SC7 Bayes Methods

Fourth problem sheet (Sections 9.3.2-10 of lecture notes).

Section A questions

1. (RJ-MCMC) For $m \in \{1, 2\}$ and $x \in (0, 1)$ let

$$\pi_{X,M}(x, m) = \pi_{X|M}(x|m)\pi_M(m)$$

with $\pi_M(m = 1) = 1/3$, $\pi_M(m = 2) = 2/3$ and

$$\pi_{X|M}(x|m = 1) = \mathbb{I}_{x=1/2}$$
$$\pi_{X|M}(x|m = 2) = 2x.$$

In the joint $\pi_{X,M}(x, m)$, we have $(x, m) \in \Omega^*$ with $\Omega^* = \{(1/2, 1)\} \cup ((0, 1) \times \{2\})$.

The marginal distribution for X is

$$\pi_X(x) = \sum_{m=1}^{2} \pi_{X,M}(x, m).$$

Let $F_X(x) = \Pr(X \le x),\ x \in (0, 1)$ be the CDF of $X \sim \pi_X(\cdot)$.

   (a) Show that $F_X(x) = \frac{2}{3}x^2 + \frac{1}{3}\mathbb{I}_{x \ge 1/2}$ and give a simple algorithm realising iid $X \sim F_X$.

   (b) Give a RJ-MCMC algorithm targeting $\pi(x, m)$ and say how you would use it to simulate $X \sim F_X$.

   Hint: See code. This gave Figure 1 at the end of the PS.

2. (Dirichlet process) Let H be a continuous distribution on $\Omega = \mathbb{R}^p,\ p \ge 1$ and suppose $G \sim \Pi(\alpha, H)$ is a DP with $\alpha > 0$ a real parameter.

   (a) Let $A \subseteq \Omega$. Calculate $\mathrm{var}(G(A))$. Briefly interpret α and H as model "parameters".

   (b) Suppose for $i = 1, 2, 3, ...,\ \theta_i \sim G$ are iid, with $G \sim \Pi(\alpha, H)$. Recall (lectures) that marginally $\theta_1 \sim H$ and $G|\theta_1 \sim \Pi(\alpha + 1, (\alpha H + \delta_{\theta_1})/(\alpha + 1))$. Show that for $n \ge 1$,

$$G|\theta_{1:n} \sim DP\left(\alpha + n, \frac{\alpha H + \sum_{i=1}^{n} \delta_{\theta_i}}{\alpha + n}\right).$$

   (c) Let $\theta_1^*, ..., \theta_K^*$ denote the distinct values of θ with associated partition $S = (S_1, ..., S_K)$, $S_k = \{i : \theta_i = \theta_k^*, i \in [n]\}$ for $k = 1, ..., K$. Show that

$$E(K) = \sum_{i=1}^{n} \frac{\alpha}{\alpha + i - 1}$$

Section B questions

3. (Reversible jump MCMC) The skew-normal distribution[1] with density $Q(y; \mu, \sigma^2, \xi)$ is obtained from the normal by skewing it with a weight $\xi > 0$. The skewing is negative for $0 < \xi < 1$, positive for $\xi > 1$ and absent for $\xi = 1$, ie $N(y; \mu, \sigma^2) = Q(y; \mu, \sigma^2, 1)$.

   The Shoshoni data $y = (y_1, ..., y_{20})$ give the values of 20 scalar width-to-length ratios of beaded rectangles used by the Shoshoni Indians. They are available here,

   `https://gksmyth.github.io/ozdasl/general/shoshoni.html`

   You can see them and an example of the skew-normal in `ProblemSheet4.R`. Consider using Bayesian inference and RJ MCMC to carry out model selection and model averaging over skewed and normal models for the Shoshoni data.

   (a) Suppose the prior probability for normal (model $m = 1$) or skew-normal (model $m = 2$) is 1/2. Write down the joint posterior distribution $\pi(\theta, m|y)$ for the model index $m = 1, 2$ and parameters $\theta = (\mu, \sigma, \xi)$ in as much detail as you can, though without eliciting priors for the parameters.

   (b) Give a reversible jump MCMC algorithm targeting $\pi(\theta, m|y)$. You can omit the fixed dimension updates.

