depmixS4: an R-package for hidden Markov models

Ingmar Visser[1] & Maarten Speekenbrink[2]

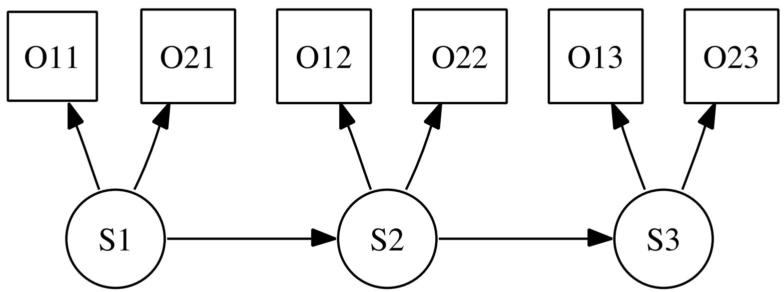
[1]Department of Psychology
University of Amsterdam

[2]Department of Psychology
University College London

Psychometric Computing, February 2011, Tuebingen

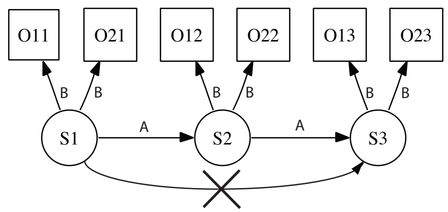
Outline

Example model



- $S_1, S_2, \ldots$: discrete states (latent or hidden)
- $O_{11}, O_{21}, O_{12}, \ldots$: observations (yes/no, RT, ...)
- For example: O_{11}, O_{21} are items on a balance scale task
- States represent different strategies that change through learning
- Dependency between S and O forms the measurement model
- Dependency between S's forms the dynamic part of the

Dependent mixture model formulation



1. $S_t = \mathbf{A}S_{t-1} + \xi_t$, $\mathbf{A}$, a transition matrix
2. $O_t = \mathbf{B}(S_t) + \zeta_t$, $\mathbf{B}$, an observation density
3. $Pr(S_t|S_{t-1}, \ldots, S_1) = Pr(S_t|S_{t-1})$ (Markov assumption)

Likelihood

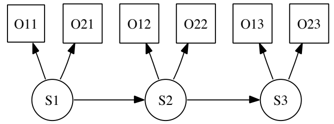
$$Pr(\mathbf{O}_1, \ldots, \mathbf{O}_T) = \sum_q \prod_{t=1}^{T} Pr(\mathbf{O}_t | S_t, \mathbf{A}, \mathbf{B}),$$

q an arbitrary hidden state sequence

- q: an enumeration of all possible state sequences (n^T)
- Leave out the sum over q (S_t known): complete data likelihood
- Note: likelihood is not computed directly (impractical for large T)

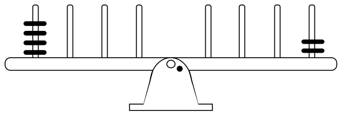
Relationship to other models

1. Latent Markov model
2. Dependent Mixture model
3. Bayesian network (with latent variables)
4. State-space model (discrete)
5. Symbolic dynamic model
6. Regime switching models
7. ...

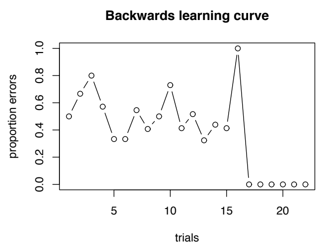


Why do we need them?

