



MCMC II

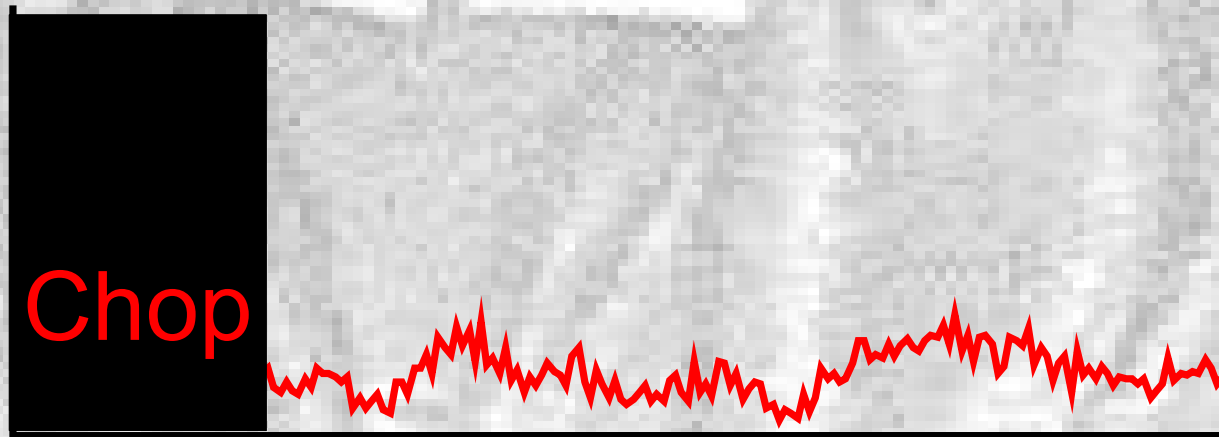
Practicalities

Practical Questions with MCMC

- If we take a lot of samples from the MCMC sampler, they should have the correct distribution
- How do we decide if we have reached the distribution?
- How well do the samples approximate the target distribution?
- How many values do we need?

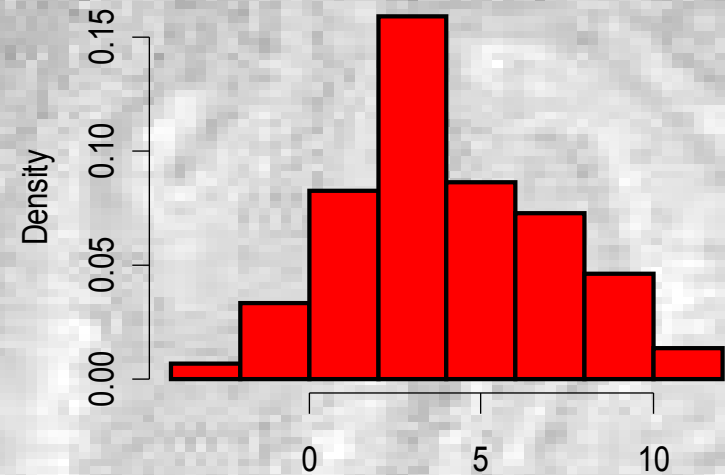
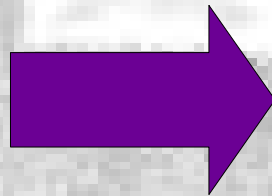
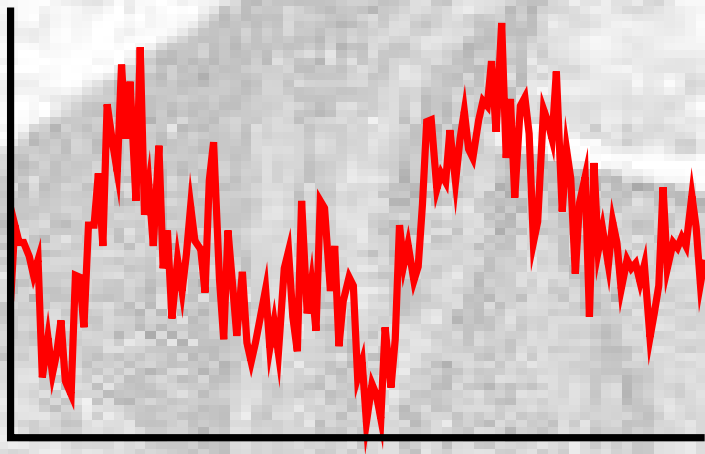
The burn-in

