SC7 Bayes Methods

Problem sheet 0 (background).

I expect you can solve these problems using the knowledge you bring into the course, except perhaps question 1(d), question 2 and question 4. We will review all this in the course, but fairly briefly, so if this is new I recommend some reading. You will find the Decision Theory material in Q2 is covered in SB2.1 Foundations of Statistical Inference (and see Sec 1.3.3 of my lecture notes) and the MCMC material in Q4 is covered in A12 Simulation and Statistical Programming (and in Sec 5.1.1-5.1.6 of my notes).

I include some R in my answers - this is just for illustration. Programming is not assessed in any way in this course.

1. (a) Consider tossing a drawing pin [see figure at end]. Define the result of a toss to be "heads" if the point lands downwards, and "tails" otherwise. Write p for the probability that a toss will land point downwards. Think about p, and choose a, b, so that a Beta(a, b) prior distribution approximates your subjective prior distribution for p. [I used $a = 2$ and $b = 3$ but you may differ.]

   (b) Now collect data. Toss a drawing pin 100 times and keep track of the number of heads after 10, 50, and 100 tosses. You may find the result depends on the surface you use. [I got 4, 16 and 26 heads after 10, 50 and 100 tosses.]

   (c) Ask someone else what prior they chose. Think of your respective priors as a hypotheses about p. Who's beliefs were better supported by the data? Compute a Bayes factor comparing your priors. [for me the other person used $a = 3$ and $b = 2$.]

   (d) Estimate a 95% HPD credible interval for p for each of the two priors you are considering, for the case when $n = 10$ trials. Write down the posterior averaged over models, stating any assumptions you make, and estimate a 95% HPD credible interval for p from the model averaged posterior.

2. Prof Wynn has a hole in his pocket. He walks home from work and finds his keys are not in his pocket. He is $p \times 100\%$ certain he left his keys on his desk (at $\theta = 0$). If they are not there then they could have fallen out anywhere between home and work (uniformly distributed between $\theta = 0$ and $\theta = 1$).

   (a) Let $\theta \in [0, 1]$ be the unknown key location and let $U, V \sim U(0, 1)$ be independent uniform random variables. Verify that Prof Wynn's prior for the key location has the same distribution as $\theta = V\mathbb{I}_{U>p}$.

   (b) Show that the prior variance for θ is $\mathrm{var}(\theta) = (1 - p)(1/12 + p/4)$ and hence show that as p approaches one (desk-certainty) the prior variance $\mathrm{var}(\theta)$ approaches zero.

(c) Prof Wynn returns to work and finds the keys are not on the desk. Explain why the posterior for θ given this new data has variance $1/12$.

(d) Can the posterior have greater variance than the prior? Can conditioning on data *increase* uncertainty?

3. Let $\theta \sim \pi(\cdot)$ and $y \sim p(\cdot|\theta)$ be a prior for a scalar parameter and the observation model for data $y \in \mathbb{R}^n$ respectively. Let $\hat{\theta}(y)$ be an estimator for θ.

(a) Suppose our loss function for estimating $\hat{\theta}$ when the truth is θ is $l(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$. Show that the Bayes estimator is the posterior mean.

(b) Suppose θ is discrete and we have the zero-one loss $l(\theta, \hat{\theta}) = \mathbb{I}_{\hat{\theta}(y) \neq \theta}$. Find the Bayes estimator for θ.

Hint: If you are not familiar with Decision Theory - just the basics are needed - then please read Section 1.3.3 of the lecture notes first.

4. Let $\theta \sim \pi(\cdot|M = m)$ and $y \sim p(\cdot|\theta, M = m)$ be the prior and observation model when the model is $M = m$ and suppose there are just two models, so $M \in \{1, 2\}$. When the model is m the parameter space is $\theta \in \Omega_m$. Consider model selection.

(a) Write down the Bayes factor $B_{1,2}$ in terms of the model elements.

(b) Is it necessary for the models to be nested in order that the Bayes factor (which is after all a likelihood rato) is a model selection criterion?

(c) Suppose the models *are* nested with $\Omega_1 \subseteq \Omega_2$, so we get the $M = 1$ prior by taking the $M = 2$ prior and conditioning on $\theta \in \Omega_1$,

$$\pi(\theta|M = 1) = \frac{\pi(\theta|M = 2, \theta \in \Omega_1)}{\pi(\Omega_1|M = 2)}$$

where

$$\pi(\Omega_1|M = 2) = \int_{\Omega_1} \pi(\theta|M = 2)\, d\theta,$$

and $p(y|\theta, M = 1) = p(y|\theta, M = 2)$ for $\theta \in \Omega_1$. Show that

$$B_{1,2} = \frac{\pi(\Omega_1|y, M = 2)}{\pi(\Omega_1|M = 2)}$$

and briefly interpret.

5. (for those who know some MCMC - we cover this in the course so not strictly background)

(a) Specify a Metropolis-Hastings Markov chain Monte Carlo algorithm targeting $p(x|\theta)$ where $x \in \{0, 1, ..., n\}$ and

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

Prove that your chain is irreducible and aperiodic.

(b) Suppose now that the unknown true success probability for the Binomial random variable X in part (a) is a random variable Θ which can take values in $\{1/2, 1/4, 1/8, ...\}$ only. The prior is

$$\pi(\theta) = \begin{cases} \theta & \text{for } \theta \in \{1/2, 1/4, 1/8, ...\}, \text{ and} \\ 0 & \text{for } \theta \text{ otherwise.} \end{cases}$$

An observed value $X = x$ of the Binomial variable in part (a) is generated by simulating $\Theta \sim \pi(\cdot)$ to get $\Theta = \theta^*$ say, and then $X \sim p(x|\theta^*)$ as before. Specify a Metropolis-Hastings Markov chain Monte Carlo algorithm simulating a Markov chain targeting the posterior $\pi(\theta|x)$ for $\Theta|X = x$.

Statistics Department, University of Oxford
Geoff Nicholls: `nicholls@stats.ox.ac.uk`