

마코프모형의 계량정보학적 응용연구

A Study on Markov Chains Applied to Informetrics

문 경 화*
(Kyung-Hwa Moon)

초 록

계량정보학 연구영역의 하나인 운영연구(Operation Research ; OR)중 미래예측이라는 목적을 가지고 있는 마코프모형(Markov Chains)의 통계기법을 활용한 두가지 실험사례를 살펴보고, 최근의 연구경향을 분석함으로써 도서관 시스템 운영과 설계에 마코프모형을 응용할 수 있는 네가지 방안을 제시하였다. 계량정보학의 한 분야로 적용되고 있는 마코프모형에 관한 국내연구가 활발하지 못한 상태이므로, 국내 계량정보학에서의 마코프모형 연구의 필요성과 활성화를 제안하였다.

키 워 드

마코프모형, 마코프체인, 계량정보학, 운영연구, 확률과정

ABSTRACT

This paper is done by studying two experimental cases which utilize the stochastic theory of Markov Chains, which is used for forecasting the future and by analyzing recent trend of studies. Since the study of Markov Chains is not applied to the Informetrics to a high degree in Korea, It is also proposed that there is a necessity for further study on Markov Chains and its activation.

KEYWORDS

Markov Chains, Informetrics, Operation Research, Stochastic Process

* 하나은행 경영전략본부 전략기획팀
(Hana Bank Strategy Development Team)

1. 서 론

학문의 계량적 분석연구는 학문의 이론과 현실간의 괴리를 제거하기 위한 목적으로, 인문사회과학 여타 분야에서와 마찬가지로 계량정보학에서도 그 중요성이 강조되고 있다. 이러한 수량적 접근방법을 통한 확률이론인 마코프모형의 연구를 통해 국내 계량정보학 연구영역에 있어서 도서관 시스템의 여러부분에서도 효과적으로 응용할 수 있는 가능성을 제시하고자 하였다. 본 연구에서는 계량정보학의 연구영역중 시스템의 운영, 계획 및 설계에 관한 문제를 계량적, 체계적으로 분석, 해결해 보고자 하는 운영연구(OR)의 연구기법인 수리계획법, 네트워크모형, 대기행렬이론, 마코프모형, 모의실험, 결정이론 중 마코프모형(Markov Chains)을 중심으로 도서관 시스템의 운영, 설계에 적용하기 위한 방안을 제시하였다.

본 연구의 목적은 점차 계량적 연구의 중요성이 강조되고 있는 상황에서 경제학, 경영학, 통계학, 전산학 등에서 끊임없이 활발히 연구되고 있는 마코프모형의 최근 연구동향을 분석하여, 마코프모형의 계량정보학에서의 응용이 미비한 우리의 문헌정보학에 응용할 수 있는 방안을 모색하고 연구 활성화의 필요성을 제안하고자 함에 있다.

연구방법은 마코프모형에 대한 이론적 배경에 대한 기본적인 이해를 가지고, 최근 국내외에서 발표된 연구

논문을 학문분야별로 분석하고 두 가지 실험사례를 분석함으로써 도서관 시스템에의 적용가능성을 제안하였다. 마지막으로 계량정보학에서의 네 가지 응용방안은 도서관내에서 마코프모형을 응용할 수 있는 분야에 대해서 상태의 설정과 초기확률 및 추이확률을 구하는 선에서 연구하였으며, 이에 따른 실지 실험이나 검증은 향후 연구로 실시하고자 하였다.

2. 마코프모형의 연구동향

2.1 최근의 연구동향

마코프모형에 관한 최근의 연구동향을 보면 주로 수학, 통계학, 컴퓨터 관련 학문에서 많은 응용 논문들이 발표되고 있다. 특히 경영, 비즈니스, 마케팅 분야에서 활발한 응용이 이루어지고 있음이 조사결과 나타나고 있다. 이는 마코프모형의 응용 분야가 계량정보학뿐이 아니라 확률론의 한 이론으로서 여타학문의 발전에 기여되고 있으며 앞으로도 꾸준한 연구가 이루어질 것이라는 전망을 가능케 한다. 이것은 마코프모형 응용의 필요성과 그 연구범위가 광범위하다는데 의미가 있다. 지금까지 국내 문헌정보학의 계량정보학 영역에서 마코프모형의 응용연구는 활발하게 이루어지지 못하였다. 국내에서 마코프모형 연구는 한정된 전공자에 의한 연구가 수행되고 있을뿐 미비한 상태이다. 이에 연구 필요성을 가지고 본 연구에서는 최근에 발표된 국내외 마코프모형의

학문분야별 연구동향을 소개하고, 새로이 응용되고 있는 은둔 마코프모형(Hidden Markov Chains)에 대해서 살펴보았다. 먼저 이러한 연구동향 분석을 통해서 문헌정보학의 계량정보학 영역에서의 응용가능성을 탐진해 볼 수 있는 기회를 갖고자 하였다.

2.2 문헌정보학 분야에서의 연구동향

먼저 문헌정보학과 관련된 분야의 최근논문을 보면 국내에서는 김현희와 손승희의 논문이 있다. Chen은 공동 하이퍼텍스트의 저술에 대해서 컴퓨터 기반 공공저술을 분석하는 모델링을 구성하여 이용자의 행동구조를 기술하는데 있어 마코프모형을 응용하는 논문을 JASIS 1997년 12월에 발표하였다. 이 연구에서는 하이퍼미디어 데이터베이스를 이용한 이용자의 상호작용 행동모델을 정립하는데 마코프 이론을 응용하였다. 1994년에는 마코프모형을 이용한 하이퍼텍스트 정보검색시스템에 대한 탐색 상태 패턴이라는 논문을 Qiu이 JASIS에 발표하였다. 이 연구는 이용자 탐색행동을 연구하기 위하여 마코프모형을 하이퍼텍스트 정보검색시스템에 응용한 것이었다. 이외에도 외국에서 그동안 연구되어 왔던 연구들을 살펴보면 다음과 같다.

1984년 Reginald P. Coady는 “Testing for Markov-chain properties in the circulation of social-science monographs”라는 논문을 *Behavioral and*

*Social Sciences Librarian*에 발표하였는데, 사회과학 단행본의 대출업무에 있어서 마코프모형의 특성을 고찰한 연구논문이었다. 이 연구에서는 도서관 대출데이터에 마코프모형의 특성을 적용하는 가설을 세우고 실험하였다. 오하이오 주립대학도서관에서 1975년부터 1978년 동안에 얻어진 자료가 대출데이터로 수행되었다. 그 데이터의 시간 비의존성 실험은 데이터설정에서 시간 의존성으로 규명되었다. 대출과정에서의 마코프특성 실험은 마코프 특성에 부합되는 14개 데이터그룹 모두를 대상으로 규명, 실험되었다. 이 실험에서 도서관의 대출업무를 마코프모형으로 실험하는 것은 상태와 시간을 동일한 기반으로 모델이 구성되기 이전에 수행되어야 한다는 사실을 밝히고 있다. 그 이전 1982년에 Henryk Rybiniski와 Mieczyslaw Muraszkiewicz가 공동으로 연구하여 *Information Systems*에 발표한 “Reorganising the files in data base management systems”이라는 논제의 연구가 있었다. 이 논문에서는 파일의 동적 재인식 형태를 포함하고 있는 Data Base Management System(DBMS)모델을 마코프모델을 이용하여 기술하고 있다. 마코프모형은 DBMS의 환경변화를 기술하기 위해서 가정되어 실험되었으며 파일 재조직의 문제점은 동적 프로그래밍 개념에서 나타나고 있다고 밝히고 있다. 1978년에는 Nancy L. Geller가 *Information Processing & Management*에 “On the citation in-

fluence methodology of Pinski and Narin”라는 논문을 발표하였는데 이 연구는 마코프모형 기법을 이용하여 Pinski와 Narin에서의 방법론적 영향을 단위별 통령화률과 관련해서 규명하였다. 마코프모형 이론은 또한 단위별 영향범위에 존재하기 위해서 충분한 조건을 주고 있다고 밝히고 있다. 이렇듯 도서관 업무와 관련된 국내에서의 마코프모형 연구가 제한적으로 이루어져 온 것에 비해서, 국외에서의 연구는 오래전부터 활발히 진행되어 왔음을 알 수 있다.

2.3 다른분야에서의 연구동향

최근 통계학분야의 연구동향을 살펴보면 마코프모형이 기본적으로 확률과정을 다루는 것이기 때문에 이 분야의 연구가 활발히 이루어져 왔음을 알 수 있다. 1998년 Valen E. Johnson은 “A Coupling-Regeneration Scheme for Diagnosing Convergence in Markov Chain Monte Carlo Algorithms”라는 논문을 *Journal of the American Statistical Association*에 발표하였다. 1997년 Bernard Fingleton이 *Oxford Bulletin of Economics and Statistics*라는 저널에 “Specification and Testing of Markov Chain Models : An Application to Convergence in the European Union”이라는 연구를 발표하였다. 그리고 Lou W. U. Wendy가 유한 마코프모형에 대해서 연구한 “On Runs and Longest Run Tests : A Method

of Finite Markov Chain Imbedding”이라는 논문이 있는데, 이 논문에서는 마코프모형의 롱런이론(Long run theory)을 유한 마코프모형으로 방법론적으로 적용하고 있다.

최근들어 경영학 분야에서도 마코프모형을 응용한 많은 연구가 이루어지고 있다. 국내의 연구동향을 보면, 박정수가 1997년 한국경영과학학회지에 “An algorithm for computing the fundamental matrix of a markov chain”을 발표한 바 있다. 여기에서는 마코프모형을 기반으로 한 컴퓨터 알고리듬을 규명하는데 응용하였다. 1995년 김상욱은 산업과 경영에 “A Markovian goal programming approach to manpower supply planning for large-size organizations”을 발표하여 대단위 조직에서의 적정인력 수급계획수립을 위해서 마코프모형과 다목표계획법(GP)을 혼용 적용하여 접근한 연구를 발표하였다. 이렇듯 경영학 분야의 인사정책에서도 마코프모형이 효과적으로 응용될 수 있음을 알 수가 있다.

