スコアを推定するための回答確率モデルの特定化が正しくなければならない。これら2つの条件（1）MAR 及び（2）正しい回答確率モデルは、実務においては、「欠測確率を十分に説明できる補助変数が利用可能である」という条件として考えることができる。

IPW 法に伴う実際問題として、補助変数に関して十分に広範な属性の調査客体で回答が成立していなければ、回答標本のなかに、推定される傾向スコアの値が非常に小さな調査客体が生じてしまい、最終的な推定結果が極端な値となってしまうことがある。これは、ロジットモデルやプロビットモデルのようなパラメトリックなモデル化の限界とみることもでき、また、偶々IPW 法に不向きな不完全データが得られたことの結果とみることもできる。IPW 法を実施する際は、調整後のウェイトが極端な値となっていないかを確認する必要がある。

図２－５　IPW 法の要点


傾向スコアの推定
$p(X) \equiv \widehat{Pr}(R = 1|X)$
$p(\overline{x}) = 9/10$
$p(x) = 1/2$
$p(\underline{x}) = 1/10$
$R = 1$
$R = 0$
R
X
$\times \frac{10}{9}$
$\times \frac{2}{1}$
$\times \frac{10}{1}$
Y
$X = \underline{x}$
$X = x$
$X = \overline{x}$
ウェイト調整後
ウェイト調整前
[欠測バイアスの補正]
[傾向スコアの値ごとの
ウェイト調整]
[散布図]

2.5 多重代入法

多重代入法は、確率的代入の考え方に基づいて、疑似完全データを複数作成する手法であるが、単一代入法である確率的回帰代入を繰り返し互いに独立に実行するものとは異なる。欠測値に関わる不確実性としては、第１に、欠測した値の背後にあるデータ生成過程自体に関する不確実性と、第２に、欠測値（の真の値）がある特定のデータ生成過程から発生するときの不確実性の２つが区別できる。単一代入法である確率的回帰代入を繰り返し互いに独立に実行するだけでは、第２の不確実性に対応することはできても、第１の不確実性を捉えることはできない。多重代入法は、２つの不確実性に対応した代入法であるといえる。
多重代入法の考え方を理解するためには、「分散分解」を理解しなければならない。一般的に次の命題が成り立つ。

ある条件による条件付分散は、当該条件を情報として包摂する条件による条件付分散の当該条件による条件付期待値と、当該条件を包摂する条件による条件付期待値の当該条件による条件付分散の和に等しい（確率変数(A,B,C)について$Var(A\|B)=E[Var(A\|B,C)\|B]+Var(E[A\|B,C]\|B)$）

この法則は、特に「分散分解」と呼ばれる。

分散分解の関係式によって、推定精度の評価における単一代入法の問題点を示す。一般的に推定量$\hat{\theta}$（標本平均でも標本分散でも何でもよい）による推定精度は、推定量$\hat{\theta}$の分散$Var(\hat{\theta})$（あるいはその平方根である標準誤差）によって評価できる。観測データを与件としたときの欠測データの条件付分布を「事後予測分布」と呼ぶ。欠測値を代入値で置換えることによって作成される疑似完全データに、所定の推定処理を実行する手法（すなわち代入法）においては、疑似完全データの代入データに不確実性が内在している。その不確実性は、次の3つに区別できる。

(1)事後予測分布を与件としたときの欠測データ生成に関する不確実性
(2)事後予測分布自体の不確実性
(3)事後予測分布の推定に関する不確実性

疑似完全データを完全データとみなして推定結果を解釈することは、これらの不確実性を捨象していることになる。

標本調査における所定の推定量$\hat{\theta}^*$は、観測データ(Y^O,X)と欠測データ(Y^M)の関数であるから、推定量$\hat{\theta}^*$の分散について、分散分解により、次式が成り立つ。

$$Var(\hat{\theta}^*) = E\left[Var\left(\hat{\theta}^*\begin{pmatrix}\text{観測データ}(Y^O, X)\\ \text{欠測データ}(Y^M)\end{pmatrix}\middle|\text{観測データ}(Y^O, X)\right)\right]$$
$$+ Var\left(E\left[\hat{\theta}^*\begin{pmatrix}\text{観測データ}(Y^O, X)\\ \text{欠測データ}(Y^M)\end{pmatrix}\middle|\text{観測データ}(Y^O, X)\right]\right)$$

（2-5-1）

事後予測分布を与件としたときの欠測データ生成に関する不確実性は、（2-5-1）式右辺第1項の期待値の中の条件付分散及び同第2項の分散の中の条件付期待値によって捉えられる。事後予測分布自体の不確実性は、（2-5-1）式右辺第1項の期待値及び同第2項の分散によって捉えられる。（2-5-1）式は、仮に完全データが得られたとしたときの所定の推定量の分散であるから、事後予測分布の推定に関する不確実性は存在しない。

