

Logit Models

Bill

2019-12-23

Contents

1	Basics	5
1.1	Logit	5
1.2	Probit	6
2	MLE	7
3	Twitter Example	9
3.1	Model	9
3.2	Simulating Data of Senators on Twitter	10
3.3	Simulating Data of Conservative Users on Twitter and Model Testing	11
3.4	Simulating Data of Liberal Users on Twitter and Model Testing .	13
4	Linear Mixed Models	17
4.1	Calculate mean	17
4.2	Test the treatment effect	19
4.3	Another example	20
4.4	Full LMM model	22
4.5	Serial correlations in time and space	23
5	Generalized Linear Mixed Models	27
5.1	Basics	27
5.2	Some References	27

Chapter 1

Basics

1.1 Logit

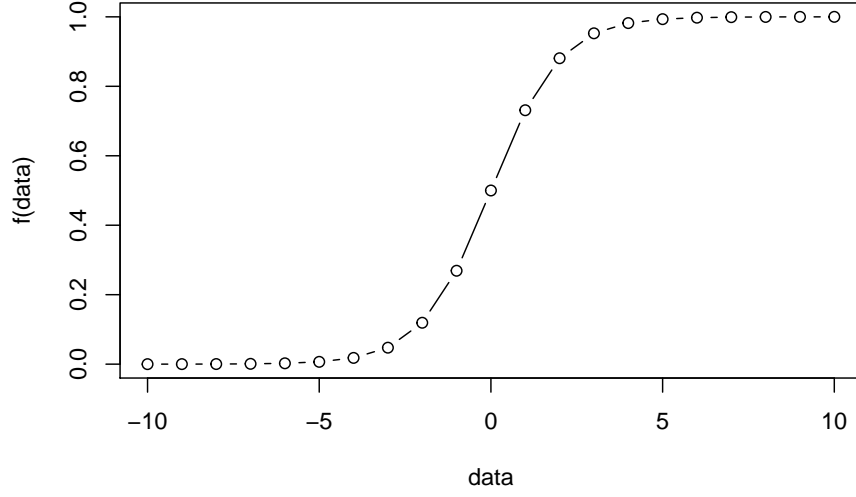
$$f(x) = \log\left(\frac{p(y=1)}{1-p(y=1)}\right)$$

The basic idea of logistic regression:

$$p(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$

Thus, $e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}$ can be from $-\infty$ to $+\infty$, and $p(y=1)$ will be always within the range of $(0, 1)$.

```
f<-function(x){exp(x)/(1+exp(x))}  
data<-seq(-10,10,1)  
plot(data,f(data),type = "b")
```



We can also write the function into another format as follows:

$$\log \frac{p(y=1)}{1-p(y=1)} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

1.2 Probit

$$\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = \Phi^{-1}(p)$$

Thus,

$$\Phi(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n) = p(y=1)$$

Chapter 2

MLE

The probability of $y = 1$ is as follows:

$$p = p(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$

Thus, the likelihood function is as follows:

$$\begin{aligned} L &= \prod p^{y_i} (1-p)^{1-y_i} = \prod \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}} \right)^{1-y_i} \\ &= \prod (1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)})^{-y_i} (1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n})^{-(1-y_i)} \end{aligned}$$

Thus, the log-likelihood is as follows:

$$\log L = \sum (-y_i \cdot \log(1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}) - (1-y_i) \cdot \log(1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}))$$

Typically, optimisers minimize a function, so we use negative log-likelihood as minimising that is equivalent to maximising the log-likelihood or the likelihood itself.

