# EnCPos

Fernanda Lang Schumacher

23/11/2020

Motivation: sleepstudy data

- ▶ The average reaction time per day for subjects was evaluated by Gregory et al. (2003) in a sleep deprivation study.
- ▶ On day 0 the subjects had their normal amount of sleep and starting that night they were restricted to 3 hours of sleep per night for 9 days, and the reaction time basead on a series of tests was measured on each day for each subject.
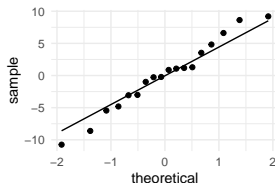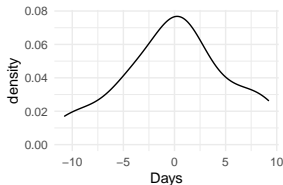- ▶ The data are avaliable at the R package *lme4*.
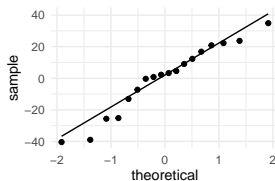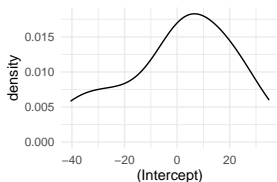
Motivation: sleepstudy data
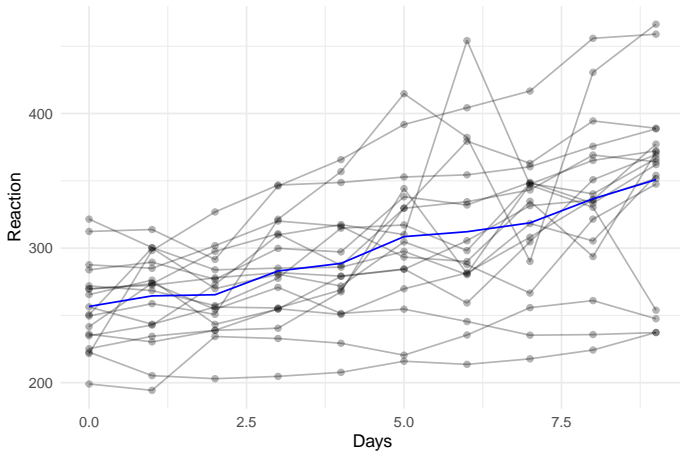
Introduction

- Bullet 1
- Bullet 2
- Bullet 3

Linear mixed models

Fitting a LMM to the sleepstudy dataset

```
fitlme <- lme(Reaction~Days,data=sleepstudy,
              random=~Days|Subject)
```

Model formulation

Within-subject dependence structures

Tools for model evaluation

The R package *skewlmm*

- ▶ The package *skewlmm* implemets an EM-type algorithm in R using S3 class, containing methods for estimating and predicting the SM(S)N-LMM.

- ▶ It has an user-friendly interface with generic R functions `print`, `summary`, `plot`, `fitted`, `residuals` and `predict` implemented.

- ▶ The main functions in the package are `smsn.lmm()` and `smn.lmm()`, which fit a SMSN-LMM and a SMN-LMM, respectively.

The R package *skewlmm*

The basic syntax of these functions is as follows:

```
smsn.lmm(data, formFixed, groupVar, formRandom,
         depStruct, distr, ...)
smn.lmm(data, formFixed, groupVar, formRandom,
        depStruct, distr, ...)
```

where

- ▶ data: A data frame containing all the variables to be used in the model.

- ▶ formFixed: A two-sided linear formula object describing the fixed effects part of the model.

- ▶ groupVar: A character containing the name of the variable which represents the subjects or groups in data.

- ▶ `formRandom`: A one-sided linear formula object describing the random effects part of the model.
- ▶ `depStruct`: A character indicating which dependence structure should be used.
- ▶ `distr`: A character indicating which distribution should be used.

- ▶ `formRandom`: A one-sided linear formula object describing the random effects part of the model.

- ▶ `depStruct`: A character indicating which dependence structure should be used.

- ▶ `distr`: A character indicating which distribution should be used.

Some other useful arguments:

- ▶ `timeVar`: A character containing the name of the variable which represents the time in data.

- ▶ `pAR`: The order of the autoregressive process that should be used (if depStruct="ARp").

- ▶ `initialValues`: A named list containing initial parameter values.

