i Info

Department of Mathematical Sciences, NTNU

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

Examination date: June 10th 2021

Examination time (from-to): 09:00 - 13:00

Academic contact during examination: Sara Martino
Phone: 99403330

Technical support during examination: Orakel support services
Phone: 73 59 16 00

Permitted examination support material: A/All support material is allowed

Other information:

- **Make your own assumptions:** 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.

- **Answers**: All answers must be justified, and necessary derivations and calculations must be included.Read all subproblems before you start.

ABOUT SUBMISSION

File Upload: 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*

NB! You are responsible to ensure that you upload the correct file(s) for all questions. Check the file(s) you have uploaded by clicking “Download” when viewing the question. All files can be removed or replaced as long as the test is open.

The additional 30 minutes are reserved for submission. If you experience technical problems during upload/submission, you must contact technical support before the examination time expires. If you can’t get through immediately, hold the line until you get an answer

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 Direct Sampling

This problem consists of two parts: a) and b). The answers to these parts should be uploaded in one pdf file

Consider the standard Weibull distribution with parameters α and $\beta = 1.$ The density function is:

$$f_0(x|\alpha) = \alpha x^{\alpha-1}e^{-x^{\alpha}}.$$

Part a) [10 %]
Explain how to sample from such distribution using one of the methods learned in the course.

Part b) [10 %]
Assume that we want to simulate two dependent random variable that are marginally Weibull distributed. One way to do that is through the so called copula approach.

Let $\Phi(\cdot)$ be the cumulative distribution function for the standard normal distribution. Show that if $Y \sim N(0,1)$ then

$$X = F_0^{-1}(\Phi(Y))$$

where $F_0(x) = \int_0^x f_0(u)du$, has a standard Weibull distribution.

Assume now that you can simulate $\mathbf{y} = (y_1, y_2)$ from a bivariate normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

Explain how you can use this to simulate two dependent Weibull distributed variables.

Last opp filen her. Maks én fil.

Alle filtyper er tillatt. Maksimal filstørrelse er **50 GB**.

Velg fil for opplasting

Maks poeng: 20

2 MCMC

This problem consists of two parts: a) and b). The answers to these parts should be uploaded in one pdf file.

Let $\mathbf{x} \sim \pi(\mathbf{x})$ and assume that we want to use the following algorithm to simulate from $\pi(\mathbf{x})$:

Given the value $\mathbf{x}^{(t)}$ at step t:

1. Generate a proposal $\mathbf{y} \sim Q(\mathbf{x}^{(t)}, \mathbf{y})$. We assume that $Q(\mathbf{x}, \mathbf{y}) > 0$ for all possible $\mathbf{x}, \mathbf{y}$.
2. Sample $U \sim \text{Unif}(0, 1)$ and update $\mathbf{x}^{(t+1)}$ as

$$\mathbf{x}^{(t+1)} = \begin{cases} \mathbf{y}, & \text{if } U < \alpha_1(\mathbf{x}^{(t)}, \mathbf{y}) \\ \mathbf{x}^{(t)}, & \text{otherwise.} \end{cases}$$

where

$$\alpha_1(\mathbf{x}, \mathbf{y}) = \frac{\pi(\mathbf{y})Q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})Q(\mathbf{x},\mathbf{y}) + \pi(\mathbf{y})Q(\mathbf{y},\mathbf{x})}$$

Part a)

Show that $\mathbf{x}^{(t)}, t = 1, 2, \dots$ is a Markov chain that converges to the target distribution $\pi(\mathbf{x})$

Use the following theorem to reason about the efficiency of the above algorithm when compared to the traditional Metropolis-Hastings algorithms given that both approaches use the same proposal distribution $Q(\mathbf{x}, \mathbf{y})$.

Theorem: Consider two transition kernels P_1 and P_2 with the same stationary distribution π. Let $\hat{\mu_1}$ and $\hat{\mu_2}$ be the estimates of $\mu = \int h(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}$ based on the chains generated from P_1 and P_2 and let v_1 and v_2 be the variances of $\hat{\mu_1}$ and $\hat{\mu_2}$.
If $P_1(\mathbf{x}, \mathbf{y}) \geq P_2(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x} \neq \mathbf{y}$ then $v_1 \leq v_2$

Part b)

Assume now that $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is a four dimensional vector and assume that you have run the algorithm above for N steps. Explain how you can use the generated samples to estimate the quantity

$$q = \text{Prob}(X_2 > \frac{X_1 + X_3}{X_4})$$

Last opp filen her. Maks én fil.

Alle filtyper er tillatt. Maksimal filstørrelse er **50 GB**.

Velg fil for opplasting

Maks poeng: 20

3 Rejection Sampling

This problem consists of two parts: a) and b). The answers to these parts should be uploaded in one pdf file

We consider here a modification of the rejection sampling algorithm for sampling from $\pi(x) = \tilde{\pi}(x)/Z_\pi$ where Z_π is a normalizing constant, such that:

$$h(x) \leq \tilde{\pi}(x) \leq K\tilde{g}(x)$$

where $h(\cdot)$is a non-negative function, $K > 0$ and $g(x) = \tilde{g}(x)/Z_g$ is a density easy to sample from (here Z_g is a normalizing constant).
The algorithm is defined as follows:

1. Draw $X \sim g$ and $U \sim \mathrm{Unif}(0,1)$ independently from each other
2. If $U \leq h(X)/(K\tilde{g}(x))$ accept X
3. If X is not accepted in step (2), sample $V \sim \mathrm{Unif}(0,1)$ independently and accept X if

$$V \leq \frac{\tilde{\pi}(X) - h(X)}{K\tilde{g}(X) - h(X)}$$

Part a)
Show that the probability of accepting X is

$$\frac{\tilde{\pi}(x)}{K\tilde{g}(x)}$$

and use this to conclude that the distribution of the samples accepted by the above algorithm is $\pi(x)$.

Part b)
Show that the probability that step (3) has to be carried out is
$$1 - \frac{\int_X h(x)dx}{KZ_g}$$
Can you think of cases when such algorithm is convenient over the standard rejection sampling?

Last opp filen her. Maks én fil.

Alle filtyper er tillatt. Maksimal filstørrelse er **50 GB**.

Velg fil for opplasting

Maks poeng: 20

4 INLA

This problem consists of two parts: a) and b). The answers to these parts should be uploaded in one pdf file.

Consider the model

$$\begin{aligned} \mathbf{y}|\mathbf{x} &\sim \pi(\mathbf{y}|\mathbf{x}) \\ \mathbf{x}|\theta &\sim \pi(\mathbf{x}|\theta) \\ \theta &\sim \pi(\theta) \end{aligned}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors of length n while θ is scalar.

Part a

Explain which characteristic needs the above model to fulfill in order to be amenable to inference through the Integrated Nested Laplace Approximation (INLA)

Part b

List at least two advantages and two disadvantages of using INLA to perform Bayesian inference compared to using MCMC.

Last opp filen her. Maks én fil.

Alle filtyper er tillatt. Maksimal filstørrelse er **50 GB**.

Velg fil for opplasting

Maks poeng: 20

5 Bootstrap

This problem consists of two parts: a) and b). The answers to these parts should be uploaded in one pdf file

Part a)
Explain the general idea of bootstrapping and when it is used.

Part b)
Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$be an original sample of size n from some distribution and assume we would like to use bootstrapping to generate a new sample. Which are the most and least likely re-sampled observation vectors. Please argue.

Last opp filen her. Maks én fil.

Alle filtyper er tillatt. Maksimal filstørrelse er **50 GB**.

Velg fil for opplasting

Maks poeng: 20