USING MIXED HIDDEN MARKOV MODELS TO CAPTURE DYNAMIC EFFECTS OF AMBIENT NOISE ON INDIVIDUAL STRESS AT WORKPLACE

Final project report as part of course fulfillment for
STAT 675 - Statistical Computing (Spring 2017)
University of Arizona
Instructor: Chengcheng Hu

May 10, 2017

Karthik Srinivasan

1 INTRODUCTION

A growing literature demonstrates the impact of the built environment on human health and wellbeing. Environmental factors such as ambient noise, temperature, air quality and humidity have been shown to have effects on an individual's emotional and physiological state of being [1, 2]. Ambient noise in work environment has a direct effect on cognitive ability, productivity and psychological stress [3]. Survey questionnaires have been the most common method of assessing the impact of noise on psychological stress. Recently, instruments such as heart rate variability are being explored to model effects of ambient noise on physiological stress [4]. Due to wearable sensor technology such as portable heart rate monitors, it has become possible to investigate effects of environmental factors such as ambient noise on instantaneous stress response [5]. Though effects of noise types and high noise on physiological stress have been investigated [4], the effect of noise in triggering switching between stressful states (such as low stress, high stress) has not yet been studied. In this project, we will investigate the dynamic effects of ambient noise on human stress using models that use statistical computing methods.

We use the data from a field experiment conducted as part of a multidisciplinary research program called Wellbuilt for Wellbeing (WB2), supported by the General Services Administration (GSA). The overarching goal is to study the impact of the workplace environment on individual wellbeing. The experimental setup consisted of participants wearing two sensors

for three days while carrying out their day-to-day activities: (a) A heart rate monitor and (b) A personal environment quality sensor-based device. Two hundred and thirty-one participants across multiple locations participated in the experiment during July 2015 through November 2016. The data for this study was hierarchical, where participants are the secondary level of data abstraction. Instantaneous physiological stress response of individuals is measured using heart rate variability (HRV) scores, observable measures that is inversely proportional to the a person's stress [6].

In this project, we address the problem of modeling dynamic effects of ambient noise on stress using Mixed Hidden Markov Models (MHMM). MHMM is a special case of the family of generative Dynamic Bayesian networks called Hidden Markov models (HMM). HMMs describe the relationship between two stochastic processes: an observed process and an underlying hidden or unobserved process. The hidden process is assumed to follow a Markov chain, and the observed data are modeled as independent conditional on the sequence of hidden states [7]. Mixed HMM is useful for modeling dynamic time series of multilevel or panel data, as observed in our case.

We posit that MHMM is a suitable model for our study due to the following reasons:

1. Heart rate variability (HRV) is a continuous variable that is a manifestation of instantaneous stress response in humans. But, discrete physiological stress states such as low stress and high stress are latent and cannot be be directly determined using HRV or other observable measures.

2. HRV for a person has a time-series pattern and its variability can reflect switching of latent stress states [6].

3. For given panel data, it is important to model multi-level effects (fixed and random effects within and across participants) and hence the single HMM is insufficient.

Concepts taught in STAT 675 class in Spring 2017 are applied to implement an MHMM that is designed to capture the dynamic effects of ambient noise on individual stress in a workplace. The data for this project was generated in a field experiment as part of on ongoing study called Wellbuilt for Wellbeing (WB2), details about which are given in the Data section.

In the following section, we will briefly describe the literature on MHMM followed by some example application in other studies that have implemented MHMM. The methods section will describe the design of the MHMM in this project, followed by summary of analysis and results completed until now. Current work and limitations will be listed in discussions section, with focus on statistical computing methodology applied in this paper. The conclusion will summarize the learnings from the project, from a functional as well as technical perspective.