#Source of R code: <https://www.r-bloggers.com/logistic-regression/>

```
mle.logreg = function(fmla, data)
{
  # Define the negative log likelihood function
  logl <- function(theta,x,y){
    y <- y
```

```

x <- as.matrix(x)
beta <- theta[1:ncol(x)]

# Use the log-likelihood of the Bernoulli distribution, where p is
# defined as the logistic transformation of a linear combination
# of predictors, according to logit(p)=(x%%beta)
loglik <- sum(-y*log(1 + exp(-(x%%beta))) - (1-y)*log(1 + exp(x%%beta)))
return(-loglik)
}

# Prepare the data
outcome = rownames(attr(terms(fmla), "factors"))[1]
dfrTmp = model.frame(data)
x = as.matrix(model.matrix(fmla, data=dfrTmp))
y = as.numeric(as.matrix(data[,match(outcome,colnames(data))]))

# Define initial values for the parameters
theta.start = rep(0,(dim(x)[2]))
names(theta.start) = colnames(x)

# Calculate the maximum likelihood
mle = optim(theta.start,logl,x=x,y=y, method = 'BFGS', hessian=T)
out = list(beta=mle$par,vcov=solve(mle$hessian),ll=2*mle$value)
}

mydata = read.csv(url('https://stats.idre.ucla.edu/stat/data/binary.csv'))
mylogit1 = glm(admit~gre+gpa+as.factor(rank), family=binomial, data=mydata)

mydata$rank = factor(mydata$rank) #Treat rank as a categorical variable
fmla = as.formula("admit~gre+gpa+rank") #Create model formula
mylogit2 = mle.logreg(fmla, mydata) #Estimate coefficients

print(cbind(coef(mylogit1), mylogit2$beta))

##                [,1]      [,2]
## (Intercept)    -3.989979073 -3.772676422
## gre             0.002264426  0.001375522
## gpa             0.804037549  0.898201239
## as.factor(rank)2 -0.675442928 -0.675543009
## as.factor(rank)3 -1.340203916 -1.356554831
## as.factor(rank)4 -1.551463677 -1.563396035

```


Chapter 3

Twitter Example

The following is part of my course project for Stat 536. It aims to replicate part of the findings from Barbera (2015) Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. Political Analysis 23 (1). Note that, the following model is much simpler than that in the original paper.

3.1 Model

Suppose that a Twitter user is presented with a choice between following or not following another target $j \in \{1, \dots, m\}$. Let $y_j = 1$ if the user decides to follow j , and $y_j = 0$ otherwise.

$$y_j = \begin{cases} 1 & \text{Following} \\ 0 & \text{NotFollowing} \end{cases}$$

$$p(y_j = 1|\theta) = \frac{\exp(-\theta_0|\theta_1 - x_j|^2)}{1 + \exp(-\theta_0|\theta_1 - x_j|^2)}$$

We additionally know the priors of θ .

$$\theta_i \sim N(0, 10^2)(i = 0, 1)$$

The likelihood function is as follows.

$$L(Y|\theta) = \prod_{j=1}^m \left(\frac{\exp(-\theta_0|\theta_1 - x_j|^2)}{1 + \exp(-\theta_0|\theta_1 - x_j|^2)} \right)^{y_j} \left(1 - \frac{\exp(-\theta_0|\theta_1 - x_j|^2)}{1 + \exp(-\theta_0|\theta_1 - x_j|^2)} \right)^{(1-y_j)}$$

Thus, the posterior is as follows.

$$L(Y|\theta) \cdot N(\theta_0|0, 10) \cdot N(\theta_1|0, 10)$$

$$\propto \prod_{j=1}^m \left(\frac{\exp(-\theta_0|\theta_1 - x_j|^2)}{1 + \exp(-\theta_0|\theta_1 - x_j|^2)} \right)^{y_j} \left(1 - \frac{\exp(-\theta_0|\theta_1 - x_j|^2)}{1 + \exp(-\theta_0|\theta_1 - x_j|^2)} \right)^{(1-y_j)} \cdot \exp\left(-\frac{1}{2}\left(\frac{\theta_0}{10}\right)^2\right) \cdot \exp\left(-\frac{1}{2}\left(\frac{\theta_1}{10}\right)^2\right)$$