The functions return an object of the class SMSN and SMN, respectively, containing a list of elements, and the following methods/functions are available to these classes:

- ▶ `print`
- ▶ `summary`
- ▶ `fitted`
- ▶ `plot`
- ▶ `predict`
- ▶ `residuals`
- ▶ `ranef`
- ▶ `acfresid`
- ▶ `healy.plot`
- ▶ `lr.test`
- ▶ `mahalDist`

Example: sleepstudy data

```
#fitting a (CI)-SMN-LMM, default is distr='norm'
fit0<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
              formRandom = ~Days,groupVar = "Subject",quiet = T)
```

```
fit0
```

```
## Linear mixed models with distribution norm and dependency structure CI
## Call:
## smn.lmm(data = sleepstudy, formFixed = Reaction ~ Days, groupVar = "Subject",
##     formRandom = ~Days, quiet = T)
##
## Fixed:Reaction ~ Days
## Random:~Days
##    Estimated variance (D):
##            (Intercept)       Days
## (Intercept)   560.68002 11.91827
## Days           11.91827 32.52525
##
## Estimated parameters:
##      (Intercept)     Days  sigma2  Dsqrt1 Dsqrt2 Dsqrt3
##        251.4051 10.4673 655.4400 23.6752 0.4059 5.6886
## s.e.     7.2257  1.5541  31.3784 10.0856 1.7514 1.4270
##
## Model selection criteria:
##   logLik     AIC      BIC
## -875.97 1763.94 1783.098
##
## Number of observations: 180
## Number of groups: 18
```

```
#fitting a (CI)-SMSN-LMM, default is distr='sn'
fitskew0<-smsn.lmm(data=sleepstudy,formFixed = Reaction~Days,
            formRandom = ~Days,groupVar = "Subject",quiet = T)
```

```
fitskew0
```

```
## Linear mixed models with distribution sn and dependency structure CI
## Call:
## smsn.lmm(data = sleepstudy, formFixed = Reaction ~ Days, groupVar = "Subject",
##     formRandom = ~Days, quiet = T)
##
## Fixed:Reaction ~ Days
## Random:~Days
##   Estimated variance (D):
##            (Intercept)      Days
## (Intercept)  1432.65701 35.22718
## Days           35.22718 33.76855
##
## Estimated parameters:
##      (Intercept)     Days   sigma2 Dsqrt1 Dsqrt2 Dsqrt3 lambda1 lambda2
##         251.4073 10.4657 652.6598 37.8418 0.8080 5.7546 -4.2917 -0.2477
## s.e.     11.6328  2.1610  32.7878 28.9642 5.2063 1.6311      NA      NA
##
## Model selection criteria:
##    logLik      AIC      BIC
##  -875.354 1766.709 1792.253
##
## Number of observations: 180
## Number of groups: 18
```

Changing the distribution

```
fit1<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 't',
              formRandom = ~Days,groupVar = "Subject",quiet = T)

fitskew1<-smsn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 'st',
                   formRandom = ~Days,groupVar = "Subject",quiet = T)

fit2<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 'sl',
              formRandom = ~Days,groupVar = "Subject",quiet = T)

fitskew2<-smsn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 'ssl',
                   formRandom = ~Days,groupVar = "Subject",quiet = T)

fit3<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 'cn',
              formRandom = ~Days,groupVar = "Subject",quiet = T)

fitskew3<-smsn.lmm(data=sleepstudy,formFixed = Reaction~Days,distr = 'scn',
                   formRandom = ~Days,groupVar = "Subject",quiet = T)
```
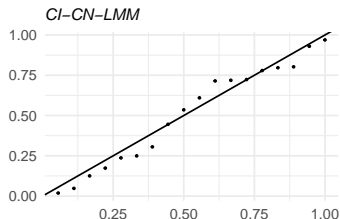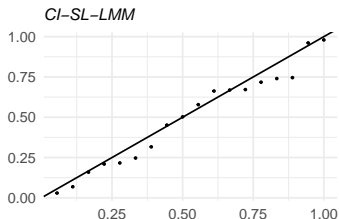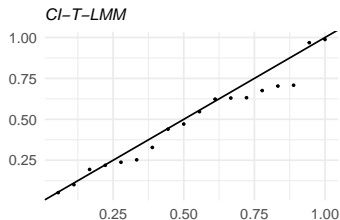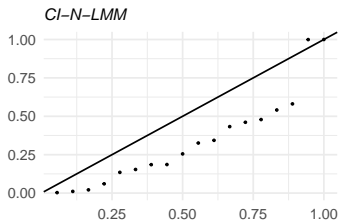
Comparing the fitted models

