

HIGH-DIMENSIONAL HYPOTHESIS TESTING USING RANDOM SUBSPACES

Måns Thulin¹

¹ Uppsala University, Sweden

We discuss multivariate hypothesis testing problems in the high-dimensional setting where the dimensionality p is larger than the number of observations n , with emphasis on tests for the equality of two mean vectors. Most high-dimensional tests for the equality of two mean vectors rely on naive diagonal or trace estimators of the covariance matrix, ignoring dependences between variables. We show how random subspaces can be used to construct a computer-intensive test that incorporates dependences.

Several two-sample tests are compared in settings motivated by microarray gene expression data, related to the genetic problem of testing whether a set of highly dependent gene expressions differ between two populations. The random subspaces test is seen to offer higher power than the competing tests when the gene expressions are dependent.

Finally, we discuss problems related to invariance properties of high-dimensional tests, and how random subspaces can be used in other hypothesis testing problems.

Keywords: Gene expression, high-dimensional, hypothesis testing, random subspace, two-sample test.

References:

Thulin, M. (2014). A high-dimensional two-sample test for the mean using random subspaces. *Comput. Stat. Data An.*, 74, 26–38.