경제학분야의 연구동향을 보면 1996년 Chib Siddhartha와 Green-berg Edward가 경제학에 있어서 마코프모형 몬테 카를로 시뮬레이션에 대해서 연구한 “Markov Chain Monte Carlo Simulation Methods in Econometrics”를 *Econometric Theory*에 발표하였다.

금융분야에 있어서 마코프모형이 응용된 연구를 역시 상당수 찾아 볼

수 있다. 먼저 국내에서 1995년 조성표가 경영학연구에 발표한 내용은 “Markov chain model for evaluating present realizable value of credit accounts”라는 제목으로 신용채권의 현재 실현가치평가를 위한 마코프모형을 연구하였다. 1993년 정종락은 마코프모형과 주식가격의 확률보행성이라는 연구논문을 발표하였고, 1994년 채종홍은 마코프 전환체제 모형의 환율 변동성 예측이라는 논문에서 금융분야의 민감한 사안인 환율의 변동추이와 환율예측을 마코프모형을 응용하여 기술하고 있다. 한완선은 1993년 증권학회지에 다국면마코프 전환모형을 이용한 주가의 동태적 분석 및 예측을 연구하여 주식가격의 변동사항을 마코프모형을 응용하여 예측할 수 있음을 밝히고 있다. 1997년 이진의 석사 논문에서는 Markov switching ARCH 모형에 의한 단기 이자율과 주가지수 수익률을 연구한 바 있으며, 1997년 최희식은 마코프모형을 응용하여 실질 이자율의 기간구조 변동과 그 원인분석을 연구하였다.

국외에서 금융분야에 마코프모형이 응용된 연구동향을 살펴보면 다음과 같다. 1997년 Raymond와 Rich Robert는 *Journal of Money, Credit, and Banking*이라는 저널에서 “Oil and Macroeconomy : A markov state-switching approach”라는 연구를 발표하여 원유가와 거시경제학을 마코프 모형을 통한 접근방법으로 연구하였다. 동년 Jarrow Robert 등이 신용

리스크 관리를 위한 마코프 모형 연구를 검토하여 *The Review of Financial Studies*에 “A markov model for the term structure of credit risk pareads”라는 논제로 발표하였다.

전산정보과학 분야에서 발표된 논문으로는 1995년 최영일과 박광채가 정보과학논집에 다중 우선 순위를 갖는 부분버퍼공유방식의 성능분석이라는 논문을 발표하였다. 이 연구논문은 다중의 우선순위를 지원하는 통신망에 확장시켜 적용할 수 있는 부분 버퍼공유 우선순위 트래픽 제어 매커니즘의 해석방법을 마코프모형을 통해서 제안하였다. 여기서 마코프모형을 이용하여 제안한 해석방법은 전체버퍼의 크기, 임계치의 크기, 각 우선순위 트래픽의 세기 및 우선순위 트래픽의 혼합비, 그리고 각 우선순위 패킷이 점유할 수 있는 버퍼 영역의 임계치에 따라서 각 우선 순위 패킷의 손실 확률을 해석할 수 있도록 일반화하였다.

컴퓨터공학에서 마코프모형의 연구동향을 보면 1997년 F. Ayadi와 J.F. Hayes가 “Performance analysis of the bilayered ShuffleNet and thd SR-Net under deflection routing”이라는 논문을 *Computer Communications*지에 발표하였다. 이 논문은 멀티홉 네트워크를 위한 새로운 논리형상인 bilayered ShuffleNet에 마코프모형을 응용하고 있다. 이 네트워크는 회로수리화에 기호흐름 방법을 적용하고 그 시스템에 마코프모형을 응용함으로서 연구되었다. 1997년 스콧트

랜드의 F. C Berger와 Weide, 그리고 Hofstede는 “Supporting query by navigation”이라는 연구를 *Proceedings of the 16th Research Colloquium of the British Computer Society Information Retrieval Specialist Group*지에 발표하였는데, 마코프모형 접근방법을 통해서 일반적인 정보검색과 하이퍼텍스트 혹은 하이퍼미디어에서의 정보검색을 연구한 것이었다.

사회학분야의 연구로 *Sociological Methodology*에 실린 논문 두편이 1998년에 발표되었는데, 하나는 Fahrmeir Ludwig와 Knorr Held 그리고 Leonhard가 함께 연구한 “Dynamic Discrete-time Duration Models : Estimation via Markov Chain Monte Carlo”라는 연구발표가 있으며, 또 하나는 Humphreys Keith가 연구한 “The latent Markov chain with multivariate random effects : an evaluation of instruments measuring labor market status in the British Household Panel study”가 있다. 이 논문은 사회과학분야에서 노동인력시장 평가와 측정을 위한 수단으로 마코프모형이 응용된 예로 들 수 있다. 동저널에 1997년 일본인 야마구찌 카주오가 “Some log-linear fixed -effect latent-trait Markov-chain models : a dynamic analysis of personal efficacy under the influence of divorce/widowhood”를 발표했다. 이 연구에서는 이혼자와 미망인의 문제를 마코프모형을 통해서 분석하고

있다. 응용사회심리학 분야에서 Moodie가 “The meanings of the community and of the individual”이라는 논문을 1997년에 발표하였다. 이렇게 노동문제나 가정문제등 사회 각분야의 현상에도 마코프모형이 응용될 수 있음을 알 수가 있다.

이외에도 여러분야에서 연구가 이루어졌는데 스포츠 분야의 다양한 운동경기에 마코프모형을 응용한 경우가 있고, 언어학 연구에 도입된 연구논문도 있다. 1997년 발표된 “A Markov Chain Approach to Baseball”은 Bukiet Bruce와 Harold Elliotte Rusty 그리고 Palacios Jos가 야구에 마코프모형을 응용하여 연구한 내용으로 *Operations Research*에 발표된 바도 있다. 그리고 1996년 Pai와 Ravishanker가 공동연구한 “Bayesian Modelling of ARFIMA Processes by Markov Chain Monte Carlo Methods”가 *Journal of Forecasting*에 발표되었다. 농업분야에서 연구된 것으로 1998년 이준배가 준마코프 의사결정 모형에 의한 농기계의 최적화 정비정책이라는 논문을 발표하여 농업분야에서도 마코프모형의 유용성과 가능성을 보여주고 있다. 정치분야에서는 배일섭이 1994년 Markov 기법 활용방안 모색을 선거전략을 중심으로 하여 연구하였고, 심리학에서는 Baum이 “A markov model for measuring storage loss and retrieval”을 발표하였고, 남봉현의 석사논문에서는 Markov chain을 이용하여 인구이동 경향 분석을 하기도 하

였다. 수산물에 대한 응용연구도 있었는데 김진석은 1992년 수산물의 국내 균형수급량 및 지역별 수급량 예측적 위방정식 및 Markov chain 확률모형의 적용을 연구하였다. 음향학 분야에서도 연구가 이루어졌는데 이강성과 김순협은 한국음향학회지에 변이음 HMM을 이용한 고립단어 인식을 연구하는데 마코프모형을 활용하였다. 건설분야에서는 1996년 안일상의 Markov chain을 이용한 교량손상모델을 연구하였고 수자원분야에서 이재준은 조건부 종속성에 대한 Markov 연쇄의 최적차수 추정을 연구 발표하였다. 의학분야에서도 치료를 위한 응용연구가 이루어져서 조일구는 Markov 모형에 의한 암환자의 의료추구 경로 분석방법이라는 논문을 발표한 바 있었다. 기상학에서도 마코프모형은 응용되어 기상예보에 실질적인 영향을 미칠수도 있다는 논문이 강용균의 “Markov chain model for daily amount of cloud in Korea”에서 밝히고 있다.

각 분야에서의 마코프모형이 응용 연구된 사례를 분석해 보면, 그 응용 범위의 광범함에 우선 의미를 둘 수 있겠다. 통계, 경영, 경제, 정치, 노동, 금융, 농업, 언어학, 의학등 수많은 분야에서 활발히 연구되고 있는 마코프모형이 국내 도서관 업무에 있어서 응용되고 있는 비중은 극히 미비한 실정이다. 이러한 마코프모형에 대한 폭넓은 연구 범위를 인식하고, 계량정보학적 측면에서 응용해 볼 수 있는 가능성을 검토하였다.

