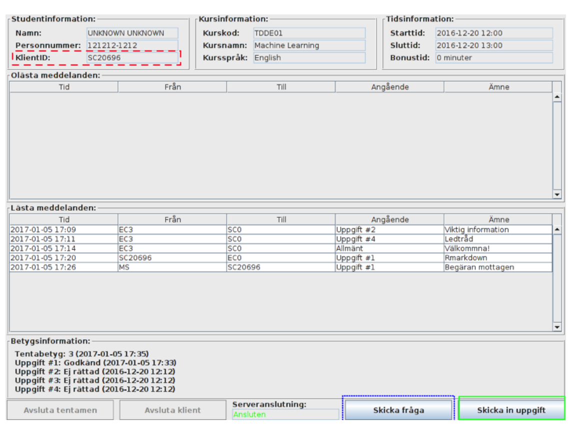
LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Mattias Villani

2017-05-30
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:	Mattias Villani. Phone: 070 – 0895205 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*
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 credit requires clear and well motivated answers.



1. Bayesian inference for the rice distribution

A commonly occuring distribution for positive data is the *Rice distribution*, which we denote by $\text{Rice}(\theta,\psi)$. The PDF for a Rice distribution is of the form

$$p(x|\theta,\psi) = \frac{x}{\psi}\exp\left(\frac{-\left(x^2+\theta^2\right)}{2\psi}\right)\cdot I_0\left(\frac{x\theta}{\psi}\right) \quad \text{for } x > 0.$$

where $\theta \geq 0$ is the location parameter and $\psi > 0$ is related to the variance. $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero, which is implemented in R as `BesselI`. We will assume for simplicity that $\psi = 1$.

(a) Write a function in R that computes the log posterior distribution of θ based on iid observations $\mathbf{x} = (x_1, ..., x_n)$ from $\text{Rice}(\theta, \psi = 1)$. Use that function to plot the posterior distribution of θ for the $n = 10$ observations in the data vector `riceData` in the supplied file `ExamData.R`.

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

(c) Explain on **Paper** how the predictive distribution for a new observation x_{n+1} is computed by integration. You don't need to actually compute the integral, just give a general formula for the predictive distribution. Now, compute the predictive distribution for a new observation x_{n+1} by simulation. You can use the normal approximation of the posterior from 1b). A simulator (`rRice`) for the Rice distribution is provided in the file `ExamData.R`.

2. Modeling count data

The data set `bids` which is loaded by the code in `ExamData.R` contains data on the number of bids in 1000 eBay auctions for collectors coins. Let $x_1, ..., x_n$, for $n = 1000$, denote the data points.

(a) Assume the Poisson model $x_1, ..., x_n|\theta \overset{iid}{\sim} \text{Pois}(\theta)$ for the data, and use the prior $\theta \sim \text{Gamma}(1, 1)$. Compute the posterior distribution for θ and plot it.

(b) Use graphical methods to investigate if the Poisson model fits the data well.

(c) Use the supplied function `GibbsMixPois.R` in the file `ExamData.R` to do Gibbs sampling for a mixture of Poissons model

$$p(x) = \sum_{k=1}^{K} w_k \cdot \text{Pois}(x|\theta_k),$$

where $w_1, ..., w_K$ are the weights (probabilities) of the mixture components (sometimes also called denoted $\pi_1, ..., \pi_K$). $\text{Pois}(x|\theta_k)$ is here used as a shorthand for the probability function (density for a discrete variables) of a Poisson distribution with mean θ_k in the kth mixture component. Use the same $\theta \sim \text{Gamma}(1, 1)$ prior for all the K components, and a uniform prior on the weights $w_1, ..., w_K$. Estimate the mixture of Poissons both with $K = 2$ and $K = 3$. Use `nIter=500` draws, and no burn-in.

(d) Use graphical methods to investigate if the mixture of Poissons with $K = 2$ fits the data well. Note that `GibbsMixPois.R` returns the posterior mean of the mixture density (`GibbsResults$mixDensMean`). Is $K = 2$ enough, or would you recommend $K = 3$?

(e) The number of mixture components, K, is usually unknown. Discuss on **Paper** how a Bayesian could do inference for K. You do not need to compute anything here, just discuss the principles.

3. Regression

`BayesLinReg.R` in the file `ExamData.R` samples from the joint posterior of $\boldsymbol{\beta}$ and σ^2 in the Gaussian linear regression with conjugate prior

$$\boldsymbol{\beta}|\sigma^2 \sim N\left(\boldsymbol{\mu}_0, \sigma^2\Omega_0^{-1}\right)$$
$$\sigma^2 \sim \text{Inv}-\chi^2(\nu_0, \sigma_0^2).$$

(a) The file `cars` which is loaded by the code in `ExamData.R` contains data on 32 cars. For each car we have observations on how many miles that car can travel on a gallon of gasoline (mpg), the weight of the car (weight) and two dummy variables that indicate if the car's engine has four cylinders (sixcyl=0 and eightcyl=0), six cylinders (sixcyl=1 and eightcyl=0) or eigth cylinders (sixcyl=0 and eightcyl=1). The dataframe also contains a column intercept with ones to get an intercept in the model. Now, use `BayesLinReg.R` to sample from the joint posterior distribution in the Gaussian linear regression

$$\text{mpg} = \beta_0 + \beta_1 \cdot \text{weight} + \beta_2 \cdot \text{sixcyl} + \beta_3 \cdot \text{eightcyl} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

Analyze the dataset by simulating 1000 draws from the joint posterior. Use the prior with $\boldsymbol{\mu}_0 = (0,0,0,0)$, $\Omega_0 = 0.01 \cdot I_4$, $\nu_0 = 1$ and $\sigma_0^2 = 36$ (which is the data variance).

i. Plot the marginal distributions of each parameter.

ii. Compute point estimates for each regression coefficient assuming the linear loss function $L(\beta_k, a) = |\beta_k - a|$, where β_k is the kth regression coefficient.

iii. Construct 95% equal tail probability intervals for each parameter and interpret them.

(b) Investigate if the effect on mpg is different in cars with six cylinders compared to cars with 8 cylinders.

(c) Compute by simulation the predictive distribution for a new car 4 cylinders and weight = 3.5.

4. Geometric data and decisions

Let $x_1, ..., x_n|\theta \overset{iid}{\sim} \text{Geometric}(\theta)$. The Geometric distribution has probability function

$$p(x|\theta) = (1-\theta)^x\theta, \text{ for } x = 0, 1, 2, ...,$$

and zero otherwise.

(a) Derive the posterior distribution $p(\theta|x_1, ..., x_n)$ on **Paper** using the conjugate Beta(α, β) prior.

(b) Show on **Paper** that the predictive distribution for a new observation x_{n+1} is of the form

$$p(x_{n+1}|x_1, ..., x_n) \propto \frac{\Gamma(x_{n+1} + \sum_{i=1}^n x_i + \beta)}{\Gamma(x_{n+1} + \sum_{i=1}^n x_i + n + \alpha + \beta + 1)}.$$

(c) Your favorite sports team has had the following result in its first $n = 10$ games of the season (W=won, L=lost): $W, L, L, W, W, L, L, L, W, W$. Assume that the games are independent and that the team has the same chance of winning in every game. Your local bookie has introduced a new game where you win $2^k - 1$ dollars if your team loses the k subsequent games and then wins the $(k+1)$th game. The game costs $2 dollars to play. Should you play it? Use a uniform prior wherever needed. [Hint: one way to solve this problem uses the results from 4b) above.]

Good luck!

Mattias