Mixed models - assorted topics

Rasmus Waagepetersen

April 19, 2024

- Bayesian analysis for AR(1)
- Cucumber data: use random effects or not ?
- Mixed models and randomized trials: evaluation of new mathematics teaching method in primary school
- Mixed models with correlated errors

Estimation for AR(1)

Likelihood for zero-mean stationary AR(1) ($|a| < 1$):

$$L(a, \tau^2) = \frac{1}{\sqrt{\frac{\tau^2}{(1-a^2)}}} \exp(-\frac{(1-a^2)}{2\tau^2} X_1^2) \prod_{i=2}^{n+1} \frac{1}{\sqrt{\tau^2}} \exp(-\frac{1}{2\tau^2}(X_i - aX_{i-1})^2)$$

Often conditional likelihood given X_1 used instead:

$$\begin{aligned} L(a, \tau^2 | X_1) &= \prod_{i=2}^{n+1} \frac{1}{\sqrt{\tau^2}} \exp(-\frac{1}{2\tau^2}(X_i - aX_{i-1})^2) \\ &\equiv (\tau^2)^{-n/2} \exp(-\frac{1}{2\tau^2} \| Y - \mathbf{X} a \|^2) \end{aligned}$$

with $Y := (X_2, \ldots, X_{n+1})^{\mathsf{T}}$ and $\mathbf{X} := (X_1, \ldots, X_n)^{\mathsf{T}}$.

Then we immediately get (least squares)

$$\hat{a} = \frac{\mathbf{X}^{\mathsf{T}} Y}{\mathbf{X}^{\mathsf{T}} \mathbf{X}} = \frac{\sum_{i=2}^{n+1} X_i X_{i-1}}{\sum_{i=1}^{n} X_i^2} \quad \hat{\tau}^2 = \frac{\| Y - \mathbf{X}\hat{a} \|^2}{n}$$

Estimation easy - but what is distribution of $\hat{a}$?

Recall

$$X_i = aX_{i-1} + \nu_i$$

where ν_i *iid* $N(0, \tau^2)$ (in fact, normality not needed).

Then

$$\hat{a} = \frac{\sum_{i=2}^{n+1} X_i X_{i-1}}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=2}^{n+1} aX_{i-1}X_{i-1}}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=2}^{n+1} \nu_i X_{i-1}}{\sum_{i=1}^{n} X_i^2} = a + \frac{\sum_{i=2}^{n+1} \nu_i X_{i-1}}{\sum_{i=1}^{n} X_i^2}$$

Thus,

$$\sqrt{n}(\hat{a} - a) = \frac{n^{-1/2} \sum_{i=2}^{n+1} \nu_i X_{i-1}}{n^{-1} \sum_{i=1}^{n} X_i^2}$$

The sequence $\nu_i X_{i-1}$, $i = 2, 3, \ldots$ is a so-called martingale difference sequence for which CLT exists:

$$n^{-1/2} \sum_{i=2}^{n+1} \nu_i X_{i-1} \to N(0, \mathbb{V}\mathrm{ar}(\nu_2 X_1))$$

in distribution. Note $\mathbb{V}\mathrm{ar}(\nu_2 X_1)) = \tau^2 \mathbb{V}\mathrm{ar} X_1$

By weak law of large numbers (for weakly correlated sequence)

$$n^{-1} \sum_{i=1}^{n} X_i^2 \to \mathbb{E} X_1^2 = \mathbb{V}\mathrm{ar}(X_1)$$

in probability.

In conclusion,

$$\sqrt{n}(\hat{a} - a) \to N(0, \tau^2 / \mathbb{V}\mathrm{ar}(X_1)) = N(0, 1 - a^2)$$

This result does not rely on normality of ν_is but quite technical (martingale CLT for correlated sequence, law of large numbers, Slutsky's theorem).

Bayesian approach

Assume $X_1 \sim N(\mu_1, 1)$ and AR(1) specification for rest of X_i's.

Use prior $p(a, \tau^2, \mu_1) \propto \frac{1}{\tau^2} p(\mu_1)$. Then posterior is

$$p(a, \tau^2, \mu_1 | X_1, \dots, X_n) \propto L(a, \tau^2 | X_1) \frac{1}{\tau^2} f(X_1 | \mu_1) p(\mu_1)$$

We focus on (a, τ^2) (not interested in μ_1 here):

$$p(a, \tau^2 | X_1, \dots, X_n) \propto L(a, \tau^2 | X_1) \frac{1}{\tau^2} = \frac{1}{\tau^2} (\tau^2)^{-n/2} \exp(-\frac{1}{2\tau^2} \| Y - \mathbf{X} a \|^2)$$

Note, conditional on $(X_1, \dots, X_n)$ this is exactly equivalent to the posterior for a normal linear model with data vector Y and design matrix $\mathbf{X}$. Hence by our Bayesian derivation (Bayes 1, slide 14) for linear normal model, we immediately get that

$$a | X_1, \dots, X_n, \tau^2 \sim N\left(\hat{a}, \tau^2 (\mathbf{X}^\mathsf{T} \mathbf{X})^{-1}\right) = N\left(\hat{a}, \frac{\tau^2}{\sum_{i=1}^{n-1} X_i^2}\right)$$

Note similarity with frequentist result ! However, no need of CLT or stationarity. On the other hand, normality is needed.