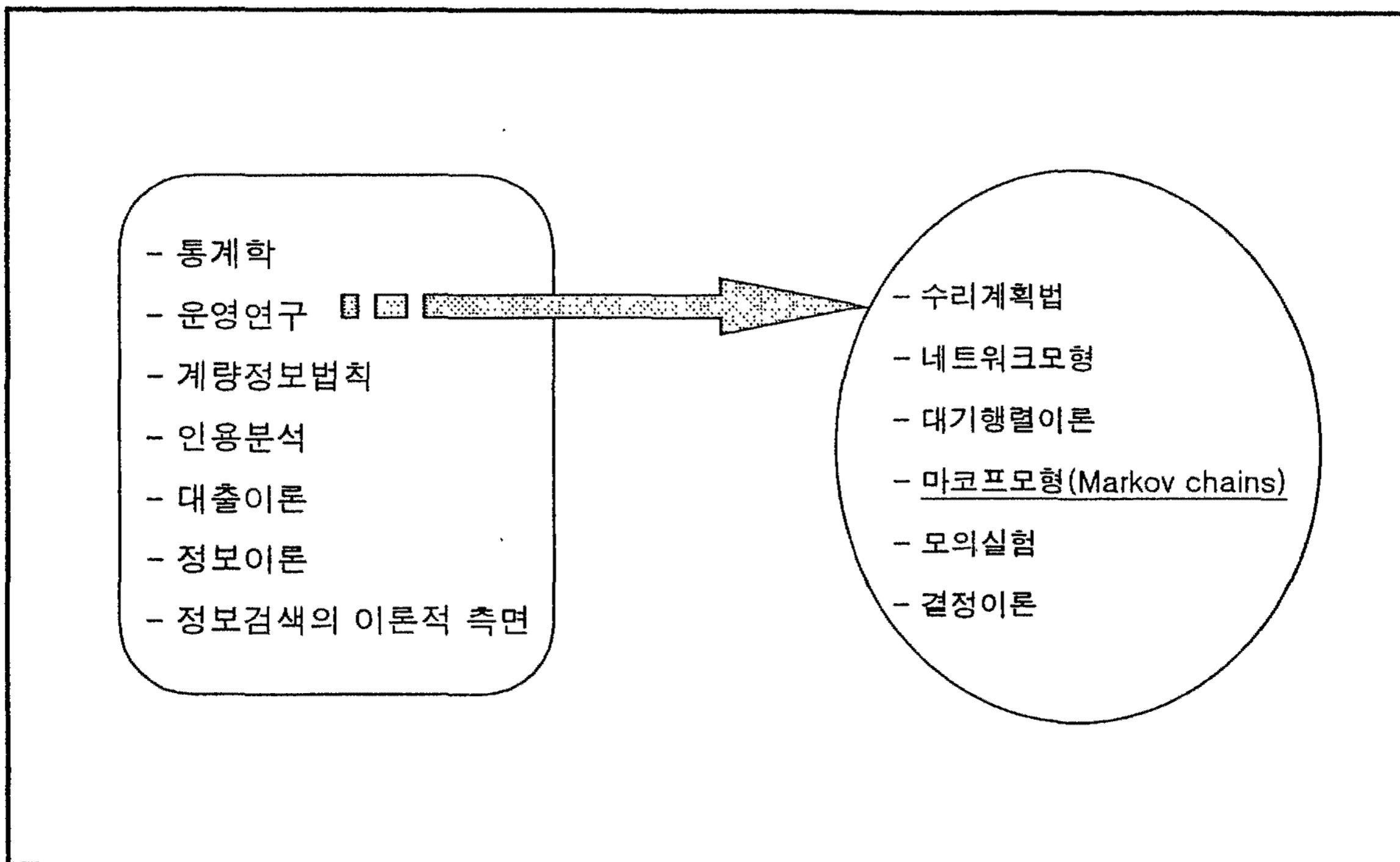
3. 이론적 배경

3.1 계량정보학의 연구영역과 확률과정

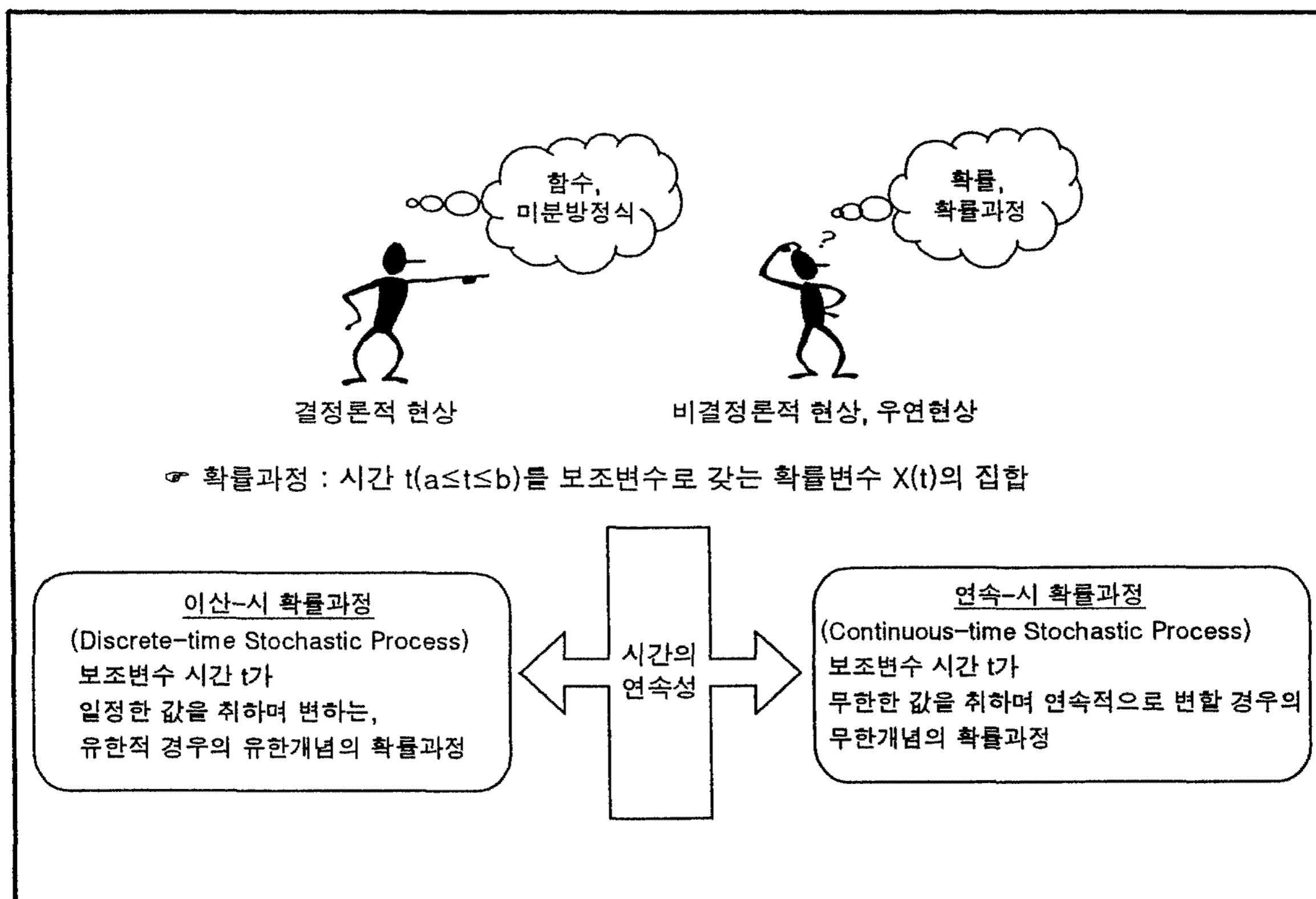
계량정보학(Informetrics)라는 용어는 계량정보학 이전의 “통계서지”라는 용어와 “계량서지학” 그리고 “계량정보화” “계량과학화”이라는 발전 과정을 역사적으로 거쳐오면서 현재 사용되고 있다. 통계서지란 “문현의 계수화방법으로 과학기술의 발전과정을 조명하는 것”이라는 정의로 사용되어 오다가, 주로 서지의 통계연구로부터 생겨나 주로 서양에서 발전하게 된 계량서지학(Bibliometrics)이라는 용어가 “도서와 다른 커뮤니케이션 미디어에 수학 및 통계방법을 적용하는 학문”으로 정의되어 사용되었다. 그후 “측정, 정보현상에 관한 제변수 간의 수리모형화, 그리고 정보의 축적과 검색을 다루는 학문”이라는 정의로써 계량정보학(Informetrics)이 사용되었다. 한편 계량과학학(Scientometrics)라는 용어는 주로 동구에서 이용되었는데 “과학과 기술발달의 측정을 연구하는 학문”으로 정의되었다. 계량정보학(Informetrics)라는 용어의 metric이란 측정(measurement)을 의미하는 것으로 이러한 수량적 접근방법을 통하여 학문연구에 있어서 이론과 현실간의 차이를 제거하고자 하는 것이다.

계량정보학의 연구영역으로는 통계학, 운영연구(OR), 계량정보법칙, 인용분석, 대출이론, 정보이론, 정보검

[그림 1] 계량정보학 연구영역



[그림 2] 확률과정



색의 이론적 측면 등이 있으며, 여기서는 마코프모형을 운영연구의 하나로서 선정하여 연구하였다.

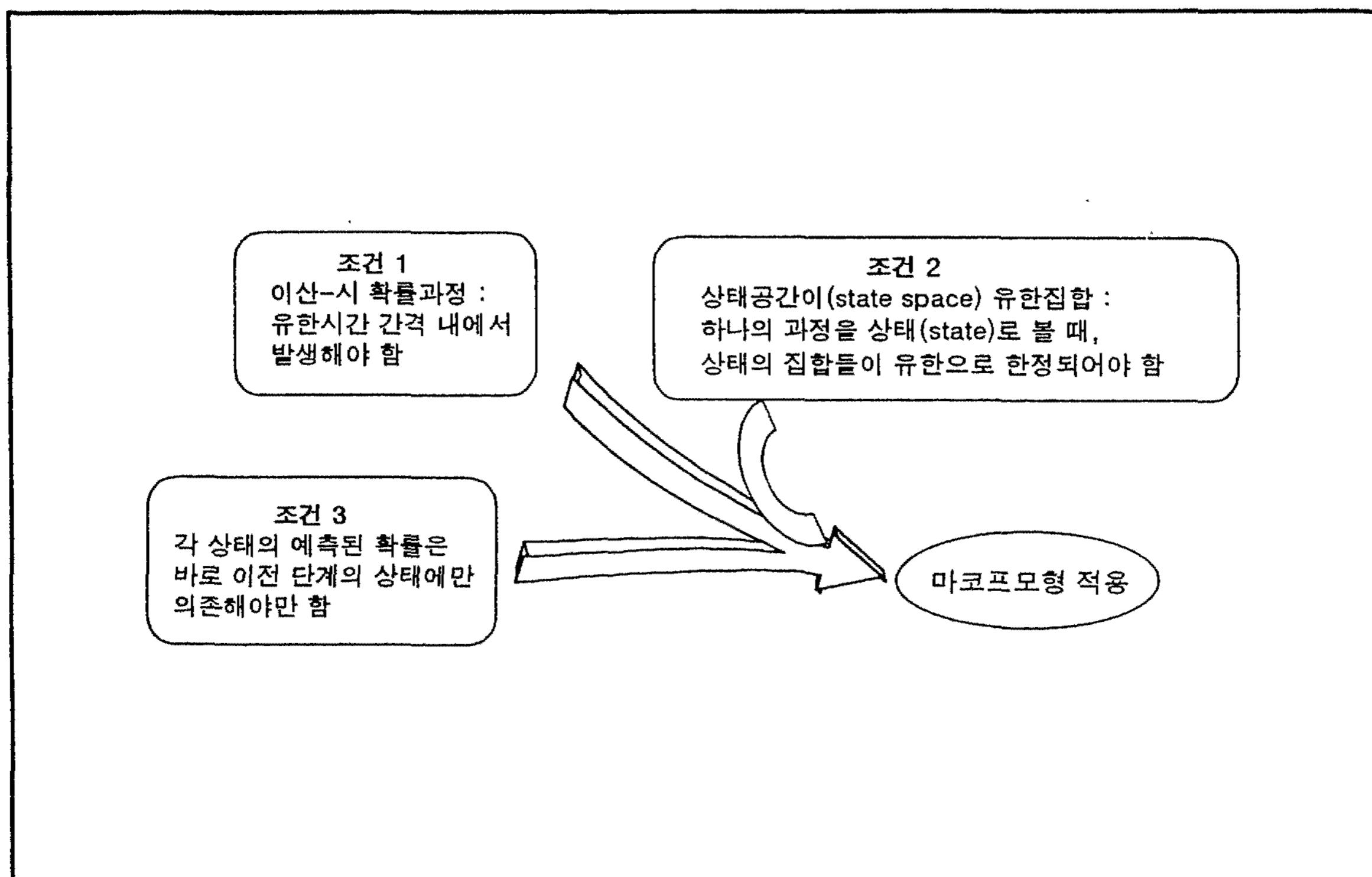
이러한 계량정보학에서 마코프체인을 설명하기 위해서는 먼저 확률과정을 이해해야 하는데 다음과 같이 살펴 볼 수 있다. 자연, 현상, 사회현상이나 인위적 실험을 통하여 결과가 얻어지게 되는 과정은 두 가지 경우로 귀착된다. 첫째는 매회 관찰 또는 측정에서 일정한 결과만이 얻어지게 되는 결정론적 현상이고, 둘째는 주사위를 던지는 게임에 있어서와 같이 매회 얻어지게 되는 결과들이 일정치 않을 경우이다. 이런 현상을 비결정론적 현상 또는 우연 현상이라고 부른다. 전자의 경우를 해석하기 위한 이론모형으로서는 함수, 미분방적식 등이 있으며, 후자의 경우에는 결과들이 얻어지게 될 경향성 내지는 법칙성을 주는 이론모형인 확률, 확률과정 등이 있다. 여기서 확률과정이란 일반적으로 시간 $t(a \leq t \leq b)$ 를 보조변수로 갖는 확률변수 $X(t)$ 의 집합 $\{x(t), a \leq t \leq b\}$ 를 말하며 시간의 연속성 여부에 따라 이산, 시확률과정과 연속, 시확률과정으로 구분된다. 이산, 시확률(discrete-time stochastic process)은 확률과정에서 보조변수 t 가 정수값 $t=0, \pm 1, \pm 2, \dots$ 만을 취하며 변화할 때 이것을 이산, 시확률과정이라 하며, 연속, 시확률과정(continuous-time stochasticprocess)은 t 가 연속적으로 변할 때를 의미한다.

