Mixed models and Bayesian statistics - various topics

Rasmus Waagepetersen

May 1, 2026

- Bayesian analysis time series (AR(1))
- Cucumber data: use random effects or not ?
- Mixed models with correlated errors

Estimation for AR(1)

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

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

Often conditional likelihood given X_1 used instead:

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

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 $\nu_i X_{i-1}$, $i = 2, 3, \dots$ 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)

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

in probability.

In conclusion,

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

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

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

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

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

Note, conditional on $(X_1, \ldots, X_n)$ 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

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

Comparison with frequentist result

From previous slide it follows that for any n,

$$\sqrt{n}(a - \hat{a}_n) \sim N\left(0, \frac{\tau^2}{\frac{1}{n}\sum_{i=1}^{n-1} X_i^2}\right),$$

where $\frac{1}{n}\sum_{i=1}^{n-1} X_i^2$ converges to $\mathbb{V}\mathrm{ar} X_1 = \tau^2/(1-a^2)$ almost surely by SLLN for dependent observations.

By Scheffe's theorem, the sequence of normal distributions $N(0, \tau^2/(\frac{1}{n}\sum_{i=1}^{n-1} X_i^2))$ converges to $N(0, 1-a^2)$ in total variation norm (and thus in distribution).

Bayesian approach simpler than frequentist but relies on assumption of normality.

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 :	0.39	0.60	0.80
Linear (only fixed):	0.84	0.68	0.82

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).

Example from Department of Health Technology

25 subjects where exposed to electric pulses of 11 different durations using two different electrodes (pin or patch).

Dependent variable: electric perception

The durations were applied in random order.

In total 550 measurements of response to pulse exposure.

Fixed effects in the model: electrode, Pulseform (duration), order of 22 measurements for each subject.

Order: to take into account habituation effect

Random effects: one random effect for each subject-electrode combination (50 random effects).

Mixed model with random intercepts

Model:

$$y_{ijk} = \mu_{ij} + U_{ij} + \epsilon_{ijk}$$

where $i = 1, \ldots, 25$ (subject), $j =$ pin,patch (electrode), and $k = 1, \ldots, 11$ measurement within subject-electrode combination.

U_{ij}'s and ϵ_{ijk}'s independent random variables.

μ_{ij} fixed effect part of the model depending on electrode, Pulseform and order of measurement. In `R`-code:

```
y~electrode*Pulseform+electrode*Order
```

Note `electrode`, `Pulseform` categorical, `Order` nominal (numerical)

Using lmer

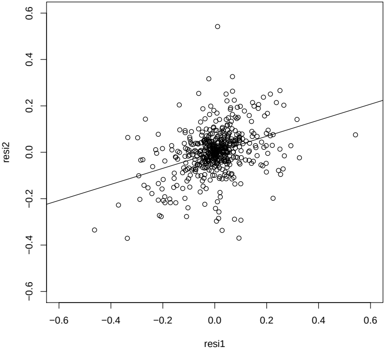
```
fit=lmer(transfPT~electrode*Pulseform+
     electrode*Order+(1|electrsubId),data=perception)

Random effects:
 Groups      Name        Variance Std.Dev.
 electrsubId (Intercept) 0.03479  0.1865
 Residual                0.01317  0.1148
Number of obs: 550, groups:  electrsubId, 50
```

Large subject-electrode variance 0.03479. Noise variance 0.01317

Serial correlation in measurement error ?

Maybe error ϵ_{ijk} not independent of previous error $\epsilon_{ij(k-1)}$ since measurements carried out in a sequence for each subject ?

For each subject-electrode combination ij plot residual r_{ijk} (`resid(fit)`) against previous residual $r_{ij(k-1)}$ for $k = 2, \ldots, 11$.



Correlation

```
cor.test(resi1,resi2)

 Pearson's product-moment correlation

data:  resi1 and resi2
t = 8.5284, df = 498, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2779966 0.4311777
sample estimates:
      cor
0.3569848
```

Mixed model with correlated errors

Analysis of residuals r_{ijk} (which are estimates of errors ϵ_{ijk}) suggests that ϵ_{ijk} are correlated (not independent).

Recall general mixed model formulation:

$$Y = X\beta + ZU + \epsilon$$

where ϵ normal with mean zero and covariance Σ.

So far $\Sigma = \sigma^2 I$ meaning noise terms uncorrelated and all with same variance σ^2

Extension: Σ not diagonal meaning $\mathbb{C}\mathrm{ov}[\epsilon_i, \epsilon_{i'}] \neq 0$.

Many possibilities for Σ - we will focus on autoregressive covariance structure that is useful for serially correlated error terms.

Implementation

Not possible in `lmer` :(

However `lme` (from package `nlme`) can do the trick:

```
fit=lme(transfPT~electrode*Pulseform+electrode*Order,
            random=~1|electrsubId,correlation=corAR1())
```

`lme` predecessor of `lmer` - both have pros and cons - but here `lme` has the upper hand.

