ABSTRACT

Many researchers encounter the missing data problem. The phenomenon may be occasioned by data omission, non-response, death of respondents, recording errors, among others. It is important to find an appropriate data imputation technique to fill in the missing positions. In this study, the Expectation Maximization (EM) algorithm and two of its stochastic variants, stochastic EM (SEM) and Monte Carlo EM (MCEM), are employed in missing data imputation and parameter estimation in multivariate t distribution with unknown degrees of freedom. The imputation efficiencies of the three methods are then compared using mean square error (MSE) criterion. SEM yields the lowest MSE, making it the most efficient method in data imputation when the data assumes the multivariate t distribution. The algorithm's stochastic nature enables it to avoid local saddle points and achieve global maxima; ultimately increasing its efficiency. The EM and MCEM techniques yield almost similar results. Large sample draws in the MCEM's E-step yield more or less the same results as the deterministic EM. In parameter estimation, it is observed that the parameter estimates for EM and MCEM are relatively close to the simulated data's maximum likelihood (ML) estimates. This is not the case in SEM, owing to the random nature of the algorithm.