The University of Newcastle

Kerrie Mengersen

Introduction to

Bayesian Methods for

QTL Analysis - 1

Armidale 2004

Plan: Day 1

1. Review of Bayesian modelling, MCMC and mixtures
2. Bayesian QTL mapping: overview
3. Workshop: estimating mixtures
4. Bayesian QTL mapping: examples

Plan: Day 2

1. Bayesian QTL mapping: examples
2. Literature review
3. Workshop: Up close to the literature
4. Software for Bayesian QTL analysis

General Bayesian approach

Data y

likelihood L(y|θ)

eg, $y \sim N(\mu,\sigma)$

or $y \sim Bin(n,p)$

Info. about parameters θ

prior p(θ)

$\mu \sim N(a,b)$ or $\mu \sim U(a,b)$,
$\sigma^{-2} \sim Ga(c,d)$

$p \sim Beta(0,1)$

General Bayesian approach

Data y

likelihood $L(y|\theta)$

eg, $y \sim N(\mu,\sigma)$
or $y \sim Bin(n,p)$

Info. about parameters θ

prior $p(\theta)$

eg, $\mu \sim N(a,b)$ or $\mu \sim U(a,b)$,
$\sigma^{-2} \sim Ga(c,d)$

Posterior dist'n of $\theta \propto$ likelihood * prior

$p(\theta|y) \propto L(y|\theta)\, p(\theta)$

Joint posterior: $p(\mu, \sigma, \ldots|y) \propto L(y|\, \mu, \sigma)\, p(\mu)\, p(\sigma)\ldots$

Bayesian Computation

- We want to estimate the expected value of some function of our parameters θ (eg means of μ, σ)
- When the modelling becomes more complex or the distributions are not 'easy' or we want more complicated expectations we can use simulation: simulate each parameter *given* the other parameters and the data

MCMC
Markov chain Monte Carlo

MCMC Algorithms

- Need to ensure conditions, eg *detailed balance, reversibility*

BUGS

Three current trends:

- Complex hierarchical (random-effects) models being analysed using S-plus, SAS etc
- Graphical models used in multivariate analysis
- Markov chain Monte Carlo (MCMC) methods turning Bayesian into mainstream statistics

Brought together in BUGS:

Bayesian Inference Using Gibbs Sampling

CODA

- **Output processor for BUGS**
- **Menu-driven set of S-Plus functions for:**
- **Convergence diagnosis**
  - specific methods
  - autocorrelations and cross-correlations
- **Summary statistics**
  - empirical mean, sd, quantiles
  - standard error of the mean
- **Graphical**
  - sample trace for each variable
  - kernel density
  - plots of some convergence diagnostics

Bayesian Mixed Models

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

β is a fixed vector

$\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ are uncorrelated random vectors

$\mathbf{X}$, $\mathbf{Z}$ are incidence matrices

$\mathbf{G}$, $\mathbf{R}$ are variance-covariance matrices, which are functions of (known) dispersion parameters.

The vector of random effects $\mathbf{u}$ can include herd effects, breeding values, permanent environmental deviations common to all records of the same (or of a set) of animals etc.

Source: D. Gianola, Inferences about Breeding Values. In Balding et al (eds) Handbook of Statistical Genetics, 2001.

Joint density

$$p(\mathbf{u},\mathbf{y}|\beta,\mathbf{G},\mathbf{R}) \propto p(\mathbf{y}|\mathbf{u},\beta,\mathbf{R})\, p(\mathbf{u}|\mathbf{G})$$

$$\propto \exp\{-\tfrac{1}{2}[(\mathbf{y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{u})+\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}]\}$$

Clear link between mixed models and hierarchical Bayesian models!
Connections between BLUP and Bayes.

Bayesian view of BLUP

- Assume **G** and **R** are known.
- Uniform prior for β over p-dimensional space (p is the order of β)

→ Joint posterior is Gaussian, so marginals and conditional distributions are Gaussian.

→ Any linear combination of β and **u** also have a Gaussian posterior distribution.

Example

- Suppose we want to infer a vector of merits or ‘aggregate genetic values **h=Mu**, of a set of candidates.
  M is a constant matrix reflecting the relative economic importance of traits
  u is a vector of multitrait genetic values.
- Posterior distribution of **h** is Gaussian, with mean vector **h=Mu** and covariance $\mathbf{MC}_u\mathbf{M}'$, where $\mathbf{C}_u$ is a submatrix of **u**.

