LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Bertil Wegmann

2021-10-21
Bayesian Learning, 6 hp
732A73/732A91/TDDE07

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

Time:	8-12
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:	Bertil Wegmann. Contact via email at bertil.wegmann@liu.se
Exam scores:	Maximum number of credits on the exam: 40.
Grades (732A73/732A91):	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 distance 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
- A PDF named *PaperSol* containing your hand-written solutions (when asked to answer on **Paper**)

Full score requires clear and well motivated answers.

---

1. Derivations and comparing posterior distributions (*Total credits: 21p*)

   Problems (a), (b) and (c) should only be solved on **Paper**. Let $x|\theta$, where $x = (x_1, \ldots, x_n)$, be an independent sample from a Poisson distribution and assume that the prior for θ has the following density:

   $$p(\theta) = \frac{1}{16}\theta^2 e^{-0.5\theta}, \theta > 0.$$

   (a) *Credits: 3p.* Derive the posterior distribution for θ.

   (b) *Credits: 2p.* Compute the Bayes point estimate of θ for $n = 15, \sum_{i=1}^{n} x_i = 75$, by assuming the quadratic loss function.

   (c) *Credits: 5p.* Derive the marginal likelihood for an independent sample $x|\theta$ from the model above. An expression is enough, you don't need to recognize any distributional family from the end result.

   (d) *Credits: 4p.* Write a function in R that computes the log posterior distribution of θ for an independent sample $x|\theta$. Then, use this function to plot the posterior distribution of θ based on the data in 1(b).

   (e) *Credits: 3p.* Use numerical optimization to obtain a normal approximation to the posterior distribution of θ based on the data in 1(b). Use the `lines` command in R to plot this approximate posterior in the same graph as the posterior obtained in 1(d). [Hints: use the argument `lower=3` in `optim`, and `method=c("L-BFGS-B")`]. Is the posterior approximation accurate?

   (f) *Credits: 4p.* Assume that a person named Gunnar told you that the maximum value of an independent Poisson distributed sample $x|\theta$ of size $n = 15$ is $T(x) = \max_{x_i} = 14$. Calculate the posterior predictive p-value, $\Pr(T(x^{rep}) \geq T(x))$, by simulating from the posterior predictive density $p(T(x^{rep})|x)$ using the model and data in this problem 1. Is it reasonable to think that the maximum value of 14 from Gunnar originates from the Poisson distribution in this problem?

2. Logit regression problems (*Total credits: 9p*)

   The file `Disease.RData`, which is loaded by the code in `ExamData.R` contains data on 32 patients in a country. For each patient i we have observed $y_i = 1$ *if patient i has a certain disease* and 0 otherwise, $x_{1i} =$ *measured value on a cell property for patient i* and $x_{2i} = 1$ *if patient i has a certain property* and 0 otherwise. The variable x_1 has been standardized to mean 0 and variance 1. The data matrix also contains a column *const* with ones to include an intercept in the model. Now, use `BayesLogitReg.R` in `ExamData.R` to sample from the joint posterior distribution of the model parameters in the following logistic regression model:

   $$\Pr(y_i = 1 \mid x_1, x_2) = p_i = \frac{\exp(\beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i})}{1 + \exp(\beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i})}$$

   Simulate 10000 draws from the joint posterior. Use the following prior distribution for $\beta = (\beta_0, \beta_1, \beta_2)$:

   $$\beta \sim N(0, 4^2 I_3)$$

   (a) *Credits: 1p.* Compute the 90% equal tail credible interval for β_1 and interpret it.

   (b) *Credits: 1p.* Compute the posterior probability that $\beta_2 > 0$ and interpret it in terms of the effect from variable x_2 on p.

   (c) *Credits: 1p.* Compute the joint posterior probability that both $\beta_1 > 0$ and $\beta_2 > 0$.

   (d) *Credits: 3p.* Consider a patient j with $x_{1j} = x_{2j} = 0$. Plot the posterior distribution of p_j and compute the posterior probability that $p_j > 0.5$ for this patient.

   (e) *Credits: 3p.* Consider a patient k with $x_{2k} = 1$. Compute 95 % equal tail posterior probability intervals for p_k on a grid of values of x_1. The grid of values of x_1 shall span between the lowest and highest values of x_1 with the distance 0.01 between any pair of values on the grid. Plot the lower and upper limits of the posterior probability intervals as a function of x_1.

3. NORMAL MODELING (*Total credits: 10p*)

Problem (a) should only be solved on **Paper**. Let $x_1, \ldots, x_n | \mu$ be an independent sample from a normal distribution with unknown mean μ and variance 50. Assume that a Bayesian data analysis of μ with $n = 10$ observations and $\bar{x} = 90$ resulted in the following posterior: $\mu | x_1, \ldots, x_n \sim N\left(92, 2^2\right)$.

(a) *Credits: 3p.* Derive the normal prior distribution that have been used for this Bayesian data analysis.

(b) *Credits: 2p.* Simulate draws from the posterior predictive density of a new observation x_{n+1} and plot the posterior predictive distribution.

(c) *Credits: 5p.* Assume that x_i is the revenue for company i under a given time period in a certain industry. The utility function for a company in the industry spending c MSEK on advertisements is given by:

$$U(\mu, c) = 60 + \sqrt{c} \log(\mu) - c.$$

How much money c should a company spend on advertisements from a Bayesian perspective? Use at least 10000 draws in your simulations. Motivate your answer with a figure.

GOOD LUCK!
BEST, BERTIL