

## 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.

The screenshot shows the Communication Client interface with the following sections:

- Studentinformation:** Namn: UNKNOWN UNKNOWN, Personnummer: 121212-1212, KlientID: SC20696 (highlighted in a red dashed box).
- Kursinformation:** Kurskod: TDDE01, Kursnamn: Machine Learning, Kurspråk: English.
- Tidsinformation:** Starttid: 2016-12-20 12:00, Sluttid: 2016-12-20 13:00, Beroettid: 0 minuter.
- Olästa meddelanden:** A table with columns Tid, Från, Till, Ämålände, and Ämne.
- Senast meddelanden:** A table with columns Tid, Från, Till, Ämålände, and Ämne, showing a list of messages.
- Betygsinformation:** Tentabeltyg: 3 (2017-01-05 17:30), Uppgift #1: Godkänd (2016-12-20 17:30), Uppgift #2: Ej rättad (2016-12-20 12:12), Uppgift #3: Ej rättad (2016-12-20 12:12), Uppgift #4: Ej rättad (2016-12-20 12:12).
- Buttons:** Avsluta tentamen, Avsluta klient, Serveranslutning: Ansluten, Skicka fråga (blue dotted box), Skicka in uppgift (green solid box).

## 1. POLITICAL ELECTION

A political party has performed a Bayesian data analysis before an upcoming election. According to their model, the posterior of the vote share  $\theta$  that the party will get in the election is  $\text{Beta}(\sqrt{c}, 20)$ -distributed, where  $c$  is the amount (in million SEK) that the party spends on the campaign.

- (a) *Credits: 2p.* Sample at least 10000 draws from the posterior of  $\theta$  for the cases when  $c = 4$  and  $c = 16$ . Based on the samples, plot the posterior of  $\theta$  for both cases.
- (b) *Credits: 2p.* Compute the probability that the party gets at least 10% of the votes for both cases.
- (c) *Credits: 6p.* To evaluate the election results, the leaders of the party have decided on a utility function  $u(\theta, c)$  that reflects the value of a specific election outcome  $\theta$  relative to the campaign cost  $c$

$$u(\theta, c) = 100 + 20 \log(\theta) - c.$$

Assuming the party leaders are Bayesians, how much money  $c$  should they spend on the campaign? Base your computations on simulations with at least 10000 draws each. Assume that the maximum campaign budget is 20 million SEK, so consider only values of  $c$  on the grid `seq(4, 20, 0.5)`. Motivate your answer with a figure.

## 2. MODELING COUNT DATA

The dataset `ebay` which is loaded by the code in `ExamData.R` contains data on the number of bids in 100 eBay auctions for collectors coins. Let  $x_1, \dots, x_n$ , for  $n = 100$ , denote the data points.

- (a) *Credits: 4p.* Assume a binomial model  $x_1, \dots, x_n | N, \theta \stackrel{iid}{\sim} \text{Bin}(N, \theta)$  for the data, where the parameter  $N = 50$  can be considered as known. Use the prior

$$p(\theta) \propto (\theta - 1)^2, \quad \text{for } 0 \leq \theta \leq 1.$$

Write a function in R that computes the (unnormalized) log posterior density function of  $\theta$ . Use the function to plot the normalized posterior density function of  $\theta$  on the interval  $0 \leq \theta \leq 1$  with at least 1000 grid points. Report the (approximate) value of the posterior mode  $\hat{\theta}$  based on the computed values needed for the plot.