- Once the chain has got to the stationary state, all values are from the required distribution
- We remove the first few values, the burn-in



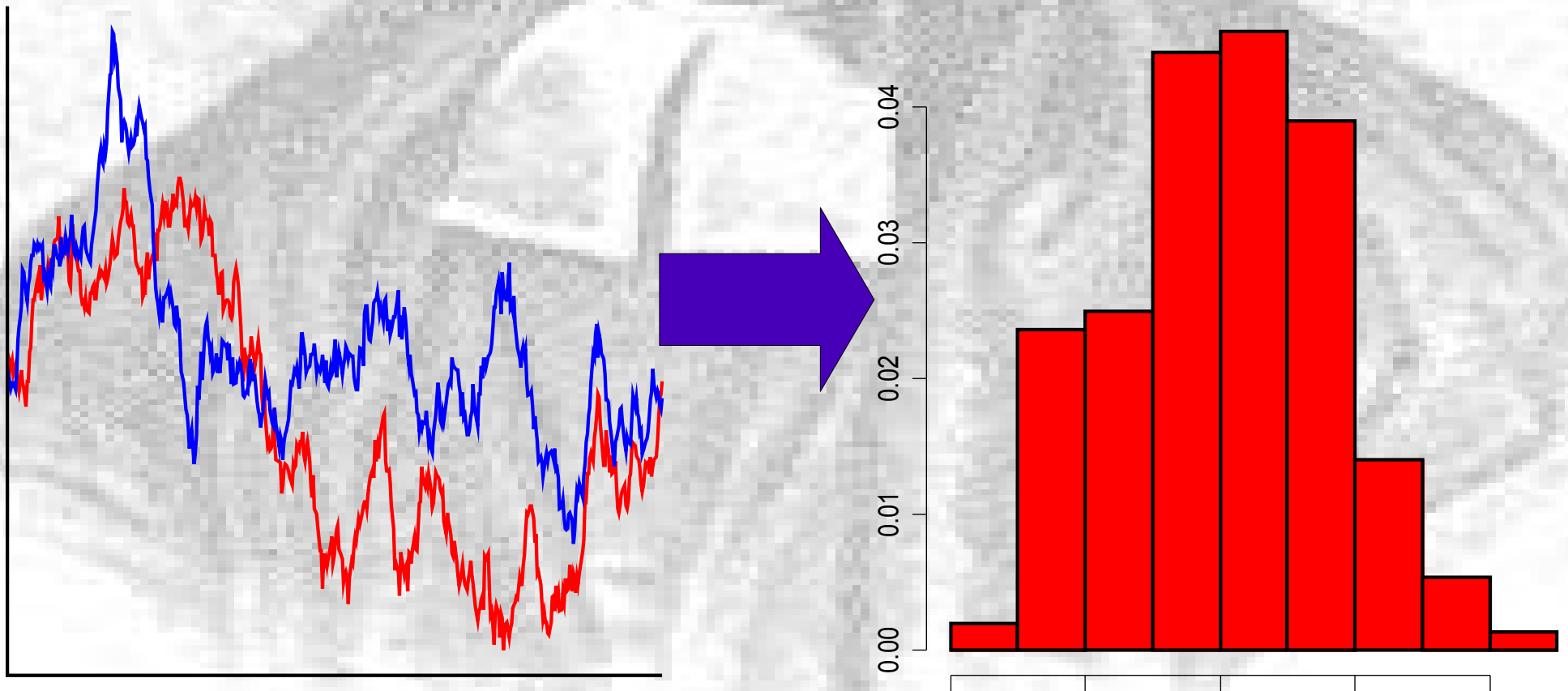
The Rest

- . After the burn-in, the values are from the target distribution



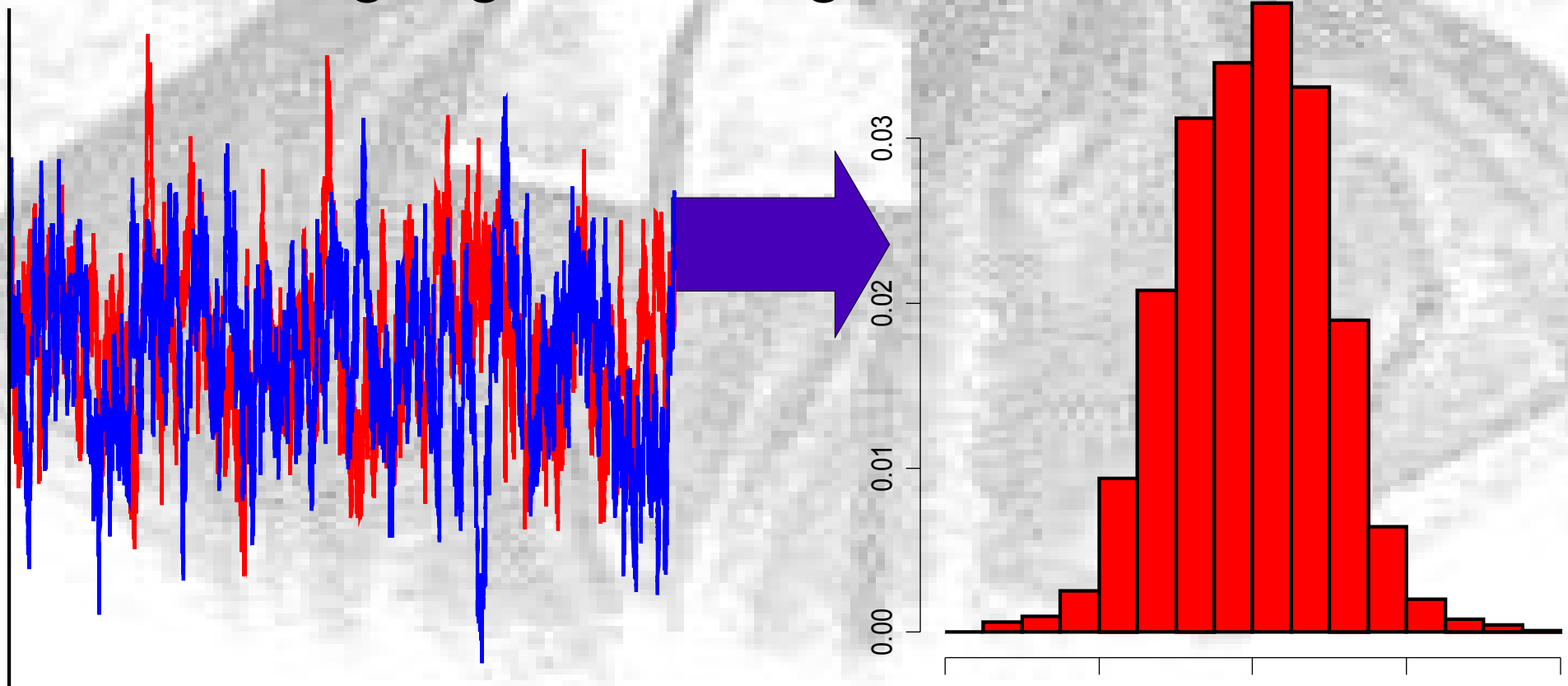
But...

- Even after we have reached convergence, the chain might not be good:



A Solution

- Run the chain for a longer period
 - e.g. 50 times as long
- Want to get good mixing



Long Chains

- If we run an chain for a long time, then we get lots of numbers
 - takes up a lot of memory
- We often only take some iterations
 - e.g. take every 10
 - Called **thinning**
- Taking iterations at an even interval reduces the autocorrelation between iterations

How many points?

- To get a good coverage of the target distribution, we need a lot of points
 - I typically use 10 000
- The precise number will depend on what you want and what resources you have
 - 1000 iterations might be OK for some problems
- How many will depend on the quality of the chain, and the distribution
 - thick tails will need more iterations

