

平成22年 3月 2日現在

研究種目：基盤研究(C)  
 研究期間：2006～2009  
 課題番号：18540148  
 研究課題名(和文) 不規則系列上の離散分布モデルに関する統計的推測の研究  
 研究課題名(英文) Statistical inference for discrete distributions on random sequences

研究代表者  
 安芸 重雄 (AKI SIGEO)  
 関西大学・システム理工学部・教授  
 研究者番号：90132696

研究成果の概要(和文)：確率変数列などの不規則な構造上に現われる連などの離散パターンやスキャンと呼ばれる複数の離散パターンの起こる回数やそれらが観測されるまでの待ち時間などの厳密な確率分布の研究とそれに基づいた統計的推測の研究を行い、さまざまな新しい成果を得た。この研究では、とくに離散パターンの待ち時間問題や離散パターンの生起数などの問題を、いくつかの停止時間を用いて統一かつ有効に扱い、解析的な方法を用いて理論的な成果を得た。また、そのうちのいくつかの結果は、数値計算にも有効であることを示した。

研究成果の概要(英文)：Exact discrete distributions of numbers of occurrences and waiting times of some patterns and scans on random sequences are investigated. Based on our theoretical new results we can make statistical inference of some parameters such as success probability. By extending the method of conditional probability generating functions we have derived new results analytically, some of which are also suitable for numerical calculations.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	500,000	0	500,000
2007年度	600,000	180,000	780,000
2008年度	500,000	150,000	650,000
2009年度	600,000	180,000	780,000
年度			
総計	2,200,000	510,000	2,710,000

研究分野：数物系科学

科研費の分科・細目：数学・数学一般(含確率論・統計数学)

キーワード：離散分布論、確率母関数、離散パターン、条件付き期待値、最尤推定量

## 1. 研究開始当初の背景

1980年代に、start-up demonstration test、連続システム(consecutive systems)の信頼性問題、品質管理、DNA系列の解析など、さまざまな応用分野から刺激を受けて行われていた離散パターンに関する離散確率分布

論の研究が、1990年頃に数え上げ組合せ論的な手法の限界とともに停滞を始めた。1990年代に入り、「有限マルコフ連鎖埋め込み法」や「条件付き確率母関数法」などの新しい手法の開発とともに再び急速に発展してきた。ただ、2000年ころからは、新手法の直接の

応用だけではうまく行かず、重要ではあるが、相当な工夫をしないと解けない問題が目立ち始めた。

## 2. 研究の目的

上で述べたような背景の下で、実際に計算が可能で、統計的推測に応用できるという視点から問題を整理して、次のような具体的な研究目的を立てた。

- (1) **Start-up demonstration test** について、従来、理論的な困難さから見送られてきたような制約、たとえば一定の長さの過去の履歴では捉えることができないような制約の下で研究を行う。
- (2) 2次元の格子点上に配置された離散確率変数族の中にある2次元の離散パターンが出現する確率の計算を行う。
- (3) いくつかの停止時間の関数として記述できるような離散確率変数の確率分布を統一的に計算することのできるアルゴリズムを開発する。
- (4) いくつかの離散パターンの待ち時間の同時分布の研究を行う。その際、「各離散パターンは他のパターンの部分パターンではない」という制約を外して議論する。
- (5) その他、不規則系列上の離散分布論の基礎理論へ貢献する。

## 3. 研究の方法

- (1) 複雑な離散分布の計算を行うためには、計算機のハード面だけでなく、数式処理やグラフィックスなども使用する必要があるため、数値計算および数式処理とグラフィックスを統合的に利用できるソフトウェアを入手して、現在の計算機環境で可能な離散分布研究に関する手法を開発してきた。
- (2) 確率論、数理統計学、グラフ理論など、この研究に関連する書籍を入手し、必要な理論を整理した。
- (3) この研究に興味をもつ国内の研究者と議論を積極的に行った。

## 4. 研究成果

- (1) **Start-up demonstration test** について、一定の長さの過去の履歴では捕らえることができない制約の下で、厳密分布の導出およびアルゴリズムの開発を行った。これは、研究の目的の(1)に当たる成果であり、この結果は学術論文5の③に発表した。
- (2) 2つの離散パターンが現われるまでの待ち時間の厳密分布を、「1つのパターンが他のパターンを含まない」という制約を置かずに導出することができた。これは、研究の目的の(4)に当たる成果で

