

Math 598

Markov chain Monte Carlo

Dr David A. Stephens

Department of Mathematics & Statistics
Room 1225, Burnside Hall

Model-selection via trans-dimensional MCMC

Model selection via MCMC exploits trans-dimensional moves

- Discrete time chains
 - Carlin and Chib (1995): extend the state space to consider all possible models simultaneously
- Continuous time chains
 - Grenander and Miller (1994), Phillips and Smith (1995): jump diffusions.
 - Geyer and Moller (1994), Stephens (M) (2000): birth-death in continuous time

Model-selection via trans-dimensional MCMC

In the contributed Discussion of Grenander and Miller (1994), Green laid out the basic construction of

Reversible Jump MCMC.

- ▶ Green pointed out that whereas their Gibbs sampler construction might only be rarely of use, their MH algorithm they also suggested could provide a general mechanism for jumping between subspaces of different dimension.
- ▶ The suggested general algorithm retained the reversibility and detailed balance properties of the usual MH algorithm, but required “dimension-matching” terms to account for the differences in dimension of different subspaces.

Reversible Jump MCMC

- ▶ target distribution π
- ▶ parameter vector $\theta \in \Theta \subseteq \mathcal{R}^d$;
- ▶ proposal kernel $Q(\theta, d\theta') = \Pr(\theta' \in d\theta' | \theta)$,
- ▶ transition kernel $P(\theta, d\theta')$ with π as its unique stationary distribution.

Let π and q denote the densities of π and Q wrt to Lebesgue measure.

Reversible Jump MCMC

By the usual arguments, the acceptance probability for values generated from q is

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

Under mild conditions on $Q(.,.)$, variates generated from this Markov chain with transition kernel $P(.,.)$ form a dependent sample from π , and the ergodicity of the chain permits Monte Carlo estimation of functionals of π .

Reversible Jump MCMC

Consider a countable collection of Bayesian models, $\{M_k, k = 1, 2, \dots\}$, where model M_k is parameterized via parameter θ_k with parameter space $\Theta_k \subset \mathbb{R}^{d_k}$. We formulate a joint prior on the parameters and the unknown model $\mathcal{M}$ by setting a discrete prior

$$\pi_0(M_k) = \Pr(\mathcal{M} = M_k)$$

say, and then the conditional prior

$$\pi_0(\theta_k|M_k)$$

so that the full joint posterior can be written

$$\pi_n(M_k, \theta_k) = \frac{\mathcal{L}_n^{M_k}(\theta_k) \pi_0(\theta_k|M_k) \pi_0(M_k)}{\sum_j \left\{ \int \mathcal{L}_n^{M_j}(t_j) \pi_0(t_j|M_j) dt_j \right\} \pi_0(M_j)}$$

Reversible Jump MCMC

Our objective is to construct a Metropolis-Hastings chain to that is aperiodic and irreducible on the **union** parameter space

$$\Theta = \bigcup_k \Theta_k.$$

This requires the specification of a proposal transition density, $q(., .)$, but, in contrast to the usual fixed-dimension case, we face the difficulty that the arguments of q are potentially of different dimensions, rendering the reversibility requirement difficult to meet.

Reversible Jump MCMC

At a specific iteration, suppose that

- the chain is in model M with parameter θ having dimension d ,
- the proposal is to move to model M' with parameter θ' having dimension d' .

We envisage this move as first selecting a move between models, $M \rightarrow M'$, and then the proposal of a θ' possibly dependent on the current θ .

Reversible Jump MCMC

For a reversible proposal mechanism, we need equally dimensioned arguments to $q(.,.)$: consider the introduction of collections of latent variables $\mathbf{u}$ and $\mathbf{v}$ of dimension d_u and d_v respectively, so that $d + d_u = d' + d_v$.

