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

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

Solution:

$$\begin{aligned}\Pr(X \leq x) &= \pi_M(1)\mathbb{I}_{x\geq 1/2} + \pi_M(2)\int_0^x \pi_{X|M}(x|m = 2)dx \\ &= \frac{1}{3}\mathbb{I}_{x\geq 1/2} + \frac{2}{3}x^2.\end{aligned}$$

Set $X = 1/2$ with probability $1/3$ and otherwise take $Y \sim \pi_{X|M}(\cdot|M = 2)$ and set $X = Y$. If you want to be specific (not needed) on how to do $Y \sim \pi_{X|M}(\cdot|M = 2)$ then the CDF of Y is $F_Y(y) = y^2$ so $F^{-1}(u) = \sqrt{u}$, and then take $U \sim U(0, 1)$ and $Y = \sqrt{U}$.

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

Solution: Suppose the MCMC state is $(X_t, M_t) = (x, m)$.

Add a dimension: $(1/2, 1) \to (x', 2)$

Suppose $(x, m) = (1/2, 1)$. If $m = 1$ propose $m' = 2$ with probability $\rho_{1,2} = 1$ and simulate $x' \sim q(x')$ where $q(x')$ is a proposal distribution we can choose.

In the notation of notes, $u \sim g_{1,2}(u)$ with $x' = u$ and $g_{1,2}(x') = q(x')$. We set $x' = \psi_1(x, u) = u$ and $u' = \psi_2(x, u) = \emptyset$.

Choose $q(x')$ anything irreducible. Since $\Omega_2 = (0, 1)$, I use

$$q(x') = \text{Beta}(x'; \alpha = 1/2, \beta = 1/2) \qquad (!)$$

to emphasise that any irreducible choice works. In this case $\mathcal{U}_{1,2} = (0, 1)$.

Decrease dimension $(x, 2) \to (1/2, 1)$:

Suppose $(x, m) = (x, 2)$ with $x \in (0, 1)$. If $m = 2$ propose $m' = 1$ with probability $\rho_{2,1} = 1$. Set $x' = 1/2$ and $u' = x$.

This is $u \sim g_{2,1}$ with $g_{2,1}(\emptyset) = 1$ so $u = \emptyset$ and we set $x' = \psi_1(x, \emptyset) = 1/2$ and $u' = \psi_2(x', \emptyset) = x$ as that is the value of u' that would "take us back" to x. In this case $\mathcal{U}_{2,1} = \{\emptyset\}$. Here are the *acceptance probabilities*:

Increase dimension, $(1/2, 1) \to (x', 2)$: the acceptance probability is,

$$\alpha(x', m'|x, m) = \min\left\{1,\ \frac{\pi(x', m')\rho_{m',m} g_{2,1}(u')}{\pi(x, m)\rho_{m,m'} g_{1,2}(u)} J_\psi(x, u)\right\}.$$

We set $(x', u') = \psi(x, u)$ with $\psi(x, u) = (u, \emptyset)$ so $J_\psi(x, u) = |\partial x'/\partial u| = 1$. Substituting $g_{2,1} = 1$, $g_{1,2}(u) = q(x')$ etc,

$$\alpha(x', m'|x, m) = \min\left\{1,\ \frac{4x'/3}{\text{Beta}(x'; \alpha, \beta)/3}\right\}.$$

Decrease dimension $(x, 2) \to (1/2, 1)$: similarly

$$\alpha(x', m|x, m') = \min\left\{1,\ \frac{\text{Beta}(x; \alpha, \beta)/3}{4x/3}\right\}.$$

Iteration: we generate our chain (X_t, M_t) iterating proposals and acceptance steps using the formulae above.

The dimension of the proposal distribution matches the *change in dimension* in the target state

$$\dim(\Omega_1 \times \mathcal{U}_{1,2}) = \dim(\Omega_2 \times \mathcal{U}_{2,1})$$

since $\dim(\{1/2\} \times (0, 1)) = \dim((0, 1) \times \{\emptyset\}) = 1$. Our algorithm is stationary for $\pi_{X,M}$ so if we have reached equilibrium then $(X_t, M_t) \sim \pi_{X,M}$ and then $X_t \sim F_X$.

