論 文 の 内 容 の 要 旨

論文題目 Algebraic Statistical Methods for Conditional Inference of Discrete Statistical Models
（離散統計モデルの条件付き推測問題に対する代数統計的手法）

氏　　名　　小川　光紀

統計的推測問題では，十分多くのデータが得られる場合に適用可能な漸近理論に基づく手法が，データがスパースであるなどの理由で適切でない場合がある．そのような場合の解決法として，条件付き尤度に基づく方法が挙げられる．しかし，この方法を実行しようとすると，条件付き尤度の規格化定数の取り扱いが計算量の観点から難しくなる場合が多い．この問題に対応するために，本論文ではいくつかの離散統計モデルの条件付き推測問題を対象として，代数統計的手法を構築した．

代数統計学は，可換環論，組合せ論，代数幾何学などの理論を用いて統計学の諸問題の解決を目指す分野横断的学問領域である．この分野は，DiaconisとSturmfels (1998)の論文や，PistoneとWynn (1996)の論文を契機として急速に発展している分野である．これら二つの論文では，いずれも多項式環におけるグレブナー基底の理論を利用している点が特徴的である．特に前者では，条件付き分布に基づく検定をマルコフ連鎖モンテカルロ法により実行する際に有用なマルコフ基底が，数学的にはトーリックイデアルの生成系と等価であることを示した点が重要である．トーリックイデアルとは，配置行列という行列から定まる多項式環のイデアルである．この事実から，具体的な統計モデル(トーリックモデル)に付随するトーリックイデアルの生成系を求める問題が盛んに研究されている．

本論文の第2章および第3章では，分割表およびランダムグラフの具体的な統計モデルに対して，条件付き分布に基づく検定に必要なマルコフ連鎖の構成法について，DiaconisとSturmfelsの枠組みを用いて議論した．マルコフ基底やそれを含む集合であるグレイバー基底を求める計算は，原理的には計算ソフトウェアを利用することで実行可能である．しかし，一般的なアルゴリズムによる計算は計算量の観点から現実的でなく，具体的な統計モデルごとに必要な数学的道具を理論的に導出する必要がある．

第2章では，二元分割表の統計モデルとして，独立構造に副表の効果を追加した統計モデルを考え，対応するマルコフ基底の構造を解析した．二元分割表の独立モデルにつ

いては，2次のmoveの集合がマルコフ基底をなすことが基本的事実として知られている．しかし，副表の効果を一つでも加えると，対応する統計モデルのマルコフ基底の構造は非自明なものとなることがある(Hara et al. (2009))．先行研究のHara et al. (2009)では一つの副表の効果を加味した場合について議論されていた．本章では，より一般に複数の副表の効果を加味した場合に対して，実用上自然ないくつかの設定の下でマルコフ基底の具体形を導出した．

第3章では，ランダムグラフの基本的な統計モデルであるベータモデルに対して，条件付き尤度に基づく検定手法を構築した．その枠組みはDiaconisとSturmfels (1998)のマルコフ連鎖モンテカルロ法に基づくものであるが，条件付き分布から観測され得るグラフに制約を伴うため，通常のマルコフ基底のみでは既約なマルコフ連鎖を構成できない場合がある．本章では，マルコフ基底を包含するグレイバー基底を利用する方法を構築した．そのために，無向グラフに付随するトーリックイデアルのグレイバー基底に対するグラフ上での特徴づけを与え，グレイバー基底の元をランダムに生成するアルゴリズムを構築した．

以上が具体的な統計モデルに対するマルコフ基底の構造解析に関する結果である．一方，トーリックイデアルの理論では，元となる配置行列から新たな配置行列の系列を作り，その操作によって代数的性質がどのように維持されるかに興味がもたれることがある．その中で重要な例が，ローレンス持ち上げである．この操作は，統計モデルの観点からも重要な意味をもち，ローレンス持ち上げのマルコフ基底およびグレイバー基底の複雑さを測る量として，マルコフ複雑度およびグレイバー複雑度という量が提案されている(SantosとSturmfels (2003))．

本論文の第4章では，配置行列のファイバーを新たな配置行列ととらえ，そのマルコフ基底の性質について議論した．すなわち，ファイバーに対するマルコフ基底の次数の上界を与え，具体的ないくつかの配置行列に対して，ファイバーのマルコフ基底の次数の値や元の配置行列のマルコフ複雑度の値について議論した．具体的な例の中には，輸送多面体に付随する配置行列のように，数学的に重要なものも含まれている．

統計的推測の基本的な問題設定として，パラメータの推定問題がある．通常の最尤推定を行うためには，パラメータの数に対して十分多くのデータが得られていることが必要であるが，主効果は層ごとに異なるがオッズ比は共通であるモデルのようにNeyman-Scott問題の状況に陥る場合には，最尤推定量が一般的には一致性も有しない．また，上述の条件付き分布に基づく検定を行って帰無仮説が棄却された際には，対立仮説におけるパラメータの推定手法として条件付き尤度に基づいた手法が自然な場合がある．これらの問題を動機付けとして，本論文では条件付き最尤推定の問題も取り扱った．

すなわち，本論文の第5章では，トーリックモデルを基礎として，さらに興味あるパラメータに関するモデル構造を追加した形の統計モデルを考え，そのパラメータ推定の

問題を扱った．基礎とするトーリックモデルのパラメータの影響を，対応する十分統計量を所与とした条件付き尤度を考えることによって除去し，興味あるパラメータのみを推定する手法について議論した．その際，条件付き尤度の規格化定数の計算が問題となるが，本論文では規格化定数をA-超幾何多項式を介して扱うことにより，様々な計算手法が適用できることを示した．特に二元分割表の独立モデルを基礎とする場合について，級数展開法，ホロノミック勾配法，差分版のホロノミック勾配法の三種類の方法について論じた．