The proposal density q is then considered for the extended parameter vectors $(\theta, \mathbf{u})$ and $(\theta', \mathbf{v})$ such that $q((\theta, \mathbf{u}), (\theta', \mathbf{v}))$ is reversible; this is most easily constructed using a 1-1 differentiable mapping, that is

$$(\theta', \mathbf{v}) = \mathbf{g}(\theta, \mathbf{u}) \quad \Longleftrightarrow \quad (\theta, \mathbf{u}) = \mathbf{h}(\theta', \mathbf{v}).$$

Reversible Jump MCMC

Standard (fixed-dimension) moves are special cases of this general proposal procedure; for example, for a Metropolis move, we might set

$$\theta' = \theta + \mathbf{u} \quad \mathbf{u} \sim \text{Normal}_d(0, \Sigma)$$

and for Metropolis-Hastings moves, we typically use a conditional generation $q(\theta'|\theta)$ taking the conditioning variable as a constant parameter in a suitably chosen density; the stochastic elements $\mathbf{u}$ in the MH move can be thought of as the uniform variates used to perform basic random number generation from this conditional density.

Reversible Jump MCMC

To establish the acceptance probability for the reversible jump move, consider a “hybrid” MH algorithm comprising a mixture of move types, with

- move m
- transition proposal Q_m
- move m selected with probability r_m

for a finite collection of move types, some of which may be trans-dimensional, but each of which retains the detailed balance property.

If ψ and ψ' represent the latent-augmented parameter vectors, and let the augmented posterior density be denoted

$$\tilde{\pi}_m(\psi) = \tilde{\pi}_m(M, \theta, \mathbf{u}) = \pi_n(M, \theta) p_U(\mathbf{u})$$

Reversible Jump MCMC

Let $\alpha_m(\psi, \psi')$ denote the acceptance probability for move type m , so that for arbitrary sets A and B

$$\int_A \tilde{\pi}_m(d\psi) \int_B \alpha_m(\psi, \psi') Q_m(\psi, d\psi') = \int_B \tilde{\pi}_m(d\psi') \int_A \alpha_m(\psi', \psi) Q_m(\psi', d\psi)$$

which implies as usual that

$$\alpha_m(\psi, \psi') \tilde{\pi}_m(\psi) q_m(\psi, \psi') = \alpha_m(\psi', \psi) \tilde{\pi}_m(\psi') q_m(\psi', \psi)$$

or,

$$\alpha_m(\psi, \psi') f_m(\psi, \psi') = \alpha_m(\psi', \psi) f_m(\psi', \psi)$$

say where f_m is defined with respect to a common, symmetric measure on the product space.

Reversible Jump MCMC

Consider the forward move; in this case $f_m(\psi, \psi')$ is given by

$$f_m(\psi, \psi') = \tilde{\pi}_m(\psi)q_m(\psi, \psi') = \pi_n(M, \theta)p_U(\mathbf{u})\overrightarrow{r}_m.$$

For the reverse move, to preserve symmetry, we must set

$$f_m(\psi', \psi) = \tilde{\pi}_m(\psi')q_m(\psi', \psi) = \pi_n(M', \theta')p_V(\mathbf{v})\left|\frac{\partial(\theta', \mathbf{v})}{\partial(\theta, \mathbf{u})}\right|\overleftarrow{r}_m$$

where the term

$$\left|\frac{\partial(\theta', \mathbf{v})}{\partial(\theta, \mathbf{u})}\right| = \left|\frac{\partial \mathbf{g}(\mathbf{t}_1, \mathbf{t}_2)}{\partial(\theta, \mathbf{u})}\right|_{\mathbf{t}_1=\theta, \mathbf{t}_2=\mathbf{u}}$$

is the Jacobian for the 1-1 mapping $\mathbf{g} : (\theta, \mathbf{u}) \mapsto (\theta', \mathbf{v})$.

Reversible Jump MCMC

This term arises as in the augmented posterior, we have

$$\tilde{\pi}(M', \theta', \mathbf{v}) = \tilde{\pi}(M, \mathbf{h}(\theta', \mathbf{v})) |J(\theta', \mathbf{v})| = \tilde{\pi}(M, \theta, \mathbf{u}) |J(\theta, \mathbf{u})|^{-1}$$

under the bijection; here

$$|J(\theta', \mathbf{v})| = \left| \frac{\partial(\theta, \mathbf{u})}{\partial(\theta', \mathbf{v})} \right| = \left| \frac{\partial(\theta', \mathbf{v})}{\partial(\theta, \mathbf{u})} \right|^{-1} = |J(\theta, \mathbf{u})|^{-1}$$

is the Jacobian.

Reversible Jump MCMC

In the expressions above, the two terms $\vec{r}_m$ and $\overleftarrow{r}_m$ represent the probabilities of choosing to make move m (from M to M') and the probability of the reverse move (from M' to M).