2 Related work

The class of Hidden Markov Models (HMM) provides several different strategies for dealing with populations of time series. It has latent or unobservable states and the observable or

state-dependent processes as a probabilistic model [8]. HMM were developed to analyze time-series data and Mixed HMM (MHMM) is an methodological extension to simultaneously model multiple time-series or panel data [9]. The Mixed HMM for incorporating random effects are useful for modeling grouped data. The parameters for regular HMM can be estimated using Baum Welch Expectation maximization (EM) method, but MHMM require different parameter estimation approaches due to presence random effects in grouped data. Altman's seminal paper in 2007 [7] lists three main approaches for fitting MHMM — Monte Carlo EM using Gibbs sampling, direct maximization of likelihood using numerical integration and simulated maximum likelihood. MHMM has been proposed for research applications thereafter. Inputs of covariates into the model can be introduced either in modeling the state dependent probability (or emission probability function), or modeling the elements of the transition probability matrix. In one application, users' check-in into location-based social network was modeled as an MHMM with spatial-temporal features as covariates in the state dependent probability [10]. The researchers use MCMC Metropolis Hastings sampling with a multivariate normal distribution as prior for the parameter set. In other applications, researchers model spatial-temporal trajectory into ecological outcomes such as feeding behavior of grey mouse lemurs [11], foraging behavior of woodpeckers [12] and diving behavior of whales [13]. The underlying approach for the MHMM design for these applications is numerical optimization of likelihood, proposed by Zuccini and MacDonald [14], with improvements in computational efficiency using approximations.

We closely follow the method used by McKellar et al. [12] in the woodpecker study, and fit MHMM using the basic numerical optimization approach in this project. The R programming for the MHMM is inspired from the forward algorithm for numerical optimization presented in Zuccini and MacDonald's book [14].

3 Method

An observed outcome vector $Y_t,\ t \in 1,...,n$ follows a HMM if (a) The hidden states $Z_t,\ t \in 1,...,n$, follows a Markov chain, and (b) Given Z_t, Y_t is independent of $Y_1,...,Y_{t-1},Y_{t+1},...,Y_n$ and $Z_1,...,Z_{t-1},Z_{t+1},...,Z_n$. The HMM is fully specified by the initial and transition probabilities of the hidden Markov chain and by the distribution of Y_t given Z_t. Typically, the latter would be chosen from a family of distributions with mean depending on Z_t [7]. The likelihood of an HMM is given by:

$$L_T = Pr(X^{(T)} = x^T) = \delta P(x_1)\Gamma P(X_2)\Gamma P(X_3)...\Gamma P(X_T)1' \tag{3.1}$$

where δ is the initial distribution, $Gamma$ is the m x m transition probability matrix (t.p.m) and P(x) is the m x m diagonal matrix with i^{th} diagonal element the state-dependent probability or density $p_i(x)$ of i^{th} latent state.

Zuccini and MacDonald [14] specify the following routine for estimation of HMM parameters related to transition probabilities and emission probabilities:

1. For a stationary Markov chain, represent the equation as $L_T = \alpha_T 1'$ where $\alpha_t = \alpha_{t-1}\Gamma P(x_t)$ and $\alpha_0 = \delta.\alpha_t$ is vector of forward probabilities.

2. Handling numerical underflow of α_t using scaling of likelihood computation. That is, define $\phi_t = \alpha_t / w_t$ such that $w_t = \sum_i \alpha_t(i) = \alpha_t 1'$. We get $log(L_T) = \sum_{t=1}^{T} log(w_t / w_{t-1})$. This procedure avoids underflow.

3. Reparametrization to avoid constraints. Transform natural parameters that are constrained to a limited range to working parameters with range $(-\inf, \inf)$ using functions such as logit functions (parameters from range [0,1] to $(-\inf, \inf)$, log function for positive real numbers, etc. These parameters are re-transformed to original scale post maximization of likelihood.

4. Avoid convergence to local maxima by repeating the process for range of starting values and checking the standard error for each estimated parameter.

McKeller et al. 2015 [12] implement an MHMM to model woodpecker foraging behavioral states as follows:

1. They consider a multi-level 2-state HMM for analyzing observed step lengths and turning angles of woodpeckers associated with different territories and different foraging sessions within territories. The two hidden or latent behavioral states of woodpecker considered are foraging and resting.