Cucumbers and random effects

For the cucumbers data example we initially include random plot, section and block effects to account for variations in soil, temperature, light etc. across greenhouse.

However, not much evidence for positive variances for these random effects ?

Should we remove random effects ? Pros and cons:

Cucumbers and random effects

For the cucumbers data example we initially include random plot, section and block effects to account for variations in soil, temperature, light etc. across greenhouse.

However, not much evidence for positive variances for these random effects ?

Should we remove random effects ? Pros and cons:

- if there is indeed random variation associated with plots, sections or blocks (other than noise) then we get invalid F-tests if just using ordinary linear model.
- if we keep random effects but there is actually no random variation for plots, sections and blocks we use an overly complex model and may loose power when investigating fixed effects of climate, variety and fertilizer

Power calculations (simulation study)

Suppose we want to assess effects of fixed effects climate, variety and fertilizer.

Suppose variance components for plot, section and block are all zero but we still include plot, section and block as random effects along with fixed main effects (not zero) of climate, fertilizer, and variety.

Power (probability of rejecting null hypothesis of no main effects) at the 5% significance level:

Model	Climate	Fert	Variety
Mixed :	0.84	0.68	0.82
Linear (only fixed):	0.39	0.60	0.80

Considerably lower power for Climate with mixed model. F-test uses denominator $\tilde{\lambda}_{BxC}$ which incorporates $\sigma^2, \sigma^2_{BxCxF}, \sigma^2_{BxC}$ and has fewer denominater degrees of freedom than for linear model. Variety is in "noise" stratum so not much difference between mixed and linear model ($\lambda_I = \sigma^2$).

Type I error (simulation study)

Suppose instead that there is indeed random variation for plot, section and block but no fixed effects of climate, fertilizer, and variety.

Type I error (probability of rejecting at 5% level):

Model	Climate	Fert	Variety
Mixed :	0.05	0.05	0.05
Linear (only fixed):	0.28	0.08	0.004

Correct significance level for mixed model (as guaranteed by theory). No control of type I error rate for wrong linear model without random effects.

Type I error (simulation study)

Suppose instead that there is indeed random variation for plot, section and block but no fixed effects of climate, fertilizer, and variety.

Type I error (probability of rejecting at 5% level):

Model	Climate	Fert	Variety
Mixed :	0.05	0.05	0.05
Linear (only fixed):	0.28	0.08	0.004

Correct significance level for mixed model (as guaranteed by theory). No control of type I error rate for wrong linear model without random effects.

Too large residual variance estimate for linear model without random effects.

Block random or fixed ?

Block only has three levels. Not much room for estimating the associated variance.

However, for testing fixed effects of climate, fertilizer and variety it does not matter whether block is included as random or fixed.

Cucumber data

For cucumber data p-values and conclusions for main effects are similar for models with and without random effects.

In practice choice of random effects should be guided by knowledge about the specific experiment conducted (I am not fan of deciding by testing hypotheses - this leads to issues with multiple testing).

TRACK (Teaching Routines and Content Knowledge) project

123 Danish primary schools randomized into treatment and control group.

Treatment: new mathematics teaching method inspired by Singaporean practice.

Follow pupils over three years starting with 4th grade.

Data available now: mathematics test result at beginning of study (baseline).

My colleagues from Aarhus wanted to test adequate randomization by assessing treatment effect at baseline.

Is this a good idea ?

OLS

They used ordinary least squares (OLS) but with adjusted standard errors taking into account correlation between schools and classes.

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y \quad \mathbb{V}\mathrm{ar}\hat{\beta} = (X^\top X)^{-1} X^\top V X (X^\top X)^{-1}$$

Plug-in empirical estimate of V.

Note: this is not BLUE !

They found slightly significant negative treatment effect !! (p slightly less than 5%)

Conclusion ??

Mixed models analysis

I used linear mixed models.

Mixed models analysis

I used linear mixed models.

School and class random effects.

In this case

$$\hat{\beta} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} Y \quad \mathbb{V}\mathrm{ar}\hat{\beta} = (X^\top V^{-1} X)^{-1}$$

Is this BLUE ?

Possible disadvantage of mixed model analysis ?

Results regarding fraction (brøk-regning) test results

Estimates of treatment effects and p-values:

OLS	OLS p-value	permutation p-value
-2.39	0	0.014
Mixed (BLUE)	approx t p-value	permutation
-1.34	0.151	0.145

Permutation: randomly permute schools into treatment and control and assess treatment effect for each permuted data set. Under null-hypothesis, data should come from this permutation distribution.