3.2 마코프모형의 특성

마코프모형은 러시아 수학자 안드레이 안드레비치 마코프(Andrei Andreyevich Markov)의 이름을 따서 명명된 확률이론이다. 마코프는 마코프모형이라는 확률과정으로 널리 알려지게 되었으며, 마코프는 초창기에 주로 수이론과 분석에 관한 연구활동에 몰두하였다. 1923년 Wiener는 연속마코프모형을 다루는 첫 후계자가 되어 열성적으로 연구를 이어갔다. 그리고 마코프의 일반이론적 기초가 1930년대 Andrei Kolmogorov에 의해 연구가 진행되어왔으며, 1903년 태어난 마코프의 아들은 마코프라는 동일한 이름을 가지고서 그 아버지의 뒤를 이어서 널리 알려진 수학자가 되어 그 뒤를 계승하였다. 1990년 이후에는 연속개념의 은닉마코프모형이 응용되었다. 이러한 마코프모형의 중요성은 과학, 기술, 경영분야에서 최근 수년간 실제적인 응용에 의해서 발전되어왔다. 마코프과정(Markov process)은 $n+1$ 번째 단계상태확률(state probability)이 n 번째 단계상태확률의 영향만 받고 그 이전단계의 상태확률의 영향을 받지 않는다는 특성을 가지는 이산, 시확률과정이다.

마코프모형의 대표적인 특성은 첫째, 그 과정이 이산, 시확률과정(discrete-time process)이어야 한다. 즉 유한시간 간격내에서 발생해야 한다. 둘째, 상태공간(state space)이 유한집합이어야 한다. 이것은 하나의 과정을 상태(state)로 볼 때 그 상태의 총합체를

[그림 3] 마코프모형의 특성



상태공간(state space)이라 하는데 일련의 상태집합들이 무한상태가 아니라 유한으로 한정되어야 한다는 것이다. 셋째, 그 과정이 각 상태의 예측된 확률은 바로 이전단계의 상태에 의존해야만 한다는 마코프성질 (Markov property)을 가져야만 하는 것이다. 이러한 조건들이 충족된다면, 마코프모형 이론은 초기확률(initial probabilities)이 주어지고 a상태로부터 b상태로 이동하는 추이확률(conditional probabilities of transition)이 주어졌을 때 비로서 마코프과정이 완전히 설명될 수 있다.

이러한 마코프모형의 특성을 설명하기 위해서 연구기관 내에서 연구자들의 이직과정을 예로 들어보면, 세 개의 연구기관 a, b, c가 있고 여기서

근무하는 네명의 연구자 A, B, C, D 가 있다고 가정한다. 각 연구자는 어느 한 연구기관에 고용되며, 년초마다 다른 기관으로 옮기거나 그대로 머물 수 있다고 할 때, 연구자들이 선택할 수 있는 연구기관을 “상태”, 상태의 집합을 “상태공간”이라고 할 수 있다. 여기에서 1년 간격으로 연구기관을 옮길수 있고 그대로 머무를 수 있기 때문에 이산, 시확률과정이며, 두 번째로 상태공간이 a, b, c로 한정되어 있는 유한집합이며, 세 번째 현재의 연구성과에 의해서 연구기관을 옮기는데 영향을 받게 될 것이므로 각 연구자의 다음 상태는 현상태에 의해서 결정된다고 볼 수 있는 것이다.

연구자 C는 연구기관 b에서 근무하기 시작하여, 다음해에 연구기관 a로

〈표 1〉 최근 4년간의 연구자들 이동과정

상태	a	b	c
	A_1	C_1	B_1
	C_2	D_1	D_2
	A_2	B_2	D_3
	A_3	B_3	D_4
	C_3	B_4	A_4
	C_4		

〈표 2〉 연구자들의 연구기관 이동추이

	a	b	c	계
a	4	0	1	5
b	1	2	1	4
c	0	1	2	3
계	5	3	4	12

옮겼고, 2~3년 동안 남아있었다. 연구자 A는 연구기관 a에서 근무하기 시작하여, 다음해와 3년 동안 남아 있었고, 4년째에 연구기관 c로 옮겼다. A_1 은 연구자 A의 1차 근무년을 나타내고, A_2 는 연구자 A의 2년차, A_3 은 3년차 근무년수를 나타낸다.

3.3 초기상태확률과 추이확률

이상의 세가지 특성을 충족시키는 것 이외에 각 상태의 초기확률과 추이확률이 필요하다. 실험을 시작할 때, 각 연구기관에 대한 연구자들의 비율에 대한 합리적인 가정을 해야 하는데, 각 상태의 초기확률은 첫해에 각 연구기관에 고용된 연구자수를 총 연구자수로 나눈 값으로 얻어진다. 각 연구기관에 대한 연구자들의 비율에 대한 초기확률이 구해지고 나면, 한 연구자가 현 연구기관에서 세가지 연

구기관중 어느하나로 이동할 수 있는 가능성의 추이확률이 필요하다. 추이 확률은 현재상태 i에서 다음상태 j로 이동될 확률로 일반적으로 P_{ij} 로 나타낸다. P_{ab} 의 조건부 추이확률은 상태 a에서 상태 b로 이동할 확률이기 때문에, 연구기관 a에서 연구기관 b로 이동하는 경우를 계산함으로써 얻어질 수 있다.

즉 〈표 1〉에서 보면, 첫 해에는 연구기관 a에 연구자 A 한 명이, 연구기관 b에 연구자 C와 D 두 명이, 그리고 연구기관 c에 연구자 C 한명이 고용되어 있었다. 따라서 기관 a, b, c에 대한 초기확률은 아래의 〈표 3〉과 같게 된다.

마찬가지로 추이확률은 〈표 4〉와 같이 나타날 수 있는데 제1행을 보면 a라는 상태에서 각각 a, b, c라는 상태로 옮겨갈 확률을 보여주고 있다. 상태 a에서 한번 이동할 때 a, b, c중 어느 한 상태로 옮겨가게 되므로 반드시 그 행렬에서의 확률합은 1이 성립된다. 즉 $P_{aa}+P_{ab}+P_{ac}=1$ 이 성립된다.

