様式 C-19

科学研究費補助金研究成果報告書

平成　21 年　6 月　12 日現在

研究種目：若手研究（B）
研究期間：2007～2008
課題番号：19700269
研究課題名（和文）　汎用証明支援システムを用いた多次元分割表における群構造に関する研究
研究課題名（英文）　Group structure of n-way contingency tables with a proof assistant system

研究代表者
小野　陽子(ONO YOKO)
新潟国際情報大学・情報文化学部・講師
研究者番号：60339140

研究成果の概要：本研究の目的は、対話型汎用証明支援システムである Isabelle を用い、群構造を考慮した多次元分割表の性質を明らかにすることである。特に、これまでに蓄積されている群論等の証明成功過程ライブラリを知識として用い、分割表の数え上げに関する問題を群の問題として置き直した命題を作成し、この命題を計算機上で Isabelle を用いて証明を行った。この証明とは、計算機を集中的に酷使する全列挙的なものではなく、論理構造を辿るものである。

交付額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	1,700,000	0	1,700,000
2008 年度	1,300,000	390,000	1,690,000
総　計	3,000,000	390,000	3,390,000

研究分野：総合領域
科研費の分科・細目：情報学・統計科学
キーワード：計算機集約的統計、汎用証明支援システム、形式化手法、半自動証明、分割表

１．研究開始当初の背景

(1)統計学の分野では、分割表の数えあげに関する研究が盛んに行われており、Markov Chain Monte Carlo 法を用いた 2 次元分割表問題に関する様々な成果が報告されてきた。しかし、多次元の分割表に関する性質は、厳密な証明が与えられていなかった。また、2 次元分割表の数えあげ手法の中には、人間の直観理解を得やすい手続き的な手法が提案されており、多次元への拡張を考慮する上で、手続きだけでない厳密な証明構造を明らかにする必要があった。

(2)形式化手法とは、システムの仕様記述、開発、検証の技術であり、論理学や離散数学等を基盤としている。CPU 回路設計などのハードウェア検証、暗号プロトコルの安全性確

認といった、高度な信頼性を求められるシステムで利用されている。また、形式言語による記述では数学的な表記ができるよう記法が定義されており、論理的な記述により事象を明確に定義できることから、開発工程でエラーが入り込まないことが保障されている。自然言語での記述に見られる曖昧さを排除しているとも言えるだろう。また、近年では抽象数学の証明支援などにも利用されている。Isabelle/HOL は形式手法言語のひとつである。本研究では、Isabelle/HOL を統計分野の問題である分割表数え上げの証明に適用し、厳密な証明を与えることを試みた。

(3) 海外、特に EU 圏では、Isabelle や Mizar などの汎用証明支援システムに関する研究に対し、多くの基金が投入されている。数学者が情報学研究を行い新たな技術の発展に関与する一方、情報学者が数学を論じ、計算機を酷使した証明を行うことで数学の新たな証明を導く、といった学問分野の融合が行われつつある。しかし、残念ながら日本ではこれらのテーマに関する研究が未だ活発に行われているとは言えないのが現状であった。また、国内外を問わず、これらの研究グループは Isabelle や Mizar といった個々のシステムの周辺にのみ点在しており、システムを越えた共同・協調研究は見られず、各々のシステムに依存した証明成功例のライブラリ化・データベース化に留まっていた。

また、汎用証明支援システムによる証明の対象の多くは抽象数学であり、統計学に関する証明を扱った例は見られなかった。このことは、統計学にて扱う証明が汎用証明支援システムを利用するに不適切であるとの理由からではない。人間が証明を行う際に用いると言われている直観を含まざるをえず、計算機証明として扱うことが困難であったためである。人間は直観と過去の証明成功の蓄積を利用し、数学を証明するが、現状の計算機システムでは直観をそのままに取り込むことは不可能に近いこともその要因のひとつと考えられた。他分野が有する技術を利用し、小規模ながらもある意味での分野の垣根を越えた連携研究を行い、統計学における問題のひとつを解決したいと考慮し、本研究を申請した。

(4) 申請者は、2 × 2 ×・・・×{1, 2, …, m} 分割表 (yes/no データに被験者のプロフィールを併せたアンケートデータの解析などに用いられる) について、パスカップリングを用いたランダム生成法を提案した。
この提案法の収束時間が、
・分割表の列数
・分割表中の数値の総和に対して自然対数をとった値
・求めようとする精度の逆数に対して自然対数をとった値
の多項式で抑えられることを証明している。この成果を情報として、計算機上の証明成功過程ライブラリに追加し、証明への利用を検討した。

２．研究の目的

本研究の目標は、対話型汎用証明支援システムである Isabelle を用い、群構造を考慮した多次元分割表の性質を明らかにすることである。そのために、明らかにすべき点として次の事柄を鑑みた。