Thus, the acceptance probability for move type m is

$$\alpha_m((M, \theta), (M', \theta')) = \min \left\{ 1, \frac{\pi_n(M', \theta') p_V(\mathbf{v}) \overleftarrow{r}_m}{\pi_n(M, \theta) p_U(\mathbf{u}) \vec{r}_m} \left| \frac{\partial(\theta', \mathbf{v})}{\partial(\theta, \mathbf{u})} \right| \right\}.$$

Reversible Jump MCMC

These general calculations simplify if one of $\mathbf{u}$ or $\mathbf{v}$ is a null vector. If $\dim(\theta) < \dim(\theta')$, and the proposed move attempts to increase the dimension of the current model, then only the augmenting variables $\mathbf{u}$ are needed to match dimension, that is, the bijection can be constructed by setting $\theta' = \mathbf{g}(\theta, \mathbf{v})$. In this case

$$\alpha_m((M, \theta), (M', \theta')) = \min \left\{ 1, \frac{\pi_n(M', \theta') \overleftarrow{r}_m}{\pi_n(M, \theta) p_U(\mathbf{u}) \overrightarrow{r}_m} \left| \frac{\partial(\theta')}{\partial(\theta, \mathbf{u})} \right| \right\}.$$

Conversely, if $\dim(\theta) > \dim(\theta')$, then only the augmenting variables $\mathbf{v}$ are needed, and

$$\alpha_m((M, \theta), (M', \theta')) = \min \left\{ 1, \frac{\pi_n(M', \theta') p_V(\mathbf{v}) \overleftarrow{r}_m}{\pi_n(M, \theta) \overrightarrow{r}_m} \left| \frac{\partial(\theta', \mathbf{v})}{\partial(\theta)} \right| \right\}.$$

Notes:

1. The proposal mechanisms can be generalized by allowing the generation for $\mathbf{u}$ or $\mathbf{v}$ to depend on the values of θ or θ' respectively. Under this generalization, $p_U(\mathbf{u})$ and $p_V(\mathbf{v})$ are replaced by $p_U(\mathbf{u}|\theta)$ and $p_V(\mathbf{v}|\theta')$ in the acceptance probabilities. This merely corresponds to an alternative construction of the augmented posterior $\tilde{\pi}$.

Reversible Jump MCMC

2. Each move type m comprises a pair of moves operating in each direction between models M and M' .
 - if move m is a move from M to M' , constructed by generation of augmenting variables $\mathbf{u}$ and transformation, there should exist in our collection of potential moves the reverse move, indexed m' , say, which utilizes the augmenting variables $\mathbf{v}$, the two moves being selected with probabilities $\vec{r}_m$ and $\overleftarrow{r}_m$

Reversible Jump MCMC

3. In some situations, the reversible jump acceptance probability simplifies even further.
 - In a move that increases dimension by d_U through the variables $\mathbf{u}$, it may be feasible to set the new parameters precisely equal to $\mathbf{u}$, that is, the proposed parameter vector θ' is formed by concatenating θ and $\mathbf{u}$. In this case, the Jacobian of the transformation is 1.
4. It may be possible to generate $\mathbf{u}$ from a prior distribution, which facilitates cancellation in the Hastings ratio.

Prior specification

Let $k = 1, 2, \dots$ index models $M_1, M_2, \dots$ under consideration, where k represents the dimension of $\theta_k = (\theta_{k1}, \dots, \theta_{kk})$. Suppose that the model specification is such that identical, independent priors are used for the components of the parameter vector

$$\pi_0(\theta_k) = \prod_{l=1}^k \pi_0(\theta_{kl}).$$

For a proposed move from M_j to M_k with $j < k$, suppose that the elements of $\mathbf{u}$ are generated independently from p_0 , with

$$p_U(\mathbf{u}) = \prod_{l=1}^{k-j} \pi_0(u_l),$$

say, and suppose that $\theta_k = (\theta_j, \mathbf{u})$.

