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

2020-08-19
Bayesian Learning, 6 hp
732A91/TDDE07/732A73

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

Time:	14-18
Allowable material:	All aids are permitted during the exam with the following exceptions: • You may not communicate with anyone else except the responsible teacher • You may not look at the solutions of any other students.
Teacher:	Per Sidén. Contact via email at per.siden@liu.se
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:
For full information, see the document *Info take home exam.pdf.*
You should submit your solutions via the Submission in LISAM.
Your submitted solutions must contain the following two things:

- A PDF named *ComputerSol.pdf* containing your computer-based solutions
- Files named *PaperSol* containing your hand-written solutions (when asked to answer on **Paper**)

Full score requires clear and well motivated answers.

---

1. PATIENTS

   The number of visiting patients at a health center in a given month is assumed to be Poisson distributed with unknown parameter θ. A Bayesian analysis has resulted in a posterior distribution for θ based on historical data D such that $\theta|D \sim N(10000, 500^2)$. If simulation is used in your solutions, use at least `10000` samples.

   (a) *Credits: 2p.* Compute a 90% equal-tail credible interval for θ.
   Solution: See `Exam732A91_200819_Sol.Rmd.`

   (b) *Credits: 3p.* Compute a 90% equal-tail prediction interval for the number of patients in the next month.
   Solution: See `Exam732A91_200819_Sol.Rmd.`

   (c) *Credits: 5p.* The health centre is considering hiring more doctors to increase its capacity. Currently there are 20 doctors employed. Each doctor can treat 500 patients each month, so the current capacity is 10000 patients. If the number of patients exceeds the capacity in a given month, the health centre has to pay overtime salary which is expensive. The health centre has the following loss function

   $$L(p, d) = (d + 0.04\max(p - 500d, 0))^2,$$

   where d is the number of doctors employed, p is the number of patients in a month, and $\max(a, b)$ is the function that returns the largest value of a and b. Should the health centre hire more doctors, and if so, how many, according to the Bayesian decision? Assume that all doctors work full-time.
   Solution: See `Exam732A91_200819_Sol.Rmd.`

2. REGRESSION

   The dataset `muscle` which is loaded by the code in `ExamData.R` contains experimental data on strips of heart tissue from rats. Each row of data contains the shortening of a strip (stored in the variable `Length`) after being dipped in a calcium chloride solution of a certain concentration (stored in the variable `Conc`). Type `?muscle` for more information about this data. Consider the following two linear regression models

   $$\begin{aligned} M_1 : &\quad \text{Length} = \beta_1 \cdot \text{Conc} + \varepsilon \\ M_2 : &\quad \text{Length} = \beta_0 + \beta_1 \cdot \text{Conc} + \beta_2 \cdot \text{Conc}^2 + \varepsilon, \end{aligned}$$

   For each model ε is iid. Gaussian measurement noise with variance σ^2, and β_0, β_1, β_2 and σ^2 are unknown parameters. For each model, use the function `BayesLinReg` to sample from the joint posterior distribution of the parameters, using 5000 draws. Analyze the dataset by simulating 5000 draws from the joint posterior. Use the prior with $\boldsymbol{\mu}_0$ as a zero vector, $\Omega_0 = 0.001 \cdot I$, $\nu_0 = 1$ and $\sigma_0^2 = 10$. [Hint: The input `X` to `BayesLinReg` must be a matrix, you can use `as.matrix()`.]

   (a) *Credits: 2p.* Compute the 95% equal tail credible intervals for all β-parameters for both models.
   Solution: See `Exam732A91_200819_Sol.Rmd.`

   (b) *Credits: 5p.* Produce two figures, one for each model, with a scatter plot of the data and overlay a curve for the posterior predictive mean for values of Conc on the grid given by `seq(0,5,0.01)`. Also overlay curves representing the 95% posterior prediction bands for the different values of Conc.
   Solution: See `Exam732A91_200819_Sol.Rmd.`

   (c) *Credits: 3p.* On **Paper**, briefly discuss advantages and disadvantages with the two models for this dataset. In particular, discuss which model gives the best Length predictions for new measurements at previously unobserved Conc values and whether the models can be used for drawing conclusions about whether larger Conc values leads to larger Length values.
   Solution:
   M_1 :
   $+$ A linear model is simple and in general less prone to overfitting.

- The linear model is too simple for this dataset which appears to have a non-linear relationship between the variables. The predictive intervals are too wide/wrongly located for many Conc values and the residuals are clearly non-Gaussian.

M_2 :
+ The quadratic model better captures the non-linear relationship in the data and gives a reasonable fit, at least for $Conc < 4$, where the data is. The predictive intervals look reasonable.
- The predictions for $Conc > 4$ do not look reasonable.

For making predictions for new measurements at previously unobserved Conc values, M_2 is clearly better for $Conc < 4$, but for larger values none of the models can be trusted. For drawing conclusions about whether larger Conc values leads to larger Length values, none of the models can be used satisfyingly. Since M_1 is linear, one could naively interpret the fact that the credible interval for β_1 is greater than zero to a positive association of the variables, however, the assumptions of linearity, Gaussian residuals etc. does not hold. M_2 is hard to interpret directly for this purpose and some other model is necessary.

3. DELIVERY TIMES

The file `delivery` which is loaded by the code in `ExamData.R` contains data on delivery times (in days) for products from an online sales company. The data $x_1, \ldots, x_n$ are assumed to be independent and Weibull distributed with parameters k and λ according to the following density function

$$p\left(x_i|k,\lambda\right) = \frac{k}{\lambda} x_i^{k-1} \exp\left(-x_i^k/\lambda\right), \;\; x_i > 0,\, k > 0,\, \lambda > 0.$$

Note that this parameterisation is different than the one in R and in the table of distributions. It is further assumed that the value of k is known and that the prior for λ is Inverse-Gamma(α, β).

