Advabced Statistical Methods in Bioinformatics
Bioinformatics Research Center **Asger Hobolth**
Aarhus University **September 19, 2018**

Handout 5

Lectures in Week 39

Monday, September 24:

You should work on the exercise below or the mandatory project. I will be available for questions and, dependent on your progress, I can provide background material, key derivations etc. Please do take advantage of this opportunity.

Wednesday, September 26:

Cancelled because my PhD student Guo Qianyun has her PhD defence. I strongly encourage you to attend the PhD defence in the lecture theatre on the 2nd floor in the BiRC building (building 1110) at 13.15. A main topic in Guo's PhD thesis is non-negative matrix factorization; this is also a topic in the course, so the defence is highly relevant.

Exercises in Week 39

Analysis of Fetal Lamb Movements
The data in the table below are numbers of movements by a fetal lamb observed by ultrasound in successive 5-second intervals. Changes in activity may be due to physical changes in the uterus, or to the development of the central nervous system.

The following three models seem relevant and natural: (i) A single Poisson, (ii) A mixture of two Poisson distributions (corresponding to low or high activity), and (iii) A HMM with two hidden states (corresponding to low or high activity) and Poisson emissions. Fit and compare these three models, and discuss how they are different. Classify (i.e. decode) the observations into low or high activity in model (ii) and (iii). Are the classifiers different? Why? Which models do you prefer? Which classifications do you prefer?

```
ObsSeq <- c(0,0,0,0,0,1,0,1,0,0,0,0,0,0,1,
            0,1,0,0,0,0,2,2,0,0,0,0,1,0,0,
            1,1,0,0,1,1,1,0,0,1,0,0,0,0,0,
            0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,
            1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
            0,0,0,1,0,0,0,0,0,7,3,2,3,2,4,
            0,0,0,0,1,0,0,0,0,0,0,0,1,0,2,
            0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,
            0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,
            2,1,0,0,1,0,0,0,1,0,1,1,0,0,0,
            1,0,0,1,0,0,0,1,2,0,0,0,1,0,1,
            1,0,1,0,0,2,0,1,2,1,1,2,1,0,1,
            1,0,0,1,1,0,0,0,1,1,1,0,4,0,0,
            2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
            0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
```

The data is from Exercise D9 page 123 in Guttorp (1995).

References

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. Chapman and Hall, Suffolf, Great Britain.