

Markov chain Monte Carlo

A few details from last time

1. Suppose $k_j \sim \text{Poisson}(\lambda)$ and $\lambda \sim \text{Gamma}(\alpha, \beta)$.

$$\log p(k|\lambda) = -n\lambda + \sum k_j \log \lambda + K$$

$$\log p(\lambda) = -\beta\lambda + (\alpha - 1) \log \lambda + K'$$

$$\log p(\lambda|k) = -(n + \beta)\lambda + (\sum k_j + \alpha - 1) \log \lambda + K''$$

$$\Rightarrow \lambda|k \sim \text{Gamma}(\sum k_j + \alpha, n + \beta)$$

2. Suppose $y \sim \text{Normal}(X\beta, \sigma^2 I)$ and $p(\beta, \sigma^2) \propto 1/\sigma^2$.

$$\log p(\beta, \sigma^2|X, y) = \frac{\|y - X\beta\|^2}{-2\sigma^2} - \left(\frac{n}{2} + 1\right) \log \sigma^2 + K$$

Note that $(y - X\beta) = (y - X\hat{\beta}) + X(\hat{\beta} - \beta)$ and these two bits are orthogonal. ($\hat{\beta} = (X'X)^{-1}X'y$)

$$\text{Thus } \|y - X\beta\|^2 = \|y - X\hat{\beta}\|^2 + \|X(\beta - \hat{\beta})\|^2$$

$$\text{As a result, } \log p(\beta|\sigma^2, X, y) = \frac{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}{-2\sigma^2} + K'$$

$$\Rightarrow \beta|\sigma^2, X, y \sim \text{Normal}(\hat{\beta}, \sigma^2(X'X)^{-1})$$

Details from last time (continued)

3. Posterior of σ^2 :

$$\begin{aligned}\log p(\sigma^2|X, y) &= \log p(\beta, \sigma^2|X, y) - \log p(\beta|\sigma^2, X, y) \\ &= \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{-2\sigma^2} - \left(\frac{n-p}{2} + 1\right) \log \sigma^2 + K \\ &= \frac{(n-p)s^2}{-2\sigma^2} - \left(\frac{n-p}{2} + 1\right) \log \sigma^2 + K\end{aligned}$$

$\Rightarrow \sigma^2 \sim \text{scaled inv-}\chi^2(n-p, s^2)$

T cell example (continued)

I re-ran the MCMC sampler using a $\text{Gamma}(\alpha = 0.01, \beta = 0.01)$ prior on the λ 's rather than the $\text{Gamma}(\alpha = 1, \beta = 1)$ used previously, to assess the influence of the prior on the results.

The following table shows how things are mixing; the rows correspond to the original prior and the new prior.

n	\sqrt{R}							
	λ_0	λ_1	λ_2	λ_3	a	b	σ^2	log lik
100	4.64	8.72	8.02	9.57	3.90	30.05	2.13	1.28
100	10.59	14.14	14.41	16.58	9.17	87.97	3.63	1.63
500	1.64	2.07	2.03	2.19	1.48	3.75	1.13	1.10
500	2.24	3.95	3.61	4.10	1.96	6.43	1.31	1.15
1000	1.20	1.25	1.25	1.27	1.16	1.43	1.05	1.03
1000	1.46	1.77	1.72	1.83	1.38	2.35	1.31	1.16
2500	1.07	1.11	1.11	1.12	1.05	1.24	1.02	1.01
2500	1.10	1.22	1.21	1.23	1.08	1.25	1.05	1.04
5000	1.02	1.03	1.03	1.03	1.02	1.05	1.01	1.00
5000	1.05	1.07	1.07	1.08	1.04	1.10	1.03	1.02
10000	1.01	1.01	1.01	1.01	1.01	1.02	1.00	1.00
10000	1.03	1.02	1.02	1.02	1.04	1.05	1.04	1.01
2nd half	1.01	1.04	1.04	1.06	1.00	1.11	1.01	1.00
2nd half	1.05	1.05	1.05	1.06	1.07	1.16	1.08	1.03

T cell example (continued)

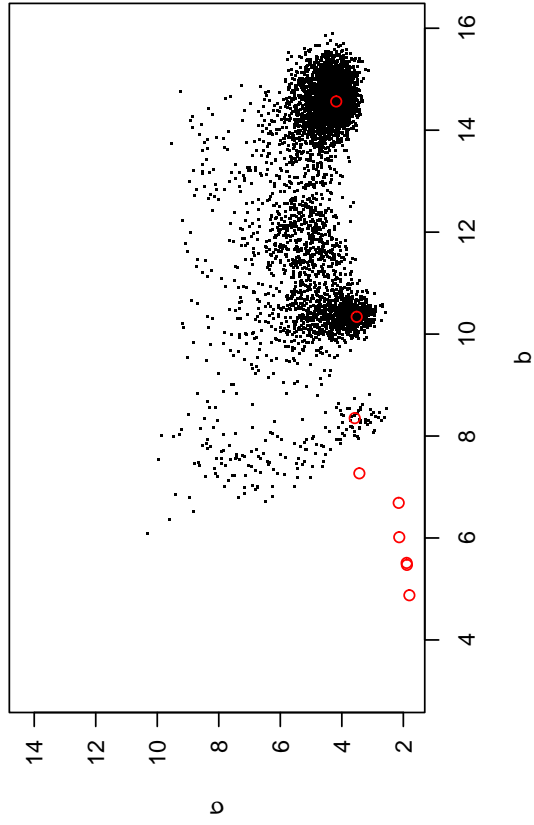
The posteriors are quite different:

	λ_0	λ_1	λ_2	λ_3	a	b	σ
ave	0.21	2.14	1.98	3.11	17.57	14.07	4.51
	0.07	3.52	3.13	5.32	19.72	8.59	7.47
sd	0.11	0.41	0.38	0.55	0.87	1.40	0.73
	0.15	1.02	0.92	1.51	1.80	2.33	1.94