Example

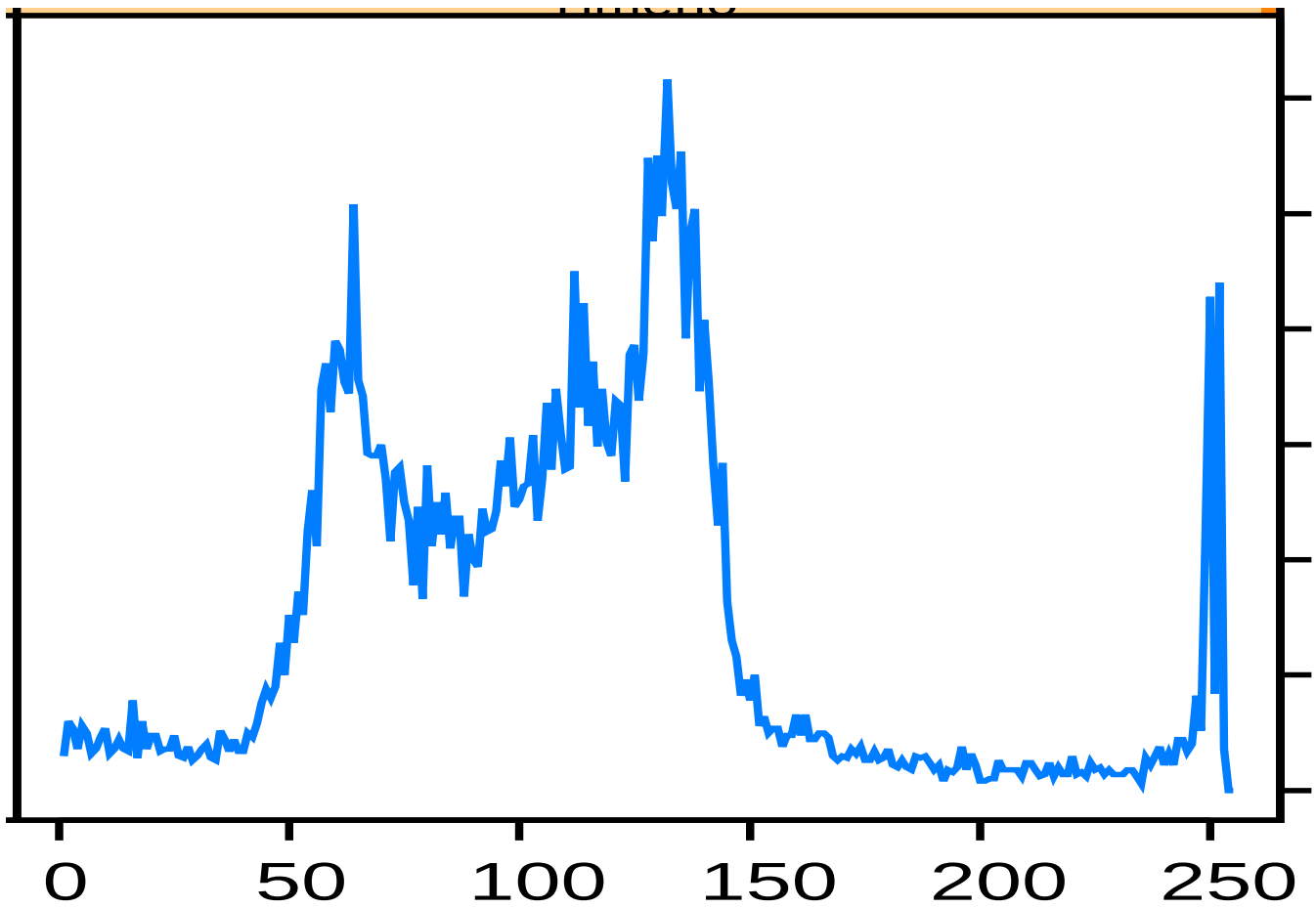
- Infer nonlinear merit, eg

  $$h = \mathbf{m}'\mathbf{u} + \mathbf{u}'\mathbf{Q}\mathbf{u}$$

  m and **Q** known.
- Now the posterior distribution of h does not have a closed form, but we can estimate it via MCMC.

(Dan's course: Animal Breeding Summer School)

Estimating Mixtures



Bayesian mixture representation

$y \sim \Sigma_{j=1:k}\, p_j\, \mathrm{N}(\mu_j, \tau_j)$

$p \sim$ Dirichlet

$\mu \sim$ Normal

$\tau \sim$ Gamma

Introduce z for (unobserved) component group

1. Given z, estimate p, μ, τ

2. Given p, μ, τ, estimate z

Normal mixture example

$p\,\mathcal{N}(\mu_1, 1) + (1-p)\,\mathcal{N}(\mu_2, 1)$ p unknown

For the mixture (1.7), the parameter space is two-dimensional, which means that the posterior surface can be easily plotted. Under a normal prior $\mathcal{N}(\delta, 1/\lambda)$ ($\delta \in \mathbb{R}$ and $\lambda > 0$ are known hyper-parameters) on both μ_1 and μ_2, with $s_j^x = \sum_{i=1}^n \mathbb{I}_{z_i=j} x_i$, it is easy to see that μ_1 and μ_2 are independent, given $(\underline{z}, \underline{x})$, with conditional distributions

$$\mathcal{N}\left(\frac{\lambda\delta + s_1^x}{\lambda + n_1}, \frac{1}{\lambda + n_1}\right) \quad \text{and} \quad \mathcal{N}\left(\frac{\lambda\delta + s_2^x}{\lambda + n_2}, \frac{1}{\lambda + n_2}\right)$$

respectively. Similarly, the conditional posterior distribution of $\underline{z}$ given (μ_1, μ_2) is easily seen to be a product of Bernoulli rv's on $\{1, 2\}$, with $(i = 1, \ldots, n)$

$$\mathbb{P}(z_i = 1 | \mu_1, x_i) \propto p \exp\left(-0.5\,(x_i - \mu_1)^2\right).$$

Gibbs for Normal mixture

0. **Initialization.** Choose $\mu_1^{(0)}$ and $\mu_2^{(0)}$.

