SC7 Bayes Methods

Third problem sheet (Sections 6-8.3 of lecture notes).

Section A

1. Suppose $X = x$ is a draw from an Ising model distribution $X \sim \pi(\cdot|\theta)$ on an $m \times m$ lattice, so that for $i = 1, 2, ..., n$ with $n = m^2$, $x_i \in \{0, 1\}$ and $x = (x_1, ..., x_n)$ with $n = m^2$. Let N_i be the set of neighbors of pixel i on the square lattice. Let $\#x$ denote the number of disagreeing neighbors, that is

$$\#x = \frac{1}{2}\sum_{i=1}^{m^2}\sum_{j \in N_i} \mathbb{I}_{x_i \neq x_j}.$$

Under the Ising model, $\pi(x|\theta) = \exp(-\theta\#x)/Z(\theta)$ where $Z(\theta)$ is a normalising constant.

Suppose we don't observe X itself but instead observe $Y = y_{obs}$ with $y_{obs} = (y_{obs1}, ..., y_{obsn})$ and $Y_i|x_i \sim N(x_i, \sigma^2)$ iid for $i = 1, 2, ..., n$. Here $\sigma > 0$ is known and a prior $\theta \sim \text{Exp}(2)$ is elicited for θ.

   (a) Write down the posterior $\pi(\theta, x|y_{obs})$ in terms of the model elements and explain why it is doubly intractable when $m \gg 1$.

   (b) Consider the statistic for $y \in R^n$

$$S(y) = \frac{1}{2}\sum_{i=1}^{m^2}\sum_{j \in N_i}(y_i - y_j)^2$$

   and distance measure $d(s - s') = |s - s'|$. Briefly motivate this choice of of the ABC statistic $S(y)$ [Hint: what happens when $\sigma \ll 1$.].

   (c) An MCMC algorithm for $X \sim \pi(x|\theta)$ is available. Give an ABC algorithm targeting $\pi(\theta|d(S(Y), S(y_{obs})) < \delta)$ using the MCMC algorithm to simulate $X \sim \pi(x|\theta)$.

2. Consider a linear regression $y_i = x_i^T\beta + \epsilon_i$ with covariates $x_i \in R^p$ and $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$. There are p parameters $\beta \in R^p$ with prior $\pi(\beta) = \prod_j \pi(\beta_j)$. For $j = 1, \ldots, p$ let $z_j \in \{0, 1\}$ be indicator variables. The component priors are

$$\pi(\beta_j|z_j) = (1 - z_j)N(\beta_j; 0, \sigma_1^2) + z_jN(\beta_j; 0, \sigma_2^2)$$

where $0 < \sigma_1 \ll \sigma_2$ are prior hyperparameters which we suppose are fixed. The prior for $z = (z_1, \ldots, z_p)$ is $\pi_Z(z) = \prod_j \pi_Z(z_j)$ where $\pi_Z(z_j) = \Pr(Z_j = 0) = w$ for some $w \in (0, 1)$.

(a) By considering the limit $\sigma_1 \to 0$, explain the relation between this prior and the spike and slab prior (end Section 7.2.1 in lecture notes). Recall that the regression parameters in that setup are $\tilde{\theta}_j = z_j\theta_j$ and $y_i = \sum_i x_i^T\tilde{\theta} + \epsilon_i$ and suppose the prior for θ is $N(0, \sigma_2^2 I_p)$ and the prior for z is the same as above.

(b) What are the relative merits of the two priors? When is one prefered to the other?

Section B (Part C/OMMS to hand in solutions to Section B)

3. (ABC) We considered a version of ABC related to the rejection algorithm. Consider the following MCMC-ABC algorithm[1], targeting $\pi(\theta|y)$ (approximately) using the statistics $S(y)$, distance $d(S, S')$ and threshold δ. The observation model is $p(y|\theta)$ and the prior is $\pi(\theta)$. Suppose $X_t = \theta$.

   Step 1. Simulate $\theta' \sim q(\theta'|\theta)$ and $y' \sim p(y'|\theta')$.

   Step 2. If $d(S(y), S(y')) < \delta$ then accept θ' (set $X_{t+1} = \theta'$) with probability

$$\alpha(\theta'|\theta) = \min\left\{1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)}\right\}$$

   and otherwise reject θ' (set $X_{t+1} = \theta$).

   (a) Show this algorithm targets $\pi(\theta|d(S(Y), S(y)) < \delta)$ (y is fixed here and $Y \sim p(\cdot|\theta)$).