3.2 Simulating Data of Senators on Twitter

Assume that we have 100 senators, 50 Democrats and 50 Republicans, who we know their ideology. Assume that Democrats have negative ideology scores to indicate that they are more liberal, whereas Republicans have positive scores to indicate that they are more conservative. The following is data simulation for senators.

```
# Republicans are more conservative, and they have positive numbers.
Republicans<-c()
Republicans<-rnorm(50,1,0.5)
No_Republicans<-rep(1:50,1)
Part_1<-cbind(No_Republicans,Republicans)

# Democrats are more liberal, and they have negative numbers.
Democrats<-c()
Democrats<-rnorm(50,-1,0.5)
No_Democrats<-rep(51:100,1)
Part_2<-cbind(No_Democrats,Democrats)
Data_Elites<-rbind(Part_1,Part_2)
Data_Elites<-as.data.frame(Data_Elites)
colnames(Data_Elites) <- c("Elite_No","Elite_ideology")

head(Data_Elites)
```

```
##   Elite_No Elite_ideology
## 1         1         1.1475584
## 2         2         1.1754757
## 3         3         0.6836400
## 4         4        -0.2846277
## 5         5         1.0901397
## 6         6         1.0112953
```

3.3 Simulating Data of Conservative Users on Twitter and Model Testing

Assume that we observe one Twitter user, who is more conservative. To simulate Twitter following data for this user, I assign this user to follow more Republican senators. Thus, if the Metropolis Hastings algorithm works as intended, we would expect to see a positive estimated value for their ideology. Importantly, as we can see in the histogram below, the estimated value indeed is positive, providing preliminary evidence for the statistical model and the algorithm. In addition, for the acceptance rate, we can see that the constant has a lower number than ideology, since we only accept a constant when it is positive.

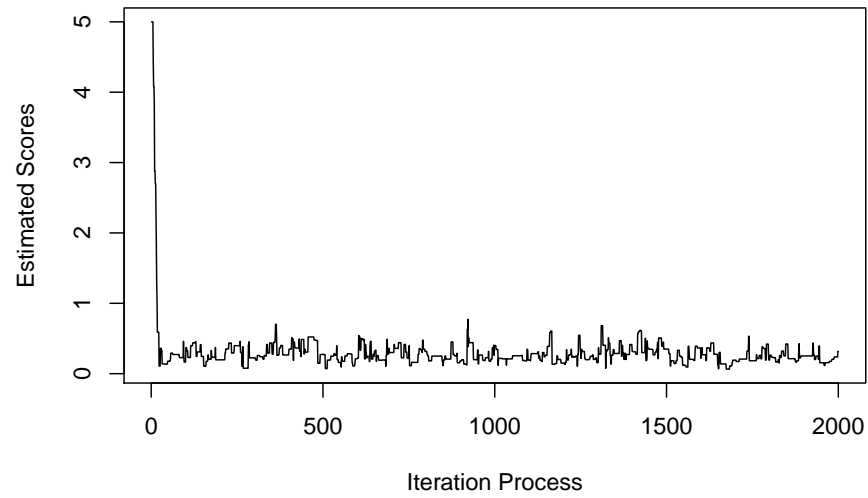
```
#This user approximately follows 45 Republican Senators and 10 Democrat Senators.
Data_user<-as.data.frame(matrix(c(ifelse(runif(50)<.1,0,1),ifelse(runif(50)<.8,0,1))), 100, 1)
colnames(Data_user)<-c("R_User")
Data_combined<-cbind(Data_Elites,Data_user)

X_data<-Data_combined$Elite_ideology
Y_data<-Data_combined$R_User

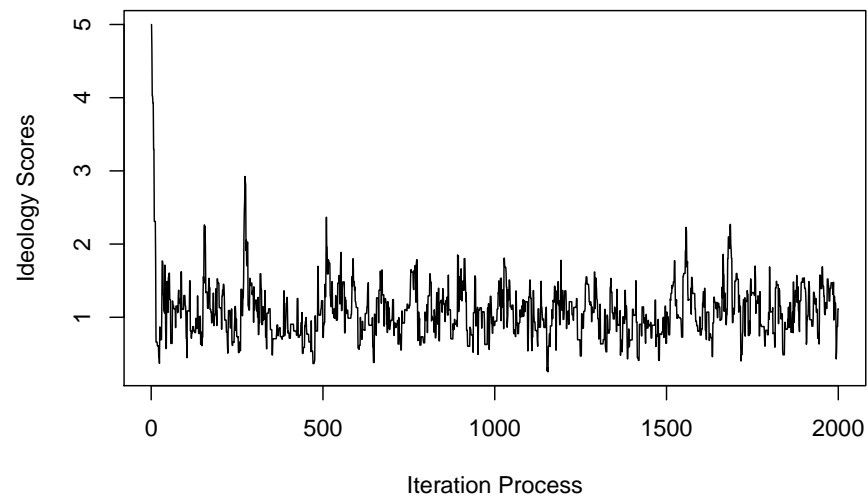
fit_C<-Bayes_logit(Y_data,X_data)
fit_C$acceptance_rate
```

```
## [1] 0.2036018 0.4847424
```

