ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
DEPARTMENT OF STATISTICS

ΚΥΚΛΟΣ ΣΕΜΙΝΑΡΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ ΝΟΕΜΒΡΙΟΣ 2018

Dr. Leonardo Egidi

Postdoctoral Researcher, Department of Economics, Business, Mathematics and Statistics, University of Trieste

Pivotal methods for Bayesian mixture models and classical k-means algorithm initialization

ΠΕΜΠΤΗ 29/11/2018
13:00

ΑΙΘΟΥΣΑ Τ101, 1ος ΟΡΟΦΟΣ
ΝΕΟ ΚΤΙΡΙΟ ΟΠΑ, (ΤΡΟΙΑΣ 2)

ΠΕΡΙΛΗΨΗ

In many statistical applications, such as mixture models and clustering, there is a natural allocation of the statistical units into a given number of groups. In order to summarize the groups' structure, one can try to detect those pivotal units that should be representative of the groups they belong to. This R package implements some of these procedures, which turned out to be well suited for dealing with Bayesian mixture models estimation (Egidi et al., 2018a) and k-means clustering (Egidi et al., 2018b).

Label switching is a well-known and fundamental problem in Bayesian estimation of mixtures, occurring when the likelihood and posterior distribution are invariant to permutations of the parameters. Due to this property, the Markov Chain Monte Carlo (MCMC) samples simulated from the posterior distribution result to be non-identifiable. The package provides a pivotal procedure for relabelling the MCMC chains and retrieving good posterior estimates.

K-means algorithm is one of the most popular procedures in data clustering. Despite its large use, one major criticism is the impact of the initial seeding on the final solution. Some pivotal methods proposed in this package result to be beneficial for initializing the centers of the classical k-means in order to obtain a better clustering solution.

All the methods will be illustrated using both real and simulated data and graphical tools will be adopted for assessing the quality of the solutions. The package is freely available at www.github.com/leoegidi/pivmet.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
DEPARTMENT OF STATISTICS

AUEB STATISTICS SEMINAR SERIES NOVEMBER 2018

Dr. Leonardo Egidi

Postdoctoral Researcher, Department of Economics, Business, Mathematics and Statistics, University of Trieste

Pivotal methods for Bayesian mixture models and classical k-means algorithm initialization

THURSDAY 29/11/2018
13:00

ROOM T101, 1ST FLOOR,
NEW AUEB BUILDING, (TROIAS 2)

ABSTRACT

In many statistical applications, such as mixture models and clustering, there is a natural allocation of the statistical units into a given number of groups. In order to summarize the groups' structure, one can try to detect those pivotal units that should be representative of the groups they belong to. This R package implements some of these procedures, which turned out to be well suited for dealing with Bayesian mixture models estimation (Egidi et al., 2018a) and k-means clustering (Egidi et al., 2018b).

Label switching is a well-known and fundamental problem in Bayesian estimation of mixtures, occurring when the likelihood and posterior distribution are invariant to permutations of the parameters. Due to this property, the Markov Chain Monte Carlo (MCMC) samples simulated from the posterior distribution result to be non-identifiable. The package provides a pivotal procedure for relabelling the MCMC chains and retrieving good posterior estimates.

K-means algorithm is one of the most popular procedures in data clustering. Despite its large use, one major criticism is the impact of the initial seeding on the final solution. Some pivotal methods proposed in this package result to be beneficial for initializing the centers of the classical k-means in order to obtain a better clustering solution.

All the methods will be illustrated using both real and simulated data and graphical tools will be adopted for assessing the quality of the solutions. The package is freely available at www.github.com/leoegidi/pivmet.