Prior specification

Then the acceptance probability $\alpha_j \equiv \alpha_j((M_j, \theta_j), (M_k, \theta_k))$ is

$$\begin{aligned}\alpha_j &= \min \left\{ 1, \frac{\pi_n(M_k, \theta_k) \overleftarrow{r}_j}{\pi_n(M_j, \theta_j) p_U(\mathbf{u}) \overrightarrow{r}_j} \left| \frac{\partial(\theta_k)}{\partial(\theta, \mathbf{u})} \right| \right\} \\ &= \min \left\{ 1, \frac{\mathcal{L}_n^{M_k}(\theta_k) \pi_0(\theta_k | M_k) \pi_0(M_k) \overleftarrow{r}_j}{\mathcal{L}_n^{M_j}(\theta_j) \pi_0(\theta_j | M_j) \pi_0(M_j) p_U(\mathbf{u}) \overrightarrow{r}_j} \left| \frac{\partial(\theta_k)}{\partial(\theta, \mathbf{u})} \right| \right\} \\ &= \min \left\{ 1, \frac{\mathcal{L}_n^{M_k}(\theta_k) \pi_0(M_k) \overleftarrow{r}_j}{\mathcal{L}_n^{M_j}(\theta_j) \pi_0(M_j) \overrightarrow{r}_j} \right\}\end{aligned}$$

as $\pi_0(\theta_k | M_k)$ can be written

$$\prod_{l=1}^k \pi_0(\theta_{kl}) = \left\{ \prod_{l=1}^j \pi_0(\theta_{jl}) \right\} \left\{ \prod_{l=1}^{k-j} \pi_0(u_l) \right\} = \pi_0(\theta_j | M_j) p_U(\mathbf{u})$$

For the reverse move that attempts to decrease the model complexity by moving from model k to model j , the deterministic proposal that sets the last $k - j$ components of θ_k to zero is used. The approach described in this example can be useful in some applications, but can also lead to low acceptance rates.

Example

Consider two models M_1 and M_2 with parameters $\theta^{(1)} = \theta_1$ and $\theta^{(2)} = (\theta_{21}, \theta_{22})$, with all parameters taking values on $\mathbb{R}$. We consider four move types:

1. $m = 1$: move **within** Model M_1 ,
2. $m = 2$: move **within** Model M_2 ,
3. $m = 3$: move **from** Model M_1 **to** Model M_2 ,
4. $m = 4$: move **from** Model M_2 **to** Model M_1 .

For the within model moves, standard MH acceptance calculations proceed as usual; Moves 3 and 4 are a forward/reverse move pair.

Clearly, if the current state of the chain is in M_1 , only moves 1 or 3 can be selected, and similarly for M_2 .

Example

We need only consider the relative magnitudes of the reversible jump moves selection probabilities; for example, if the probability of selecting move 3 is 0.3 and the probability of selecting move 4 is 0.1, then

$$\frac{\overleftarrow{r}_3}{\overrightarrow{r}_3} = \frac{0.1}{0.3} \qquad \frac{\overleftarrow{r}_4}{\overrightarrow{r}_4} = \frac{0.3}{0.1}$$

are the ratios that enter into the acceptance probability calculations for move types 3 and 4 respectively.

Example

When the current state of the chain is in M_1 , to propose a move to M_2 after selecting move 3, we must introduce a single random variate u ; suppose that $u \sim \text{Normal}(0, 1)$, and let

$$\theta_{21} = \theta_1 + u \quad \theta_{22} = \theta_1 - u \quad \Longleftrightarrow \quad \theta_1 = \frac{\theta_{21} + \theta_{22}}{2}.$$

In this case the acceptance probability takes the form

$$\alpha_3 = \min \left\{ 1, \frac{\pi_n(M_2, (\theta_{21}, \theta_{22}))}{\pi_n(M_1, \theta_1) \phi(u)} \frac{\overleftarrow{r}_3}{\overrightarrow{r}_3} \left| \frac{\partial(\theta_{21}, \theta_{22})}{\partial(\theta_1, u)} \right| \right\}.$$

where $\phi(\cdot)$ is the standard normal pdf, and

$$\left| \frac{\partial(\theta_{21}, \theta_{22})}{\partial(\theta_1, u)} \right| = \left| \begin{array}{cc} \frac{\partial \theta_{21}}{\partial \theta_1} & \frac{\partial \theta_{21}}{\partial u} \\ \frac{\partial \theta_{22}}{\partial \theta_1} & \frac{\partial \theta_{22}}{\partial u} \end{array} \right| = \left| \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right| = 2.$$

