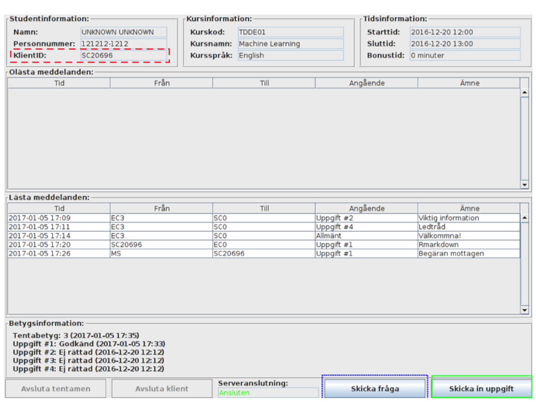
LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Per Sidén

2019-06-04
Bayesian Learning, 6 hp
732A91/TDDE07/732A73

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

Time:	8-12
Allowable material:	- The allowed material in the folders 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. Political election

A political party has performed a Bayesian data analysis before an upcoming election. According to their model, the posterior of the vote share θ that the party will get in the election is Beta$(\sqrt{c}, 20)$-distributed, where c is the amount (in million SEK) that the party spends on the campaign.

(a) *Credits: 2p.* Sample at least `10000` draws from the posterior of θ for the cases when $c = 4$ and $c = 16$. Based on the samples, plot the posterior of θ for both cases.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

(b) *Credits: 2p.* Compute the probability that the party gets at least 10% of the votes for both cases.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

(c) *Credits: 6p.* To evaluate the election results, the leaders of the party have decided on a utility function $u(\theta, c)$ that reflects the value of a specific election outcome θ relative to the campaign cost c

$$u(\theta, c) = 100 + 20 \log(\theta) - c.$$

Assuming the party leaders are Bayesians, how much money c should they spend on the campaign? Base your computations on simulations with at least `10000` draws each. Assume that the maximum campaign budget is 20 million SEK, so consider only values of c on the grid `seq(4,20,0.5)`. Motivate your answer with a figure.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

2. Modeling count data

The dataset `ebay` which is loaded by the code in `ExamData.R` contains data on the number of bids in 100 eBay auctions for collectors coins. Let $x_1, ..., x_n$, for $n = 100$, denote the data points.

(a) *Credits: 4p.* Assume a binomial model $x_1, ..., x_n | N, \theta \overset{iid}{\sim} \text{Bin}(N, \theta)$ for the data, where the parameter $N = 50$ can be considered as known. Use the prior

$$p(\theta) \propto (\theta - 1)^2, \quad \text{for } 0 \le \theta \le 1.$$

Write a function in R that computes the (unnormalized) log posterior density function of θ. Use the function to plot the normalized posterior density function of θ on the interval $0 \le \theta \le 1$ with at least 1000 grid points. Report the (approximate) value of the posterior mode $\hat{\theta}$ based on the computed values needed for the plot.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

(b) *Credits: 3p.* Use the supplied function `GibbsMixPois` in the file `ExamData.R` to do Gibbs sampling for a mixture of Poissons model where each datapoint is modeled as independent with density

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

where $w_1, ..., w_K$ are the weights (probabilities) of the mixture components (sometimes also denoted $\pi_1, ..., \pi_K$). $\text{Pois}(x|\theta_i)$ 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 $\theta_k \sim \text{Gamma}(1, 1)$ prior for all the K components, and a uniform prior on the weights $w_1, ..., w_K$. Set the seed using `set.seed(100)` and run the Gibbs sampling algorithm with $K = 2$ and `nIter=500` draws. Plot trajectories and cumulative means over the iterations for the θ_k parameters, based on the produced samples, which can be found in the output of the function (`thetaSample`). Based on the plots, suggest a sufficient number of initial burn-in iterations, after which the chain has reached its stationary distribution.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

(c) *Credits: 3p.* 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 (`mixDensMean`). Also, use graphical methods to investigate if the binomial model with parameters

$N = 50$ and $\theta = \hat{\theta}$ fits the data well. $\hat{\theta} = 0.1$ can be assumed if a) was not solved. Which model would you recommend and why?
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

3. NORMAL MODEL WITH NORMAL PRIOR

Let $x_1, \ldots x_n | \theta, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$ be normally distributed observations and assume that σ^2 is known. Assume a normal prior $\theta \sim \mathrm{N}\left(\mu_0, \tau_0^2\right)$. The whole problem should be solved on **Paper**.

(a) *Credits: 3p.* Show that the likelihood function can be written as

$$p\left(x_1, \ldots x_n | \theta, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[n\left(\theta - \bar{x}\right)^2 + \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]\right),$$

where $\bar{x}$ is the mean of the observations.
Solution: Due to the independent observations, the likelihood function can be written as

$$\begin{aligned}
p\left(x_1, \ldots x_n | \theta, \sigma^2\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_i - \theta\right)^2\right) \\
&= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right) - \left(\theta - \bar{x}\right)\right)^2\right) \\
&= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 + n\left(\theta - \bar{x}\right)^2 + 2\left(\theta - \bar{x}\right)\underbrace{\left(\sum_{i=1}^{n} x_i - n\bar{x}\right)}_{=0}\right]\right) \\
&= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[n\left(\theta - \bar{x}\right)^2 + \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]\right). \qquad \square
\end{aligned}$$