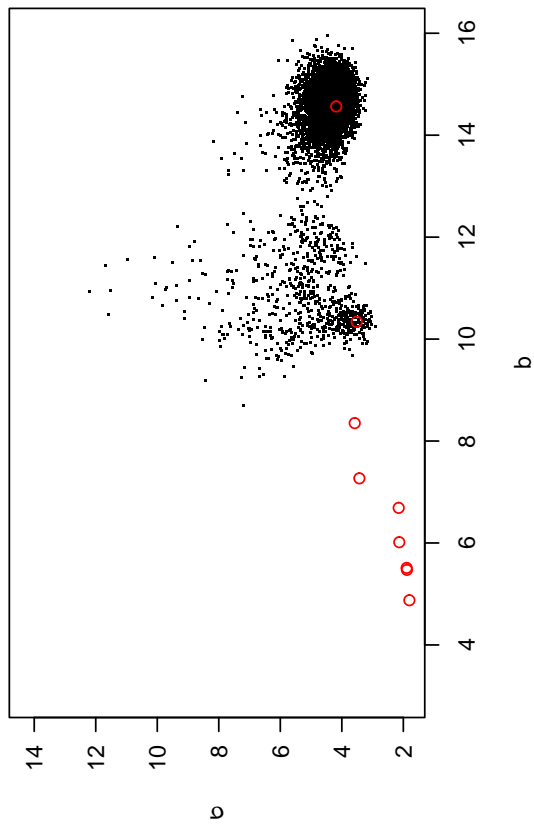
The plots on the following pages show that the change in prior has resulted in *radical* changes in the posteriors. There also seems to be some mixing problems which were not detected by Gelman's \sqrt{R} statistic.

The plot of σ versus λ_0 suggests that $\lambda_0 \approx 0$ is a problem.