Example

For the reverse move 4, we have

$$\alpha_4 = \min \left\{ 1, \frac{\pi_n(M_1, \theta_1) \phi(u)}{\pi_n(M_2, (\theta_{21}, \theta_{22}))} \frac{\overleftarrow{r}_4}{\overrightarrow{r}_4} \left| \frac{\partial(\theta_1, u)}{\partial(\theta_{21}, \theta_{22})} \right| \right\}.$$

where

$$\left| \frac{\partial(\theta_1, u)}{\partial(\theta_{21}, \theta_{22})} \right| = \frac{1}{2}.$$

Pharmacokinetic example

Consider two competing pharmacokinetic (PK) models for response data $y(t)$ measured at different time points $t_1, \dots, t_n$:

1. **Model M_1 :** One compartment, elimination only;

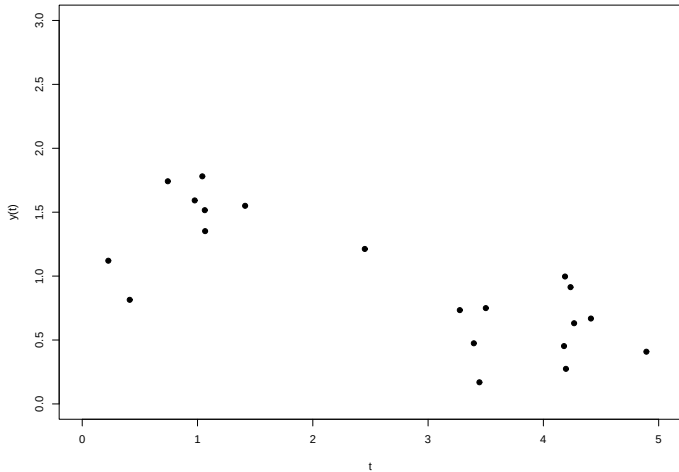
$$\mathbb{E}[Y(t)] = A_1 \exp\{-\lambda_1 t\} \quad t \geq 0$$

2. **Model M_2 :** One compartment, absorption and elimination;

$$\mathbb{E}[Y(t)] = A_2 (\exp\{-\lambda_{21} t\} - \exp\{-(\lambda_{21} + \lambda_{22})t\}) \quad t \geq 0$$

where (A_1, λ_1) and $(A_2, \lambda_{21}, \lambda_{22})$ are positive parameters. Under an assumption of additive, heteroscedastic Normal errors, we have two competing explanations for the observed data; both models can be fitted using ordinary least-square

Pharmacokinetic example



Pharmacokinetic example

These models are nested, but the nesting structure is complicated, as we require $\lambda_{22} \rightarrow \infty$, that is, a boundary point in the parameter space, which leads to non-regular frequentist asymptotic theory.

Bayes factors can be used:

$$BF(M_1, M_2) = \frac{\int \mathcal{L}_n^{M_1}(\theta_1) \pi_0(\theta_1 | M_1) d\theta_1}{\int \mathcal{L}_n^{M_2}(\theta_2) \pi_0(\theta_2 | M_2) d\theta_2}$$

but this requires numerical integration.

Pharmacokinetic example

We consider a reversible jump MCMC solution. First, we use a log-scale parameterization and set

$$\theta_1 = (\log A_1, \log \lambda_1) \qquad \theta_2 = (\log A_2, \log \lambda_{21}, \log \lambda_{22}).$$

We place equal prior probabilities on M_1 and M_2 , and then place independent $N(0, \tau)$ priors on the components of θ_1 and θ_2 .

The prior on the residual error variance σ^2 is Inverse Gamma with parameters 10 and 4.

Pharmacokinetic example

The ML estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ can be computed easily, as can the Hessian matrices $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$; these likelihood-based results yield reasonable approximations to the posterior distributions to produce independence MH algorithms.

Specifically, at the ML estimates for σ under the two models, we may approximate the conditional posterior for θ by the Normal density

$$p(\theta_k | \hat{\sigma}, \mathbf{y}) \simeq \text{Normal}(\hat{\theta}_k, n^{-1} \hat{\sigma}_k^2 \hat{\mathbf{I}}_k^{-1}) \quad (1)$$

On fitting using ML, the estimates of σ under the two models are found to be quite similar ($M_1 : \hat{\sigma}_1 = 0.252, M_2 : \hat{\sigma}_2 = 0.329$).

