

# 6643 HW4 Key

Nichole Carlson

(25 points) Since the complex MLE estimation in LMM requires computational optimization approaches, you will implement one of the classical approaches here to investigate how it works. The goal 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.

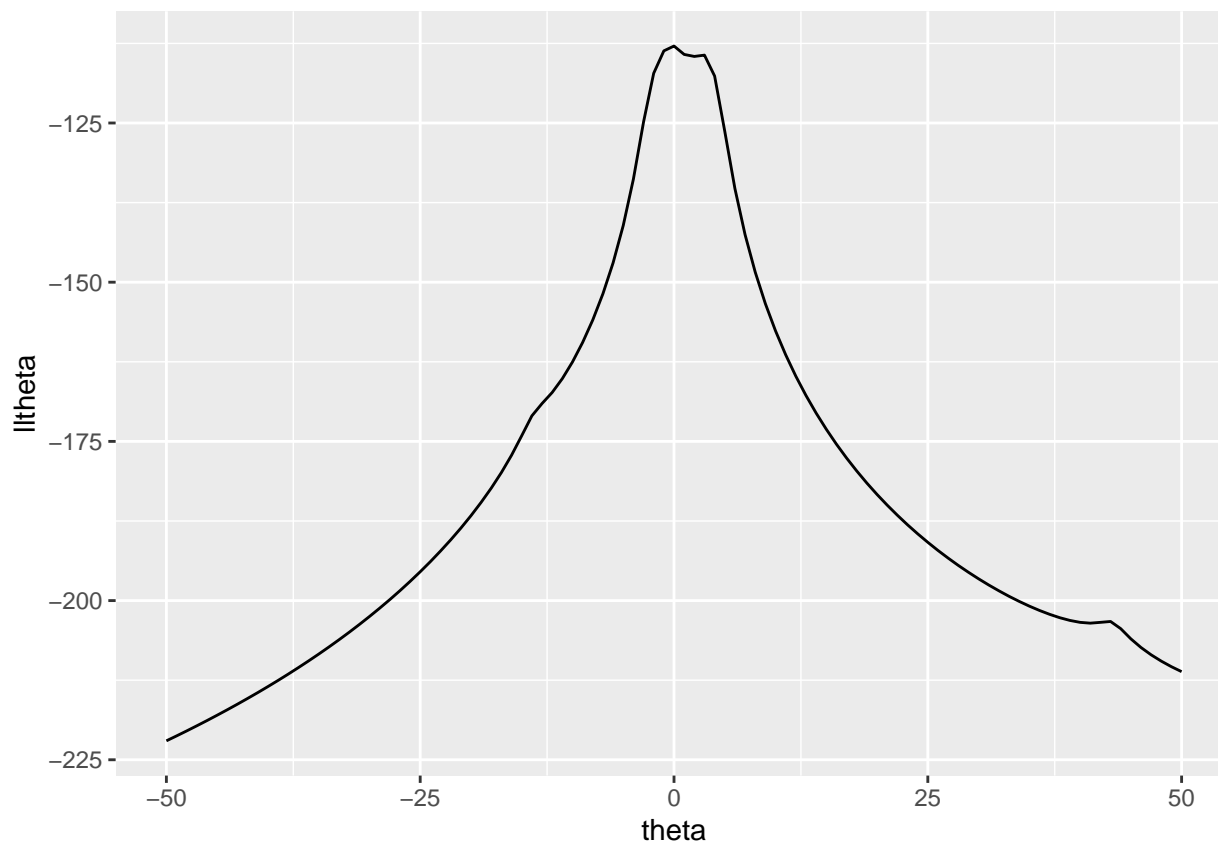
To complete this problem we need the likelihood, its derivative, and the information (second derivative of the likelihood). Recall that the N-R method for maximizing a function  $f(\alpha)$  is

$$\alpha_{t+1} = \alpha_t + \frac{f(\alpha_t)}{f'(\alpha_t)}.$$

However, in our case  $f(\alpha) = l'(\alpha)$  the derivative of the likelihood. This is the form of the N-R that I gave in the class notes. The Cauchy pdf is:  $\frac{1}{\pi[1+(x-\theta)^2]}$  and the likelihood is  $\prod_{i=1}^n \frac{1}{\pi[1+(x_i-\theta)^2]}$ . The log likelihood is  $-\sum_{i=1}^n [\pi + \log(1 + (x_i - \theta)^2)]$ .

