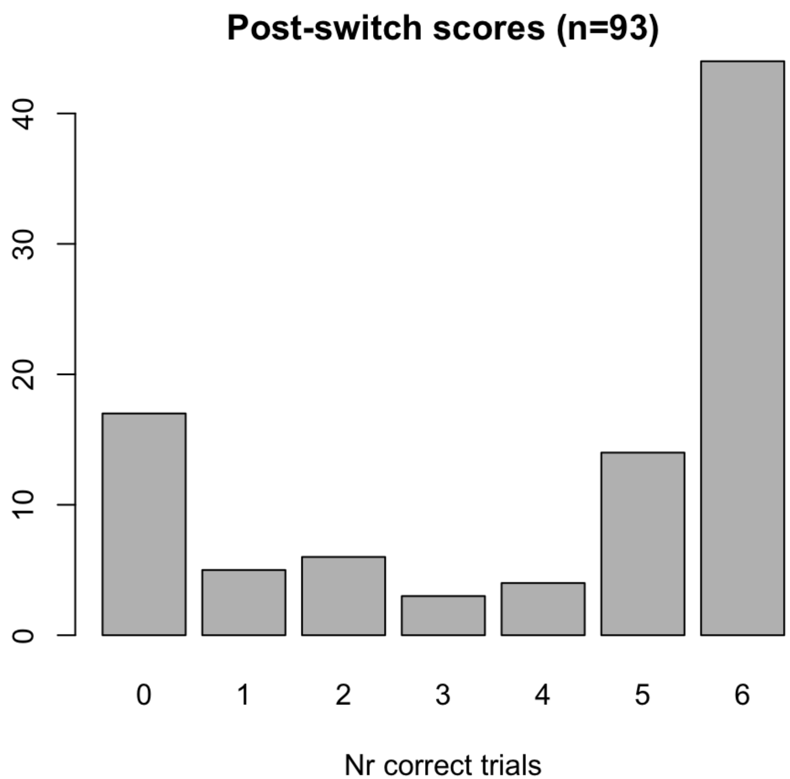
Dimensional Change Card Sort task

Binomial mixture model

Ingmar Visser (with contributions by Maarten Speekenbrink & Bianca van Bers)

DCCS mixture model: assumptions

1. Model with two components for the sum scores
2. Assume that the data are binomial (with n=6 items)



DCCS mixture model: specification

$$P(y_t) = \sum_{i=1}^{N} \pi_i P(y_t | \theta_i)$$

where:

- π_i are the mixture proportions
- $P(y_t | \theta_i)$ are the component specific distributions:

$$P(y_t | \theta_i) = \text{Binom}(\theta_i, n = 6),\ i = 1, 2,$$

with

- θ_i the success probabilities for the binomial distribution in each component

Fitting the mixture model with `depmixS4`

```r
library(depmixS4)
library(hmmr)
data(dccs)
head(dccs)
```

```
  pp ageM ageY ageGr    sex nTrPre nCorPre t1Post t2Post t3Post t4Post t5Post
1  1   71    5     4 female      6       6      1      1      1      1      1
2  4   66    5     4 female      6       6      1      1      1      1      1
3  5   70    5     4 female      7       6      1      1      1      1      1
4  6   62    5     4   male      6       6      0      1      1      1      1
5  7   66    5     4   male      6       6      1      1      1      1      1
6  8   66    5     4   male      7       6      1      1      1      1      1
  t6Post nCorPost passPost
1      1        6        1
2      1        6        1
3      1        6        1
4      1        5        1
5      1        6        1
6      1        6        1
```

Fitting the mixture model with `depmixS4`

```
m2 <- mix(response=cbind(nCorPost,6-nCorPost)~1,nstates=2,data=dccs,family=binomial())
```

The `mix()` function takes the following arguments:

1. `response`: a formula specifying the response models
2. `nstates`: the number of states/components
3. `data`: the data frame containing the variables in the response models
4. `family`: the family of the response models (as in the `glm` function), in this case a binomial distribution with, by default, a logit link

Fitting the mixture model with `depmixS4`

```r
seed <- 1234
set.seed(seed)
fm2 <- fit(m2)
```

```
converged at iteration 11 with logLik: -147
```

```r
fm2
```

```
Convergence info: Log likelihood converged to within tol. (relative change)
'log Lik.' -147 (df=3)
AIC:  300
BIC:  308
```

