NTNU
Norwegian University of Science and Technology

Department of Mathematical Sciences

Examination paper for
TMA4300 Computer Intensive Statistical Methods

Academic contact during examination: Håkon Tjelmeland

Phone: 4822 1896

Examination date: May 24th 2017

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

Permitted examination support material: C:

- Calculator HP30S, CITIZEN SR-270X, CITIZEN SR-270X College or Casio fx-82ES PLUS with empty memory.
- Statistiske tabeller og formler, Akademika.
- One yellow, stamped A5 sheet with own handwritten formulas and notes.

Other information:

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

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1

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

a) Use the probability integral transform approach (the inverse cumulative distribution function technique) to sample from the distribution with probability density

$$g(x) = \begin{cases} 2xe^{-x^2} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Use pseudo code to specify the simulation algorithm.

In the following you can also assume that you are able to sample from the distribution considered in **a**).

b) Use rejection sampling with a proposal distribution with density $g(x)$ as specified in **a**) to sample from the distribution with probability density

$$f(x) = \begin{cases} cxe^{-x^3} & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a normalising constant. In particular simplify the expression for the acceptance probability as much as possible and use pseudo code to specify the simulation algorithm.

Problem 2

Consider a random sample $x_1, \ldots, x_n$ from an exponential distribution with (unknown) intensity λ, i.e. with density

$$p(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Write $p(x;\lambda)$ on the form of a one-parameter exponential family,

$$p(x;\lambda) = a(x)e^{\phi(\lambda)t(x)+b(\lambda)},$$

i.e. identify functions $a(x)$, $\phi(\lambda)$, $t(x)$ and $b(\lambda)$.

Use this to write down a formula for the conjugate prior distribution for λ and identify what type of distribution this is.

Problem 3

In this problem we will consider a hierarchical Bayesian model for analysing observed life times for a specific type of electronic components. Assume we have available such components from N different producers. We number the producers from 1 to N. For producer number $i \in \{1, \ldots, N\}$ we test n_i components, and number these components from 1 to n_i. For component number j from producer number i we observe how long time x_{ij} this component works before it fails. We assume x_{ij} to be exponentially distributed with intensity λ_i, i.e.

$$x_{ij}|\lambda_i \sim \text{Exponential}(\lambda_i).$$

The intensity λ_i is thus characterising the quality of the components produced by producer number i. We assume the various x_{ij}'s to be conditionally independent given the intensities $\lambda_1, \ldots, \lambda_N$.

Apriori we assume the λ_i's to be conditionally independent and identically distributed given two hyper-parameters α and β, and we assume

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

As the last step in the hierarchical model we assume α and β to be apriori independent, β to be inverse gamma distributed,

$$\beta|\alpha \sim \text{InvGamma}(a, b),$$

where a and b are treated as fixed constants, and α is assumed to have an (improper) uniform distribution on $[0, \infty)$.

The exponential distribution Exponential(λ) has density as given in Problem **2**, the gamma distribution Gamma(α, β) has density

$$p(x|\alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and the inverse gamma distribution InvGamma(a, b) has density

$$p(x|a, b) = \begin{cases} \frac{1}{b^a \Gamma(a)} \frac{e^{-\frac{1}{xb}}}{x^{a+1}} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

a) Write down an expression for the posterior distribution $p(\lambda_1, \ldots, \lambda_N, \alpha, \beta|x)$, where $x = (x_{ij}; j = 1, \ldots, n_i, i = 1, \ldots, N)$ is the observed failure times. It is sufficient to find an expression that is proportional to the posterior distribution.

Derive the full conditional distributions for each component in the parameter vector $(\lambda_1, \ldots, \lambda_N, \alpha, \beta)$. If possible specify what parametric family each full conditional belongs to, and their parameter values.

b) Use pseudo code to outline how you would generate samples from the posterior distribution using Markov chain Monte Carlo (MCMC). Specify in particular what your proposal distributions are and simplify as much as possible the expressions for the corresponding acceptance probabilities.

Assume you have run the MCMC algorithm for M iterations and denote the generated states by $\{(\lambda_1^m, \ldots, \lambda_N^m, \alpha^m, \beta^m)\}_{m=0}^M$. In particular $(\lambda_1^0, \ldots, \lambda_N^0, \alpha^0, \beta^0)$ is the initial state. It is of interest to use the MCMC output to estimate the following three quantities.

1. The posterior mean of λ_i, i.e.
$$\mathrm{E}[\lambda_i|x].$$
2. The posterior probability that the quality of the components from producer number i is better than the quality of the components from producer number j, i.e.
$$P(\lambda_i < \lambda_j|x).$$
3. The posterior probability that a new component from producer number i will work at least until a given time t, i.e.
$$P(x_{i,\text{new}} > t|x),$$
where x is the observed failure times and $x_{i,\text{new}}$ is the failure time for a new component from producer number i.

c) Specify how you would estimate each of the three quantities above based on the output from the MCMC algorithm.

Problem 4

Assume we have observed n values $x_1, \ldots, x_n$ which we assume are independent samples from some unknown distribution $F(x)$. Assume we are interested in the variance of this distribution,

$$\sigma^2 = \mathrm{Var}_F[X] = \mathrm{E}_F\left[(X - \mu_F)^2\right],$$

where μ_F is the mean value in the distribution F.

a) Explain briefly what we mean by the plug-in principle.

Show that if we apply the plug-in principle for estimating σ^2 we obtain the estimator

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the empirical mean.

As you know the most commonly used estimator for σ^2 is not $\widehat{\sigma}^2$, but

$$\widehat{\widehat{\sigma}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

b) Define the bias of $\widehat{\widehat{\sigma}}^2$.

Define the ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$.

Use pseudo code to show how one can use stochastic simulation to estimate the ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$.

The ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$ can in fact be found analytically.

c) Derive an easy to compute formula for the ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$.