2. The state dependent probability function is bivariate distribution conditioned on the latent state. They consider turn angle to have a Von mises distribution and step length to have a zero-inflated gamma distribution and take their product as joint density of the bivariate distribution.

3. The likelihood of MHMM is the product of likelihood across each session and territory.

4. Covariates such as number of birds in group and the type of pine stems in each territory are hypothesized to effect switching between two latent states. The elements of the transition probability matrix or state transition probabilities are therefore modeled as a logit function of linear combination of the covariates.

5. Territory-specific random effects are introduced into the model for state transition probabilities by replacing the intercept with random variables $\epsilon^{(c)}$ that take different values for different territories c. They additionally enforce an approximation of discrete support for $\epsilon^{(c)}$ using a meta-parameter K that is determined using enumeration and comparing AIC of model fits.

In this report, we take a simplified version of McKeller et al.'s approach. We use the univariate observed outcome as HRV and model the state transition probabilities as logit function of noise-level and other covariates. We do not include random effects as mentioned in step (5)

of McKeller et al.'s approach at this stage.

The MHMM in this study is represented in Figure 3.1

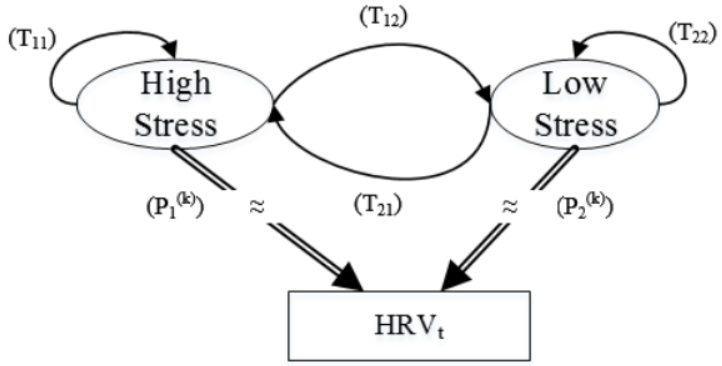

Figure 3.1: Representation of Mixed Hidden Markov Model. T_{ii} are transition probabilities that are modeled as logit functions of covariates and $P_i^{(k)} \forall k \in K$ are the state dependent probabilities for grouped data of K participants

We hypothesize that the transition between low and high stressful states of individuals at workplace is effected by ambient sound levels. In addition, the covariates — *Age*, *Gender*, *noise sensitivity* and *Time of day* are introduced as control variables in the logit model as shown below:

$$\begin{aligned} logit(T_{ii}) = &\beta_{0,i} + \beta_{1,i} Sound_t + \beta_{2,i} Age + \beta_{3,i} Gender + \beta_{4,i} NoiseSensitivity + \\ &\beta_{5,i} TimeOfDay \end{aligned} \tag{3.2}$$

where, the transition probability matrix is given as:

$$\Gamma = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \tag{3.3}$$

Note that $T_{12} = 1 - T_{11}$ and $T_{21} = 1 - T_{22}$. Hence logit functions of only T_{11} and T_{22} are fit as shown in Equation 3.2.

The m environment-based input variables are inputs to the logit model that estimates the parameters of the transition probability function between two states i =1,2 in the above equation. The likelihood function to the parameter estimates of the mixed Hidden Markov Model is as follows:

$$L_T = \prod_{i=1}^{i=K} L_i = \prod_{i=1}^{i=K} \delta P_i(x_1) \Gamma P_i(X_2) \Gamma P_i(X_3) ... \Gamma P_i(X_T) 1' \tag{3.4}$$

