論文の内容の要旨

論文題目　Exact methods and Markov chain Monte Carlo methods of conditional inference for contingency tables
（正確法およびマルコフ連鎖・モンテカルロ法による分割表の条件付推測問題の解法）

氏名　青木 敏

分割表に集約されるデータに対する統計的推測は，伝統的に，漸近分布論に頼る方法が主流であるといえる．例えば，尤度比検定統計量のカイ二乗近似や，未知母数の最尤推定量の正規近似などは，連続量を含む広いクラスのデータ解析に用いられる一般的な手法であるが，分割表データの解析においても幅広く用いられている．一方，分割表データ解析における特有の問題として，十分大きい標本が得られないような場合，あるいは，標本数はある程度大きくても，観測値が疎な分割表となるような場合には，漸近分布論の当てはまりが不良となることが古くから指摘されている．臨床医学研究，特に癌のような稀な疾患を扱うような状況は，疎な分割表が観測される典型例であり，このような場合には，漸近分布論に頼らない解析手法が重要となる．以上のような状況を念頭におき，本論文では，漸近分布論に頼らない分割表解析の手法のいくつかの既存の研究に対し，その改良と新たな提案を行なう．

漸近分布論に頼らない分割表解析の手法は，歴史的には，R. A. Fisher により 1934 年に提案された 2×2 分割表の正確検定が最初である．Fisher 以降，その考え方は一般化され，一般のサイズの 2 元分割表，さらには 3 元以上の高次分割表のさまざまな統計的推測の問題に対して，正確法が提案されている．これらの手法の多くは，条件付推測問題の枠組，具体的には，さまざまな相似検定に対する，有意確率の正確計算の手法として提案されており，局外母数に対する十分統計量の値を条件付けたもとでの標本空間の要素を，いかにして効率的に列挙するか，が，計算時間の上で重要となる．本論文の前半では，正確法の手法を扱ったいくつかの既存の研究に対し，その改良と新たな提案を行なう．

1983 年に Mehta and Patel によって提案されたネットワークアルゴリズムは，一般化 Fisher 正確検定の有意確率の計算アルゴリズムである．これは，2 元分割表に対する行と列との独立性の仮説の正確検定に対して，現在，もっとも広く用いられている正確計算アルゴリズムのひ

とつである．ネットワークアルゴリズムは，分割表のセル頻度を列ごとに順次条件付け，残りの部分分割表に対する統計量の値の最大値，最小値を評価することにより，標本空間を効率的に縮小（枝刈り）するものである．この最適化問題の評価は，ネットワークアルゴリズムの本質であるといえる．本論文では，この最大化問題の最適解に対する新たな上界を提案し，その性質を調べるとともに，提案する手法によってより効率的に枝刈りが行なわれることを，数値計算によって確認する．提案する上界は，与えられるデータの種類に関わらず良い性質を持ち，情報幾何学の観点から説明付けられるものである．

さらに，提案する上界は，類似の任意の正確検定に対応する最適化問題に適用可能である．本論文では，集団遺伝学において重要な役割を持つ，ハーディー・ワインバーグ平衡仮説の正確検定を取り上げ，有意水準の正確計算のためのネットワークアルゴリズムを定式化し，その最大化問題の上界評価として，先ほどと同様の手法を適用する．また，この問題に関しては，最小化問題の最適解を陽に求めることができることを指摘する．論文では，これらの結果を用いて構築したネットワークアルゴリズムの有効性を，数値計算によって示す．遺伝解析の研究においては，本論文で提案するような組合せ論的手法はあまり知られておらず，本論文の結果から，さらなる問題が提起されること，また，より大量の遺伝データの解析が可能になることが期待される．

一方，ネットワークアルゴリズムのような，効率的な正確計算アルゴリズムの構築が困難な問題に対しては，正確法の代わりにモンテカルロ法を用いた有意水準の不偏推定が有効となる．本論文の後半では特に，マルコフ連鎖・モンテカルロ法による推定を扱う．この際に重要となるのが，パラメータに対数線形性を仮定した階層モデルの，分解可能性の概念である．分解可能でない帰無分布からのサンプリングは，一般には容易でなく，マルコフ連鎖・モンテカルロ法のための連結なマルコフ連鎖の構成も困難であることが知られている．分解可能でないモデルの最も簡単な例は，3 元分割表における無 3 因子交互作用の仮説であるが，この場合も，すべての 2 次元周辺度数が任意に固定された 3 元分割表の空間に，連結なマルコフ連鎖を構成するのは，一般には容易でない．