Pharmacokinetic example

A reversible jump MCMC algorithm can be constructed as follows: we again consider four move types:

1. $m = 1$: move **within** M_1 ; update θ_1 from $\pi_n(\theta_1|M_1, \sigma)$
2. $m = 2$: move **within** M_1 ; update θ_2 from $\pi_n(\theta_2|M_2, \sigma)$
3. $m = 3$: move **from** M_1 **to** M_2 ; propose a new θ_2 , and carry out an accept/reject step.
4. $m = 4$: move **from** Model M_2 **to** Model M_1 ; propose a new θ_1 , and carry out an accept/reject step.

with the remaining parameter σ^2 being updated in a Gibbs sampler algorithm at each iteration.

Pharmacokinetic example

Moves $m = 3, 4$ are a forward/reverse move pair. For move 3, several options are available; for example, we could adopt the earlier strategy, and generate a new variate u from the prior for the additional parameter, and then merely use the mapping

$$(\theta_{11}, \theta_{12}, u) \longmapsto (\theta_{21} = \theta_{11}, \theta_{22} = \theta_{12}, \theta_{23} = u)$$

with reverse move setting $\theta_{23} = 0$.

This approach may be adequate, but more probably would not facilitate good mixing across the models.

Pharmacokinetic example

A perhaps better strategy is to consider a different augmentation, where we generate $\mathbf{u} = (u_1, u_2, u_3)$ from the model in (1) for $k = 2$, and map $(\theta_{11}, \theta_{12}, u_1, u_2, u_3)$ to

$$(\theta_{21} = u_1, \theta_{22} = u_2, \theta_{23} = u_3, v_1 = \theta_{11}, v_2 = \theta_{12})$$

with the paired reverse move being to generate $\mathbf{v} = (v_1, v_2)$ from the model in (1) for $k = 1$.

This guarantees that the proposed value θ_2 lies in a region with reasonably high posterior support under model M_2 , although it does not guarantee that the move will be accepted with high probability.

Pharmacokinetic example

In the Hastings ratio, the Jacobian of the transformation is 1, and under equal probabilities of forward/reverse moves, we have that

$$\frac{\pi_n(M_2, \theta_2) p_V(v_1, v_2)}{\pi_n(M_1, \theta_1) p_U(u_1, u_2, u_3)}$$

can be written

$$\frac{\mathcal{L}_n^{M_2}(\theta_2) \left\{ \prod_{j=1}^3 \phi(\theta_{2j}/\tau)/\tau \right\} \phi_2(\theta_{11}, \theta_{12}; \hat{\theta}_1, \hat{\mathbf{I}}_1)}{\mathcal{L}_n^{M_1}(\theta_1) \left\{ \prod_{j=1}^2 \phi(\theta_{1j}/\tau)/\tau \right\} \phi_3(\theta_{21}, \theta_{22}, \theta_{23}; \hat{\theta}_2, \hat{\mathbf{I}}_2)}$$

where τ is the prior variance for the regression parameters.

Pharmacokinetic example

The logic of this construction is that numerically

$$\mathcal{L}_n^{M_1}(\theta_1) \simeq \text{Normal}_2(\hat{\theta}_1, \hat{\mathbf{I}}_1^{-1}) \quad \mathcal{L}_n^{M_2}(\theta_2) \simeq \text{Normal}_3(\hat{\theta}_2, \hat{\mathbf{I}}_2^{-1}).$$

A different approximation that incorporates the Normal asymptotic likelihood in equation (1) as well as the Normal prior distribution for the parameters can be constructed.

Pharmacokinetic example

The algorithm was run for 100000 iterations. In this run with $\tau = 4$, the chain spent about 66 % of the time in model M_1 , indicating the posterior probabilities are

$$\pi_n(M_1) \simeq 0.66 \quad \pi_n(M_2) \simeq 0.34.$$

The model posterior probabilities vary with the choice of τ ; this is as expected, as the model probabilities are closely related to the marginal likelihood, or prior predictive distribution, which is the expected value of the likelihood for the observed data with respect to the prior distribution.