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

2. (Dirichlet process) Let H be a continuous distribution on $\Omega = \mathbb{R}^p$, $p \geq 1$ 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".

> **Solution:** By the definition of a DP and per lecture notes
>
> $$G(A) \sim \mathrm{Beta}(\alpha H(A), \alpha(1 - H(A))).$$
>
> Now if $X \sim \mathrm{Beta}(a, b)$ then
>
> $$\mathrm{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$
>
> so
>
> $$\mathrm{var}(G(A)) = \frac{H(A)(1 - H(A))}{(\alpha + 1)}$$
>
> Now if $G \sim \Pi(\alpha, H)$ then G is a random distribution. H is the centering distribution (since $E(G(A)) = H(A)$) and α is a precision parameter (larger α is lower variance for $G(A)$ for every set $A \subset \Omega$). In particular $\alpha + 1$ is like the effective sample size of the information in the prior.

   (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 $n \geq 1$,

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

> **Solution:** In showing
>
> $$G|\theta_{1:n} \sim DP\left(\alpha + n, \frac{\alpha H + \sum_{i=1}^{n} \delta_{\theta_i}}{\alpha + n}\right)$$
>
> We use the result from lectures for $G|\theta_1$ as the base case. Let $\tilde{G}_{n-1} = G|\theta_1, ..., \theta_{n-1}$. The induction hypothesis is
>
> $$\tilde{G}_{n-1} \sim \Pi(\tilde{\alpha}_{n-1}, \tilde{H}_{n-1})$$
>
> with $\tilde{\alpha}_{n-1} = \alpha + n - 1$ and
>
> $$\tilde{H}_{n-1} = \frac{\alpha H + \sum_{i=1}^{n-1} \delta_{\theta_i}}{\alpha + n - 1}.$$

We can use the same update that took us from G to $G|\theta_1$. We have

$$\tilde{G}_{n-1}|\theta_n \sim DP\left(\tilde{\alpha}_{n-1}+1, \frac{\tilde{\alpha}_{n-1}\tilde{H}_{n-1}+\delta_{\theta_n}}{\tilde{\alpha}_{n-1}+1}\right),$$

with

$$\begin{aligned}\frac{\tilde{\alpha}_{n-1}\tilde{H}_{n-1}+\delta_{\theta_n}}{\tilde{\alpha}_{n-1}+1} &= \frac{(\alpha+n-1)\frac{\alpha H+\sum_{i=1}^{n-1}\delta_{\theta_i}}{\alpha+n-1}+\delta_{\theta_n}}{\alpha+n-1+1}\\ &= \frac{\alpha H+\sum_{i=1}^{n}\delta_{\theta_i}}{\alpha+n}\end{aligned}$$

and $\tilde{\alpha}_{n-1}+1=\alpha+n-1+1$ so that $\tilde{\alpha}_{n-1}+1=\alpha+n$. We have the distribution of $\tilde{G}_{n-1}|\theta_n$. But $\tilde{G}_{n-1}|\theta_n = G|\theta_1,...,\theta_n$ and we are done.

(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

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

Solution: Let $\mathcal{E}_j$ be the event that the j'th θ-value starts a new cluster. This happens with probability $\alpha/(\alpha+j-1)$. The final total number of clusters is $K=\sum_{j=1}^{n}\mathbb{I}_{\mathcal{E}_j}$ so

$$\begin{aligned}E(K) &= E\left(\sum_{j=1}^{n}\mathbb{I}_{\mathcal{E}_j}\right)\\ &= \sum_{j=1}^{n}\Pr(\mathcal{E}_j)\\ &= \sum_{j=1}^{n}\frac{\alpha}{\alpha+j-1}\end{aligned}$$

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

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

$0 < \xi < 1$, 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.

(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

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

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.