distr	AIC	BIC
norm	1763.9	1783.1
sn	1766.7	1792.3
t	1737.5	1759.8
st	1739.8	1768.6
sl	1736.2	1758.5
ssl	1738.4	1767.2
cn	1733.7	1759.3
scn	1735.9	1767.8

Assessing the goodness of fit using a Healy-type plot

```
grid.arrange(healy.plot(fit0),
             healy.plot(fit1),
             healy.plot(fit2),
             healy.plot(fit3))
```

LR test for $H_0 : \lambda = 0$ in the CN-LMM

```
lr.test(fitskew3,fit3)
```

```
##
## Model selection criteria:
##              logLik      AIC      BIC
## fitskew3  -857.926 1735.851 1767.781
## fit3      -858.861 1733.722 1759.266
##
##      Likelihood-ratio Test
##
## chi-square statistics =  1.871248
## df =  2
## p-value =  0.392341
##
## The null hypothesis that both models represent the
## data equally well is not rejected at level  0.05
```
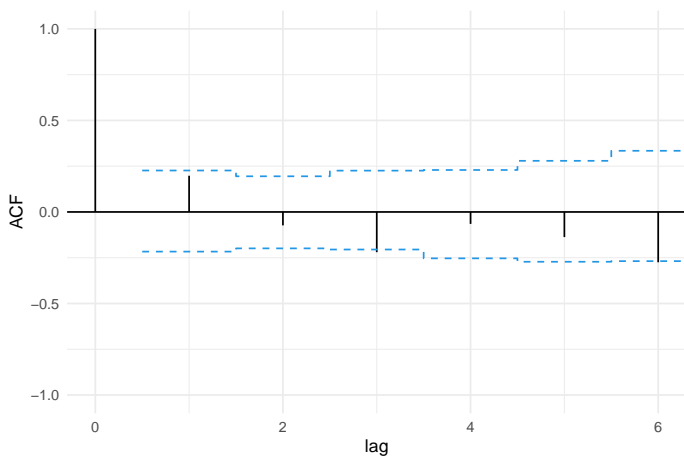
Computing the ACF of the residuals from CN-LMM

```
acfresid(fit3)

##     lag          ACF n.used
## 1     0  1.00000000    180
## 2     1  0.19793878    162
## 3     2 -0.07329748    144
## 4     3 -0.21972223    126
## 5     4 -0.06519806    108
## 6     5 -0.13723727     90
## 7     6 -0.27485055     72
## 8     7 -0.08778128     54
## 9     8  0.19912222     36
## 10    9  0.58707921     18
```

Plotting the ACF (CI-CN-LMM)

```r
plot(acfresid(fit3,calcCI = T,maxLag = 6))
```

Fitting an AR(p)-SMN-LMM

```r
#sl
fit2ar1<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="sl",depStruct = "ARp",pAR=1,quiet=T)

fit2ar2<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="sl",depStruct = "ARp",pAR=2,quiet=T)

#cn
fit3ar1<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="cn",depStruct = "ARp",pAR=1,quiet=T)

fit3ar2<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="cn",depStruct = "ARp",pAR=2,quiet=T)
```

Fitting a DEC-SMN-LMM

```
#sl
fit2dec<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="sl",depStruct = "DEC",quiet=T,
                 timeVar = 'Days')
```

```
#cn
fit3dec<-smn.lmm(data=sleepstudy,formFixed = Reaction~Days,
                 formRandom = ~Days,groupVar = "Subject",
                 distr="cn",depStruct = "DEC",quiet=T,
                 timeVar = 'Days')
```

Since the data are equally spaced and sorted by time, the use of `timeVar`
in here is optional (the function will use the position if `timeVar` is not
provided).

