Report on the PhD thesis:

Selected aspects of functional estimation and testing

by Luboš Prchal

The thesis concerns two very modern fast developing statistical fields: functional data analysis and statistical analysis of ROC curves. The main contribution of the thesis lies in developing statistical models suitable for rather complex data sets, investigation of properties of these models, conducting extensive simulation studies in order to check appropriateness of these models on the artificial data sets. Finally, the developed procedures are applied to the real data sets, where the application is not at any case straightforward, moreover, a suitable software has to be also developed and a number of tuning constants has to be suitably chosen.

The thesis is based on 7 papers, 6 of them written with the coauthors ([1] – [6]) and one by Luboš Prchal ([7]) . Chapter I is introductory. Chapters II – IV concern Functional data analysis while Chapters V – VII are devoted to ROC curves. The mani results are contained in Chapters III, IV, VI, VII.
Chapter II contains an introduction into the area, basic notions are introduced, the main problems are formulated and basic achievements in the area related to the thesis obtained so far are surveyed.
Chapter III concerns application of spline estimators in functional linear regression and autoregression. Particularly, alternative estimators based on B-splines are introduced and their applicability is studied. In the second part of the chapter the developed estimation procedures is applied to the real data set– electricity consumption in Sardegna, where the problem is to predict electricity consumption during coming weekend or weekdays based on the actual one.
In Chapter IV permutation tests for testing of no effect of the covariate on a functional response versus there is an effect in the functional regression setup are developed and studied. The procedures are based on nonparametric kernel estimators. The developed tests are based on are permutational principle. It appears to be a quite strong tool for the problem. Attention is paid to the choice of various tuning constants. A simulation study is conducted that includes also results on power and finally application to the real data set concerning atmospheric radiation (data provided by Czech Hydrometeorogical Institute).
Chapter V provides an introduction into the area of the kernel ROC curves estimation. Basic notions, basic results as well as results obtained so far are presented in a comprehensive way. .
Chapter VI focuses on data driven kernel ROC curve estimators. The author develops a modification of the kernel ROC curve estimation suitable in situation when the diagnostic variables have highly skewed marginal distributions and the standard procedures usually fail. Simulation study is conducted and, moreover, the developed procedure is applied to the set of linguistic data.
In Chapter VII a test of equivalence of two ROC curves is developed and its asymptotic properties under the null hypothesis are studied. Theoretical results are then applied to the linguistic data. Simulation study is also conducted, where choice of various tuning constant is deeply discussed.

The thesis covers quite large area of modern statistical modeling. a number of new statistical procedures (tests and estimators) are developed. Some theoretical properties are studied, simulation studies are conducted in order to check the behavior of the developed statistical procedures on artificial data sets, i.e., when the true model is known. Afterwards the procedures are on quite complex real data (electricity consumption in Sardegna (Italy), atmospheric radiation profiles (Czech republic) and linguistics data sets (Czech Republic)).
The main contribution of the thesis lies in developing useful statistical procedures for various situations, studying their properties with he emphasis on computationally feasible procedures with still reasonable data fit. Also computational aspects are deeply discussed, some suitable algorithms are developed.

Here a few comments and questions:

• Simulation study in Chapter III is made for only very smooth functions. There is a naturel question: is the behavior similar for functions with certain nonsmoothness?

• In Chapter IV, Section 3.2: Are really the residuals uncorrelated?

• In Chapter IV the approximation to the critical values for the developed tests are obtained via permutation principle. Just simulations are made. Are the approximations to the critical values asymptotically correct? Is the test procedure consistent?

• In Chapter VII the developed approximation to the critical values is based on the limit distribution under the null hypothesis. Therefore the approximation is asymptotically correct when data follow the null hypothesis. But having a real data set it is not known whether data follow the null hypothesis or some alternatives. The question: is the resulting test consistent w.r.t what alternatives? Theorem VII.1 holds true only for $1 \leq j \leq J$, where J is fixed. The question is how to choose J in particular situation, is there any recommendation?

The above questions and remarks are just small comments and have no any influence on the positive evaluation of the thesis.

Prchal's thesis is clearly written, mathematics is correct. But a big attention is paid to practical implementation of the developed procedures on a computer. A number of new statistical procedures is introduced, their basic theoretical properties are studied. The focus is not on the technical part but on developing statistical procedures, motivations and application to some quite complex real data sets, where the application of the developed procedures is not at all straightforward but demands adjustment of the models as well procedures. It was necessary to choose properly various tuning constants that influence considerably the final results. He shows the ability to handle practical implementation of the developed procedures.

The thesis clearly proves author's ability for creative scientific work in mathematical statistics particularly in part connected with computational aspects. From my point of view, there is no doubt that this work fulfills all requirements of a PhD. thesis in mathematical statistics. I strongly recommend acceptance of this work.

Prof. RNDr.Marie Hušková,
Department of Statistics, Charles University,
Sokolovská 83, 186 75 Praha,
Czech Republic
huskova@karlin.mff.cuni.cz
April 4, 2008

References

[1] Spline estimator for the functional linear regression (coauthor Pascal Sarda) (2008). Submitted to the J. Multivariate Analysis.

[2] Electricity consumption prediction with functional linear regression using splines estimators (coauthors Jaromr Antoch, Rosaria de Rosa and Pascal Sarda)(2008) Submitted to the Journal of Applied Statistics.

[3] Changes in atmospheric radiation from the statistical point of view (coauthor Daniel Hlubinka) (20080 J.Computational Statistics 51, 4926 – 4941.

[4] Statistical methods for analysis of meteorogical measurements. (coauthors J.Antoch and D. Hlubinka)(2008) Submitted to Computational Statistical data Analysis.

[5] No effect and lack-of-fit permutation tests for functional regression. To appear in Computational Statistics, 2008 (coauthors Harv Cardot and Pascal Sarda)

[6] Nonparametric comparison of ROC curves: testing equivalence and clustering (coauthors Jaromr Antoch and Pascal Sarda)

[7] Data-driven kernel ROC curve estimators for skewed diagnostic variable. (2008)