   (b) Suppose we have $y_i \sim$ Poisson(Λ), $i = 1, 2, ..., n$ with $n = 5$. Prior $\lambda \sim \Gamma(\alpha = 1, \beta = 1)$. Give the ABC-MCMC algorithm targeting $\pi(\lambda|y)$ (approximately). Take $S(y) = \bar{y}$, $d(\bar{y}', \bar{y}) = |\bar{y}' - \bar{y}|$ and $\delta = 0.5$.

4. (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 $p \sim$ Beta$(1, 9)$, $\beta_i/2.5 \sim t(1)$, iid for $i = 1, ..., p$ and $z_i \sim$ Bern(p), iid for $i = 1, ..., n$ and $\sigma \sim 1/\sigma$.

[1] Marjoram et al, "Markov chain Monte Carlo without likelihoods", PNAS (2003).

(c) The columns of $X_{2:p}$ are centred to mean zero. 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 $|\hat{\beta}_i - \beta_i|^2$.

5. (MCMC with a Jacobian) Consider an MCMC algorithm targeting $\pi(\theta) \propto \theta^{-1/2}/(1+\theta^2)$ with $\theta > 0$ a scalar random variable. 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.

6. If $\pi(\theta)$ is a prior for θ then an inference scheme is a rule $\psi(\theta; \pi, y)$ for updating belief for θ given data y. For example in Bayesian inference $\psi_{Bayes}(\theta; \pi, y) = \pi(\theta|y)$ but in ABC at fixed δ, $\psi_{\Delta,\delta}(\theta; \pi, y) = \pi(\theta|Y \in \Delta_y(\delta))$.

   For $1 \le j < n$ let $y_{1:j} = (y_1, ..., y_j)$ and $y_{j+1:n} = (y_{j+1}, ..., y_n)$ so we split the data into two sets. Suppose the data are conditionally independent, so

$$p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta).$$

   A belief update is *order-coherent* for conditionally independent data if

$$\psi(\theta; \pi, y) = \psi(\theta; \psi(\theta; \pi, y_{1:j}), y_{j+1:n})$$

   for all $j \in \{1, 2, ..., n-1\}$ (the posterior from the first data set is the prior for the next).

   (a) Show that Bayesian inference is order-coherent.

   (b) Show that ABC with fixed δ is not in general order coherent. *Hint: take summary statistic $S(y) = y$ and Euclidean distance measure $d(y, y') = ||y - y'||$ and give a counter-example.*

   (c) Let $C_y(\delta)$ be the rectangular prism $C_y(\delta) = \{y' \in R^n : |y_i - y_i'| < \delta\ \forall\ i = 1, ..., n\}$. Show that inference with $\psi_{C,\delta}(\theta; \pi, y) = \pi(\theta|Y \in C_y(\delta))$ is order-coherent.

Section C

7. (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".

8. Continuing question 4 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 `ProblemSheet3-25.R` is MCMC code for this problem. Run the MCMC, test for outliers and give an 95% HPD interval for the outlier probability p.

9. Continuing question 5 in Section B, implement the MCMC and check your answer!

10. (RJ targeting inhomogeneous Poisson point process) Let $[L, U]$ be an interval and $\lambda : [L, U] \to \mathbb{R}^+$. In an inhomogeneous Poisson point process (IPP) the probability for an event in a small interval δt is $\lambda(t)\delta t + o(\delta t)$. The sample space is $\Omega = \cup_{n=0}^{\infty}\Omega_n$ where $\Omega_0 = \{\emptyset\}$ and

$$\Omega_n = \{y \in [L, U]^n : L < y_1 < y_2 < \cdots < y_n < U\}.$$

The probability density to realise any particular point pattern $y \in \Omega,\ y = (y_1, \ldots, y_n)$ is

$$p(y|\lambda) = \exp(-\Lambda)\prod_{i=1}^{n}\lambda(y_i)$$

where $\Lambda = \int_L^U \lambda(t)dt$. Denote by $|y|$ the number of points in the ordered set y.

(a) Let $Y \sim p(\cdot|\lambda)$ and let $N = |Y|$. Show that $N \sim \text{Poisson}(\Lambda)$.

(b) Give a reversible jump MCMC algorithm to sample $Y \sim p(\cdot|\lambda)$.

(c) Implement this and check the samples you realise satisfy $|Y| \sim \text{Poisson}(\Lambda)$.

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