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

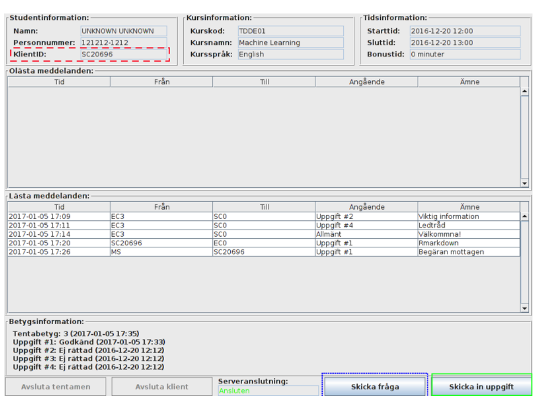
2019-10-31
Bayesian Learning, 6 hp
732A91/TDDE07/732A73

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. Financial decisions

   A large bank has to make a decision every morning whether or not to buy a certain financial option, in order to insure a stock portfolio against financial risks. The bank's utility outcome depending on whether the portfolio value goes up or down is given by the table below. Let $x \sim \text{Bern}(\theta)$ denote the portfolio outcome, with $x = 1$ meaning that the portfolio value goes up and $x = 0$ meaning that the portfolio value goes down.

	Portfolio value goes up	Portfolio value goes down
Buy the option	30	-10
Don't buy the option	90	-120

   (a) *Credits: 2p.* Assume that it is known that $\theta = 0.6$. Compute the Bayesian decision, whether the bank should buy the option or not.
   Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

   (b) *Credits: 5p.* Now assume that θ is unknown, and that there are 100 independent observations $x_1, ..., x_{100} \overset{iid}{\sim} \text{Bern}(\theta)$ from previous days. Assume that the bank has a $\text{Beta}(3, 2)$ prior for θ. Derive the predictive distribution for x_{101} on **Paper**. Assume that the portfolio value went up in 62 of the 100 previous days.
   [Hints: It may be used that the posterior distribution for θ is $\text{Beta}(\alpha + s, \beta + f)$-distributed, with α and β being prior parameters and s and f being the number of observed successes and failures in the Bernoulli trials. The Gamma function has the property $\Gamma(y + 1) = y\Gamma(y)$.]
   Solution:

$$
\begin{aligned}
p(x_{n+1}|x_{1:n}) &= \int p(x_{n+1}|\theta)\, p(\theta|x_{1:n})\, d\theta \\
&= \int \theta^{x_{n+1}} (1-\theta)^{1-x_{n+1}} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s)\Gamma(\beta+f)} \theta^{\alpha+s-1} (1-\theta)^{\beta+f-1}\, d\theta \\
&= \frac{\Gamma(105)}{\Gamma(65)\Gamma(40)} \int \theta^{x_{101}+65-1} (1-\theta)^{1-x_{101}+40-1}\, d\theta \\
&= \frac{\Gamma(105)}{\Gamma(65)\Gamma(40)} \frac{\Gamma(x_{101}+65)\Gamma(1-x_{101}+40)}{\Gamma(106)} \\
&= \frac{\Gamma(x_{101}+65)\Gamma(1-x_{101}+40)}{105\Gamma(65)\Gamma(40)},
\end{aligned}
$$

   using $\Gamma(y + 1) = y\Gamma(y)$. So,

$$
p(x_{101} = 1|x_{1:100}) = \frac{\Gamma(66)\Gamma(40)}{105\Gamma(65)\Gamma(40)} = \frac{65}{105} = \frac{13}{21}
$$

   and conversely $p(x_{101} = 1|x_{1:100}) = 8/21$, which gives the conclusion

$$
x_{n+1}|x_{1:n} \sim \text{Bern}(13/21).
$$

   (c) *Credits: 3p.* Compute the Bayesian decision for day 101 based on the information in (b). Solutions based on simulation and solutions based on the answer from (b) are both accepted.
   Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

2. Speed limits

   The dataset `Traffic` which is loaded by the code in `ExamData.R` contains data from an experiment with speed limits on the motorway that was performed in Sweden in 1961-1962. Typing `?Traffic` will present the help file with information on the variables. Let y_i denote the number of traffic incidents in day i.

(a) *Credits: 4p.* Consider the model $y_i \overset{iid}{\sim} \text{Poisson}(\theta)$ for all observations. Use a conjugate prior for θ with $E(\theta) = Var(\theta) = 20$. Plot the posterior density function of θ on the grid defined by `seq(18,24,.01)`. Compute the posterior probability that θ is smaller than 21.
Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

(b) *Credits: 4p.* Now instead consider two independent Poisson models, one for the days when there was a speed limit (`limit==''yes''`) and one for the days when there was no speed limit (`limit==''no''`). These two models can be written as $y_{L,i} \overset{iid}{\sim} \text{Poisson}(\theta_L)$ and $y_{N,i} \overset{iid}{\sim} \text{Poisson}(\theta_N)$ where the subscripts indicate which observations and parameters that correspond to the days with the limit or with no limit. Use the same prior as in (a) for both θ_L and θ_N. Do posterior inference in both models and answer the question whether a speed limit leads to a lower amount of accidents. Clearly motivate your answer using a Bayesian statistical language.
Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

(c) *Credits: 2p.* A politician claims that the experiment proves that introducing a speed limit decreases the number of accidents by at least 15%. Do you agree with this statement? Clearly motivate your answer using a Bayesian statistical language.
Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

3. Gibbs sampling

Let $x \sim \text{Bin}(\nu, \pi)$ be a single observation from the Binomial distribution. Assume independent priors for the unknown parameters $\nu \sim \text{Poisson}(\lambda)$ and $\pi \sim \text{Beta}(\alpha, \beta)$.

