Residual covariance derivation from lecture.

Miratrix

October 22, 2016

We have a random slopes model with 5 timepoints:

$$
\begin{aligned}
Y_{ti} &= \pi_{0i} + \pi_{1i} a_t + \tilde{\epsilon}_{ti} \\
\pi_{0i} &= \beta_0 + u_{0i} \\
\pi_{1i} &= \beta_1 + u_{1i} \\
\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ & \tau_{11} \end{pmatrix} \right]
\end{aligned}
$$

Let $\tilde{\epsilon}_i \sim N(0, \sigma^2)$. Let our intercept correspond to our first timepoint, so $a_1 = 0, a_2 = 1, ..., a_5 = 4$. I.e., our a_t is number of years since onset of study. Then β_0 is the average outcome at onset of the study and β_1 is the rate of growth (per year) in the population.

The reduced form is

$$
\begin{aligned}
Y_{ti} &= \beta_0 + \beta_1 a_t + u_{0i} + u_{1i} a_t + \tilde{\epsilon}_{ti} \\
&= \beta_0 + \beta_1 a_t + \epsilon_{ti}
\end{aligned}
$$

with $\epsilon_{ti} = u_{0i} + u_{1i} a_t + \tilde{\epsilon}_{ti}$.

Note that $\tilde{\epsilon}_{ti}$ is the specific time-individual residual after the individual random effects, and ϵ_{ti} is the *overall* residual (deviation from what we expect from the population).

Now we assume for any subject i, the covariance matrix of their residuals is

$$
\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} & \delta_{15} \\ & \delta_{22} & \delta_{23} & \delta_{24} & \delta_{25} \\ & & \delta_{33} & \delta_{34} & \delta_{35} \\ & & & \delta_{44} & \delta_{45} \\ & & & & \delta_{55} \end{pmatrix} \right]
$$

The big matrix is our V_i from lecture. The diagonal are our variances at each timepoint (so in fact, if we took the variance of all the $t = 5$ observations we should get something close to δ_{55}).

Let's calculate $\delta_{13} = cov(\epsilon_{i1}, \epsilon_{i2})$. (In lecture I started δ_{12}.)

First we need a math fact about random quantities A, B, and C:

$$cov(A + B, C) = cov(A, C) + cov(B, C).$$

Also if you multiply something by a constant k you have

$$cov(k_1 A, k_2 B) = k_1 k_2 cov(A, B).$$

Also we need that $a_1 = 0$ and $a_3 = 2$. Then we have:

$$\begin{aligned}
\delta_{13} &= cov(\epsilon_{i1}, \epsilon_{i3}) \\
&= cov(u_{0i} + u_{1i}a_1 + \tilde{\epsilon}_{1i}, u_{0i} + u_{1i}a_3 + \tilde{\epsilon}_{3i}) \\
&= cov(u_{0i} + \tilde{\epsilon}_{1i}, u_{0i} + 2u_{1i} + \tilde{\epsilon}_{3i}) \\
&= cov(u_{0i}, u_{0i}) + cov(u_{0i}, 2u_{1i}) + cov(u_{0i}, \tilde{\epsilon}_{3i}) + cov(\tilde{\epsilon}_{1i}, u_{0i}) + cov(\tilde{\epsilon}_{1i}, 2u_{1i}) + cov(\tilde{\epsilon}_{1i}, \tilde{\epsilon}_{3i}) \\
&= \tau_{00} + 2\tau_{01} + 0 + 0 + 0 + 0 \\
&= \tau_{00} + 2\tau_{01}
\end{aligned}$$

Note how we multiple out the individual components, and this gives an expression for the overall covariance of our two residuals. If we did this for each $\delta_{tt'}$ we could fill in our 5×5 matrix. Fun!

A core idea here is the independence of the different residual pieces makes a lot of the terms go to 0, giving short(er) expressions than we might have otherwise. The random slope model dictates the overall covariance of the residuals.

For the variances, you would just calculate covariance of a quantity with itself. Let's do δ_{11}, the variance of timepoint 1:

$$\begin{aligned}
\delta_{11} &= var(\epsilon_{i1}) = cov(\epsilon_{i1}, \epsilon_{i1}) \\
&= cov(u_{0i} + u_{1i}a_1 + \tilde{\epsilon}_{1i}, u_{0i} + u_{1i}a_1 + \tilde{\epsilon}_{1i}) \\
&= cov(u_{0i} + \tilde{\epsilon}_{1i}, u_{0i} + \tilde{\epsilon}_{1i}) \\
&= cov(u_{0i}, u_{0i}) + cov(u_{0i}, \tilde{\epsilon}_{1i}) + cov(\tilde{\epsilon}_{1i}, u_{0i}) + cov(\tilde{\epsilon}_{1i}, \tilde{\epsilon}_{1i}) \\
&= \tau_{00} + 0 + 0 + \sigma^2
\end{aligned}$$

Now let's do δ_{55}, the variance of timepoint 5:

$$
\begin{aligned}
\delta_{55} &= var(\epsilon_{i5}) = cov(\epsilon_{i5}, \epsilon_{i5}) \\
&= cov(u_{0i} + u_{1i}a_5 + \tilde{\epsilon}_{5i}, u_{0i} + u_{5i}a_5 + \tilde{\epsilon}_{5i}) \\
&= cov(u_{0i} + 4u_{1i} + \tilde{\epsilon}_{5i}, u_{0i} + 4u_{1i} + \tilde{\epsilon}_{5i}) \\
&= cov(u_{0i}, u_{0i}) + cov(u_{0i}, 4u_{1i}) + cov(u_{0i}, \tilde{\epsilon}_{5i}) + \\
&\quad\; cov(4u_{1i}, u_{0i}) + cov(4u_{1i}, 4u_{1i}) + cov(4u_{1i}, \tilde{\epsilon}_{5i}) \\
&\quad\; cov(\tilde{\epsilon}_{1i}, u_{0i}) + cov(\tilde{\epsilon}_{1i}, 4u_{1i}) + cov(\tilde{\epsilon}_{1i}, \tilde{\epsilon}_{1i}) \\
&= \tau_{00} + 4\tau_{01} + 0 + 4\tau_{01} + 16\tau_{11} + 0 + 0 + 0 + \sigma^2 \\
&= \tau_{00} + 16\tau_{11} + 8\tau_{01} + \sigma^2.
\end{aligned}
$$

Note how the variance around the intercept (at time 1 where $a_1 = 0$) is less than far out. As the different slopes diverge, the overall variance increases as we move away from the intercept point. One interesting aspect of random slope models is the marginal (at each time point) heteroskedasticity. This can be offset by a negative τ_{01}, but generally as you move away from the intercept variance goes up. This is even more reason for picking the intercept in a principled way, and for picking it near the center of your data if possible.