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

2020-06-04
Bayesian Learning, 6 hp
732A91/TDDE07/732A73

Take Home Computer Exam
Bayesian Learning (732A91/TDDE07/732A73), 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:	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 in via the Submission in LISAM.
Your submitted solutions must contain the following three 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**)
- A file named *Honorary* containing your hand-written honorary statement.

Full score requires clear and well motivated answers.

---

1. Binomial distribution

   Let x be a single observation from the binomial distribution $x|N,\theta \sim \text{Binomial}(N,\theta)$, where $N = 50$.

   (a) *Credits: 3p.* Assume a uniform prior, $\theta \sim U(0,1)$. On **Paper**, show that the posterior distribution is $\theta|x \sim \text{Beta}(x+1, 51-x)$.
   Solution: Bayes' theorem gives

   $$p(\theta|x) \propto p(x|\theta)p(\theta) \propto \theta^{x}(1-\theta)^{50-x} \cdot 1 \propto \theta^{x+1-1}(1-\theta)^{51-x-1},$$

   which can be identified as the form of a Beta-distribution with parameters $\alpha = x+1$ and $\beta = 51-x$.

   (b) *Credits: 5p.* It is known that the true value of θ is close to 0.5, so instead assume the uniform prior $\theta \sim U(0.3, 0.7)$. Use R to compute the (normalized) posterior distribution of θ when the observed value is $x = 33$. Plot the posterior distribution of θ on the interval $(0,1)$. Plot the posterior distribution when the prior in a) is used in the same figure.
   Solution: See `Exam732A91_200604_Sol.Rmd`.

   (c) *Credits: 2p.* Compute the posterior probability that $\theta < 0.5$ when the observed value is $x = 33$ for both priors.
   Solution: See `Exam732A91_200604_Sol.Rmd`.

2. Probit regression

   The file `titanic` which is loaded by the code in `ExamData.R` contains the information about 1316 passengers on the ship Titanic that sunk in the Atlantic ocean in April 1912. This was a big disaster. Descriptions of the variables presented in the table below. A passenger could travel in first, second or third class.

Variable	Data type	Notation	Meaning	Role
Survived	Binary	y	Whether or not the passenger survived	Response
Intercept	1	x_1	Constant to the intercept	Feature
Adult	Binary	x_2	Whether or not the passenger was an adult	Feature
Man	Binary	x_3	Whether or not the passenger was a man	Feature
Class1	Binary	x_4	Whether or not the passenger travelled in first class	Feature
Class2	Binary	x_5	Whether or not the passenger travelled in second class	Feature

   (a) *Credits: 3p.* Consider the probit regression

   $$\Pr(y = 1|\mathbf{x}) = \Phi\left(\mathbf{x}^T\boldsymbol{\beta}\right),$$

   where y is the binary variable with $y = 1$ if the passenger survived and $y = 0$ otherwise. $\mathbf{x}$ is a 5-dimensional vector containing the five features (including a one for the constant term that models the intercept) and $\boldsymbol{\beta}$ is the corresponding vector of coefficients. The provided function `BayesProbReg()` in `ExamData.R` performs Gibbs sampling for probit regression using data augmentation as described in Lecture 7. Use the function to produce `1000` samples from the *joint* posterior distribution of $\boldsymbol{\beta}$. Use the prior $\boldsymbol{\beta} \sim \mathcal{N}(0, \tau^2 I)$, with $\tau = 50$. Plot the marginal posterior distribution of each parameter.
   Solution: See `Exam732A91_200604_Sol.Rmd`.

   (b) *Credits: 2p.* Assume the linear loss function $L(\beta_k, a) = |\beta_k - a|$, where β_k is the kth regression coefficient, and compute point estimates for each parameter.
   Solution: See `Exam732A91_200604_Sol.Rmd`.

   (c) *Credits: 2p.* Compute the posterior probability that $\beta_2 + \beta_5 > 0$ and interpret this probability.
   Solution: See `Exam732A91_200604_Sol.Rmd`.

   (d) *Credits: 3p.* The Gibbs sampling method is asymptotically exact, but can be time consuming. On **Paper**, mention two different methods for approximating the posterior of $\boldsymbol{\beta}$, and briefly describe the methods.

Solution: Method 1: Normal approximation $\boldsymbol{\beta}|\mathbf{y},\mathbf{X} \sim N(\hat{\boldsymbol{\beta}}, \mathbf{J}_{\mathbf{y}}^{-1}(\hat{\boldsymbol{\beta}}))$, where $\hat{\boldsymbol{\beta}}$ is the posterior mode and $\mathbf{J}$ is the observed Hessian at the mode. These can be obtained by numerical optimization of the posterior.
Method 2: Variational inference. Use the data augmentation (latent variable) formulation as in the Gibbs sampler, with augmented variable $\mathbf{u}$. Using a mean field (factorized) variational posterior $q(\mathbf{u},\boldsymbol{\beta}) = q_{\mathbf{u}}(\mathbf{u})q_{\boldsymbol{\beta}}(\boldsymbol{\beta})$, which results in a normal approximation $q_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ with variational parameters that can be optimized by minimizing the KL divergence.