あり、この結果は学術論文5の④に発表した。

- (3) 高次マルコフ連鎖における長さ  $k$  の連の待ち時間分布において、従来計算が困難であった、 $k$  の値がマルコフ性の次数  $m$  に比べて大きい場合に、条件付き確率母関数の連立方程式のサイズを  $m$  まで下げることができることを理論的に示した。これは研究の目的の(5)に当たる成果である。この結果は、学術論文5の⑨に発表した。
- (4) 連続変量をあるレベル  $t$  で切って離散化する場合に、離散化されてきた系列上の離散パターンの数は、レベル  $t$  を時間とする確率過程と捉えることができる。この確率過程の漸近的性質を調べ、それが、離散パターンに依存するガウス過程に弱収束することを示した。これは、それまでに知られていた **empirical process** の **Brownian bridge** への弱収束を真に含む結果であり、それ自身興味あるものでもある。その結果、複雑な離散分布の新しい近似法が得られた。これも研究の目的の(5)に当たる成果である。この結果は学術論文5の⑥に発表した。
- (5) マルコフ依存性をもった多値試行系列において、連の数の同時分布や待ち時間の同時分布などについて統一的な手法により厳密分布を導いた。これは研究の目的の(3)と(5)に当たる成果である。この結果は、学術論文の5の⑦、⑧、⑩において発表した。
- (6) 高次マルコフ木 (higher order Markov tree) によって与えられるグラフィカルモデル上で、連やスキャン統計量の厳密分布を導出するための理論的な結果を与えた。これは、不規則系列上で研究されている条件付き確率母関数の方法を有向木のグラフィカルモデルへ拡張したものであるが、所謂線形から非線形への拡張であるので、易しい拡張ではない。この結果の1つの応用として、有向木の形をしたシステムの信頼度の計算があるが、そのためのアルゴリズムおよび具体的な計算例も提供することができた。これは研究の目的の(2)と(5)に関連しており、学術論文5の①に発表した。
- (7) 独立な多値試行列上で複数の離散パターンを同時に考え、与えられた試行数までのパターンの観測数の確率分布と、複数のパターンに対してそれぞれが与えられた回数だけ起こるまでの待ち時間の確率分布の間の関係を、それぞれの2重確率母関数によって明らかにした。この結果は非常に一般的な結果であり、一般化された誕生日問題、クーポン収集問題などの複雑な問題に理論的な見通し

を与えるだけでなく、計算にも有効である。これは研究の計画の(5)に当たる成果である。この結果は、学術論文の5の②に発表した。

以上、述べた研究成果は離散分布論を中心とした数理統計学の基礎理論に貢献すると考えられる。また、これらはすべて国際的な査読付き学術雑誌に英文で公表することができたので、少なくとも国際的な基準で客観的な評価は得られていると期待できる。

#### 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 10 件)

- ① Inoue, K. and Aki, S. (2009). Distributions of runs and scans on higher-order Markov trees, Communications in Statistics-Theory and Methods, Vol. 38, 621-641 (査読有)
- ② Inoue, K. and Aki, S. (2009). On waiting time distributions associated with compound patterns in a sequence of multi-state trials, Annals of the Institute of Statistical Mathematics, Vol. 61, 499-516 (査読有)
- ③ Aki, S. and Hirano, K. (2008). Waiting time distributions for a run with additional constraints, Journal of Statistical Planning and Inference, 138, 3492-3501 (査読有)
- ④ Aki, S. and Hirano, K. (2008). Joint distributions of waiting time random variables for patterns, Journal of the Japan Statistical Society, Vol. 38, No. 1, 97-105 (査読有)
- ⑤ Inoue, K. and Aki, S. (2008). Methods for studying generalized birthday and coupon collection problems, Communications in Statistics-Simulation and Computation, 37:5, 844-862 (査読有)
- ⑥ Aki, S. (2008). Joint distributions of numbers of occurrences in a discrete pattern and weak convergence of an empirical process for the pattern, Journal of Multivariate Analysis, Vol. 99, 1460-1473 (査読有)
- ⑦ Inoue, K. and Aki, S. (2007). Joint distributions associated with compound patterns in a sequence of Markov dependent multistate trials and estimation problems, Journal of the Japan Statistical Society, Vol. 37, 285-298 (査読有)
- ⑧ Inoue, K. and Aki, S. (2007). Joint distributions of numbers of runs of specified lengths in a sequence of Markov dependent trials, Annals of the Institute of Statistical Mathematics, Vol. 59, 577-595 (査読有)
- ⑨ Aki, S. and Hirano, K. (2007). On the waiting time for the success run, Annals of the Institute of Statistical Mathematics, Vol. 59, 597-602 (査読有)
- ⑩ Inoue, K. and Aki, S. (2007). On

generating functions of waiting times and numbers of occurrences of compound patterns in a sequence of multistate trials, Journal of Applied Probability, Vol. 44, 71-81 (査読有)

〔学会発表〕(計 7 件)

- ① 井上潔司, 安芸重雄、「有向木上の連の分布」、2009年度統計関連学会連合大会、2009年9月9日、同志社大学(京田辺市)
- ② 安芸重雄、「離散パターンの数に関する empirical process(こついて)」、日本数学会 2009年度年会、2009年3月27日、東京大学(東京都)
- ③ 井上潔司, 安芸重雄、「一般化された誕生日問題」、日本数学会 2009年度年会、2009年3月27日、東京大学(東京都)

## 6. 研究組織

### (1) 研究代表者

安芸 重雄 (AKI SIGEO)

関西大学・システム理工学部・教授

研究者番号：90132696