혼합분포모형의 매개변수 추정방법 비교

Comparison of Three Parameter Estimation Methods for Mixture Distributions

신주영*, 김수영**, 김태림***, 허준행****

Ju-Young Shin, Sooyoung Kim, Taereem Kim, Jun-Haeng Heo

요　　지

상이한 자연현상으로 발생된 자료들은 때때로 통계적으로 다른 특성을 가지는 경우가 있다. 이런 자료들은 다른 두 개 이상의 모집단에서 자료가 발생한 것으로 가정할 수 가 있다. 기존에 널리 사용되어온 분포형 모형의 경우 단일한 모집단으로부터 자료가 발생한다는 가정하에서 개발된 모형들로 위에서 언급한 자료들을 적절히 모의할 수 없다. 이런 상이한 모집단에서 발생된 자료를 모형화 하기 위해서 혼합분포모형(mixture distribution)이 개발되었다. 홍수나 가뭄 등과 같은 극치 사상의 경우 다양한 자연현상들로부터 발생하기에 혼합분포모형을 적용할 경우 보다 정확한 모의가 가능하다. 혼합분포모형은 두 개 이상의 비혼합분포모형들을 가중합하여 만들어진다. 혼합분포모형의 형태로 인하여 기존의 분포형 모형의 매개변수 추정 모형으로 널리 사용되던 최우도법 (maximum likelihood method), 모멘트법(method of moment), 확률가중모멘트법 (probability weighted moment method) 등을 이용하여 혼합분포모형의 매개변수를 추정하는 것이 용이 하지 않다. 혼합분포모형의 매개변수 추정 방법으로는 Expectation-Maximization (EM) 알고리즘, Meta-Heuristic Maximum Likelihood (MHML) 방법, Markov Chain Monte Carlo (MCMC) 방법 등이 적용되고 있다. 현재까지 수자원 분야에서 사용되는 극치 자료를 혼합분포모형을 이용하여 모의할 때 매개변수 추정방법에 따른 특성에 대한 연구가 진행되지 않았다. 본 연구에서는 우리나라 연최대강우량 자료를 이용하여 혼합분포모형의 매개변수 추정방법 (EM 알고리즘, MHML 방법, MCMC 방법) 들의 특성들을 비교 분석하였다. 혼합분포모형으로는 Gumbel-Gumbel 혼합분포모형을 적용하였다. 본 연구의 결과는 향후 혼합분포모형을 이용한 연구에 좋은 기초자료로 사용될 수 있을 것으로 판단된다.

핵심용어 : 혼합분포모형, Expectation-Maximization 알고리즘, Metaheuristic maximum likelihood 방법, Marcov Chain Monte Carlo, 매개변수 추정

* 정회원・연세대학교 대학원 사회환경시스템공학부 토목공학과 박사후연구원・E-mail : ausran@yonsei.ac.kr

** 정회원・연세대학교 대학원 사회환경시스템공학부 토목공학과 박사후연구원・E-mail : palewords@gmail.com

*** 정회원・연세대학교 대학원 사회환경시스템공학부 토목공학과 통합과정・E-mail : taereem@yonsei.ac.kr

**** 정회원・연세대학교 대학원 사회환경시스템공학부 토목공학과 교수・E-mail : jhheo@yonsei.ac.kr