

# The leave- $p$ -out estimator of the prediction error as a $U$ -statistic and its asymptotic tests

Mathias Fuchs

Institut für Medizinische Informationsverarbeitung Biometrie und Epidemiologie,  
Ludwig-Maximilians-Universität München,  
Marchioninstr. 15, 81377 München, Germany

**Abstract.** We outline some consequences of identifying the leave- $p$ -out estimator of the generalization error rate of a machine learning algorithm or the prediction error of a model as a  $U$ -statistic. In particular, asymptotic normality holds true, in contrast to usual cross-validation.

Furthermore, an appropriate variance estimator leads to an asymptotically exact test of the null hypothesis that two such algorithms have equal error rate. Such a test seems to be new even for cases of few variables.

We present an application to tuning parameter choice in lasso regression on a gene expression data set and conclude with a discussion of computational aspects of the leave- $p$ -out-statistic.