(a) *Credits: 2p.* Show on **Paper** that the full conditional posterior for π is $\pi|x, \nu \sim \text{Beta}(\alpha + x, \beta + \nu - x)$.
Solution:

$$
\begin{aligned}
p(\pi|x,\nu) &\propto p(x|\nu,\pi)\, p(\pi|\nu) \\
&\propto \binom{\nu}{x} \pi^x (1-\pi)^{\nu-x} \pi^{\alpha-1} (1-\pi)^{\beta-1} \\
&\propto \pi^{\alpha+x-1} (1-\pi)^{\beta+\nu-x-1}.
\end{aligned}
$$

Since this has the form of a Beta-distribution with parameters $\alpha + x$ and $\beta + \nu - x$ it is clear that $\pi|x, \nu \sim \text{Beta}(\alpha + x, \beta + \nu - x)$.

(b) *Credits: 3p.* Show on **Paper** that the full conditional posterior for ν can be written as proportional to the following expression

$$
p(\nu|x,\pi) \propto \frac{[\lambda(1-\pi)]^{\nu-x}}{(\nu-x)!}, \quad \text{for } \nu = x, x+1, x+2, \ldots
$$

Solution:

$$
\begin{aligned}
p(\nu|x,\pi) &\propto p(x|\nu,\pi)\, p(\nu|\pi) \\
&\propto \binom{\nu}{x} \pi^x (1-\pi)^{\nu-x} \frac{\lambda^\nu}{\nu!}, \quad \text{for } x = 0, 1, \ldots, \nu \\
&\propto \frac{\nu!}{(\nu-x)!x!} (1-\pi)^{\nu-x} \frac{\lambda^\nu}{\nu!} \lambda^{-x} \\
&\propto \frac{[\lambda(1-\pi)]^{\nu-x}}{(\nu-x)!}, \quad \text{for } \nu = x, x+1, x+2, \ldots
\end{aligned}
$$

(c) *Credits: 5p.* Make simulations from the joint posterior of ν and π using Gibbs sampling. Assume $x = 20$, $\lambda = 30$, $\alpha = 2$ and $\beta = 2$. Draw at least 2000 posterior samples and use 500 samples for burnin. Plot the marginal posterior of each variable based on the simulations. Evaluate the convergence of the Gibbs sampler by suitable graphical methods.
[Hint: It may be used that the transformed variable $z = \nu - x$ has the full conditional $z|x, \pi \sim$

Poisson($\lambda(1-\pi)$), so to sample from $p(\nu|x,\pi)$ one can first sample $z_s \sim p(z|x,\pi)$ and then set $\nu_s = z_s + x$.]

Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

4. Car stopping distances

The `cars` data contains 50 recorded measurements of the speed of cars and the distances taken to stop when breaking. The dataset is loaded by the `ExamData.R` file. Typing `?cars` will present the help file with information on the variables. Consider the following simple regression model for the stopping distances y_i

$$
\begin{aligned}
y_i =& \alpha + \beta x_i + \varepsilon_i \\
\varepsilon_i \overset{iid}{\sim}& \mathrm{N}\left(0, \sigma^2\right),
\end{aligned}
$$

where x_i is the speed. The supplied stan model `LinRegModel` in `ExamData.R` can be used to simulate from the posterior of α, β and σ^2, using non-informative priors for α and β. The prior for σ^2 is given by $\sigma^2 \sim \text{Inv-}\chi^2\left(\nu = 5, s^2 = 100\right)$.

(a) *Credits: 4p.* Use the supplied stan model to do Bayesian inference. Draw at least 2000 posterior samples and use 500 samples for burnin. Produce a figure that contains a scatter plot of the data, and overlay a curve for the mean of the posterior predictive distribution $p(\tilde{y}|\tilde{x}, y)$ of a new observation $\tilde{y}$ given values of $\tilde{x}$ in the range $[0, 25]$. Also, overlay curves corresponding to a 90% equal tail interval for the same posterior predictive distribution given values of $\tilde{x}$ in the range $[0, 25]$.

Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

(b) *Credits: 2p.* Compute a 95% equal tail credible interval for α. Give a real-world interpretation of the interval, and discuss how one could change the model in order to make the values of α more realistic.

Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

(c) *Credits: 4p.* When inspecting a scatter plot of the data, it seems like the variation in stopping distances increases with the speed. With this motivation, implement a different regression model using Stan which has *heteroscedastic* variance. This means that each observation should have its own error variance σ_i^2, which depends on the speed x_i according to

$$
\sigma_i^2 \overset{iid}{\sim} \text{Inv-}\chi^2\left(\nu = 5, s^2 = \left(\exp\left(\gamma + \phi x_i\right)\right)^2\right),
$$

where γ and ϕ are unknown parameters. Reproduce the same figure as in (a), with a scatter plot of the data and the mean and 90% interval of the posterior predictive distribution, using the new model. Assume non-informative priors for γ and ϕ. Compare the models in (a) and (c) based on the figures, in terms of data fit.

Solution: See the code in `Exam732A91_191031_Sol.Rmd`.

Good luck!

Mattias and Per