Solution: Let $y_{S_k} = (y_i)_{i \in S_k}$.

$$\begin{aligned}
f(y|S,\mu^*,\sigma^*) &= \prod_{k=1}^{K} f(y_{S_k}|\theta_k^*) \\
\pi(\theta^*,S) &= P^M_{\alpha,[n]}(S)\pi_M(\theta^*|S) \\
\pi_M(\theta^*|S) &= \prod_{k=1}^{K(S)} h(\theta_k^*) \\
P^M_{\alpha,[n]}(S) &= \frac{\Gamma(\alpha)}{\Gamma(\alpha/M)^K}\frac{M!}{(M-K)!}\frac{\prod_{k=1}^K \Gamma(\alpha/M+n_k)}{\Gamma(\alpha+n)}
\end{aligned}$$

in terms of which

$$\pi(S,\mu^*,\sigma^*|y) \propto P^M_{\alpha,[n]}(S)\pi_M(\theta^*|S)f(y|S,\mu^*,\sigma^*).$$

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

Solution: When n is large the number of clusters may become unphysical in a DP (which tends to generate a small number of very large clusters and a large number of very small clusters). We saw that the expected number of clusters grows without bound with increasing n. Bounding the number of clusters may lead to a prior more in line with prior knowledge.

(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 $\pi_M(S,\theta^*|,y)$.

Solution: The y_i values are conditionally independent (of each other and everything else) given the θ_i values, so it is sufficient to show the θ_i values are exchangeable for $i = 1,...,n$. But the generative model is (1) $G_M \sim \Pi_M(\alpha, H)$, then (2) $\theta_i \sim G_M$ iid and finally (3) $y_i \sim f(y_{k_i}|\theta_i)$, so it doesnt matter what order we simulate (θ_i, y_i). For the Gibbs sampler we want the conditional distribution of i in S given y and θ^*. Relabel so that i is in S_K. If i is the last arrival and $n_K = 1$ then $S^{-i} = (S_1,...,S_{K-1})$. Sample $k' \sim p,\ p = (p_1,...,p_K)$ where

$$p_k \propto \begin{cases} (|S_k| + \alpha/M)f(y_i|\theta_k^*) & 1 \le k < K \\ \alpha(1-(K-1)/M)f(y_i|\theta_K^*) & k = K \end{cases}.$$

If $n_K > 1$ then simulate $\theta^*_{K+1} \sim h(\cdot)$ and $k' \sim p,\ p = (p_1, ..., p_{K+1})$ where

$$p_k \propto \begin{cases} (|S_k| + \alpha/M) f(y_i|\theta^*_k) & 1 \le k < K \\ (|S_K| - 1 + \alpha/M) f(y_i|\theta^*_K) & k = K \\ \alpha(1 - K/M) f(y_i|\theta^*_{K+1}) & k = K + 1 \end{cases}.$$

Normalise these for the Gibbs sampler distributions. Take i out of S_K and put it in $S_{k'}$, deleting S_K and θ^*_K if $S_K \setminus \{i\} = \emptyset$. Finally, restore the standard labelling of S ordering by least entry. Note that if $K = M$ then $\alpha(1 - K/M) = 0$ so we will never choose to increase K past M.

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 $m \ge 1$ 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 $\lambda_i \ge 0$ give the disaster rate over the interval $(\theta_{i-1}, \theta_i]$. The rate function $\lambda(t) = \lambda(t; \theta, \lambda)$ for y is

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

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.