How many points II

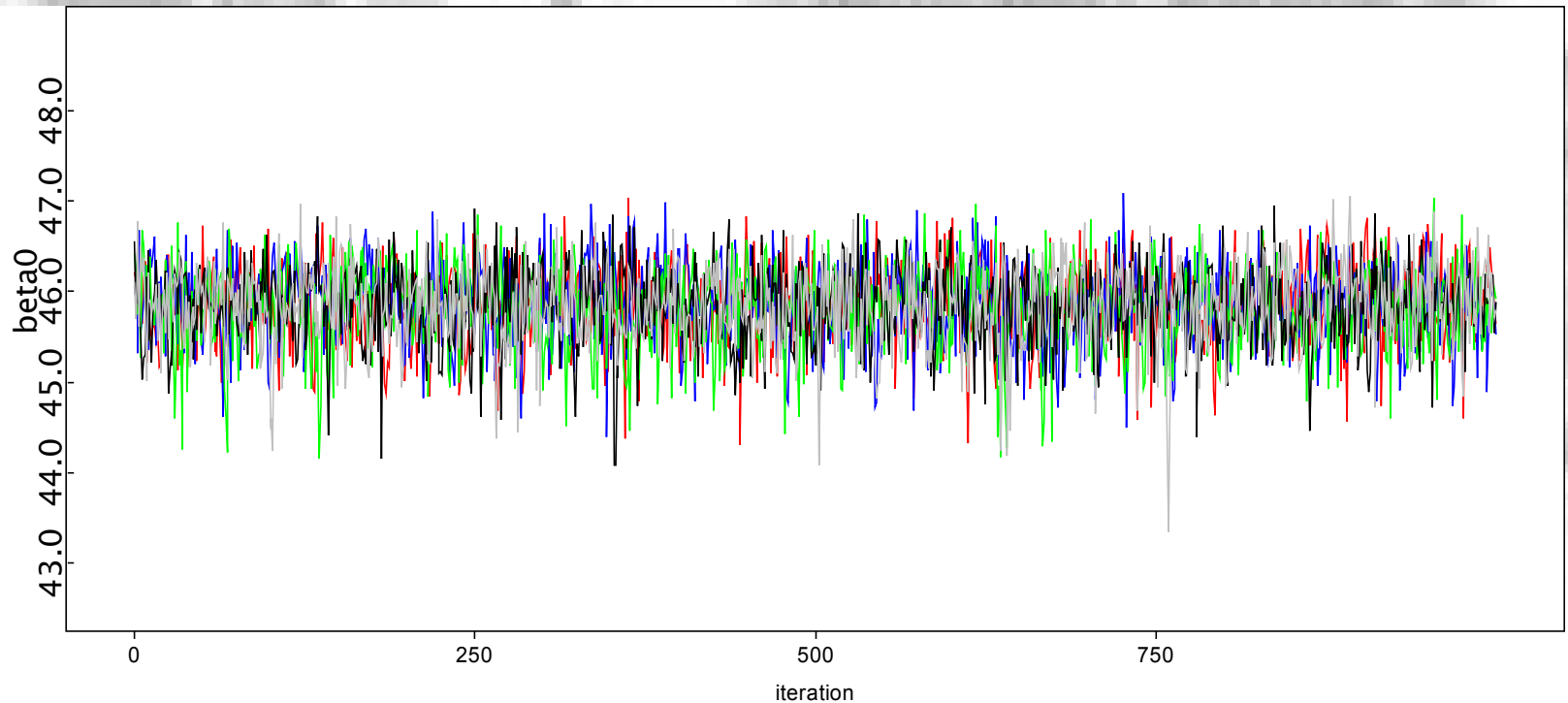
- We can assess how well the estimation is doing by comparing the mean of the samples, and the true posterior mean
 - have to use estimates!
 - Called the MC error
- Rule of thumb: we are OK if this is less than 5% of the true error
 - i.e. if the error in the estimation swamped by the uncertainty in the true value

Checking the Chain

- We need to check if we have converged
- There are formal methods, mostly based on running several chains
- If we have not converged, then we may have to remove our points, as a burn-in
- Or, if we have converged, but the mixing is not good, then we have to either thin the chain, or improve the sampling

