Matias Quiroz and Mattias Villani
Division of Statistics and Machine Learning
Dept. of Computer and Information Science
Linköpings universitet
matias.quiroz@liu.se

2016-04-19

Bayesian Learning, 6 hp

Computer lab 2

You can use any programming language for the labs, but my hints, help and solutions will be in R.
You are supposed to work and submit your labs in pairs, but do make sure that both of you are contributing. Submit your solutions by Lisam no later than **April** 29 at **midnight**. You should submit the report (pdf only) and an executable file containing your code.

1. *Multinomial model with Dirichlet prior*
   In May 2014, Statistics Sweden conducted a Party Preference Survey (PSU) asking Swedish voters to state which party they would vote if there were a general election in September 2014. A total of 9085 voters were asked and 4757 voters responded. The choices offered were 1. M - Moderate Party; 2. C - Center Party; 3. FP - Liberal People's Party; 4. KD - Christian Democrats; 5. MP - Green Party; 6. S - Swedish Social Democratic Party; 7. V - Left Party; 8. SD - Sweden Democrats; 9. Others - Other parties. The data is summarized in four age-categories (excluding non-response)

Age	M	C	FP	KD	MP	S	V	SD	Others	Total
18-29	208	45	46	35	110	189	34	53	88	808
30-49	403	58	74	42	146	413	127	93	57	1413
50-64	370	51	60	47	67	401	59	61	15	1131
65+	383	89	86	65	45	567	74	79	17	1405
Total	1364	243	266	189	368	1570	294	286	177	4757

   Assume that voters of each age group are independent random samples from the population. Model the data with four different multinomial distributions, one for each age group:

$$y_{i1}, ..., y_{ik} \sim \text{Multinomial}(y_i; \theta_{i1},..., \theta_{ik}) \qquad \text{for} \quad i = 1, ..., 4; \quad k = 9$$

   where y_{ij} is the number of voters in age group i that responded that they would vote for party j, $y_i = \sum_{j=1}^{k} y_{ij}$, and θ_{ij} is the probability that a randomly selected voter in age group i states that he/she would vote for party j. Our prior distribution for $(\theta_{i1}, ..., \theta_{ik})$ is the same for all age groups, which is Dirichlet$(\alpha_1, ..., \alpha_k)$, and the age groups are assumed to be independent a priori. Based on the election results in 2010, we specify the prior hyperparameters as

α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
30	6	7	6	7	30	6	6	2

(a) Show that the posterior distribution of $(\theta_{i1}, ..., \theta_{ik})$ for the ith voting group is also Dirichlet$(\alpha_1 + y_{i1}, ..., \alpha_k + y_{ik})$.

(b) *Analyzing the impact of voting age.* Calculate the posterior distribution of $(\theta_{i1}, ..., \theta_{ik})$ separately for each age group. Compare the voting behaviour between age groups in a suitable graph.

(c) For the voters of age 18-29, what is the posterior probability that the Red-Greens (S, MP, V) will win against the Alliance (M, C, FP, KD)? Now compute the same probability for all four age groups.
[Hint: Use simulation methods to compute $\Pr(\sum_{j=5}^{7} \theta_{1j} > \sum_{j=1}^{4} \theta_{1j} | \mathrm{y})$ for each age group.]

(d) Who actually wins the elections clearly depend on the number of voters in each age group that actually vote on the day of the election. Let Y_i, $i = 1, 2, 3, 4$, be the number of voters in each age group that actually vote on the election day. Suppose that $Y_1, ..., Y_4 \sim$ Multinomial$(Y, 0.2, 0.3, 0.3, 0.2)$, where $Y = 6,300,000$ is the size of the population that are allowed to vote. Now, given the total number of voters in age group i, Y_i, we model the vote counts for each party in this age group as $Y_{i1}, ..., Y_{ik} \sim$ Multinomial$(Y_i; \theta_{i1}, ..., \theta_{ik})$. What is your prediction that the Red-Greens will win the election?
[Hint: 1. For $i = 1, ..., 4$, simulate a draw of $\theta_{i1}, ..., \theta_{ik}$ and Y_i; 2. Simulate $Y_{i1}, ..., Y_{ik}$; 3. Evaluate the probability $\Pr(\sum_{j=5}^{7} \sum_{i=1}^{4} Y_{ij} > \sum_{j=1}^{4} \sum_{i=1}^{4} Y_{ij})$.]

2. *Linear and polynomial regression*
The data set `JapanTemp.dat` contains daily temperatures (in Celcius degrees) at some Japanese location over the course of a year. The response variable is *temp* and the covariate is

$$time = \frac{\text{the number of days since beginning of year}}{365}.$$

The task is to perform a Bayesian analysis of a quadratic regression

$$temp = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2 + \varepsilon,\ \varepsilon \overset{iid}{\sim} N(0, \sigma^2).$$

(a) *Determining the prior distribution of the model parameter.* Use the conjugate prior for the linear regression model. Your task is to set the prior hyperparameters μ_0, Ω_0, ν_0 and σ_0^2 to sensible values. You may not be an expert in Japanese temperatures, and I don't expect any deep expert knowledge, but do come up with something. You may simplify by assuming that Ω_0 is a diagonal matrix, if you want.
[Hint: it may be useful as a preliminary exploratory step to use the `lm()` command. The command `lm(y ~ x + I(x^2))` fits a quadratic model using plain least squares]

(b) *Check if your prior from a) is sensible.* One way to check if a suggested prior is reasonable is to simulate draws from the joint prior of all parameters and for every draw compute the regression curve. This gives a collection of regression curves, one for each draw from the prior. Do the curves look reasonable? If not, change the prior hyperparameters until the collection of prior regression curves do agree with your prior beliefs about the regression curve.
[Hint: the R package `mvtnorm` will be handy.]

(c) Write a program that *simulates from the joint posterior distribution* of β_0, β_1,β_2 and σ^2. Try it out on the model in a).
[Hint: the R package `mvtnorm` will be handy.]

(d) It is of interest to locate the day with the highest expected temperature (that is, the *time* where $E(temp|time)$ is maximal). Let's call this value $\tilde{x}$. Use the simulations in d) to simulate from the *posterior distribution of the day with highest temperature*, $\tilde{x}$. [Hint: the regression curve is a quadratic. You can find a simple formula for $\tilde{x}$ given β_0, β_1 and β_2.]

(e) Say now that you want to *estimate a polynomial model of order* 7, but you are worried that higher order terms may not be needed, and you worry about over-fitting. Suggest a suitable prior that mitigates this potential problem.

MAY BAYES BE WITH YOU!