Bayesian statistics project – Dec. 13, 2016

Project deadline: January, 27, 2016.

Article: Tanner, M.A. and The Wong, W.H. (1987), Calculation of Posterior Distributions by Data Augmentation.

What we have to do?
1. Explain the theoretical, computational and / or empirical method
2. Emphasize the main points of the paper
3. Make Monte Carlo simulations / use new data

What is in this paper?
Article goal: iterative method for the computation of posterior distributions.
The method consists of iterating the following steps:
  a) Given the current guess of the posterior distribution of θ given y, generate a sample of $m > 0$ latent data patterns from the predictive distribution of z given y
  b) Update the posterior of θ given y, to be the mixture of the m augmented data posteriors.

Article plan:
- *Section 1:* Introduction, data augmentation as a general tool for the analysis of data in complex models.

- *Section 2:* Presentation of the basic algorithm and illustrate its steps in the context of a simple example.

- *Section 3:* Apply the method to the problem of inference on the covariance matrix of the multivariate normal distribution with missing values.

- *Section 4:* Dirichlet Sampling Procedure (DSP).

- *Section 5:* DSP applied to the study of social survey data modeled by a log-linear model with a latent variable.

- *Section 6:* Back to the basic algorithm: uniqueness of the fixed point characterization that motivates the basic algorithm + convergence results. *(can be skipped)*

- *Section 7:* Variations on the basic algorithm + issues in its practical implementation.

Examples:
- Genetic linkage: Exo 5.26 / Example 5.21 from course "computational statistics"
- Social survey data

Proposition of report table of contents:

1. Introduction *(1 page) (At the end)*
   a. What is the main topic of the article?

2. EM algorithm VS article's algorithm *(3 pages) (Alexis)*
   a. Explain what is the EM algorithm
   b. Explain how does the article's algorithm work
   c. Compare both algorithms

3. What is Dirichlet sampling? *(2 – 3 pages) (Gaston)*
   a. How does this work?
   b. What does this thing bring to us in this article?

4. Some examples *(1 – 2 pages) (Thomas)*
   a. Genetic linkage
   b. Social survey data

5. Monte Carlo simulations / EM algorithm *(3 pages) (Thomas)*
   a. Reproduce the results of the paper using Python (multiprocessing?)

Thomas:
- Create a git private repository
- Add Python / Latex in it
- Share the link to other group members