1. **Step t.** For $t = 1, \ldots$

   1.1 Generate $z_i^{(t)}$ $(i = 1, \ldots, n)$ from

$$\mathbb{P}\left(z_i^{(t)} = 1\right) = 1 - \mathbb{P}\left(z_i^{(t)} = 2\right) \propto p \exp\left(-\frac{1}{2}\left(x_i - \mu_1^{(t-1)}\right)^2\right)$$

   1.2 Compute $n_j^{(t)} = \sum_{i=1}^{n} \mathbb{I}_{z_i^{(t)}=j}$ and $(s_j^x)^{(t)} = \sum_{i=1}^{n} \mathbb{I}_{z_i^{(t)}=j} x_i$

   1.3 Generate $\mu_j^{(t)}$ $(j = 1, 2)$ from $\mathcal{N}\left(\frac{\lambda\delta + (s_j^x)^{(t)}}{\lambda + n_j^{(t)}}, \frac{1}{\lambda + n_j^{(t)}}\right)$.

BUGS code

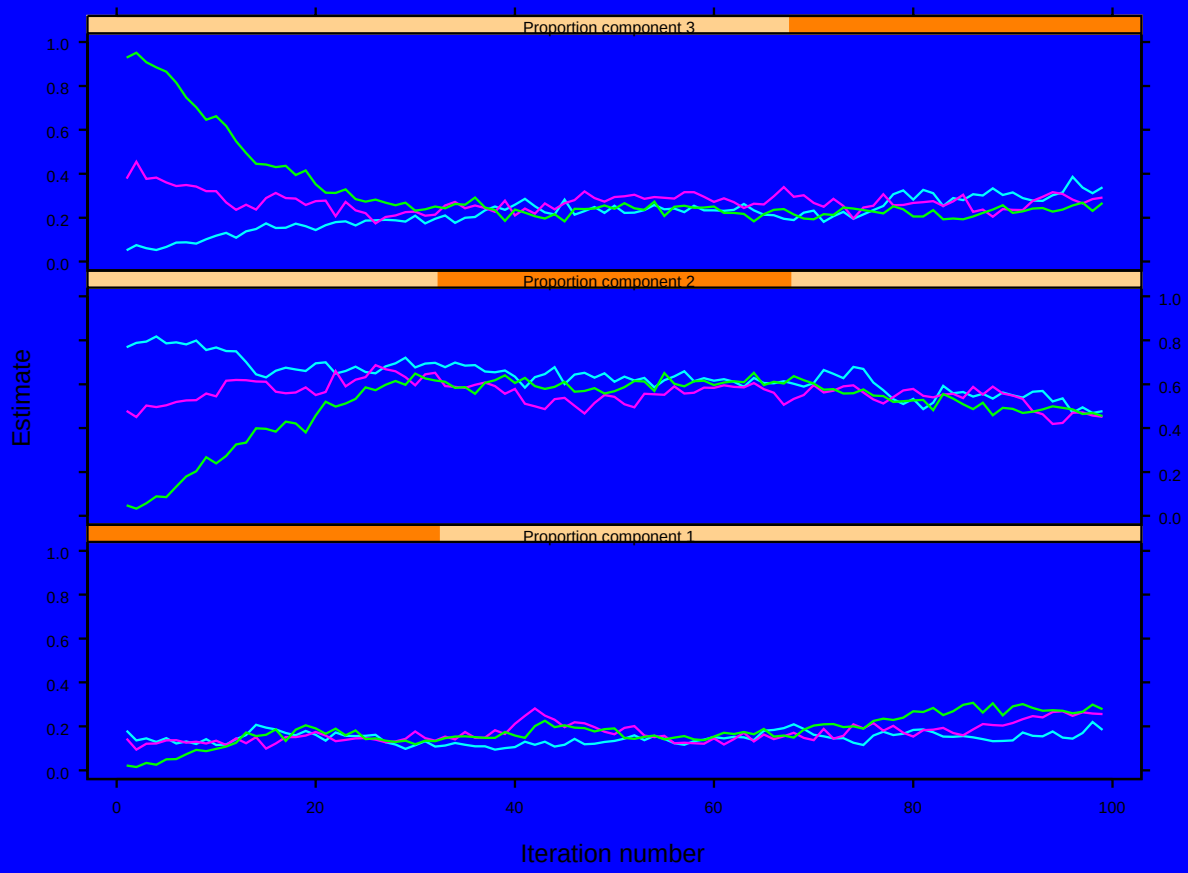
```
model
    {
        for( i in 1 : N ) {
            y[i] ~ dnorm(mu[i], tau[T[i]])
            mu[i] <- lambda[T[i]]
            T[i] ~ dcat(P[])
        }
        P[1:3] ~ ddirch(alpha[])
        lambda[3] ~ dnorm(0.0, 1.0E-6)I(lambda[2], )
        lambda[2] ~ dnorm(0.0, 1.0E-6)I(lambda[1], )
        lambda[1] ~ dnorm(0.0, 1.0E-6)
        tau[3] ~ dgamma(0.001, 0.001)  sigma[3] <- 1 / sqrt(tau[3])
        tau[2] ~ dgamma(0.001, 0.001) sigma[2] <- 1 / sqrt(tau[2])
        tau[1] ~ dgamma(0.001, 0.001) sigma[1] <- 1 / sqrt(tau[1])
}
```

BUGS Results

A 1000 update burn in followed by a further 20000 updates using 3 chains gave the parameter estimates

Param	Mean	SD	Actual
λ_1	0.18	0.053	0.18
λ_2	0.57	0.068	0.55
λ_3	0.25	0.042	0.27
μ_1	34.8	1.333	35.6
μ_2	45.9	0.699	45.8
μ_3	63.3	2.263	61.7
σ_1	4.03	0.683	4.59
σ_2	5.32	0.535	5.93
σ_3	11.34	1.106	11.89

Trace Plots for some parameters



How many components?