DCCS mixture model: results

```r
depmixS4::summary(fm2)
```

```
Mixture probabilities model
  pr1   pr2
0.328 0.672

Response parameters
Resp 1 : binomial
    Re1.(Intercept)
St1           -1.86
St2            2.70
```

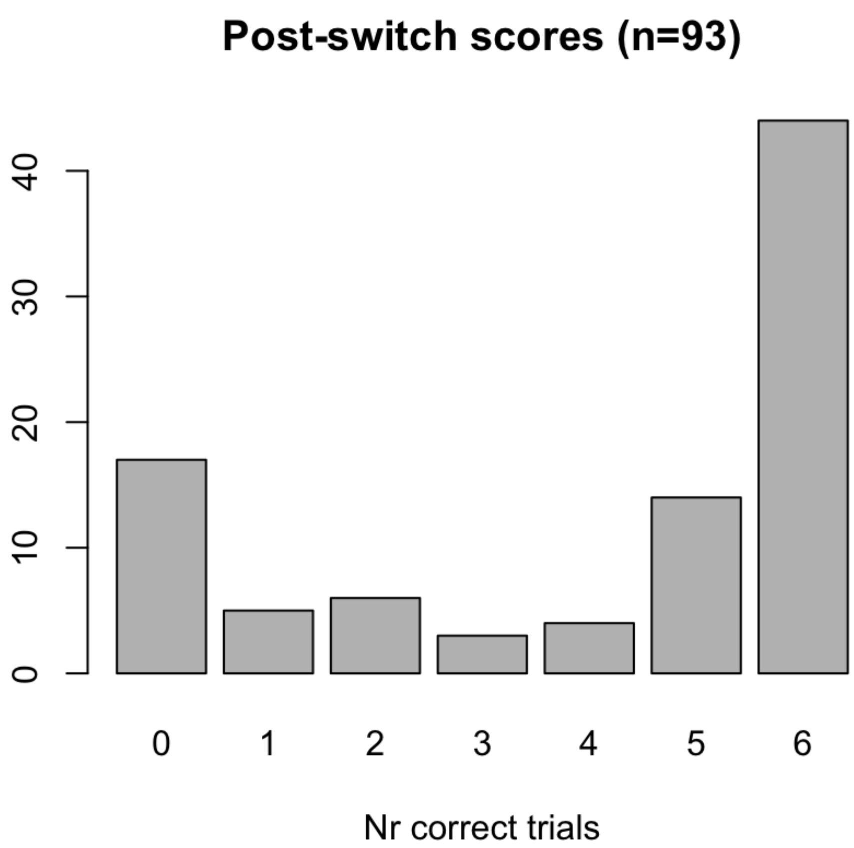
Note that the response parameters are in terms of the logit link. They can be transformed into probabilities as

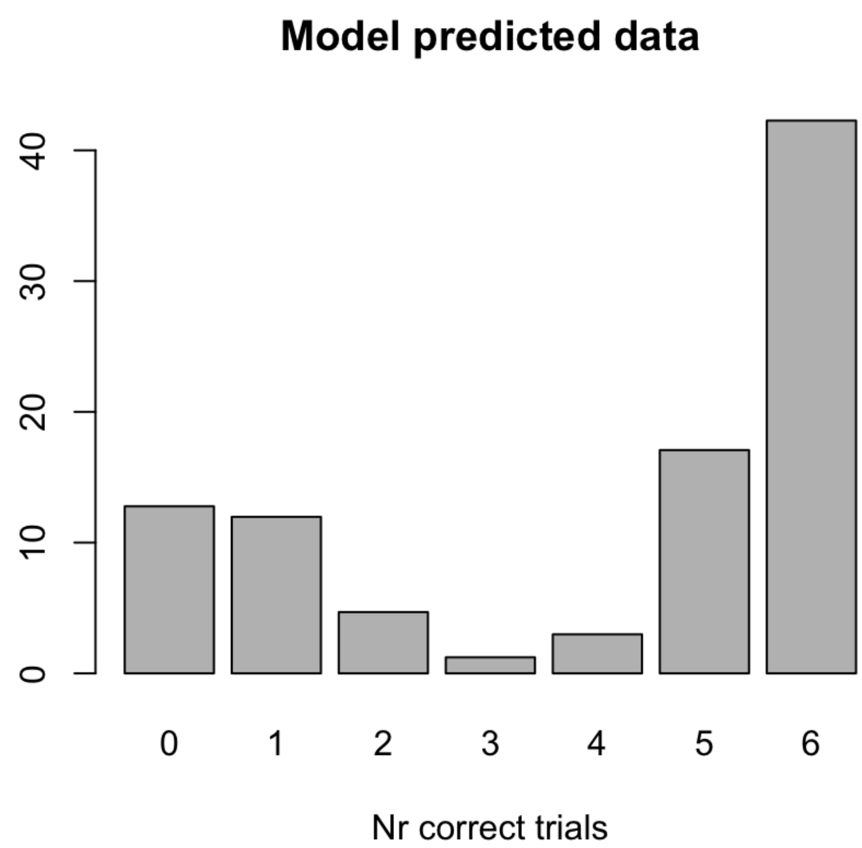
$$\theta_1 = \frac{-1.857}{1 + \exp(-1.857)} = 0.135$$