確率的回帰代入法の代入値は観測値及び回帰モデルの誤差項の関数であるから、確率的回帰代入法によって作成される疑似完全データの代入データ$\overline{Y^M}_{SSI}$は観測データ(Y^O, X)及び回帰モデルの誤差項ε^{SSI}の関数である。ただし、観測データと代入データの関係のうち、観測データから推定される回帰モデルのパラメータ$\hat{\beta} = \hat{\beta}\left(\text{観測データ}(Y^O, X)\right)$を介した部分を明示して、代入モデルを次式で表す。

$$\text{代入データ}\left(\overline{Y^M}_{SSI}\right) = g^{SSI}\left(\begin{matrix}\text{観測データ}(Y^O, X)\\ \text{誤差項}(\varepsilon^{SSI})\end{matrix}\middle|\text{パラメータの推定値}(\hat{\beta})\right)$$

確率的回帰代入法による推定量$\hat{\theta}^{SSI}$は、疑似完全データを完全データとみなして算出する推定量$\hat{\theta}^*$である。従って、推定量$\hat{\theta}^{SSI}$は、観測データと推定された回帰モデルの誤差項の関数である。

$$\hat{\theta}^{SSI} = \hat{\theta}^*\begin{pmatrix}\text{観測データ}(Y^O, X)\\ \text{代入データ}(\overline{Y^M}_{SSI})\end{pmatrix}$$
$$= \hat{\theta}^*\left(\begin{matrix}\text{観測データ}(Y^O, X)\\ g^{SSI}\left(\begin{matrix}\text{観測データ}(Y^O, X)\\ \text{誤差項}(\varepsilon^{SSI})\end{matrix}\middle|\text{パラメータの推定値}(\hat{\beta})\right)\end{matrix}\right)$$

そこで、確率的回帰代入法による推定量$\hat{\theta}^{SSI}$の分散については、分散分解により、次式が成り立つ。

$$
\begin{aligned}
&Var\left(\hat{\theta}^{SSI}\right) \\
&= E\left[Var\left(\hat{\theta}^{*}\left(g^{SSI}\left(\begin{matrix}\text{観測データ}(Y^{O},X)\\ \text{誤差項}(\varepsilon^{SSI})\end{matrix}\middle|\ \text{パラメータの推定値}(\hat{\beta})\right)\right)\middle|\ \text{観測データ}(Y^{O},X)\right)\right] \\
&+ Var\left(E\left[\hat{\theta}^{*}\left(g^{SSI}\left(\begin{matrix}\text{観測データ}(Y^{O},X)\\ \text{誤差項}(\varepsilon^{SSI})\end{matrix}\middle|\ \text{パラメータの推定値}(\hat{\beta})\right)\right)\middle|\ \text{観測データ}(Y^{O},X)\right]\right)
\end{aligned}
\qquad (2\text{-}5\text{-}2)
$$

(2-5-2)式右辺第1項は、観測データを与件としたときの推定量$\hat{\theta}^{SSI}$の条件付分散の期待値である。推定量$\hat{\theta}^{SSI}$は観測データだけでなく回帰モデルの誤差項にも依存するので、観測データを与件としたときの推定量$\hat{\theta}^{SSI}$の条件付分散自体は回帰モデルの誤差項に由来する変動を反映する。

(2-5-2)式右辺第2項は、観測データを与件としたときの推定量$\hat{\theta}^{SSI}$の条件付期待値の分散である。回帰モデルの誤差項に関する積分又は積算の演算が、代入モデルの誤差項によって表現される欠測データ生成に関する不確実性を表しており、所定の推定量$\hat{\theta}^{*}$の分散の分散分解第2項における欠測データに関する積分又は積算の演算が(観測データを与件としたときの)欠測データ生成に関する不確実性を表していることに対応している。ただし、この誤差項が本来の欠測データに由来する不確実性を過不足なくとらえているかは自明ではない。

(2-5-2)式右辺では、パラメータの推定に関わる不確実性も生じている。これは、(2-5-2)式右辺には存在しなかったものである。非確率的単一代入の実施者は、代入データを欠測データの真の値であるとみなしており、従って、事後分布のパラメータの推定値を真の値とみなしているため、(2-5-2)式右辺の分散を算出する際は、パラメータを確率変数とはみなさない。このように事後予測分布の推定に関する不確実性を適切に評価しないことも、推定精度を過大評価させる効果をもつ。

