Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for
TMA4300 Computer Intensive Statistical Methods

Academic contact during examination: Andrea Riebler

Phone: 4568 9592

Examination date: June 1st, 2016

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

Permitted examination support material: C:

- Calculator HP30S, CITIZEN SR-270X or CITIZEN SR-270X College, Casio fx-82ES PLUS with empty memory.
- Statistiske tabeller og formler, Tapir.
- K. Rottmann: Matematisk formelsamling.
- One yellow, stamped A5 sheet with own handwritten formulas and notes.
- Dictionary in any language.

Other information:

- All 10 sub-problems in this exam count approximately the same.
- All answers must be justified!!!
- In your solution you can use English and/or Norwegian.

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1

a) Based on an ordinary multiple regression model, explain what we understand by "bootstrapping the residuals" and how it is used. Contrast it to "paired bootstrap".

Problem 2

Assume you are only able to sample from a uniform distribution Unif(0,1).

a) Assume we would like to generate samples from a continuous distribution with density $f(x)$ and cumulative distribution function $F(X)$. Explain how the inverse transform technique works and why it generates samples from $f(x)$.

b) Use the inverse transform technique to sample from a standard double exponential distribution:

$$f(x) = \frac{1}{2}\exp(-|x|), \quad x \in \mathbb{R}$$

Problem 3

Assume we have conditionally independent Poisson count data y_i, $i = 1, \dots, n$, with mean θ_i:

$$y_i \mid \theta_i \sim \text{Poisson}(\theta_i)$$

where means θ_i are gamma distributed

$$\theta_i \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta).$$

Assume priors $\alpha \sim \text{Exp}(a)$ and $\beta \sim \text{Gamma}(b, c)$, where a, b and c are treated as fixed constants.

The gamma distribution Gamma(α, β) has density function:

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x), \quad \text{with } x \geq 0 \text{ and } \alpha, \beta > 0,$$

and the exponential distribution Exp(a) has density function:

$$p(x) = a\exp(-ax), \quad \text{with } x \geq 0 \text{ and } a > 0.$$

a) Derive the full conditional distributions of the individual components of the parameter vector $(\alpha, \beta, \theta_1, \ldots, \theta_n)$. If possible define the parametric distribution and its parameters.

b) Use pseudo code to outline how you would obtain samples from the posterior distribution using Markov chain Monte Carlo (MCMC).

Problem 4

a) For each of the two models below, explain whether the INLA methodology could be used to get parameter estimates. Give reasons if you think it cannot be used.

- Model 1: Assume we have conditionally independent binomial distributed data y_i, $i = 1, \ldots, I$, with fixed sample sizes n_i and success probability p_i,

$$y_i \mid p_i \sim \text{Binomial}(n_i, p_i).$$

Using a logit link function we assume that

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i + u_i.$$

Here x_i, $i = 1, \ldots, I$, are given covariate values. The prior distributions for the second stage of this hierarchical model are

$$\begin{aligned} \beta_0 &\sim \mathcal{N}(0, 0.001^{-1}) \\ \beta_1 &\sim \mathcal{N}(0, 0.001^{-1}) \\ u_i \mid \tau &\sim \mathcal{N}(0, \tau^{-1}) \end{aligned}$$

There is one hyperparameter to which we assign a gamma prior distribution:

$$\tau \sim \text{Gamma}(1, 0.05)$$

- Model 2: Assume we have observations y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, that are conditionally independent normal distributed

$$y_{ij} \mid \mu_{ij}, \tau \sim \mathcal{N}(\mu_{ij}, \tau^{-1}).$$

The mean is modelled as

$$\mu_{ij} = \beta_0 + u_i + v_{ij}.$$

The prior distributions for the second stage of this hierarchical model are

$$\begin{aligned} \beta_0 &\sim \mathcal{N}(0, 0.001^{-1}) \\ u_i \mid \kappa &\sim \mathcal{N}(0, \kappa^{-1}) \\ v_{ij} &\sim \text{Bernoulli}(0.3) \end{aligned}$$