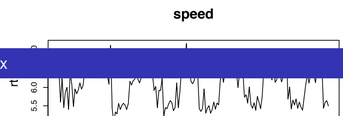
1. Piagetian development, conservation, balance scale
2. Concept identification learning
3. Strategy switching: Speed-accuracy trade-off
4. Iowa Gambling task
5. Weather Prediction task
6. Climate change



[Jansen and Van der Maas, 2002]



[Schmittmann et al., 2006]



DepmixS4

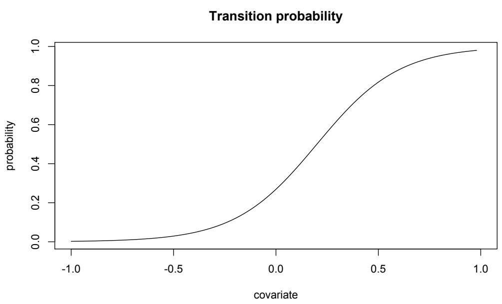
- `R`-package
- `depmixS4` fits `dep`endent `mix`ture models
- mixture components are generalized linear models (and others ...)
- Markov dependency between components

In short: `depmixS4` fits hidden Markov models of generalized linear models in both large N, small T as well as N=1, T large samples.
[Visser and Speekenbrink, 2010]

Transition & initial model

Each row of the transition matrix and the initial state probabilities:

- is modeled as a multinomial distribution
- uses the logistic link function to include covariates
- can have time-dependent covariates



Response models

Current options for the response models are models from the generalized linear modeling framework, and some additional distributions.

From `glm`:

- normal distribution; continuous, gaussian data
- binomial (logit, probit); binary data
- Poisson (log); count data
- gamma distribution

Additional distributions:

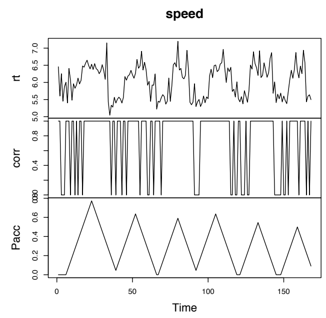
- multinomial (logistic or identity link); multiple choice data
- multivariate normal
- exgaus distribution (from the `gamlss` package); response time data
- it is easy to add new response distributions

Optimization

`depmixS4` uses:

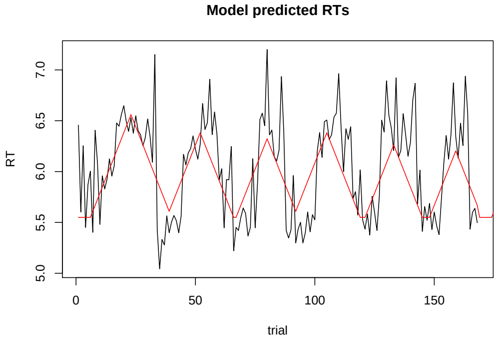
- EM algorithm (interface to `glm` functions in R)
- Direct optimization of the raw data log likelihood for fitting contrained models (using `Rsolnp`)

Speed-accuracy: data

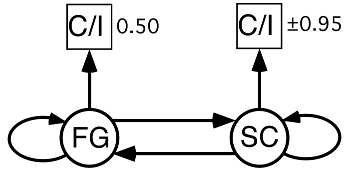


- Three blocks with N=168,134,137 trials (first block shown)
- Speeded reaction time task
- Speed and accuracy manipulated by reward variable
- Question: is there a single (linear) relationship between responses and covariate or switching between regimes?

Speed-accuracy: linear model predictions

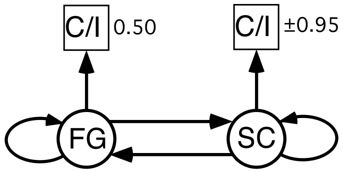


Speed-accuracy: two-state model



- FG=fast guessing
- SC=stimulus controlled
- Response times also modeled
- Pay-off for accuracy as covariate on the transition probabilities