$$\theta_2 = \frac{2.699}{1 + \exp(2.699)} = 0.937$$

DCCS mixture model: model checking

The 2-component mixture model gives a good description of the data





Posterior state probabilities

```r
pst <- posterior(fm2)
head(pst,12)
```

```
   state       S1      S2
1      2 4.38e-06 1.00000
2      2 4.38e-06 1.00000
3      2 4.38e-06 1.00000
4      2 4.16e-04 0.99958
5      2 4.38e-06 1.00000
6      2 4.38e-06 1.00000
7      2 4.38e-06 1.00000
8      1 9.97e-01 0.00278
9      2 4.16e-04 0.99958
10     2 4.38e-06 1.00000
11     1 9.97e-01 0.00278
12     2 4.38e-06 1.00000
```

Average posterior state probabilities

```r
mpost <- aggregate(posterior(fm2)[,2],by=list(age=dccs$ageY),mean)
```

age	n	P(S = 1)	P(S = 2)
3	43	0.571	0.429
4	27	0.22	0.78
5	23	0	1

Table of posterior state assignments

age	n	S = 1	S = 2
3	43	25	18
4	27	6	21
5	23	0	23

Is a 2-component mixture model enough?

nstates	AIC	BIC
1	566.45	568.99
2	300.15	307.74
3	294.9	307.57
4	298.7	316.43
5	302.68	325.47

A 3-component model fits best according to the AIC and BIC. In addition, a bootstrap LR test rejects the 2-component model in favour of a 3-component model, $P(\text{LR} \geq 9.24) = .001$.

Results of a 3-component mixture

```
Mixture probabilities model
  pr1   pr2   pr3
0.163 0.186 0.650

Response parameters
Resp 1 : binomial
    Re1.(Intercept)
St1          -8.149
St2          -0.793
St3           2.877
```

