

Homework 4 - BIOS 6643 Analysis of Longitudinal Data

Dominic Adducci

2023-09-20

Question 1

Complex MLE estimation in LMM requires computational optimization approaches, the goal here is to implement a basic Newton-Raphson algorithm in R. The following data are an i.i.d sample from a $\text{Cauchy}(\theta, 1)$ distribution: 1.77, -0.23, 2.76, 3.80, 3.47, 56.75, -1.34, 4.24, -2.44, 3.29, 3.71, -2.40, 4.53, -0.07, -1.05, -13.87, -2.53, -1.75, 0.27, 43.21.

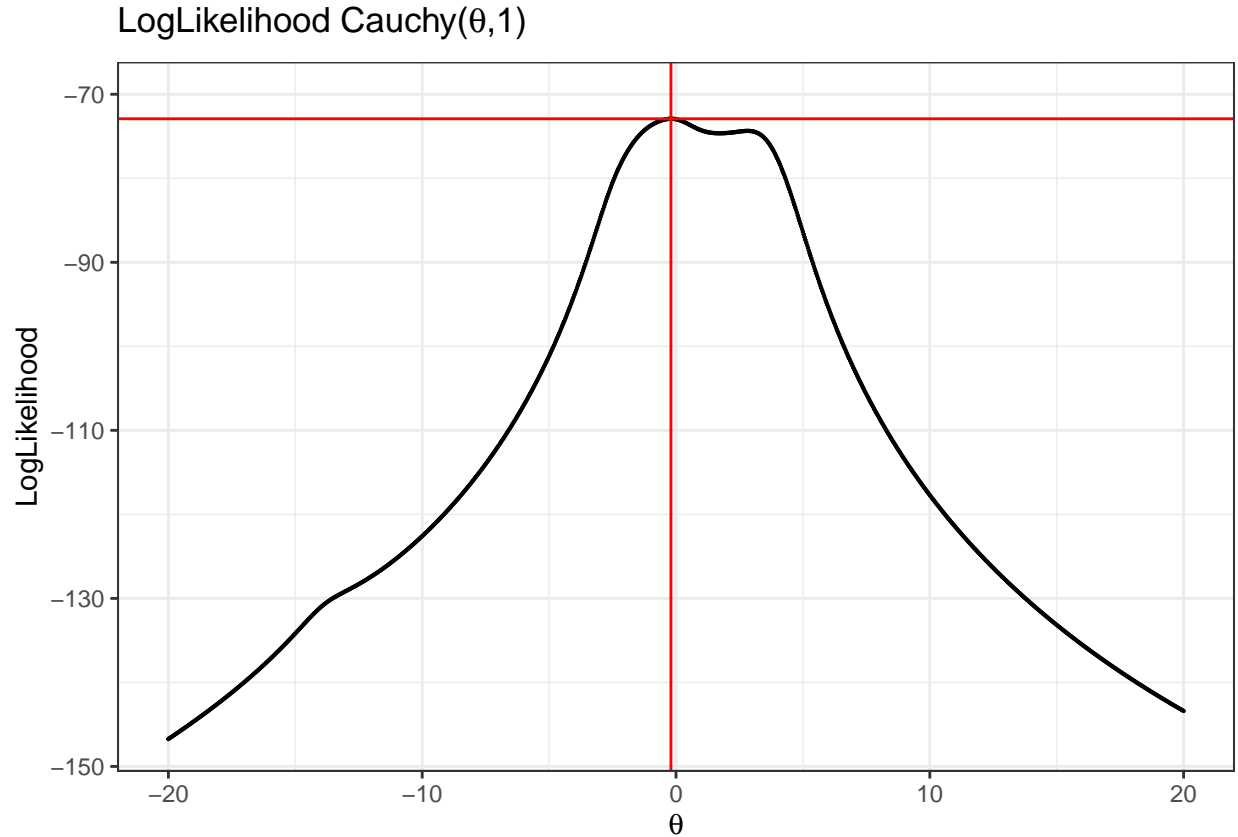
Part A: Graph the log likelihood function.