- Separate model estimation: Bayes factors, posterior odds, BIC, DIC, …
- Simultaneous model estimation: Carlin and Chib, reversible jump MCMC, birth-and-death
- "Distance" measures, loss functions, etc

Marginal likelihood

- Bayes formula: $p(\theta|y) = p(y|\theta)p(\theta)/p(y)$

  ie $\log[p(y)] = \log[p(y|\theta)] + \log[p(\theta)] - \log[p(\theta|y)]$

Marginal likelihood *conditional likelihood* *penalty favours parsimony*

- Marginal likelihood $p(y|M_k)$ is probability of data given model M_k, averaged over the priors assigned to the parameters in that model.

$$p(y|M_k) = \int P(y|M_k,\theta_k)p(\theta_k|M_k)d\theta_k, \quad k=1, 2, \ldots, K$$

Bayes factors

- Consider models $M_1, \ldots, M_K$ (not necessarily nested)
- The Bayes factor for model 2 compared to model 1 is the ratio of marginal likelihoods

$$B_{21} = p(y|M_2) / P(y|M_1)$$

- $2\log(B_{21})$ gives same scale as usual deviance and LR statistics.

Guidelines for Bayes Factors (arbitrary!)

B_{21}	$2\log(B_{21})$	Interpretation
<1	Negative	Supports M_1
1 to 3	0 to 2	Weak support for M_2
3-20	2-6	Supports M_2
20-150	6-10	Strong evidence for M_2
>150	>10	Very strong support for M_2

Posterior odds

- Posterior probability of a model:
  $P(M_k|y) = P(M_k)\, P(y|M_k) / P(y)$
- Posterior odds of model 1 compared to model 2:

$$\frac{P(M_1 \mid y)}{P(M_2 \mid y)} = \frac{P(M_1)}{P(M_2)} \frac{P(y \mid M_1)}{P(y \mid M_2)}$$

ie, the ratio of the prior probabilities for each model, multiplied by the Bayes factor.

(BF is only defined when the marginal density of y under each model is proper.)

Posterior odds

- Posterior probability of a model:
  $P(M_k|y) = P(M_k)\, P(y|M_k) / P(y)$
- Posterior odds of model 1 compared to model 2:
$$\frac{P(M_1 \mid y)}{P(M_2 \mid y)} = \frac{P(M_1)}{P(M_2)} \frac{P(y \mid M_1)}{P(y \mid M_2)}$$
- To estimate θ, we can 'model-average':

$$E(\theta|y) = \Sigma\, w_k\, \mu_k$$

$$Var(\theta|y) = \Sigma_k\, [\, var(\theta_k|y,M_k) + \mu_k^2] - \{E(\theta|y)\}^2$$

$\mu_k = E(\theta_k|y,M_k)$ (posterior mean for θ under model k)

$w_k = Pr(y|M_k)/\, \Sigma_k Pr(y|M_k)$ (weight for model k)

Example

Human chromosomes: males XY, females XX

Haemophilia exhibits X-chromosome-linked recessive inheritance, so a male who inherits the gene on the X chromosome is affected but a female who carries the gene on only one of the two X chromosomes is unaffected. The disease is usually fatal for women who inherit two such genes, and this is very rare, since the frequency of occurrence of the gene is low in human populations.

Source: Gelman et al (1995) Bayesian Data Analysis

Example: the prior distribution

A woman has an affected brother, which implies that her mother must be a carrier of the haemophilia gene with one 'good' and one 'bad' haemophilia gene.

Her father is not affected.

Thus the woman has a 50-50 chance of having the gene.

Unknown quantity of interest: whether the woman is a carrier of the gene ($\theta=1$) or not ($\theta=0$).

Based on the information provided so far, the prior distribution for the unknown θ is

$$\Pr(\theta=1) = \Pr(\theta=1) = 1/2$$

Example: model and likelihood

We need some data: the woman has two sons, neither of whom is affected.

Let y_i=*1 or 0* denote affected/unaffected son.

The outcomes of the two sons are *exchangeable* and, conditional on the unknown θ, are independent: we assume the sons are not identical twins.

→likelihood function

$$\Pr(y_1=0, y_2=0|\theta=1) = (0.5)(0.5) = 0.25$$

$$\Pr(y_1=0, y_2=0|\theta=0) = (1)(1) = 1$$

(OK, there is a nonzero probability due to mutation but we will ignore this)

Example: posterior distribution

$$Pr(\theta=1|y) = p(y|\theta=1)p(\theta=1) / p(y)$$

$$p(y) = p(y|\theta=1)p(\theta=1) + p(y|\theta=0)p(\theta=0)$$
$$= \Sigma\, p(y|\theta)p(\theta)$$

So:

$$Pr(\theta=1|y) = (0.25)(0.5) / \{(0.25)(0.5)+(1.0)(0.5)\}$$
$$= 0.125 / 0.625 = 0.20$$

In terms of odds:

Prior odds of woman being a carrier is 0.5/0.5=1.

