



Title	Statistical Modeling for Unobserved Variables and Incomplete Data
Author(s)	高井, 啓二
Citation	大阪大学, 2009, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/49608
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 ＜a href="https://www.library.osaka-u.ac.jp/thesis/#closed">https://www.library.osaka-u.ac.jp/thesis/#closed >大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	たか い けい じ 高 井 啓 二
博士の専攻分野の名称	博 士（工 学）
学 位 記 番 号	第 2 3 0 2 3 号
学 位 授 与 年 月 日	平 成 21 年 3 月 24 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 基礎工学研究科システム創成専攻
学 位 論 文 名	Statistical Modeling for Unobserved Variables and Incomplete Data （非観測変数と不完全データに対する統計的モデリング）
論 文 審 査 委 員	（主査） 教 授 狩 野 裕 （副査） 教 授 白 旗 慎 吾 教 授 名 和 範 人

論 文 内 容 の 要 旨

本論文では、統計学における潜在変数、欠測値、そして最適化アルゴリズムについて研究を行なった。潜在変数、欠測値は共に直接観測できない量であるという共通の観点から、数学的には統一的な議論を展開することができる。

欠測値に関する研究では、データが欠測値を含むとき、統計的に正しい結論を得る方法を定式化した。欠測値とは、データ収集の際に観測されなかった値である。欠測値の発生原因は、様々あるが、特に欠測値が自分自身の値に依存して生じる場合がある。例えば、体重の調査では体重が重い人は重い故に答えない。この場合には、正しい結論を得るためには欠測値を生じさせる機構をモデル化しなければならない。本研究では、このような欠測値のある分割表（クロス集計表）における独立性の検定に関する問題を解決した。具体的には、(i) 欠測機構のグラフィカルモデルによるモデル化とその問題、(ii) パラメタの識別性・推定法（最適化アルゴリズム）、である。

また、この研究で得られた推定法を使い、仮説下(制約式下)での潜在変数を含む統計モデル(潜在クラスモデル)のパラメタ推定に適用した。データ解析では、解析者の仮説が定数制約や線形制約という形で、表現される。潜在クラスモデルのパラメタ推定では、このような制約の下での最適化を行うことが難しかった。本研究では、上の欠測値の問題で使用した推定法を用いることにより、最適化を容易に可能にした。

潜在変数に関する研究では、時系列因子分析における因子不確定性の問題を解決した。潜在変数(因子とも呼ぶ)とは、実際は観測されないが、データの解釈のために想定される確率変数である。時系列因子分析は、時系列データの変動を潜在変数によって説明する。例えば、各家電会社の株価(観測できる)の変動を、家電業界の景気(直接は観測できない)によって説明する。この分析法には、潜在変数に一意的に値を与えられないという因子不確定性の問題があった。本研究では、ある条件下で、因子不確定性が消滅することを示した。

論 文 審 査 の 結 果 の 要 旨

サイエンスにおいて実証研究は重要な位置を占める。実験や調査を繰り返しい行い結果を記述的に要約・表示することは常套手段である。ランダムネスが大きい場合は確率的評価、すなわち統計的推測が不可欠になる。実験や調査において得られるべきデータが得られないことがあり、これを欠測または欠損という。欠測は記述的であれ推測的であれ分析結果とその解釈を歪めることが知られている。近年、欠測値を含むデータ（不完全データという）から適切に情

報を引き出すための方法論が注目されている。欠測値は直接観測されない潜在変数と考えることもできる。潜在変数を扱う潜在変数モデルは社会科学を中心に発展してきた統計モデルであるが、欠測値問題とも密接に関連する。

本論文は、特に離散型データ解析における欠測値問題と潜在変数モデル、そして、不完全データに対する離散型モデルに内在する数値的な困難を研究対象とする。欠測には比較的扱いやすい無視可能と呼ばれる場合とそうでない無視不可能の場合がある。本論文は第 2 章と第 3 章において、無視不可能な欠測値を含む 2×2 分割表の統計的推測方法に貢献をした。グラフィカルモデルによって欠測メカニズムを規定し独立性の検定統計量を構成した。最尤推定値を解く際、数値的な困難に直面しそれは多くの文献で指摘されていることを知る。そして、最尤推定値を求めるための有力な手段である EM アルゴリズムを改良することで、この問題において最尤推定値を安定的に求めることに成功した。第 4 章では、前章で開発したアルゴリズムを制約のある潜在クラスモデルの推定問題に応用し、ここでも、最尤推定値が安定的に求められることを示した。第 5 章では潜在変数モデルの根源的な課題である因子不確定性の問題が時系列因子分析モデルにおいても重要であることを指摘し、因子が確定するための数学的条件を提示することに成功した。

統計科学におけるこれらの貢献は、博士（工学）の学位論文として価値のあるものと認める。