

Modeling Compositional Data Using Dirichlet Regression Models

Rafiq H. Hijazi

Department of Statistics, UAE University, Al-Ain, UAE.

Email: rhijazi@uaeu.ac.ae

Robert W. Jernigan

Department of Mathematics and Statistics, American University, Washington, USA.

Email: jernigan@american.edu

Abstract

Compositional data are non-negative proportions with unit-sum. These types of data arise whenever we classify objects into disjoint categories and record their resulting relative frequencies, or partition a whole measurement into percentage contributions from its various parts. Under the unit-sum constraint, the elementary concepts of covariance and correlation are misleading. Therefore, compositional data are rarely analyzed with the usual multivariate statistical methods. Aitchison [1] introduced the logratio analysis to model compositional data. Campbell and Mosimann [9] suggested the Dirichlet Covariate Model as a null model for such data. In this paper, the maximum likelihood estimation methods in Dirichlet regression are developed and the sampling distributions of these estimates are investigated. Measures of total variability and goodness of fit are proposed to assess the adequacy of the suggested models in analyzing compositional data.

Keywords: Compositional data, Dirichlet distribution, logratio analysis, maximum likelihood estimation.

2000 Mathematics Subject Classification: 62F12, 62J02, 62J20, 62P12.