Estimates of variance parameters

With uncorrelated errors: $\tau^2 = 0.035$ $\sigma^2 = 0.013$ BIC -511

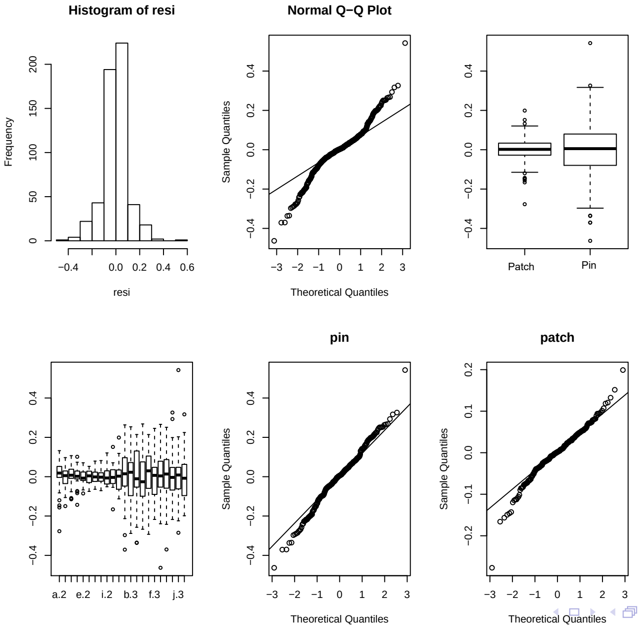
With autoregressive errors: $\tau^2 = 0.030$ $\sigma^2 = 0.018$ BIC -646 $\rho = 0.626$

Variance parameters not so different but quite big estimated correlation for errors. BIC clearly favors model with autoregressive errors.

Quite similar (with/without autoregressive errors) fixed effects estimates.

No clear pattern regarding sizes of standard errors of parameter estimates.

Model assessment - residuals



Much larger residual variance for pin electrode than for patch electrode.

Still room for improvement of model !

Can add `weights=varIdent(form=~1|factor(electrode)` in `lme`. Seems not available in SPSS or `lmer`.

Separate analyses for two electrodes ?

Software summary:

	F-test	correlated residuals	variance heterogeneity
lmer	yes	no	no
lme	no	yes	yes
SPSS	yes	yes	no

Exercises

1. Fit model with autoregressive errors to the Orthodont dataset.
2. Given the model (1) verify that $\epsilon_{ij(k+1)} \sim N(0, \sigma^2)$ and $\mathbb{C}\mathrm{orr}(\epsilon_{ijk}, \epsilon_{ij(k+1)}) = \rho$.

Basic model for serial correlation: autoregressive

Consider sequence of noise terms: $\epsilon_{ij1}, \epsilon_{ij2}, \ldots, \epsilon_{ij11}$.

Model for variance/covariance:

$$\mathbb{C}\mathrm{ov}(\epsilon_{ijk}, \epsilon_{ijk'}) = \sigma^2 \rho^{|k-k'|} \quad \mathbb{C}\mathrm{orr}(\epsilon_{ijk}, \epsilon_{ijk'}) = \rho^{|k-k'|} \quad |\rho| < 1$$

Thus

$$\mathbb{V}\mathrm{ar}\epsilon_i = \sigma^2$$

and ρ is correlation between two consecutive noise terms,

$$\rho = \mathbb{C}\mathrm{orr}(\epsilon_{ijk}, \epsilon_{ij(k+1)})$$

Interpretation of autoregressive model

Covariance structure arises from following autoregressive model:

$$\epsilon_{ij(k+1)} = \rho\epsilon_{ijk} + \nu_{ij(k+1)} \tag{1}$$

where $\epsilon_{ij1} \sim N(0, \sigma^2)$, and

$$\nu_{ijl} \sim N(0, \omega) \quad \omega = \sigma^2(1 - \rho^2) \quad l = 2, \ldots, 11$$

$\epsilon_{ij1}, \nu_{ij2}, \ldots, \nu_{ij11}$ assumed to be independent.

Yet another example of building correlation using independent building blocks !

Model with different variances for pin and patch electrodes

Much larger residual variance for pin electrode than for patch electrode.

Fit model with variance heterogeneity:

```
fithetcorr=lme(transfPT~electrode*Pulseform+electrode*Order
random=~1|electrsubId,data=perception,
weights=varIdent(form=~1|electrode),correlation=corAR1())
```

```
Random effects:
 Formula: ~1 | electrsubId #correlation within electrode-su
        (Intercept)  Residual
StdDev:   0.1732323 0.1691177

Correlation Structure: AR(1)
 Formula: ~1 | electrsubId

Parameter estimate(s):
     Phi
```

```
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | electrode
 Parameter estimates:
      pin     patch
1.0000000 0.4122666
```

BIC -814

Subject variance $0.1732^2 = 0.029$
Variance for pin electrode: $0.1691^2 = 0.028$
Variance for patch electrode: $0.1691^2 \cdot 0.4122^2 = 0.0049$