Povzetek

V diplomskem delu so predstavljene osnove statistične analize manjkajočih podatkov. Uvod zajema sam problem, vzorce in mehanizme manjkajočih podatkov ter metode, ki se uporabljajo pri reševanju. V naslednjih dveh poglavjih so opisana osnovna orodja iz verjetnosti in inferenčne statistike, ki jih pri analizi manjkajočih podatkov potrebujemo. Podrobneje so predstavljene cenilke maksimalnega verjetja in iterativni EM algoritem za iskanje le teh. V zadnjem poglavju je EM algoritem natančneje izpeljan za primer polinomske in dvorazsežne normalne porazdelitve. Za oba primera so dodani programi v Matlabu, skupaj z različnimi numeričnimi primeri. Za zaključek je dokazana konvergenca k rešitvi.

Math. Subj. Class. (2000): 65C60, 62N02, 62F10.
Ključne besede: Manjkajoči podatki, maksimalno verjetje, MLE cenilke, EM algoritem.

Abstract

In this work the basic theory of analysis with missing data is introduced. In the beginning the problem of missing data, missing data patterns, mechanisms that lead to missing data and different methods that handle them are presented. After the introduction the basic probability and statistical inference theory are reviewed. The attention is on maximum likelihood estimators and iterative EM algorithm. In the last chapter EM algorithm is implemented and tested with multinomial and bivariate normal distribution in Matlab. The last part covers the proof of its convergence property.

Math. Subj. Class. (2000): 65C60, 62N02, 62F10.
Keywords: Missing Data, Maximum Likelihood, Maximum Likelihood Estimators, EM algorithm.

Literatura

[1] Geoffrey J. McLachlan and Thriyambakam Krishnan, *The EM algorithm and Extensions*. Wiley series in probability and statistics, 1946.

[2] Roderick J. A. Little and Donald B. Rubin, *Statistical Analysis with Missing Data*. Wiley series in probability and statistics, 2nd Edition, 2002.

[3] George Castella and Roger L. Berger, *Statistical Inference*. Duxbury, Thomas Learning, 2nd Edition, 2001.

[4] Paul D. Allison, *Missing Data*. Sage University Papers Series, Thousand Oaks, 2001.

[5] Thomas Alan Severini, *Likelihood methods in Statistics*. Oxford University Press, New York, 2000.

[6] Douglas C. Montgomery and George C. Runger, *Applied Statistics and Probability for Engineers*. Wiley, 4nd Edition, 2006.

[7] Rajko Jamnik, *Verjetnostni račun*. Mladinska knjiga, 1971.

[8] David C. Howell, *Statistical Methods for Psychology*. Wadsworth Publishing, 7 edition, 2009.

[9] Matej Mlakar, *Jensenova neenakost*. 2009, <http://www2.arnes.si/∼mmlaka10/Clanki/Jensen.pdf>.

[10] Stanford Weisberg, *Applied Linear Regression*. Wiley series in probability and statistics, New Yersey, 3nd Edition, 2005.

[11] Brani Vidakovic, *EM Algorithm and Mixtures*. <http://www2.isye.gatech.edu/ brani/isyebayes/bank/handout12.pdf>.

[12] Vladimir Batagelj, *Verjetnostni račun in statistika*. 2005, <http://vlado.fmf.uni-lj.si/vlado/VeSt/vest03.pdf>.