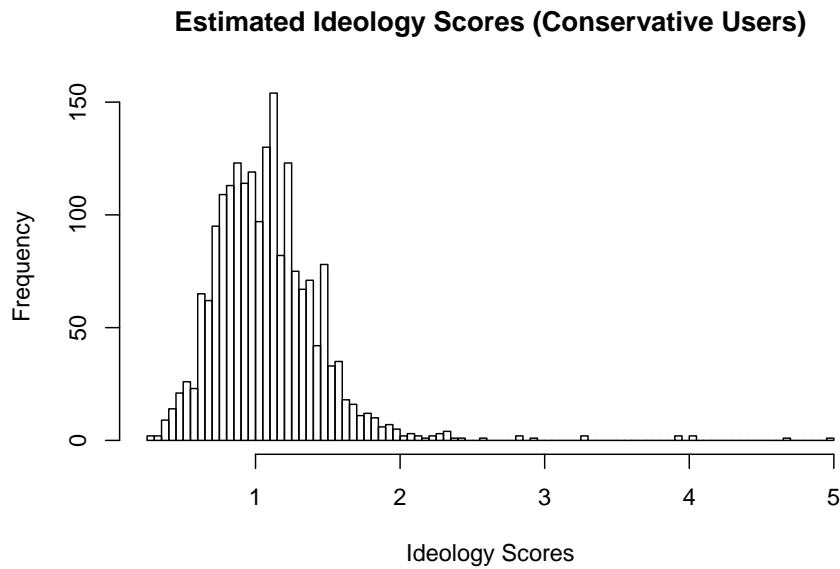
```
plot(fit_C$theta[,1],main="Constant (Conservative Users)",
     xlab="Iteration Process",ylab="Estimated Scores",type="l")
```

Constant (Conservative Users)

```
plot(fit_C$theta[,2],main="Estimated Ideology Scores (Conservative Users)",  
     xlab="Iteration Process",ylab="Ideology Scores",type="l")
```

Estimated Ideology Scores (Conservative Users)

```
hist(fit_C$theta[,2],main="Estimated Ideology Scores (Conservative Users)",
     xlab="Ideology Scores",breaks = 100)
```



3.4 Simulating Data of Liberal Users on Twitter and Model Testing

To further verify the Metropolis Hastings algorithm, I plan to test the opposite estimate. Specifically, assume that we observe another user, who is more liberal. To simulate Twitter following data for this user, I assign this user to follow more Democrat senators. In this case, we would expect to see a negative value for their estimated ideology. As we can see in the histogram shown below, as expected, the estimated value is negative, providing convergent evidence for the model and the algorithm.

