Take home exam for Bayesian Inference in Theory and Practise
2007-03-27

Instructions: You are not allowed to get help from others inside or outside the class. Solutions should either be e-mailed to me (preferred) (mattias.villani@riksbank.se) or be placed in my 'postfack' at the department of statistics **no later than 17.00 on April 9, 2007**.

1.a. Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. Your collegue takes a random sample of 30 items. Three were defective in the sample. Assume a uniform prior for θ. Compute the posterior of θ.

1.b. Your collegue 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 happend that the 30th item was the third one to be defective. Redo the posterior calculculation, this time under the new sampling scheme. Discuss the results.

1. c. Repeat 1a and 1b, but this time using Jeffreys' prior. Discuss the results.

1. d. Do a Bayesian test of the hypothesis: $H_0 : \theta = 0.05$. Assume a uniform prior, and two other priors (to learn about the sensitivity with respect to the prior).

2.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 θ.

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

2.c. You have now managed to obtain a third sample $z_1, ..., z_{10} \overset{iid}{\sim} N(\theta, 3)$, with mean $\bar{z} = 1.221$. Unfortunatly, the measuring device for this latter sample was defective: any measurement above 3 was recorded as 3. There were 2 such measurements. Compute the posterior distribution based on all three samples (x, y and z).

3.a. You want to estimate the annual cost of your mobile phone bill. In the month of january 2007, you made 23 calls, with average duration of 114 seconds per call. The call rate is 5 SEK per minute (hence, every second costs 5/60 SEK, there are no initial costs. Same cost day and night). Assume an exponential distribution for the duration of the calls and that every month has 30 days. Use Bayesian methods to compute the probability that the annual call cost for the year 2007 exceeds 3000 SEK. Make whatever assumptions you find necessary/plausible.

4.a. Assume that the data on the next page (can also be downloaded from the course web site) follows a quadratic regression with an intercept

$$y_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + \varepsilon_i,$$

where ε_i is assumed to be iid normal noise with variance σ^2. Use simulation methods to generate a sample from the joint posterior of β_0, β_1,β_2 and σ^2. Use a relatively vague conjugate prior (simple is good, do not complicate). Compute the predictive distribution for $x = 1.1$. [Hint: use BACC for the simulation].

4.b. Fit a linear spline with 10 fixed knots to the data. Compare the fit of the spline to the quadratic fit in 4.b. Compute the predictive distribution for $x = 1.1$. [Hint: construct the linear splines in R, then use BACC on the constructed spline covariates].

4.c. Consider the following alternative model for the regression data:

$$y_i = \sin(x_i\alpha_1) + \varepsilon_i,$$

where $\varepsilon_i \overset{iid}{\sim} N(0, 0.1^2)$. Compute the posterior distribution of α_1. Use a uniform prior between -10 and 10. Compute the predictive distribution for $x = 1.1$. [Hint: evaluate the posterior on a grid. Alternatively, use the Metropolis algorithm].

Have a good learning experience! Good luck!

Mattias Villani

Data for Exercise 4

x	y	x	y
0.1000	0.5089	0.5500	0.4712
0.1100	0.3891	0.5600	0.4081
0.1200	0.6361	0.5700	0.3453
0.1300	0.7675	0.5800	0.2433
0.1400	0.5750	0.5900	0.2581
0.1500	0.7674	0.6000	0.1980
0.1600	0.8428	0.6100	0.0659
0.1700	0.5919	0.6200	0.0038
0.1800	0.6392	0.6300	-0.0380
0.1900	0.8705	0.6400	-0.2059
0.2000	0.8015	0.6500	-0.1316
0.2100	0.9364	0.6600	-0.1459
0.2200	0.9728	0.6700	-0.1754
0.2300	0.9840	0.6800	-0.1112
0.2400	1.0611	0.6900	-0.3386
0.2500	1.0158	0.7000	-0.2885
0.2600	1.0826	0.7100	-0.3172
0.2700	0.8555	0.7200	-0.3484
0.2800	0.9835	0.7300	-0.5860
0.2900	0.9770	0.7400	-0.5086
0.3000	0.8371	0.7500	-0.5478
0.3100	1.0255	0.7600	-0.7126
0.3200	0.8939	0.7700	-0.7248
0.3300	1.1384	0.7800	-0.5795
0.3400	0.9112	0.7900	-0.7363
0.3500	1.0369	0.8000	-0.7178
0.3600	0.9958	0.8100	-0.7797
0.3700	0.8691	0.8200	-0.8818
0.3800	0.7292	0.8300	-0.9019
0.3900	0.9230	0.8400	-0.8272
0.4000	0.8082	0.8500	-0.9900
0.4100	0.9488	0.8600	-0.8380
0.4200	0.9140	0.8700	-0.8782
0.4300	1.0061	0.8800	-1.0338
0.4400	0.8676	0.8900	-0.9923
0.4500	0.7137	0.9000	-1.0963
0.4600	0.7837	0.9100	-1.2071
0.4700	0.6106	0.9200	-0.8951
0.4800	0.6735	0.9300	-1.0499
0.4900	0.6329	0.9400	-0.9672
0.5000	0.5985	0.9500	-0.9759
0.5100	0.5259	0.9600	-0.9940
0.5200	0.6250	0.9700	-1.0909
0.5300	0.2846	0.9800	-1.0772
0.5400	0.4702	0.9900	-1.0093
		1.0000	-1.0775