Online Supplement: Time-Evolving Psychological Processes Over Repeated Decisions

David Gunawan[1,5], Guy E. Hawkins[2], Robert Kohn[3,5], Minh-Ngoc Tran[4], Scott D. Brown[2]

1: School of Mathematics and Applied Statistics, University of Wollongong

2: School of Psychology, University of Newcastle, Australia

3: Australian School of Business, University of New South Wales, Sydney, Australia

4: Discipline of Business Analytics, University of Sydney Business School

5: Australian Research Council Center of Excellence in Mathematical and Statistical Frontiers

Bayesian Estimation Methods for the Markov Switching Model

This section discusses the proposed Markov Switching LBA model. Let $y_{i,j} = (RE_{i,j}, RT_{i,j})$ be the observed data for $i = 1, ..., N$ trials and $j = 1, ..., J$ subjects, where $RE_{i,j} \in \{1, 2\}$ is the response choice and $RT_{i,j}$ is the response time for the ith trial and jth subject. We assume that for subject j the observations $y_{i,j}$ depend on a hidden discrete Markov process $S_{i,j}$ with state space $\{1, ..., K\}$. The state switching stochastic process $S_{i,j}$ is described by the $(K \times K)$ transition matrix ξ_j, where $\xi_{k,l,j} := (\xi_j)_{kl}$ is the transition probability from state k to state l:

$$\xi_{k,l,j} = \Pr\left(s_{i,j} = l | s_{i-1,j} = k\right), \quad \text{for all} \quad k, l \in \{1, ..., K\}. \tag{1}$$

Conditional on knowing the states $s_j = (s_{1,j}, ..., s_{N,j})$, the observations $y_{i,j}$ are independent. For each trial i, the distribution of $y_{i,j}$ depends on one of the K random effects $\alpha_j = (\alpha_{1,j}, ..., \alpha_{K,j})$:

$$(y_{i,j} | s_{i,j} = k, \alpha_j) \sim LBA\left(y_{i,j} | \exp\left(\alpha_{k,j}\right)\right). \tag{2}$$

To account for the dependence of the random effects in different states, the prior distribution of the vector α_j is modelled as

$$\alpha_j | \mu, \Sigma \sim N\left(\alpha_j | \mu, \Sigma\right). \tag{3}$$

In our applications, the number of states is set to $K = 2$. The prior for μ is $N\left(0, I_D\right)$, with D the dimension of α_j; the prior for Σ is Inverse Wishart $IW\left(s_0 = I_D, v_0 = 30\right)$. The prior for the elements $\xi_{1,1,j}$ and $\xi_{2,2,j}$ of the transition matrix is Beta$(4, 1)$ for all j. The 95% probability interval for this prior is $(0.3993, 0.9939)$ implying a large range of values are possible.

Bayesian Inference

Let θ be the unknown group-level parameters and $p(\theta)$ the prior for θ. Let $y_j := (y_{1,j}, ..., y_{N,j})$ be the vector of observations for the jth subject and define $y := y_{1:J} = (y_1, ..., y_J)$ as the vector of observations for all J subjects. Let $\alpha_j := (\alpha_{1,j}, \alpha_{2,j})$ be the vector of random effects for subject j, and $p(\alpha_j|\theta)$ its density under the group-level distribution. Now define $\alpha := \alpha_{1:J} := (\alpha_1, ..., \alpha_J)$ as the vector of random effects for all J subjects.

We assume that the random effects α_j are independent given θ and that the observations $y_{i,j}$ are independent given θ, α_j, s_j, and ξ_j, i.e.,

$$p(\alpha_{1:J}|\theta) = \prod_{j=1}^{J} p(\alpha_j|\theta) \text{ and } p(y|\theta, \alpha_{1:J}, s_{1:J}) = \prod_{j=1}^{J}\prod_{i=1}^{N} p(y_{i,j}|s_{i,j}, \theta, \alpha_j, \xi_j). \tag{4}$$

The density of the sampling distribution on the state process s is

$$p(s_{1:J}|\xi_{1:J}) = \prod_{j=1}^{J} p(s_{0,j}) \prod_{i=1}^{N} p(s_{i,j}|s_{i-1,j}, \xi_j), \tag{5}$$

where the initial distribution $p(s_{0,j})$, for $j = 1, ..., J$, follows a uniform distribution over the states.