まとめると、確率的回帰代入法による推定量$\hat{\theta}^{SSI}$は、観測データだけでなく代入モデルの確率項にも依存する($\hat{\theta}^{SSI}=\hat{\theta}^{*}\left(\text{観測データ},\text{代入データ}\right)=\hat{\theta}^{*}\left(\text{観測データ},h^{SSI}(\text{観測データ},\text{確率項})\right)=\hat{\theta}^{SSI}\left(\text{観測データ},\text{確率項}\right)$)。つまり、不完全データに対して、欠測データに関する推定の不確実性を代入モデルの確率項で捉えようとしている。しかし、一般的な推定量$\hat{\theta}$に関して、(2-5-2)式を計算することは容易ではないという問題がある。また、確率的回帰代入法においては、推定された回帰モデルから生成する欠測データの、乱数としての不確実性は捉えられているが、回帰モデルの推定自体に伴う不確実性、すなわち観測データを所与としたときの欠測データの条件付分布に関わる不確実性は捉えられていない。

第 2.2 節で述べたとおり、単一代入法は、MAR の下であれば、１次モーメントに関する点推定については欠測バイアスを緩和できる。しかし、MAR の下でも、標準誤差や１次超のモーメントの推定については下方バイアスを伴う（このバイアスは欠測バイアスではなく、処理に由来するバイアスである）。これに対して、多重代入法は、（2-5-1）及び（2-5-2）式に示した分散分解に基づいて、またデータ生成の不確実性のみならずデータ生成過程自体に関する不確実性も考慮に入れて推定精度の評価を可能にする手法である。

○多重代入のたとえ話

多重代入法の正確な説明は本節後半部に示し、本節前半ではまず直感的な理解を目指す。分かりやすい図解がないので、やや散文的になるが、たとえ話で説明する。多重代入法は、図２－６に示すような処理である。

図２－６　多重代入法のたとえ話

多重代入法のたとえ話	実際
1. 不完全データをよくみる	・事後予測分布$f(Y^M\|Y^O,X)=\int f(Y^M,\delta\|Y^O,X)d\delta=\int f(Y^M\|Y^O,X,\delta)f(\delta\|Y^O,X)d\delta$をモデル化
2. （いかにも背後の完全データを生成しそうな）サイコロをひとつ作る（**不確実性1**）	・分布$f(\delta\|Y^O,X)$からの乱数発生で値$\delta^{(h)}$を得る
3. 作ったサイコロを振る（**不確実性2**）	・値$\delta^{(h)}$で評価した分布$f\left(Y^M\middle\|Y^O,X,\delta^{(h)}\right)$からの乱数発生で値$Y^M_{(h)}$を得る
4. 出た目を代入値としてひとつの疑似完全データができる	・疑似完全データ$\left(Y^O,Y^M_{(h)},X\right)$を得る
2～4をH回繰り返す	
5. H個の疑似完全データのそれぞれに分析を適用	・H個の疑似推定結果$\left(\hat{\theta}^{(h)},\hat{V}^{(h)}\right)_{h=1,\cdots,H}$を得る
6. H個の分析結果をRubin則に従って統合	$\hat{\theta}_{MI}=\frac{1}{H}\sum_{h=1}^{H}\hat{\theta}^{(h)}$ $\hat{V}_{MI}=W+(1+1/H)B$ $W=\frac{1}{H}\sum_{h=1}^{H}\hat{V}^{(h)},\quad B=\frac{1}{H-1}\sum_{h=1}^{H}\left(\hat{\theta}^{(h)}-\hat{\theta}_{MI}\right)^2$

不完全データをよく眺めたうえで、その不完全データの背後にある完全データを生み出しそうな“サイコロ”をひとつ作る。ここで“サイコロ”といっているのは、データ生成過程のことである。ここで、“サイコロ”（データ生成過程）というものに関して２通りの考え方がある。第１は、データの背後には真の“サイコロ”（データ生成過程）がただひとつ存在しており、データからそれについて推定しなければならないという世界観である。第２は、“サイコロ”（データ生成過

程）は、いわゆる「可能存在」であり、データに基づく限りで許される範囲における可能性の広がり、としてのみ捉えうるという世界観である。図２－６多重代入法のたとえ話における「“サイコロ”をひとつ作る」というステップは、後者の世界観で理解される。“サイコロ”の可能性の広がりが、データ生成過程に関する不確実性に対応する。図２－６の第２のステップで、いろいろな可能性のなかから無作為に選び出されたひとつの“サイコロ”を、図２－６の第３のステップで振る。この「“サイコロ”を振る」というステップが、欠測値（の真の値）が、ある特定のデータ生成過程から発生するときの不確実性に対応する。図２－６の第４のステップで、特定の偶然性をもつ疑似完全データがひとつ作成される。このような疑似完全データを、互いに独立に複数作成することで、疑似完全データの標本が得られる。図２－６の第６ステップの具体的な内容については、本節後半を参照のこと。

