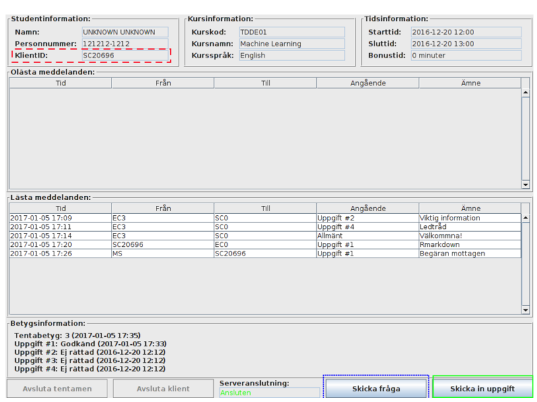
LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Mattias Villani

2017-10-27
Bayesian Learning, 6 hp
732A91/TDDE07

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

Time:	8-12 AM
Allowable material:	- The allowed material in the folder given_files in the exam system. - Calculator with erased memory.
Teacher:	Måns Magnusson. Phone: $070-5889715$ 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. Bayesian inference for proportions data

   The file `ExamData.R` loads a vector `yProp` with $n = 20$ observed proportions (between 0 and 1). Let's assume that the these data can be modelled as independent variables following a symmetric Beta distribution: $y_1, ..., y_{20}|\theta \overset{iid}{\sim} \text{Beta}\ (\theta, \theta)$. Assume the prior $\theta \sim \text{Exp}(1)$.

   (a) *Credits: 3p.* Plot the posterior distribution over the grid `thetaGrid = seq(0.01,15, length =1000)`. Use the previous computations to approximate the optimal Bayes point estimator assuming a zero-one loss.
   Solution: See `Exam732A91_171027_Sol.R`.

   (b) *Credits: 5p.* Now assume a general Beta distribution (possibly non-symmetric): $y_1, ..., y_{20}|\theta \overset{iid}{\sim} \text{Beta}\ (\theta_1, \theta_2)$. Assume that θ_1 and θ_2 are independent apriori with priors $\theta_1 \sim \text{Exp}(1)$ and $\theta_2 \sim \text{Exp}(1)$. Use numerical optimization to obtain a normal approximation of the *joint* posterior distribution of θ_1 and θ_2. You don't need to plot the distribution, just provide its mean and covariance matrix. [Hints: use the argument `lower` in `optim`, and `method=c("L-BFGS-B")`].
   Solution: See `Exam732A91_171027_Sol.R`.

   (c) *Credits: 2p.* Discuss how a Bayesian can determine if the symmetric model in 1a) or the non-symmetric model in 1b) is most appropriate for this data. No need to compute anything here, just discuss.
   Solution: See `Exam732A91_171027_Sol.R`.

2. 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 modelling the response variable `medv` (median 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

$$\beta|\sigma^2 \sim N(0, 10^2\sigma^2 I)$$
$$\sigma^2 \sim Inv - \chi^2(1, 6^2).$$

   (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. Compute 95% Highest Posterior Density (HPD) credible intervals for the β of the variable 'lower status of the population (percent)' (`lstat`), and give a correct Bayesian interpretation. Explain how this interpretation is different from a frequentist confidence interval. [Hint: obtaining the HPD interval is easy in this problem...]
   Solution: See `Exam732A91_171027_Sol.R`.

   (b) *Credits: 3p.* The commune of Boston wants to know the effects on house prices from a policy that lowers `lstat` by 30%. Present this effect to the policy makers by comparing the predictive distribution of the price for house no.9 in the dataset, before and after the policy change. The commune believes that the policy change will almost certainly increase the value for a house like no.9. Do you agree?
   Solution: See `Exam732A91_171027_Sol.R`.

   (c) *Credits: 3p.* The response variable in the Boston housing data is truncated at `medv`=50. So far we have ignored this. Discuss on **Paper** how a Bayesian can proceed if we want to handle this data truncation properly. You do not need to perform the actual analysis, but you should discuss how one can proceed.
   Solution: We need to modify the likelihood for the observations that are truncated. For the truncated observations we replace the normal density $\phi(y_i|\mathbf{x}_i^T\beta, \sigma^2)$ with what we actually have observed: that y is at least 50. This term is $1 - \Phi(y_i|x_i^T\beta, \sigma^2)$, where $\Phi(y_i|x_i^T\beta, \sigma^2)$ is the CDF.

The posterior is no longer analytically available in closed form, but we can always use Metropolis-Hastings as we did in Lab 4.

