LINKÖPING UNIVERSITY
Division of Statistics and Machine Learning
Department of Computer and Information Science
Mattias Villani

Bayesian Learning
Advanced level course

## Mathematical Exercises 1

---

Try to solve the problems before class. Don't worry if you fail, the important thing is trying.
You should not hand in any solutions.
This part of the course is not obligatory and is not graded.

---

1. What matters is the likelihood

   (a) Assume that you want to investigate the proportion ($\theta$) of defective items manufactured at a production line. Your colleague takes a random sample of 30 items. There were three defective items in the sample. Assume a uniform prior for $\theta$. Compute the posterior of $\theta$.

   (b) Your colleague now tells you that he did not decide on the sample size before the sampling was performed. His sampling plan was to keep on sampling items until he had found three defective ones. It just happened that the 30th item was the third one to be defective. Redo the posterior calculation, under the new sampling scheme. Compare the results with Problem 1(a). [Hint: negative binomial distribution]

2. Normal with a normal is normal

   (a) Let $x_1, ..., x_n \overset{iid}{\sim} N(\theta, \sigma^2)$ and assume that $\sigma^2$ is known. Assume a uniform prior:

   $$p(\theta) \propto c.$$

   Derive the posterior distribution of $\theta$.

   (b) Assume now a normal prior

   $$\theta \sim N(\mu_0, \tau_0^2).$$

   Derive the posterior distribution of $\theta$.

3. IT ENDS WITH AN ABNORMAL TWIST

   (a) Let $x_1, ..., x_{10} \overset{iid}{\sim} N(\theta, 1)$. Let the sample mean be $\bar{x} = 1.873$. Assume that $\theta \sim N(0, 5)$ apriori. Compute the posterior distribution of $\theta$.

   (b) Assume now that you have a second sample $y_1, ..., y_{10} \overset{iid}{\sim} N(\theta, 2)$, where $\theta$ is the same quantity as in 3a. The sample mean in this second sample is $\bar{y} = 0.582$. Compute the posterior distribution of $\theta$ using both samples (the $x$'s and the $y$'s) under the assumption that the two samples are indepedent.

   (c) You have now managed to obtain a third sample $z_1, ..., z_{10} \overset{iid}{\sim} N(\theta, 3)$, with mean $\bar{z} = 1.221$. Unfortunately, the measuring device for this latter sample was defective: any measurement above 3 was recorded as 3. There were two such measurements. Compute the posterior distribution based on all three samples ($x, y$ and $z$). [Hint: in this case the posterior distribution is not a known distribution (it is not normal for example). It is enough to give an expression for the (unnormalized) posterior. You can also plot this over a grid on your computer, if you like.]

4. SOLVING THIS PROBLEM MAY TAKE EXPONENTIALLY LONG TIME

   (a) Let $x_1, ..., x_n \overset{iid}{\sim} Expon(\theta)$. We use the parametrization of the exponential distribution where if $X \sim Expon(\theta)$ then $E(X) = 1/\theta$.

   (b) Show that the conjugate prior for the exponential model is $\theta \sim Gamma(\alpha, \beta)$. Derive the posterior distribution for $\theta$

Have fun!

- Mattias

LINKÖPING UNIVERSITY
Division of Statistics and Machine Learning
Department of Computer and Information Science
Mattias Villani and Per Sidén

Bayesian Learning
Advanced level course

## Mathematical Exercises 2

Try to solve the problems before class. Don't worry if you fail, the important thing is trying.
You should not hand in any solutions.
This part of the course is not obligatory and is not graded.

1. TAU-CHI.

   (a) Let $x_1, ..., x_n \overset{iid}{\sim} N(\theta, \sigma^2)$. Assume that $\theta$ is known, but $\sigma^2$ unknown. Derive the posterior distribution for $\sigma^2$. Use the conjugate prior.

   (b) Assume that $\theta = 1$ and that you have observed the data $x_1 = 0.6, x_2 = 3.2, x_3 = 1.2$. Compute the posterior of $\sigma^2$ based on these three data points. Use a prior with very little information (it is up to you how to define little information).

