Review of depmixS4: An R-package for hidden Markov Models

January 22, 2010

This is a review of the paper "depmixS4: An R-Package for hidden Markov Models" by Ingmar Visser and Maarten Speekenbrink. The `depmixS4 R` package provides a comprehensive collection of functions for estimating Markovian models.

The paper provides a nice overview of the basic idea behind dependent mixture models and their practical applications. Following a brief review of the methodology the authors provide several examples illustrating the functionality of the package. They then go on to show how users can easily build their own functions on top of the `depmixS4` package structure.

There are several minor points which should be addressed/revised.

- In equation (1) on page 3 the authors use $\mathbf{O}_t$, which more than likely is equal to $(O_t^1, \ldots, O_t^m)$, but it is not explicitly defined.

- In equation (9) `ln` is used whereas in the other components of the equation `log` is used.

- In Section 3 the authors provide some examples of the `depmix` function. There several brief comments here.

  1. In Section 3.1 it would be helpful if the authors made a clearer connection between the observation that fast participant response seems to lead to a chance level of accuracy and slower participant response seems to lead to a higher level of accuracy and why a two state process would be appropriate in this context.

  2. A bit more detail on the `balance` data discussed in Section 3.5.

  3. More discussion on the arguments `instart`, `respstart` and `trstart`, corresponding to the starting values for the prior, transition and response models.

  4. In Section 3.5 at the top of page 12, a clearer description of the results of the model. Specifically the prior probabilities and the relationship between age and response accuracy.

All told I found the paper and package to be quite interesting and timely. HMM's see wide use in a number of applications and having a very general, comprehensive package will be very helpful. Excepting the points above I believe that this paper should be approved for publication.

That being said I believe that this package will draw a great deal of interest both inside and outside the statistics community. For this reason, while certainly *not* requiring it, I believe elaborating on the applications and results pertaining the `speed` and `balance` data sets would contribute considerably to the content.