この問題は，Diaconis and Sturmfels の 1998 年の論文で，任意に固定された周辺度数に対して連結な連鎖を構成するために必要な基底（マルコフ基底）の導出，という形で定式化された．彼らは，多項式環上のあるイデアルの生成元がマルコフ基底に対応することを示し，その計算のためには，初項イデアルの生成系（グレブナ基底）を求めることができれば良いことを示唆した．この結果により，計算時間の問題を無視すれば，既存の代数計算アルゴリズムによって，理論上，任意のサイズの問題に対するマルコフ基底が求められることになるため，画期的な研究であるといえる．しかしこの代数算法には，計算時間が現実的でないという問題に加え，被約グレブナ基底がマルコフ基底としての冗長な出力を数多く含むこと，さらに，本来は対称性がある問題に対しても，固定した項順序に依存する非対称な出力となること，などの問題点があることが指摘されている．特に，計算時間の問題は重要であり，現在においては，現実的に実行可能な問題のサイズは，かなり小さなものに限られてしまう．これに対し本論文では，いくつかの問題に対するマルコフ基底を，代数アルゴリズムを用いずに導出した結果を紹介する．さらに，マルコフ基底の極小性とその一意性に関するいくつかの結果も紹介する．

まず，本論文では，応用上きわめて重要な，2 次元周辺度数が固定された 3 元分割表の空間

に連結な連鎖を構成する，という問題を扱う．本論文では，各軸の水準数が比較的少ない場合，具体的には，$3 \times 4 \times K, 4 \times 4 \times 4$ 分割表に対する極小基底の具体形を導出する．本論文のアプローチは，代数アルゴリズムを用いない初等的なものであり，その適用範囲は限られるが，得られたこれらの結果から，より大きなサイズの表に対するマルコフ基底の要素の一部を導出する指針も同時に与える．論文では，得られたマルコフ基底の膨大なリストを与えており，この結果は，3 元分割表解析のさまざまな問題に対して幅広く適用されることが期待される．さらにこの問題は，計算代数学においても興味深い問題のひとつとして捉えられており，本論文が，統計学と計算機代数学の橋渡しとなり，新たなアルゴリズムの構築や問題提起が期待される．

また，行と列を固定した 2 元分割表からのサンプリングは容易である一方，2 元分割表が構造的ゼロセルを含む場合には，マルコフ基底の構成は自明でない．本論文では，この問題についても，極小マルコフ基底の具体形を導出する．この結果は，二部グラフの理論と密接に関連しており，やはり統計学と代数学，組合せ論の橋渡しとなっている．

さらに，理論面での結果として，極小マルコフ基底の構造と，極小マルコフ基底が一意的に存在するための必要十分条件，および，各軸における水準の入れ替えを対称群の作用として定式化したときに，群の作用に関するマルコフ基底の不変性と，不変極小マルコフ基底の構造，不変極小マルコフ基底が一意的に存在するための必要十分条件，などを与える．これらの結果もまた，代数アルゴリズムを用いない，初等的な考察から得られるものである．

マルコフ基底の構造や，その極小性，一意極小性に関しては，未解決な問題が数多く存在する．これらの未解決問題については，本論文の最後で言及する．例としては，先ほどの，2 次元周辺度数が固定された 3 元分割表に関するマルコフ基底の構成問題に対して，$3 \times 4 \times K, 4 \times 4 \times 4$ 分割表については，一意極小基底の存在が証明されたが，一般の $I \times J \times K$ 分割表に対しても同様な一意極小基底が存在するのか，という問題がある．同様に，与えられた問題に対し，マルコフ基底を具体的に計算することなく，極小マルコフ基底の一意性を判定するアルゴリズムの構築にも興味があるが，これも現段階では未解決である．不変性との関連からは，すべての一意極小マルコフ基底は同時に一意不変極小マルコフ基底であるため，極小マルコフ基底が一意的に定まらず，かつ，一意不変極小マルコフ基底が存在する例が存在するかどうかに興味があるが，現在のところ，そのような例は，自明な例を除いては見付かっていない．また，より高次の問題に対する極小基底の導出も，重要な問題である．本論文では，3 元分割表に対する結果を拡張して，水準数がすべて 2 であるような 4 元分割表に関して，すべての階層モデルに対する極小マルコフ基底，不変極小マルコフ基底を計算し，その具体形を与える．