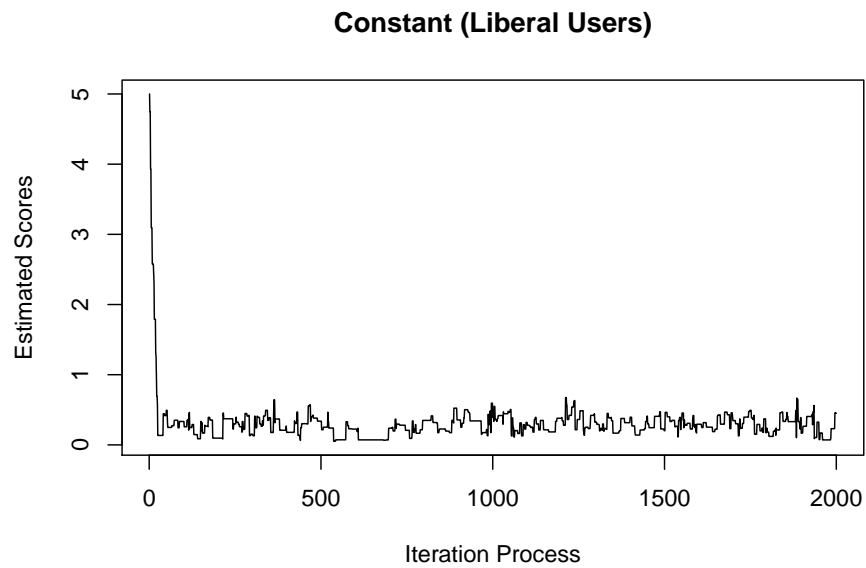
```
#This user approximately follows 10 Republican Senators and 45 Democrat Senators.
Data_user<-as.data.frame(matrix(c(ifelse(runif(50)<.8,0,1),ifelse(runif(50)<.1,0,1))), 100, 1)
colnames(Data_user)<-c("L_User")
Data_combined<-cbind(Data_Elites,Data_user)

X_data<-Data_combined$Elite_ideology
Y_data<-Data_combined$L_User
```

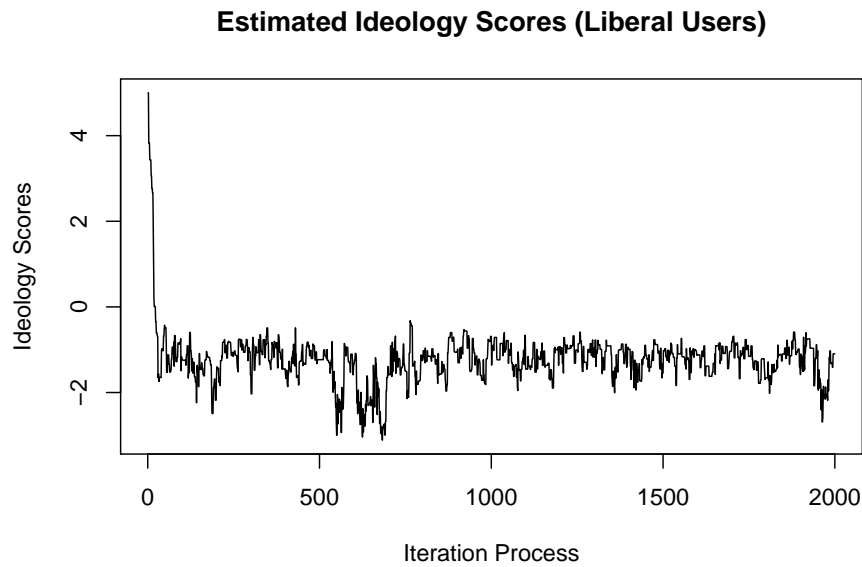
```
fit_L<-Bayes_logit(Y_data,X_data)
fit_L$acceptance_rate
```

```
## [1] 0.1815908 0.4672336
```