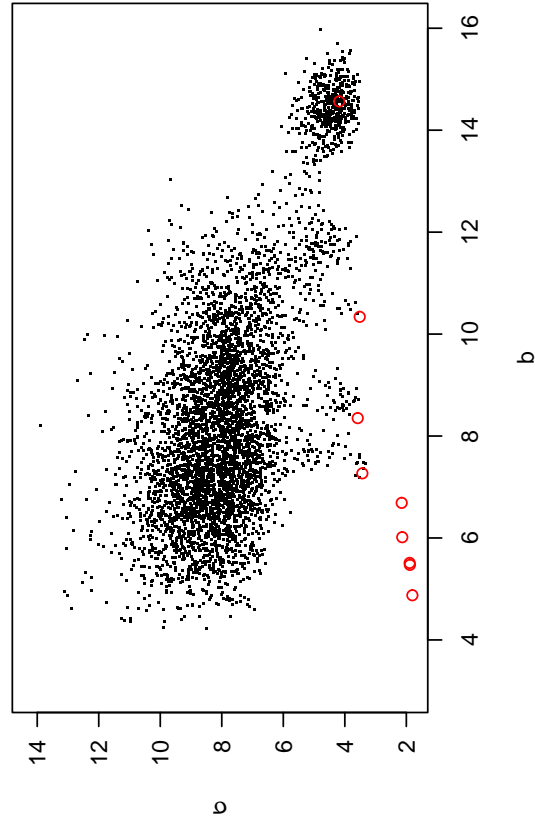
orig prior; chain 1



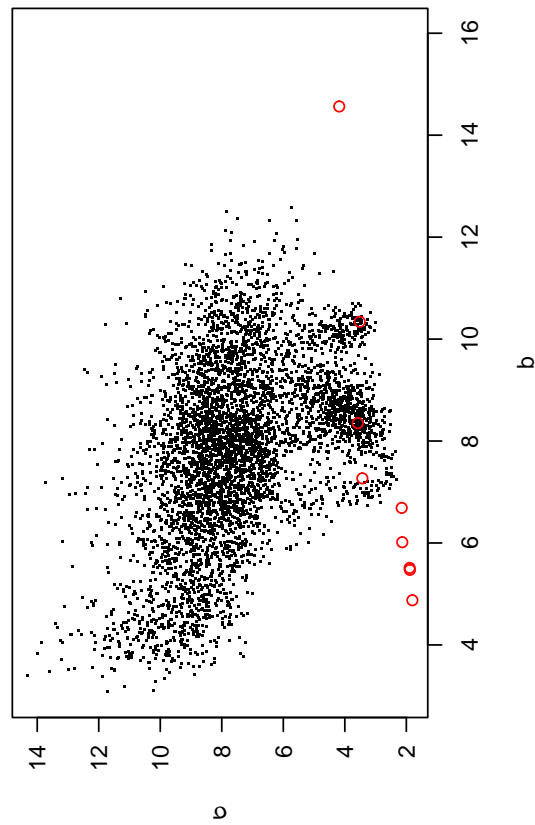
orig prior; chain 9



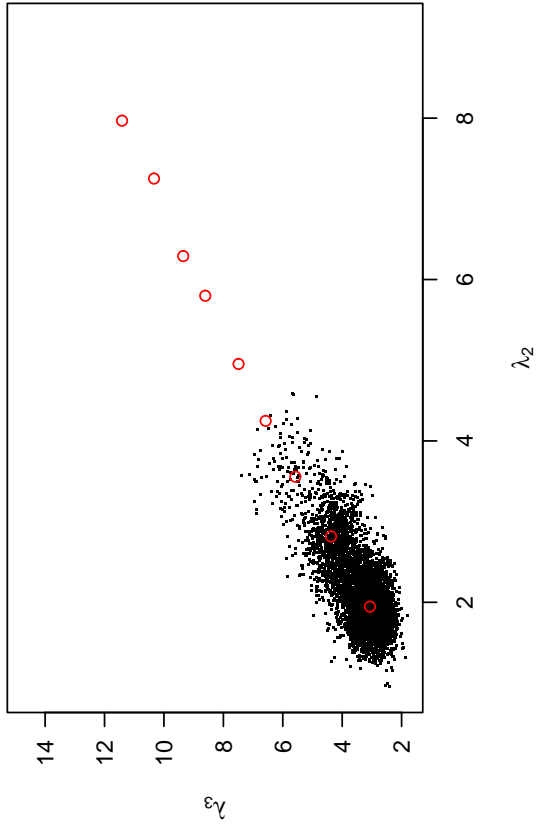
new prior; chain 1



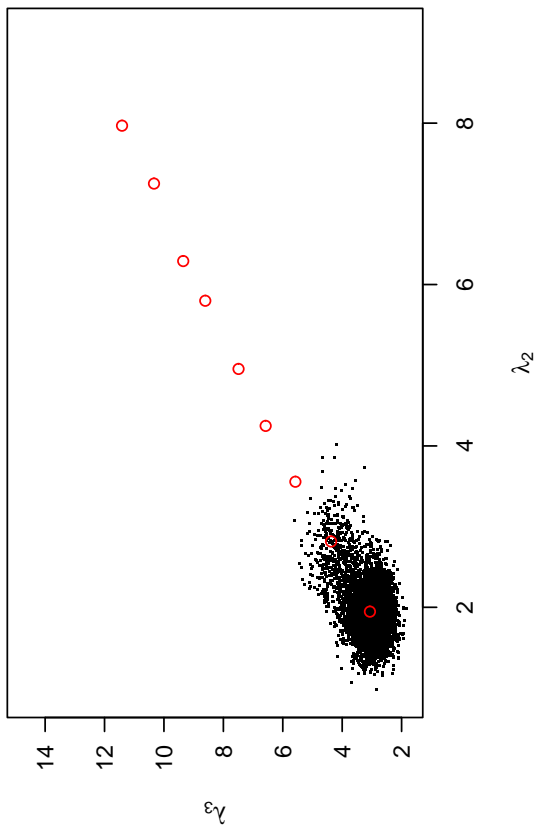
new prior; chain 9



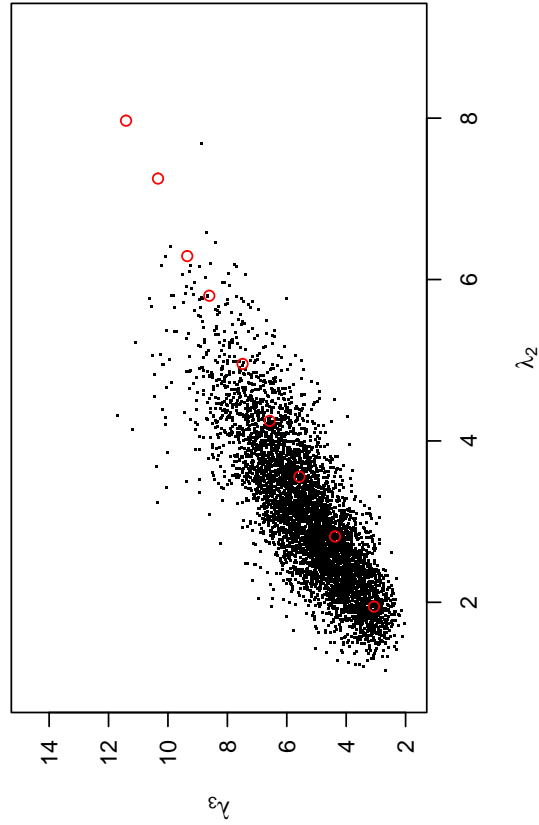
orig prior; chain 1



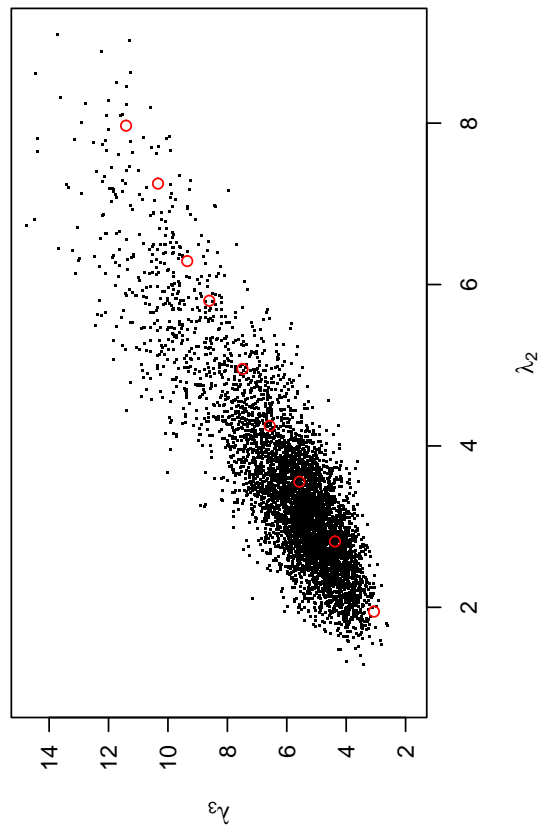
orig prior; chain 9



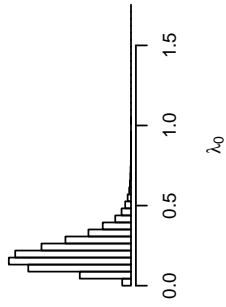
new prior; chain 1



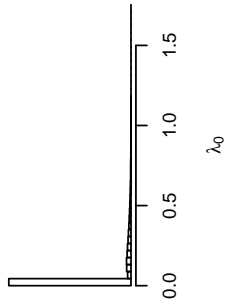
new prior; chain 9



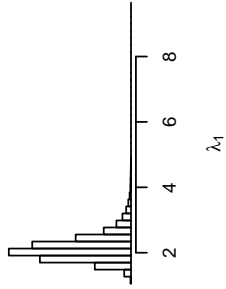
orig prior



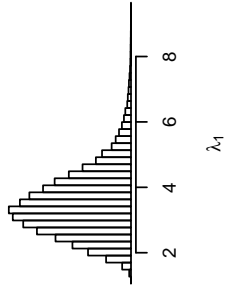
new prior



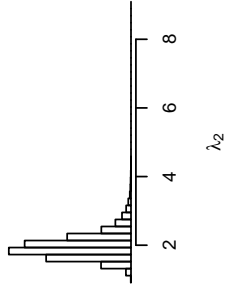
orig prior



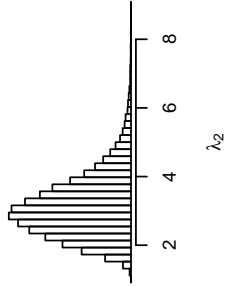
new prior



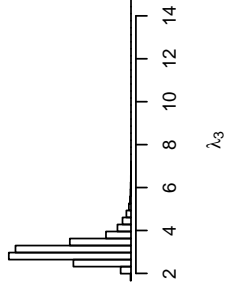
orig prior



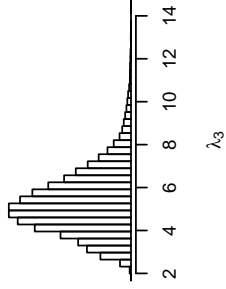
new prior



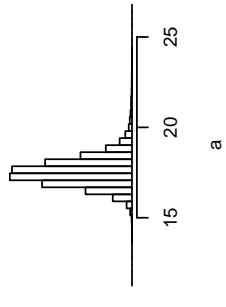
orig prior



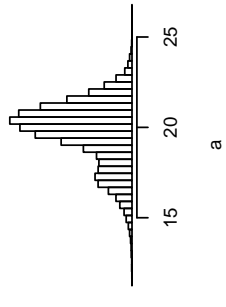
new prior



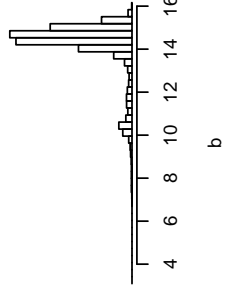
orig prior



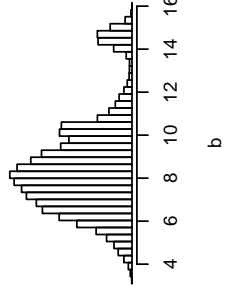
new prior



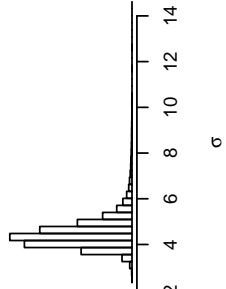
orig prior



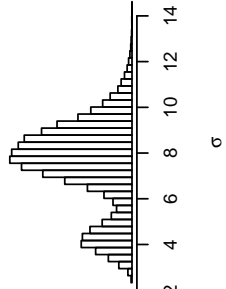
new prior



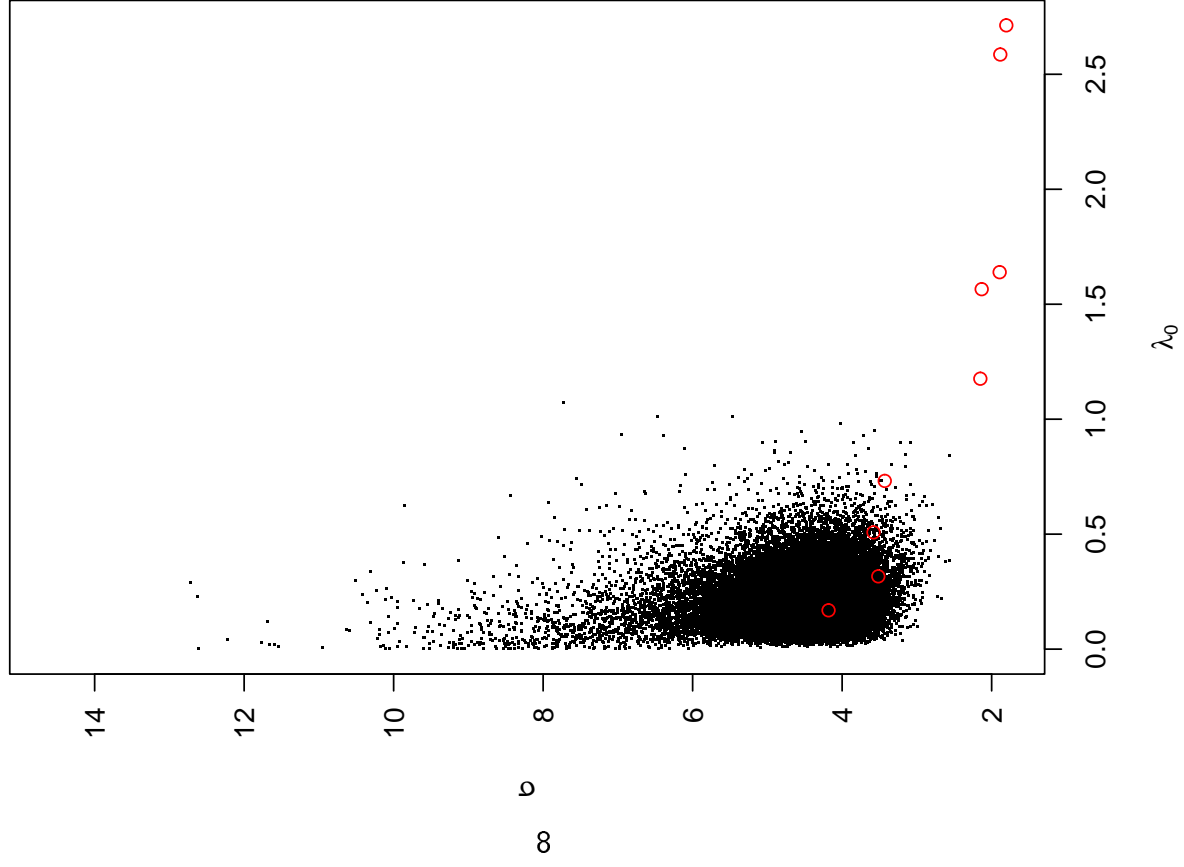
orig prior



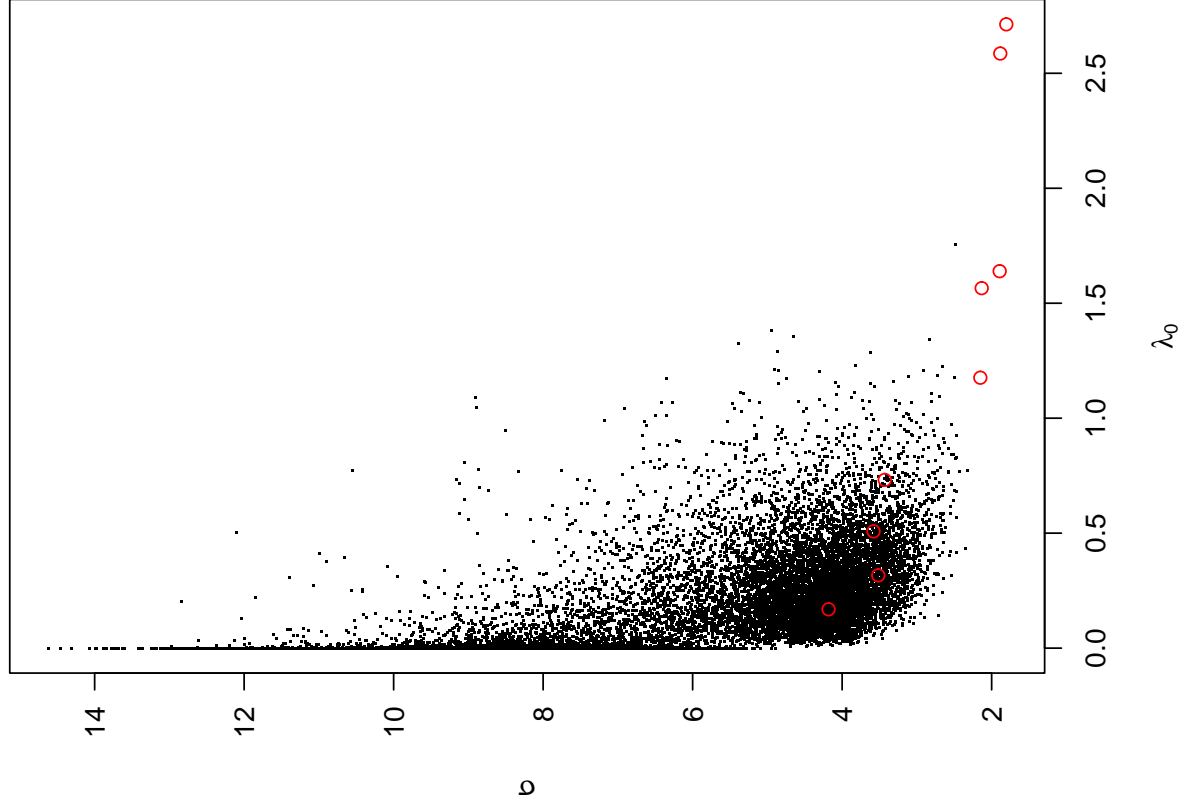
new prior



orig prior



new prior