〇多重代入法の実行例

図２－７は、第 2.2.1 節の図２－２－１～２－２－８で用いた人工的な不完全データに対して、多重代入法の実行例を示したものである。ここでは図２－７に示した多重代入法のたとえ話では捨象されていた、補助変数の役割に焦点を当てる。

まず、図２－７中の①及び②の処理は、それぞれ“サイコロ”に関する不確実性、及びデータ発生の不確実性に対応している（図２－６の「“サイコロ”を作る」ステップが図２－７の①、図２－６の「“サイコロ”を振る」ステップが図２－７の②にそれぞれ対応している）。特に、欠測データの事後分布を特定するパラメータδ_h^*が、第h疑似完全データ作成用の“サイコロ”に対応する。ここで欠測が生じていないレコードの補助変数(x_{1i}, x_{2i})は、観測データy_i^Oとともに、“サイコロ”作成における投入要素となっている（補助変数には欠測は生じないが、図２－７では、目的となる変数に欠測が生じているレコードの補助変数(x_1^O, x_2^O)と目的となる変数に欠測が生じているレコードの補助変数(x_1^M, x_2^M)を区別している）。

図２－６の第１のステップで「不完全データをよくみる」のは、第２のステップで「サイコロをひとつ作る」ための情報収集であるが、そこでは欠測が生じていないレコード$(y_i, x_{1i}, x_{2i})_{i \in S^R}$だけが対象となっている。欠測が生じていないレコードを考慮して“サイコロ”を作るのであるが、出来上がった“サイコロ”を振るときには欠測が生じているレコードの補助変数の情報が利用される。たとえて言えば、“サイコロの振り方”は欠測が生じているレコードの補助変数(x_1^M, x_2^M)の値に依存する（現実世界のサイコロは、強く振るか弱く振るか、角度をつけるかといった振り方によって無作為性が変化するとは考えられないが、ここでは説明の便宜上振り方に応じて目の出方が変わってくる“サイコロ”を考えている。そ

