RTI Hierarchical Markov Models

Russell G. Almond
Florida State University

December, 2016. Do not cite or quote.

1 Common Data Layout

Let I be the number of students, and T_i be the number of measurements made on Student i. Let $T_{max} = \max_{i \in I} T_i$.

Let $Obs_{t,i}$ be the observation for Student i on the tth measurement occasion. Let $Time_{t,i}$ be the elapsed time between measurement occasion t and $t+1$ for Student i and let $Dose_{t,i}$ be the dosage of treatment received by Student i between times t and $t+1$. In general the dose will be the treatment intensity multiplied by the elapsed time. Note that the indexes are backward from the usual description so these can be described as a one-dimensional array of vectors in Stan.

2 Common Evidence Model

It is assumed that the measurement instruments are all vertically scaled and on the same scale. This eliminates a potential identifiability issue between the growth parameters and the operating characteristics of the instruments.

The students proficiencies are represented by a single latent variable $\theta_{t,i}$ (once again the indexes are backwards so this can be represented as an array of vectors in Stan). The relationship between theta and the observation is given by the following equation:

$$Obs_{t,i} \sim N(obs_{int} + obs_{slope}\theta_{t,i}, res_{std}) \tag{1}$$

The three parameters which control equation 1 are further defined in terms of other parameters. Let obs_{rel} be the reliability of the instrument, $obs_{std,t1}$ be the standard deviation of the scores at the first measurement occasion, and $obs_{mean,t1}$ be the mean of those scores. To identify the latent scale, $\theta_{i,1}$ is assume to have a standard normal distribution. Therefore,

$$obs_{int} = obs_{mean,t1}; \qquad obs_{slope} = obs_{std,t1}\sqrt{obs_{rel}}; \qquad res_{std} = obs_{std,t1}\sqrt{1 - obs_{rel}} \tag{2}$$

This should ensure that the scale at the initial time point is properly identified.

3 Variable Slopes Model 2

This model assumes that students ability grows according to a Wiener process with drift. That is, between each time point there is an independent increment to each student's ability, and those increments accumulate over time. The process is assumed to have drift as the students are often actively receiving instruction, and the average trend will depend on the instruction received.

The average growth (or drift) has two components a natural growth component and a treatment effect. It is assumed that the students are in a RTI-type program where they are divided into two tiers. Students in Tier I receive the normal instruction and only exhibit normal growth. Students in Tier II receive both normal instruction and some kind of supplemental instruction; thus, their growth with have both natural and treatment effects. The variable $Dose_{t,i}$ indicates how much supplemental instruction each student receives between measurement points t and $t+1$. It is zero for students in Tier I and positive for students in Tier II.[1]

Using this decomposition for the average learning gain, the change in the latent proficiency can be decomposed as:

$$\theta_{t+1,i} = \theta_{t,i} + slope_i * Time_{t,i} + treat_{eff} * Dose_{t,i} + \epsilon_{t,i} \quad (3)$$

Note that in this equation, the natural growth rate, $slope_i$, varies by person, but the treatment effect does not. Also, it is assume that the treatment effect and natural growth rate are additive. Finally, to make this a Wiener process, the variance of the innovation term, $\epsilon_{t,i}$ depends on the elapsed time, $Time_{t,i}$; in particular, $\epsilon_{t,i} \sim N(0, \sqrt{var_{innov}Time_{t,i}})$.

? (?) notes that there is often a correlation between the slope and the initial value in growth curves. This is because the first measurement occasion is often not the true time zero. Consider a growth curve for Reading in Kindergarten students. Most students will have received some kind of pre-Reading instruction either through home or pre-school. So even if the first measurement occasion is the first day of class, they still will have received prior instruction. Students who naturally grow at a faster rate are likelier to then be at a higher level when first measured. This is complicated by the fact that time zero may also vary from student to student. For example, entering Kindergarten student vary considerably in the amount of pre-school they may have attended and the number of reading related activities that they do in their home life.

To capture this idea, the slope distribution is characterized with three parameters, $slope_{mu}$, $slope_{std}$ and $slope_{r2}$. The last parameter is the correlation between the $slope_i$ and $\theta_{i,1}$. To capture this relationship, the slopes are made dependent on the initial proficiencies as follows:

$$slope_i = slope_{mu} + slope_{std}(\sqrt{1 - slope_{r2}^2}\phi_i + slope_{r2} * \theta 1, i, \quad (4)$$

where both $\theta 1, i$ and ϕ_i have unit normal distributions.

[1] Tier III can be accommodated by using a higher intensity for the dosage parameter.