Simulated annealing

Refs: Lange, §24.7; Numerical Recipes in C, §10.9

We wish to form a stochastic optimizer of $\pi(\theta)$, where θ is high-dimensional.

First, we need a symmetric jump distribution $J(\theta^*|\theta^{(t)})$

Second, we form a “temperature schedule,” $\tau(t)$, starting high and decreasing to 0.

At step t , we “power up” $\pi(\theta)$ by $1/\tau(t)$, and do a Metropolis step:

1. Pick θ^* from $J(\theta^*|\theta^{(t)})$.
2. Pick $U \sim \text{uniform}(0, 1)$
3. Step to θ^* if $U < \exp\{[\log \pi(\theta^*) - \log \pi(\theta^{(t)})]/\tau(t)\}$

If $\pi(\theta^*) \geq \pi(\theta^{(t)})$, we always step to θ^* .

If $\pi(\theta^*) < \pi(\theta^{(t)})$, we step to θ^* with probability related to the ratio of the function values.

At early, “high” temperatures, we are much more liberal in stepping down in π as well as up. As the temperature “cools,” we tend only to step upward in π .

Reversible jump MCMC

Green, *Biometrika* 82:711–32, 1995.

We consider the problem of *Bayesian model selection*.

Consider models $\{\mathcal{M}_k, k \in \mathcal{K}\}$ where \mathcal{K} is countable. Suppose that model \mathcal{M}_k has parameter vector $\theta^{(k)} \in \mathbb{R}^{n_k}$.

We decompose the joint probability of the model index, parameter vector and prior into the model probability, prior and likelihood:

$$p(k, \theta^{(k)}, y) = p(k) p(\theta^{(k)}|k) p(y|k, \theta^{(k)})$$

We want to use MCMC to sample from the posterior

$$p(k, \theta^{(k)}|y) = p(k|y) p(\theta^{(k)}|k, y)$$

Jumping through $(k, \theta^{(k)})$ -space can be tricky, since when k changes, the size of $\theta^{(k)}$ changes. Green (1995) describes how to do it. Jumping between $(k, \theta^{(k)})$ and $(k+1, \theta^{(k+1)})$ requires care to keep the chain reversible; the Jacobian crops up.

Rather than work through Green's paper (which is not easy), I'll sketch one of his examples.

A 1D multiple change-point problem

Suppose we observe y_1, \dots, y_n in $(0, L)$ from an inhomogeneous Poisson process with rate $x(t)$, where $x(t)$ is a step function.

Let $x(t)$ have k change-points at positions $0 \equiv s_0 < s_1 < s_2 < \dots < s_k < s_{k+1} \equiv L$. Let h_j be the height of $x(t)$ in the interval (s_j, s_{j+1}) . Thus $\theta^{(k)}$ has dimension $2k + 1$ (the s_j and the h_j).

Prior:

$k \sim \text{Poisson}(\lambda)$ with $k \leq k_{\max}$.

Given k , s_1, \dots, s_k are distributed according to the even order statistics of $2k + 1$ uniformly dist'd points on $(0, L)$.

Given k , $h_0, \dots, h_k \sim \text{iid Gamma}(\alpha, \beta)$.