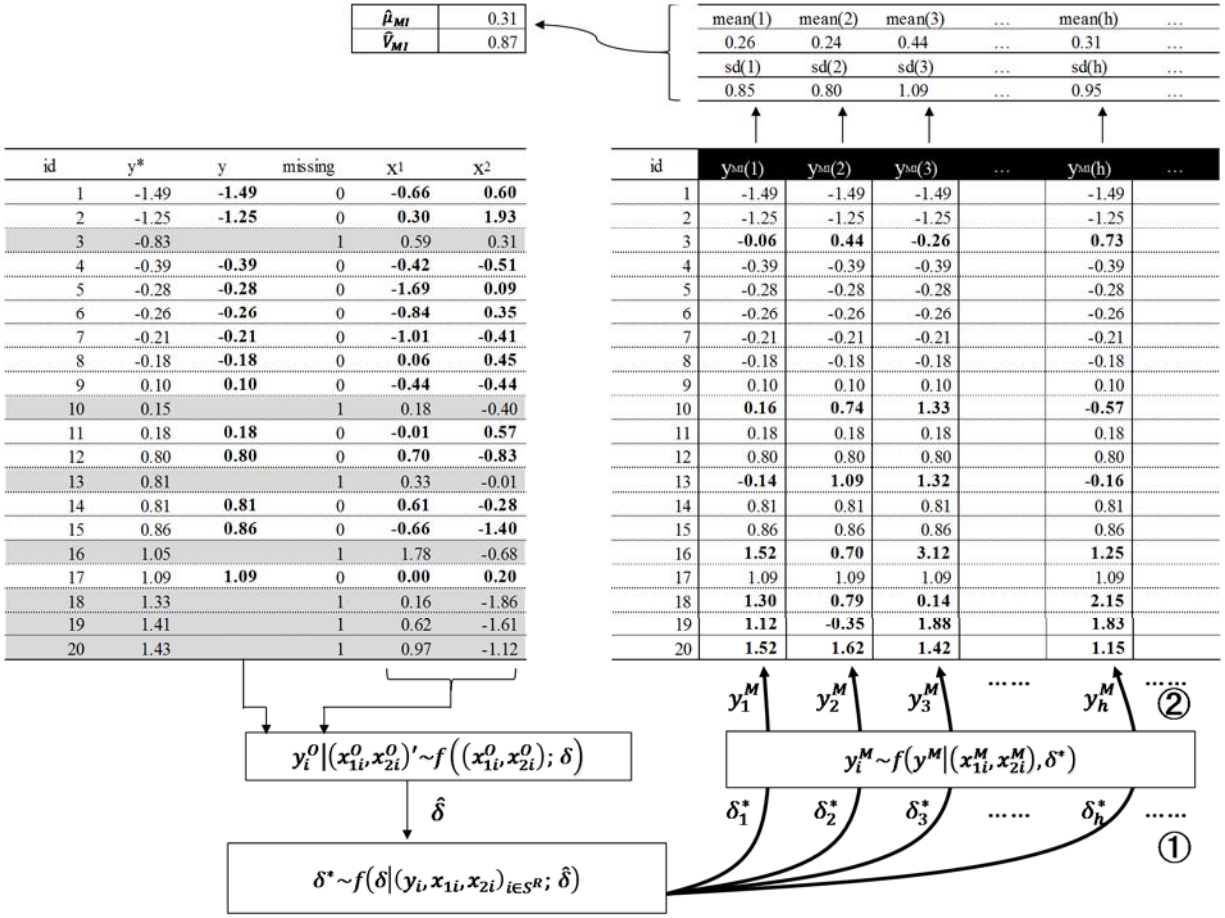
もそも正多面体のサイコロを考えているわけでもない)。

まとめると、多重代入法では、データ生成過程の可能性の広がりのなかから、互いに独立に複数の偶然性を取り出し、それらの個々の偶然性のそれぞれに対して解析を適用し、それらの個々の結果を統合することで、データ生成過程自体に関する不確実性と、データ発生に関わる不確実性を捉えた推定を行う。その際、補助変数に含まれる情報のうち、欠測値の推定に資する情報が活用される（この点は、単一代入法も同じである)。それらの情報は不確実性の範囲を狭めるといえる。欠測が生じていないレコードの情報は、データ生成過程自体の可能性の広がりに制約を課す役割を果たし、欠測が生じているレコードの補助変数の値は、データ発生の不確実性の範囲を狭める役割を果たす。

○図２－４－３の補足説明

第 2.2.2 節の図２－３－１～２－３－３には、単一代入法に加えて多重代入法の実行結果の一例を示す。ただしこの例示では、多重代入法によって作成される疑似完全データの数は１であり、本来多重代入法が想定する適用方法ではない。ここでは、多重代入法における代入値自体の特徴を確認するためにこの図を示す。結果は、確率的回帰代入と同様のものであることが分かる。補助変数が正負を問わず目的となる変数との相関を示す場合（図２－３－１及び２－３－２）は、（図から読み取れる情報に関する限り）真の姿（各図のパネル（A))に似た代入結果となるが、補助変数が目的となる変数と無相関である場合（図２－３－３）は、代入結果は真の姿から大きく乖離する。多重代入法においても、用いる補助変数が目的となる変数と無相関である場合は、欠測バイアスを緩和する効果が期待できない。

図２－７　多重代入法の処理手順



y*: 真の値、y: 観測データ、missing: 欠測指標、(x_1, x_2): 補助変数、y_{MI}: 多重代入法による代入値

2.6 尤度法

欠測データメカニズムが MAR である場合は、補助変数の情報を活用することで、推定における欠測バイアスを緩和することができる。これに対して、欠測データメカニズムが MNAR である場合は、補助変数の活用だけでは欠測バイアスを緩和できない（すなわち、欠測バイアスの緩和に資する補助変数が利用できない）。そこで、不完全データの背後にあるデータ生成過程をモデル化することで、欠測バイアスの緩和を図る方法として、尤度法がある。

不完全データの分析手法としての尤度法は、通常の最尤推定法を、欠測の生じるデータへ拡張したものである。最尤推定法では、データ生成過程をモデル化することで、データが発生する確率を導出し、データ発生確率をデータ生成過程のパラメータの関数とみなす。この関数は「尤度関数」と呼ばれる。最尤推定法は、与えられたデータの尤度関数を最大化するパラメータを推定量とする推定方法である。最尤推定法は、「発生したデータは最も高い確率で発生したものであろう」という推定原理に基づいている。最尤推定量が、一致性（標本サイズを増加させていくと、推定量が真の値に確率的に収束するという性質）、漸近正規性（標本サイズを増加させていくと、推定量の分布が正規分布に収束するという性質）、漸近効率性（標本サイズを増加させていくと、推定量の分散が理論的下限に収束するという性質）という望ましい性質をもつための十分条件が知られており、それらの条件のうち、モデル化が正しいという条件以外は緩やかな条件である。