Comparing the fitted models

distr	depStruct	AIC	BIC
sl	CI	1736.2	1758.5
sl	AR1	1716.9	1742.5
sl	AR2	1717.3	1746.0
sl	DEC	1718.2	1746.9
cn	CI	1733.7	1759.3
cn	AR1	1715.5	1744.3
cn	AR2	1714.9	1746.9
cn	DEC	1716.3	1748.2

Healy plot for the AR(p)-SMN-LMM

```
grid.arrange(healy.plot(fit2ar1),
             healy.plot(fit3ar1),
             healy.plot(fit3ar2),ncol=2)
```
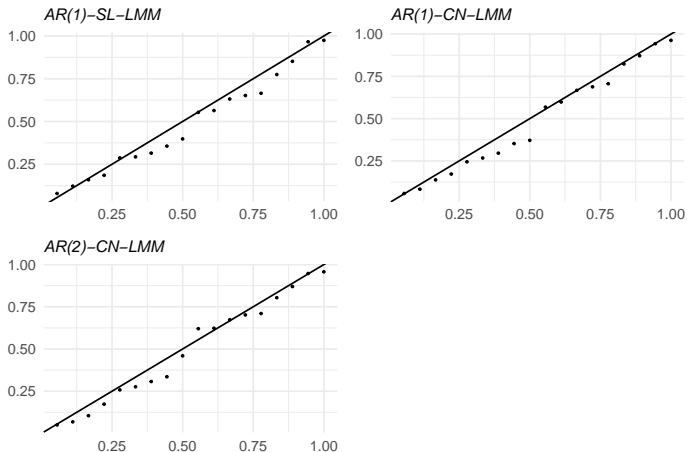
LR test for $H_0 : \phi_2 = 0$ in the CN-LMM

```
lr.test(fit3ar1,fit3ar2)
```

```
##
## Model selection criteria:
##             logLik      AIC        BIC
## fit3ar1 -848.760 1715.520 1744.257
## fit3ar2 -847.463 1714.925 1746.855
##
##      Likelihood-ratio Test
##
## chi-square statistics =  2.594949
## df =  1
## p-value =  0.1072049
##
## The null hypothesis that both models represent the
## data equally well is not rejected at level  0.05
```

ACF plot for AR(1)-SMN-LMM

```
grid.arrange(plot(acfresid(fit2ar1,calcCI = T,maxLag = 6))+
             ggtitle("AR(1)-SL-LMM"),
             plot(acfresid(fit3ar1,calcCI = T,maxLag = 6))+
             ggtitle("AR(1)-CN-LMM"), ncol=2)
```
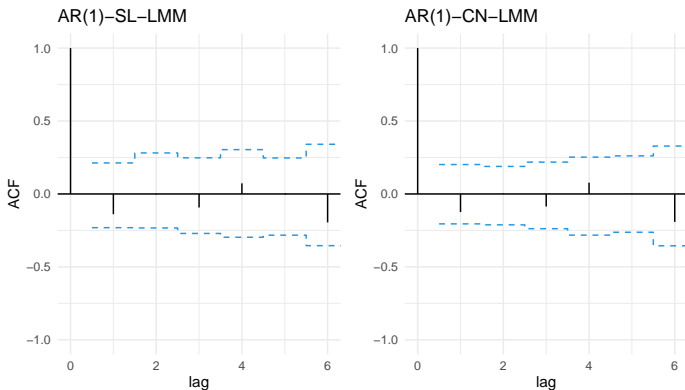
Fitted models

```
cbind(c(fit2ar1$theta,NA),c(fit2ar1$std.error,NA),
      fit3ar1$theta,fit3ar1$std.error) %>%
  knitr::kable(digits = 2,col.names =
                 c('estimate','std.error','estimate','std.error'))
```

|	estimate	std.error	estimate	std.error
(Intercept)	251.55	7.56	252.55	7.02
Days	9.81	1.66	9.99	1.57
sigma2	266.24	63.43	417.84	78.77
phiAR1	0.57	0.14	0.56	0.14
Dsqrt1	14.21	10.15	16.90	11.82
Dsqrt2	2.58	4.16	2.82	4.76
Dsqrt3	2.17	6.45	2.89	6.31
nu1	1.52	NA	0.12	NA
	NA	NA	0.16	NA