(b) *Credits: 4p.* Derive the posterior distribution of $\theta | \sigma^2, x_1, \ldots, x_n$. The result in (a) may be used without proof.
Solution: The posterior can be derived as

$$\begin{aligned}
p\left(\theta | x_1, \ldots, x_n\right) &\propto p\left(x_1, \ldots, x_n | \theta\right) p\left(\theta\right) \\
&\propto \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[n\left(\theta - \bar{x}\right)^2 + \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]\right) \cdot \exp\left(-\frac{1}{2\tau_0^2}\left(\theta - \mu_0\right)^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2} n\left(\theta - \bar{x}\right)^2 + \frac{1}{\tau_0^2}\left(\theta - \mu_0\right)^2\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\theta^2\left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right) - 2\theta\left(\frac{n}{\sigma^2}\bar{x} + \frac{\mu_0}{\tau_0^2}\right) + \text{const}\right]\right).
\end{aligned}$$

We identify the normal form of this distribution and by comparing to

$$\exp\left(-\frac{1}{2\tau_n^2}\left(\theta - \mu_n\right)^2\right) = \exp\left(-\frac{1}{2}\left[\theta^2 \frac{1}{\tau_n^2} - 2\theta\frac{\mu_n}{\tau_n^2} + \text{const}\right]\right),$$

and μ_n and τ_n^2 can be identified as

$$\tau_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1}, \quad \mu_n = \tau_n^2\left(\frac{n}{\sigma^2}\bar{x} + \frac{\mu_0}{\tau_0^2}\right) = w\bar{x} + \left(1 - w\right)\mu_0$$

with

$$w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}.$$

So the posterior is $\theta|\sigma^2, x_1, \ldots, x_n \sim \mathrm{N}\left(\mu_n, \tau_n^2\right)$ with μ_n and τ_n^2 defined above. Just providing the expression of the posterior without derivation gives 1.5p.

(c) *Credits: 3p.* For the values $\sigma^2 = 2$, $\mu_0 = 0$, $\tau_0^2 = 1$, $\bar{x} = 1$ and $n = 18$, compute the predictive distribution of a new observation $x_{n+1}|x_1, \ldots, x_n$ given this model.
Solution: According to the slides of Lecture 4, the resulting predictive distribution is normal with mean μ_n and variance $\sigma^2 + \tau_n^2$, so

$$w = \frac{\frac{18}{2}}{\frac{18}{2} + \frac{1}{1}} = \frac{9}{10}, \quad \mu_n = \frac{9}{10}1 + \frac{1}{10}0 = \frac{9}{10}$$
$$\tau_n^2 = \left(\frac{18}{2} + \frac{1}{1}\right)^{-1} = \frac{1}{10}, \quad \sigma^2 + \tau_n^2 = 2.1.$$

The predictive distribution is $x_{n+1}|x_1, \ldots, x_n \sim \mathrm{N}(0.9, 2.1)$.

4. Exponential model with gamma prior

Assume that the lifetimes of cellphones are modeled as exponentially distributed $x_1, \ldots, x_n|\theta \overset{iid}{\sim} \mathrm{Exp}(\theta)$, with the prior $\theta \sim \mathrm{Gamma}(\alpha, \beta)$. Consider the following two models, which have the same exponential likelihood and only differ in the prior parameters:

$$M_1\colon\ \theta \sim \mathrm{Gamma}(2, 1)$$
$$M_2\colon\ \theta \sim \mathrm{Gamma}(10, 10).$$

(a) *Credits: 2p.* Which of these two priors is more informative? Motivate your answer.
Solution: See the code in `Exam732A91_190604_Sol.Rmd`.

(b) *Credits: 4p.* Write a function in `R` that can be used to compute the marginal likelihood for each of the two models for the dataset `cellphones` which is loaded by the code in `ExamData.R`. Use the function to compare the models based on this dataset, assuming the prior probabilities $p(M_1) = p(M_2) = 0.5$. Which model is more probable? [Hint: It may be used that the posterior for the exponential model with the conjugate Gamma prior is $\theta|x_1, \ldots, x_n \sim \mathrm{Gamma}\left(\alpha + n, \beta + \sum_{i=1}^n x_i\right)$.]
Solution: Use Bayes' theorem to write the marginal likelihood as

$$p(x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\theta)\, p(\theta)}{p(\theta|x_1, \ldots, x_n)}.$$

Note that this holds for any value of theta. Then use that

$$p(M_i|x_1, \ldots, x_n) \propto p(x_1, \ldots, x_n|M_i)\, p(M_i).$$

Alternatively, one can use

$$\begin{aligned} p(x_1, \ldots, x_n) &= \int p(x_1, \ldots, x_n|\theta)\, p(\theta)\, d\theta \\ &= \int \prod_{i=1}^n \theta \exp(-\theta x_i) \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)\, d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + n)}{(\beta + n\bar{x})^{\alpha+n}}. \end{aligned}$$

See the code in `Exam732A91_190604_Sol.Rmd`.

(c) *Credits: 4p.* Describe on **Paper** how Bayesian model averaging can be used to compute the posterior predictive density function of a new observation $\tilde{x}$ using both models. Compute the 90% posterior predictive interval of $\tilde{x}$ given the `cellphones` dataset, again assuming the prior probabilities $p(M_1) = p(M_2) = 0.5$. If you did not solve b), you may assume that the models are equally probable.
Solution: The posterior predictive density can be decomposed as

$$p(\tilde{x}|\mathbf{x}) = p(\tilde{x}|\mathbf{x}, M_1)\, p(M_1|\mathbf{x}) + p(\tilde{x}|\mathbf{x}, M_2)\, p(M_2|\mathbf{x}),$$

where $p(M_1|\mathbf{x})$ and $p(M_2|\mathbf{x})$ are the probabilities computed in b) and

$$p(\tilde{x}|\mathbf{x}, M_i) = \int p(\tilde{x}|\theta)\, p(\theta|\mathbf{x}, M_i)\, d\theta$$

is the posterior predictive density for one of the models. See the code in `Exam732A91_190604_Sol.Rmd`.

GOOD LUCK!

MATTIAS AND PER