Pharmacokinetic example

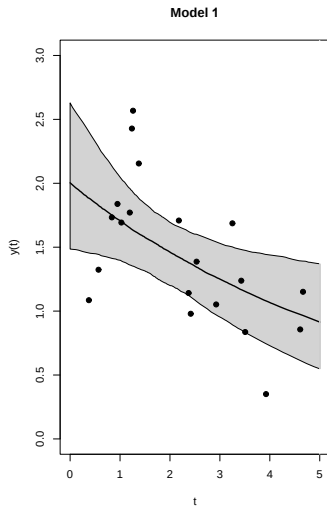
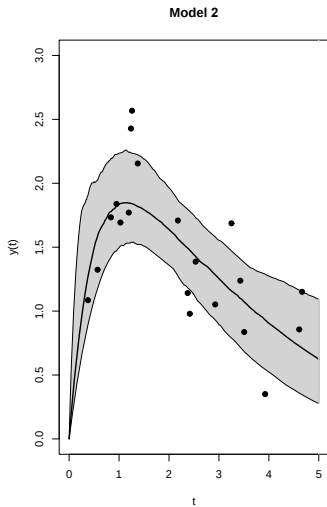
It is evident from the discussion that the prior specification acts as a penalty for complexity. For illustration, if $\tau = 1$, the model probabilities change to $(0.80, 0.20)$; if $\tau = 10$, the model probabilities are $(0.26, 0.74)$.

Pharmacokinetic example

Conditional on M_1 or M_2 being true, we can perform inference about the parameters of the two models, and also reconstruct estimates and posterior credible intervals for $\mathbb{E}[Y]$.

The figure below displays the reconstructed posterior intervals for the two models.

Pharmacokinetic example



Pharmacokinetic example

It is evident that both models are plausible explanations of the observed data, but offer potentially different predictions of future responses, especially near $t = 0$. Note that the possible nesting structure, incorporated by the assumption that $\lambda_{22} \rightarrow \infty$, is not supported by the priors studied.

Finally, note that the BIC values for the two models, computed as

$$\text{BIC}_k = -2 \log \mathcal{L}_n^{M_k}(\hat{\theta}_1) + d_k \log n$$

where d_k is the total number of parameters fitted, are 25.553 and 33.227 respectively for models 1 and 2, indicating strong support for model 1.

Appendix: Trans-dimensional examples

Finite Mixture Model

The finite mixture model with K components has density

$$f(y|\omega, \theta, K) = \sum_{k=1}^K \omega_k f_k(y|\theta_k)$$

where θ_k are the parameters for component density k , and

$$0 < \omega_k < 1, \forall k \quad \& \quad \sum_{k=1}^K \omega_k = 1.$$

Finite Mixture Model

Most typically the component densities are from the same parametric location-scale family (say Gaussian) which differ in location and scale.

For exchangeable data $(y_1, \dots, y_n)$ from this model, the likelihood arising from this model does not permit analytical calculation of the posterior distribution, so MCMC is commonly used.

Finite Mixture Model

If K is known, MCMC is typically implemented using auxiliary variables $z_1, \dots, z_n$, and a completed data model

$$f(y, z; \omega, \theta, K) = \omega_z f_z(y | \theta_z) \quad z \in \{1, 2, \dots, K\}$$

and a Gibbs sampler strategy updating the $\mathbf{z}$ and (ω, θ) from their full conditional posterior distributions.

Conditional on (ω, θ) , the posterior distribution for each z_i is a discrete distribution on $\{1, 2, \dots, K\}$, whereas conditional on the $\mathbf{z}$, the posterior for the mixture weights ω is a density on the K dimensional simplex, and the posterior for the components of θ_k proceeds using only those y_i for which $z_i = k$, for $k = 1, \dots, K$.

Finite Mixture Model

If K is also treated as an unknown parameter, then transdimensional MCMC is also needed. A discrete prior on K , typically on some finite set, completes the posterior specification, and yields a posterior distribution $\pi_n(K, \theta^{(K)}, \omega^{(K)})$.

Finite Mixture Model

In this model, dimension-changing moves correspond to changing the value of K , and in order to construct an efficient and tunable MCMC algorithm, moves which perturb K by ± 1 are typically considered.