Likelihood ratio based on information about unaffected sons is 0.25/1 = 0.25

So posterior odds are $0.25 \times 1 = 0.25$.

Converting back to a probability: 0.25/(1+0.25) = 0.2

Example: Bayes factor

- Models: M_1: woman is affected ($\theta=1$)
  M_2: woman is unaffected ($\theta=0$)
- Prior odds: $P(M_2)/p(M_1) = 1$
- Bayes factor of the data that the woman has two unaffected sons is
  $p(y|M_2) / p(y|M_1) = 1.0 / 0.25$
- Posterior odds are
  $p(M_2|y) / p(M_1|y) = 4$
- Clear accumulation of evidence supporting M_2.
- Note: BF make sense in this example because each of the discrete alternatives makes sense and the marginal distributions of the data under each model, $p(y|M_i)$ are proper.

Bayesian Information Criterion (BIC)

- Approximate the Bayes factor by a Laplace approximation to exploit standard output from GLIM, SAS etc
- Assume the prior $p(\theta|M)$ is $MVN(\theta^*,I)$ (I=expected information matrix for a single observation, so the prior is equivalent to a single extra observation). Then if p is the dimension of the model and with n observations (or an appropriate definition of n):

$$BIC = \log P(y|\theta^*,M) - p/2\log n$$

Bayesian Information Criterion as described by Ball (2002?)

Firstly the paper gives the Bayesian information criterion, from which approximate probabilities for models can be found

$$\mathrm{BIC} = n\log(1 - R^2) + k\log(n)$$

where here k is the number of parameters in the model, and n is the number of observations. It has been shown that with no prior, then the posterior probability p has the property

$$p \propto \exp(-\mathrm{BIC}/2)$$

however this relationship relies on large sample sizes. A modification has been previously suggested, BIC-δ,

$$\text{BIC-}\delta = n\log(1 - R^2) + k\delta\log(n)$$

with δ a constant. The choice of δ is dependent on sample size and other unreserached factors, however BROMAN (1997) suggests $\delta = 2$ or $\delta = 3$.

Discussion of BIC

- BIC penalises models which improve fit at the expense of more parameters (encourages parsimony).
- Problem is that the true dimensionality (number of parameters p) of the model is not known, and also that the number of parameters may increase with sample size n.
- Can approximate using the effective number of parameters (Speigelhalter et al, 1999).
- Alternatives are DIC (deviance information criterion), conditional posterior predictive probabilities, etc.

Carlin and Chib (1995)

Problem:

- several models $M_1,\ldots, M_k$ with dimensions $d_1,..,d_k$
- Prior probability $p(j)$ that model j is the true one
- We want best model index M and posterior densities of $\theta_1,\ldots,\theta_k$

Solution:

At each iteration, estimate superparameter $\theta.S=\{\theta_1,.., \theta_k\}$ and 'best model' index $M=j$

Carlin and Chib (1995)

Joint density: $P(y,\theta.S,j) = p(y|\theta.S,j)p(\theta,S|j)P(j)$

- $P(j)$ is prior probability that model *j* is the true one. Typically we might take $P(1)=P(2)=\ldots=P(K)=1$
- Given non-overlapping parameters, $P(y|\theta.S,j)=p(y|\theta_j,j)$, and if the parameters of different models are conditionally independent given one of them is selected, then $P(\theta.S|j)=\Pi_{i=1,..,K}P(\theta_i|j)$, so $P(\theta.S|j)=\Pi p(\theta_i|j)$ (pseudo-prior). (Usually, take common pseudo-priors for all models or use pilot runs to provide parameters.)

Example

- Onion bulb growth data (Gelfand et al, 1992): choose between a Gompertz and logistic growth model for the onion bulb evolution through time.
- Carry out separate pilot runs with a nonlinear regression to estimate Gompertz growth curve parameters θ^G and another to estimate logistic parameters θ^L with estimated precisions T^G, T^L.
- Use an informative prior based on these estimates as the pseudo-prior and a (considerably) less prior centred on these estimates as the true prior. Thus the true prior for the Gompertz (when the Gompertz model is selected) might be $\theta \sim N(\theta^G, C/T^G)$, C=1000 (say), and the pseudo prior for the Gompertz parameters when the logistic model is selected is $\theta \sim N(\theta^G, 1/T^G)$.
- With equal prior model odds, the BF is 0.957/0.043=22.2, in favour of the logistic model.

Congdon, pp.482-4)

RJMCMC

- RJMCMC (Green 1995): called *reversible jump* because it is based on a *reversibility* constraint on the dimension-changing moves that bridge the different spaces.
- The only real difficulty compared with previous algorithms is to construct moves between the dimensions.
- Reversibility can be processed at a local level: since the model indicator μ is an integer-valued random variable, we can impose reversibility for each pair (k_1,k_2) of the model space.
- Core idea: supplement each of the spaces 1 and 2 with adequate artifical spaces in order to create a *bijection* ("bridge") $T_{k1 \to k2}$ between them.