The goal is to obtain samples from the posterior density

$$p(\alpha_{1:J}, \xi_{1:J}, \theta, s_{1:J}|y_{1:J}) \propto p(y|\theta, \alpha_{1:J}, s_{1:J})\, p(s_{1:J}|\xi_{1:J})\, p(\alpha_{1:J}|\theta)\, p(\theta)\, p(\xi_{1:J})$$

We extend the particle Metropolis within Gibbs (PMwG) sampler of Gunawan, Hawkins, Tran, Kohn, and Brown (2020) to estimate the Markov switching LBA model. The main requirement of the sampler is to construct an augmented target distribution that includes the discrete states, the model parameters and multiple copies of the random effects and has the required posterior as its marginal.

Let $\{m_j(\alpha_j|\theta, y_j); j = 1, ..., S\}$ be the proposal densities used for approximating the conditional densities $\{p(\alpha_j|\theta, y_j); j = 1, ..., S\}$; see Gunawan et al. (2020) for further details about the proposal densities. Let α_j^r be the rth sample from the proposal density $m_j(\alpha_j|\theta, y_j)$ for subject j. Define $\alpha_{1:J}^{1:R} := \left\{\alpha_1^{1:R}, ..., \alpha_J^{1:R}\right\}$ and $\alpha_j^{1:R} := \left\{\alpha_j^1, ..., \alpha_j^R\right\}$. The joint density of the particles $\alpha_{1:J}^{1:R}$ conditional on θ and $y_{1:J}$ is

$$\psi\left(\alpha_{1:J}^{1:R}|\theta, y\right) = \prod_{j=1}^{J}\prod_{r=1}^{R} m_j\left(\alpha_j^r|\theta, y_j\right). \tag{6}$$

Let $p = (p_1, ..., p_J)$ with each $p_j \in \{1, ..., R\}$, $\alpha_{1:J}^{(-p)} = \left\{\alpha_1^{(-p_1)}, ..., \alpha_J^{(-p_J)}\right\}$, with $\alpha_j^{(-p_j)} = \left(\alpha_j^1, ..., \alpha_j^{p_j-1}, \alpha_j^{p_j+1}, ..., \alpha_j^R\right)$ be the collection of particles excluding the selected particles and $\alpha_{1:J}^{(p)} = (\alpha_1^{p_1}, ..., \alpha_J^{p_J})$ be the vector of selected random effects for all J subjects. The augmented target density is

$$\overline{\pi}_R\left(\theta, \alpha_{1:J}^{1:R}, s_{1:J}, \xi_{1:J}|y\right) = \frac{p(\alpha_{1:J}^p, \theta, s_{1:J}, \xi_{1:J}|y)}{R^S} \frac{\psi\left(\alpha_{1:J}^{1:R}|\theta, y\right)}{\prod_{j=1}^{S} m_j\left(\alpha_j^{p_j}|\theta, y_j\right)}. \tag{7}$$