(a) *Credits: 3p.* On **Paper**, show that the posterior for λ is Inverse-Gamma$(\alpha + n, \beta + \sum_{i=1}^n x_i^k)$.
Solution: From the table of distributions we get the prior density function

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} \exp(-\beta/\lambda).$$

and the likelihood function is

$$p\left(x|k,\lambda\right) = \prod_{i=1}^n p(x_i|k,\lambda) = \prod_{i=1}^n \frac{k}{\lambda} x_i^{k-1} \exp\left(-x_i^k/\lambda\right).$$

Bayes' theorem gives

$$\begin{aligned} p(\lambda|x) &\propto p(x|k,\lambda)p(\lambda) \\ &\propto \prod_{i=1}^n \frac{1}{\lambda} \exp\left(-x_i^k/\lambda\right) \cdot \lambda^{-(\alpha+1)} \exp(-\beta/\lambda) \\ &\propto \lambda^{-(\alpha+n+1)} \exp\left(-\frac{1}{\lambda}\left[\beta + \sum_{i=1}^n x_i^k\right]\right), \end{aligned}$$

which can be identified as the form of an Inverse-Gamma distribution. So $\lambda|x \sim$ Inverse-Gamma$(\alpha + n, \beta + \sum_{i=1}^n x_i^k)$.

(b) *Credits: 3p.* Assume that $\alpha = 2$ and $\beta = 2$. Consider the following three values for k: 0.5, 1.5 and 2.5. For each value of k, draw samples from the posterior of λ and use the samples to compute the posterior mode of λ using the `density()` function. [Hint: To sample from the Inverse-Gamma distribution, one can use that if $Y \sim$ Gamma(a, b) then $1/Y \sim$ Inverse-Gamma(a, b).]
Solution: See `Exam732A91_200819_Sol.Rmd`.

(c) *Credits: 4p.* Make a figure that contains a histogram of the data and three Weibull densities for the data based on the three values for k and the corresponding posterior modes for λ. Based on the figure, which value for k would you say gives the best fit to the data? Motivate your answer.
Solution: See `Exam732A91_200819_Sol.Rmd.`

4. More delivery times

Consider the same data and model as in Assignment 3. That is, the file `delivery` which is loaded by the code in `ExamData.R` contains data on delivery times (in days) for products from an online sales company. The data $x_1, \ldots, x_n$ are assumed to be independent and Weibull distributed with parameters k and λ according to the following density function

$$p(x_i|k,\lambda) = \frac{k}{\lambda} x_i^{k-1} \exp\left(-x_i^k/\lambda\right), \quad x_i > 0,\, k > 0,\, \lambda > 0.$$

Note that this parameterisation is different than the one in R and in the table of distributions. It is further assumed that the value of k is known and that the prior for λ is $\lambda \sim \text{Inverse-Gamma}(\alpha, \beta)$.

(a) *Credits: 4p.* On **Paper**, show that the marginal likelihood of the data can be expressed as

$$p(x_1, \ldots, x_n) = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \frac{\beta^\alpha k^n \prod_{i=1}^n x_i^{k-1}}{\left(\beta + \sum_{i=1}^n x_i^k\right)^{\alpha+n}}.$$

Solution: Using the posterior in 3a, the marginal likelihood can be computed as

$$\begin{aligned}
p(x) &= \frac{p(x|\lambda)p(\lambda)}{p(\lambda|x)} \\
&= \frac{\prod_{i=1}^n \frac{k}{\lambda} x_i^{k-1} \exp\left(-x_i^k/\lambda\right) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} \exp(-\beta/\lambda)}{\frac{\left[\beta+\sum_{i=1}^n x_i^k\right]^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{-(\alpha+n+1)} \exp(-\left[\beta + \sum_{i=1}^n x_i^k\right]/\lambda)} \\
&= \frac{k^n \prod_{i=1}^n x_i^{k-1} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+n+1)} \exp(-\left[\beta + \sum_{i=1}^n x_i^k\right]/\lambda)}{\frac{\left[\beta+\sum_{i=1}^n x_i^k\right]^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{-(\alpha+n+1)} \exp(-\left[\beta + \sum_{i=1}^n x_i^k\right]/\lambda)} \\
&= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \frac{\beta^\alpha k^n \prod_{i=1}^n x_i^{k-1}}{\left[\beta + \sum_{i=1}^n x_i^k\right]^{\alpha+n}}.
\end{aligned}$$

The marginal likelihood can also be computed through $p(x) = \int_0^\infty p(x|\lambda)p(\lambda)d\lambda$.

(b) *Credits: 4p.* Now, instead view k as an unknown parameter, and assume the prior $k \sim \text{Exp}(1)$. Write a function in R that computes the log unnormalized posterior distribution of k. Use that function to plot the normalized posterior distribution of k on the interval $(0, 3)$. Assume that $\alpha = 2$ and $\beta = 2$.
Solution: The likelihood of the data given k is identical to the marginal likelihood in 3a. See `Exam732A91_200819_Sol.Rmd.`

(c) *Credits: 2p.* Compute a 95% equal-tail credible interval for k based on the posterior in (b).
Solution: See `Exam732A91_200819_Sol.Rmd.`

Good luck!

Mattias and Per