不完全データの分析手法としての尤度法が、欠測の生じないデータに対する通常の最尤推定法と異なる点として、次の2つが挙げられる。第1に、不完全データの尤度関数は、データ生成過程のパラメータの関数であるだけではなく、欠測値の関数でもある。第2に、不完全データでは、欠測パターン自体がデータの構成要素となる。つまり、不完全データは、完全データとは異なる次元の情報を追加的に含んでいるため、不完全データのデータ生成過程ないし尤度関数は、完全データのデータ生成過程ないし尤度関数とは異なる次元の引数をもつ。

第1の点については、不完全データの尤度法では、尤度関数を欠測データに関して積分又は積算することによって、最尤推定法における最大化の目的関数を導出する。尤度関数を欠測データに関して積分又は積算するという処理は、「発生したデータは最も高い確率で発生したものであろう」という、最尤推定法の推定原理に即したものである。このことを、簡単な例によって示す。

硬貨を投げて表が出れば値 1、裏が出れば値 0 をとる2値変数を考える。2枚の硬貨、たとえば百円玉と五十円玉のそれぞれに、この2値変数を定義し、百円玉に対しては2値変数A、五十円玉に対しては2値変数Bとする。また2つの2値変数A及びBの和を変数Cとする（$C \equiv A + B$）。3つの変数A、B及びCのうち、任意の2つの値が分か

れば残りの値も分かるので、任意の2つの変数として、変数B及びCに注目する。当該百円玉で表が出る確率をα、当該五十円玉で表が出る確率をβとすると、2つの変数として変数B及びCの同時分布は、$\Pr(B=0,C=0)=(1-\alpha)(1-\beta)$、$\Pr(B=0,C=1)=\alpha(1-\beta)$、$\Pr(B=1,C=1)=(1-\alpha)\beta$及び$\Pr(B=1,C=2)=\alpha\beta$（また、$\Pr(B=0,C=2)=\Pr(B=1,C=0)=0$）である。2つの硬貨を100回投げて、各回の変数$B$及び$C$のデータを収集したとする。100回のうち30回は$(B,C)=(0,0)$、20回は$(B,C)=(0,1)$、25回は$(B,C)=(1,1)$、25回は$(B,C)=(1,2)$というデータが得られたとする。この欠測が生じていないデータに対する対数尤度関数（尤度関数の対数値）$\ln L^*$は、$\ln L^*=30\ln(1-\alpha)(1-\beta)+20\ln\alpha(1-\beta)+25\ln(1-\alpha)\beta+25\ln\alpha\beta$であるから、対数尤度関数を最大化する解の必要条件は、$\partial\ln L^*/\partial\alpha=-30/(1-\alpha)+20/\alpha-25/(1-\alpha)+25/\alpha=0$及び$\partial\ln L^*/\partial\beta=-30/(1-\beta)-20/(1-\beta)+25/\beta+25/\beta=0$となり、最尤推定の結果は、$(\hat{\alpha},\hat{\beta})=(0.45,0.5)$である。

次に、どうしたわけか変数Cの値が、一部の回について観測されなかった場合を考える。上記の試行結果で、4通りのパターンのそれぞれで、5回分について変数Cの値が観測されていないとする。この場合、100回のうち25回は$(B,C)=(0,0)$、15回は$(B,C)=(0,1)$、20回は$(B,C)=(1,1)$、20回は$(B,C)=(1,2)$、10回は$(B,C)=(0,NA)$、10回は$(B,C)=(1,NA)$という結果である（ちなみに、ここで変数Cではなく変数Bに欠測が生じていれば、定義上$C=0$ならば$B=0$であり、$C=2$ならば$B=1$であるから、事実上欠測を減らすことができる。このように欠測を減らすために利用できるデータ以外の情報源を「expert knowledge」と呼ぶ）。尤度を求める際、この欠測を含む20回分については、確率測度を欠測値に関して積算する。計算としては、$(B,C)=(0,NA)$となった10回分の各々については、変数Cが値 0、1、2 のそれぞれをとる可能性を考慮して、値$\Pr(B=0,C=0)+\Pr(B=0,C=1)+\Pr(B=0,C=2)=(1-\alpha)(1-\beta)+\alpha(1-\beta)+0=1-\beta$、また、$(B,C)=(1,NA)$となった10回分の各々については、変数$C$が値 0、1、2 のそれぞれをとる可能性を考慮して、値$\Pr(B=1,C=0)+\Pr(B=1,C=1)+\Pr(B=1,C=2)=0+(1-\alpha)\beta+\alpha\beta=\beta$が、それぞれの欠測値に関して積算された尤度となる。$(B,C)=(0,NA)$となった10回分のそれぞれの尤度$1-\beta$は、単に当該五十円玉で裏が出る確率であり、また、$(B,C)=(1,NA)$となった10回分のそれぞれの尤度βは、単に当該五十円玉で表が出る確率である。つまり、欠測値に関する積算（連続変数の場合は積分）という処理は、観測された変数のみの分布に基づいて尤度を求めることにほかならない。この欠測が生じているデータに対する対数尤度関数（尤度関数の対数値）$\ln L$は、$\ln L=25\ln(1-\alpha)(1-\beta)+15\ln\alpha(1-\beta)+20\ln(1-\alpha)\beta+20\ln\alpha\beta+10\ln(1-\beta)+10\ln\beta$であるから、対数尤度関数を最大化する解の必要条件は、$\partial\ln L/\partial\alpha=-25/(1-\alpha)+15/\alpha-20/(1-\alpha)+20/\alpha=0$及び$\partial\ln L/\partial\beta=-25/(1-\beta)-15/(1-\beta)+20/\beta+$