- (b) *Credits: 3p.* Use the supplied function `GibbsMixPois` in the file `ExamData.R` to do Gibbs sampling for a mixture of Poissons model where each datapoint is modeled as independent with density

$$p(x) = \sum_{k=1}^K w_k \cdot \text{Pois}(x | \theta_k),$$

where  $w_1, \dots, w_K$  are the weights (probabilities) of the mixture components (sometimes also denoted  $\pi_1, \dots, \pi_K$ ).  $\text{Pois}(x | \theta_i)$  is here used as a shorthand for the probability function (density for a discrete variables) of a Poisson distribution with mean  $\theta_k$  in the  $k$ th mixture component. Use the same  $\theta_k \sim \text{Gamma}(1, 1)$  prior for all the  $K$  components, and a uniform prior on the weights  $w_1, \dots, w_K$ . Set the seed using `set.seed(100)` and run the Gibbs sampling algorithm with  $K = 2$  and `nIter=500` draws. Plot trajectories and cumulative means over the iterations for the  $\theta_k$  parameters, based on the produced samples, which can be found in the output of the function (`thetaSample`). Based on the plots, suggest a sufficient number of initial burn-in iterations, after which the chain has reached its stationary distribution.

- (c) *Credits: 3p.* Use graphical methods to investigate if the mixture of Poissons with  $K = 2$  fits the data well. Note that `GibbsMixPois.R` returns the posterior mean of the mixture density (`mixDensMean`). Also, use graphical methods to investigate if the binomial model with parameters  $N = 50$  and  $\theta = \hat{\theta}$  fits the data well.  $\hat{\theta} = 0.1$  can be assumed if a) was not solved. Which model would you recommend and why?

### 3. NORMAL MODEL WITH NORMAL PRIOR

Let  $x_1, \dots, x_n | \theta, \sigma^2 \stackrel{iid}{\sim} N(\theta, \sigma^2)$  be normally distributed observations and assume that  $\sigma^2$  is known. Assume a normal prior  $\theta \sim N(\mu_0, \tau_0^2)$ . The whole problem should be solved on [Paper](#).

- (a) *Credits: 3p.* Show that the likelihood function can be written as

$$p(x_1, \dots, x_n | \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \left[ n(\theta - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x})^2 \right]\right),$$

where  $\bar{x}$  is the mean of the observations.

- (b) *Credits: 4p.* Derive the posterior distribution of  $\theta | \sigma^2, x_1, \dots, x_n$ . The result in (a) may be used without proof.
- (c) *Credits: 3p.* For the values  $\sigma^2 = 2$ ,  $\mu_0 = 0$ ,  $\tau_0^2 = 1$ ,  $\bar{x} = 1$  and  $n = 18$ , compute the predictive distribution of a new observation  $x_{n+1} | x_1, \dots, x_n$  given this model.

### 4. EXPONENTIAL MODEL WITH GAMMA PRIOR

Assume that the lifetimes of cellphones are modeled as exponentially distributed  $x_1, \dots, x_n | \theta \stackrel{iid}{\sim} \text{Exp}(\theta)$ , with the prior  $\theta \sim \text{Gamma}(\alpha, \beta)$ . Consider the following two models, which have the same exponential likelihood and only differ in the prior parameters:

$$M_1: \theta \sim \text{Gamma}(2, 1)$$

$$M_2: \theta \sim \text{Gamma}(10, 10).$$

- (a) *Credits: 2p.* Which of these two priors is more informative? Motivate your answer.
- (b) *Credits: 4p.* Write a function in R that can be used to compute the marginal likelihood for each of the two models for the dataset `cellphones` which is loaded by the code in `ExamData.R`. Use the function to compare the models based on this dataset, assuming the prior probabilities  $p(M_1) = p(M_2) = 0.5$ . Which model is more probable? [Hint: It may be used that the posterior for the exponential model with the conjugate Gamma prior is  $\theta | x_1, \dots, x_n \sim \text{Gamma}(\alpha + n, \beta + \sum_{i=1}^n x_i)$ .]
- (c) *Credits: 4p.* Describe on [Paper](#) how Bayesian model averaging can be used to compute the posterior predictive density function of a new observation  $\tilde{x}$  using both models. Compute the 90% posterior predictive interval of  $\tilde{x}$  given the `cellphones` dataset, again assuming the prior probabilities  $p(M_1) = p(M_2) = 0.5$ . If you did not solve b), you may assume that the models are equally probable.

GOOD LUCK!

MATTIAS AND PER