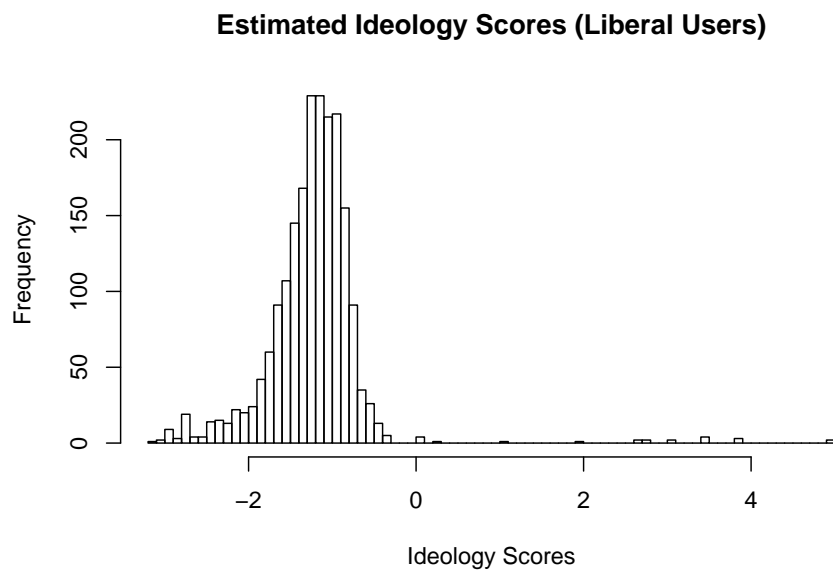
```
plot(fit_L$theta[,1],main="Constant (Liberal Users)",
      xlab="Iteration Process",ylab="Estimated Scores",type="l")
```



```
plot(fit_L$theta[,2],main="Estimated Ideology Scores (Liberal Users)",
      xlab="Iteration Process",ylab="Ideology Scores",type="l")
```



```
hist(fit_L$theta[,2],main="Estimated Ideology Scores (Liberal Users)",
     xlab="Ideology Scores",breaks = 100)
```



Chapter 4

Linear Mixed Models

The following is a shortened version of Jonathan Rosenblatt's LMM tutorial.
<http://www.john-ros.com/Rcourse/lme.html>.

In addition, another reference is from Douglas Bates's R package document.
https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf?fbclid=IwAR1nmmRP9A0BrhKdgBibNjM5acR_spTpXV8QlQGdmTWyQz3ZtV3LYn6kCbQ

Assume that y is a function of x and u , where x is the fixed effect and u is the random effect. Thus, we can get,

$$y|x, u = x'\beta + z'u + \epsilon$$

For random effect, one example can be that you want to test the treatment effect, and sample 8 observations from 4 groups. You measure before and after the treatment. In this case, x represents the treatment effect, whereas z represents the group effect (i.e., random effect). Note that, in this case, it reminds the paired t-test. Remember in SPSS, why do we do paired t-test? Typically, it is the case when we measure a subject (or, participant) twice. In this case, we can consider each participant as an unit of random effect (rather than as group in the last example.)

4.1 Calculate mean

The following code generates 4 numbers ($N(0, 10)$) for 4 groups. Then, replicate it within each group. That is, in the end, there are 8 observations.

Note that, in the following code, there are no “independent variables”. Both the linear model and mixed model are actually just trying to calculate the mean. Note that `$lmer(y~1+1|groups)` and `lmer(y 1|groups)` will generate the same results.

```

set.seed(123)
n.groups <- 4 # number of groups
n.repeats <- 2 # samples per group
#Generating index for observations belong to the same group
groups <- as.factor(rep(1:n.groups, each=n.repeats))
n <- length(groups)
#Generating 4 random numbers, assuming normal distribution
z0 <- rnorm(n.groups, 0, 10)
z <- z0[as.numeric(groups)] # generate and inspect random group effects
z

```

```

## [1] -5.6047565 -5.6047565 -2.3017749 -2.3017749 15.5870831 15.5870831 0.7050839
## [8] 0.7050839

```

```

epsilon <- rnorm(n,0,1) # generate measurement error
beta0 <- 2 # this is the actual parameter of interest! The global mean.
y <- beta0 + z + epsilon # sample from an LMM

# fit a linear model assuming independence
# i.e., assume that there is no "group things".
lm.5 <- lm(y~1)

# fit a mixed-model that deals with the group dependence
#install.packages("lme4")
library(lme4)
lme.5.a <- lmer(y~1+1|groups)
lme.5.b <- lmer(y~1|groups)
lm.5

```

```

##
## Call:
## lm(formula = y ~ 1)
##
## Coefficients:
## (Intercept)
##          4.283

```

```
lme.5.a
```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 + 1 | groups
## REML criterion at convergence: 36.1666
## Random effects:
## Groups      Name          Std.Dev.

```

```
## groups (Intercept) 8.8521
## Residual          0.8873
## Number of obs: 8, groups: groups, 4
## Fixed Effects:
## (Intercept)
##          4.283
```

```
lme.5.b
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 | groups
## REML criterion at convergence: 36.1666
## Random effects:
## Groups Name Std.Dev.
## groups (Intercept) 8.8521
## Residual          0.8873
## Number of obs: 8, groups: groups, 4
## Fixed Effects:
## (Intercept)
##          4.283
```

4.2 Test the treatment effect

As we can see that, LLM and paired t-test generate the same t-value.