$20/\beta - 10/(1-\beta) + 10/\beta = 0$となり、最尤推定の結果は$(\hat{\alpha}, \hat{\beta}) = (0.4375, 0.5)$である。欠測値に関して積算（連続変数の場合は積分）した尤度関数は、「観測データ尤度（observed-data likelihood）」関数と呼ばれる。

第2の点は説明を要する。不完全データとそれに対応する完全データの相違を理解するためには、一見逆説的な次の事実が重要である。すなわち、不完全データは、それに対応する完全データの情報の一部に覆いを掛けたものに等しいが、不完全データにはそれに対応する、完全データには含まれない情報が追加的に含まれている。その追加的に含まれる情報とは、「覆いの掛けられ方に関する情報」、つまり欠測パターンに関する情報である。覆いの掛けられていない完全データでは、どのように覆いが掛けられる可能性が高いか、あるいはそもそも覆いが掛けられる可能性があるのか、ということ（つまり欠測パターンの確率分布）に関して、推定する手掛かりとなる情報は一切含まれていない。

不完全データの分析手法としての尤度法では、MNAR の欠測データメカニズムに対して、不完全データに追加的に含まれた「覆いの掛けられ方に関する情報」を、欠測パターンの分布に関する推定に資する情報として有効に活用する。もちろん統計調査の目的は、興味の対象となる変数の分布に関する推定であって、欠測パターンの分布に関する推定ではない。それでも本来の目的のために「欠測パターンの分布に関する推定に資する情報」が活用できるのは、欠測データメカニズムとして MNAR を想定するからである。MNAR のもとでは、欠測パターンがどのように発生するかということと、興味の対象となる変数の値がどのような値をとるかということとの間に相互依存関係があるため、欠測パターンがどのように発生するかということを推定する上で役に立つ情報は、興味の対象となる変数の値がどのように発生するかということを推定する上でも役に立つのである。

ここで、MAR と MNAR のそれぞれで活用する情報を比較すると、次のようになる。欠測データメカニズムが MAR である場合は、補助変数の情報を活用することで、興味の対象となる変数の値と欠測パターンとの間の相互依存性を取り除くことができるので、補助変数の活用だけで興味の対象となる変数に関する推定から欠測バイアスは除かれる。他方、欠測データメカニズムが MNAR である場合は、補助変数の値で条件付けてもなお興味の対象となる変数の値と欠測パターンとの間に相互依存関係が残るので、その残された相互依存関係をモデル化したうえで、補助変数の情報だけではなく上述の「欠測パターンの分布に関する推定に資する情報」を活用することで興味の対象となる変数に関する推定から欠測バイアスは除かれる。

次に、上で説明した不完全データの尤度法が、通常の最尤推定法と異なる2つの点（通常の尤度関数が欠測値の関数となること及び欠測パターン自体がデータとなること）を踏まえて、不完全データの尤度関数を導く考え方を説明する。不完全データは、それに対応する完全データの情報の一部に覆いを掛けたものに等しいので、不完全

データのデータ生成過程は、(1)それに対応する完全データのデータ生成過程と(2)完全データに覆いを掛ける確率的過程という2つのデータ生成過程が合成されたものとみることができる。前者を「興味の対象となるデータ生成過程」と呼ぶことにして、後者は「欠測データメカニズム」に他ならない。このようにみた不完全データのデータ生成過程を、「全データのデータ生成過程(generating process of full-data)」と呼び、全データのデータ生成過程から導かれる尤度関数を「全データ尤度(full-data likelihood)」関数と呼ぶ。

