

Case Study 1: Clustering the epileptic.qol Dataset

Bayesian mixture model using the mixAK package

```
# install.packages("joineRML")
library(joineRML)
data(epileptic.qol)
# convert days to months
epileptic.qol$time_month <- epileptic.qol$time/30.25
# sort by id and time
epileptic.qol <- epileptic.qol[order(epileptic.qol$id,epileptic.qol$time_month),]

# scaling the clustering variables prior to analysis
epileptic.qol$anxiety_scale <- as.numeric(scale(epileptic.qol$anxiety))
epileptic.qol$depress_scale <- as.numeric(scale(epileptic.qol$depress))
epileptic.qol$aep_scale <- as.numeric(scale(epileptic.qol$aep))
```

Bayesian latent class mixed effect model (mixAK package)

```
# install.packages("mixAK")
library(mixAK)
```

```
## Warning: package 'lme4' was built under R version 4.2.2
```

```
#set.seed(22)
# not run to reduce compiling time
#PED <- NULL
#for (kk in 1:7){ # note that when kk=8, convergence issues appeared
#modK <- GLMM_MCMC(y = epileptic.qol[,c("anxiety_scale", "depress_scale", "aep_scale")],
#  dist = c("gaussian", "gaussian", "gaussian"),
#  id = epileptic.qol[, "id"],
#  z = list(anxiety_scale = epileptic.qol[, c("time_month")],
#    depress_scale = epileptic.qol[, c("time_month")],
#    aep_scale = epileptic.qol[, c("time_month")]),
#  random.intercept = c(TRUE, TRUE, TRUE),
#  prior.b = list(Kmax = kk),
#  nMCMC = c(burn = 1000,
#    keep = 1000,
#    thin = 10,
#    info = 1000),
#  parallel = TRUE)
#  PED <- c(PED, modK$PED[3])
#}
# print the best number of clusters with the smallest PED
#num.clust.mixAK <- which.min(PED); num.clust.mixAK
#-----#
```

```

num.clust.mixAK <- 2 # optimal number of clusters based on PED
# note that even seed is used, each time running the model, the
# clustering results (e.g., cluster proportions and membership) are
# slightly different
set.seed(2022)
ptm <- proc.time()
fit_mixAK <- GLMM_MCMC(y = epileptic.qol[,c("anxiety_scale","depress_scale","aep_scale")],