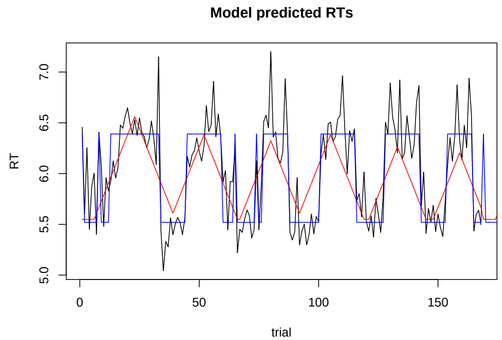
Speed-accuracy: two-state model



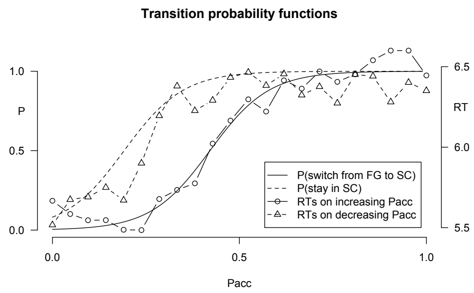
Fitting this in `depmixS4`:

```
mod1 <- depmix(list(rt~1,corr~1),
+ data=speed, transition=~Pacc, nstates=2,
+ family=list(gaussian(), multinomial("identity")),
+ ntimes=c(168,134,137))
fm1 <- fit(mod1)
```


Speed-accuracy: switching model



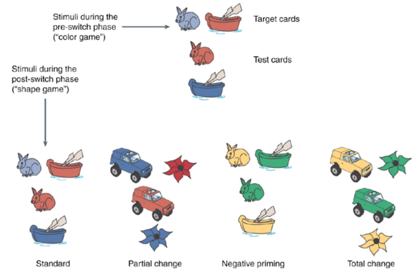
Speed-accuracy: transition probabilities



- transition probabilities as function of covariate
- hysteresis: asymmetry between switching from FG to SC and vice versa

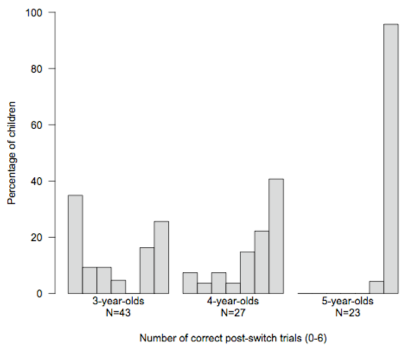
What is DCCS?

- task 1: sort by color
- task 2: sort by shape
- measures: ability to switch/flexibility



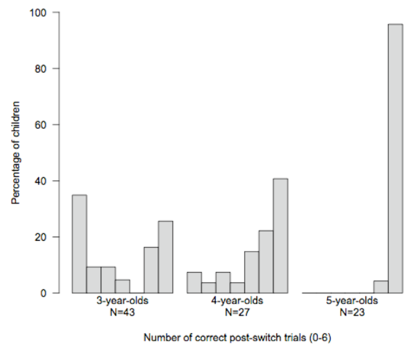
DCCS: data

- data consists of 6 trials (task 2)
- traditional analyses:
  1. 0/1 correct: perseveration
  2. 5/6 correct: switching



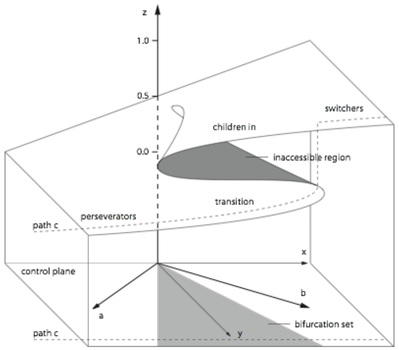
DCCS: research questions

- can we characterize the remaining group?
- are there children in transition, shifting from one strategy to another?
- alternative: are they simply guessing?



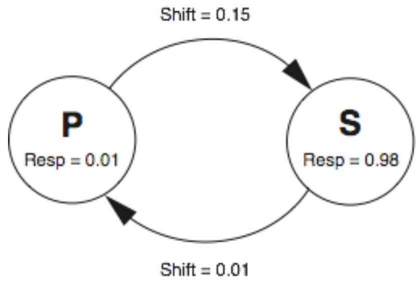
DCCS: theory

- cusp model predicts instability in the transitional phase
- shifting back and forth between ‘strategies’
- hysteresis: assymetry in transition probabilities