There are two precision parameters (inverse variance parameters), τ and κ, in the model to which we assign gamma priors,

$$\begin{aligned} \tau &\sim \text{Gamma}(1, 0.05) \\ \kappa &\sim \text{Gamma}(1, 0.05). \end{aligned}$$

Problem 5

Cole et al. (2012) propose a rejection sampling approach to sample from a posterior distribution as a simple and sometimes efficient alternative to MCMC. They summarise their approach as follows:

1. Define a model with likelihood $f(y \mid \theta)$ and prior distribution $f(\theta)$.
2. Compute the maximum likelihood estimator $\hat{\theta}_{\text{ML}}$.
3. To obtain a sample from the posterior distribution:
   (a) Draw $\theta^\star$ from the prior distribution.
   (Note, this must cover the range of the posterior distribution).
   (b) Compute the ratio $p = f(y \mid \theta^\star)/f(y \mid \hat{\theta}_{\text{ML}})$
   (c) Draw u from a uniform distribution Unif(0,1).
   (d) If $u \leq p$ then accept $\theta^\star$. Otherwise, reject $\theta^\star$ and repeat from step a).

a) Using Bayes' rule write out the posterior density $f(\theta \mid y)$. What are the target distribution $h(\theta)$ and the proposal distribution $g(\theta)$ used in the rejection sampler?

b) For which value of a is the acceptance probability

$$p = \frac{h(\theta^\star)}{a \cdot g(\theta^\star)}$$

used in the rejection sampler equal to $f(y \mid \theta^\star)/f(y \mid \hat{\theta}_{\text{ML}})$. Explain why the inequality $h(\theta) \leq a \cdot g(\theta)$ is guaranteed by the approach of Cole et al. (2012). Could the choice of a be improved in order to get a more efficient algorithm? If so, describe how, if not explain why.

Problem 6

Assume we are interested to sample from a target density $h(x)$, but we are only able to sample from a proposal distribution $g(x)$. Suppose that $a \geq 1$ is a known constant, but that $h(x)$ is not less or equal than $a \cdot g(x)$ for all x; that is $a \cdot g(x)$ does not dominate $h(x)$ over the whole range. We define the set C where domination occurs as

$$C = \{x : h(x) \leq a \cdot g(x)\}$$

and C^c as the set where $a \cdot g(x)$ does not dominate. Figure 1 shows an illustration of a nondominating density and the C region. Assume we apply the rejection

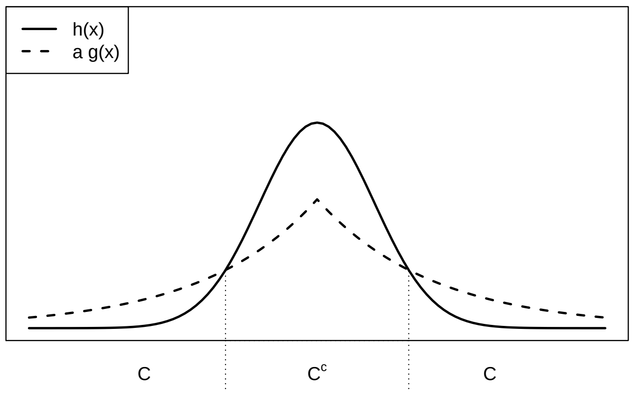

Figure 1: Rejection sampling when $a \cdot g(x)$ is not dominating $h(x)$ over the whole support range.

sampling approach as if $a \cdot g(x)$ would be dominating over the whole range.

a) Derive the density of the random variable Y that comes out of this algorithm and show that it is equal to:

$$\begin{aligned} q(y) &= \frac{h(y)}{a \cdot d}, \qquad \text{if } y \in C \\ &= \frac{g(y)}{d}, \qquad \text{if } y \neq C, \end{aligned}$$

with $d = P[U \leq h(Z)/(a \cdot g(Z))]$ where Z is distributed according to g and $U \sim \text{Unif}(0,1)$.

b) Derive an expression for d and show that it reduces to $1/a$ if $a \cdot g(x)$ would be a dominating density.