Solution: Let $T = U - L$. We have $\pi(m|\rho) = \text{Poisson}(m - 1; T\rho)$ and for $m \ge 1$ and $\theta \in \Theta_{m-1}$ the θ's are uniform in (L, U) so

$$\pi_\Theta(\theta|m) = (m - 1)!/T^{m-1}$$

(recall they are ordered), with $\Theta_0 = \emptyset$ (when $m = 1$) and

$$\Theta_{m-1} = \{\theta \in (L, U)^{m-1} : L < \theta_1 < ... < \theta_{m-1} < U\}.$$

The full prior is

$$\begin{aligned}\pi(\theta,\lambda,m,\rho) &= \pi_\Theta(\theta|m)\pi_\Lambda(\lambda|m)\pi(m|\rho)\pi(\rho)\\ &= \frac{(m-1)!}{T^{m-1}}\pi_\Lambda(\lambda|m)\exp(-T\rho)\frac{(T\rho)^{m-1}}{(m-1)!}\pi(\rho)\\ &\propto \exp(-T\rho)\rho^{m-1}\pi_\Lambda(\lambda|m)\pi(\rho)\end{aligned}$$

with $\pi_\Lambda(\lambda|m)\pi(\rho)$ given. The parameter space for (θ,λ,m,ρ) is

$$\Omega = [0,\infty)\times\bigcup_{m=1}^{\infty}\{m\}\times\Lambda_m\times\Theta_{m-1}$$

where $\Lambda_m = [0,\infty)^m$.

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

Solution: Let $n_j = \sum_i \mathbb{I}_{\theta_{j-1}<y_i\le\theta_j}$ be the number of events in the interval $(\theta_{j-1},\theta_j]$. This number is Poisson$(\lambda_j(\theta_j-\theta_{j-1}))$ and the points are otherwise scattered uniformly at random. The likelihood is

$$\begin{aligned}L(\theta,\lambda;y) &= \prod_{j=1}^m e^{-\lambda_j(\theta_j-\theta_{j-1})}\frac{(\lambda_j(\theta_j-\theta_{j-1}))^{n_j}}{n_j!}\times\frac{1}{(\theta_j-\theta_{j-1})^{n_j}}\\ &\propto \prod_{j=1}^m e^{-\lambda_j(\theta_j-\theta_{j-1})}\frac{\lambda_j^{n_j}}{n_j!},\end{aligned}$$

(which alternatively follows from exponential intervals between y's but is fiddly to show as you have to allow for intervals that cross a change point) so the posterior is

$$\pi(\lambda,\theta,m,\rho|y)\propto\pi(\theta,\lambda,m,\rho)L(\theta,\lambda;y)$$

with the likelihood and prior given above and parameter space $(\lambda,\theta,m,\rho)\in\Omega$ with Ω defined above.

(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.

Solution: Relabel $\theta' = (\theta'_1, ..., \theta'_{m'-1})$ and $\lambda' = (\lambda'_1, ..., \lambda'_{m'})$ and consider now generating $(\lambda, \theta, m, \rho)$ from $(\lambda', \theta', m', \rho)$ by some sort of deletion move. The death update is chosen with probability $p_{m',m}$. In the death update choose $i \sim U\{1, ..., m'-1\}$, delete θ'_i and λ'_i and λ'_{i+1}. Simulate a new value $\tilde{\lambda}_i \sim \text{Exp}(1)$ say. Set $\lambda = (\lambda'_1, ..., \lambda'_{i-1}, \tilde{\lambda}_i, \lambda'_{i+2}, ..., \lambda'_{m'})$ and $\theta = (\theta'_1, ..., \theta'_{i-1}, \theta'_{i+1}, ..., \theta'_{m'})$.

The probability distribution for the birth proposal is

$$q(\lambda', \theta', m', \rho|\lambda, \theta, m, \rho) = p_{m,m+1} m^{-1} (\theta_i - \theta_{i-1})^{-1} \text{Exp}(\lambda_{i,1}; 1) \text{Exp}(\lambda_{i,2}; 1)$$

and for the death proposal is

$$q(\lambda, \theta, m, \rho|\lambda', \theta', m', \rho) = p_{m+1,m} m^{-1} \text{Exp}(\lambda_i; 1).$$

We get m^{-1} because we select a change-point to delete from $m' - 1 = m$ choices. The death proposal has to hit the old state, so the proposal density for $\tilde{\lambda}_i$ is evaluated at $\tilde{\lambda}_i = \lambda_i$.

