SC7 Bayes Methods

Third problem sheet (Sections 4.3-7.2 of lecture notes).

Section A questions

1. Let $x_1, x_2, x_3, ...$ be an infinite exchangeable sequence of binary random variables. Show that $\mathrm{cov}(x_i, x_j) \geq 0$ for all $i, j \in \{1, 2, 3, ...\}$.

2. The Savage axioms (as formulated by DeGroot) characterise coherent prior preference for events stated in terms of inequalities, so that $A \leq B$ implies B is at least as likely as A.

   (a) Write down the first three axioms (see Lecture notes).

   (b) Suppose a probability space $(S, \mathcal{S}, \pi)$ expressing prior preferences exists. For $A, B \in \mathcal{S}$ let A^c, B^c give the complements of A and B. Show $A \leq B \;\; \Rightarrow \;\; A^c \geq B^c$ from the Axioms of Probability.

Section B questions

3. Continuing question 2 in Section A, suppose preferences over sets in $\mathcal{B}$ satisfy the first three Savage Axioms. Show (from the Savage Axioms) that if $A \leq B$ then $A^c \geq B^c$.

4. Let X_1, X_2 be binary random variables. Table entries below give probabilities, $p(x_1, x_2) = \Pr(X_1 = x_1, X_2 = x_2)$, for outcomes $(X_1, X_2) = (x_1, x_2)$ indicated by row and column.[1]

	$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	0	1/2
$X_2 = 1$	1/2	0

   (a) Show that X_1 and X_2 are exchangeable.

   (b) Show that there does not exist a distribution F such that

$$p(x_1, x_2) = \int_0^1 \prod_{i=1,2} p^{x_i}(1-p)^{1-x_i} dF(p),$$

   ie, de Finetti's theorem need not hold if the exchangeable sequence is finite.

5. For $x_i \in \{0, 1\}, \;\; i = 1, 2, 3, ...,$ John elicits a separate prior $p_n(x_1, x_2, ..., x_n)$ for each value of $n = 1, 2, ...$ John's priors are not marginally consistent, in particular,

$$p_n(x_1, ..., x_n) \neq p_{n+1}(x_1, ..., x_n, 0) + p_{n+1}(x_1, ..., x_n, 1).$$

[1] From P. Diaconis and Freedman D. (1980). *Finite Exchangeable Sequences.* Ann. Probab. v8 p745–764.

(a) Can you think of a well-known family of distributions on binary random variables that is not marginally consistent? *Hint, a model for binary images...*

(b) Show that John's priors do not satisfy the Savage axioms (consider the first three).

(c) Show that, under John's priors, $x_1, x_2, x_3, ...x_{n+1}$ is not part of an infinite exchangeable sequence. *Hint: show infinite exchangeable sequences are marginally consistent.*

6. (Model averaging) Consider a normal linear model allowing for outliers. Let X be an $n \times p$ design matrix with rows $x_i = (x_{i,1}, ..., x_{i,p})$ and first column $X_{i,1} = 1, i = 1, ..., n$. Let β be a p-component parameter vector with β_1 the regression intercept. Let z be a latent indicator variable with $z_i = 1$ if (y_i, x_i) is an outlier and $z_i = 0$ otherwise. The response $y_i \sim N(x_i\beta, \sigma^2)$ if $z_i = 0$ and $y_i \sim N(x_i\beta, \rho\sigma^2)$ if $z_i = 1$. Here ρ is a variance inflation factor defining outliers (and ρ is fixed, so for eg we take $\rho = 9$ in Q10 below). Let p be the probability that any single given data point is an outlier.

   (a) The model parameters are β, σ, p and the n-component vector z. The choice of ρ defining outliers is fixed. Write down the likelihood $L(\beta, \sigma, z; y)$.