Example: Mixture of normals

Start with jth component in a model with k components
Split a component, to give a model with k+1 components

$$\mathfrak{M}_k : \sum_{j=1}^{k} p_{jk} N(\mu_{jk}, \sigma_{jk}^2)$$

[Richardson & Green, 1997]

Moves:

(i). Split

$$\begin{cases} p_{jk} &= p_{j(k+1)} + p_{(j+1)(k+1)} \\ p_{jk}\mu_{jk} &= p_{j(k+1)}\mu_{j(k+1)} + p_{(j+1)(k+1)}\mu_{(j+1)(k+1)} \\ p_{jk}\sigma_{jk}^2 &= p_{j(k+1)}\sigma_{j(k+1)}^2 + p_{(j+1)(k+1)}\sigma_{(j+1)(k+1)}^2 \end{cases}$$

(ii). Merge (*reverse*)

Example (cont)

Additional Birth and Death moves for empty components (created from the prior distribution)

Equivalent

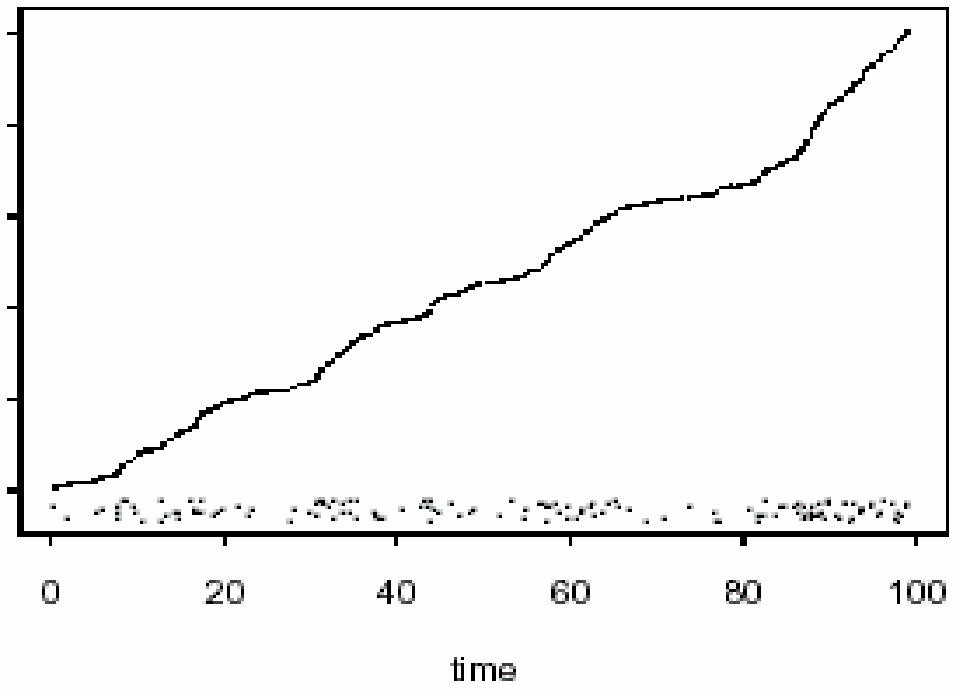
(i). Split

$$(T)\begin{cases} u_1, u_2, u_3 & \sim & \mathcal{U}(0,1) \\ p_{j(k+1)} & = & u_1 p_{jk} \\ \mu_{j(k+1)} & = & u_2 \mu_{jk} \\ \sigma^2_{j(k+1)} & = & u_3 \sigma^2_{jk} \end{cases}$$

Example: Change-point analysis

Example:
cyclones hitting the Bay of Bengal

141 cyclones over a period of 100 years
(a cyclone is a storm with winds > 88 km h^{-1}).



Example: Change-point analysis

- Observations (y_1, y_2, …, y_n) at time periods $t=1,..,n$.
  We want a 'step function' to represent the intensity of the cyclone.
  The step function will comprise a set of step heights h and step positions s.
- Likelihood: Poisson process $p(y|t)$.
- Prior for step function: $x(t)$ = flat?
- Prior for the number of steps k: $Poisson(\lambda)$

MCMC algorithm

I will use four moves:

(a) Metropolis change to a randomly chosen step height h_j.

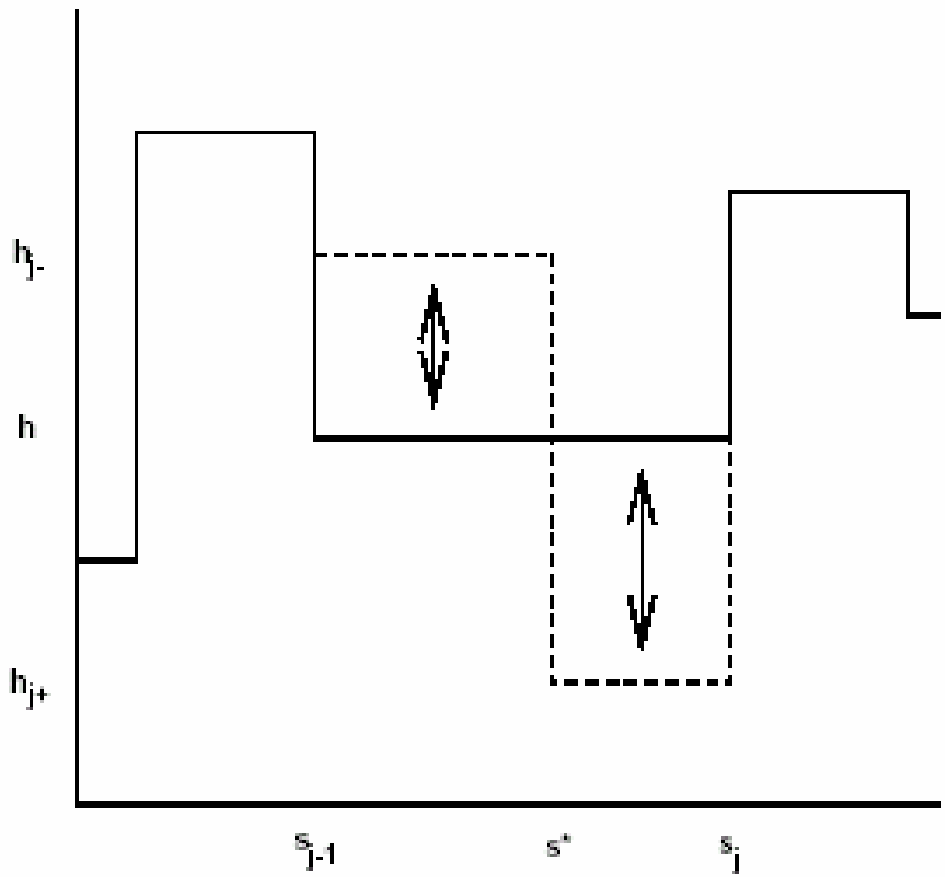
(b) Metropolis change to a randomly chosen step position s_j.