The likelihood function of the $\text{Cauchy}(\theta, 1)$ distribution is:

$$\frac{1}{\pi^n} \frac{1}{\prod [1 + (x_i - \theta)^2]}$$

and the loglikelihood is:

$$-n \log(\pi) - \sum_i^n \log(1 + (x_i - \theta)^2)$$



From the plot of log-likelihood the maximum log-likelihood values is -72.92, and the optimal theta is -0.2.

Part B

Find (and write an R program) to find the MLE for θ using the Newton-Raphson method.

The equation for the Newton-Raphson method in general form is as follows:

$$x_i = x_{i-1} - \frac{f(x_{i-1})}{f'(x_{i-1})}$$

Translating this to finding the MLE of θ :

$$\hat{\theta}_i = \hat{\theta}_{i-1} - \frac{\log L'(\hat{\theta}_{i-1})}{\log L''(\hat{\theta}_{i-1})} = \hat{\theta}_{i-1} - \frac{S(\hat{\theta}_{i-1})}{I(\hat{\theta}_{i-1})}$$

What finding an MLE the score $S(\theta)$, the derivative of the log-likelihood, should equal 0, $S(\theta) = 0$. $I(\theta)$ is the Fisher observed information from the data.

Check the code appendix for the function.

```
##      Estimate Difference
## 1  2.938901  -1.061099
## 2  2.834520  -1.165480
## 3  2.817895  -1.182105
## 4  2.817472  -1.182528
## 5  2.817472  -1.182528
```

```
## 6 2.817472 -1.182528
## 7 2.817472 -1.182528
## 8 2.817472 -1.182528
## 9 2.817472 -1.182528
## 10 2.817472 -1.182528
## 11 2.817472 -1.182528
## 12 2.817472 -1.182528
## 13 2.817472 -1.182528
## 14 2.817472 -1.182528
## 15 2.817472 -1.182528
## 16 2.817472 -1.182528
## 17 2.817472 -1.182528
## 18 2.817472 -1.182528
## 19 2.817472 -1.182528
## 20 2.817472 -1.182528
```

Part C Try all of the following starting points: -11,-1,0,1.5,4,4.7,7,8,and 38.

Question 2

In a paragraph explain the difference between a general linear model or multiple regression (GLM; not a generalized linear model like a logistic regression or (GLM; not a generalized linear model like a logistic regression or Poisson, which will be discussed later) and a linear mixed model (LMM).

A general linear model only has fixed effects, while a linear mixed model includes both fixed and random effects. The random effects of the linear mixed model allows the model to account for differences between subjects. A simple example would be measuring something like cholesterol levels through time. A general linear model (fixed effects only) may include covariates such as time, BMI, smoking status, sex, race, etc. Every individual will follow the same regression line based on those covariates. In a linear mixed model random effect which account for subject differences, such as someone tending to have higher or lower cholesterol at start (random intercept) can be included to better model change over time. A better fit may be found by including random slopes as well for each subject if cholesterol trajectory is found to be different between subjects.

Question 3

In a short paragraph, explain the difference between a profiled likelihood and a restricted likelihood for a linear mixed model, and how and why they are used.

In a profile likelihood you maximize the likelihood by fixing every other parameter and only allowing one to vary. Doing this maximizes the single parameter you allowed to vary. You can then repeat this process for every other parameter incrementally, plugging in the estimates for parameters which have already been maximized. The downside to this method is that variance estimates are biased downward. The restricted likelihood (REML) allows for estimating parameters which are not biased regardless of sample size. The downside for the REML method compared to profile likelihood is that you can only compare REML model using a likelihood ratio of both models have the same set of fixed effects.

Profile likelihood - maximize likelihood by fixing every other parameter and only allow one to vary and maximize that using a grid search. After getting that maximum repeat each process for each parameter in the set of likelihood parameter.

Restricted maximum likelihood - REML has property that your standard error are unbiased regardless of sample size, so generally is in situations where were trying to get unbiased estimates of errors when sample

sizes are small. Loglikelihood are not valid for reduced in full REML models. Can only compare REML models that have the same set of fixed effects.