$$\begin{aligned} \frac{dl}{d\theta} &= 2 \sum_{i=1}^n \left[ \frac{(x_i - \theta)}{1 + (x_i - \theta)^2} \right] \\ \frac{dl'}{d\theta} &= 2 \sum_{i=1}^n \frac{-[1 + (x_i - \theta)^2 + 2(x_i - \theta)^2]}{[1 + (x_i - \theta)^2]^2} \\ \frac{dl'}{d\theta} &= 2 \sum_{i=1}^n \frac{[(x_i - \theta)^2 - 1]}{[1 + (x_i - \theta)^2]^2} \end{aligned}$$

Graph the log likelihood function.



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

```

## create function to implement NR algorithm
my.nrfun<-function(x,theta.start,eps=1.e-4,max.iter=200) {
  if (missing(theta.start)) theta.start <- median(x)
  theta <- theta.start
  n<-length(x)
  score <- sum(2 * (x - theta)/((x-theta)^2 + 1)^2)
  iter<-1
  conv <- T
  while (abs(score)>eps && iter<=max.iter) {
    inform<-2 * sum((1 - (x - theta)^2)/((x-theta)^2 + 1)^2)
    theta <-theta + score/inform
    iter <- iter+1
    score <- sum(2 * (x - theta)/((x-theta)^2 + 1)^2)
  }

  if (abs(score)>eps) {
    print("No convergence")
    conv <- F
  }

  loglik <- -(length(x)*pi + sum(log(1+ (theta - x)^2)))
  inform <- 2 * sum((1 - (x - theta)^2)/((x-theta)^2 + 1)^2)
  output <- list(iter = iter,theta=theta,loglik=loglik,score = score, inform = inform, convergence = conv)
  output
}

test.cauchy <- my.nrfun(xdata,theta.start = median(xdata))

```

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

```

## [1] "No convergence"
## [1] "No convergence"
## [1] "No convergence"

##           [,1]      [,2]      [,3]      [,4]      [,5]
## iter      65       31       40       8       10
## theta     numeric,20 numeric,20 numeric,20 numeric,20 numeric,20
## loglik    -103.2451 -55.49843 -55.49841 -54.76436 -55.88027
## score     9.432695e-05 7.806729e-05 8.38028e-05 4.471007e-05 4.23822e-05
## inform    -0.4323075 3.525717 3.525711 -1.828038 7.004381
## convergence TRUE     TRUE     TRUE     TRUE     TRUE
##           [,6]      [,7]      [,8]      [,9]
## iter      10       201      201      201
## theta     numeric,20 numeric,20 numeric,20 numeric,20
## loglik    -55.8802 -127.5685 -127.7982 -138.8603
## score     -4.463069e-05 0.003478195 0.003389556 -0.0003774876
## inform     7.004357 -0.08405274 -0.08300806 -0.05564811
## convergence TRUE     FALSE     FALSE     FALSE

```

Discuss your results. Is the mean of the data a good starting point?

No, the mean is ill defined for the Cauchy and influenced by the outliers that will show up in the Cauchy.

(5 points) 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 Poisson, which will be discussed more later) and a linear mixed model (LMM).

The GLM assumes independence in errors, which results in independence in the outcomes. The LMM allows correlation between observations and a better choice for repeatedly collected data on an individual. The LMM has a very flexible approach to incorporating correlation between observations including directly modeling the correlation through a marginal model of the outcome and specifying between subject variation in intercepts and slope values (ie random effects) or even a combination of the two.

(5 points) 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.