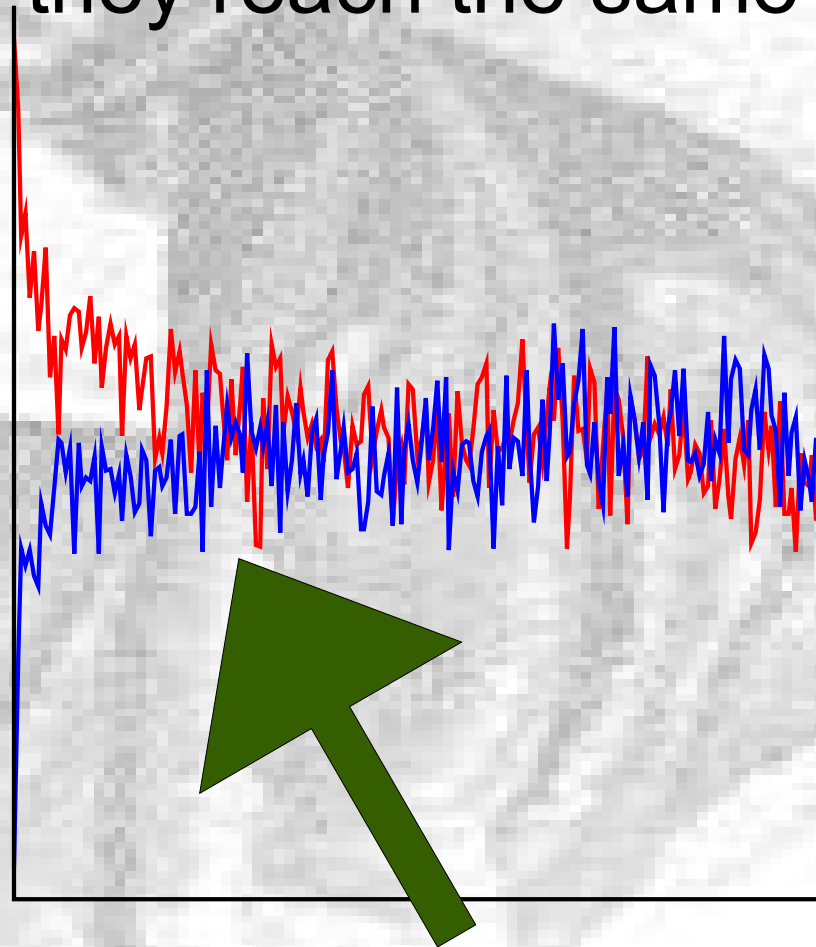
Visuals

- The simplest way to check chains is to look at a graph
 - WinBugs history



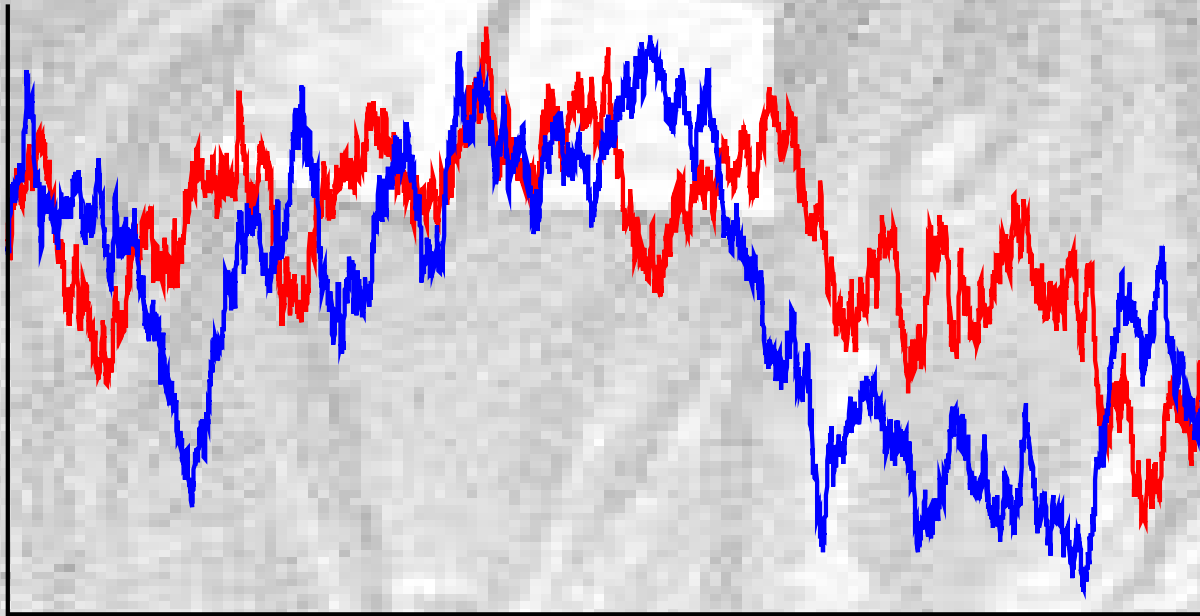
When shall we 2 meet again?