3. Olympic Poisson regression

A country is interested in the connection between how much money it spends on its Olympic program (giving athletes better opportunities to train) and the number of Olympic medals it wins. Let y denote the number of medals won and x is the logarithm of the amount it spends on the Olympic program in million dollars. In the last four Olympic Games the country has won 5, 3, 17 and 8 medals after spending respectively 20, 20, 50 and 40 million on the Olympic program (corresponding to x-values $\log 20, \log 20, \log 50$ and $\log 40$). Consider the Poisson regression model

$$y_i|\beta \sim \text{Poisson}\left[\exp\left(x_i\beta\right)\right],\ i = 1, ..., 4,$$

where β is an unknown regression parameter. Note that this model has no intercept.

(a) *Credits: 4p.* Compute a normal approximation for the posterior of β by using numerical optimization, assuming the prior $\beta \sim N(1, (1/10)^2)$. Report the posterior mean and standard deviation of β using the approximation.
Solution: See `Exam732A91_200604_Sol.Rmd`.

(b) *Credits: 6p.* For the next Olympic Games, the country needs to decide how much it should spend on the Olympic program. The country considers decreasing the spendings to 20 million or keeping it at 40 million. It has a loss function described by

$$L(y_5, x_5) = 4 + \frac{\exp(x_5)}{50} - \sqrt{y_5},$$

where x_5 is the log spendings and y_5 the number of medals won in the next Olympic Games. Compute the Bayesian decision, using your posterior approximation in a) and simulation with at least `10000` random samples.
Solution: See `Exam732A91_200604_Sol.Rmd`.

4. Classification

A certain type of fish is known to have an unequal gender balance. Females are more common than males. 20 random fish are caught and examined. Out of these 20 fish, 16 turn out to be female and 4 turn out to be male.

(a) *Credits: 3p.* Assume these observations are iid Bernoulli distributed with unknown parameter θ. Assume that the prior for θ is Beta(2, 2). A new fish is caught. What is the predictive probability that the new fish is a female? Answer on **Paper** and clearly motivate your answer.

Solution: This is Bernoulli trials with a Beta prior with $\alpha = 2, \beta = 2, s = 16, f = 4$, so the posterior is $\theta|x \sim Beta(18, 6)$, assuming $x_1, \ldots x_{20}$ are the binary data with $x_i = 1$ meaning fish i is female. Furthermore, the predictive distribution can be computed as in the solution to Math Exercise 2.2a, showing it is $Bernoulli(\frac{2+16}{2+2+20}) = Bernoulli(0.75)$. Another way to see this is that the predictive distribution is obviously Bernoulli since x_{21} is binary with

$$E(x_{21}|x_{1:20}) = E(E(x_{21}|\theta)|x_{1:20}) = E(\theta|x_{1:20}) = \frac{18}{18+6} = 0.75,$$

using the law of iterated expectations and the mean of a Beta distribution. **Answer:** The predictive probability is 0.75.

(b) *Credits: 3p.* The length and weight of each fish are also measured. The average length and weight of the female and male fish are reported in the table below. Assume that the measurements are independent and follow normal distributions with unknown mean parameters μ_{FL} (female length), μ_{ML} (male length), μ_{FW} (female weight) and μ_{MW} (male weight), and known standard deviation parameters $\sigma_{FL} = \sigma_{ML} = 2$ (length) and $\sigma_{FW} = \sigma_{MW} = 50$ (weight). Assume uniform priors for the mean parameters. What is the predictive distribution of the length of the new fish if it is assumed that the fish is a male? Answer on **Paper** and clearly motivate your answer.

	Average length (cm)	Average weight (g)
Female	14	300
Male	12	280

Solution: Assume without loss of generality that the first 16 observations are the females. Let y denote length. We then have 4 iid observations $y_{17:20}$ from $N(\mu_{ML}, \sigma^2_{ML})$, with $p(\mu_{ML}) \propto c$ and $\sigma^2_{ML} = 4$. According to Lecture 4, the predictive distribution in this case is $y_{21}|y_{17:20}, x_{21} = 0 \sim N(\bar{y}_{17:20}, \sigma^2_{ML}(1 + 1/4)) = N(12, 5)$.

(c) *Credits: 4p.* Construct a Naive Bayes classifier based on the data in a) and b). What is the predictive probability that the new fish is a female given that it is 10 cm long and weighs 250 g?

Solution: Let z denote weight. Naive Bayes assumes $p(y_{21}, z_{21}|x_{21}, data) = p(y_{21}|x_{21}, data)p(z_{21}|x_{21}, data)$. Bayes theorem gives

$$\begin{aligned} p(x_{21}|y_{21}, z_{21}, data) &\propto p(y_{21}, z_{21}|x_{21}, data)p(x_{21}|data) \\ &\propto p(y_{21}|x_{21}, data)p(z_{21}|x_{21}, data)p(x_{21}|data), \end{aligned}$$

where the first two factors can be computed as i b) and the last factor as in a). See the code in `Exam732A91_200604_Sol.Rmd`.

Good luck!

Mattias and Per