From this target distribution, we can construct the PMwG sampler that generates samples $\left(\theta, \alpha_{1:J}^{1:R}, s_{1:J}, \xi_{1:J}\right)$ from the posterior distribution $p\left(\alpha_{1:J}^{p}, \theta, s_{1:J}, \xi_{1:J} | y\right)$. See Gunawan et al. (2020) for further details about the properties of the augmented target densities. Algorithm 1 describes the PMwG sampling scheme for the hierarchical Markov switching LBA model. The sampler starts at an initial set of group level parameters $\theta = (\mu, \Sigma)$, individual random effects $\alpha_{1:J}$, the transition matrix $\xi_{1:J}$, and the hidden state process $s_{1:J}$. Step (2) is the conditional MC algorithm given in Algorithm 2 that generates $R-1$ particles while keeping the particles $\alpha_{1:J}^{p}$ fixed. The conditional MC algorithm gives the collection of particles $\alpha_{1:J}^{1:R}$ and the normalised weights $W_{1:J}^{1:R}$. Step (3) samples the index vector $p = (p_1, ..., p_J)$ with probability given by equation (8). We then update the selected particles $\alpha_{1:J}^{(p)} = (\alpha_1^{p_1}, ..., \alpha_J^{p_J})$, and discard the rest of particles. Step (4) samples the group level parameters using Gibbs steps, conditional on the selected particles $\alpha_{1:J}^{(p)}$. Step (5) samples the hidden Markov process s_j, for $j = 1, ..., J$, using Algorithm 3. Step (6) samples the elements of the transition matrix.

Algorithm 1 PMwG algorithm for the Markov switching LBA model

1. Select the initial values for $s_{1:J}$, $\alpha_{1:J}$, $\xi_{1:J}$, θ and set $\alpha_j^1 = \alpha_j^{p_j}$ for $j = 1, ..., J$.

2. Sample $\alpha_{1:J}^{(-p)} \sim \overline{\pi}_R \left(\cdot | p, \alpha_{1:J}^p, \theta, y, s_{1:J}, \xi_{1:J}\right)$ using the conditional Monte Carlo algorithm.

3. Sample the index vector $p = (p_1, ..., p_J)$ with probability

$$\overline{\pi}_R \left(p_1 = l_1, ..., p_J = l_J | \alpha_{1:J}^{1:R}, \theta, y, s_{1:J}, \xi_{1:J}\right) = \prod_{j=1}^{J} W_j^{l_j}, \tag{8}$$

   where W_j is the normalised weight in Equation (11).

4. Sample the group-level parameters:

   (a) $\mu | p, \alpha_{1:J}^p, \xi, y, s_{1:J}, \Sigma$ from $N\left(\mu^*, \Sigma^*\right)$, where $\Sigma^* = \left(J\Sigma^{-1} + I\right)^{-1}$ and $\mu^* = \Sigma^* \left(\Sigma^{-1} \sum_{j=1}^{J} \alpha_j^{p_j}\right)$.

   (b) $\Sigma | p, \alpha_{1:J}^p, \xi, y, s_{1:J}, \mu$ from $IW\left(k, B\right)$, where $k = v_0 + J$ and $B = s_0 + \sum_{j=1}^{J} \left(\alpha_j^{p_j} - \mu\right) \left(\alpha_j^{p_j} - \mu\right)'$

5. Sample the hidden Markov process s_j for $j = 1, ..., J$ using Algorithm 3.

6. Sample the elements of the transition matrix. For $K = 2$, the transition matrix ξ_j depends on $\xi_{1,1,j}$ and $\xi_{2,2,j}$ for $j = 1, ..., J$ subjects.

   (a) Sample $\xi_{1,1,j} | p, \alpha_{1:J}^p, \mu, y, s_{1:J}, \Sigma$ from Beta$\left(e_{11} + N_{11,j}^*, e_{12} + N_{12,j}^*\right)$ for $j = 1, ..., J$, where $N_{lk,j}^*$ counts the number of transitions from state l to state k for subject J.

   (b) Sample $\xi_{2,2,j} | p, \alpha_{1:J}^p, \mu, y, s_{1:J}, \Sigma$ from Beta$\left(e_{21} + N_{21,j}^*, e_{22} + N_{22,j}^*\right)$ for $j = 1, ..., J$. We set $e_{11} = e_{22} = 4$ and $e_{12} = e_{21} = 1$.

