Department of Mathematical Sciences, NTNU

Examination paper for **TMA4300 Computer Intensive Statistical Methods**

Academic contact during examination: Sara Martino
Phone: 44903330

Examination date: May 12th 2020
Examination time (from-to): 09:00 - 13:00

Permitted examination support material: All support material is allowed

Other information:

- For the exercises that are not multiple choice, all your answers should be justified and the hand-in material should contain sufficient calculations to make your reasoning clear.
- If a question is unclear/vague – make your own assumptions and specify in your answer the premises you have made. Only contact academic contact in case of errors or insufficiencies in the question set.
- **Saving:** Answers written in Inspera are automatically saved every 15 seconds. If you are working in another program remember to save your answer regularly
- **Cheating/Plagiarism:** The exam is an individual, independent work. Examination aids are permitted. All submitted answers will be subject to plagiarism control. *Read more about cheating and plagiarism here.*
- **Notifications:** If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspera. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.
- **Weighting:** The weight in the grading of each subtask is indicated at the bottom of the exercises.
- All files must be uploaded before the examination time expires. 30 minutes are added to the examination time to manage the sketches/calculations/files. (The additional time is included in the remaining examination time shown in the top left-hand corner.)
  - *How to digitize your sketches/calculations*
  - *How to create PDF documents*
  - *Remove personal information from the file(s) you want to upload*

ABOUT SUBMISSION

Your answer will be submitted automatically when the examination time expires and the test closes, if you have answered at least one question. This will happen even if you do not click “Submit and return to dashboard” on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted.

Withdrawing from the exam: If you wish to submit a blank test/withdraw from the exam, go to the menu in the top right-hand corner and click “Submit blank”. This can not be undone, even if the test is still open.

Accessing your answer post-submission: You will find your answer in Archive when the examination time has expired.

/Language: English

1 **Introduction**: Consider the density
$f(x) = C\exp\{-2x\}|\sin(x)|$
where C is a normalising constant and x is a positive real number, $x \in \mathcal{R}^+$.

Exercise:

- Explain how to use rejection sampling to simulate n random samples from this density.
- How can we use the samples to estimate the constant C?

Fill in your answer here

Format B I U x_2 x^2 I_x Ω Σ

Words: 0

Maximum marks: 10

2(a)

Introduction: The code refers to a MCMC program used to generate samples from a target distribution (not provided). The target density in log-scale is defined in the **R** function **dtarget(x, log = TRUE).**

Exercise:

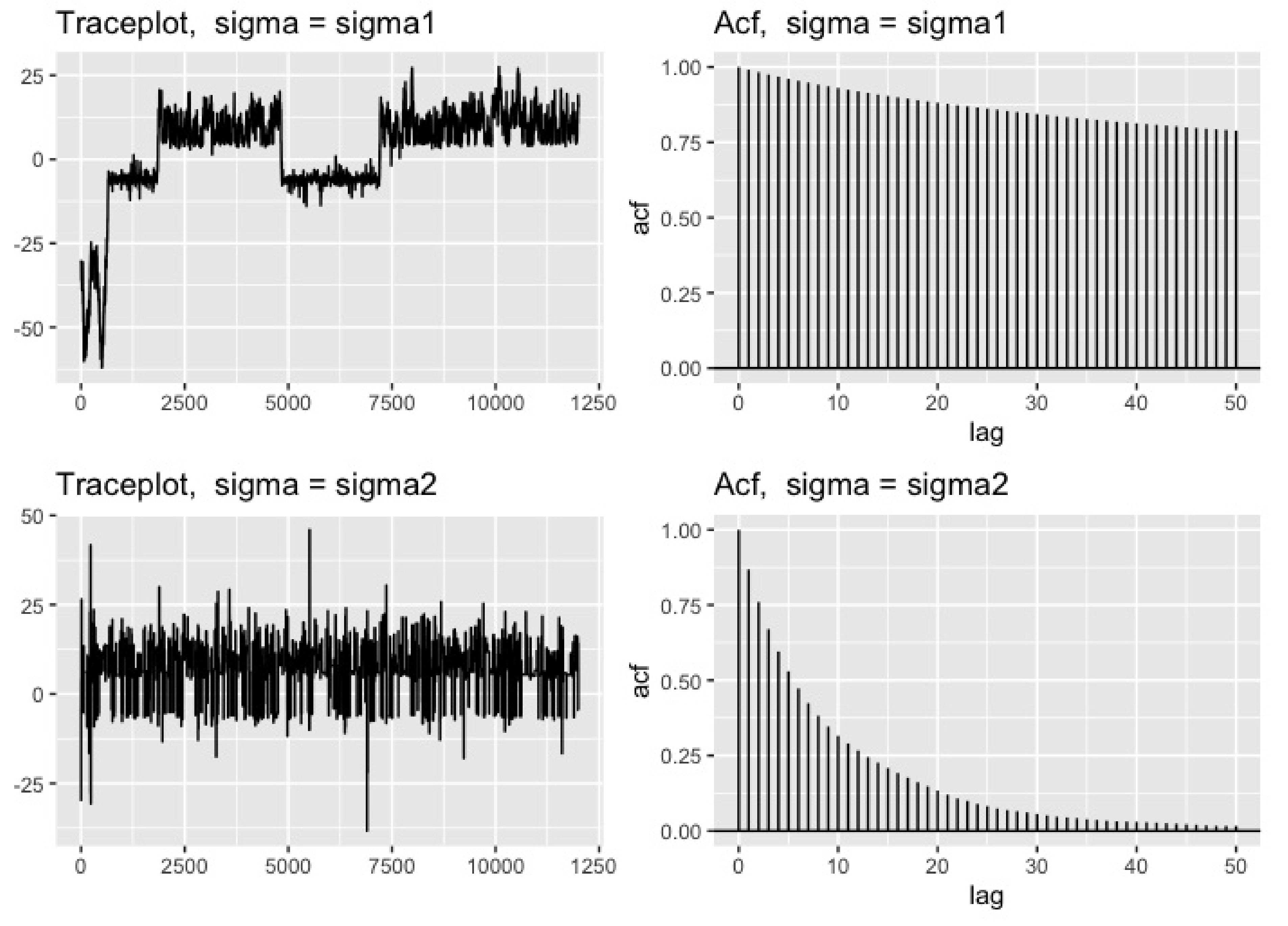
- What type of proposal has been used?

Select one alternative:

- Random walk proposal
- Independence proposal
- Impossible to decide

Maximum marks: 5

2(b) **Introduction:** The code in 2c was used to generate samples using two different value for the parameter **sd.** The plot below shows the traceplot and the autocorrelation function for the two runs



Exercise:

- What is the role of the sd parameter?
- Compare the two runs in term of burn in, acceptance rate and mixing properties.

Fill in your answer here

Format

Words: 0

Maximum marks: 10

3 **Introduction:** Assume that, given the vector $\eta = \{\eta_1, \dots, \eta_n\}$ the observations $y_i, i = 1, \dots, n$ are independent and Poisson distributed with parameter $\lambda_i = \exp(\eta_i)$ i.e.

$$y_i|\eta_i = \text{Poisson}(\lambda_i); i = 1, \dots, n$$

Exercise: Consider the following possible models for the linear predictor η_i (in each case x_i represents the value of a known covariate). For which models can we use INLA to perform posterior inference on the model?

Select one or more alternatives:

- $\eta_i = \alpha + \beta x_i + U_i V_i$
  where
  $\alpha, \beta \sim \mathbf{N}(0,1)$
  $U_i \sim \mathbf{N}(0,1)$ for all $i = 1, \dots, n$
  $V_i \sim \mathbf{N}(0,1)$ for all $i = 1, \dots, n$

- $\eta_i = \alpha + \beta x_i + U_i$
  where
  $\alpha, \beta \sim \mathbf{N}(0,1)$
  $U_i \sim \mathbf{N}(0,1)$ for all $i = 1 \dots, n$

- $\eta_i = \alpha + \beta x_i + V_i$
  where
  $\alpha, \beta \sim \mathbf{N}(0,1)$
  $V_i \sim \text{Bernoulli}(0.3)$ for all $i = 1 \dots, n$

- $\eta_i = \alpha + \beta x_i$
  where
  $\alpha, \beta \sim \mathbf{N}(0,1)$

Maximum marks: 10

4 **Introduction:** Assume that $x_1, \ldots, x_n$ is a sample from a normal distribution with mean θ and variance 1. Consider the following two statements:

1. "A 95% confidence interval for θ is the interval [2, 2.5]."
2. "A 95% credibility interval for θ is the interval [2 2.5]."