(c) Jump move: birth/death of steps

- birth: choose new step position s^* at random, split current step height h into two (h_-, h_+)
- death: choose step at random to kill, combine current step heights (h_-, h_+) into one: h

(d) Update hyperparameters α, β

Birth and death of steps



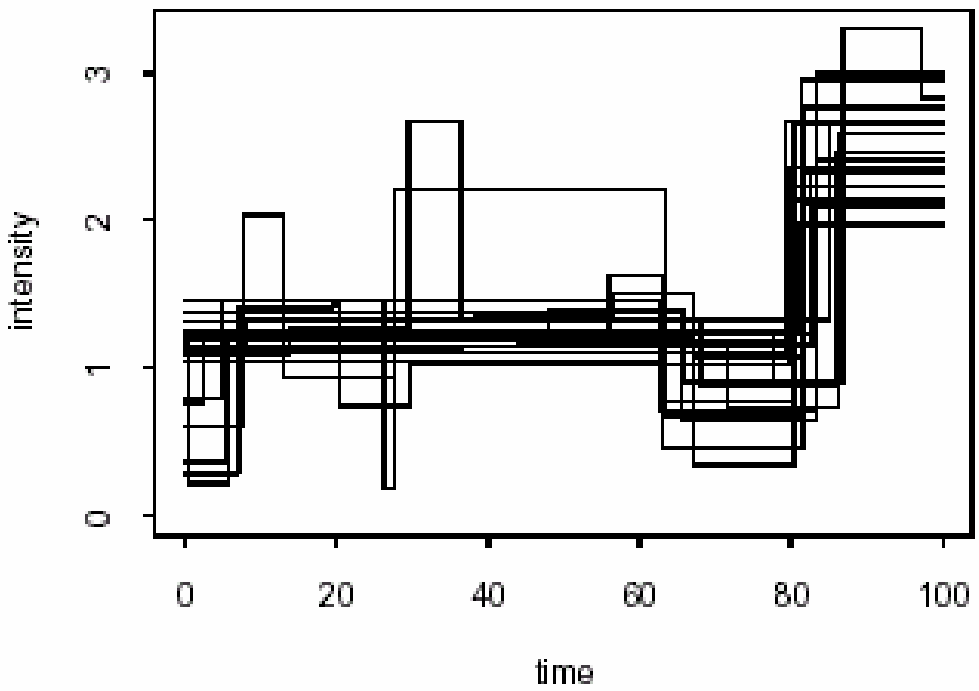
$$h_-^{w_-} h_+^{w_+} = h^{w_- + w_+}$$

$$(h, w, s^*, v) \leftrightarrow (h_-, h_+, w_-, w_+)$$

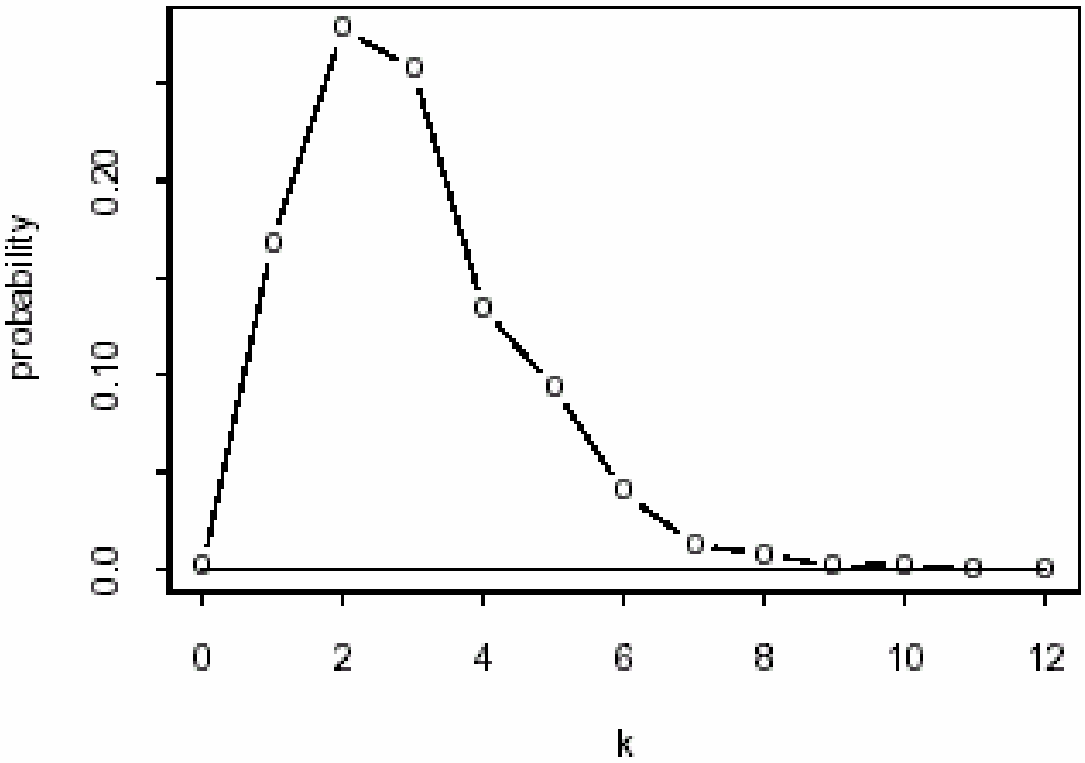
Choices of hyperparameters:

- Prior on k: Poisson(λ), with $\lambda = 3$.
- Prior on h_j: Gamma(α,β), with
  - $\alpha \sim \Gamma(2, 2)$
  - $\beta \sim \Gamma(1, n/L)$

Sample of step functions from the posterior:

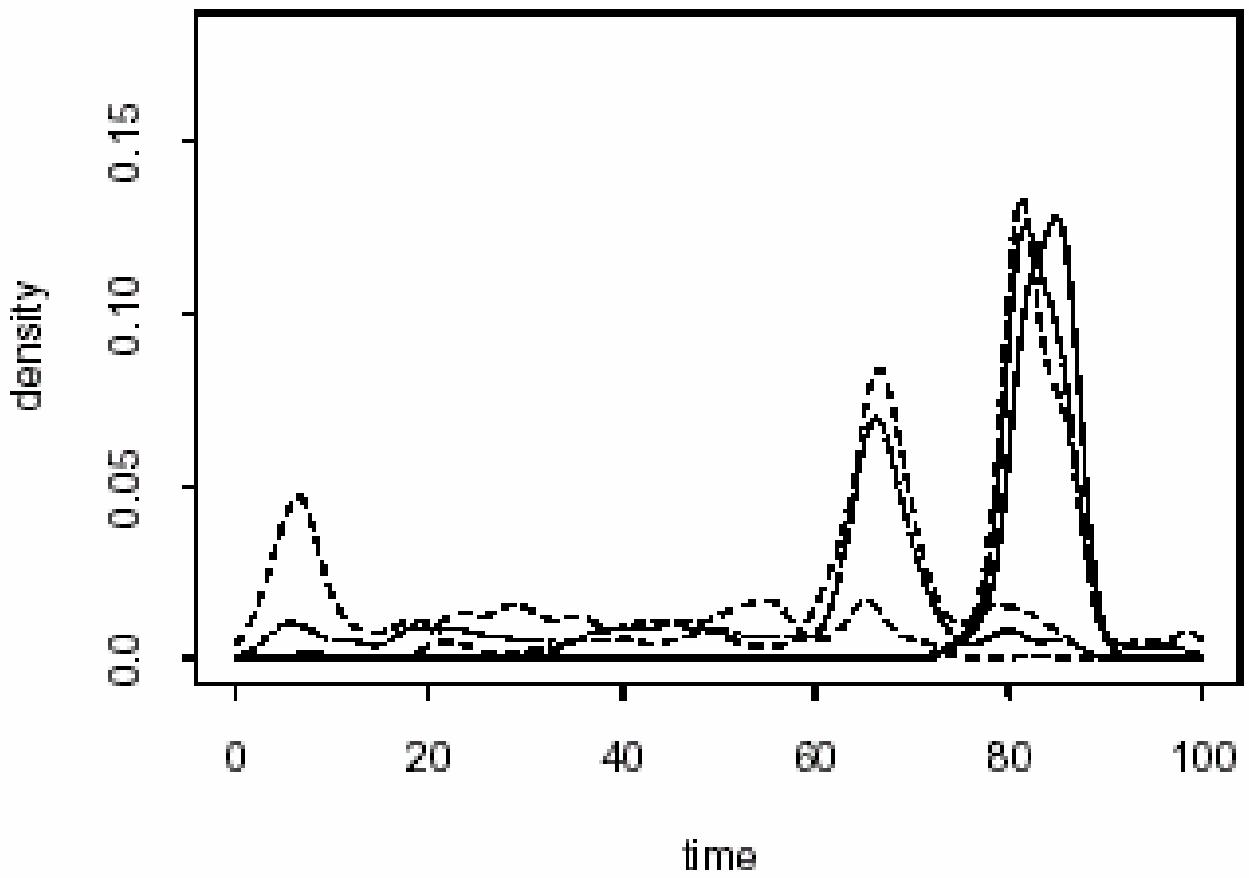


Posterior for the number of change points k:



Zero change points is ruled out; $k = 1$ or 2 more probable than under the prior.

Posterior density estimates for change-point positions



Model-averaged estimate: $E(x(\cdot)|y)$