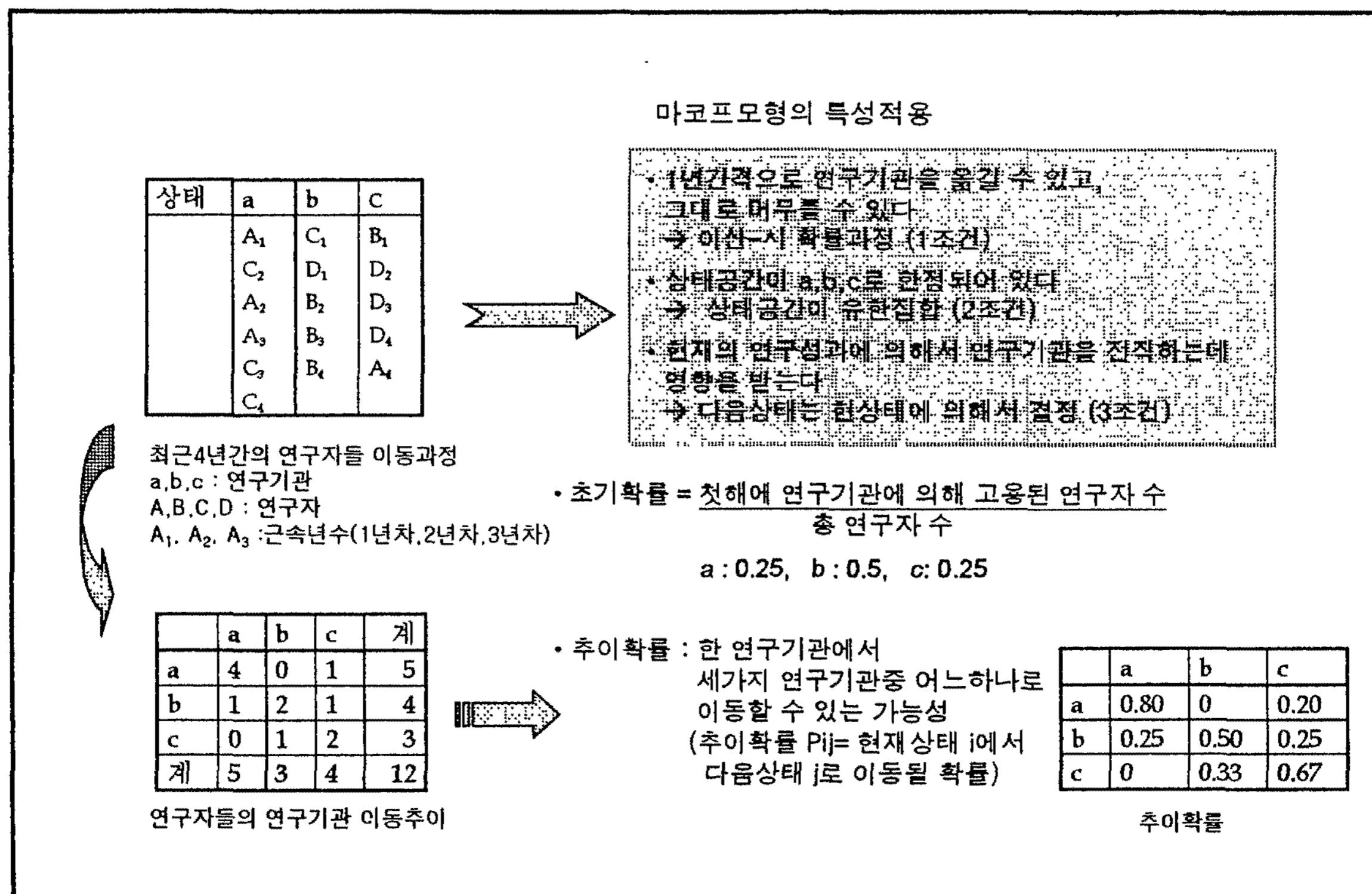
〈표 3〉 초기 확률

a	b	c
0.25 (1/4)	0.50 (2/4)	0.25 (1/4)

〈표 4〉 추이 확률

	a	b	c
a	0.80	0	0.20
b	0.25	0.50	0.25
c	0	0.33	0.67

(그림 4) 초기확률과 추이확률



다. 이 경우에 P_{aa} 의 조건확률은 상태 a에서 상태 a로 이동할 확률이기 때문에 이 경우는 $A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$, $C_2 \rightarrow C_3$, $C_3 \rightarrow C_4$ 로 총 4번이며 연구기관 a에서 연구기관 b로 옮겨간 사람은 아무도 없고 연구기관 a에서 연구기관 c로 옮긴 경우는 $A_3 \rightarrow A_4$ 로 단 한번 있다. 따라서 상태 a에서 a, b, c로의 추이확률은 각각 $4/5 (=0.80)$, $0, 1/5 (=0.20)$ 이 된다.

3.4 미래예측과 고정확률

마코프모형의 중요한 목적은 시스템의 미래 행위를 예측할 수 있다는 것이며, 그 결과는 추이확률과 현 단계의 상태확률이 결합되어 구해지는 다음 단계의 상태확률과 추이확률을 n제곱하여 구할 수 있는 시스템의

장기적인 안정상태의 확률을 나타내고 있다. 예에서 다음기간에 각 연구기관에 고용될 연구자들의 비율 즉 상태확률을 예측하고자 하는데, 이 예측은 초기확률과 추이확률을 곱하여 구할 수 있다. 다음기간의 상태확률이 계산된 후 그 다음기간의 상태확률의 계산도 가능한 것이다. 두 기간후의 상태확률은 다음 기간의 상태확률과 추이확률을 곱해서 구해지며 <표 5>와 같다.

마코프모형이론에 의하면 추이확률 P 가 정규 마코프모형이면, 여러 단계를 거친 후 각행에 있는 확률값들의 분포가 똑같은 고정확률행렬 L 을 얻게 되는데, 일단 고정확률행렬 L 에 이르면 행렬의 확률값들은 시간에 따라 더 이상 변화하지 않는다. 이 경우

〈표 5〉 미래예측의 추이확률(2승)

$$M^2 = b \begin{bmatrix} a & .640 & .067 & .293 \\ c & .325 & .333 & .342 \\ & .083 & .389 & .528 \end{bmatrix}$$

$$\begin{array}{ccc} a & b & c \\ a & b & c \\ [0.25 0.5 0.25] \times \begin{bmatrix} 0.80 & 0.00 & 0.20 \\ 0.25 & 0.50 & 0.25 \\ 0.00 & 0.33 & 0.67 \end{bmatrix}^2 & = & [0.33 0.33 0.34] \end{array}$$

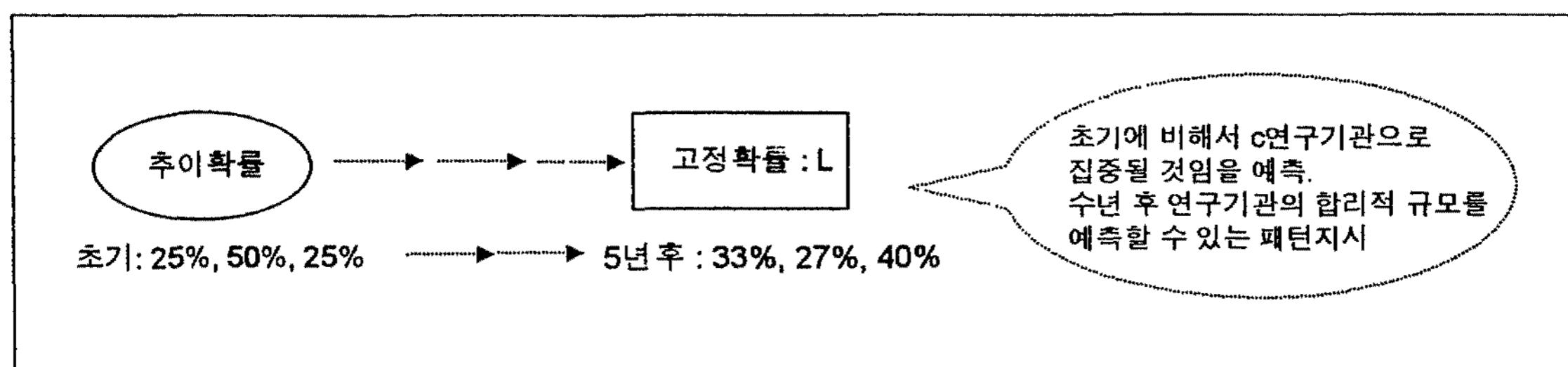
〈표 6〉 고정확률L(5승)

$$M^5 = b \begin{bmatrix} a & .3335 & .2662 & .3996 \\ c & .3332 & .2664 & .3997 \\ & .3328 & .2667 & .3997 \end{bmatrix}$$

$$0.33 \quad 0.27 \quad 0.40$$

$$\begin{array}{ccc} 0.80 & 0.00 & 0.20 \\ [0.25 0.5 0.25] \times \begin{bmatrix} 0.80 & 0.00 & 0.20 \\ 0.25 & 0.50 & 0.25 \\ 0.00 & 0.33 & 0.67 \end{bmatrix}^5 & = & [0.33 0.33 0.34] \end{array}$$

〔그림 5〕 마코프모형의 미래예측



는 2의 5승 단계를 거친후 고정확률 행렬에 이르게 된다. 고정확률행렬은 몇 년후에 연구자들의 33%, 27%, 40%가 연구기관 a, b, c에서 각각 균무하게 될 수 있음을 예측하고 있다. 즉 초기확률과 고정확률을 비교하면 연구기관 c로 연구자들이 몰릴 것을

예측할수 있는 것이다. 이는 연구기관 c의 고정확률이 40%로 초기확률인 25 %보다 훨씬 높기 때문이다(〈표 6〉). 결국 마코프모형은 가상의 연구기관 실험에서 수년후에 세 연구기관의 합리적 규모를 예측할 수 있는 패턴을 지시해 주는 것이다.

3.5 은닉마코프모형 (Hidden Markov Chains)

은닉 마코프모형(HMM : Hidden Markov Model)은 이중적으로 결합된 stochastic process로 구성된 확률적 함수이다. HMM의 내부에 존재하는 것으로 가정되는 마코프모형은 유한한 개수의 상태와 각 상태와 결부되어있는 난수 함수들의 집합을 가지고 있다. 각각의 이산시간에서 프로세스는 어떤 상태에 있고, 그 상태를 결정하는 난수 함수로부터 한 출력벡터가 관측된다고 가정한다. 그리고 나서 내부의 마코프모형은 전이확률행렬에 따라 다음 상태로 전이하게 된다. 따라서 관측자의 입장에서는 상태로부터 출력벡터만을 관측할 수 있을 뿐이며, 내부의 상태는 관측할 수 없는 상황이다. HMM은 한 상태에서 관측가능한 벡터들에 따라 이산분포, 준연속분포, 연속분포 등으로 분류된다. 즉 은닉마코프모형은 지금까지의 이산-시(discret-time)개념 이론과는 달리 연속적인(continuous)개념의 이론을 적용한 것이다. 이러한 은닉 마코프모형은 군집개념을 구성하는데 클러스터링으로서 많이 활용되고 있다. 예를 들어서 빈민촌과 부촌, 선거전에서는 여당역세지역과 야당역세지역에 관한 군집을 예측하는데 이용되어 1990년 초부터 활발한 연구가 이루어졌다. 1997년 7월 일본 정보처리학회논문지에 竹内孔一와 松本欲治는 은둔 마코프모형에 의한 일본어 형태소 분석의 패러미터 학습이라는 은닉 마코프모

형에 대한 연구를 발표한 바 있다. 이는 마코프모형을 활용하여 일본어의 형태소를 분석하고, 자연어처리를 하기 위한 패러미터추정을 연구하고 있다. 그리고 국내에서 이종진이 Hidden Markov Model을 이용한 연속숫자 인식이라는 연구를 발표하였으며, 박현상이 Diphone 단위의 Hidden Markov Model을 이용한 한국어 단어인식을 연구하였고, 김우성과 이기돈이 Hidden Markov Model을 이용한 다중 활자체 영, 숫자 인식을 1992년에 발표한 바 있다.