  dist = c("gaussian","gaussian","gaussian"),
  id = epileptic.qol[, "id"],
  z = list(anxiety_scale = epileptic.qol[, c("time_month")],
    depress_scale = epileptic.qol[, c("time_month")],
    aep_scale = epileptic.qol[, c("time_month")]),
  random.intercept = c(TRUE,TRUE,TRUE),
  prior.b = list(Kmax = num.clust.mixAK),
  nMCMC = c(burn = 1000, keep = 1000, thin = 10, info = 1000),
  parallel = TRUE)

## Parallel MCMC sampling of two chains started on Tue Jun  6 21:52:35 2023.
## Parallel MCMC sampling finished on Tue Jun  6 21:53:06 2023.
##
## Computation of penalized expected deviance started on Tue Jun  6 21:53:06 2023.
## Computation of penalized expected deviance finished on Tue Jun  6 21:53:29 2023.

run.time.mixAK <- as.numeric((proc.time() - ptm)[3])

fit_mixAK <- NMixRelabel(fit_mixAK,type = "stephens",keep.comp.prob=TRUE)

##
## Re-labelling chain number 1
## =====
## MCMC Iteration (simple re-labelling) 1000
## Stephens' re-labelling iteration (number of labelling changes): 1 (0)
##
## Re-labelling chain number 2
## =====
## MCMC Iteration (simple re-labelling) 1000
## Stephens' re-labelling iteration (number of labelling changes): 1 (0)

cluster.mixAK <- apply(fit_mixAK[[1]]$poster.comp.prob,1,which.max);

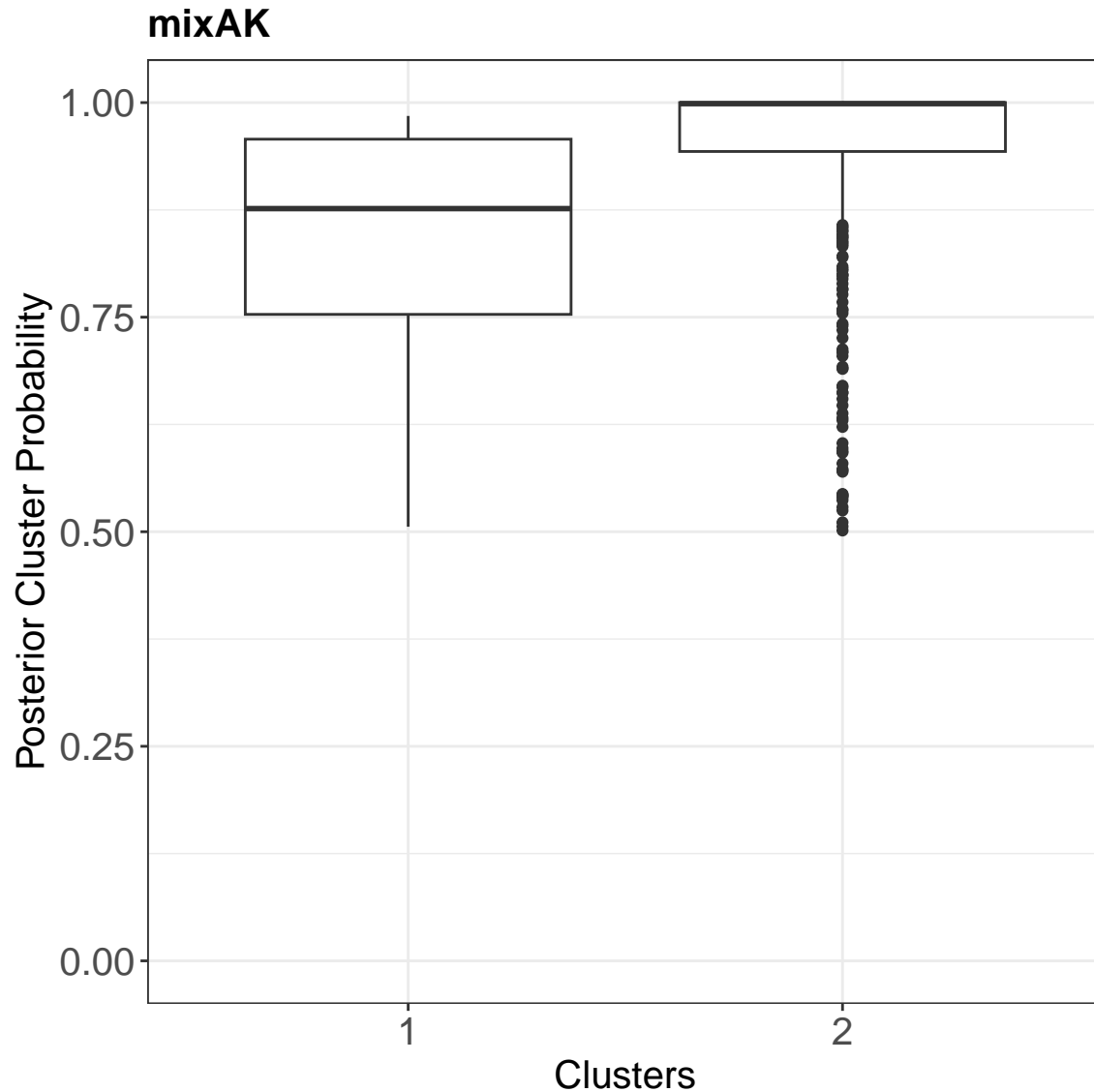
# Keep last observation per id
dnew_uq <- epileptic.qol[!duplicated(epileptic.qol$id, fromLast=TRUE),]
dnew_uq$postprob <- apply(fit_mixAK[[1]]$poster.comp.prob,1,max);
dnew_uq$cluster.mixAK <- cluster.mixAK
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.2
# Posterior cluster probability
bp.mixAK <- ggplot(dnew_uq, aes(x=factor(cluster.mixAK), y=postprob)) +
  geom_boxplot() + ggtitle("mixAK") +
  xlab("Clusters") + ylab("Posterior Cluster Probability") +
  ylim(c(0,1)) +
  theme_bw() +
  theme(legend.position = "none",
    plot.title = element_text(size = 15, face = "bold"),