7. Repeat steps 2 to 6 for the required number of iterations.

Algorithm 2 Conditional Monte Carlo (Markov switching) Algorithm

1. Fix $\alpha_{1:J}^{1} = \alpha_{1:J}^{p}$
2. For $j = 1, ..., J$
   (a) Sample α_j^r from the proposal density $m_j\left(\alpha_j|\theta, y_j\right)$ for $r = 2, ..., R$.
   (b) Compute the importance weights
   $$w_j^r = \frac{p\left(y_j|\theta, \alpha_j^r, s_j\right) p\left(\alpha_j^r|\theta\right)}{m_j\left(\alpha_j^r|\theta, y_j\right)}, \tag{9}$$
   where
   $$p\left(y_j|\theta, \alpha_j^r, s_j\right) = \prod_{i=1}^{N} p\left(y_{i,j}|s_{i,j}, \alpha_{s_{i,j}}^r\right). \tag{10}$$
   (c) Compute the normalised weights
   $$W_j^r = w_j^r / \sum_{k=1}^{r} w_k^r. \tag{11}$$
   .

Algorithm 3 Sampling the hidden state process s_j

1. Run the filtering algorithm and store the filtered state probability $p\left(s_{i,j}=k|y_j^i,\alpha_j,\theta,\xi_j\right)$ for $k=1,...,K$ and $i=1,...,N$ trials, where $y_j^i=(y_{1,j},...,y_{i,j})$. The following steps are carried out recursively for $i=1,...,N$

   (a) Compute one-step ahead prediction for $s_{i,j}$ by using

   $$\Pr\left(s_{i,j}=l|y_j^{i-1},\alpha_j,\theta,\xi_j\right)=\sum_{k=1}^{K}\xi_{k,l,j}\Pr\left(s_{i-1,j}=k|y_j^{i-1},\alpha_j,\theta,\xi_j\right),\tag{12}$$

   for $l=1,...,K$, where $\xi_{k,l,j}=\Pr\left(s_{i,j}=l|s_{i-1,j}=k,y_j^{i-1},\alpha_j,\theta,\xi_j\right)$.

   (b) Filtering for $s_{i,j}$ by using

   $$\Pr\left(s_{i,j}=l|y_j^i,\theta,\alpha_j,\theta,\xi_j\right)=\frac{p\left(y_{i,j}|s_{i,j}=l,y_j^{i-1},\alpha_j,\theta,\xi_j\right)\Pr\left(s_{i,j}=l|y_j^{i-1},\alpha_j,\theta,\xi_j\right)}{\sum_{k=1}^{K}p\left(y_{i,j}|s_{i,j}=k,y_j^{i-1},\alpha_j,\theta,\xi_j\right)\Pr\left(s_{i,j}=k|y_j^{i-1},\alpha_j,\theta,\xi_j\right)}.\tag{13}$$

2. Sample $s_{N,j}$ from the filtered state probability distribution $\Pr\left(s_{N,j}=k|y_j,\theta,\alpha_j,\theta,\xi_j\right)$.

3. For $i=N-1,N-2,...,0$, sample $s_{i,j}$ from the conditional distribution

$$\Pr\left(s_{i,j}=k|s_{i+1,j}=l,y_j^i,\theta,\alpha_j,\theta,\xi_j\right)=\frac{\xi_{k,l,j}\Pr\left(s_{i,j}=k|y_j^i,\theta,\alpha_j,\theta,\xi_j\right)}{\sum_{r=1}^{K}\xi_{r,l,j}\Pr\left(s_{i,j}=r|y_j^i,\theta,\alpha_j,\theta,\xi_j\right)}.\tag{14}$$

References

Gunawan, D., Hawkins, G. E., Tran, M.-N., Kohn, R., & Brown, S. (2020). New estimation approaches for the hierarchical linear ballistic accumulator model. *Journal of Mathematical Psychology*, *96*, 102368.