4. 실험 사례

4.1 민속음악 주제의 저자데이터 실험

마코프모형이 민속음악 주제에 관한 저자데이터 비교 실험에 응용된 것을 살펴보면 다음과 같다. 민속음악 주제에 대한 전반적인 저자데이터 비교는 10년동안 2,018명의 저자가 발표한 총 3,302건의 발행물을 대상으로 하였다. 마코프모형의 세가지 조건을 충족시키기 위해서 동일저자의 두 연속적인 발행물의 저작사이 시간을 각 단계에 대한 단위를 선택했다.

세가지 마코프모형의 조건을 충족시키기 위해 이산, 시확률과정이어야 한다는 첫 번째 조건에 대해서는 동시에 두 개 논문을 발행할 수 있고, 시간상 차이가 있다. 그 단계는 이산, 유한적인 것이다.

그리고 상태공간이 유한집합이어야 한다는 마코프모형의 두번째 조건을 충족시키기 위해서는 민속음악 주제 내에 원론, 일반, 아프리카 민속음악, 아시아 민속음악, 유럽 민속음악, 북미 민속음악, 남미 민속음악, 오스트레일리아 민속음악, 대중음악 등 총9개의 한정된 하부주제 분야가 있으며, 지리적 위치에 의해서 지역구분되어 있다. 각 논문이 초록저널 *Repertoire Internationale de Litteratur Musicale*의 분류에 있어서 중복되지 않는다. 그 조건은 포괄적이며, 상호배타적으로 적용된다.

또한 바로 이전단계의 상태에 의존한다는 세번째 조건충족을 위해서는, 한 하부주제에서 다른 것에 대한 저자의 추이는 그 저자의 현재의 연구관심사에 의존하게 되므로 모든 조건을 충족하게 된다. 재정, 저자의 교육배경, 자원의 활용, 자료와 기회가 유용할지라도, 그리고 저자의 결합은 요소들에 영향을 줄 수 있으며, 저자의 현재 연구관심사가 주요요소가 된다. 이상을 요약해 보면 다음과 같다.

- (1) 9개 하부주제분야사이의 저자 이동과정은 유한하다.
- (2) 9상태 혹은 하부주제의 모든 리스트는 유한하다.
- (3) 각 상태추이는 오직 현재상태에 의존한다.

따라서 이 실험은 마코프모형을 위한 조건을 충족시키고 있다고 볼 수 있다. 이것은 모든 다른 상태로부터 다다를 수 있는 상태의 각각을 허용하는 9개 하부주제에서 발생되는 것

으로부터 한 논문을 방해하는 외부가 공물은 존재하지 않기 때문에 정규마코프모형이라 할 수 있다. 9개 하부주제사이에서 각 저자들이 발표한 논문들의 이동경로를 추적하였고, 원론주제에 대해서는 a, 일반 주제에 대해서는 b, 아프리카 민속음악에 대해서는 c, 아시아 민속 음악에 대해서는 d, 유럽 민속음악에 대해서는 e, 북미 민속음악에 대해서는 f, 남미 민속음악에 대해서는 g, 오스트레일리아 민속음악에 대해서는 h, 대중 민속음악에 대해서는 i라고 명명하였다. 각 저자의 첫 논문은 초록서비스에 의해서 사용된 표목에 의해서 하부주제에 대한 관련성을 예측하여 조사하였다. 각 논문은 9개 하부주제에 각기 하나씩만 할당되었다. a에서 i 주제까지 52, 52, 148, 409, 614, 194, 175, 35 그리고 339건의 해당 주제의 발표건수들이 집계되었다. 여기서 총저자수인 2,018을 첫해의 발표건수로 나눈 값인 초기확률은 <표 7>과 같이 구할 수 있다.

<표 7> 민족음악학내 9개 하부주제의 이동에 대한 초기확률

a : Discipline	=0.0258 (52)
b : General	=0.0258 (52)
c : Africa	=0.0733(148)
d : Asia	=0.2027(409)
e : Europe	=0.3043(614)
f : North America	=0.0961(194)
g : South America	=0.0867(175)
h : Australia	=0.0173 (35)
i : Popular Music	=0.1680(339)

〈표 8〉 9개 하부주제의 이동추이

	a	b	c	d	e	f'	g	h	i	Number of Transitions
a	11	10	6	11	28	0	3	1	0	70
b	7	4	5	15	10	3	1	0	2	47
c	7	2	82	19	5	2	2	1	9	129
d	18	5	13	78	21	1	2	1	3	142
e	32	14	0	11	135	44	0	1	5	202
f	2	4	3	2	0	17	4	0	12	44
g	3	1	1	1	1	1	24	0	2	34
h	4	2	1	2	0	0	0	14	0	23
i	1	1	5	1	6	14	4	1	26	59
Number of Transitions	85	43	116	140	206	42	40	19	59	750

〈표 9〉 추이행렬

	a	b	c	d	e	f	g	h	i
a	.1571	.1429	.0857	.1571	.4000	.0000	.0429	.0143	.0000
b	.1489	.0851	.1064	.3191	.2128	.0638	.0213	.0000	.0426
c	.0543	.0155	.6357	.1473	.0388	.0155	.0155	.0078	.0698
d	.1268	.0352	.0915	.5493	.1479	.0070	.0141	.0070	.0211
e	.1584	.0693	.0000	.0545	.6683	.0198	.0000	.0050	.0248
f	.0456	.0909	.0682	.0455	.0000	.3864	.0909	.0000	.2727
g	.0882	.0294	.0294	.0294	.0294	.0294	.7059	.0000	.0588
h	.1739	.0870	.0435	.0870	.0000	.0000	.0000	.6087	.0000
i	.0169	.0169	.0847	.0169	.1017	.2373	.0678	.0169	.4407

다음은 한 주제에서 다른 주제로의 이동추이를 산출하였다. 즉 아프리카 민속음악을 연구하던 저자가 다른 주제의 민속음악을 발표한 건수를 집계한 것이다. 1,441명이 그 주제에 대해서 단 한건의 발표물을 기고했고, 따라서 그것들의 본래의 주제로부터 어떤 추이를 산출해낼 수 없으므로 제외시켰다. 아래의 〈표 8〉에서 보면 Pca라는 것은 하부주제 c(Africa)로부터 하부주제 a(Discipline)로의 추이를 보여준다. 즉 아프리카 민속음악

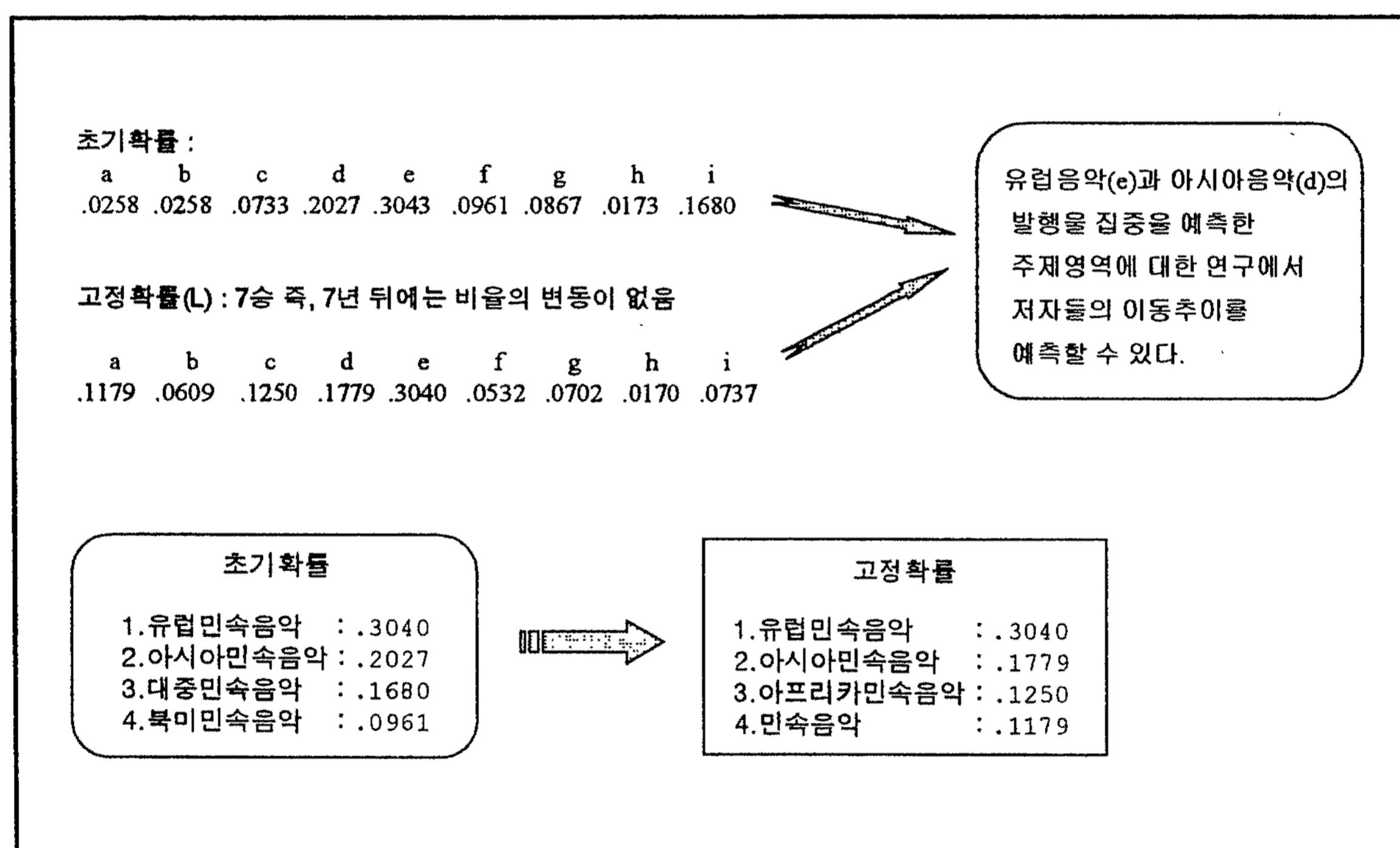
을 연구하던 저자가 원론에 해당하는 연구물을 발표하게 된 이동추이이다. 각각의 민속음악 주제에서 다른 주제로 이동추이를 살펴보면 아래 표와 같다.