Log likelihood:

For a given $x(t)$, the log likelihood is

$$l(k, s_j, h_j | y) = \sum_j h_j \#\{i : s_j < y_i < s_{j+1}\} \\ - \sum_j h_j (s_{j+1} - s_j)$$

The Markov chain

Markov chain: four types of moves

1. Change the height of a random step
2. Change the position of a random step
3. Add a new step (“birth”)
4. Remove a step (“death”)

At each step in the chain, we will choose one of these four possible moves at random, depending on the current value of k (the number of change-points).

Let η_k , π_k , b_k and d_k be the probabilities of the four moves, given k .

Of course, we must have $\eta_k + \pi_k + b_k + d_k = 1$.

Also, $\pi_0 = d_0 = 0$ and $b_{k_{\max}} = 0$.

In order to ensure that $b_k p(k) = d_{k+1} p(k+1)$, we take $b_k = c \min\{1, p(k+1)/p(k)\}$ and $d_{k+1} = c \min\{1, p(k)/p(k+1)\}$, where c is chosen so that $b_k + d_k \leq 0.9$ for all k .

Further, for $k \neq 0$, we take $\eta_k = \pi_k$.

Height and position moves

To change the height of a step:

1. Choose one of the h_0, \dots, h_k at random (say h_j).
2. Propose a change from h_j to h'_j such that $\log(h'_j/h_j) \sim \text{uniform}(-\frac{1}{2}, \frac{1}{2})$.
3. Accept the change with probability

$$\min\{1, \text{LR} \times (h'_j/h_j)^\alpha \exp[-\beta(h'_j - h_j)]\}$$

where LR is the likelihood ratio.

To change the position of a step:

1. Choose one of the h_1, \dots, s_k at random (say s_j).
2. Propose $s'_j \sim \text{uniform}(s_{j-1}, s_{j+1})$.
3. Accept with probability

$$\min\{1, \text{LR} \times \frac{(s_{j+1} - s'_j)(s'_j - s_{j-1})}{(s_{j+1} - s_j)(s_j - s_{j-1})}\}$$

Birth and death steps

The birth step:

1. Choose a position s^* uniformly on $(0, L)$ (say between s_j and s_{j+1}).
2. Propose new heights h'_j and h'_{j+1} so that $h'_{j+1}/h'_j = (1 - U)/U$ with $U \sim \text{uniform}(0, 1)$ and satisfying
$$(s^* - s_j) \log h'_j + (s_{j+1} - s^*) \log h'_{j+1} = (s_{j+1} - s_j) \log h_j$$
3. We'll defer the acceptance probability for now.
4. If the move is accepted, we move s_{j+1}, \dots, s_k over to s_{j+2}, \dots, s_{k+1} and set $s_{j+1} = s^*$. (Similarly for the h_j 's.)

Death step:

1. Pick one of s_1, \dots, s_k at random (say s_{j+1}).
2. Pick the new height, h'_j , to satisfy
$$(s_{j+1} - s_j) \log h_j + (s_{j+2} - s_{j+1}) \log h_{j+1} = (s_{j+2} - s_j) \log h'_j$$
3. The acceptance probability is again deferred.

Birth/death acceptance probabilities

In a birth step, we're going from dimension $2k+1$ to $2k+3$.

The acceptance probability for the birth step is of the following form:

$$\min\{1, \text{LR} \times \text{prior ratio} \times \text{proposal ratio} \times \text{Jacobian}\}$$

$$\text{Prior ratio} = \frac{p(k+1)}{p(k)} \times \frac{2(k+1)(2k+3)}{L^2} \times \frac{(s_{j+1}-s^*)(s^*-s_j)}{s_{j+1}-s_j} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{h'_j h'_{j+1}}{h_j}\right)^{\alpha-1} \exp[-\beta(h'_j + h'_{j+1} - h_j)]$$

$$\text{Proposal ratio} = \frac{d_{k+1}L}{b_k(k+1)}$$

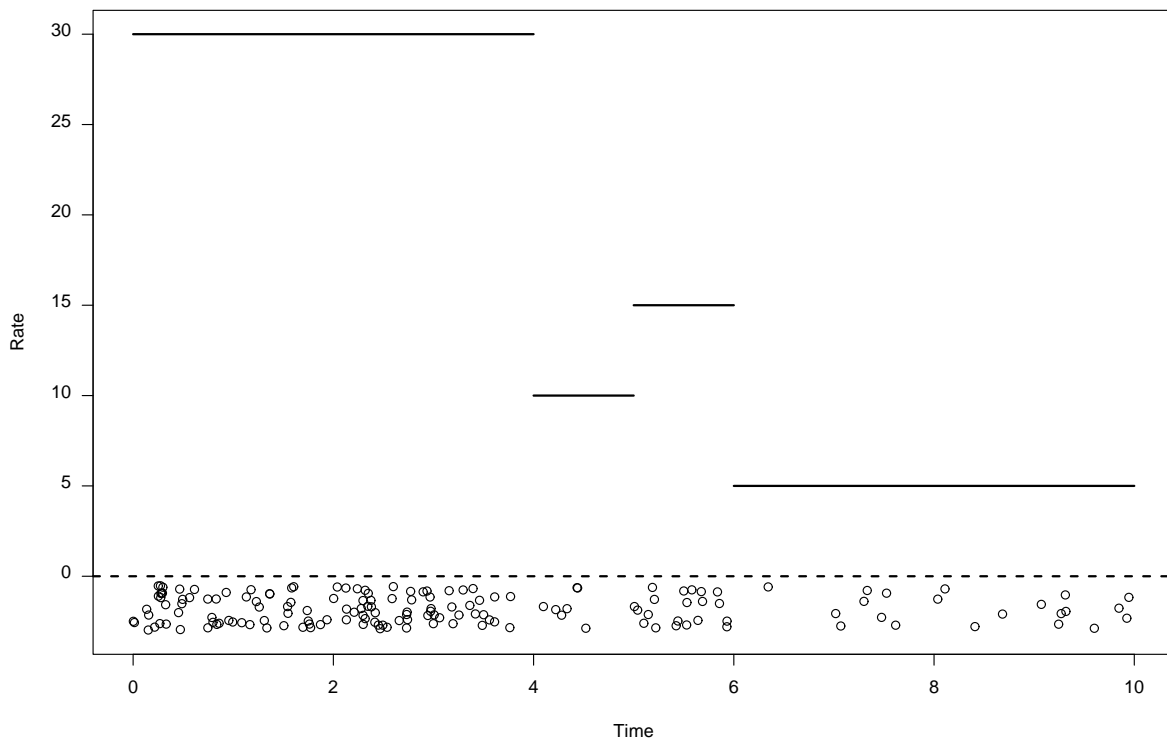
$$\text{Jacobian} = \frac{(h'_j + h'_{j+1})^2}{h_j}$$

For the death step, use the reciprocal of the above, with things appropriately re-labelled.

