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, β, the noise variance σ^2 and the regularization/shrinkage hyperparameter λ using Gibbs sampling. In particular, I claim that the full conditional posterior of λ is Scaled Inv-χ^2. Derive this full conditional posterior of λ. [Hint: start by writing up the expression for the joint posterior of β, σ^2 and λ using the Tattoo-version of Bayes Theorem. The full conditional posterior of λ 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 θ. Derive an expression for the (repeated) sampling variance of $\hat{\theta}$.
   (b) Derive the posterior distribution for θ 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 θ 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 θ. 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 θ? 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 θ.

(b) Approximate the posterior distribution of θ 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 θ.

(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!