

# Homework #02

Statistical Methods in Data Science II & Lab

2020/2021

**deadline 06/15/2021**

**Your Last+First Name** \_\_\_\_\_ **Your Matricola** \_\_\_\_\_

1a) Illustrate the characteristics of the statistical model for dealing with the *Dugong*'s data. Lengths ( $Y_i$ ) and ages ( $x_i$ ) of 27 dugongs (see cows) captured off the coast of Queensland have been recorded and the following (non linear) regression model is considered in Carlin and Gelfand (1991):

$$\begin{aligned} Y_i &\sim N(\mu_i, \tau^2) \\ \mu_i = f(x_i) &= \alpha - \beta\gamma^{x_i} \end{aligned}$$

Model parameters are  $\alpha \in (1, \infty)$ ,  $\beta \in (1, \infty)$ ,  $\gamma \in (0, 1)$ ,  $\tau^2 \in (0, \infty)$ . Let us consider the following prior distributions:

$$\begin{aligned} \alpha &\sim N(0, \sigma_\alpha^2) \\ \beta &\sim N(0, \sigma_\beta^2) \\ \gamma &\sim Unif(0, 1) \\ \tau^2 &\sim IG(a, b) (InverseGamma) \end{aligned}$$

1b) Derive the corresponding likelihood function

1c) Write down the expression of the joint prior distribution of the parameters at stake and illustrate your suitable choice for the hyperparameters.

1d) Derive the functional form (up to proportionality constants) of all *full-conditionals*

1e) Which distribution can you recognize within standard parametric families so that direct simulation from full conditional can be easily implemented ?

1f) Using a suitable Metropolis-within-Gibbs algorithm simulate a Markov chain ( $T = 10000$ ) to approximate the posterior distribution for the above model

1g) Show the 4 univariate trace-plots of the simulations of each parameter

1h) Evaluate graphically the behaviour of the empirical averages  $\hat{I}_t$  with growing  $t = 1, \dots, T$

1i) Provide estimates for each parameter together with the approximation error and explain how you have evaluated such error

1l) Which parameter has the largest posterior uncertainty? How did you measure it?

1m) Which couple of parameters has the largest correlation (in absolute value)?

1n) Use the Markov chain to approximate the posterior predictive distribution of the length of a dugong with age of 20 years.

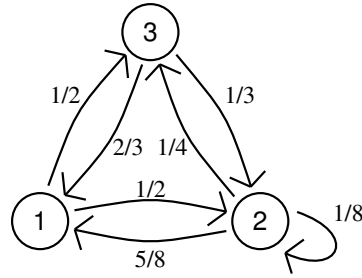
1o) Provide the prediction of a different dugong with age 30

1p) Which prediction is less precise?

*(write your answers and provide your R code for the numerical solution)*

```
#  
#  
#  
#  
#  
#  
#  
#  
#
```

- 2) Let us consider a Markov chain  $(X_t)_{t \geq 0}$  defined on the state space  $\mathcal{S} = \{1, 2, 3\}$  with the following transition



- 2a) Starting at time  $t = 0$  in the state  $X_0 = 1$  simulate the Markov chain with distribution assigned as above for  $t = 1000$  consecutive times
- 2b) compute the empirical relative frequency of the two states in your simulation 2c) repeat the simulation for 500 times and record only the final state at time  $t = 1000$  for each of the 500 simulated chains. Compute the relative frequency of the 500 final states. What distribution are you approximating in this way? Try to formalize the difference between this point and the previous point.
- 2d) compute the theoretical stationary distribution  $\pi$  and explain how you have obtained it
- 2e) is it well approximated by the simulated empirical relative frequencies computed in (b) and (c)?
- 2f) what happens if we start at  $t = 0$  from state  $X_0 = 2$  instead of  $X_0 = 1$ ?

```

#
#
#
#
#
#
#
#
#

```

```

## This homework will be graded and it will be part of your final evaluation
##
##

```

```

## Last update by LT: Wed May 26 16:14:49 2021

```