I'll leave it to you to read Green (1995) and figure out where this comes from.

Example data

I consider (simulated) data on the interval $(0, 10)$. I observed 162 points. The following figure shows the data and the true underlying rate function, $x(t)$.



In the prior, I'll take $\lambda = 1$, $\alpha = 0.2$, $\beta = 0.01$ and $k_{\max} = 30$. (Note that this choice of α and β result in $E(h_j) = 20$ and $SD(h_j) \approx 44$.)

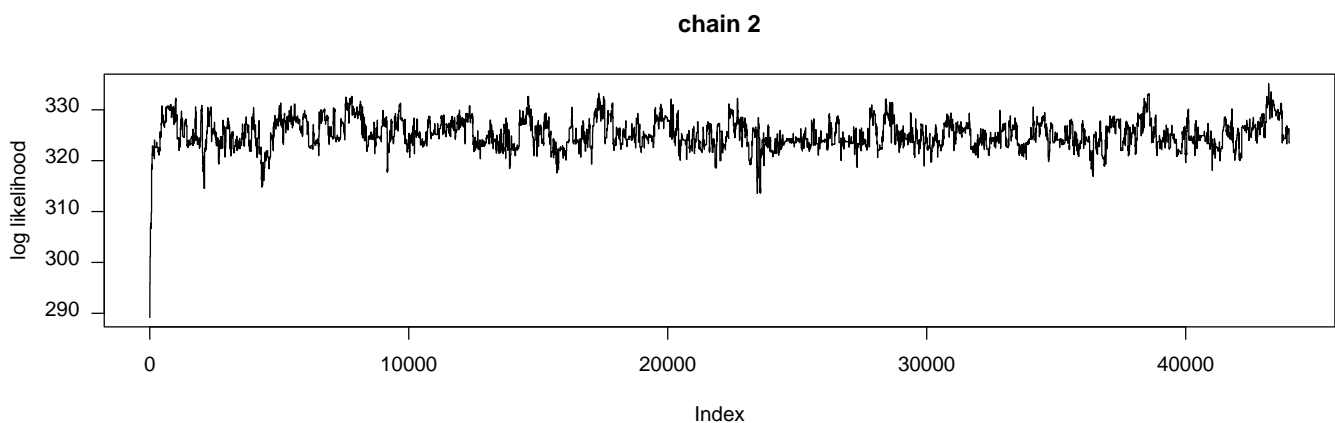
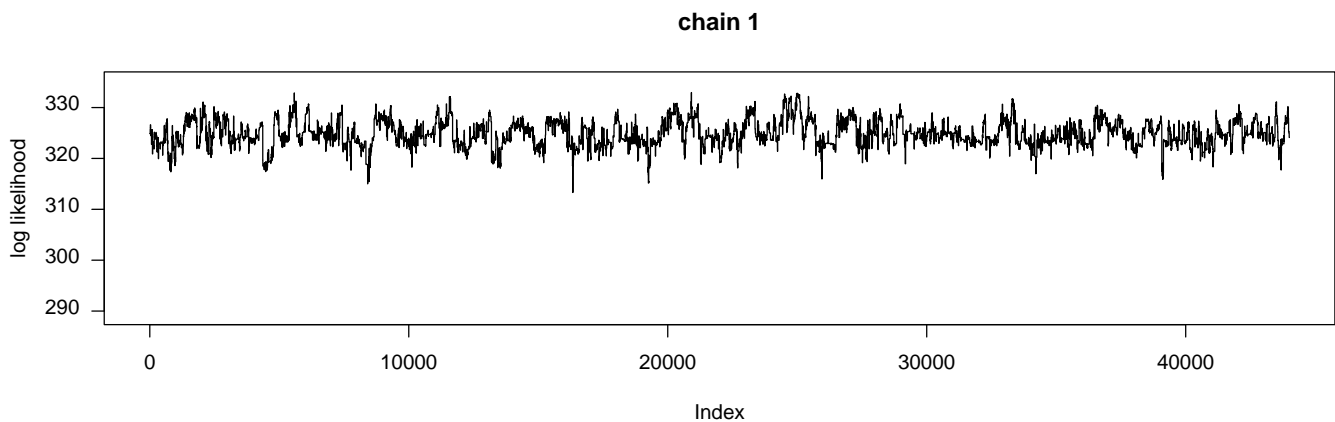
Note that Green (1995) considered some real data, and used 40,000 steps of a Markov chain with a 4000 step burn-in. I'll run two chains, starting at the truth and at a model with $k = 0$, though with a 24,000 step burn-in.

Results

The following are the estimated posteriors for k (the number of change-points); these are percentages.

	1	2	3	4	5
chain 1	47.3	40.3	10.9	1.4	0.0
chain 2	41.2	40.6	15.2	2.7	0.2

Here are some plots tracking the log likelihood (*not* the posterior...I was lazy).