Where L_i is the likelihood function for a single participant case as seen in Equation 3.1. The likelihood for MHMM is just a product of individual likelihoods. For a two state HMM with K

participants, considering that the state dependent probabilities of the observable outcomes conditioned on the latent states are normally distributed, we therefore have to estimate $K(2+2M)$ parameters in a basic model. Here M is the total number of coefficients in the logit function for the transition probability, including the intercept term. The factor 2 is for estimating the mean and standard deviation of the normal distributions. M is multiplied by two owing to the two states in the HMM. A pooling strategy is adopted where the random-effects determines the state dependent probability but the transition probability remains constant or is pooled across participants; hence reducing the parameter estimation to a count of $2K+2M$. It is important to note that we assume participant fixed effects in this study for model simplicity; as compared to McKeller et al.'s[12] special treatment of intercepts of transition probability models (See Step 5 of their study summary above). The intention is to check the performance of MHMM parameter estimation using the numerical optimization approach using the simpler model, before adding additional complexities.

4 Data

We measure noise level using personal environment quality sensor-based devices worn by all participants. The outcome variables is SDNN (Standard deviations of NN interval),a heart rate variability score measured at 5 minute intervals. Noise sensitivity is a personality trait variable constructed from series of self-reported survey questions completed by each participant during the beginning of the experiment. Time of day is a discrete variable with three levels — Morning, Afternoon, and Evening included in the model to capture circadian variability of HRV. After data integration, preprocessing and cleaning, the complete dataset contained approximately 40,000 observations or 200,000 minutes of heart rate monitor and environment quality data streams for 231 participants. We fit MHMM model for first 5000 observations in this project. The objective is to evaluate performance of MHMM estimation methods with different numerical optimizers and approximations for this sample. The most efficient and robust method can then be used for fitting the complete data. The data and R code for this project can be accessed in this github page

5 Analysis and Results

We follow the MHMM model parameter estimation process proposed by Zuccini and MacDonald [14] (Refer steps in previous section).

The R program in Appendix has following functions:

1. **Transformation of natural to working parameters:** Probability-based parameters such as delta are transformed using logit function from [0,1] to $(-\inf, \inf)$ range. Standard deviation is transformed using log transformation and other parameters such as mean and model efficients are not transformed as they are already unconstrained.

2. **Inverse transformation of working parameters:** Working parameters are inverse transformed after the optimization routine.

3. **Likelihood:** The likelihood function as given in 3.4 is implemented here.

4. **Maximum likelihood estimation using optimization:** Optimization procedure such as Nelder Mead in *nlm*, BFGS and Conjugate gradient in *optim* packages in R are used for optimization of likelihood for all parameters in model. Control parameters of these methods are varied and model fit is compared for time to convergence. AIC and BIC values for final model fit are recorded

The *nlm* package has an optim to print the interim model fit and the gradient for each iteration. The methods in optim package ran for a really long time (More than 3 days for given data). Reducing tolerance of step and gradient in the *nlm* routine either resulted in poor fit or failed at the first iteration itself. Convergence is highly affected by initialization values. As dataset size is increased, the computation time increases exponentially. All versions of program were run using an Intel Core i7 2.5 GHz processor and 16GB RAM.

Parameter estimates for state dependent probability distributions are given in Tables 5.1, 5.2. The means of the distributions. Based on our previous experience on comparing variance in HRV across participants, we feel that the model may not have converged to a global maxima for most of the participant's grouped data. However, the standard deviation how some amount of dispersion suggesting that a pooling strategy with two latent states of stress for all participants may not be optimal. That is, some participants may have more than two latent states, in which case, the MHMM is forced to fit a normal distribution to a multimodal distribution, which is suboptimal. Possible methodological improvements based on the results are given in the discussions section. The logit functions for the transition probabilities are:

$$\begin{aligned} logit(T_{11}) =& 0.4034 + 0.0275 * Sound_t - 0.0353 * Age + 0.8574 * NoiseSensitivity + \\ & 0.2441 * I(TimeOfDay = Afternoon) + \\ & 0.1854 * I(TimeOfDay = Evening) + 0.2178 * I(Gender = Female) \end{aligned}$$

$$\begin{aligned} logit(T_{22}) =& 1.0695 + 0.0083 * Sound_t - 0.0667 * Age + 1.029 * NoiseSensitivity - \\ & 0.0309 * I(TimeOfDay = Afternoon) + \\ & 0.1112 * I(TimeOfDay = Evening) - 0.0039 * I(Gender = Female) \end{aligned}$$