3. Bayes for the Binomial model

   Let $X_1, ..., X_n|\theta \overset{iid}{\sim} \text{Bin}(m, \theta)$ be binomially distributed data, where m is known. This problem should only be solved on **Paper**.

   (a) *Credits: 3p.* What is the conjugate prior for θ in this model? Derive the posterior distribution for θ.

   Solution: This is essentially a Bernoulli model with mn trials (the permutation part $\begin{pmatrix} m \\ x \end{pmatrix}$ in the Binomial distribution is a constant that does not depend on θ). The conjugate prior is $Beta(\alpha, \beta)$ and the posterior is $Beta(\alpha + \sum_{i=1}^{n} x_i, \beta + nm - \sum_{i=1}^{n} x_i)$.

   (b) *Credits*: 2p. Give a formula for the Bayes point estimator assuming a quadratic loss function.

   Solution: This is the posterior mean, which for a $Beta(\alpha+\sum_{i=1}^{n} x_i, \beta+nm-\sum_{i=1}^{n} x_i)$ distribution is

$$E(\theta|x_1, ..., x_n) = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + nm}$$

   (c) *Credits: 5p.* Derive the predictive distribution of a new observation x_{n+1} from the same model as in 3(a). An expression is enough, you don't need to recognize the distributional family.

   Solution: The predictive distribution of a new observation x_{n+1} is

$$\begin{aligned}
p(x_{n+1}|x_1, ..., x_n) &= \int p(x_{n+1}|\theta)p(\theta|x_1, ..., x_n)d\theta \\
&= \int \begin{pmatrix} m \\ x_{n+1} \end{pmatrix} \theta^{x_{n+1}}(1-\theta)^{m-x_{n+1}} \\
&\times \frac{1}{B(\alpha + \sum_{i=1}^{n} x_i, \beta + nm - \sum_{i=1}^{n} x_i)} \theta^{\sum_{i=1}^{n} x_i + \alpha - 1}(1\text{-}\theta)^{nm - \sum_{i=1}^{n} x_i + \beta - 1} d\theta \\
&\propto \begin{pmatrix} m \\ x_{n+1} \end{pmatrix} \int \theta^{x_{n+1} + \sum_{i=1}^{n} x_i + \alpha - 1}(1-\theta)^{nm - \sum_{i=1}^{n} x_i + m - x_{n+1} + \beta - 1} d\theta \\
&= \begin{pmatrix} m \\ x_{n+1} \end{pmatrix} B(x_{n+1} + \sum_{i=1}^{n} x_i + \alpha, nm + m - \sum_{i=1}^{n} x_i - x_{n+1} + \beta)
\end{aligned}$$

   This distribution can be further simplified, but this is enough for full credit.

4. Prediction and decision

   A city consider building a new bridge at a main road. The weight that the bridge can hold (y) at any given time is related to the build cost (a) as follows

$$y = 10a$$

   The city has collected data on the maximal weight on the road during five different days: $y_1 = 195, y_2 = 191, y_3 = 196, y_4 = 197$ and $y_5 = 189$. Assume the following model for these measurements: $y_1, ..., y_5|\theta \sim N(\theta, \sigma^2)$, where σ^2 is assumed known at $\sigma^2 = 10^2$. Assume a non-informative prior.

   (a) *Credits: 3p.* Simulate 1000 draws from the predictive distribution of the maximal weight on a given future day.

   Solution: The predictive distribution is $N(\bar{y}, \sigma^2/5+\sigma^2) = N(193.6, 10.95^2)$. See `Exam732A91_171027_Sol.R`.

   (b) *Credits: 2p.* Use simulation to approximate the predictive probability that the weight on any of the coming 365 days will exceed 230.

   Solution: See `Exam732A91_171027_Sol.R`.

(c) *Credits: 5p.* The loss function for the building project is

$$L(a,\theta) = \begin{cases} a & \text{if the bridge does not fall down in its first 365 days } (\theta = 0) \\ a + 100 & \text{if the bridge does fall down in its first 365 days } (\theta = 1) \end{cases}$$

(what happens after a year is not of interest to the politicians ...). Compute the optimal build cost (a) using a Bayesian approach.

Solution: We choose a to minimize Posterior expected loss

$$E\left[L(a,\theta)\right] = a \cdot \Pr(\theta = 0|a, y_1, ..., y_5) + (100 + a)\Pr(\theta = 1|a, y_1, ..., y_5)$$

See `Exam732A91_171027_Sol.R`.

Good luck!

Mattias