```
times<-rep(c(1,2),4) # first time and second time
times
```

```
## [1] 1 2 1 2 1 2 1 2
```

```
data_combined<-cbind(y,groups,times)
data_combined
```

```
##           y groups times
## [1,] -3.4754687     1     1
## [2,] -1.8896915     1     2
## [3,]  0.1591413     2     1
## [4,] -1.5668361     2     2
## [5,] 16.9002303     3     1
## [6,] 17.1414212     3     2
## [7,]  3.9291657     4     1
## [8,]  3.0648977     4     2
```

```
lme_diff_times<- lmer(y~times+(1|groups))
```

```
t_results<-t.test(y~times, paired=TRUE)
```

```
lme_diff_times
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ times + (1 | groups)
## REML criterion at convergence: 35.0539
## Random effects:
##   Groups   Name      Std.Dev.
##   groups   (Intercept) 8.845
##   Residual                1.013
## Number of obs: 8, groups: groups, 4
## Fixed Effects:
##   (Intercept)      times
##           4.5691      -0.1908
```

```
print("The following results are from paired t-test")
```

```
## [1] "The following results are from paired t-test"
```

```
t_results$statistic
```

```
##           t
## 0.2664793
```

4.3 Another example

```
data(Dyestuff, package='lme4')
attach(Dyestuff)
```

```
## The following objects are masked from Dyestuff (pos = 5):
##
##   Batch, Yield
```

```
Dyestuff
```

```
##      Batch Yield
## 1      A  1545
## 2      A  1440
## 3      A  1440
## 4      A  1520
## 5      A  1580
## 6      B  1540
## 7      B  1555
## 8      B  1490
## 9      B  1560
## 10     B  1495
## 11     C  1595
## 12     C  1550
## 13     C  1605
## 14     C  1510
## 15     C  1560
## 16     D  1445
## 17     D  1440
## 18     D  1595
## 19     D  1465
## 20     D  1545
## 21     E  1595
## 22     E  1630
## 23     E  1515
## 24     E  1635
## 25     E  1625
## 26     F  1520
## 27     F  1455
## 28     F  1450
## 29     F  1480
## 30     F  1445
```

```
lme_batch<- lmer( Yield ~ 1 + (1|Batch) , Dyestuff )
summary(lme_batch)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Yield ~ 1 + (1 | Batch)
##      Data: Dyestuff
##
## REML criterion at convergence: 319.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.4117 -0.7634  0.1418  0.7792  1.8296
##
```

```
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   Batch    (Intercept) 1764      42.00
##   Residual                    2451      49.51
## Number of obs: 30, groups:  Batch, 6
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  1527.50      19.38    78.8
```

4.4 Full LMM model

```
data(sleepstudy, package='lme4')
attach(sleepstudy)
```

```
## The following objects are masked from sleepstudy (pos = 5):
##
##   Days, Reaction, Subject
```

```
head(sleepstudy)
```

```
##   Reaction Days Subject
## 1 249.5600    0     308
## 2 258.7047    1     308
## 3 250.8006    2     308
## 4 321.4398    3     308
## 5 356.8519    4     308
## 6 414.6901    5     308
```

```
nrow(sleepstudy)
```

```
## [1] 180
```

```
#lmer ( Reaction ~ Days + ( Days | Subject ) , data= sleepstudy )
lmer ( Reaction ~ Days + ( Days| Subject ) , data= sleepstudy )
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (Days | Subject)
##   Data: sleepstudy
## REML criterion at convergence: 1743.628
```

```
## Random effects:
## Groups   Name                Std.Dev. Corr
## Subject  (Intercept) 24.737
##          Days        5.923   0.07
## Residual                25.592
## Number of obs: 180, groups: Subject, 18
## Fixed Effects:
## (Intercept)                Days
##          251.41             10.47
```

4.5 Serial correlations in time and space

The hierarchical model of $y|x, u = x'\beta + z'u + \epsilon$ can work well for correlations within blocks, but not for correlations in time as the correlations decay in time. The following uses nlme package to calculate time serial data.