   (c) Explain how to estimate the Bayes Factor comparing skew-normal and normal models from MCMC output $\theta^{(t)} = (\mu^{(t)}, \sigma^{(t)}, \xi^{(t)})$ and $m^{(t)}, t = 1, 2, ..., T$. How you would simulate data y' from the model averaged posterior predictive distribution $p(y'|y)$?

   (d) (Section C) The code in the R-file `ProblemSheet4.R` implements RJ-MCMC for these data. Use the code to estimate the Bayes factor mentioned above.

4. Let $\Xi_{[n]}$ be the set of partitions of $[n] = \{1, ..., n\}$. The CRP realises $S \in \Xi_{[n]}$ with probability

$$P_{\alpha,[n]}(S) = \frac{\Gamma(\alpha)\alpha^K}{\Gamma(\alpha + n)} \prod_{k=1}^{K} \Gamma(|S_k|).$$

   Let $\mathcal{P}_{[n]}$ be the permutations of $\{1, ..., n\}$.

   (a) For $\sigma \in \mathcal{P}_n$ let $P_{\alpha,\sigma}(S)$ be the distribution over partitions we get if the customers arrive in the order $\sigma = (\sigma_1, \ldots, \sigma_n)$ and let $S(\sigma)$ be the partition obtained by permuting the customer labels in S according to σ. For example if $S = (\{1, 2\}, \{3\})$ and $\sigma = (3, 2, 1)$ then $S(\sigma) = (\{1\}, \{2, 3\})$ because the new partition is $\{\{\sigma_1, \sigma_2\}, \{\sigma_3\}\} = \{\{3, 2\}, \{1\}\}$ and recall the convention $\min(S_k) < \min(S_{k'}) \Leftrightarrow k < k'$.

   Show that $P_{\alpha,[n]}(S) = P_{\alpha,[n]}(S(\sigma)) = P_{\alpha,\sigma}(S)$ for all $S \in \Xi_{[n]}$, so CRP outcomes don't depend on customer arrival order.

[1] Fernandez & Steel "*Bayesian Modeling of Skewness and Fat Tails*", JASA, 1998

(b) Let $S \sim P_{\alpha,[n]}$ and $S^{-i} = (S_1^{-i}, ..., S_{K^{-i}}^{-i})$ be the partition we get if we realise S and then remove some $i \in \{1, ..., n\}$. Here $K^{-i} = K - 1$ if we create an empty cluster when we remove i and otherwise $K^{-i} = K$. For example if $S = (\{1, 2\}, \{3\})$ then $K = 2$ and $S^{-3} = (\{1, 2\})$ so $K^{-3} = 1$.

Let $P_{\alpha,[n]\setminus\{i\}}(S')$ be the probability to realise $S' \in \Xi_{[n]\setminus\{i\}}$ if i is removed from the list of customers before S' is simulated from the CRP. Show that $S^{-i} \sim P_{\alpha,[n]\setminus\{i\}}(S^{-i})$.

5. Consider the following prior for the cluster labels $z = (z_1, ..., z_n)$ of data $y = (y_1, ..., y_n)$ in a mixture model with a fixed number M of components. Let $w = (w_1, ..., w_M)$ be a vector of probabilities $\sum_m w_m = 1$ giving the mixture-component weights.

$$\begin{aligned} w &\sim \text{Dirichlet}(\alpha_1, ..., \alpha_M), \qquad \text{with } \alpha > 0 \text{ and } \alpha_m = \alpha/M,\ m = 1, ..., M \\ z_i &\sim \text{Cat}(w), \qquad \text{iid for } i = 1, ..., n. \end{aligned}$$

In this model $z_i \in \{1, ..., M\}$ is the label of the cluster to which y_i belongs, and the notation $z_i \sim \text{Cat}(w),\ i = 1, ..., n$ means that for $m \in \{1, ..., M\}$ we have $z_i = m$ with probability w_m. Suppose the list $z_1, ..., z_n$ of cluster labels contains $K \le M$ unique distinct values $m_1, ...m_K$. For $k = 1, ..., K$ let $S_k = \{i : z_i = m_k, i = 1, ..., n\}$ give the label-grouping determined by z and let $S = (S_1, ..., S_K)$.