2. FEEL THE BERN.

   (a) Let $x_1, ..., x_n \overset{iid}{\sim} Bern(\theta)$, with a $Beta(\alpha, \beta)$ prior for $\theta$. Derive the predictive distribution for $x_{n+1}$.

   (b) You need to decide if you bring your umbrella during your daily walk. It has rained on two days during the last ten days, and you assess those ten days to be representative also for the weather today, the 11th day. Your utility for the action-state combinations are given in the table below. Assume a $Beta(1, 1)$ prior for $\theta$. Compute the Bayesian decision.

   (c) How sensitive is your decision in (b) to the changes in the prior hyperparameters, $\alpha$ and $\beta$?

| | Rainy | Sunny |
|---|---|---|
| Bring umbrella | 10 | 20 |
| Leave umbrella | -50 | 50 |

3. Campaign or no campaign - that is the question.

(a) Let $x_i$ be the number of sales of a product on month $i$. Let $x_1, ..., x_n \overset{iid}{\sim} N(\theta, \sigma^2)$ be the (approximate) distribution for the sales, and let $\theta \sim N(200, 50^2)$ a priori. Assume that $\sigma^2 = 25^2$ and that we have observed $n = 5$ and $\bar{x} = 320.4$. Compute the predictive distribution for $x_6$.

(b) The company has the choice of performing a marketing campaign for their product. The marketing campaign costs $300 and is believed to increase sales by 20% compared to when no campaign is performed. The company sells the product for $p = 10$ dollar and the cost of producing the product is $q = 5$ dollar. There are no fixed production costs. Assume that the company's utility is described by $U(y) = 1 - \exp(-y/1000)$, where $y$ is the total profit from sales in the next month. Should the company perform the marketing campaign? [Hint: the expected value of the exponential function of a normal random variable $S \sim N(\mu, \sigma^2)$ is $E(\exp(S)) = \exp(\mu + \sigma^2/2)$.]

4. Predictive distribution for a Poisson model

(a) Do Exercise 13(a) in Chapter 2 of the course book. That is, assume that the number of fatal accidents on scheduled airline flights each year are independent with a Poisson($\theta$) distribution. Set a prior distribution for $\theta$ and determine the posterior distribution based on the data from 1976 through 1985, given below. Under this model, give a 95% predictive interval for the number of fatal accidents in 1986. You can use the normal approximation to the gamma and Poisson or compute using simulation.

24, 25, 31, 31, 22, 21, 26, 20, 16, 22

Have fun!

- Mattias

LINKÖPING UNIVERSITY
Division of Statistics and Machine Learning
Department of Computer and Information Science
Mattias Villani and Per Sidén

2018-04-11
Bayesian Learning
Advanced level course

## Mathematical Exercises 3

---

Try to solve the problems before class. Don't worry if you fail, the important thing is trying.
You should not hand in any solutions.
This part of the course is not obligatory and is not graded.

---

1. Fill in the blanks
   (a) Show that the full conditional posterior of $I_i$ on Lecture 7, Slide 22, is correct.
   (b) On Lecture 7, Slide 26, I argue that one can simulate from the joint posterior distribution of the regression coefficients, $\beta$, the noise variance $\sigma^2$ and the regularization/shrinkage hyperparameter $\lambda$ using Gibbs sampling. In particular, I claim that the full conditional posterior of $\lambda$ is Gamma distributed. Derive this full conditional posterior of $\lambda$. [Hint: start by writing up the expression for the joint posterior of $\beta$, $\sigma^2$ and $\lambda$ using the Tattoo-version of Bayes Theorem. The full conditional posterior of $\lambda$ is proportional to this expression.]