Profile likelihood is an approach to jointly maximize the full likelihood of the observed data by maximizing each parameter with the other parameter(s) fixed at a particular value. Then a set of values (grid search) is used to derive a set of maximization values for the parameter being maximized over. The maximum of all the profile maximums is then the joint maximum. Restricted maximum likelihood is derived by rotating the data according to an orthogonal transformation. The result is a restricted distribution (restricted by the "rules" of the transformation). It is a distribution that is also only defined by the variance terms (the mean is zero) so the variance estimates serve as a plug in for the full maximum likelihood estimator of the  $\beta$ 's in the LMM.

In this next part of the homework you will now fit the models that you wrote up in homework 3. The data can be found in Week 1 out of class materials and as a link in Homework 4. Recall, an investigator was interested in understanding whether Cortisol (a stress hormone) secretion differs in women suffering from depression. The investigator measured Cortisol levels 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), for this analysis the investigator was interested in differences in the circadian pattern between the groups. The data were divided into 6 blocks of 4 hours and averaged to obtain a set of "block means".

(10 points)

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.

*# Fit a multiple linear regression with no correlation assumed*

```
lr.fit<-lm(Cortisol ~ as.factor(Time) + as.factor(casecontrol), data=Cort_P1)
lr.summary <- summary(lr.fit)
```

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.

```
lr.summary$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	3.0501178	0.5087862	5.994891	5.745148e-09
## as.factor(Time)2	3.9697332	0.6661575	5.959151	6.991077e-09
## as.factor(Time)3	10.0790961	0.6661575	15.130200	5.710893e-39
## as.factor(Time)4	6.1978910	0.6661575	9.303943	2.727306e-18
## as.factor(Time)5	4.4042123	0.6661575	6.611368	1.701672e-10

```
## as.factor(Time)6          1.7518341  0.6661575  2.629760 8.977900e-03
## as.factor(casecontrol)Patient 0.9210898  0.3846062  2.394891 1.722849e-02
```

I will interpret Time2 and Time3. The mean cortisol levels are 3.97 (SE=0.67) units higher in time block 2 compared to time block 1. In addition, The mean cortisol levels are 10.08 (SE=0.67) units higher in time block 3 compared to time block 1.

Will these standard errors be too big or small and why?

I thought that these SE would be too small compared to models that account for correlation because they are treating all the data (52\*6) as the number in the SE calculation. The actual DF should be smaller because of the correlation between observations on the same individual. This was the case for the group variable but not time variable. Note the two sources of variation differ here (between subject for case/control and within for Time).

(15 points)

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.

```
# Fit a mixed effects model with a no random effects and an unstructured var-cov

lmm.un.fit<-glms(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                 data=Cort_P1, correlation = corSymm(form = ~ 1 | SubjectId))
lmm.un.summary <- summary(lmm.un.fit)

un.varcov<-lmm.un.fit$modelStruct$corStruct
```

The correlation structure has weaker correlations the farther apart the times are. Note, time 1 and 6 are actually as close in time as 1 and 2 based on the 24 hour day. The correlations range from 0.2-0.86 (wide range). Mostly there are modest correlations between observations.

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.

```
lmm.un.summary$tTable
```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.1311256	0.6060442	5.166497	4.317354e-07
## as.factor(Time)2	3.9697332	0.5073507	7.824436	8.429569e-14
## as.factor(Time)3	10.0790961	0.5243558	19.221865	1.686143e-54
## as.factor(Time)4	6.1978910	0.5849491	10.595607	1.525294e-22
## as.factor(Time)5	4.4042123	0.6098695	7.221565	4.122203e-12
## as.factor(Time)6	1.7518341	0.3212775	5.452713	1.025850e-07
## as.factor(casecontrol)Patient	0.7590742	0.7254012	1.046420	2.961958e-01

I will interpret Time2 and Time3. The mean cortisol levels are 3.97 (SE=0.51) units higher in time block 2 compared to time block 1. In addition, The mean cortisol levels are 10.08 (SE=0.52) units higher in time block 3 compared to time block 1.

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.

Mean cortisol levels are 0.76 (SE=0.73) units higher in the depressed patients compared to the controls.

(15 points) 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

```
# Fit a mixed effects model with a no random effects and an unstructured var-cov

lmm.cs.fit<-glsl(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                 data=Cort_P1, correlation = corCompSymm(form = ~ 1 |SubjectId))
lmm.cs.summary <- summary(lmm.cs.fit)

cs.varcov<-lmm.cs.fit$modelStruct$corStruct
cs.varcov
```

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

The correlation between observations is moderate at 0.49. Based on the unstructured this seems like a reasonable compromise between all the values we saw above.

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.

