Mathematical Methods

The purpose of this note is to establish the theoretical basis for extending the work by Ashwood et al. I will first highlight the mathematical model of the hidden Markov model (HMM) used and then briefly cover the reasoning behind using (approximate) leave-one-out cross-validation to obtain the pointwise predictive probability to compare different models as opposed to normalized likelihoods in the case of Ashwood et al.

The Model

Observation Model

We assume that observations are generated from a $Bernoulli(p)$ distribution with the probability p given by

$$p = \Pr\left(y_t = 1 \mid z_t = k, u_t, \gamma_k\right) = \frac{1}{1 + \exp\left(u_t \gamma_k^\top\right)}$$

Here, $y = 1$ indicates a rightward decision and $y = 0$ indicates a leftward decision, z_t is the hidden state at time t, u_t represents the vector of predictors at time t and γ_k is the vector of coefficients for state k.

Thus, p is given by a sigmoid function and can be viewed as a logistic regression. The time-varying emission vector $\tilde{\mathbf{b}}(t)$ reflecting the probability of observing y_t in state j is given by

$$b_j(t) = \Pr\left(y_t \mid z_t = j\right)$$

Using the probability mass function of the Bernoulli distribution, we can write the observation model as

$$b_j(t) = \begin{cases} \frac{1}{1+\exp\left(u_t \gamma_j^\top\right)} & \text{if } y_t = 1, \\ 1 - \frac{1}{1+\exp\left(u_t \gamma_j^\top\right)} & \text{if } y_t = 0. \end{cases}$$

Transition Model

The time-varying transition matrix $\mathbf{A}(t) \in \mathbb{R}_+^{K \times K}$ is given by

$$a_{i,j}(t) = \Pr(z_t = j | z_{t-1} = i, \vec{u}_t) = \operatorname{softmax}\left(d_{i,j} + u_t \beta_i^\top\right) = \frac{\exp\left(d_{i,j} + u_t \beta_i^\top\right)}{\sum_{k=1}^{i} \exp\left(d_{i,j} + u_t \beta_i^\top\right)}$$

In other words, for each of the K states, a multinomial logistic regression is used to predict the probability of the next possible K states. Importantly, while we allow the intercept to vary between states, the coefficients are fixed and only depend on the previous state ($z_{t-1} = i$).

Fitting the Model

The Forward Algorithm

The forward algorithm is used to iteratively compute the probability of an observation sequence given the initial state probabilities π, the transition matrix $\mathbf{A}(t)$ and the emission vector $\tilde{\mathbf{b}}(t)$. By integrating (marginalizing) over all possible state sequences, we obtain the probability of the observation sequence up to time t, **only** conditional on the hidden state at time t (in theory, one could just naïvely sum over the probabilities under all possible state sequences, but this is intractable due to combinational explosion).

$$\alpha_{\mathrm{t}}(i) = \Pr\left(y_1 \ldots y_t, z_t = i\right)$$
$$\alpha_{t+1}(j) = \sum_{i=1}^{K} \alpha_t(i) a_{ij} b_j(t+1)$$

The initial value $\alpha_1(i)$ is calculated using π_i.

$$\alpha_{t=1}(i) = \pi_i \cdot b_i(1)$$

The sum over all possible end states at t is sufficient to compute the likelihood of the observation sequence $y_1, \ldots, y_t$.

$$\mathrm{L}(y_1, \ldots, y_t) = \sum_{i=1}^{K} \alpha_t(i)$$

Point Estimates

The forward algoritm can be combined with the so-called backward algorithm to compute the probability of a particular latent state at time t (the process is then fittingly termed forward-backward-algorithm). This information can then be used in an interative procedure known as expectation maximization (EM) to find the maximum likelihood (ML) or maximum a posteriori (MAP) estimate of the observation sequence. EM works by repeatedly finding the most likely state sequence (E-step) and subsequently maximizing the likelihood by adjusting the other parameters (M-step).

Other options include direct Maximal Likelihood estimation a Variational Bayes estimations.

Full Baysian Inference

Markov Chain Monte Carlo Sampling

Markov Chain Monte Carlo (MCMC) procedures such as Gibb's sampling or Hamitonian Monte Carlo (HMC) can be used to sample from the posterior distributions.

- K: Number of different states
- P: Number of predictors for transitions model.
- M: Number of predictors for observation model.