From these equations, one can infer that participants continue to remain in High stress with higher sound by a factor of 0.0275 ms/dbA. Females transition less from High stress to low stress than they transition from low to high stress conditions as the coefficient in the $logit(T_{22})$ model is negative. With increase in Age, dynamic switching between stressful states increases, as both coefficients for $logit(T_{11})$ and $logit(T_{22})$ are negative. During afternoons and evenings, switching between stress states is lesser than morning. Since we have not scaled the variables, one cannot compare the coefficients between different inputs in this study.

Table 5.1: Mean values of state dependent probability normal distribution across participants

Participant	High stress	Low Stress	Participant	High stress	Low Stress
P1	47.81	88.13	P24	47.7	88.17
P2	47.8	88.18	P25	47.7	88.17
P3	47.76	88.17	P26	47.88	88.17
P4	47.81	88.17	P27	47.87	88.17
P5	47.81	88.14	P28	47.8	88.17
P6	47.83	88.13	P29	47.8	88.17
P7	47.8	88.17	P30	47.78	88.17
P8	47.74	88.17	P31	47.76	88.17
P9	47.87	88.15	P32	47.87	88.17
P10	47.85	88.17	P33	47.78	88.17
P11	47.77	88.17	P34	47.88	88.14
P12	47.8	88.17	P35	47.8	88.2
P13	47.8	88.15	P36	47.92	88.15
P14	47.73	88.17	P37	47.73	88.17
P15	47.81	88.16	P38	47.79	88.17
P16	47.82	88.13	P39	47.72	88.17
P17	47.82	88.17	P40	47.82	88.13
P18	47.75	88.17	P41	47.83	88.17
P19	47.71	88.17	P42	47.8	88.16
P20	47.84	88.18	P43	47.8	88.21
P21	47.7	88.17	P44	47.81	88.17
P22	47.8	88.17	P45	47.8	88.17
P23	47.79	88.17	P46	47.8	88.19

Table 5.2: Standard deviations of state dependent probability normal distribution across participants

Participant	High stress	Low Stress	Participant	High stress	Low Stress
P1	17.89	21.58	P24	14.6	25.26
P2	25.39	37.1	P25	18.83	25.48
P3	11.03	25.15	P26	13.56	15.56
P4	13.44	26.95	P27	6.53	22.03
P5	16.85	24.12	P28	24.55	25.15
P6	16.58	19.27	P29	14.27	25.39
P7	24.28	25.43	P30	13.36	25.51
P8	17.78	22.99	P31	23.82	25.5
P9	12.96	23.28	P32	11.61	21.19
P10	15.53	25.74	P33	15.09	25.38
P11	20.94	25.5	P34	13.72	20.03
P12	23.16	25.53	P35	25.5	35.67
P13	21.25	19.42	P36	12.45	20.51
P14	25.86	25.51	P37	12.91	25.37
P15	14.7	22.96	P38	11.99	25.4
P16	23.67	22.42	P39	23.79	25.49
P17	8.89	21.73	P40	16.95	18.4
P18	18.96	25.27	P41	13.08	25.29
P19	15.19	24.53	P42	24.72	20.36
P20	11.9	36.59	P43	25.65	20.65
P21	12.67	25.67	P44	20.13	25.47
P22	25.35	29.58	P45	24.82	23.29
P23	15.64	25.5	P46	25.47	43.91

6 Discussions

In this study, we study the dynamic effects of ambient noise on individual stress at workplace using mixed hidden markov models (MHMM). The parameter estimation is challenging and involves using packages for unconstrained optimization that were introduced to us in STAT 675 class. The basic idea for the model was given in McKellar et al. 2015 [12] and the R program for numerical optimization for basic HMM was adopted from the popular book "HMM with R" by MacDonald and Zuccini[14]. We implemented the basic version for of the mixed hidden models, approximating the state dependent probabilities for each participant as normal distributions, and the transition probabilities between two latent states as a logit function that has a linear dependency on input covariates. We have neither conducted any diagnostics nor done any test data evaluation for our models, as the objective is to try implementation of MHMM for the given problem in this study.

