Web-based Supporting Materials for "Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms" by R. Langrock, B. J. Swihart, B. S. Caffo, N. S. Punjabi and C. M. Crainiceanu

A second simulation study

The simulation study given in the main manuscript considers a model which is quite different from the model fitted to the SHHS data: for example, it involves only two states, whereas in the application to the SHHS data we focus on the results for a five-state model. The reason for considering a model with far fewer parameters in that simulation study is simply that it allows us to compare the performance of the MLE and the two-stage method: in the case of five states the former method is too time-consuming for a simulation study.

In order to better evaluate the performance of the two-stage method in a setting similar to the SHHS application, we here give the results of another simulation study that focuses exclusively on the much faster two-stage method. We simulated from the model for sleep EEG outlined in Section 2.1 of the main manuscript, generating $T = 300$ observations – each drawn from one of $N = 5$ possible Dirichlet distributions as selected by an underlying Markov chain – for each of $M = 20$ individuals. To these twenty time series we then applied the first stage of the two-stage approach. This exercise was repeated 500 times, and the performance of the method was analyzed based on point and interval estimates of the Dirichlet parameters.

To replicate, at least to some extent, the setting of the application given in Section 4.3 of the main manuscript, the following parameter values were used in the simulations. We used Dirichlet distributions with parameter vectors

$$\begin{aligned}
\lambda^{(1)} &= (76.29,\ 7.99,\ 4.62,\ 2.47) \quad (\text{state } 1),\\
\lambda^{(2)} &= (68.83, 14.18,\ 8.81,\ 4.24) \quad (\text{state } 2),\\
\lambda^{(3)} &= (\ 7.80,\ 4.87,\ 6.08,\ 3.90) \quad (\text{state } 3),\\
\lambda^{(4)} &= (25.20,\ 8.74,\ 6.85,\ 4.13) \quad (\text{state } 4)\\
\text{and } \lambda^{(5)} &= (\ 0.83,\ 0.57,\ 0.56,\ 1.00) \quad (\text{state } 5),
\end{aligned}$$

i.e., precisely the Dirichlet distributions that were estimated in Section 4.3 of the main manuscript. For each run and individual the transition probability matrix of the associated underlying five-state Markov chain was drawn uniformly from the set of all 102 transition probability matrices that were estimated in the SHHS application. The difference between the setting here and that analyzed in Section 4.3 of the main manuscript is that here we consider fewer individuals and also less data per individual, and therefore much less data in total. This was done in order to ensure a reasonable computational effort in the simulation study.

Table 1 provides the following summary statistics for the parameter estimates: sample means of the estimates, sample standard errors of the estimates, and coverage proportions of 95% confidence intervals. The confidence intervals for the Dirichlet parameters were obtained in Stage I of the two-stage method based on the Hessian of the log-likelihood for the parameter estimates.

All confidence intervals have coverage proportions close to the desired value 0.95. For some estimates, there is an indication of a very small bias for some of the Dirichlet parameter estimators. However, the maximum (absolute) relative bias, $|(\hat{\theta} - \theta)/\theta|$ where $\hat{\theta}$ denotes the estimator of a parameter θ, over all Dirichlet parameter estimators is 0.9% (for $\lambda_2^{(5)}$) and is thus negligible. These results strongly indicate that the two-stage method works very well in the SHHS scenario analyzed in the main manuscript.

Table 1: Summary statistics for the Dirichlet parameter estimates in the (additional) simulation study. Columns in the table are: state-dependent Dirichlet parameter (with superscript denoting the state), true parameter value used to simulate data, ME = mean estimate, SE = standard error of estimates, and CC = coverage proportion of confidence intervals.

Parameter	True value	ME	SE	CC
State 1				
$\lambda_1^{(1)}$	76.289	76.744	3.176	0.942
$\lambda_2^{(1)}$	7.988	8.024	0.305	0.928
$\lambda_3^{(1)}$	4.617	4.635	0.165	0.944
$\lambda_4^{(1)}$	2.471	2.480	0.092	0.948
State 2				
$\lambda_1^{(2)}$	68.833	68.774	2.802	0.956
$\lambda_2^{(2)}$	14.179	14.171	0.583	0.954
$\lambda_3^{(2)}$	8.813	8.803	0.351	0.945
$\lambda_4^{(2)}$	4.239	4.233	0.158	0.944
State 3				
$\lambda_1^{(3)}$	7.801	7.799	0.301	0.950
$\lambda_2^{(3)}$	4.868	4.885	0.199	0.946
$\lambda_3^{(3)}$	6.082	6.098	0.264	0.958
$\lambda_4^{(3)}$	3.902	3.906	0.164	0.946
State 4				
$\lambda_1^{(4)}$	25.204	25.220	0.629	0.960
$\lambda_2^{(4)}$	8.737	8.752	0.213	0.952
$\lambda_3^{(4)}$	6.852	6.866	0.163	0.966
$\lambda_4^{(4)}$	4.128	4.132	0.099	0.962
State 5				
$\lambda_1^{(5)}$	0.830	0.833	0.043	0.970
$\lambda_2^{(5)}$	0.566	0.571	0.029	0.948
$\lambda_3^{(5)}$	0.559	0.559	0.028	0.964
$\lambda_4^{(5)}$	1.003	1.009	0.056	0.950