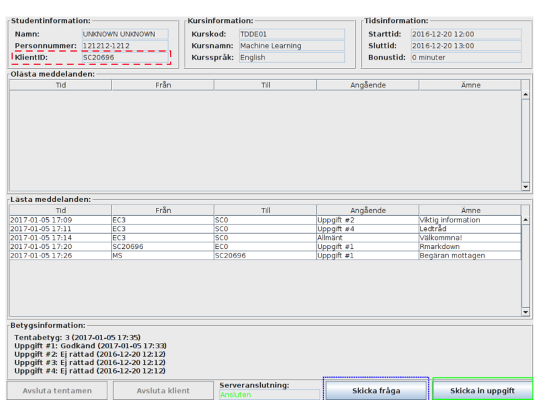
LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Bertil Wegmann

2022-10-20
Bayesian Learning, 6 hp
732A73/732A91/TDDE07

Computer Exam
Bayesian Learning (732A73/732A91/TDDE07), 6 hp

Time:	8-12
Allowable material:	- The allowed material in the folders given_files in the exam system. - Calculator with erased memory
Teacher:	Josef Wilzén. Phone: 0739 – 760354 and through the Communication client.
Exam scores:	Maximum number of credits on the exam: 40.
Grades (732A73/732A91):	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 the 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 (at each sub-problem of the PDF).
Submission starts by clicking the button in the `green` solid rectangle in the figure below.
The submitted PDF file should be named *BayesExam.pdf*
Questions can be asked through the Communication client (`blue` dotted rectangle in the figure below).
Full score requires clear and well motivated answers.



1. DERIVATIONS AND COMPARING POSTERIOR DISTRIBUTIONS (*Total credits: 17p*)

   Problems (a), (b) and (c) should only be solved on **Paper**. Let $x|\theta$, where $x = (x_1, \ldots, x_n)$, be an independent sample from a binomial distribution $\text{Bin}(N, \theta)$, where N is known and

$$f(x_i|\theta) = \binom{N}{x_i} \theta^{x_i} (1-\theta)^{N-x_i}, \; 0 < \theta < 1, \; x_i = 0, 1, 2, \ldots, N, \; i = 1, \ldots, n.$$

   Assume the following conjugate beta prior for θ with parameters $\alpha = 2$ and $\beta = 3$: $p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$.

   (a) *Credits: 3p.* Derive the posterior distribution for θ.

   (b) *Credits*: *2p.* Compute the Bayes point estimate of θ for $x_1 = 13, x_2 = 8, x_3 = 11, x_4 = 7$ and $N = 20$ by assuming the zero-one loss function.

   (c) *Credits: 5p.* Derive the posterior predictive density of a new observation x_{n+1}. An expression is enough, you don't need to recognize any distributional family from the end result.

   (d) *Credits: 4p.* Write a function in R that computes the log posterior distribution of θ as the sum of the log likelihood and the log prior distribution of θ for an independent sample $x|\theta$. Then, use this function to plot the posterior distribution of θ based on the data in 1(b).

   (e) *Credits: 3p.* Use numerical optimization to obtain a normal approximation to the posterior distribution of θ based on the data in 1(b). Use the `lines` command in R to plot this approximate posterior in the same graph as the posterior obtained in 1(d). [Hints: use the argument `lower=0.2` in `optim`, and `method=c("L-BFGS-B")`]. Is the posterior approximation accurate?

2. THE NUMBER OF WOLVES (*Total credits: 11p*)

   Problem (a) should only be solved on **Paper**. The file `Wolves.RData`, which is loaded by the code in `ExamData.R`, contains data on 156 observation weeks of wolves. For each observation week i we have that $y_i =$ *the number of wolves on observation week i*, $x_i = 1$ *if the observation week was in region A* and 0 *if the observation week was in region B.*

   (a) *Credits: 3p.* Consider the model $y_i|\theta \overset{iid}{\sim} \text{Poisson}(\theta)$, where $f(y_i|\theta) = \frac{\theta^{y_i} e^{-\theta}}{y_i!}$, for all observations. Use the following conjugate gamma prior for θ with parameters $\alpha = 40$ and $\beta = 2$: $p(\theta) \propto \theta^{\alpha-1} e^{-2\theta}$. Derive the posterior distribution for θ.

   (b) *Credits: 2p.* Plot the posterior distribution of θ and compute the posterior probability that θ is smaller than 21.

   (c) *Credits: 4p.* Now consider two independent Poisson models, one for the observation weeks in region A and one for the observation weeks in region B. These two models can be written as $y_{A,i}|\theta_A \overset{iid}{\sim} \text{Poisson}(\theta_A)$ and $y_{B,i}|\theta_B \overset{iid}{\sim} \text{Poisson}(\theta_B)$, where the subscripts indicate which observations and parameters that correspond to the observation weeks in region A or B. Use the same prior as in (a) for both θ_A and θ_B. Do posterior inference in both models. Calculate the probability that $y_{B,j} > y_{A,j}$ on a randomly picked future week j.

   (d) *Credits: 2p.* A wolf expert claims that the average number of wolves per week in region B is at least 10 % more than the average number of wolves per week in region A. Do you agree with this statement? Clearly motivate your answer from a Bayesian perspective using your posterior inference in (c).

3. Regression (*Total credits: 12p*) 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

$$
\begin{aligned}
\beta|\sigma^2 &\sim N(0, 10^2\sigma^2 I) \\
\sigma^2 &\sim Inv - \chi^2(1, 4^2).
\end{aligned}
$$

Now, use `BayesLinReg.R` in `ExamData.R` to simulate 10000 draws from the joint posterior distribution of all regression coefficients and the error variance.

(a) *Credits: 3p.* Summarize the posterior of the regression coefficients by the point estimate under the linear loss function, and by 95% equal tail credible intervals for all parameters in β. Interpret the credible interval for the regression coefficient on per capita crime rate by town (`crim`).

(b) *Credits: 1p.* Compute the posterior mean and posterior median of the standard deviation σ.

(c) *Credits: 3p.* Compute 95% equal tail posterior probability intervals for the expected value μ of a new house in 1000$ on a grid of values of per capita crime rate by town (`crim`) and with the remaining covariates as given in the vector `XNewHouse`. The grid of values of `crim` shall span between the lowest and highest values of `crim` with the distance 0.1 between any pair of values on the grid. Plot the lower and upper limits of the posterior probability intervals as a function of `crim`.

(d) *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 be sold for more than 20000$) on the project.

(e) *Credits: 2p.* Let $T(y) = \max_{y_i}$. Calculate the posterior predictive p-value, $\Pr(T(y^{rep}) \geq T(y))$, by simulating from the posterior predictive density $p(T(y^{rep})|y)$ using the model and data in this problem 3. How well can the model replicate the value of the most expensive house in this data?

Good luck!
Best, Bertil