```

```
axis.text=element_text(size=15),
axis.title=element_text(size=15),
axis.text.x = element_text(angle = 0 ),
strip.text.x = element_text(size = 15, angle = 0),
strip.text.y = element_text(size = 15,face="bold"))
```

bp.mixAK



```
N <- length(unique(epileptic.qol$id))
per <- paste(round(100*table(cluster.mixAK)/N,1),"%",sep="")
dnew_uq$cluster.mixAK <- factor(cluster.mixAK,
                                labels=paste("Cluster ",1:num.clust.mixAK," (",per,")",sep=""))
dat.cluster <- data.frame(dnew_uq$id,dnew_uq$cluster.mixAK)
colnames(dat.cluster) <- c("id","cluster.mixAK")
dnew <- merge(epileptic.qol,dat.cluster,by="id")
```

```
library(cowplot)
# plotting the first feature (anxiety)
```

```

p1.mixAK <- ggplot(data =dnew, aes(x =time_month, y = anxiety,
  color=cluster.mixAK,
  linetype=cluster.mixAK,
  fill=cluster.mixAK))+
  ggtitle("mixAK") +
  geom_smooth(aes(x =time_month, y = anxiety,
    color=cluster.mixAK,
    linetype=cluster.mixAK,
    fill=cluster.mixAK),
    method = "loess", linewidth = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "none",
    plot.title = element_text(size = 15, face = "bold"),
    axis.text=element_text(size=15),
    axis.title=element_text(size=15),
    axis.text.x = element_text(angle = 0 ),
    strip.text.x = element_text(size = 15, angle = 0),
    strip.text.y = element_text(size = 15,face="bold")) +
  guides( fill=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    color=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    linetype=guide_legend(title=NULL,nrow = 1,byrow=TRUE)) +
  xlab("Time (months)") + ylab("anxiety") +
  ylim(c(min(dnew$anxiety,na.rm=TRUE),max(dnew$anxiety,na.rm=TRUE))))+
  scale_color_manual(values=c("green", "black","blue"))+
  scale_fill_manual(values=c("green", "black","blue"))

# plotting the second feature (depress)
p2.mixAK <- ggplot(data =dnew, aes(x =time_month, y = depress,
  color=cluster.mixAK,
  linetype=cluster.mixAK,
  fill=cluster.mixAK))+
  ggtitle("mixAK") +
  geom_smooth(aes(x =time_month, y = depress,
    color=cluster.mixAK,
    linetype=cluster.mixAK,
    fill=cluster.mixAK),
    method = "loess", linewidth = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "none",
    plot.title = element_text(size = 15, face = "bold"),
    axis.text=element_text(size=15),
    axis.title=element_text(size=15),
    axis.text.x = element_text(angle = 0 ),
    strip.text.x = element_text(size = 15, angle = 0),
    strip.text.y = element_text(size = 15,face="bold")) +
  guides(fill=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    color=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    linetype=guide_legend(title=NULL,nrow = 1,byrow=TRUE)) +
  xlab("Time (months)") + ylab("depress") +
  ylim(c(min(dnew$depress,na.rm=TRUE),max(dnew$depress,na.rm=TRUE))))+
  scale_color_manual(values=c("green", "black","blue" ))+
  scale_fill_manual(values=c("green", "black","blue" ))

```

```

# plotting the third feature (aep)
p3.mixAK <- ggplot(data =dnew, aes(x =time_month, y = aep,
  color=cluster.mixAK,
  linetype=cluster.mixAK,
  fill=cluster.mixAK))+
  ggtitle("mixAK") +
  geom_smooth(aes(x =time_month, y = aep,
    color=cluster.mixAK,
    linetype=cluster.mixAK,
    fill=cluster.mixAK),
    method = "loess", linewidth = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "none",
    plot.title = element_text(size = 15, face = "bold"),
    axis.text=element_text(size=15),
    axis.title=element_text(size=15),
    axis.text.x = element_text(angle = 0 ),
    strip.text.x = element_text(size = 15, angle = 0),
    strip.text.y = element_text(size = 15,face="bold")) +
  guides(fill=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    color=guide_legend(title=NULL,nrow = 1,byrow=TRUE),
    linetype=guide_legend(title=NULL,nrow = 1,byrow=TRUE)) +
  xlab("Time (months)") + ylab("aep") +
  ylim(c(min(dnew$aep,na.rm=TRUE),max(dnew$aep,na.rm=TRUE)))+
  scale_color_manual(values=c("green", "black","blue" )) +
  scale_fill_manual(values=c("green", "black","blue" ))

#-----#
# extract a legend
legend.mixAK <- get_legend(ggplot(data =dnew, aes(x =time_month, y = depress,
  color=cluster.mixAK,
  linetype=cluster.mixAK,
  fill=cluster.mixAK))+
  ggtitle("mixAK") +
  geom_smooth(aes(x =time_month, y = depress,
    color=cluster.mixAK,
    linetype=cluster.mixAK,
    fill=cluster.mixAK),
    method = "loess", linewidth = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = c(0.5,0.5),
    legend.text = element_text(size = 12),
    plot.title = element_text(size = 15, face = "bold"),
    axis.text=element_text(size=15),
    axis.title=element_text(size=15),
    axis.text.x = element_text(angle = 0 ),
    strip.text.x = element_text(size = 15, angle = 0),
    strip.text.y = element_text(size = 15,face="bold")) +
  guides(fill=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
    color=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
    linetype=guide_legend(title=NULL,nrow = 2,byrow=TRUE)) +
  xlab("Time (months)") + ylab("depress") +
  scale_color_manual(values=c("green","black","blue" ))+