Part 2ish

Investigator wants to understand whether Cortisol (a stress hormone) secretion differs in women suffering from depression. Cortisol was measured every 10 minutes for a period of 24 hours starting at 9 am. 26 patients and 26 controls were collected in the study. Although the data were collected every 10 minutes for a period of 24 hours on each subject (144 observations), the investigators were interested in differences in the circadian pattern between the groups. Data was divided into 6 blocks of 4 hours and averaged to obtain a set of “block means”.

Question 4

Part A Fit a multiple linear regression to investigate how mean cortisol values change over the day (categorical time) and how the average cortisol levels differ by group (no interaction for this model). This will be used to anchor the comparisons later in the assignment.

Table 1: MLR Cortisol

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf.High | P-Value |
|--------------|----------|-----------|--------------|---------------|------------|
| (Intercept) | 3.0501 | 0.5088 | 2.0489 | 4.0513 | 5.7451e-09 |
| timeTime2 | 3.9697 | 0.6662 | 2.6589 | 5.2806 | 6.9911e-09 |
| timeTime3 | 10.0791 | 0.6662 | 8.7682 | 11.3899 | < 2.22e-16 |
| timeTime4 | 6.1979 | 0.6662 | 4.8870 | 7.5087 | < 2.22e-16 |
| timeTime5 | 4.4042 | 0.6662 | 3.0934 | 5.7151 | 1.7017e-10 |
| timeTime6 | 1.7518 | 0.6662 | 0.4410 | 3.0627 | 0.0089779 |
| casecontrolp | 0.9211 | 0.3846 | 0.1643 | 1.6779 | 0.0172285 |

Table 1 shows the output of the multiple linear regression where mean cortisol is the outcome and time and casecontrol are covariates. Time is factored into 6 different levels, where time 1 is the reference.

Part B Provide a table of mean differences from the 6th time period along with SE’s of the differences. Interpret two of the coefficients. You do not need to conduct inference.

Table 2: MLR Cortisol - Time 6 Reference

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf.High | P-Value |
|--------------|----------|-----------|--------------|---------------|------------|
| (Intercept) | 4.8020 | 0.5088 | 3.8008 | 5.8031 | < 2.22e-16 |
| timeTime1 | -1.7518 | 0.6662 | -3.0627 | -0.4410 | 0.00897790 |
| timeTime2 | 2.2179 | 0.6662 | 0.9071 | 3.5287 | 0.00097714 |
| timeTime3 | 8.3273 | 0.6662 | 7.0164 | 9.6381 | < 2.22e-16 |
| timeTime4 | 4.4461 | 0.6662 | 3.1352 | 5.7569 | 1.1726e-10 |
| timeTime5 | 2.6524 | 0.6662 | 1.3415 | 3.9632 | 8.5630e-05 |
| casecontrolp | 0.9211 | 0.3846 | 0.1643 | 1.6779 | 0.01722849 |

Table 2 shows the output of the multiple linear regression where time 6 is the reference group. For term timeTime1 this is the mean difference between cortisol at time 1 and cortisol at time 6. Interpreting this it means cortisol is 1.7518 units higher at time 6 compared to time 1. For the term casecontrolp this is the mean difference between the control (reference) group and women with depression. Interpreting this it means cortisol is 0.9211 units higher for women with depression compared to women without depression (control).

Part C Will these standard error be too big or small and why?

Standard error is found with the equation:

$$Std.Error = \frac{\sigma}{\sqrt{n}}$$

For a study at a single instance n is generally the number of subjects. For this study because subjects each have 6 time points (assuming balanced data) n refers to all instances where cortisol was measured for each subject, 52×6 , which mean $n = 312$.