The partition is determined by z, so that $S = S(z)$ with $S \in \Xi_{[n]}$. There are many z's giving the same S. For example, if $n = 4$ and $M = 5$ then $z = (1, 1, 3, 3)$, $z = (3, 3, 1, 1)$ and $z = (4, 4, 2, 2)$ determine the same clustering $S = (\{1, 2\}, \{3, 4\})$.

(a) (*This is an optional Section C question, but result needed below*) Let $n_k = |S_k|$ for $k = 1, ..., K$. Let $P^M_{\alpha,[n]}(S)$ be the probability to realise S. Calculate

$$P^M_{\alpha,[n]}(S) = \sum_{z:S(z)=S} P^M_{\alpha,[n]}(z),$$

where $P^M_{\alpha,[n]}(z)$ is the probability the process realises $z = (z_1, ..., z_n)$, and show

$$P^M_{\alpha,[n]}(S) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/M)^K} \frac{M!}{(M-K)!} \frac{\prod_{k=1}^K \Gamma(\alpha/M + n_k)}{\Gamma(\alpha + n)}.$$

(b) Show that, for each $S \in \Xi_{[n]}$, $\lim_{M\to\infty} P^M_{\alpha,[n]}(S) = P_{\alpha,[n]}(S)$, with $P_{\alpha,[n]}$ from Question (4).
Note: $x\Gamma(x) = \Gamma(x + 1)$ *and* $x\Gamma(x) \to 1$ *as* $x \searrow 0$.

6. A realisation, $G_M \sim \Pi_M(\alpha, H)$, of the *multinomial DP* is simulated as follows:

$$\begin{aligned} w &\sim \text{Dirichlet}(\alpha_1, ..., \alpha_M), \qquad \text{with } \alpha > 0 \text{ and } \alpha_m = \alpha/M,\ m = 1, ..., M, \\ \tilde{\theta}_m &\sim H, \quad \text{iid for } m = 1, ..., M, \end{aligned}$$

and $G_M = \sum_{m=1}^{M} w_m \delta_{\tilde{\theta}_m}$. Here, for $m = 1, ..., M$, $\tilde{\theta}_m \in \mathbb{R}^p$ is a parameter vector of dimension p and H is a base distribution with probability density h on $\mathbb{R}^p$.

(a) For $i = 1, ..., n$, let $\theta_i = \tilde{\theta}_{z_i}$ with

$$z_i \sim \text{Cat}(w), \quad \text{iid for } i = 1, ..., n.$$

Show that $\Pr\{\theta_i \in A | w, \tilde{\theta}) = G_M(A)$ for $A \subseteq \mathbb{R}^p$ and $i = 1, ..., n$.

(b) Let $\theta_1^*, ..., \theta_K^*$ denote the distinct values of θ with associated partition $S = (S_1, ..., S_K)$, $S_k = \{i : \theta_i = \theta_k^*, i \in [n]\}$ for $k = 1, ..., K$. Give the joint distribution $\pi_M(\theta^*, S)$.

(c) Consider the following process.

Step 1 Simulate $\psi_1 \sim H$

Step 2 Independently for $i = 1, ..., n-1$, and sequentially, simulate

$$\psi_{i+1} \sim \frac{\alpha(1 - K_i/M)H + \sum_{k=1}^{K_i}(n_{i,k} + \alpha/M)\delta_{\psi_k^*}}{\alpha + i}.$$

where K_i is the number of distinct ψ-values $\psi_1^*, ..., \psi_{K_i}^*$ at the time of the $i+1$'st arrival and $n_{i,k}$ is the number of times ψ_k^* appears in the list $(\psi_1, ..., \psi_i)$. Show that $\psi = (\psi_1, ..., \psi_n)$ above has the same distribution as $\theta = (\theta_1, ..., \theta_n)$ in Question 6a. *Hint: set it up as a variant of a CRP realising ψ^*, C with ψ^* the unique values in ψ and C the corresponding partition of ψ and repeat the calculation we did in lectures for $P_{\alpha,[n]}(S)$ to get $P(C) = P_{\alpha,[n]}^M(C)$.*