```

```

    scale_fill_manual(values=c("green","black","blue" ))
)

```

```
## Warning: Removed 53 rows containing non-finite values (`stat_smooth()`).
```

```

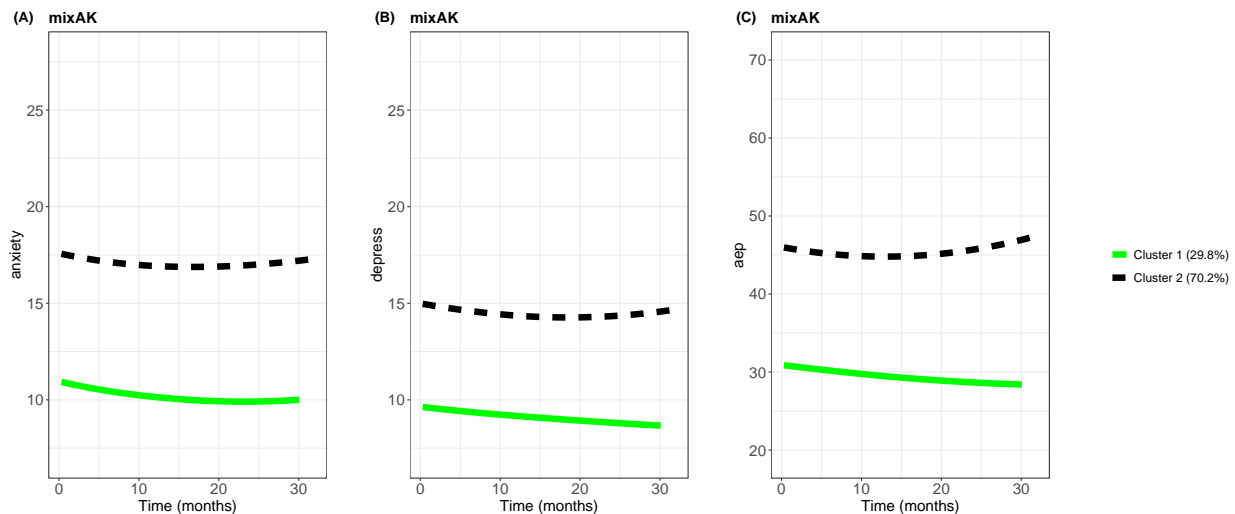
plot_grid(p1.mixAK,NULL,p2.mixAK,NULL,p3.mixAK,NULL,legend.mixAK,
  labels=c("(A)","", "(B)","", "(C)","",""),
  ncol = 7,
  rel_widths = c(1,0.1,1,0.1,1,0.1,0.5))

```

```
## Warning: Removed 57 rows containing non-finite values (`stat_smooth()`).
```

```
## Removed 53 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 93 rows containing non-finite values (`stat_smooth()`).
```



```

dnew_uq$with.time.month <- dnew_uq$with.time/30.25
fit <- survfit(Surv(with.time.month, with.status2) ~ cluster.mixAK,
  data = dnew_uq)
res.cox <- coxph(Surv(with.time.month, with.status2) ~ cluster.mixAK,
  weights=postprob, data = dnew_uq)
pvalue <- ifelse(summary(res.cox)$sctest[3] >= 0.0001,
  summary(res.cox)$sctest[3], '<0.0001')

```

```
# Visualize with survminer
```

```
library(survminer)
```

```
## Warning: package 'ggpubr' was built under R version 4.2.2
```

```

library(survival)
names(fit$strata) <- paste("Cluster ",1:num.clust.mixAK," (",per,")",sep="")
gp_survival.mixAK <- ggsurvplot(fit, data = dnew_uq, title="mixAK",
  risk.table = TRUE,
  risk.table.y.text.col = TRUE,
  pval = pvalue,
  legend = "bottom", # conf.int = TRUE,
  xlab = "Time (months)",
  legend.title="Clusters",
  ggtheme = theme_bw() +

```

```

      theme(legend.position = "none", legend.title = element_blank(),
            plot.title = element_text(size = 15, face = "bold"),
            axis.text = element_text(size = 15),
            axis.title = element_text(size = 15),
            strip.text.x = element_text(size = 15),
            strip.text.y = element_text(size = 15)))
gp_survival.mixAK$plot <- gp_survival.mixAK$plot +
  guides(fill = guide_legend(title = NULL, nrow = 1),
         color = guide_legend(title = NULL, nrow = 1),
         linetype = guide_legend(title = NULL, nrow = 1)) +
  scale_color_manual(values = c("green", "black", "blue")) +
  scale_fill_manual(values = c("green", "black", "blue"))
gp_survival.mixAK$plot

```