Table 3: MLR Cortisol - Time 6 Ref. Variances

| Intercept | Time 1 | Time 2 | Time 3 | Time 4 | Time 5 | Group P |
|-----------|----------|----------|----------|----------|----------|-----------|
| 1.294317 | 2.218829 | 2.218829 | 2.218829 | 2.218829 | 2.218829 | 0.7396097 |

Table 3 shows the variances for each estimate. Given that each estimate is relatively low compared to the variances the standard errors (which were used to calculate variance) may be too large to make a meaningful interpretation, as subjects may have significantly different mean cortisol measurements between each other.

Question 5

Part A Fit a linear mixed model assuming categorical time and group with an unstructured variance-covariance structure using method=REML (the default). Print out the correlation matrix and interpret the general patterns of the correlation in the errors for an individual.

Table 4: LMM Unstructured Var-Cov

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 3.1311 | 0.6060 | 1.9433 | 4.3190 | 4.3174e-07 |
| timeTime2 | 3.9697 | 0.5074 | 2.9753 | 4.9641 | 8.4296e-14 |
| timeTime3 | 10.0791 | 0.5244 | 9.0514 | 11.1068 | < 2.22e-16 |
| timeTime4 | 6.1979 | 0.5849 | 5.0514 | 7.3444 | < 2.22e-16 |
| timeTime5 | 4.4042 | 0.6099 | 3.2089 | 5.5995 | 4.1222e-12 |
| timeTime6 | 1.7518 | 0.3213 | 1.1221 | 2.3815 | 1.0258e-07 |
| casecontrolp | 0.7591 | 0.7254 | -0.6627 | 2.1808 | 0.2962 |

Table 4 shows the fixed effects for the linear mixed effects model with unstructured variance-covariance.

```
## Correlation structure of class corSymm representing
## Correlation:
```

```
##      1      2      3      4      5
## 2 0.454
## 3 0.417 0.528
## 4 0.274 0.455 0.555
## 5 0.211 0.445 0.557 0.864
## 6 0.781 0.321 0.515 0.521 0.525
```

The above output shows the correlation structure for the linear mixed model with unstructured variance-covariance.

Part B Provide a table of mean differences from the 6th time period along with SE's of the differences. Interpret two of the coefficients. You do not need to conduct inference.

Table 5: LMM Unstructured Var-Cov - Time 6 Ref.

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 4.8830 | 0.6060 | 3.6951 | 6.0708 | 1.7812e-14 |
| timeTime1 | -1.7518 | 0.3213 | -2.3815 | -1.1221 | 1.0258e-07 |
| timeTime2 | 2.2179 | 0.5658 | 1.1089 | 3.3269 | 0.00010954 |
| timeTime3 | 8.3273 | 0.4780 | 7.3903 | 9.2642 | < 2.22e-16 |
| timeTime4 | 4.4461 | 0.4752 | 3.5146 | 5.3775 | < 2.22e-16 |
| timeTime5 | 2.6524 | 0.4730 | 1.7253 | 3.5795 | 4.6086e-08 |
| casecontrolp | 0.7591 | 0.7254 | -0.6627 | 2.1808 | 0.29619701 |

Table 5 shows the output of the linear mixed model with unstructured variance-covariance where time 6 is the reference group.

- For term timeTime1 this is the mean difference between cortisol at time 1 and cortisol at time 6. Interpreting this it means cortisol is 1.7518 units higher at time 6 compared to time 1. This estimate is identical to that of the MLR model from question 4, but the standard error is smaller at 0.3213 compared to the MLR model's standard error of 0.6662. In other words the estimate of the LMM with unstructured variance-covariance is more precise.
- For term timeTime2 this is the mean difference between cortisol at time 2 and cortisol at time 6. Interpreting this is means cortisol is 2.2179 units higher at time 2 compared to time 6. This estimate is identical to that of the MLR model from question 4, but the standard error is smaller at 0.5658 compared to the MLR model's standard error of 0.6662. In other words the estimate of the LMM with unstructured variance-covariance is more precise.

Part C What is the estimate difference in mean cortisol levels (and SE) between the depressed and non-depressed groups. Interpret the finding in a sentence. You do not need to conduct inference.