```
lmm.cs.summary$tTable
```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.0501178	0.5937646	5.136914	4.991635e-07
## as.factor(Time)2	3.9697332	0.4770246	8.321862	2.935779e-15
## as.factor(Time)3	10.0790961	0.4770246	21.129090	1.131148e-61
## as.factor(Time)4	6.1978910	0.4770246	12.992812	5.089272e-31
## as.factor(Time)5	4.4042123	0.4770246	9.232673	4.593010e-18
## as.factor(Time)6	1.7518341	0.4770246	3.672419	2.836408e-04
## as.factor(casecontrol)Patient	0.9210898	0.7179734	1.282902	2.005003e-01

The coefficients are the same as above and the interpretation doesn't change. Note the SE's are different and the same for all time points (unlike for the unstructure) and then different for the group level factor. One is a within subject SE (time) and one is a between subject SE (group).

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.

Mean cortisol levels are 0.92 (SE=0.72) units higher in the depressed patients compared to the controls. This is a slightly different estimate from the unstructured (maybe related to bias in estimation with the changes in correlation structure).

(15 points)

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

```
# Fit a mixed effects model with a no random effects and an AR1 var-cov

lmm.ar.fit<-glsl(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                 data=Cort_P1, correlation = corAR1(form = ~ 1 |SubjectId))
```

```
lmm.ar.summary <- summary(lmm.ar.fit)

ar.varcov<-lmm.ar.fit$modelStruct$corStruct
ar.varcov
```

```
## Correlation structure of class corAR1 representing
##      Phi
## 0.554872
```

The correlation between observations is moderate at 0.55 and higher than the CS estimate. However, values that are farther apart get this value exponentiated. In this model that means that the 1-6 time difference will have a small correlation not reflective of the circadian pattern.

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.

```
lmm.ar.summary$tTable
```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.1383874	0.5645829	5.558772	5.933059e-08
## as.factor(Time)2	3.9697332	0.4484124	8.852862	7.115987e-17
## as.factor(Time)3	10.0790961	0.5591457	18.025886	5.824353e-50
## as.factor(Time)4	6.1978910	0.6120058	10.127178	5.708063e-21
## as.factor(Time)5	4.4042123	0.6394541	6.887456	3.251161e-11
## as.factor(Time)6	1.7518341	0.6541879	2.677876	7.810441e-03
## as.factor(casecontrol)Patient	0.7445505	0.6095681	1.221440	2.228629e-01

Coefficients and interpretation are the same as the other models.

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.

Mean cortisol levels are 0.74 (SE=0.61) units higher in the depressed patients compared to the controls. This is a slightly different estimate from the compound symmetry model and closer to unstructured.

(10 points) Write a paragraph comparing and contrasting the parameter estimates for  $\beta$  (the mean differences) and the standard errors across the different variance-covariance structures.

The time estimates are the same across the models. The group estimates were more sensitive to the misspecification of the correlation structure. The SE's were smallest for the CS model (which is also perhaps the most oversimplified). The SE's for independence were slightly high compared to some models. The SE's were all different for the unstructured and AR(1) models reflecting a different correlation for each pairing (Note: we only have coefficients from some of the comparisons so the AR(1) SE's will be similar for all comparisons that are the same distance apart). The SE's were impacted the most by the differences in the models (compared to the coefficient estimates).

(25 points) Refit the models using method=ML and compare and contrast the findings with the different estimation approaches. How were the  $\beta$  coefficients impacted (if at all)? How were the SE estimates (and variance-covariance structures) impacted by the different methods (if at all)? Explain.