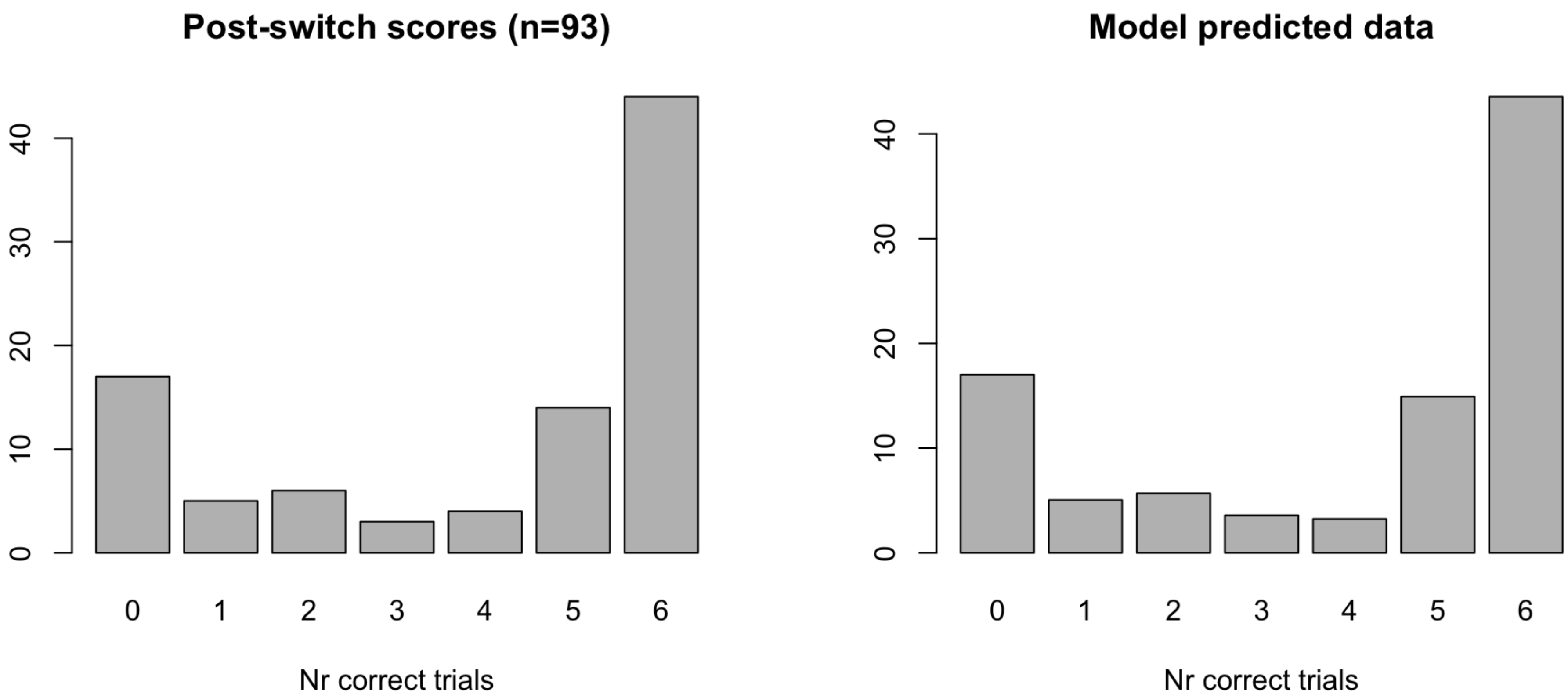
$$\theta_1 = \frac{-8.149}{1 + \exp(-8.149)} = 0$$

$$\theta_2 = \frac{-0.793}{1 + \exp(-0.793)} = 0.312$$

$$\theta_3 = \frac{2.877}{1 + \exp(2.877)} = 0.947$$

Model check 3-component mixture

The 3-component mixture model gives an excellent description of the data!



Average posterior state probabilities

age	n	P(S = 1)	P(S = 2)	P(S = 3)
3	43	0.311	0.27	0.418
4	27	0.066	0.211	0.723
5	23	0	0.001	0.999

DCCS mixture model with covariate on class membership

```r
m2a <- mix(cbind(nCorPost,6-nCorPost)~1,ns=2,data=dccs,
           family=binomial(),prior=~scale(ageM))
fm2a <- fit(m2a)
```

```
converged at iteration 10 with logLik: -129
```

```r
fm2a
```

```
Convergence info: Log likelihood converged to within tol. (relative change)
'log Lik.' -129 (df=4)
AIC:  266
BIC:  276
```

```r
fm2
```

```
Convergence info: Log likelihood converged to within tol. (relative change)
'log Lik.' -147 (df=3)
AIC:  300
BIC:  308
```

DCCS mixture model with covariate on class membership: results

```
summary(fm2a)
```

```
Mixture probabilities model
Model of type multinomial (mlogit), formula:
~scale(ageM)
Coefficients:
             St1  St2
(Intercept)    0 1.35
scale(ageM)    0 2.00
Probalities at zero values of the covariates.
0.207 0.793

Response parameters
Resp 1 : binomial
    Re1.(Intercept)
St1           -1.84
St2            2.72
```

$$\theta_1 = \frac{-1.841}{1 + \exp(-1.841)} = 0.137$$

$$\theta_2 = \frac{2.722}{1 + \exp(2.722)} = 0.938$$

DCCS mixture model: conclusions

1. It is correct to say that more 5 year olds than 3 year olds pass the task
2. It is *not* correct to say that *average* 3 year olds are worse at the DCCS than 4 and 5 year olds: the *average* 3 year old does not exist (it is like saying that the average sex is m/f)
3. In the 2-component model, there are two clear states (types) reflecting those who *can* and those who *can not* switch. In all age groups children *either* perform mostly correct *or* mostly incorrect and a mixture model captures that aspect of the data very well.
4. In the 3-component model, a third state (type) may reflect guessing behaviour. Alternatively, this could be people who learn to switch during the task. This possibility will be explored later with a hidden Markov model.