Accepted Manuscript

A mixture model-based nonparametric approach to estimating a count distribution

Chew-Seng Chee

PII: S0167-9473(16)30286-9
DOI: http://dx.doi.org/10.1016/j.csda.2016.11.012
Reference: COMSTA 6388

To appear in: *Computational Statistics and Data Analysis*

Received date: 27 February 2016
Revised date: 17 November 2016
Accepted date: 19 November 2016

Please cite this article as: Chee, C.-S., A mixture model-based nonparametric approach to estimating a count distribution. *Computational Statistics and Data Analysis* (2016), http://dx.doi.org/10.1016/j.csda.2016.11.012

A mixture model-based nonparametric approach to estimating a count distribution

Chew-Seng Chee

School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

Abstract

The density of a mixture distribution with unknown discrete mixing distribution can be a way of finding a nonparametric estimate of a density. Comparing with a standard parametric approach that runs the risk of model misspecification and a kernel-based nonparametric approach that retains all data points for constructing a density estimate, this mixture model-based nonparametric approach that naturally circumvents these problems is very appealing to density estimation. Owing to these advantages, this approach to estimating a count distribution is considered. Estimation of a count distribution via this means necessarily involves the problem of nonparametric estimation of a discrete mixing distribution. The nonparametric estimation problem is formulated using the family of power divergences, thus offering a class of nonparametric estimators of a discrete mixing distribution. A fast algorithm is presented for computing the class of nonparametric minimum power divergence estimates with which the corresponding class of mixture model-based estimates of a count distribution can be implemented. Only the use of the probability mass function of the Poisson mixture distribution with unknown discrete mixing distribution is illustrated. The simulation results show that the mixture model-based estimators on average outperform the kernel-based estimators for uncontaminated samples of larger sizes and contaminated samples of all sizes considered.

Keywords: Count data, Nonparametric minimum power divergence estimation, Poisson mixture model, Power divergence family.

1. Introduction

Mixture models (see, e.g., Titterington *et al.*, 1985; McLachlan and Peel, 2000) have marked at least a century of history with evidence, e.g., of Pearson's (1894) work. Nowadays, mixture models are commonly postulated as statistical models for various real life problems due to their flexibility and generality, and one of their important applications is the estimation of density

Email address: chee@umt.edu.my (Chew-Seng Chee)