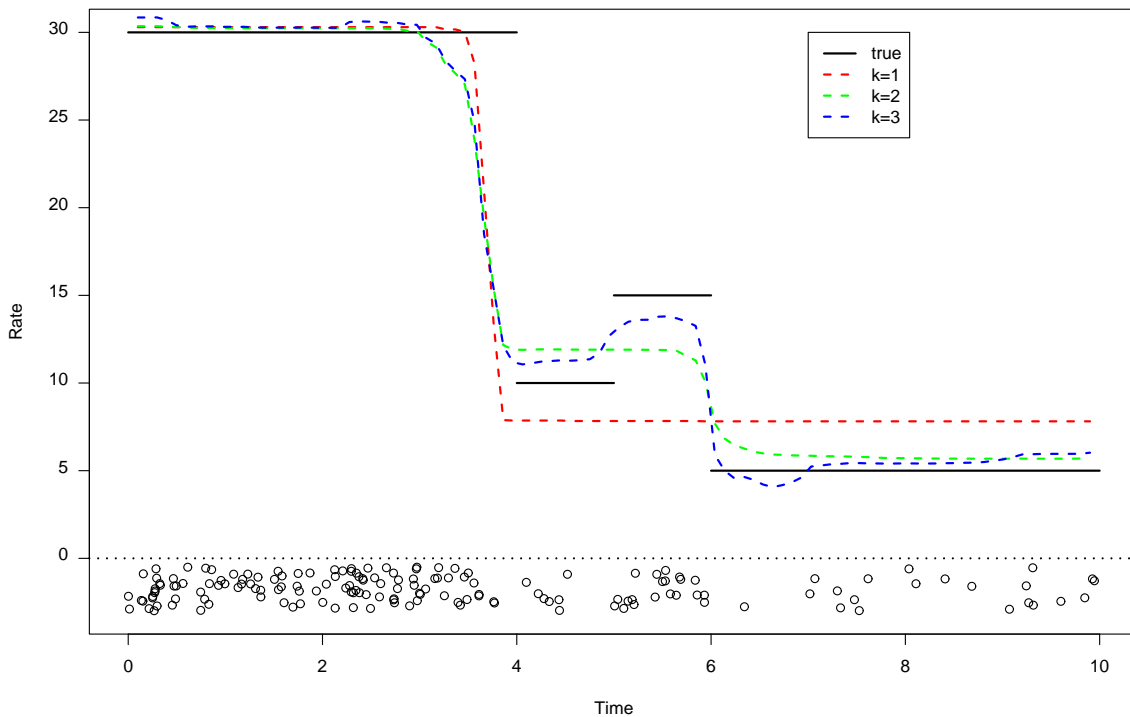


Results (continued)

It's interesting to look at the acceptance probabilities in the two chains, as functions of the step types.

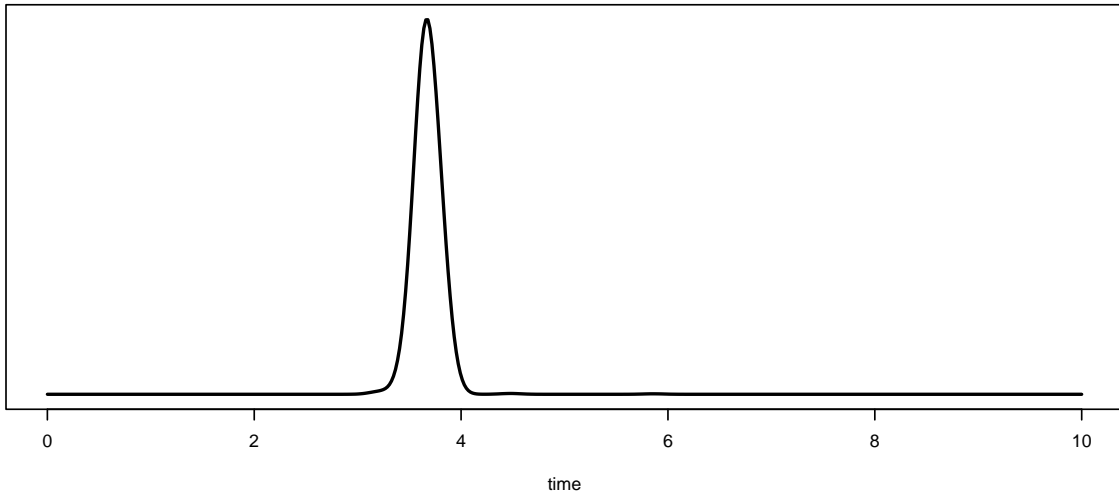
step type	% of steps	% accepted
height	14	47
position	14	10
birth	20	15
death	51	6

Finally, some fitted curves.

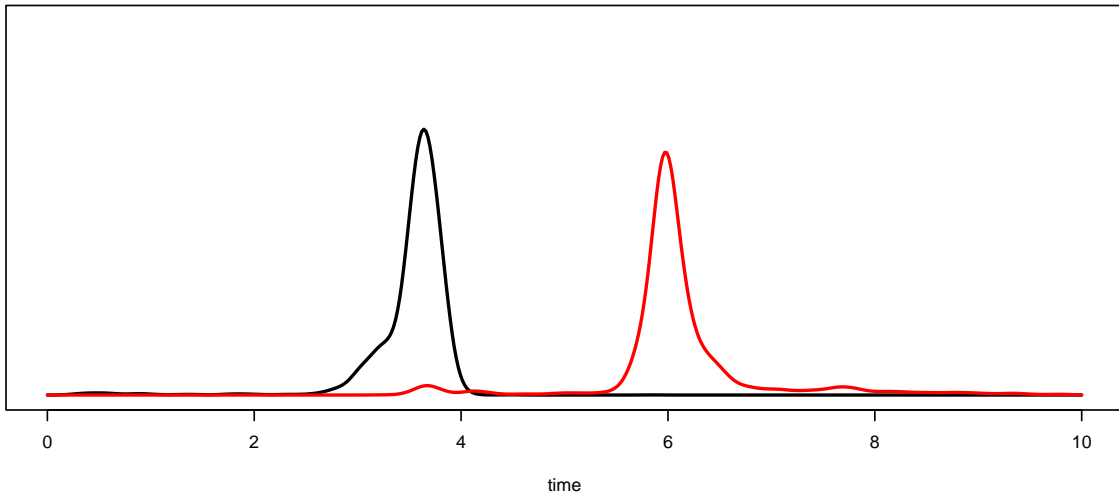


The next page shows estimated posterior densities for the locations of the change-points.

k = 1



k = 2



k = 3

