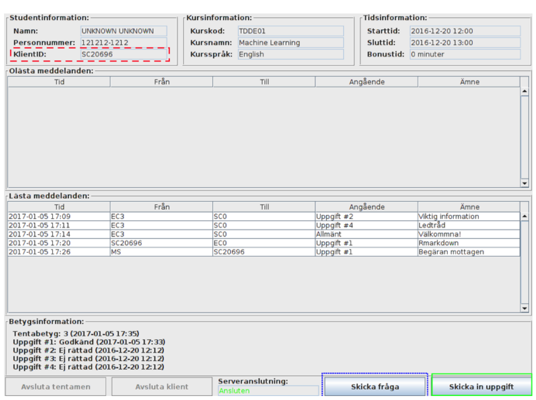
LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Per Sidén

2019-08-21
Bayesian Learning, 6 hp
732A91/TDDE07

Computer Exam - Bayesian Learning (732A91/TDDE07), 6 hp

Time:	14-18
Allowable material:	- The allowed material in the folder given_files in the exam system. - Calculator with erased memory.
Teacher:	Per Sidén. Phone: $070 - 4977175$ and through the Communication client.
Exam scores:	Maximum number of credits on the exam: 40. Maximum number of credits on each exam question: 10.
Grades (732A91):	A: 36 points B: 32 points C: 24 points D: 20 points E: 16 points F: <16 points
Grades (TDDE07):	5: 34 points 4: 26 points 3: 18 points U: <18 points

INSTRUCTIONS:
When asked to give a solution on **Paper**, give that answer on physical papers supplied with the exam.
Each submitted sheet of paper should be marked with your *Client ID* from the *Communication Client*.
The client ID is the code in the `red` dashed rectangle in figure below.
All other answers should be submitted in a single PDF file using the *Communication Client.*
Include important code needed to grade the exam (inline or at the end of the PDF).
Submission starts by clicking the button in the `green` solid rectangle in figure below.
The submitted PDF file should be named *BayesExam.pdf*
Questions can be asked through the Communication client (`blue` dotted rectangle in figure below).
Full score requires clear and well motivated answers.



1. Regression

The Boston housing data contains characteristics of 506 houses in the Boston suburbs and their selling price. The dataset is loaded by the `ExamData.R` file. The original data is in `Boston` and `?Boston` will present the help file with information on all variables. We are here interested in modeling the response variable `medv` (value of the house in 1000$) as a function of all the other variables in the dataset. The `ExamData.R` also prepares the data so that the vector `y` contains the response variable and the matrix `X` contains the covariates (with the first column being ones to model the intercept term). The vector `covNames` contains the names of all the covariates. Use the conjugate prior

$$\beta|\sigma^2 \sim N(0, 10^2\sigma^2 I)$$
$$\sigma^2 \sim Inv - \chi^2(1, 5^2).$$

(a) Credits: 4p. Use the function `BayesLinReg` supplied in `ExamData.R` to simulate 5000 draws from the posterior distribution of all regression coefficients and the error variance. Summarize the posterior of the regression coefficients by the point estimate under the quadratic loss function, and by 95% equal-tail credible intervals. Interpret the credible interval for the regression coefficient on nitrogen oxides concentration (`nox`).

(b) Credits: 3p. Based on your samples, do a kernel density estimate of the posterior of σ^2 using the density function in R with default settings and plot it. Use that kernel density estimate to compute the posterior mode and a 90% Highest Posterior Density interval for σ^2.

(c) Credits: 3p. A construction company is planning to build a new house with covariates as given in the vector `XNewHouse`. The construction cost is 20000$ and the company is planning to sell the house when it is finished. Do a Bayesian analysis (using simulation methods) to determine how probable it is that the company will make money (that the house will sell for more than 20000$) on the project.

2. Earthquakes

Let the time x (in years) between successive occurrences of earthquakes in a certain region follow a geometric distribution

$$p(x|\theta) = (1-\theta)^x \theta, \qquad x = 0, 1, 2, 3, \ldots, \qquad 0 < \theta < 1,$$

where $x = 0$ corresponds to the next earthquake occurring within the same year as the previous one. The times between the last five earthquakes were $35, 14, 4, 10$ and 2 years. Consider the data as independent and assign a Beta(α, β)-prior to θ.

(a) Credits: 5p. Derive the posterior mode for θ on **Paper**.

(b) Credits: 5p. The last earthquake occurred this year. Using Bayesian methods, compute the maximal number of years until the next earthquake has occurred with 95% probability. Use prior parameters $\alpha = 1$ and $\beta = 1$. [Hints: The posterior for θ is Beta-distributed. See `?Geometric` in R].

3. Data from different distributions

Assume that $x_1, \dots, x_m$ are independent observations from a model where each observation is distributed according to the probability density function (pdf) $p(x_i|\theta)$. Further assume that we have access to a second independent sample $y_1, \dots, y_n$ of independent observations from a different model, distributed according to the pdf $p(y_i|\theta)$ and that we use the prior distribution $p(\theta)$. Note that both distributions depend on the same parameter θ.

