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

2021-06-03
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. Customers' choice of brands (*Total credits: 11p*)

   From a random sample of $n = 100$ customers, a store has noted for a certain product with three possible brands A, B, C that 38 customers bought brand A, 27 customers bought brand B, and 35 bought brand C. Let θ_i be the probability that a random customer, who buys the product, chooses brand i, where $i = A, B, C$.

   In problems (a), (b) and (c) you assume the prior $\theta_A \sim Beta\,(\alpha = 16, \beta = 24)$ and only consider that each customer chooses brand A or not brand A.

   (a) *Credits: 3p.* Compute the posterior probability that $\theta_A > 0.4$ and plot the posterior distribution of $1 - \theta_A$.

   (b) *Credits: 2p.* Compute a 95% equal tail credible interval for the ratio $\frac{1-\theta_A}{\theta_A}$ and interpret it.

   (c) *Credits: 2p.* Compute the marginal likelihood for the model $x_1, \ldots, x_{100}|\theta_A \overset{iid}{\sim} Bernoulli\,(\theta_A)$.

   (d) *Credits: 4p.* Assume a Dirichlet prior for $(\theta_A, \theta_B, \theta_C)$ such that $E\,[\theta_A] = E\,[\theta_B] = E\,[\theta_C] = \frac{1}{3}$ and where the prior information is equivalent to a random sample of 60 customers. Compute the posterior probability that $\theta_A > \theta_C$.

2. Derivations and comparing posterior distributions (*Total credits: 17p*)

   Problems (a), (b) and (c) should only be solved on **Paper**. Let $x_1, \ldots, x_n|\theta$ be an independent sample from a distribution with the following density:

   $$f\,(x_i|\theta) = 2\theta x_i e^{-\theta x_i^2},\ x_i > 0, \theta > 0\ i = 1, \ldots, n.$$

   Assume that the conjugate prior for θ (as a special case of the gamma distribution) is following the exponential distribution with $E\,[\theta] = 2$.

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

   (b) *Credits: 2p.* Compute the Bayes point estimate of θ for $n = 13, \sum_{i=1}^{n} x_i^2 = 2.8$, by assuming the zero-one loss function.

   (c) *Credits: 5p.* Derive the posterior predictive density of a new observation x_{n+1}. 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 $x_1, \ldots, x_n|\theta$. Then, use this function to plot the posterior distribution of θ based on the data in 2(b).

   (e) *Credits: 3p.* Use numerical optimization to obtain a normal approximation to the posterior distribution of θ based on the data in 2(b). Use the `lines` command in R to plot this approximate posterior in the same graph as the posterior obtained in 2(d). [Hints: use the argument `lower=0.1` in `optim`, and `method=c("L-BFGS-B")`]. Is the posterior approximation accurate?

3. Regression problems (*Total credits: 12p*)

The file `UniversityEntrance.RData`, which is loaded by the code in `ExamData.R` contains data on 180 high school students taking a university entrance exam. For each student i we have observed $y_i =$ *test result for student* i, $x_{1i} =$ *verbal IQ for student* i, $x_{2i} =$ *socioeconomic status for student* i, $x_{3i} = 1$ *if student* i *belongs to high school B* (0 otherwise), $x_{4i} = 1$ *if student* i *belongs to high school C* (0 otherwise). All variables (except the dummy variables) have 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 `BayesLinReg.R` in `ExamData.R` to sample from the joint posterior distribution of the model parameters in the following Gaussian linear regression

$$y = \mu + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\mu = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_1 \cdot x_3 + \beta_6 \cdot x_1 \cdot x_4$$

Simulate 10000 draws from the joint posterior. Use the following prior distributions for $\beta = (\beta_0, \beta_1, \ldots, \beta_6)$ and σ^2:

$$\beta|\sigma^2 \sim N(0, 5^2\sigma^2 I_7)$$
$$\sigma^2 \sim Inv - \chi^2(1, 2^2)$$

(a) *Credits: 2p.* Compute the posterior mean and 95% equal tail credible intervals for all parameters in β and interpret the interval for β_1.

(b) *Credits: 1p.* Compute the posterior median of the standard deviation σ.

(c) *Credits: 4p.* Investigate if the effect on y from x_1 is different for students in *high school B* compared to students in *high school C*.

(d) *Credits: 3p.* Compute 90% equal tail posterior probability intervals for the expected test results μ on a grid of values of x_1 for students with $x_2 = 0.5$ who are neither in high school B nor in high school C. 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.

(e) *Credits: 2p.* Compute by simulation the posterior predictive distribution of y for a new student (not in the data) in *high school B* with $x_1 = 0.4$ and $x_2 = 1$ and plot the result.

Good luck!
Best, Bertil