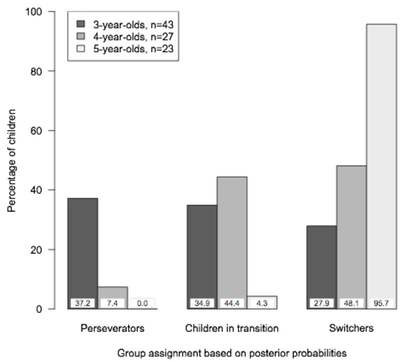
DCCS: results

- P: perseveration state
- S: switch state
- transition P->S much larger than transition S->P
- this model better than a model without transitions



DCCS: results

- 30 to 40 % of 3/4 year olds are in the transitional phase, shifting between strategies



Other applications

- Climate change data
- Learning on the Iowa Gambling Task
- Balance scale task
- Categorization learning

Take home messages

- depmixS4 can be downloaded from: http://r-forge.r-project.org/depmix/ or from CRAN
- Also models with transient and absorbing states
- Easy to add your own favorite distribution
- Paper in Journal of Statistical Software: depmixS4: An R-package for Hidden Markov Models
- This is not a psychometrics package, rather a psychodynamics package

Thanks

- Thanks to Han van der Maas for the speed-accuracy data
- Thanks to Bianca Beersma for the DCCS data (paper in press!)
- Happy mixing!

Dutilh, G., Wagenmakers, E.-J., Visser, I., and van der Maas, H. L. J. (2011).
A phase transition model for the speed–accuracy trade–off in response time experiments.
Cognitive Science.

Jansen, B. R. J. and Van der Maas, H. L. J. (2002).
The development of children's rule use on the balance scale task.
Journal of Experimental Child Psychology, 81(4):383–416.

Schmittmann, V. D., Visser, I., and Raijmakers, M. E. J. (2006).
Multiple learning modes in the development of rule-based category-learning task performance.
Neuropsychologia, 44(11):2079–2091.

Visser, I., Raijmakers, M. E. J., and Van der Maas, H. L. J. (2009).
Hidden markov models for individual time series.
In Valsiner, J., Molenaar, P. C. M., Lyra, M. C. D. P., and Chaudhary, N., editors, *Dynamic Process Methodology in the Social and Developmental Sciences*, chapter 13, pages 269–289. Springer, New York.

Visser, I. and Speekenbrink, M. (2010).
depmixS4: An R-package for hidden Markov models.
Journal of Statistical Software, 36(7):1–21.
R package, current version available from CRAN.

Future developments

1. richer measurement models, eg factor models, AR models etc
2. richer transition models, eg continuous time measurement occasions
3. explicit state durations
4. identifiability of models
5. model selection

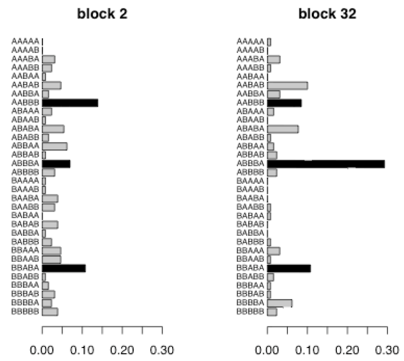
Experiment

- categorization learning experiment
- 32 learning blocks with feedback
- 5 transfer blocks without feedback

Research questions:

1. detect different patterns of generalization (rule-based vs. exemplar-based)
2. study representational shifts in learning: does representational format change with learning?

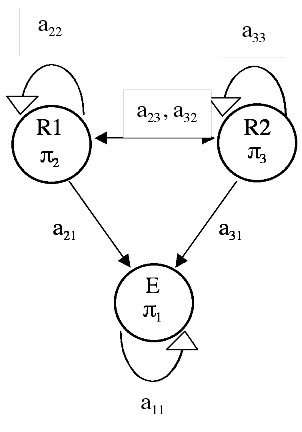
Data: transfer trials



1. 32 possible generalization patterns (for 5 binary items)
2. Rule 1: AABBB
3. Rule 2: BBABA
4. Exemplar: ABBBA
5. Occurrence of exemplar-based responding increases
6. Data fairly noisy

