

Projet R

stoehr@ceremade.dauphine.fr

Instructions.

Due to 4th January, 2019, before 8 p.m.

- Must be deposited in the Dropbox folder: a report including comments, code and output (RMarkdown, knitr, Anaconda).
- **Each day of delay is penalised by 1 point.**

The report:

- should contains answers to the questions and comments on the study lead. Clear and concise drafting will be appreciated. The report should not exceed 20 pages.
- Graphs should be carefully annotated.
- When explaining algorithms, you can use pseudo-code/Rmarkdown/knitr to refer to your code. But, **do not simply copy-paste the R code in the report.**

The R code:

- should be well commented and fairly optimised to use the specificity of the language.
- Your code should run without errors and allow to reproduce the results presented in the report (give the random seed used). In R, the function `set.seed` allows to specify the random seed to use.

Model n°6

Let Y be an output binary random variables and X an explanatory random variable distributed according to the uniform distribution on $\{-1; 1\}$. Consider the probit model with random effect

$$\mathbb{P}[Y = 1 \mid \mathbf{X} = \mathbf{x}] = \Phi(\alpha + \beta X + \epsilon),$$

where Φ is the cumulative distribution function of a standard Gaussian variable $\mathcal{N}(0, 1)$ and $\epsilon \sim \mathcal{N}(0, 1)$.

The analysis is based on observed (x_1, \dots, x_n) and (y_1, \dots, y_n) . Prior distributions are set to

$$\alpha \sim \mathcal{N}(0, 1) \quad \text{and} \quad \beta \sim \mathcal{N}\left(0, \frac{n}{\sum_{i=1}^n x_i}\right).$$

Simulation.

1. Explain how to sample from the model. Code the procedure in a function `rgen`.
2. Data may contains noise, a systematic bias or missing/censored values. Give a function `tr.dat` which transforms a sample such that it has one or several of the aforementioned peculiarities.

Parameter inference.

3. Explain how to infer model's parameters with one or several simulation based approaches and give the corresponding code.

Perform the analysis for the following situations:

4. a n -sample from the model with $n = 20$,
5. a n -sample from the model with $n = 500$,
6. a n -sample from the model, with $n = 20$, which exhibits one or several of the aforementioned peculiarities,
7. a n -sample from the model, with $n = 500$, which exhibits one or several of the aforementioned peculiarities.
8. Test the sensitivity of the prior by modifying one or several components.
9. Choose a simpler version of the model and check which type is more adequate for the data.