The numerical optimization approach for MHMM models takes a lot of time to converge. It requires newer methods of efficient and robust estimation for improved and generic applications into similar problems. The means of the state dependent probability normal distributions do not vary much, suggesting that there is a strong effect of the initial values. Standard errors can be derived using bootstrap approach if the model can be fit faster and we can therefore use resampling methods.

Alternatives to the current MHMM approach are:

- Propose a better parameter estimation method that is more robust and efficient
- Fit a single HMM model for the entire data after accounting for inter-personal variability. One very naive method can be introduce indicators for each participant into the logit model.
- Fit ensemble of HMM models for each participant grouped data and characterize the difference between model fits using inter-personal attributes. However, a pooled estimate of dynamic effects may not be directly available using this method.
- Fit other models such as generalized Dynamic Bayesian networks using Bayesian or frequentist parameter estimation approaches.

7 Conclusion

The focus of this project is to formulate a mixed MHMM that captures the dynamic relationship of ambient noise and individual stress at workplace. We implement the basic MHMM model using numerical optimization of likelihood method, analyze the problems for parameter estimation and interpret model fit. We can infer some relationships from our current model. Moving forward, we wish to improve over the parameter estimation method, as currently, the program takes a lot of time to converge and also faces the risk of converging to a local maxima. Current work is directed towards improving parameter estimation methods

for MHMM and planning diagnostics and test evaluation for improved models. Our ultimate goal is to develop an optimal model that determines dynamic effects of ambient noise-levels on stress.

REFERENCES

[1] P. MacNaughton, J. Spengler, J. Vallarino, S. Santanam, U. Satish, and J. Allen, "Environmental perceptions and health before and after relocation to a green building," *Building and Environment*, vol. 104, pp. 138–144, 2016.

[2] J. F. Thayer, B. Verkuil, J. F. Brosschot, K. Kampschroer, A. West, C. Sterling, I. C. Christie, D. R. Abernethy, J. J. Sollers, G. Cizza, A. H. Marques, and E. M. Sternberg, "Effects of the physical work environment on physiological measures of stress," *European Journal of Cardiovascular Prevention & Rehabilitation*, vol. 17, no. 4, pp. 431–439, 2010.

[3] J. Heerwagen and A. Zagreus, "Title: The human factors of sustainable building design: post occupancy evaluation of the Philip Merrill Environmental Center," 2005.

[4] C. Sun Sim, J. Hyun Sung, S. Hyeon Cheon, J. Myung Lee, J. Won Lee, and J. Lee, "The Effects of Different Noise Types on Heart Rate Variability in Men," *Yonsei Med J http*, vol. 56, no. 1, 2015.

[5] K. Srinivasan, D. Herzl, M. Lunden, S. Andrews, N. Goebel, R. Herzl, M. R. Mehl, B. Gilligan, J. Heerwagen, K. Kampschroer, K. Canada, F. Currim, S. Ram, C. Lindberg, E. Sternberg, P. Skeath, B. Najafi, J. Razjouyan, and H.-K. Lee, "A Regularization Approach for Identifying Cumulative Lagged Effects in Smart Health Applications," *Proceedings of the 6th International Conference on Digital Health Conference - DH '17*, 2017.

[6] B. Xhyheri, O. Manfrini, M. Mazzolini, C. Pizzi, and R. Bugiardini, "Heart Rate Variability Today," *Progress in Cardiovascular Diseases*, vol. 55, no. 3, pp. 321–331, 2012.

[7] R. M. Altman, "Mixed Hidden Markov Models," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 201–210, 2007.

[8] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[9] F. Bartolucci, A. Farcomeni, and F. Pennoni, "Latent Markov models : a review of a general framework for the analysis of longitudinal data with covariates," *Munich Personal RePEc Archive*, no. 39023, pp. 1–37, 2012.