다음의 〈표 9〉에서는 이 실험에 대한 추이행렬 E를 나타내며, 하부주제 i로부터 j로의 이동에 대한 조건부확률을 포함한다. 여기서 각행(row)의 합은 0이 된다. 하부주제 h를 보면 오스트레일리아 음악인데, 행(row) h와 열(column) h는 변동이 거의 없

〈표 10〉 고정확률

	a	b	c	d	e	f	g	h	i
a	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
b	.1179	.0609	.1250	.1779	.3041	.0532	.0702	.0170	.0737
c	.1179	.0609	.1250	.1779	.3041	.0532	.0702	.0170	.0737
d	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
e	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
f	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0737
g	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0738
h	.1179	.0609	.1250	.1779	.3041	.0532	.0701	.0170	.0737
i	.1179	.0609	.1250	.1779	.3040	.0532	.0702	.0170	.0737

[그림 6] 초기확률과 고정확률의 비교



게 나타나고 있다. 이는 하부주제로부
터 이 하부주제로의 이동추이는 거의
없다는 것을 의미한다. 다시 말해, 상
대적으로 하부주제 유럽음악 e에 대
해서는 보다 활발한 움직임이 있는
것이다. 〈표 9〉의 추이행렬은 〈표 8〉
에서 조사된 9개 하부주제의 이동추
이 건수를 행의 총계로로 나눈 값으

로 산출한 것이다.

다음은 위의 추이확률을 여러번 반
복 계산하는 과정에서, n의 검정(승
수")에 대한 행렬을 얻게 되는데 이
러한 고정된 수로 집결되는 상태가
고정확률이다. 이 실험에서는 일곱번
째 단계에서 즉 M의 7승 단계에서
고정된 수의 행렬을 발견하게 되는데,

그 행렬은 <표 10>에서와 같이 정상 확률분포로 집중하게 된다. 초기확률과 고정확률 비교를 통해, 대중 민속음악과 일반주제의 민속음악 영역이 덜 안정적인 반면에 유럽과 오스트레일리아 민속음악이 가장 안정적이라는 것을 알 수가 있다.

결국, 초기확률과 고정확률 수치를 비교해보면 다음과 같다. 초기확률에서 유럽 민속음악이 .3040, 아시아민속음악이 .2027, 대중민속음악이 .1680, 북미민속음악이 .0961이던 것이 몇 년이 경과한 후에는 고정확률에서 나타나는 바와 같이 유럽 민속음악 .3040, 아시아민속음악 .1779로 두 주제의 민속음악이 여전히 지배적인 우위를 점하게 될 것이라는 것을 예측가능하게 한다. 그다음 고정확률에서는 대중 민속음악과 북미 민속음악 대신에 아프리카 민속음악과 민속음악에 대한 연구가 증가될 것임을 예측가능케 하고 있음을 알 수가 있는 것이다.

몇년 후에 아시아 민속음악과 유럽 민속음악이 음악가들 사이에서 주요 관심사가 되더라도, 대중음악과 북미민속음악으로부터 아프리카 민속음악과 원론에 속하는 민속음악 쪽으로 관심이 이동될 것임을 예측할 수 있다.

북미 민속음악과 대중 민속음악 영역이 다소 감소하는 반면에, 아프리카음악으로의 저자비율은 8%~13%로 증가할 것이다. 정상분포 L의 역은 재귀성, 혹은 저자가 동일주제로 되돌아가기 전에 발표된 논문수를 알기 위해 사용되었다. 하부주제 a에서 i에 대한 가치는 8, 16, 8, 6, 3, 19, 14, 59 그리고

14로 나타난다.

한 저자가 북미민속음악을 벗어나 다른 분야의 음악을 하게 되다가 음악가가 본래의 주제음악으로 되돌아가기 전 까지는 최소한 3개의 발표작이 되는 셈이 된다. 저자당 평균발행 건수는 총 발표건수인 3,302을 총저자수인 2,018로 나눈 값, 혹은 1.6이 된다. 이런 것들이 모든 하부주제들에 대해서 사실이기 때문에, 그것은 한명의 민속음악가가 한때 그가 다른 영역을 추구하기 위해 떠났던 연구영역으로 되돌아가려하지 않는다고 결론지어진다.

4.2 한국화학자들의 이동과정 실험

두 번째 실험은 한국화학자들의 7개 하위주제간 이동실험인데 두가지 가설을 가지고 실험을 하였던 결과를 보면 다음과 같다. 시스템의 미래 행위를 예측하는 기법인 마코프모형을 이용하여 학문의 발전과정을 기술, 예측할 수 있다는 가설 1과 마코프모형의 추이확률행렬을 이용하여 학문분야의 지적구조를 다차원 공간상에 규명할 수 있다는 가설 2를 검증하기 위해서 화학분야 학술지들에 실린 3,815편의 논문을 이용한 세가지 실험을 하였다. 첫 실험은 처음의 가설 1을 검증하기 위해서 <학술총람 제33집>에서 1967년에서 1973년 동안 2편 이상의 논문을 쓴 313명의 화학자들이 저술한 총논문수 968편을 분석하여 각 하위주제의 초기 상태확률, 추이확률 및 고정확률을 구하였다. 고정확률에 의해서 예측된 각 하위주제에 대한 연구자들의 비율이 타

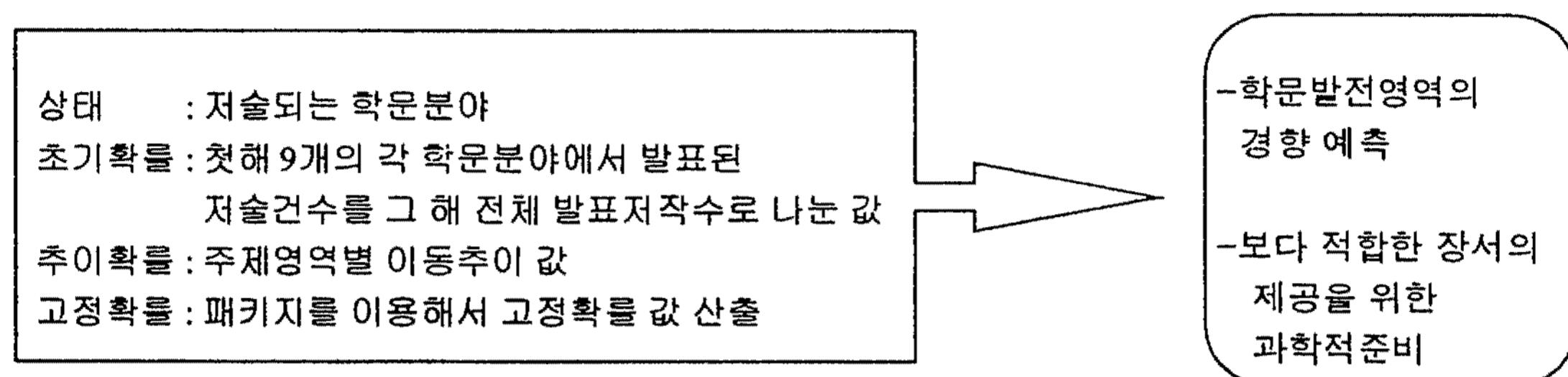
당성이 있는지를 검증하기 위해서 313명 중에서 1974년부터 1983년 동안 1편 이상의 논문을 쓴 290명을 선정하여 이들이 쓴 총논문수 1,773편을 7개의 하위주제별로 집계하여 분석한 결과 인 관측확률을 예측확률(고정확률)과 비교해 본 결과 예측대로 무기, 분석 유기화학에 관한 논문을 발표한 연구자들의 비율이 낮아졌으며, 물리화학, 생화학, 고분자화학의 비율이 높아지고 있었다. 그러나 공업화학만은 연구자들의 비율이 조금 높아지리라는 예측과는 달리 다소 낮아지고 있었다. 결론적으로 7년 동안 화학자들의 하위주제간의 추이과정을 분석하여 얻은 각 하위주제의 고정확률은 앞으로의 연구자들의 확률분포를 대체적으로 예측해 주고 있다고 말할 수 있었다. 가설 2를 위한 실험은 1983년 이후의 각 하위주제에 대한 화학자들의 비율을 예측하기 위해서 1967년부터 1983년 동안 648명의 저자가 쓴 총논문수 4,598편을 분석하여 각 하위주제의 초기상태확률과 고정확률을 구하여 비교분석하였다. 분석결과 1983년 이후에는 초기상태확률과 비교해서 무기화학, 분석화학, 유기화학, 생화학 분야의 화학자들의 관심이 고분자, 물리화학, 공업화학 분야로 쏠릴 것으로 예측된다. 세 번째 실험은 동시 인용법을 이용하여 특정분야의 지적구조를 규명하듯이 마코프모형의 추이행렬을 이용하여 2차원의 공간에 하위주제를 나타내는 점들을 매핑해 보았는데 대체로 각 하위주제가 독립된 영역임을 나타내 보이고 있었고 군집분석결과 유

기화학, 생화학, 분석화학이 모여서 하나의 군집을 이루고 공업화학, 물리화학, 무기화학이 모여서 또다른 군집을 이루었는데 고분자화학은 어느 군집에도 속하지 않고 따로 떨어져 있다. 중심에 위치한 물리화학 분야는 6개의 타하위주제들에서 물리화학에로의 추이조건확률의 합이 가장 큰 타하위주제들과의 관계가 가장 밀접한 하위주제이며 그 어느 군집에도 속하지 않고 뚝 떨어져 고립된 모습을 보이는 고분자화학은 6개의 타하위주제들에서 고분자화학으로의 추이확률의 합이 가장 작은 타하위주제들과의 관계가 가장 미미한 하위주제였다. 세 번째 실험결과를 동시인용빈도에 의한 유사성을 바탕으로 다차원축적기법과 군집분석에 의해서 규명한 화학분야의 지적구조와 비교, 분석해 볼만한 의미가 있는 실험사례가 있었다.