The acceptance probability for the birth is simply

$$\alpha(\lambda', \theta', m', \rho|\lambda, \theta, m, \rho) = \min\left\{1, \frac{\pi(\lambda', \theta', m', \rho|y) q(\lambda, \theta, m, \rho|\lambda', \theta', m', \rho)}{\pi(\lambda, \theta, m, \rho|y) q(\lambda', \theta', m', \rho|\lambda, \theta, m, \rho)}\right\}.$$

The following remarks are not needed if the above is correct. The dimensions balance since

$$\dim(\theta, \lambda) + \dim(\theta^*, \lambda_{i,1}, \lambda_{i,2}) = m - 1 + m + 3 = 2m + 2$$

and

$$\dim(\theta', \lambda') + \dim(\tilde{\lambda}_i) = m' - 1 + m' + 1 = 2m' = 2m + 2.$$

In terms of the notation for the general reversible jump algorithm, θ, u become (θ, λ) and $u = (\theta^*, \lambda_{i,1}, \lambda_{i,2})$ and θ', u' become (θ', λ') and $u' = \tilde{\lambda}_i$. The proposal densities are

$$g_{m,m',i}(\theta^*, \lambda_{i,1}, \lambda_{i,2}) = (\theta_i - \theta_{i-1})^{-1} \text{Exp}(\lambda_{i,1}; 1) \text{Exp}(\lambda_{i,2}; 1)$$

and

$$g_{m',m,i}(\tilde{\lambda}_i) = \mathrm{Exp}(\lambda_i; 1).$$

The probability to propose $m \to m'$ when $m' = m+1$ is $\rho_{m,m'} = p_{m,m'}$ and $\rho_{m',m} = p_{m',m}$. Given we know m, m', we choose the transition kernel $dK_{m,m',i}((\theta,\lambda),(\theta',\lambda'))$ with probability $\xi_{m,m',i} = 1/m$. This operates on the i'th interval and uses $g_{m,m',i}$. It is reversed by the kernel $dK_{m',m,i}((\theta',\lambda'),(\theta,\lambda))$ which is chosen with probability $\xi_{m',m,i} = 1/(m'-1) = 1/m$. This operates on the i'th change-point and uses $g_{m',m,i}$. The Jacobian is

$$\left| \frac{\partial(\theta', \lambda', \tilde{\lambda}_i)}{\partial(\theta, \lambda, \theta^*, \lambda_{i,1}, \lambda_{i,2})} \right| \;=\; 1$$

because all the elements in $(\theta', \lambda', \tilde{\lambda}_i)$ are matched to, and equal to, unique elements in $(\theta, \lambda, \theta^*, \lambda_{i,1}, \lambda_{i,2})$ so the matrix of derivatives will have exactly one 1 in each row and column and otherwise all zero entries. In the notation of lectures the acceptance probability is

$$\alpha(\theta', m'|\theta, m) = \min\left\{1,\ \frac{\pi(\theta', m'|y)\rho_{m',m}\xi_{m',m,i}g_{m',m,i}(u')}{\pi(\theta, m|y)\rho_{m,m'}\xi_{m,m',i}g_{m,m',i}(u)} \left|\frac{\partial(\theta', u')}{\partial(\theta, u)}\right|\right\}$$

which (in the notation of the present problem) is exactly $\alpha(\lambda', \theta', m', \rho|\lambda, \theta, m, \rho)$ above. Note the slight extra feature that in going from $m \to m'$ we have many ways to add a change point (ie, split any of the m intervals). Each of these is a different transition kernel which we pick with probabilities $\xi_{m,m'} = (\xi_{m,m',1}, \ldots, \xi_{m,m',m})$ so we have factors $\xi_{m,m',i}$ in the acceptance probability as we saw in the "Matched Proposals" section of lecture notes (currently Definition 8.19 in Section 8.2.3 on page 95).

A full implementation is given in the online code for this PS.