Mahalanobis distance for AR(1)-SMN-LMM

```
grid.arrange(plot(mahalDist(fit2ar1),fitobject = fit2ar1,nlabels = 0)+
             ggtitle('AR(1)-SL-LMM'),
             plot(mahalDist(fit3ar1),fitobject = fit3ar1,nlabels = 0)+
             ggtitle('AR(1)-CN-LMM'),
             ncol=2)
```
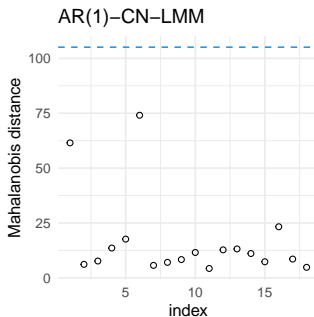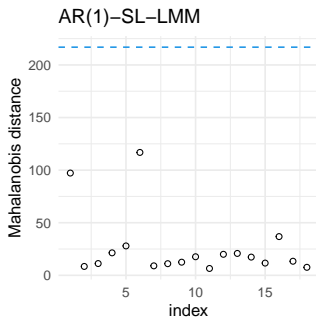
Mahalanobis distance versus $\hat{u}$ for AR(1)-SMN-LMM

```
grid.arrange(qplot(mahalDist(fit2ar1),fit2ar1$uhat)+theme_minimal()+
             ylab("uhat")+xlab("Mahalanobis distance")+
             ggtitle('AR(1)-SL-LMM'),
           qplot(mahalDist(fit3ar1),fit3ar1$uhat)+theme_minimal()+
             ylab("uhat")+xlab("Mahalanobis distance")+
             ggtitle('AR(1)-CN-LMM'),
           ncol=2)
```
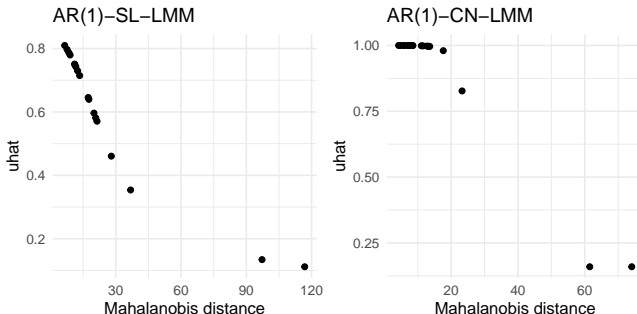
Plot fitted models

```
grid.arrange(plot(fit2ar1,type = "normalized")+ggtitle('AR(1)-SL-LMM'),
             plot(fit3ar1,type = "normalized")+ggtitle('AR(1)-CN-LMM'),
             ncol=2)
```
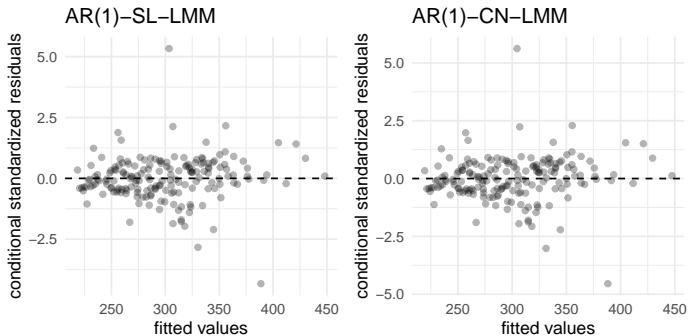
Prediction of future measurements

```
tail(sleepstudy,n=3)
```

```
##     Reaction Days Subject
## 178 343.2199   7    372
## 179 369.1417   8    372
## 180 364.1236   9    372
```

```
predData <- data.frame(Reaction=NA,Days=10:11,Subject="372")
predict(fit2ar1,newData = predData)
```

```
##   groupVar Days    ypred
## 1      372   10 377.0998
## 2      372   11 389.5745
```

```
predict(fit3ar1,newData = predData)
```

```
##   groupVar Days    ypred
## 1      372   10 377.1056
## 2      372   11 389.5576
```

References

▶ Gregory Belenky, Nancy J. Wesensten, David R. Thorne, Maria L. Thomas, Helen C. Sing, Daniel P. Redmond, Michael B. Russo & Thomas J. Balkin (2003) Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research 12, 1–12.

▶ Schumacher, F. L., Matos, L. A., & Lachos, V. H. (2020). Scale mixture of skew-normal linear mixed models with within-subject serial dependence. arXiv preprint arXiv:2002.01040.