```
library(nlme)
head(nlme::Ovary, n=50)
```

```
## Grouped Data: follicles ~ Time | Mare
##      Mare      Time follicles
## 1      1 -0.13636360         20
## 2      1 -0.09090910         15
## 3      1 -0.04545455         19
## 4      1  0.00000000         16
## 5      1  0.04545455         13
## 6      1  0.09090910         10
## 7      1  0.13636360         12
## 8      1  0.18181820         14
## 9      1  0.22727270         13
## 10     1  0.27272730         20
## 11     1  0.31818180         22
## 12     1  0.36363640         15
## 13     1  0.40909090         18
## 14     1  0.45454550         17
## 15     1  0.50000000         14
## 16     1  0.54545450         18
## 17     1  0.59090910         14
## 18     1  0.63636360         16
## 19     1  0.68181820         17
## 20     1  0.72727270         18
## 21     1  0.77272730         18
## 22     1  0.81818180         17
## 23     1  0.86363640         14
```

```
## 24    1  0.90909090    12
## 25    1  0.95454550    12
## 26    1  1.00000000    14
## 27    1  1.04545500    10
## 28    1  1.09090900    11
## 29    1  1.13636400    16
## 30    2 -0.15000000     6
## 31    2 -0.10000000     6
## 32    2 -0.05000000     8
## 33    2  0.00000000     7
## 34    2  0.05000000    16
## 35    2  0.10000000    10
## 36    2  0.15000000    13
## 37    2  0.20000000     9
## 38    2  0.25000000     7
## 39    2  0.30000000     6
## 40    2  0.35000000     8
## 41    2  0.40000000     8
## 42    2  0.45000000     6
## 43    2  0.50000000     8
## 44    2  0.55000000     7
## 45    2  0.60000000     9
## 46    2  0.65000000     6
## 47    2  0.70000000     4
## 48    2  0.75000000     5
## 49    2  0.80000000     8
## 50    2  0.85000000    11
```

```
fm10var.lme <- nlme::lme(fixed=follicles ~ sin(2*pi*Time) + cos(2*pi*Time),
                        data = Ovary,
                        random = pdDiag(~sin(2*pi*Time)),
                        correlation=corAR1() )
summary(fm10var.lme)
```

```
## Linear mixed-effects model fit by REML
## Data: Ovary
##      AIC      BIC    logLik
## 1563.448 1589.49 -774.724
##
## Random effects:
## Formula: ~sin(2 * pi * Time) | Mare
## Structure: Diagonal
##      (Intercept) sin(2 * pi * Time) Residual
## StdDev:      2.858385          1.257977 3.507053
##
```



```
## Correlation Structure: AR(1)
## Formula: ~1 | Mare
## Parameter estimate(s):
##      Phi
## 0.5721866
## Fixed effects: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
##                               Value Std.Error DF   t-value p-value
## (Intercept)          12.188089 0.9436602 295 12.915760 0.0000
## sin(2 * pi * Time)  -2.985297 0.6055968 295 -4.929513 0.0000
## cos(2 * pi * Time)  -0.877762 0.4777821 295 -1.837159 0.0672
## Correlation:
##              (Intr) s(*p*T
## sin(2 * pi * Time)  0.000
## cos(2 * pi * Time) -0.123 0.000
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -2.34910093 -0.58969626 -0.04577893  0.52931186  3.37167486
##
## Number of Observations: 308
## Number of Groups: 11
```


Chapter 5

Generalized Linear Mixed Models

5.1 Basics

The following is the note from Charle E. McCulloch's "Maximum likelihood algorithms for Generalized Linear Mixed Models"

5.2 Some References

<http://www.biostat.umn.edu/~baolin/teaching/linmods/glmm.html>

http://www.biostat.umn.edu/~baolin/teaching/probmods/GLMM_mcmc.html

<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>