From table 4 (or 5, they are equivalent for the casecontrol variable) the estimate difference between the control group and the depressed group is 0.7591. This means that cortisol is 0.7591 units higher in the depressed group. The estimate is smaller than the estimate from the MLR model, but has a higher standard error of 0.7254 compared to the MLR model's standard error of 0.3846.

Question 6

Part A Fit a linear mixed model assuming categorical time and group with a compound symmetry structure using method=REML (the default). Print out the correlation matrix and interpret the general patterns of the correlation in the errors for an individual.

Table 6: LMM Compound Symmetry Var-Cov

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 3.0501 | 0.5938 | 1.8864 | 4.2139 | 4.9916e-07 |
| timeTime2 | 3.9697 | 0.4770 | 3.0348 | 4.9047 | 2.9358e-15 |
| timeTime3 | 10.0791 | 0.4770 | 9.1441 | 11.0140 | < 2.22e-16 |
| timeTime4 | 6.1979 | 0.4770 | 5.2629 | 7.1328 | < 2.22e-16 |
| timeTime5 | 4.4042 | 0.4770 | 3.4693 | 5.3392 | < 2.22e-16 |
| timeTime6 | 1.7518 | 0.4770 | 0.8169 | 2.6868 | 0.00028364 |
| casecontrolp | 0.9211 | 0.7180 | -0.4861 | 2.3283 | 0.20050032 |

Table 6 shows the fixed effects for the linear mixed effects model with compound symmetry variance-covariance structure.

```
## Correlation structure of class corCompSymm representing
##      Rho
## 0.491355
```

The above output shows the correlation structure for the linear mixed model with compound symmetry variance-covariance structure.

Part B Provide a table of mean differences from the 6th time period along with SE's of the differences. Interpret two of the coefficients. You do not need to conduct inference.

Table 7: LMM Compound Symmetry Var-Cov - Time 6 Ref.

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 4.8020 | 0.5938 | 3.6382 | 5.9657 | 1.4528e-14 |
| timeTime1 | -1.7518 | 0.4770 | -2.6868 | -0.8169 | 0.00028364 |
| timeTime2 | 2.2179 | 0.4770 | 1.2829 | 3.1529 | 4.9628e-06 |
| timeTime3 | 8.3273 | 0.4770 | 7.3923 | 9.2622 | < 2.22e-16 |
| timeTime4 | 4.4461 | 0.4770 | 3.5111 | 5.3810 | < 2.22e-16 |
| timeTime5 | 2.6524 | 0.4770 | 1.7174 | 3.5873 | 5.8875e-08 |
| casecontrolp | 0.9211 | 0.7180 | -0.4861 | 2.3283 | 0.20050032 |

Table 7 shows the output of the linear mixed model with compound symmetry variance-covariance where time 6 is the reference group. Because the variance-covariance structure is compound symmetric each SE for time points are the same.

- For term timeTime1 this is the mean difference between cortisol at time 1 and cortisol at time 6. Interpreting this it means cortisol is 1.7518 units higher at time 6 compared to time 1. This estimate is identical to that of the MLR model from question 4, but the standard error is smaller at 0.4770

compared to the MLR model's standard error of 0.6662. In other words the estimate of the LMM with compound symmetry variance-covariance is more precise.

- For term timeTime2 this is the mean difference between cortisol at time 2 and cortisol at time 6. Interpreting this it means cortisol is 2.2179 units higher at time 2 compared to time 6. This estimate is identical to that of the MLR model from question 4, but the standard error is smaller at 0.4770 compared to the MLR model's standard error of 0.6662. In other words the estimate of the LMM with compound symmetry variance-covariance is more precise.

Part C What is the estimated difference in mean cortisol levels (and SE) between the depressed and non-depressed groups. Interpret the finding in a sentence. You do not need to conduct inference.

From table 4 (or 5, they are equivalent for the casecontrol variable) the estimate difference between the control group and the depressed group is 0.9211. This means that cortisol is 0.9211 units higher in the depressed group. This estimate is identical to the MLR model, but has a higher standard error of 0.7180 compared to the MLR model's standard error of 0.3846.

