02424, 2021: Assignment 3

This is the third of three mandatory assignments for the course 02424. It must handed in using learn.dtu (time and date is given in learn). The submissions most contain a collected attached file in Portable Document Format (PDF), .R files can be oploaded as appendix.

When writing the report please explain carefully what you did in each step, back up your statements with quantitative measures when possible, explicitly write down all models used in mathematical notation, and last keep it short and concise.

Part 1: Clothing insulation level

The level of clothing a person is wearing at office is one of the key factors influencing their level of comfort [1]. In addition, the level of comfort influences the need for cooling and/or heating.

The data (`clothingFullAss03.csv`), note that the data have been slightly manipulated for this assignment, include clothing insulation levels (`clo`) worn by subjects of three experimental studies in the Laboratory of Occupant Behavior, Satisfaction, Thermal comfort and Environmental Research (LOBSTER) [2]. In addition there are a number of other variables to be used for modeling (see Table 1).

Table 1: List of included variables in the data-set for part 1

Variable	Type	Description
`clo`	Continuous	Level of clothing
`tOut`	Continuous	Outdoor temperature
`tInOp`	Continuous	Indoor operating temperature
`sex`	Factor	Sex of the subject
`subjId`	Factor	Identifier for subject
`time`	Continuous	Time of measurement (within the day)
`day`	Factor	Day (within the subject)
`time2`	Integer	Measurement number (within the day)
`subDay`	Factor	Unique identifier for day and subject

Mixed effect models

In the first part of this assignment you should fit and develop mixed effect models with different structures, it is important that you write down the final models in maths notation (when appropriate) in each step.

1. Present the data (e.g. plots), try to make it clear that there are differences between subjects.
2. Fit mixed effects models that use subjId as a random effect. You should write down the final model in maths notation.
3. Fit a mixed effect model that include subjId and day (nested to subjId) as random effects (compare with the results from the previous questions), you may use `subDay` as an approximation to these models.
4. Fit a model including within day auto-correlation (repeated measurement set up), with `subDay` as random effect, you should only consider random intercepts in these models.
5. Give an interpretation of your model, including some graphical presentation. Also discuss if there are any features/structure that might be missing in the model set-up you have here.
6. Write a small conclusion of your findings in Part A.

Hierarchical models: Random variance

In this part you should focus on the difference in variance between the subjects. The problem will be solved though a series of sub-problems as stated below. I all questions below we assume that u_i, v_{ij}, ε_{ijk}, and γ_i are iid. and independent of each other.

1. Consider the model

$$clo_{i,j,k} = \mu + \beta(sex_i) + u_i + \varepsilon_{ijk}; \quad u_i \sim N(0, \sigma_u^2) \quad \varepsilon_{ijk} \sim (0, \sigma^2) \tag{1}$$

where $clo_{i,j,k}$ is the clothing insulation level for subject i on day j and k refer to the observation number within the day. Write the density of the observation vector of subject i as a multivariate distribution and implement the likelihood using the matrix-vector formulation.

You may check you result by comparing to

```
library(lme4)
fit0 <- lmer(clo~sex+(1|subjId),data=dat,REML=FALSE)
```

2. Consider the model

$$\begin{aligned} clo_{i,j,k} =& \mu + \beta(sex_i) + u_i + v_{ij} + \varepsilon_{ijk} \\ u_i \sim& N(0, \sigma_u^2) \\ v_{ij} \sim& N(0, \sigma_v^2) \\ \varepsilon_{ijk} \sim& N(0, \sigma^2) \end{aligned} \tag{2}$$

where $clo_{i,j,k}$ is the clothing insulation level for subject i on day j and k refer to the observation number within the day. Write the density of the observation vector of subject i as a multivariate distribution and implement the likelihood using the matrix-vector formulation.

You may check your results by comparing with

```
fit1 <- lmer(clo~sex+(1|subjId)+(1|subjId:day),data=dat,REML=FALSE)
```

3. Consider the model

$$\begin{aligned} clo_{i,j,k} =& \mu + \beta(sex_i) + u_i + v_{ij} + \varepsilon_{ijk}; \\ u_i \sim& N(0, \sigma_u^2 \alpha(sex_i)) \\ v_{ij} \sim& N(0, \sigma_v^2 \alpha(sex_i)) \\ \varepsilon_{ijk} \sim& N(0, \sigma^2 \alpha(sex_i)) \end{aligned} \tag{3}$$

Write the density of the observation vector of subject i as a multivariate distribution and implement the likelihood using the matrix-vector formulation. (Consider reasonable implementations of $\alpha(sex_i)$ e.g. domain and which parameters can be identified).

4. (Week 12) The results in Theorem 6.7 can be formulated as

$$Y_i|\gamma_i \sim N(\mu, \sigma^2/\gamma_i) \tag{4}$$
$$\gamma_i \sim G(1, \phi) \tag{5}$$

i.e. γ_i follow a Gamma distribution with mean value 1 and variance $1/\phi$. Show that the marginal distribution of Y_i is

$$f_{Y_i} \sim \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}; 2\phi\right) \tag{6}$$

where f_0 is the pdf of a student t-distributed random variable with 2ϕ degrees of freedom. (You may want to consult the proofs related to e.g. the Poisson-Gamma model).

5. (Week 12) Consider the model

$$clo_{i,j,k}|u_i, v_{ij}, \gamma_i \sim N(\mu + \beta(sex_i) + u_i + v_{ij}, \sigma^2 \alpha(sex_i)/\gamma_i) \tag{7}$$
$$\begin{aligned} u_i|\gamma_i \sim& N(0, \sigma_u^2 \alpha(sex_i)/\gamma_i) \\ v_{ij}|\gamma_i \sim& N(0, \sigma_v^2 \alpha(sex_i)/\gamma_i) \\ \gamma_i \sim& G(1, \phi) \end{aligned} \tag{8}$$

find the marginal distribution of clo_i (i.e. the vector of all observations on subject i and notation as is in previous questions).

Hint: start by using the results from the first questions to write the distribution of $clo_{i,j,k}|\gamma_i$.

Implement the likelihood and estimate the parameters.

6. (Week 11) Consider the model

$$\begin{aligned} clo_{i,j,k}|u_i, v_{ij}, \gamma_i \sim& N(\mu + \beta(sex_i) + u_i + v_{ij}, \sigma^2\alpha(sex_i)e^{-\gamma_i}) \qquad (9) \\ u_i|\gamma_i \sim& N(0, \sigma_u^2\alpha(sex_i)e^{-\gamma_i}) \\ v_{ij}|\gamma_i \sim& N(0, \sigma_v^2\alpha(sex_i)e^{-\gamma_i}) \\ \gamma_i \sim& N(0, \sigma_G^2) \qquad (10) \end{aligned}$$

   Estimate parameters of the model using the Laplace approximation, either by implementing it yourself or using TMB.

7. Compare the result from point 5 and 6, eg. by comparing the likelihood, the parameters and $\gamma_i|clo_i$ for the two models.

References

[1] Fanger, P.O. (1970). *Thermal Comfort Analysis and Applications in Environmental Engineering*. McGraw-Hill, New York.

[2] Schweiker, M. and Wagner, A. (2015). *A framework for an adaptive thermal heat balance model (ATHB)*. Building and Environment (94), Elsevier Ltd.