[10] J. Ye, Z. Zhu, and H. Cheng, "What ' s Your Next Move : User Activity Prediction in Location-based Social Networks," *Sdm*, pp. 171–179, 2013.

[11] S. Schliehe-Diecks, P. M. Kappeler, and R. Langrock, "On the application of mixed hidden Markov models to multiple behavioural time series," *Interface Focus*, vol. 2, no. 2, pp. 180–189, 2012.

[12] A. E. McKellar, R. Langrock, J. R. Walters, and D. C. Kesler, "Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird," *Behavioral Ecology*, vol. 26, no. 1, pp. 148–157, 2015.

[13] S. L. DeRuiter, R. Langrock, T. Skirbutas, J. A. Goldbogen, J. Chalambokidis, A. S. Friedlaender, and B. L. Southall, "A multivariate mixed hidden Markov model to analyze blue whale diving behaviour during controlled sound exposures," *arXiv preprint*, vol. 1602, no. 06570, pp. 1–26, 2016.

[14] W. Zucchini and I. L. MacDonald, "Hidden Markov Models for Time Series: An Introduction using R," *South African Actuarial Journal*, vol. 10, no. 1, p. 265, 2009.

8 APPENDIX

8.1 APPENDIX-A: R CODE FOR MHMM IMPLEMENTATION

```
library(boot)
library(CircStats)
library(Hmisc)

data_set_all <- read.csv("train_HMM_Sound.csv")
#data_set <- data_set_all
data_set <- data_set_all[1:5000,]
data_set$P_ID <- factor(data_set$P_ID)

data_in <- data_set[,c("P_ID", "Soundm","Age","Noise_sensitivity","ToDAfternoon",
"ToDEvening", "GenderFemale", "SDNN")]

data_in_split <- split(data_in,data_in$P_ID)

## K method: K is number of participants and that many normal distributions
#  to be estimated
## VARIABLE NAMES:
## delta = Markov chain initial state distribution
## intercept = Intercept values of transition probability stochastic function
## betaS = Coefficients of covariates in the logit regression for
#  transition probabilities
## mu = Mean of the Normal desnity assumed for state distribution probability
#  function or emission probability
## sigma = Std. deviation of the Normal desnity assumed for state distribution
#  probability function or emission probability
```

```
## Natural parameters to working parameters
pn2pw_mix <- function(delta,intercept,betaS,mu,sigma)
{
  tdelta <- logit(delta)     ## Transfrom from [0,1] to (-inf,inf)
  tsigma <- log(sigma)        ## Transfrom from [0,inf) to (-inf,inf)
  ## No need of transformation for mu, intercept, betaS as they
  #  are already unconstrained
  parvect <- c(tdelta,intercept,betaS,mu,tsigma)
  return(parvect)
}

## Working parameters back to natural parameters
pw2pn_mix <- function(parvect,K)
{
  delta <- inv.logit(parvect[1])  ##Not transformed
  intercept <- matrix(parvect[2:3],nrow=2)
  betaS <- matrix(parvect[4:15],nrow=2)
  mu <- matrix(parvect[16:(2*K+15)],nrow=2)
  sigma <- matrix(exp(parvect[(2*K+16):(15+4*K)]),nrow=2)
  return(list(delta=delta,betaS=betaS,intercept=intercept,mu=mu,sigma=sigma))
}

## Initialize the parameters
mu01<-c(c((min(data_in$SDNN)+4*mean(data_in$SDNN))/5,
          (max(data_in$SDNN)+4*mean(data_in$SDNN))/5))