(1) 2 次元分割表に関する手続き的数えあげ手法に対して厳密な証明を与えることを第 1 の目的とした。Foulkes’ lemma と呼ばれる数えあげ手法に着目し、分割表を行列と捉え、descent と呼ばれる条件を満たす置換行列と

分割表から得られる行列が１対１対応をしていることを、Isabelle を利用して証明を行うことで厳密な計算機証明を与え、その論理構造を解明することを目指した。

(2)手続き的数えあげ問題の 3 次元以上への拡張を第 2 の目的とした。この拡張は、目的(1)の研究成果を元に行われるものであり、これまでに行った、統計学での多次元分割表全列挙問題に関する研究結果から証明の経路を抽出することで、人間がいかに証明の道筋を作成しゴールへ到達するか、そこに規則性が存在するかという事柄を鑑み、新しい知識ベースの提案を視野に入れている。直観が果たす役割を明確にすることが、本研究の鍵であると推察される。

(3) 分割表の数え上げに関しては、その多くが Markov Chain Monte Carlo 法もしくは代数的アプローチによるものである。また、本研究では群論を基礎として用いるが、このこと自体は特異的なことではない。本研究では計算機を用いて証明を行うが、計算機を集中的に酷使する全列挙的なものではなく、厳密な証明を計算機上にてひとつずつ積み上げいく手法であった。このような証明手法はこれまでこの分野において見られないものである。計算機による厳密証明の適用可能性を探索することを第 3 の目的とした。具体的には、人間の証明過程と計算機の証明過程の差異から具体化することを考慮した。

３．研究の方法

(1)基礎数学等の証明成功過程のライブラリ化・データベース化

汎用証明支援システム上での証明成功過程のライブラリならびにデータベースがなければ、必要な部品を構築しながら証明を進めなければならず、証明の道筋を見失う結果となる。そのための準備として、群論、集合と位相、線形代数（特に行列や置換行列）などの基礎的な数学に関する定理を、Isabelle を用いて証明し、これらのライブラリ化とデータベース化を推進する必要があった。そのため、国内外の研究者と連携を図り、データベース構築の一員として作業を行った。

(2)Foulkes’ lemma の Isabelle による厳密証明

Foulkes’ lemma による 2 次元分割表の数えあげ手法は、研究目的(1)(2)に示したように、人間が直観的に理解し、手続き的に数えあげを行うことは容易であるが、その証明を計算機上での証明手法にそぐうように形式化することは難しい問題である。計算機はパターンマッチングを利用して蓄積された知識を組み合わせ、論理の整合性を判定した証明を行うため、計算機が持つ知識の型と完全合致するように、人間が表現しなければならないためである。本研究では、Foulkes’ lemma を形式化し、Isabelle を用いて証明を行うことで、厳密な証明を与え、人間が直観的に処理を行っている箇所の洗い出しを行った。その結果から、人間による証明と計算機による証明の差異を明確化し、分割表の数えあげ問題に関する理論的問題点の探索を試みた。

(3)2 次元分割表における形式的手法の拡張検討

研究の方法(2)の結果から、3 次元への拡張を検討するにあたり、従来の手法を一般次元へ拡張できない原因を、人間による証明と計算機による証明の違いにあると予測し、両者の比較を行った。

４．研究成果

(1)基礎数学等の証明成功過程のライブラリ化・データベース化

Foulks lemmaを形式化し、証明を行う上で必要なデータベースを準備した。このとき、次の２つの事柄を用意した。

① 数学の基礎的道具の定義

例えば、整数の区間に関する次の命題は基礎的道具のひとつである。

"[[a < y; y ≤ a + b]]
==⇒ ∃ s ∈ {1..b}.y = a + s"

このように、人間が見て明らかと言えるものであっても、計算機に用意しなければならない。また、類似した定義が用意されていても、利用者が求める型と完全一致していない場合は、自ら作成し定義した方が証明に便利なこともある。

②行列、置換および置換行列などの簡単な性質の定義

Isabelle にある数学データベースだけでは、行列などの定義が不足しているため、次のような lemma の準備も必要である。これは、f を N 文字の置換とし、P, Q を f により決められる N×N 行列と仮定したとき、P と Q は一致することを命題としたものである。

"[[permutation N f;
permutation_matrix N f P;
permutation_matrix N f Q]]
==⇒ P = Q"

(2)Foulkes’ lemma の Isabelle による厳密証明

準備したデータベースを利用し、証明を行った。証明の概要は次の２つに大別された。

①分割表を行列として与えた時に、その行列から対応する置換行列を作成することができることを示した。この命題は次のように形式化を行った。

"T_to_Perm_block i j T r c P ==
(∀ x ∈ (horizontal_strip i r).
∀ y ∈ (vertical_strip j c).
(if (∃ s ∈ {1..(T i j)}.
x = (Σk = 1..(i − 1).(r k))
+ (Σk = 1..(j − 1).(T i k)) + s ∧
y = (Σl = 1..(j − 1).(c l))
+ (Σl = 1..(i − 1).(T l j)) + s)
then (P x y = 1) else (P x y = 0)))"