   (b) Write down the posterior $p(\beta, \sigma, p, z|y)$ if the priors are $p \sim \text{Beta}(1, 9)$, $\beta_i/2.5 \sim t(1)$, iid for $i = 1, ..., p$ and $z_i \sim \text{Bern}(p)$, iid for $i = 1, ..., n$ and $\sigma \sim 1/\sigma$.

   (c) The columns of X are scaled and centred to mean zero, variance one. Show that, conditional on $z_i = 0, i = 1, ..., n$ (no outliers) and σ, β_1 is independent of $\beta_2, ..., \beta_p$ in the posterior. Why might this be desirable for MCMC analysis?

   (d) An MCMC sampler targeting $\pi(\beta, \sigma, p, z|y)$ is given. Explain (a) how you would use the MCMC output to test if a given data point is an outlier, (b) how you would sample the model averaged posterior $\pi(\beta, \sigma, p|y)$ and (c) how you would form a point estimate $\hat{\beta}_i,\ i \in \{1, ..., p\}$ for β_i if your loss function is the square error $|\hat{\beta}_i - \beta_i|^2$.

7. (MCMC with a Jacobian) Consider an MCMC algorithm targeting $\pi(\theta) \propto \theta^{-1/2}/(1 + \theta^2)$ with $\theta > 0$ a scalar parameter. In the following ν is a fixed parameter of the MCMC and $t(\nu)$ denotes the student-t distribution with ν degrees of freedom.

   (a) Calculate the acceptance probability for the MCMC proposal $u \sim t(\nu)$, $\theta' = \theta^u$.

   (b) Comment briefly on how you would decide a value for ν.

Section C questions

8. Continuing question 3 in Section B, show that prior preferences respecting the Savage Axioms are transitive, that is, if $A \leq B$ and $B \leq C$ then $A \leq C$.

9. Consider a process generating $x_1, x_2, x_3, ...$ in which $x_1 = 1$ with fixed and known probability p and for $n = 1, 2, ...,$

$$p(x_{n+1} = 1 | x_n, ..., x_1) = \frac{p + k_n}{1 + n},$$

where $k_n = \sum_{i=1}^{n} x_i$.

   (a) Is the process Markov?

   (b) Show that this process generates an infinite exchangeable sequence.

10. (MCMC with a Jacobian) Let $\theta \in \Re^p$ be a p-component parameter vector with prior $\pi(\theta)$ and $y \in \Re^n$ an n-component data vector with observation model $y \sim p(y|\theta)$. the parameters are positive, and satisfy an order constraint, $0 < \theta_1 < \theta_2 < ... < \theta_p < \infty$.

   (a) Consider the following MCMC proposal. Draw $u_1 \sim U(1/2, 2)$ and $u_2 \sim N(0, \sigma^2)$ where $\sigma > 0$ is a fixed parameter of the MCMC. Set $\theta' = u_1\theta + u_2$, that is $\theta'_i = u_1\theta_i + u_2,\ i = 1, 2, ..., p$. Calculate the acceptance probability $\alpha(\theta'|\theta)$ in as much detail as you can.

   (b) Explain qualitatively why the proposal scheme above is not irreducible (for example in the "computer measure"). A MCMC algorithm which has an update with a second distinct proposal mechanism alternates between the two updates. Outline briefly (in a sentence) a suitable "second update".

11. Continuing question 6 in Section B, the `hills` data are often used to illustrate outlier detection. The finishing time is transformed to make the response more normal, and the covariates for height climbed and distance covered are scaled and centred.

```
> data(hills); a=hills
> a$y=sqrt(a$time); a$climb=scale(a$climb); a$dist=scale(a$dist)
```

We would like to fit a normal linear model `y~climb+dist` to these data, allowing for possible outliers and carrying out model averaging over the outlier labels z. In the file `ProblemSheet3-23.R` is MCMC code for this problem. Run the MCMC, test for outliers and give an 95% HPD interval for the outlier probability p.

12. Continuing question 7 in Section B, implement the MCMC and check your answer!

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