Roy’s Largest Root Test Under Rank-One Alternatives

Iain M. Johnstone†, Boaz Nadler∗
Department of statistics
Stanford University
Stanford, CA
e-mail: imj@stanford.edu
Department of Computer Science
Weizmann Institute of Science
Rehovot, Israel
e-mail: boaz.nadler@weizmann.ac.il

Abstract: Roy’s largest root is a common test statistic in a variety of hypothesis testing problems. Despite its popularity, obtaining accurate tractable approximations to its distribution under the alternative has been a longstanding open problem in multivariate statistics. In this paper, assuming Gaussian observations and a rank one alternative, also known as concentrated non-centrality, we derive simple yet accurate approximations for the distribution of Roy’s largest root test for five of the most common settings. These include signal detection in noise, multivariate analysis of variance and canonical correlation analysis. Our main result is that in all five cases Roy’s test can be approximated using simple combinations of standard univariate distributions, such as central and non-central $\chi^2$ and $F$. Our results allow approximate power calculations for Roy’s test, as well as estimates of sample size required to detect given (rank-one) effects by this test, both of which are important quantities in hypothesis-driven research.

AMS 2000 subject classifications: Primary 62H20, 62H10; secondary 15A18, 47A55.

Keywords and phrases: Roy’s largest root test, greatest root statistic, concentrated non-centrality, MANOVA, canonical correlation, matrix perturbation.

Tests based on the largest eigenvalue of a sample covariance matrix, and extensions, have a long history in multivariate analysis, statistical signal processing and allied fields. The exact distributions of these tests have complicated forms that have perhaps limited their use in application. The aim of this paper is to use a small noise perturbation approach to derive simple and often accurate approximations to the power of a class of such largest root tests.

∗Part of this work was performed while the author was on sabbatical at UC-Berkeley and at Stanford Department of Statistics.
†This research was supported in part by NSF DMS 0906812, NIH BIB R01EB1988 and BSF 2012-159.