Question 7

Part A Fit a linear mixed model assuming categorical time and group with an AR(1) structure using method=REML (the default). Print out the correlation matrix and interpret the general patterns of the correlation in the errors for an individual.

Table 8: LMM AR(1) Var-Cov

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 3.1384 | 0.5646 | 2.0318 | 4.2449 | 5.9331e-08 |
| timeTime2 | 3.9697 | 0.4484 | 3.0909 | 4.8486 | < 2.22e-16 |
| timeTime3 | 10.0791 | 0.5591 | 8.9832 | 11.1750 | < 2.22e-16 |
| timeTime4 | 6.1979 | 0.6120 | 4.9984 | 7.3974 | < 2.22e-16 |
| timeTime5 | 4.4042 | 0.6395 | 3.1509 | 5.6575 | 3.2512e-11 |
| timeTime6 | 1.7518 | 0.6542 | 0.4696 | 3.0340 | 0.0078104 |
| casecontrolp | 0.7446 | 0.6096 | -0.4502 | 1.9393 | 0.2228629 |

Table 8 shows the fixed effects for the linear mixed effects model with AR(1) variance-covariance.

```
## Correlation structure of class corCompSymm representing
##      Rho
## 0.491355
```

The above output shows the correlation structure for the linear mixed model with AR(1) variance-covariance.

Part B Provide a table of mean differences from the 6th time period along with SE's of the differences. Interpret two of the coefficients. You do not need to conduct inference.

Table 9: LMM AR(1) Var-Cov - Time 6 Ref.

| Term | Estimate | Std.Error | 95% Conf.Low | 95% Conf. High | P-Value |
|--------------|----------|-----------|--------------|----------------|------------|
| (Intercept) | 4.8902 | 0.5646 | 3.7837 | 5.9968 | 2.7585e-16 |
| timeTime1 | -1.7518 | 0.6542 | -3.0340 | -0.4696 | 0.00781044 |
| timeTime2 | 2.2179 | 0.6395 | 0.9646 | 3.4712 | 0.00059886 |
| timeTime3 | 8.3273 | 0.6120 | 7.1278 | 9.5268 | < 2.22e-16 |
| timeTime4 | 4.4461 | 0.5591 | 3.3502 | 5.5420 | 3.6193e-14 |
| timeTime5 | 2.6524 | 0.4484 | 1.7735 | 3.5313 | 8.8967e-09 |
| casecontrolp | 0.7446 | 0.6096 | -0.4502 | 1.9393 | 0.22286293 |

Table 9 shows the output of the linear mixed model with AR(1) variance-covariance where time 6 is the reference group.

- For term timeTime1 this is the mean difference between cortisol at time 1 and cortisol at time 6. Interpreting this it means cortisol is 1.7518 units higher at time 6 compared to time 1. This estimate is identical to that of the MLR model from question 4, but the standard error is slightly smaller at 0.6542 compared to the MLR model's standard error of 0.6662. In other words the estimate of the LMM with AR(1) variance-covariance is more precise.
- For term timeTime2 this is the mean difference between cortisol at time 2 and cortisol at time 6. Interpreting this it means cortisol is 2.2179 units higher at time 2 compared to time 6. This estimate is identical to that of the MLR model from question 4, but the standard error is slightly smaller at 0.6395 compared to the MLR model's standard error of 0.6662. In other words the estimate of LMM with AR(1) variance-covariance is slightly more precise.

Part C What is the estimated difference in mean cortisol levels (and SE) between the depressed and non-depressed groups. Interpret the finding in a sentence. You do not need to conduct inference.

From table 4 (or 5, they are equivalent for the casecontrol variable) the estimate difference between the control group and the depressed group is 0.7446. This means that cortisol is 0.7446 units higher in the depressed group. This estimate is smaller compared to the MLR model, but has a higher standard error of 0.6096 compared to the MLR model's standard error of 0.3846.

```
# Extracting covariance matrix from gls
#lmm <- gls()
#lmm_summ <- summary(lmm)
#varcov <- lmm_summ$modelStruct$corStruct
```