```
# Fit a mixed effects model with a no random effects and an unstructured var-cov

lmm.un.fit.ml<-glms(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                    data=Cort_P1, correlation = corSymm(form = ~ 1 | SubjectId),method="ML")
lmm.un.summary.ml <- summary(lmm.un.fit.ml)
```

```

un.varcov.ml<-lmm.un.fit.ml$modelStruct$corStruct

# Fit a mixed effects model with a no random effects and an unstructured var-cov

lmm.cs.fit.ml<-glS(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                    data=Cort_P1, correlation = corCompSymm(form = ~ 1 |SubjectId),method="ML")
lmm.cs.summary.ml <- summary(lmm.cs.fit.ml)

cs.varcov.ml<-lmm.cs.fit.ml$modelStruct$corStruct

##4. Fit a mixed effects model with a no random effects and an AR1 var-cov

lmm.ar.fit.ml<-glS(Cortisol ~ as.factor(Time) + as.factor(casecontrol),
                    data=Cort_P1, correlation = corAR1(form = ~ 1 |SubjectId),method="ML")
lmm.ar.summary.ml <- summary(lmm.ar.fit.ml)

ar.varcov.ml<-lmm.ar.fit.ml$modelStruct$corStruct

lmm.un.summary.ml$table

```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.1324884	0.6022588	5.201233	3.637949e-07
## as.factor(Time)2	3.9697332	0.5082941	7.809914	9.279541e-14
## as.factor(Time)3	10.0790961	0.5249026	19.201842	2.007653e-54
## as.factor(Time)4	6.1978910	0.5874410	10.550661	2.166227e-22
## as.factor(Time)5	4.4042123	0.6122056	7.194008	4.899999e-12
## as.factor(Time)6	1.7518341	0.3224932	5.432158	1.139639e-07
## as.factor(casecontrol)Patient	0.7563485	0.7182411	1.053057	2.931486e-01

```
lmm.cs.summary.ml$table
```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.0501178	0.5904586	5.165676	4.334813e-07
## as.factor(Time)2	3.9697332	0.4778061	8.308251	3.223692e-15
## as.factor(Time)3	10.0790961	0.4778061	21.094531	1.522444e-61
## as.factor(Time)4	6.1978910	0.4778061	12.971560	6.085242e-31
## as.factor(Time)5	4.4042123	0.4778061	9.217572	5.127908e-18
## as.factor(Time)6	1.7518341	0.4778061	3.666412	2.900935e-04
## as.factor(casecontrol)Patient	0.9210898	0.7120630	1.293551	1.967994e-01

```
lmm.ar.summary.ml$table
```

##	Value	Std.Error	t-value	p-value
## (Intercept)	3.1374364	0.5626360	5.576316	5.415064e-08
## as.factor(Time)2	3.9697332	0.4491800	8.837734	7.926058e-17
## as.factor(Time)3	10.0790961	0.5594212	18.017007	6.295117e-50
## as.factor(Time)4	6.1978910	0.6117414	10.131555	5.520066e-21
## as.factor(Time)5	4.4042123	0.6387455	6.895098	3.103500e-11
## as.factor(Time)6	1.7518341	0.6531502	2.682131	7.714100e-03
## as.factor(casecontrol)Patient	0.7464525	0.6060828	1.231602	2.190467e-01



```
un.varcov.ml
```

```
## Correlation structure of class corSymm representing
## Correlation:
## 1      2      3      4      5
## 2 0.447
## 3 0.411 0.523
## 4 0.262 0.446 0.550
## 5 0.198 0.437 0.552 0.863
## 6 0.778 0.312 0.510 0.514 0.519
```

```
cs.varcov.ml
```

```
## Correlation structure of class corCompSymm representing
##      Rho
## 0.4855424
```

```
ar.varcov.ml
```

```
## Correlation structure of class corAR1 representing
##      Phi
## 0.5510902
```

The results are nearly identical in this problem. The unstructured correlation estimates are slightly different but very close and similar in interpretation.