②分割表から得られた行列と対応する置換行列が 1 対 1 であることを示した。次は、その命題を形式化したものである。

“[[n_matrix m n T; permutation N f; N m r;
permutation_matrix N f P; composition_n
composition_n N n c; P_to_Table N m n r c P T; D n c dc;
Comp_to_D m r dr;
Comp_to_descents N f ⊆ dr `{1..(m − 1)};
descents N (f^P_N) ⊆ dc `{1..(n − 1)}]]
==⇒ ∀ i ∈ {1..m}.∀ j ∈ {1..n}.
T_to_Perm_block i j T r c P"

これらの形式化された命題を証明し、厳密証明を行った。直観的証明への形式化手法の適用は、形式化手法に関する研究成果として稀有であり、この証明過程の分析を進めることにより、自動証明等の計算機証明問題の新しい議論がなされることと思われる。

(3)2 次元分割表における形式的手法の拡張検討

人間の直観を利用した証明手法と計算機上での証明手法の差異を比較した。この比較結果から、直観の導入と鍵となる証明過程の挿入が証明の自動化へと繋がることが明らかになった。しかし、現状では一般的な他次

元分割表の数えあげに拡張することはできず、特殊な形の分割表にのみ留まっている。

その原因として、証明過程における descent の概念が拡張困難であったことが挙げられる。そのため、多次元分割表の数えあげに関する指標作成には至らなかった。しかし、Isabelle の持つ特性を利用した隙間のない証明過程を半自動で作成するシステムの実現により、証明において、群論によるアプローチが不足している箇所や人間の直観による曖昧さを含む箇所を明確にすることが可能となるものと推測される。

(4)今後の課題

多次元分割表の数えあげ問題の証明に形式手法を用いることは、複雑化された命題を隙間なく証明していく上で重要であると推察される。本研究により、現在の形式手法の問題点が明らかになった。それは次の 4 つである。これらの問題解決がなされることにより、統計学における計算機証明の適用や、計算機自動証明などの広い意味での人口知能への応用が認められるものと思われる。

①知識データベースの構築

現時点では、形式証明のための基本概念のデータ蓄積が少ないため、個別の問題に対して一から全ての数学的な道具を作り上げなければならない。これから、形式的な方法の利用が進むにつれて、道具をライブラリとして蓄えることにより個別の問題に対応しやすくなるであろう。

②形式表現アプリケーションの互換性

現在、さまざまな形式表現アプリケーションが存在し、複数のアプリケーションで表現されたライブラリの互換性が得られない。この問題を克服しなければ、人類の知恵の大きな無駄遣いとなる。互換性を持つことで、計算機証明の知識は飛躍的に増大するものと思われる。

③形式表現の系統化

与えられた問題を、形式証明でチェックするためには形式証明が行えるように、論理的に表現する必要がある。種々の問題の形式表現を行うことで、形式表現の手法を系統化する必要がある。系統化することで複数アプリケーションでの利用が簡単になり、準備した知識ベースを共通に利用することが可能となるであろう。

④自動証明への移行

実際に形式証明を書き下すためには、現在は人力に頼っている。問題の形式化、必要な定理の探索と作成といった作業をできるだけ機械化して、労力を省く必要があると思われる。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

Yoko Ono and Hidetsune Kobayashi, Comparison a human proof in Isabelle, Proceedings of the 2nd Workshop on Programming Languages for Mechanized Mathematical Systems, 2008, 査読有.

〔学会発表〕（計 3 件）

①橋口 博樹,小野 陽子, 黒田 正博, 分割表上のメトロポリスウォークの数値精度改良,第 22 回日本計算機統計学会,2008 年 11 月 6 日,(財)先端医療振興財団・臨床研究情報センター（兵庫県神戸市).

②Hidetsune Kobayashi, Zhenbing Zeng

and Yoko Ono,
A Trial for Automated Proof,
Computers in Scientific Discovery 4,2008年3月28日,East China Normal University, Shanghai, China, 査読有.

③Yoko Ono and Hiroki Hashiguchi,
On Comparison of p-values for Contingency Tables between MCMC and Direct Sampling
Bulletin of the 56rd Session of the International Statistical Institute,2007年8月28日,Lisbon, Portugal, 査読有.

６．研究組織
(1)研究代表者

小野　陽子(ONO YOKO)
新潟国際情報大学・情報文化学部・講師
研究者番号：60339140

(2)研究分担者

該当者なし

(3)連携研究者

該当者なし