



# Discriminant analysis in small and large dimensions

April 2017

TARAS BODNAR<sup>a,1</sup>, STEPAN MAZUR<sup>b</sup>, EDWARD NGAILO<sup>a</sup> AND NESTOR PAROLYA<sup>c</sup>

<sup>a</sup> *Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden*

<sup>b</sup> *Unit of Statistics, School of Business, Örebro University, Fakultetsgatan 1, SE-70182 Örebro, Sweden*

<sup>c</sup> *Institute of Statistics, Leibniz University Hannover, Königsworther Platz 1, D-30167 Hannover, Germany*

## Abstract

We study the distributional properties of the linear discriminant function under the assumption of normality by comparing two groups with the same covariance matrix but different mean vectors. A stochastic representation for the discriminant function coefficients is derived which is then used to obtain their asymptotic distribution under the high-dimensional asymptotic regime. We investigate the performance of the classification analysis based on the discriminant function in both small and large dimensions. A stochastic representation is established which allows to compute the error rate in an efficient way. We further compare the calculated error rate with the optimal one obtained under the assumption that the covariance matrix and the two mean vectors are known. Finally, we present an analytical expression of the error rate calculated in the high-dimensional asymptotic regime. The finite-sample properties of the derived theoretical results are assessed via an extensive Monte Carlo study.

ASM Classification: 62H10, 62E15, 62E20, 60F05, 60B20

*Keywords:* discriminant function, stochastic representation, large-dimensional asymptotics, random matrix theory, classification analysis

---

<sup>1</sup>Corresponding Author: Taras Bodnar. E-Mail: [taras.bodnar@math.su.se](mailto:taras.bodnar@math.su.se). Tel: +46 8 164562. Fax: +46 8 612 6717. This research was partly supported by the Swedish International Development Cooperation Agency (SIDA) through the UR-Sweden Programme for Research, Higher Education and Institutional Advancement. Stepan Mazur acknowledges financial support from the project "Ambit fields: probabilistic properties and statistical inference" funded by Villum Fonden