Exercise:

Describe the context in which each statement is used, and and for each of statement 1 and 2 write down a precise interpretation.

Format

Words: 0

Maximum marks: 10

5(a) **Introduction:** Consider the following model for the observed data $(y_1, \dots, y_n)$:

$$y_i = \begin{cases} \mathcal{N}(0,1), \text{ if } x_i = 1 \\ \text{Gamma}(1,1), \text{ if } x_i = 0 \end{cases}$$

where $(x_1, \dots, x_n)$ are unobserved Bernoulli variables with unknown parameter θ.

Exercise:

- Explain the principles of the EM algorithm when applied to the estimation of the parameter θ

Fill in your answer here

Format B I U $\times_2$ $\times^2$ I_x Ω Σ

Words: 0

Maximum marks: 10

5(b) **Introduction**: We want to apply the EM algorithm to estimate the parameter $\theta$for the model in 5a)

Exercise:

- Derive an expression for the complete data likelihood $L(\theta; y_1, \ldots, y_n, x_1, \ldots, x_n)$
- Derive the function $Q(\theta|\theta')$
- Find a formula for the θ maximizing $Q(\theta|\theta')$

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **35 GB**

Select file to upload

Maximum marks: 10

5(c) **Introduction**: we want to estimate the variance of the EM estimator of θ derived in 5b) using bootstrap

Exercise:

- Write a pseudocode for how we can use bootstrapping to estimate the standard deviations of the maximum likelihood estimators for θ

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **35 GB**

Select file to upload

Maximum marks: 10

6 **Introduction**: Assume we have the following pdf

$f(x) = C \exp\{k_1 x - k_2 e^x\} = C \exp\{h(x)\}$

where $x \in \mathcal{R}$, C is such that the density integrates to 1. Moreover let $k_1 = 10 \text{ and } k_2 = 2$

Exercise:

- Find the mode of the density
- Use the Taylor expansion to the second order to approximate the function $h(x) = (k_1 x - k_2 e^x)$ and build a Gaussian approximation of $f(x)$ around its mode.
- Use the Gaussian approximation to estimate the value of the constant C

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **35 GB**

Select file to upload

Maximum marks: 10

7

Introduction: Given the probability density

$g(x_1, x_2) = C\ \exp(-x_1^2 x_2 + x_1 \log x_2)$

where $x_1, x_2 > 0$ and C is such that the density integrates to 1.

Exercise:

- Explain in detail how you can use Gibbs sampling to obtain samples from the distribution. Write down the full conditionals and write a pseudo-code

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **35 GB**

Select file to upload

Maximum marks: 10

Question 2

Attached

```r
my.mcmc <- function(x,  numit , sd )
  {
  xsamples <- rep(NA, numit)
  yes<-0
  no<-0
  xsamples[1] = x
  # specify a starting value x<-0.0
  for(k in 2:numit){
    # propose a new value
    proposal <- rnorm(1, mean=x, sd=sd )
    # compute log posterior ratio
    logposterior.ratio <- dtarget(proposal, log=TRUE) - dtarget(x, log=TRUE)
    # compute log proposal ratio
    logproposal.ratio <- 0
    # derive the acceptance probability (on log scale)
    alpha <- logposterior.ratio + logproposal.ratio # accept -reject step
    if(log(runif(1)) <= alpha){
      # accept the proposed value
      x <- proposal
      # increase counter of accepted values
      yes<-yes+1
    } else{
      # stay with the old value
      no <- no + 1 }
    if(k %% 100 == 0){
      # print every 100 iterations the acceptance rate
      cat("The acceptance rate is:", round(yes/(yes+no)*100,2), "%\n")
    }
    xsamples[k] = x
    }
  return(xsamples)
  }
```