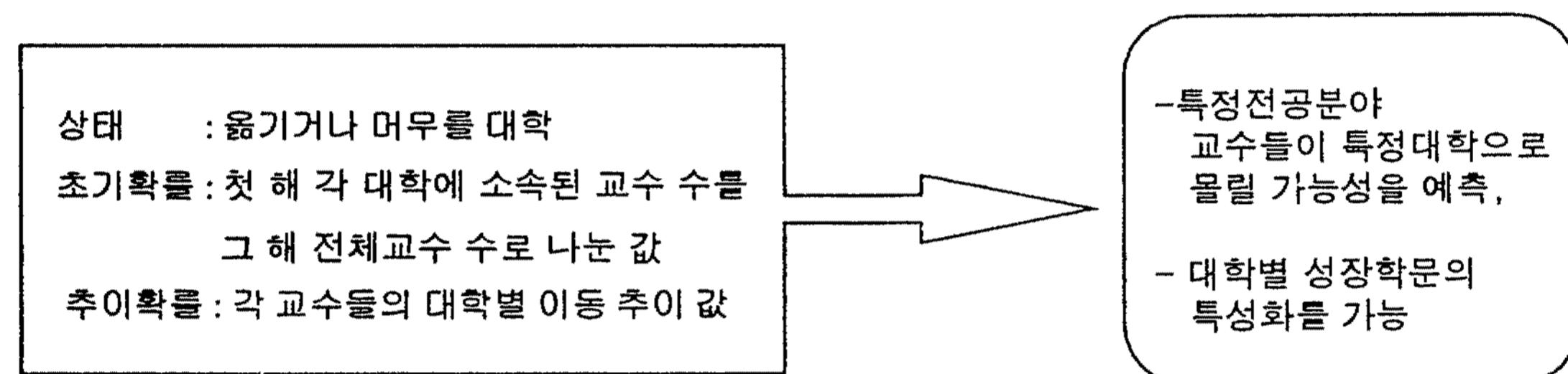
5. 계량정보학에서의 마코프모형응용

Zunde와 Slamecka은 마코프모형을 응용한 학문개발처리 모형을 정립한 바 있으며, 고프만은 문헌의 동적과정 연구에서 마코프모형을 응용하였고, 기호를 사용하여 하위주제에서 저자들의 변동상황을 설명하고 예측하기 위해 마코프모형을 사용하였다. 그는 주혈흡증병과 수두세포 주제분야에서 저자를 분석하기도 하였다. 최근에 들어서 마코프모형이론은 물리학과 사회과학 그리고 병리학, 기술과학, 경영학에서 많이 응

[그림 7] 장서개발 및 수서정책 응용



[그림 8] 저자이동 연구



용되고 있음을 볼 수 있는데 이러한 마코프모형을 계량정보학에 응용할 수 있는 몇가지 분야를 제안하였다.

5.1 장서개발 및 수서정책

도서관의 장서개발 특히 수서정책에 마코프모형을 이용하여 수서시스템의 미래 행위를 예측하고, 효과적인 운영을 하는데 응용할 수 있다. 저자의 저술실태를 조사함으로써, 장서개발에 있

어서 강화, 유지, 혹은 쇠퇴될 학문영역에 대한 수서기준의 과학적이고 보다 객관적인 근거자료로써 활용할 수 있다. 최근 10년동안의 9개 학문분야들에 대한 저술실태가 변화하는 것을 조사한다고 하면, 우선 상태를 저술할 학문분야들로 보고, 초기확률은 첫해에 9개 학문분야에서 발표된 저술건수들 각각을 그해 전체 저술건수로 나눈값이 된다. 이를 통해서 주제영역

별 변화추이값인 추이확률을 구하고, 패키지를 통해서 안정된 확률분포를 나타내는 고정확률을 구할 수 있다. 여기서 초기확률과 고정확률을 비교해보면, 가까운 미래에 발전할 학문분야와 쇠퇴 영역을 예측할 수 있다. 즉, 도서관 이용자의 현재와 미래 연구관심영역의 경향을 예측하여, 이에 대응한 도서관 장서를 구성하고 보다 적합한 서비스를 준비할 수 있다는데 그 응용효과가 있다.

5.2 저자이동연구

마코프모형 기법은 특별히 관련성이 있는 저자들그룹 사이에서 그들의 주요연구분야가 변화하는 것에 의해 특정저자의 연구분야 이동연구에 응용될 수 있다. 마찬가지로, 특정 저널그룹 사이에서 발표경향은 이러한 저널들 사이에서 인용의 이동추이에 따르게 될 것이다. 즉, 이러한 저자

및 인용 이동추이를 통해서 미래의 연구경향을 예측할 수 있는데 마코프 모형이 응용될 수 있다.

5.3 대학간 교수이동

대학에 소속된 교수들이 일정기간 동안, 일정 대학들 사이에서 이동하는 추이를 또한 마코프모형을 통해서 예측할 수 있다. 최근 십년간 a 대학에서 f 대학에 소속하여 근무하는 교수들이 매년 대학을 옮길 수 있거나 그대로 머물 수 있다고 가정하면, 상태를 이동할 수 있는 대학으로 볼 수 있다. 마코프모형의 첫번째 조건인 이산-시 확률과정을 충족시키고, 옮겨갈 수 있는 상태공간을 6개 대학(a, b, c, d, e, f)으로 한정함으로써, 마코프모형의 두번째 조건을 충족시킨다. 또한 현재의 연구성과나 관심에 따라서 옮겨가게 될 것이므로 세번째 조건까지 모두 충족된다. 여기서 초기 확률은 각 대학에 소속된 첫해 교수수를 전체 교수수로 나누어 산출할 수 있으며, 각 교수들의 대학별 이동 추이를 추적하여, 추이확률을 구할 수 있다. 이를 통해서 어느 전공분야의 교수들이 어느 특정대학으로 몰릴지 가능성을 예측해 볼 수 있게 될 것이며, 이를 통해서 대학별 성장 학문의 특성화를 가늠해볼 수 있는 연구효과가 있다.

5.4 재정정책

마코프모형은 재정정책 수립에도 영향을 미칠 수 있는데, 하부주제에서

인용이나 저술 이동과정을 추적함으로써, 한정된 학문내에서 연구관심분야의 이동연구에 관심을 가지고, 그런 정보에 근간해서 학문적으로 기대되는 성장 분야와 퇴보 분야에 대한 기준을 가지고 재정정책을 세울 수 있다. 즉 인용이 많이 되는 영역의 자료구입 예산배정을 높게 잡고, 저술실태가 저조하거나 자주 인용되지 않고 있는 영역의 자료구입 예산배정을 낮게 혹은 삭감할 수 있도록 예측하여, 도서관 시스템운영을 위한 보다 과학적이고 합리적인 예산수립을 세울 수 있다. 분야별 이용률 즉 대출율을 구하여 마찬가지로, 초기확률, 추이확률 그리고 고정확률을 구할 수 있을 것이며 이를 비교하여 예산배정을 위한 합리적이고 객관적인 예측으로 시스템을 운영할 수 있는 마코프모형의 응용효과가 있다.

이상의 응용분야에서 마코프모형 기법을 적용함에 있어서, 추이행렬의 검증을 계산하는 과정은 상당히 많은 단계를 제곱승하는 번거로움이 있는데 이런 과정은 MINITAB 패키지 등을 사용하여 보다 쉽고 간편하게 응용할 수 있다. 또한 BASIC 해석기가 유용할 수 있으며, SAS나 SPSS 그리고 최근에는 GAUSS를 보편적으로 적용할 수 있다.

6. 결 론

이상에서 마코프모형에 대한 기본적인 이론과 이 모형을 활용한 두 가지 실험사례를 통해서 계량정보학에서의

응용방안을 제안하였다. 그러나 보다 복잡한 상황 즉 휴간이나 폐간저널에 대한 문제점이 있었으며, 마코프모형이 학문발전과정을 기술하고 예측할 수 있다는 가설을 검증하는 데 있어서 보다 충분한 실험데이터를 가지고 진행될 수 있어야 할 것이다. 본 연구에서는 마코프모형의 도서관 실무분야에 대한 응용가능성을 제안하는 선에서 이루어졌으며 본 네가지 제안에 대한 실제실험은 수행되지 않았다. 이러한 가능성검토를 바탕으로 구체적인 상태를 설정하여 심도깊은 실험을 향후연구하여 시도하고자 한다.

〈참 고 문 헌〉

- 김현희. “마코브체인을 이용한 한국화학자의 공식커뮤니케이션의 구조적 분석”, 정보관리연구 20(1), pp. 66~85, 1989.
- 김현희. “학술커뮤니케이션의 수량학적 분석에 관한 연구”, 도서관학 14집, pp. 93~130, 1987.
- 사공철, 최정희. “문현정보학분야의 통계기법 사용에 관한 계량적 연구”, 도서관학논집 20, pp. 31~60, 1993.
- 윤구호. “계량서지학적 연구의 이론과 실제”, 정보관리연구 20(1), pp. 1~29, 1989.
- 윤희윤. 마르코프모형에 의한 대학도서관의 도서대출빈도예측 및 대출기간 설정:사회과학도서를 중심으로 경북대학교 석사논문, 1985.
- 정영미. “계량서지학적 연구에 관한 고찰”, 도협월보 19(1), pp. 3~9. 1978.
- Cinlar, Erhan. *Introduction to Stochastic Processes*. Englewood Cliffs : Prentice Hall, 1995.
- Feller, William. *An Introduction to Probability Theory and Its Applications*. New York : Academic Press, 1997.
- Gilks, W. R. et.al. *Markov Chain Monte Carlo in Practice*. CRC Press, 1996.
- Hogg, Robert V. & Craig, Allen T. *Introduction to Mathematical Statistics*. 4th ed. New York : Macmillan Publishing, 1978.
- <http://www.cindoc.csic.es/cybermetrics/links03.html>
- <http://sci.hkbu.edu.hk/math/markov.html>
- Kalos, Marvin H. & Whitlock, Paula. A. *Monte Carlo Methods*. John Wiley & Sons, 1996.
- Kleinrock, Leonard. *Queueing Systems*. New York : John Wiley & Sons
- Larson, Ray R. “Bibliometrics of the World Wide Web,” *JASIS* 33, 1996.
- Pao, Miranda Lee & McCreery, Laurie. “Bibliometrics Application of Markov Chains,” *Information Processing & Management*, 22(1), pp. 7~17.
- Stewart, William J. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1995.