- ▶ For example, moves of type $K \rightarrow K + 1$ are termed “birth” moves, and $K \rightarrow K - 1$ are termed “death” moves.
- ▶ The forward/reverse move pairs are then births from K and deaths from $K + 1$.

Finite Mixture Model

We consider the case of a Normal mixture model for simplicity.

- ▶ If y is one dimensional, each component density has two parameters, so a birth/death pair requires the introduction of three latent variables $\mathbf{u}$, a new mean and variance, and also a parameter to introduce a new component weight.
- ▶ If y is d -dimensional, $d + d(d+1)/2 + 1$ new scalar parameters are needed, that is, a new mean and covariance matrix, and a new component weight parameter.

Finite Mixture Model

- **New location/scale parameters:** Either these new parameters are generated independently of the current model parameters from proposal distribution, or a subset of the current components are selected and used to generate a new additional component.
- **New weight parameters:** A random split or rescaling of the currently existing weights is used.

Finite Mixture Model

In typical mixture applications, inference about the parameter K may be of primary interest in a given application, but perhaps more commonly the mixture model offers flexible modelling in the presence of heterogeneity. The Normal mixture case is the most commonly studied, but other mixtures of other parametric models have also been used.

Regression

Suppose the regression relationship between a scalar covariate x and scalar covariate y

$$y = g(x) + \epsilon$$

is the focus of interest, for independent homoscedastic Normal errors $\{\epsilon\}$. Typically, we seek the prediction $\hat{y} = \hat{g}(x)$.

Regression

Suppose we seek to approximate g locally on a finite domain $\mathcal{X}$ by a series of step functions:

$$g_K(x; \boldsymbol{\beta}_K, \kappa_K) = \sum_{k=1}^K \beta_{Kk} \mathbb{1}_{B_k}(x)$$

where $(B_1, \dots, B_K)$ form a partition of $\mathcal{X}$ defined by a series of knots $\kappa_0, \kappa_1, \dots, \kappa_{K-1}, \kappa_K$ so that $B_k \equiv [\kappa_{k-1}, \kappa_k)$. Without loss of generality, assume $\mathcal{X} \equiv [0, 1)$, with $\kappa_0 = 0, \kappa_K = 1$. In this step-function approximation, $\boldsymbol{\beta}_K = (\beta_{K1}, \dots, \beta_{KK})$ are the parameters defining the piecewise constant levels of the function.

Regression

The likelihood for data $(x_i, y_i), i = 1, \dots, n$ is

$$\mathcal{L}_n^{M_K}(\beta_K, \sigma, \kappa_K) = \prod_{i=1}^n \phi\left(\frac{y_i - g_K(x_i; \beta_K, \kappa_K)}{\sigma}\right)$$

Typically, within model M_K , a conjugate prior specification on (β_K, σ) is used so that the marginal likelihood $\mathcal{L}_n^{M_K}(\kappa_K)$ can be computed analytically by integrating out (β_K, σ) . The conjugate prior takes the form

$$\pi_0(\beta_K, \sigma) = \pi_0(\beta_K | \sigma) \pi_0(\sigma)$$

and $\pi_0(\beta_K | \sigma)$ is some multivariate Normal density.

Regression

A reversible jump MCMC algorithm can be constructed to explore the posterior $\pi_n(M_K, \kappa_K)$ which, under the conjugate specification for (β_K, σ) , depends on the marginal likelihood and the prior $\pi_0(M_K, \kappa_K)$.

One approach to the specification of this prior is to assume a Poisson process model for the change-in-level or jump locations of the step function.

- ▶ $\pi_0(M_K)$ to be a Poisson distribution with rate parameter λ ,
- ▶ within model M_K , κ_K are the order statistics derived from a Uniform random sample on $(0, 1)$, so that

$$\pi_0(\kappa_{K1}, \dots, \kappa_{K,K-1} | M_K) = (K-1)!$$

for $0 < \kappa_1 < \dots < \kappa_{K,K-1} < 1$.

Regression

Dimension-changing moves in this model correspond to the addition or removal of a knot or knots.

- a new knot can be proposed uniformly on $(0,1)$, or uniformly in an interval selected in light of the current knot positions. In the former case, the reverse move corresponds to removing a knot uniformly at random from the current collection.
- Model complexity in this model is controlled by the Poisson model hyperparameter λ , and the covariance structure in the prior model for β_K .