NIPS*95 Workshop: Missing Data: Methods and Models:

Multiple Imputation and Simulation Methods
Multiple Imputation for Neural Networks and Classification Trees

Brian D. Ripley

Professor of Applied Statistics, University of Oxford,
1 South Parks Road, Oxford OX1 3TG, UK. `ripley@stats.ox.ac.uk`
`http://www.stats.ox.ac.uk/~ripley`

Abstract

Multiple imputation methods complete missing data more than once, and average (in a suitable sense) over the predictions made with each of the possible completed datasets. To complete the data, simulation methods are used to sample from the conditional distribution of what is missing given what is known; iterative methods such as the Gibbs sampler and 'data augmentation' may need to be used.

In talk 1 these methods are illustrated on a simple logistic regression and an opinion survey in Slovenia. In talk 2 the mechanics of using multiple imputation with classification trees and neural networks (MLPs , RBFs, etc) are illustrated through a worked example.

References

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995) *Bayesian Data Analysis.* Chapman & Hall.

Little, R. J. A. & Rubin, D. B. (1987) *Statistical Analysis with Missing Data.* New York: Wiley.

Ripley, B. D. (1996, actually Nov 1995) *Pattern Recognition and Neural Networks.* Cambridge University Press. ISBN 0 521 48086 7.
[For details see `http://www.stats.ox.ac.uk/~ripley/PRbook`.]

Rubin, D. B. (1978) Multiple imputations in sample surveys: a phenomological Bayesian approach to non-response (with discussion). *Proc. Amer. Statist. Assoc. Survey Methods Section*, pp. 20–34.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys.* Wiley.

Rubin, D. B. (1995?) Multiple imputation after 18 years (with discussion). *J. Amer. Statist. Assoc.* to appear.

Rubin, D. B., Stern, H. S. & Vehovar, V. (1995) Handling "don't know" survey responses: the case of the Slovenian plebiscite. *J. Amer. Statist. Assoc.* **90**, 822–828.

Tanner, M. A. (1993) *Tools for Statistical Inference.* Second edition. Springer.

Tanner, M. A. & Wong, W. H. (1987) The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82**, 528–550.