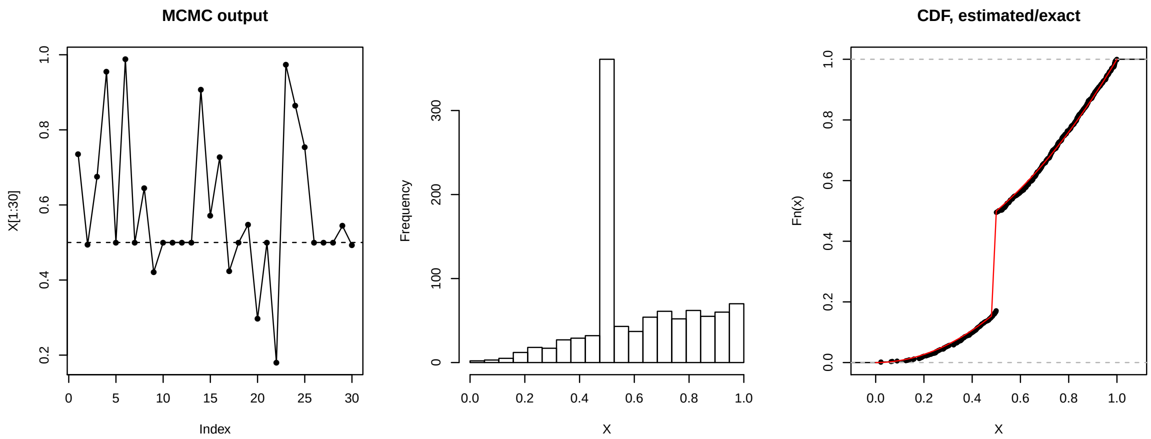

Figure 1: RJ-MCMC targeting $\pi(x, m)$: (Left) plot of x-values realised by the chain (sub-sampled every 10 steps); (Centre) histogram estimate of marginal pdf of x ($f_X(x) = \frac{4}{3}x + \frac{1}{3}\delta_{1/2}(x)$) showing the atom of probability at $x = 1/2$; (Right) Marginal CDF of x ($F_X(x)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

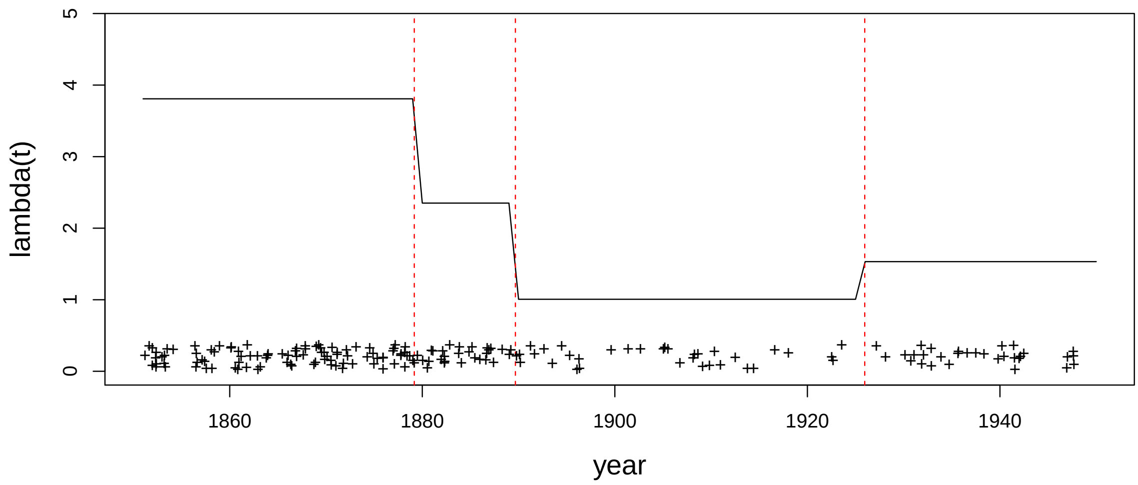

Figure 2: Coal mining disasters: event dates y (+ signs), change point times (θ vertical lines) and $\lambda(t)$ itself (piecewise constant function of year, t).

Statistics Department, University of Oxford

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