sigma01 <- c(sd(data_in$SDNN),sd(data_in$SDNN))
delta0 <- 0.1
betaS0<-matrix(rep(0.1,12),nrow=2)
intercept0 <- matrix(c(0.2,0.9),nrow=2)

K <- length(data_in_split)

mu0 <- matrix(rep(mu01,K),nrow=2)
sigma0 <- matrix(rep(sigma01,K),nrow=2)

## Parameter vector, check if we get the initial values back
# after inverse transformation
parvect <- pn2pw_mix(delta0,intercept0,betaS0,mu0,sigma0)
pw2pn_mix(parvect,K)

mllk_mix <-function(parvect,X)
{
  allprobs <- list()
  K <- length(X)
```

```
pn<-pw2pn_mix(parvect,K)
llk <- rep(NA,K)
for(part in 1:K)
{
  allprobs[[part]] <- matrix(NA,nrow=nrow(X[[part]]),ncol=2)

  for (j in 1:2)  ##j is hidden state
  {
    allprobs[[part]][,j] <- dnorm(X[[part]][,"SDNN"],mean=pn$mu[j,part],
    sd=pn$sigma[j,part])  ## Conditional probability for each state
  }

  allprobs[[part]] <- ifelse(!is.na(allprobs[[part]]),allprobs[[part]],1)
  lscale <- 0

  n <- nrow(X[[part]])

  cov1<-X[[part]][,"Soundm"]
  cov2<-X[[part]][,"Age"]
  cov3<-X[[part]][,"Noise_sensitivity"]
  cov4<-X[[part]][,"ToDAfternoon"]
  cov5<-X[[part]][,"ToDEvening"]
  cov6<-X[[part]][,"GenderFemale"]

  for (i in 1:n)
  {
    if (i==1){ foo<-c(pn$delta,1-pn$delta) } ## Initialize foo

    gamma<-matrix(NA,nrow=2,ncol=2)  ## Initialize gamma matrics
    for (j in 1:2)
    {
      gamma[j,j] <- inv.logit( pn$intercept[j]   + pn$betaS[j,1]*cov1[i] +
                  pn$betaS[j,2]*cov2[i] + pn$betaS[j,3]*cov3[i] +
              pn$betaS[j,4]*cov4[i]  + pn$betaS[j,5]*cov5[i] +
        pn$betaS[j,6]*cov6[i] )
      gamma[j,-j] <- 1 - gamma[j,j]
    }

    foo<-foo %*% gamma*allprobs[[part]][i,]
    sumfoo <- sum(foo)
    lscale<-lscale+log(sumfoo)
    foo<-foo/sumfoo
  }
  llk[part] <- - lscale
}
```

```
  mllk <- sum(llk)
  return (mllk)
}

mle_mix <- function(data_in,X,delta0,intercept0,betaS0,mu0,sigma0)
{
  parvect0<-pn2pw_mix(delta0,intercept0,betaS0,mu0,sigma0)

  ## NUMERICAL OPTIMIZATION of parameters using packages 'nlm' and methods
  # in 'optim' package.
  # We tried different methods in optim, 'BFGS' is shown below.
  # The control parameters for nlm and optim methods were also tweaked
  # to check speed of estimation

  #mod <- nlm(mllk_mix,X=X,parvect0,print.level=2,iterlim=200,stepmax=5,
  hessian=F,ndigit=4,steptol=1e-4,gradtol=1e-4)
  #mod <- nlm(mllk_mix,X=X,parvect0,print.level=2,iterlim=1500,
  # stepmax=5,hessian=F,ndigit=4)
  mod <- optim(parvect0,mllk_mix,X=X,method="BFGS",hessian=F)
  pn<-pw2pn_mix(mod$estimate,K=length(X))
  mllk <- mod$minimum
  np <- length(parvect0)
  AIC <- 2*(mllk+np)
  n <- nrow(data_in)
  BIC <- 2*mllk+np*log(n)
  list(mu=pn$mu,sigma= pn$sigma, delta = pn$delta, intercept = pn$intercept,
  betaS=pn$betaS,mllk=mllk,AIC=AIC,BIC=BIC)
}

## Repeat below code for each optimization technique
# (i.e. using nlm, optim(CG), optim(BFGS), etc)
system.time(
  HMM_mix1 <- mle_mix(data_in,data_in_split,delta0,intercept0,betaS0,mu0,sigma0)
)

HMM_mix1
```