(d) (Section C) Let $\phi_i \sim G$ iid for $i = 1, ..., n$ with $G \sim \Pi(\alpha, H)$ and $\phi = (\phi_1, ..., \phi_n)$. Let $\phi = \theta(\phi^*, S)$ with θ the usual invertible mapping between the two representations. Let $\psi_i \sim G_M$ iid for $i = 1, ..., n$ with $G_M \sim \Pi_M(\alpha, H)$ and $\psi = (\psi_1, ..., \psi_n)$. Let $\psi = \theta(\psi^*, C)$ be corresponding unique values and partition representation (ie as in the hint for Question 6c). Show that $\psi \to \phi$ in distribution as $M \to \infty$ at fixed n. *Hint show that* $\Pr\{(\psi^*, C) \in A^*\} \to \Pr\{(\phi^*, S) \in A^*\}$ *for all sets* A^*.

Section C questions

7. The observation model for data y is $y_i \sim f(\cdot|\theta_i)$, iid for $i = 1, ..., n$ with parameter vector $\theta = (\theta_1, ..., \theta_n)$ determined from the multinomial Dirichlet process model via a realisation of θ^* and S as in Question 6.

   (a) Write down the posterior $\pi_M(S, \theta^*|y)$ for $S, \theta^*|y$ in terms of the model elements.

   (b) Why might we prefer a prior derived from a multinomial Dirichlet process over a prior derived from a Dirichlet process?

(c) Show that the pairs $(\theta_i, y_i)_{i=1}^n$ are exchangeable (as pairs, *ie* preserving the association between θ_i and y_i). Give the S, θ^*-update of a Gibbs sampler targeting $\pi_M(S, \theta^*|, y)$.

8. Mining disasters were common in the period $1850 - 1950$. Let $L = 1850$ and $U = 1950$ and for $i = 1, 2, ..., n$, let $y_i \in (L, U)$ be the date of the i'th event. Let $y = (y_1, ..., y_n)$.

   Model the event times y as the arrival times of a Poisson process of piecewise constant rate $\lambda(t)$ per year. For $m \geq 1$ let $\theta_0 = L$ and $\theta_m = U$ and for $i = 1, ..., m-1$ let $\theta_i \in (L, U)$ be the sorted change-point times at which $\lambda(t)$ jumps up or down. The number of change-points is $m - 1$ so if $m = 1$ then there are no change points and the rate $\lambda(t)$ is constant for $t \in (L, U)$. For $i = 1, ..., m$ let $\lambda_i \geq 0$ give the disaster rate over the interval $(\theta_{i-1}, \theta_i]$. The rate function $\lambda(t) = \lambda(t; \theta, \lambda)$ for y is

$$\lambda(t) = \sum_{i=1}^{m} \lambda_i \mathbb{I}_{\theta_{i-1} < t \leq \theta_i} \qquad L < t < U.$$

   The data and a realisation of $\lambda(t)$ with $m = 4$ are shown in Figure 2 below.

   Let $\theta = (\theta_1, ..., \theta_{m-1})$ and $\lambda = (\lambda_1, ..., \lambda_m)$. Model the change-point times θ as arrivals in a Poisson process of unknown rate ρ per year. The number of intervals m is unknown. Prior densities $\pi_R(\rho),\ \rho \in [0, \infty)$ and $\pi_\Lambda(\lambda|m) = \prod_{i=1}^m \pi_\Lambda(\lambda_i),\ \lambda \in [0, \infty)^m$ are given.

   (a) i. Write down the prior $\pi(\theta, \lambda, m, \rho)$ in as much detail as you can. Specify its parameter space, $(\theta, \lambda, m, \rho) \in \Omega$ say.

   ii. Write down the posterior $\pi(\lambda, \theta, m, \rho|y)$ in terms of the model elements.

   (b) In a reversible jump MCMC algorithm targeting $\pi(\lambda, \theta, m, \rho|y)$, birth and death updates are chosen with probabilities $p_{m,m+1}$ and $p_{m,m-1}$ respectively. A birth proposal $(\lambda, \theta, m, \rho) \to (\lambda', \theta', m', \rho)$ with $m' = m+1$ is generated as follows: choose an interval $i \sim U\{1, ..., m\}$ uniformly; simulate a split point $\theta^* \sim U(\theta_{i-1}, \theta_i)$; simulate two new values $\lambda_{i,1}, \lambda_{i,2} \sim \text{Exp}(1)$ independently. In the candidate state

$$\lambda' = (\lambda_1, ..., \lambda_{i-1}, \lambda_{i,1}, \lambda_{i,2}, \lambda_{i+1}, ..., \lambda_m)$$
$$\theta' = (\theta_1, ..., \theta_{i-1}, \theta^*, \theta_i, ..., \theta_{m-1}).$$

   Give a matching death proposal $(\lambda', \theta', m', \rho) \to (\lambda, \theta, m, \rho)$ and the acceptance probability for the birth proposal. No simplification of expressions is required.