- . By plotting the values against iteration number, we can see if they reach the same distribution



Thinning

- We can also see if the mixing is good
- Here it is not



Formal Tests

- We can also use more formal tests of convergence
 - several in the Coda package for R
- One test in WinBugs
 - Gelman-Rubin-Brooks
- Multivariate example of the Gelman-Rubin statistic

Gelman-Rubin

- Based on an ANOVA idea
- Look at a single variable
- m chains, each of length n
- Can estimate the variance of a stationary distribution in two ways
 - variance within a single chain, W
 - variance over all chains, B/n

Gelman-Rubin II

- If the chains have converged, both estimates are unbiased, i.e. $B=W$
- If the initial values are overdispersed and have not converged, then the Between term is an overestimate
 - Use the statistic: $R = B/W$
- If $R > 1$, we have not converged
 - we estimate R by

$$\hat{R} = \frac{m+1}{m} \left(\frac{n-1}{n} + \frac{B}{W} \right) - \frac{n-1}{mn}$$

Gelman-Rubin-Brooks

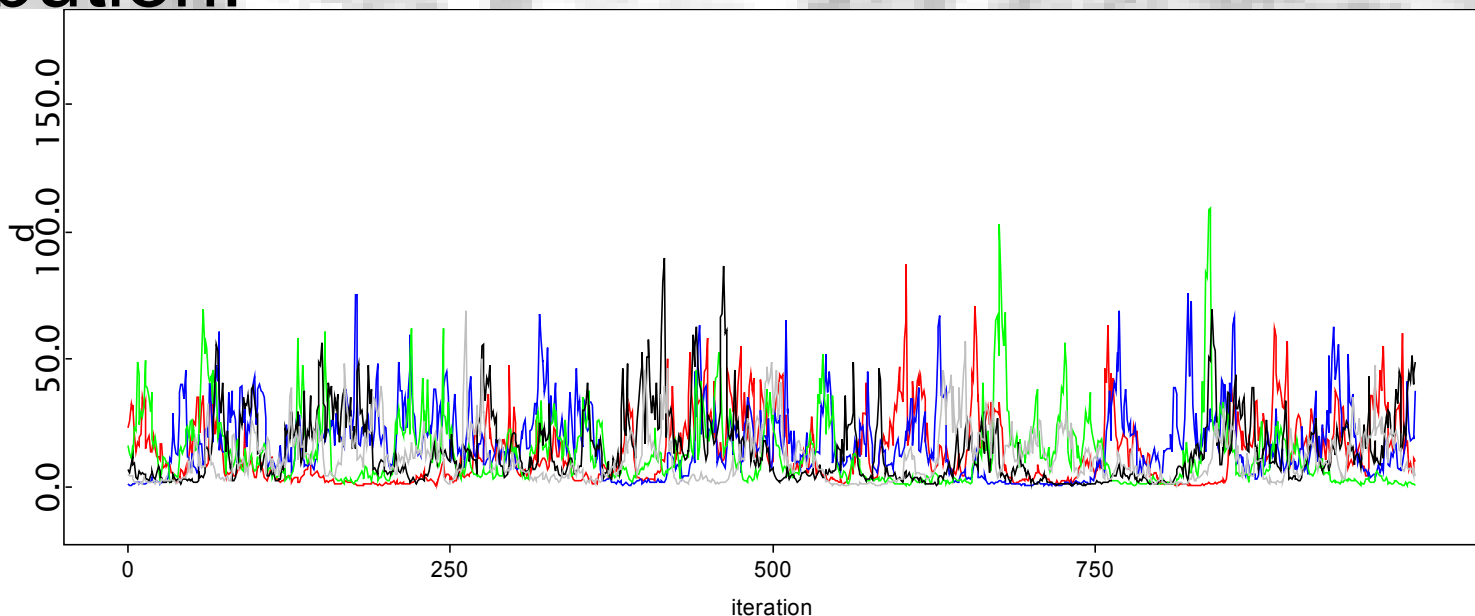
- The ratio B/W may be about 1 by chance
 - if both B and W are still moving
 - for example, if the initial values are not over-dispersed, and a chain starts to explore a new part of the parameter space
- B and W should therefore be checked as well
- WinBugs uses a graphical approach

Gelman-Rubin-Brooks: The Graph

- WinBugs plots a bgr diag
- Three components
 - Gelman-Rubin statistic
 - B variance
 - width of the central 80% interval of the pooled runs
 - W variance
 - average width of the 80% intervals within individual runs
- W and B scaled to a maximum of 1