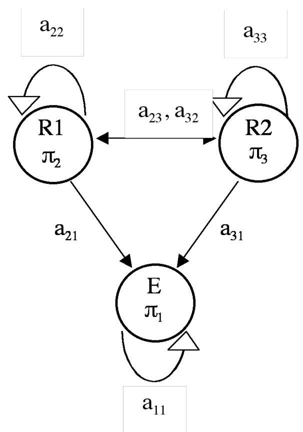
Model specification

- 3 states representing Rule 1, Rule 2 and Exemplar based responding
- models with 2 to 5 states were fitted
- model selection by BIC



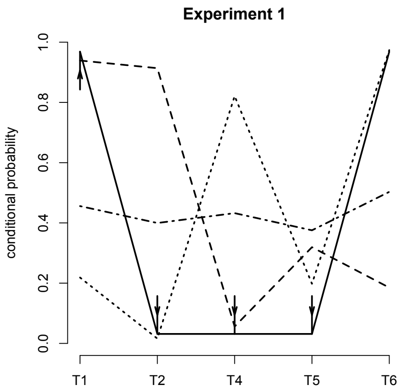
Results (1)

- the selected model has another state corresponding to guessing behavior
- the model includes a covariate on the probability 'correct' in the exemplar state
- this tests the assumption that consistency of applying this strategy increases with training



Results (2)

- 4 states representing Rule 1 (AABBB), Rule 2 (BBABA), Exemplar (ABBBA), and guessing
- covariate of block number on probability 'correct' in the exemplar state (increasing consistency)



Results (3)

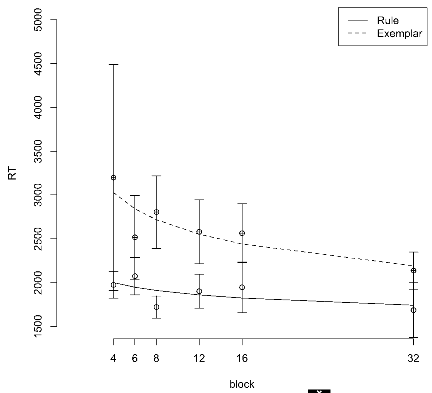
- Exemplar state is absorbing
- Transitions occur mostly from Res to R1 and R2 and from rule states to the Exemplar state

Estimated transition matrix of LMM 4a of Experiment 1.

	Transition probabilities			
State	E	R1	R2	Res
E	1.00	0.00	0.00	0.00
R1	0.02	0.91	0.03	0.04
R2	0.11	0.00	0.89	0.00
Res	0.09	0.06	0.07	0.78

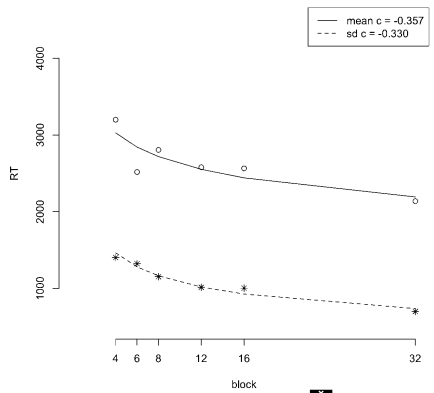
Results (4)

- Posterior assignment of responses to states
- Response times for the rule based versus exemplar based strategies over training
- Results indicate that exemplar based responding is an expression of automatization

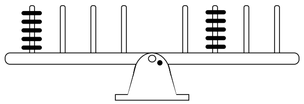


Results (5)

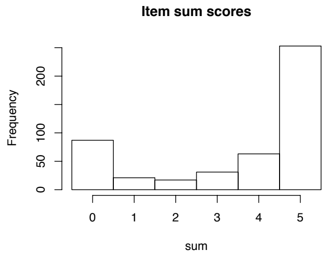
- Automatization should result in a power law of learning
- Mean and sd have identical coefficients



Data



- 5 distance items on the balance scale task
- age as covariate
- items scored binary



Data provided by Brenda Jansen (Jansen & Van der Maas, 2002)

Model

3-class model with age as covariate on the class proportions