2. Frequentist meltdown or Bayesian breakdown?
   (a) Let $x_1, ..., x_n \overset{iid}{\sim} Uniform(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Let $\hat{\theta} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be an estimator of $\theta$. Derive an expression for the (repeated) sampling variance of $\hat{\theta}$.
   (b) Derive the posterior distribution for $\theta$ assuming a uniform prior distribution. [Hint: Here it absolutely crucial to think about the support for the data distribution. Once you have observed some data, some $\theta$ values are no longer possible. I strongly suggest that you plot some imaginary data on the real line and plot the data distribution in the same graph for some made up values of $\theta$. Just to make you think in the right direction.]
   (c) Assume that you have observed three data observations: $x_1 = 1.1, x_2 = 2.09, x_3 = 1.4$. What would a frequestist conclude about $\theta$? What would a Bayesian conclude? Discuss.

3. Who doesn't want to be Normal?

(a) Let $x_1, ..., x_n \overset{iid}{\sim} \text{Bern}(\theta)$ and $\theta \sim \text{Beta}(\alpha, \beta)$ a priori. Find the posterior mode of $\theta$.

(b) Approximate the posterior distribution of $\theta$ by a normal distribution.

(c) Assume now that you have the data $n = 6$ and $s = 1$. Plot the true posterior distribution and the normal approximation in the same graph. Assume a uniform prior for $\theta$.

(d) Redo the previous exercise, but this time with twice the data size: $n = 12$ and $s = 2$.

4. Naive doctors

(a) Three diseases (A,B and C) have very common symptoms and are therefore hard to distinguish between for a doctor. A medical company has developed two different tests (T1 and T2) to discriminate between the three diseases. A training data from $n = 20$ patients was collected to learn a predictive model that can be used to classify a patient into disease A-C on the basis of the results from both T1 and T2. $n_A = 5$ of the patients had disease A, $n_B = 5$ of the patients had disease B and $n_C = 10$ of the patients had disease C. The table below gives the mean measurement in each patient group for both tests. The test measurements can be assumed to follow a normal distribution with variance $\sigma^2 = 1$ for all patient groups, and for both tests. Develop a Naive Bayes classifier based on this training data. You can assume uniform priors in any place you needs a prior. Make a prediction for a new patient with measurement 1.3 on T1 and 4.2 on T2.

| | $\bar{X}_1$ | $\bar{X}_2$ |
|---|---|---|
| Disease A | 1.2 | 2.1 |
| Disease B | 1.4 | 3.5 |
| Disease C | 0.7 | 4.7 |

Have fun!

LINKÖPING UNIVERSITY
Division of Statistics and Machine Learning
Department of Computer and Information Science
Mattias Villani and Per Sidén

2018-04-27
Bayesian Learning
Advanced level course

## Mathematical Exercises 4

---

Try to solve the problems before class. Don't worry if you fail, the important thing is trying.
You should not hand in any solutions.
This part of the course is not obligatory and is not graded.

---

1. Bernoulli meets Laplace
   (a) Let $x_1, ..., x_n \overset{iid}{\sim} Bernoulli(\theta)$ and assume the prior $\theta \sim Beta(\alpha, \beta)$. Derive the marginal likelihood for this model.
   (b) Compute the marginal likelihood of the model in a) using the Laplace approximation.
   (c) Is this approximation accurate if $\alpha = \beta = 1$ and you have observed $s = 6$ success in $n = 10$ trials?

2. Fill in the blanks - again
   (a) Derive the marginal likelihood for the Poisson model with Gamma prior at the end of Slide 7 at Lecture 10.

3. Pareto
   (a) Let $x_1, \ldots, x_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$. Let $\theta \sim \text{Pareto}(\alpha, \beta)$, that is
   $$p(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}},\ \theta \geq \beta.$$
   Show that this is a conjugate prior to this Uniform model and derive the posterior for $\theta$. [Hint: Don't forget to include an indicator function when you write up the likelihood function. The Uniform$(0, \theta)$ distribution is zero for outcomes larger than $\theta$.]
   (b) Derive the predictive distribution of $x_{n+1}$ given $x_1, \ldots, x_n$. [Hint: It is wise to break up the integrals in two parts.]

Have fun!

- Mattias