(a) Credits: 2p. Show on **Paper** that using the posterior based on the first sample $x_1, \dots, x_m$ as a prior when analyzing the second sample $y_1, \dots, y_n$ gives the same posterior as when analyzing both samples simultaneously. That is, show that

$$p(y_1, \dots, y_n|\theta)\, p(\theta|x_1, \dots, x_m) \propto p(x_1, \dots, x_m, y_1, \dots, y_n|\theta)\, p(\theta).$$

(b) Credits: 4p. Assume now that $x_1, \dots x_m$ are exponentially distributed and that θ has a Gamma$(3, 2)$ prior, that is, $p(x_i|\theta) = \theta \exp(-\theta x_i)$, for $x_i > 0$ and $p(\theta) = 4\theta^2 \exp(-2\theta)$, for $\theta > 0$. Derive the posterior $p(\theta|x_1, \dots, x_m)$ on **Paper**.

(c) Credits: 4p. In addition to the assumptions in (b), assume that $y_1, \dots, y_n$ follows a t-distribution with 5 degrees of freedom and $\sigma = 1$, centered around $\log(\theta)$, that is

$$p(y_i|\theta) = \frac{\Gamma(3)}{\Gamma(5/2)\sqrt{5\pi}} \left(1 + \frac{1}{5}(y_i - \log(\theta))^2\right)^{-3}.$$

Write a function in R that computes the log unnormalized posterior distribution of θ based on all observations $x_1, ..., x_m, y_1, \dots, y_n$. Use that function to plot the normalized posterior distribution of θ on the interval $(0, 2)$ for the observations in the data vectors `xData` and `yData` in the supplied file `ExamData.R`, containing the values of $x_1, ..., x_m$ and $y_1, \dots, y_n$.

4. Metropolis-Hastings for aircraft incidents

The file `incidents`, which can be loaded by the code in `ExamData.R`, contains data on the number of aircraft incidents each year from 1970 until 2018. Denote the observations with $x_1, ..., x_n$ and assume that they are independent and identically distributed even though this ignores the temporal dependence present in the data. Further assume that they follow a negative binomial distribution with the alternative parameterization

$$p\left(x|\mu,\phi\right) = \left(\begin{array}{c} x+\phi-1 \\ x \end{array}\right)\left(\frac{\mu}{\mu+\phi}\right)^{x}\left(\frac{\phi}{\mu+\phi}\right)^{\phi}, \quad x = 0,1,2,\ldots \quad \mu > 0, \quad \phi > 0,$$

where μ is the mean of the distribution and ϕ is a parameter that controls the overdispersion. Assume the following prior for μ and ϕ

$$p\left(\mu,\phi\right) \propto \frac{1}{\phi^2}.$$

(a) Credits: 5p. Simulate from the posterior using the Metropolis algorithm. Denote $\theta = \left(\mu,\phi\right)^T$ and use as proposal density the multivariate normal density (random walk Metropolis):

$$\theta_p|\theta^{(i-1)} \sim N\left(\theta^{(i-1)}, c\cdot\Sigma\right),$$

where and $\theta^{(i-1)}$ is the previous draw, $c = 0.1$ and $\Sigma = \text{diag}(100, 5)$ is a diagonal matrix which is close to the covariance matrix of a normal approximation of the posterior of θ at the mode. Use 50 iterations burn-in and thereafter draw 1000 samples from the posterior. Use $\theta^{(0)} = \left(200, 20\right)^T$ as initial values. Produce traceplots of the posterior samples over the iterations. Based on these plots, would you say that this MCMC sampler is efficient? Motivate your answer. If your answer is no, make a suggestion about how to improve the efficiency of the Metropolis algorithm. [Hint: The proposal distribution can be truncated to avoid proposals of $\mu \leq 0$ or $\phi \leq 0$ which can lead to numerical errors, use e.g. `theta[theta<=0]=1e-6`.]

(b) Credits: 5p. Now, instead simulate from the posterior using the Metropolis-Hastings algorithm. Use a non-symmetric proposal density which is independent between μ and ϕ (θ_1 and θ_2):

$$\theta_{j,p}|\theta_j^{(i-1)} \sim Gamma\left(c\cdot\theta_j^{(i-1)}, c\right),$$

for $j = 1$ and $j = 2$ and $c = 0.8$. Again, use 50 iterations burn-in and thereafter draw 1000 samples from the posterior and $\theta^{(0)} = \left(200, 10\right)^T$ as initial values. Produce new traceplots of the posterior samples over the iterations. Based on these plots, would you say that this MCMC sampler is efficient? Motivate your answer. If your answer is no, make a suggestion about how to improve the efficiency of the Metropolis-Hastings algorithm.

Good luck!

Mattias and Per