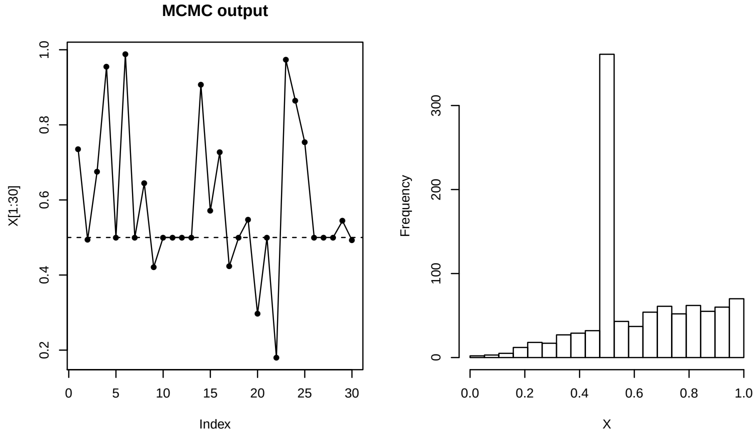

Figure 1: RJ-MCMC targeting $\pi(x, m)$: (Left) plot of x-values realised by the chain (sub-sampled every 10 steps); (Centre) histogram estimate of marginal pdf of x ($f_X(x) = \frac{4}{3}x + \frac{1}{3}\delta_{1/2}(x)$) showing the atom of probability at $x = 1/2$; (Right) Marginal CDF of x ($F_X(x) = \frac{2}{3}x^2 + \frac{1}{3}\mathbb{I}_{x\geq 1/2}$).

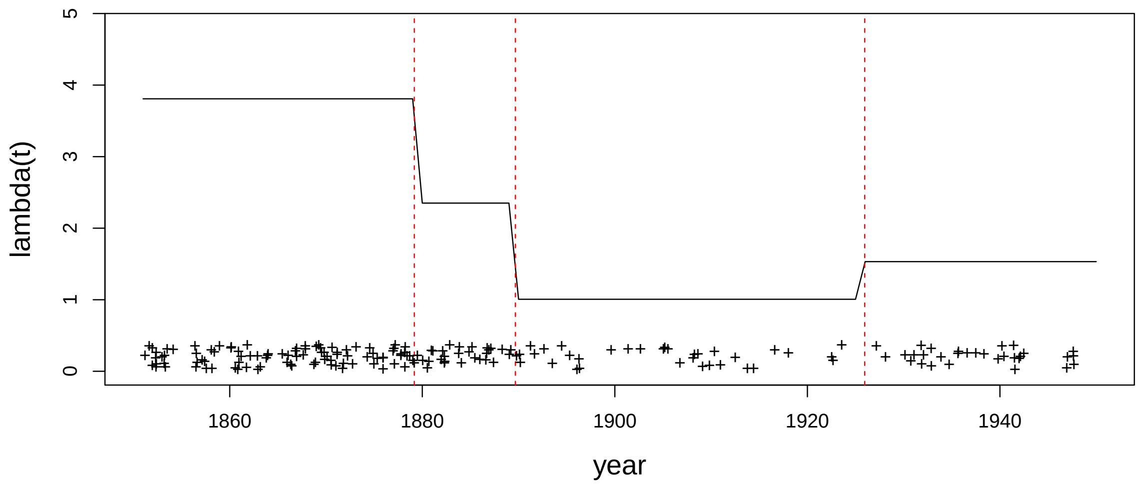

Figure 2: Coal mining disasters: event dates y (+ signs), change point times (θ vertical lines) and $\lambda(t)$ itself (piecewise constant function of year, t).

Statistics Department, University of Oxford

Geoff Nicholls: `nicholls@stats.ox.ac.uk`