全データ尤度関数は、補助変数Xの値を所与としたときの興味の対象となる変数Yとその観測指標Rの条件付同時分布の確率密度(質量)関数に等しい。従って、尤度法におけるモデル化は、当該同時分布の特定化である。この同時分布自体を特定化することは可能であるが、その場合、欠測データに関する積分又は積算という処理によって、最尤推定における最大化の目的関数となる観測データ尤度関数が複雑になることに注意を要する。全データのデータ生成過程は、興味の対象となるデータ生成過程と欠測データメカニズムを合成したデータ生成過程であるとみなしたとき、全データ尤度関数は、それを構成する2つのデータ生成過程のそれぞれに対応する尤度関数に分解することができる。興味の対象となるデータ生成過程と欠測データメカニズムをそれぞれモデル化して、それぞれのモデルから導かれる尤度関数に全データ尤度関数を分解する場合のモデルは、「選択モデル」と呼ばれる。MNAR の下では、興味の対象となるデータ生成過程と欠測データメカニズムの間に相互依存関係がある。選択モデルによる全データのデータ生成過程のモデル化においてこの相互依存関係を表す方法の一例として、興味の対象となるデータ生成過程のモデルの誤差項と欠測データメカニズムのモデルの誤差項の同時分布を特定化するという仕方がある。この場合、2つの誤差項の相関が0であれば MAR のモデルとなる。

○Heckman の選択モデル

欠測バイアスへの対処としての尤度法の好例として(ただし無回答による欠測ではなく、値が原理的に観測されないことによる欠測ではあるが)、Heckman の選択モデルによる賃金関数の推定がある。一般的に、労働者ごとに労働市場で提示される賃金は、労働者の学歴、職歴、年齢といった属性の関数である。この関数を特に「賃金関数」と呼ぶ。標本調査によって若年女性の賃金関数を推定したい場合、標本に選ばれた調査客体ごとに、学歴、職歴、年齢といった属性と、労働市場で提示される賃金の値をデータとして収集しなければならないが、若年女性のすべてが実際に労働市場に参加しているわけではないので、一部の調査客体については「労働市場で提示される賃金」(以下「提示賃金」)は観測されない。提示賃金の欠測は、無回答によるものではなく、原理的な観測不能性によるものである。提示賃金が観測されている調査客体のデータだけを用いて賃金関数を推定した場合、推定結果は、「若年女性の賃金関数」に

関するものではなく、「働いている若年女性の賃金関数」に関するものである。

ここで、標準的なミクロ経済学理論、つまり、労働によって所得を得て消費と余暇から効用水準が決まる家計による最適化問題の解として、提示賃金が留保賃金を上回る場合に働き（労働供給が正となり）、上回らない場合は働かない（労働供給は0となる）という行動が導かれる。留保賃金は経済主体の効用関数によって決まるので、モデル化する場合は、留保賃金を当該経済主体の効用関数の決定要因（たとえば家族構成、不労所得、資産水準など）の関数とみなす。まとめると、当該標本調査のデータ生成過程のモデルは次式で表される。

$$\text{提示賃金} = h(\text{学歴}, \text{職歴}, \text{年齢}, \cdots) + \text{賃金関数の誤差項}$$

$$\text{留保賃金} = g(\text{家族構成}, \text{不労所得}, \text{資産}, \cdots) + \text{留保賃金の誤差項}$$

$$\text{提示賃金の観測指標} = \begin{cases} 1 & (\text{提示賃金} > \text{留保賃金}) \\ 0 & (\text{提示賃金} \leq \text{留保賃金}) \end{cases}$$

このモデルの誤差項にパラメトリックな分布を仮定することで尤度関数が導かれ、最尤推定を行うことができる。

Heckmanの選択モデルによる賃金関数の推定では、提示賃金の観測の成否が、労働市場への参加の有無によって決まるが、通常の統計調査における無回答による欠測についても応用できる。その場合、無回答に関する意思決定の理論モデルがあれば、欠測データメカニズムに理論的な基礎付けが得られたことになる。標準的な経済学の原理によれば、「回答することから得られる便益≦回答することの機会費用」という条件が、無回答となる必要十分条件となる。回答することから得られる便益は、社会的規範や調査協力への謝礼が考えられる（現実には前者の方が大きい割合を占めている）。回答することの機械費用は、回答する時間や労力である。回答の便益と費用は調査客体ごとに異なり、例えば調査客体が企業であれば純便益（便益－機会費用）は企業規模や業種等の属性の関数であり、調査客体が個人であれば純便益は所得や年齢等の属性の関数と考えられる。この関数形を適当に決めれば、上述の賃金関数の場合と同様に、欠測データメカニズムのモデルが得られる。ただし実際に尤度法を適用する場合には、調査客体の行動モデルを明示的に考えずに選択モデルを便宜的に用いることもあり、それは理論的な基礎付けを欠くことになる。