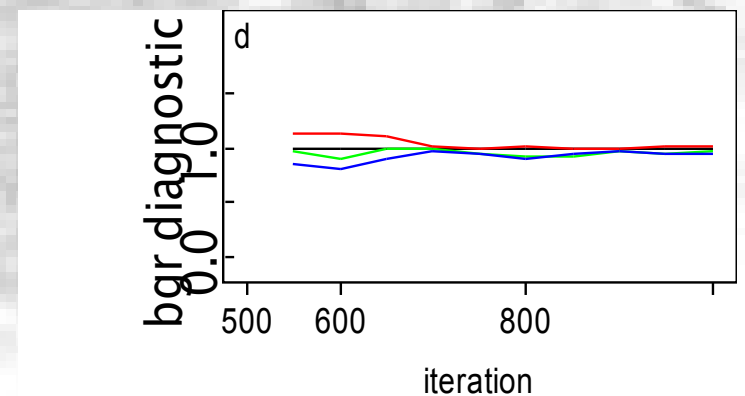
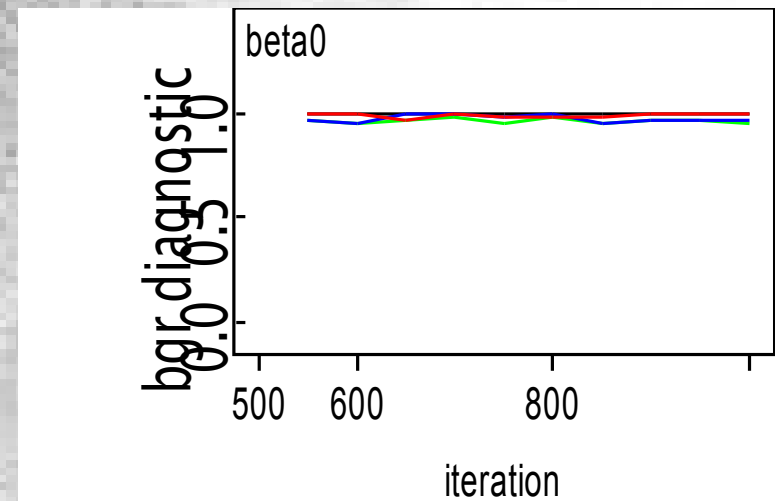
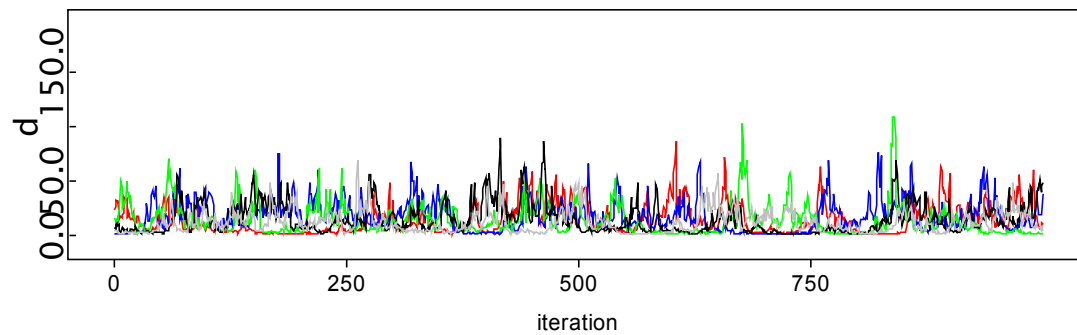
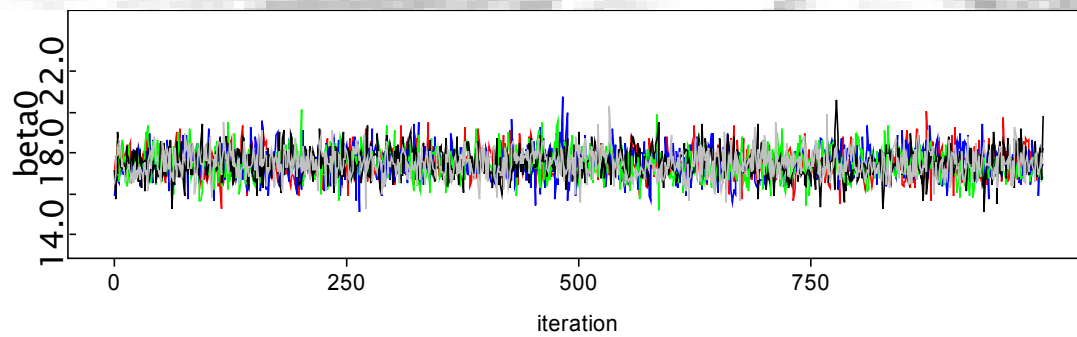
What Should Happen

- We have convergence if:
 - Gelman-Rubin statistic is about 1
 - B and W both stabilise around the same value
- For example: the degrees of freedom in a t distribution:



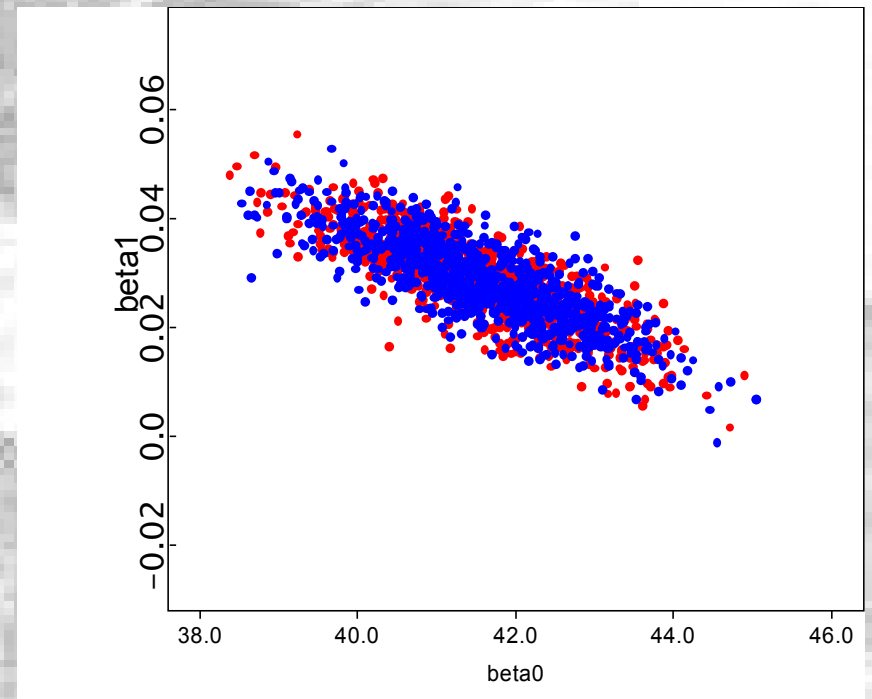
Mixing With Freedom

- Mean and degrees of freedom of a t



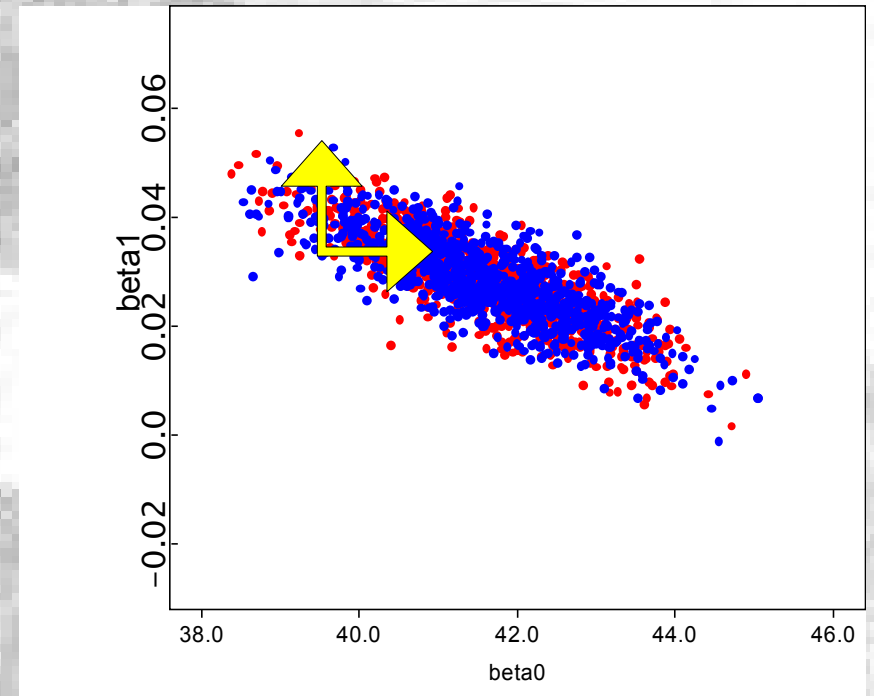
Re-Parameterisation

- In several dimensions we can have several
- parameters being correlated
- e.g. regression: place the intercept at 0
- Large correlation
- $\rho = -0.82$



Problems with Correlations

- Makes sampling difficult
 - Difficult to make large jumps
- Solution: reparameterisation
 - write model in a different form



Problems with Correlations

- In this case, we move the intercept to the mean
- i.e. remove the mean of the covariate from every covariate
- Removes the correlation

$$- \rho = 0.04$$

