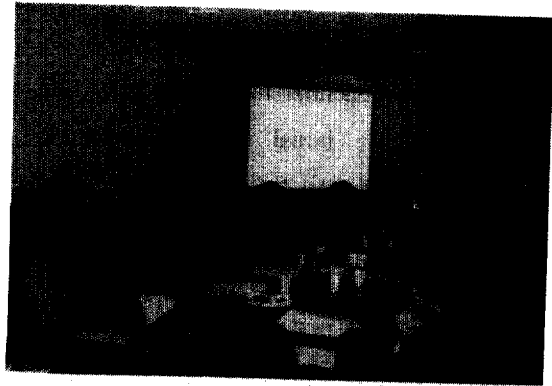


제 3 차 WORKSHOP

『EM Algorithm과 통계적 추정』

- 일 시 : 1995. 11. 9 (16:00~18:30)
- 장 소 : 충북대학교 자연과학대학(40동) 2층 교수회의실
- 주 관 : 충북개발연구원, 한국통계학회 충북지회
- 참석자 : 학계, 지역 연구기관, 충북 지역 경제유관기관의 전문가 40여명
- 발표자 : 조세현 박사(충북개발연구원 비상근연구원, 미시간주립대학교 교육심리통계학 박사)

정보(missing data) 와 EM algorithm에 입각하여 다층선형구조 모형(multilevel structural equation models)을 수립하고, ML 추론이론과, EM알고리즘을 설계하여 컴퓨터 프로그램을 개발하였다.



■ 논문발표 요지

지난 30년동안 응용통계학계에 주어진 여러가지 중요한 연구과제 가운데 하나는 여하이 다층구조의 자료(hierarchical data)를 분석하기 위한 선형구조 방정식 모형을 만들고, ML추론이론을 정립하며, 계산수행을 위한 알고리즘을 설계하여 컴퓨터 프로그램(software)을 만드느냐 하는 것이었다.

불균형자료에 대하여 그리고 부적합해결(improper solution) 문제에 대하여 본 논문은 세계 최초로 현대 통계학의 핵심 개념인 완전정보(complete data), 불완전

EM알고리즘은 우리가 불완전정보에 대한 Poosterior 분포의 해석적 표현을 얻을 수 있으며, 완전정보의 최대확률추정치(MLE)의 해석적 표현이 가능할 때는 언제나 강력한 계산도구로 사용할 수 있으며, 추정치가 안정적이다. 이 모형과 통계이론은 multivariate random coefficient (MRLM) 모델을 submodel로 사용할 수 있다는데 모형의 장점이 있다.

본 논문에서 저자는 한개의 예시를 통하여 empirical Bayes의 기본원리와 Model의 수립과정을 보이고, 인공자료를 사용하여 EM 알고리즘의 정확도를 검색하였다. 분석결과는 이 새로운 통계모형

과 이론 그리고 저자가 창안 개발한 MASEM 컴퓨터 프로그램은 변량이 Hierarchy 별로 변동하는 경우의 통계분석방법으로 광범위한 연구분야 (의학, 사회과학, 유전공학, 공학, 농화학, 농학) 에서 사용할 수 있음을 보여주고 있다.

■ 주요 토의내용

▶ initial starting values 의 지정문제

-분포가 exponential family 인 경우 EM algorithm 은 initial starting values 에 굳건한 (robust) 것으로 알려져 있다. 그리고 저자가 창안 개발한 MASEM (다층자료의 분석용 일반 프로그램) 컴퓨터 프로그램은 pseduo-MLE 로 initial starting values 를 계산하는 알고리즘을 따로 내장했다.

▶EM algorithm 에서 찾는 최적값에 대한, global maxima 와 local maxima 의 문제

-현재까지 연구된 바에 의하면 (Wu, 1983), local 또는 global 에 대한 판정을 할 수가 없다. 그러나 그것은 알고리즘 설계분야에서 중요한 주제가 아니며, 알고리즘 설치의 핵심은 수렴에 있다. 추정치가 수렴되지 않으면 제안된 추정이론과 알고리즘에 의한 추정은 불가능한 것이다.

▶model 의 적합도 검정문제

-일반 likelihood ratio test로 모델의 적합도를 계산하고 RMQR(root mean squared residuals) 을 계산할 수 있다.

▶저자의 Ph. D 논문을 지도한 Dr. Raudenbush는 본인의 논문이 역사에 기록될만한 것으로 평가한 바 있다. 그것은 저자가 세계 최초로 Unbalanced Multilevel Structural Equation Models, ML 추정이론과 EM 알고리즘을 개발하여 이 분야의 연구에 공헌했기 때문이다. LISREL 이론과 프로그램이 Sweden 의 Dr. Joreskog 에 의하여 창안 개발되어 보급된 것처럼 (LISEL 모형을 사용해 Data 를 분석하여 발표한 논문의 수가 3,000편 이상이 된다), hierarchical data 에 대하여 MASEM 이론과 프로그램이 그와 같은 역할을 할 수 있다. 본 논문의 모형, 이론, 그리고 알고리즘과 프로그램은 등록되어, 지적재산권에 관한 국제협약에 의해 보호되고 있으며, 어떠한 형태로도 논문의 어떠한 부분도 인용할 수 없다.