

[Th1b]

## MCMC sampling - part II

### Things that can go wrong (in MCMC sampling)

- No convergence
- Pseudoconvergence

Not really wrong, but ineffective  
(Approximate)

- Convergence, but correlated samples

The solution is to run the chain longer

There is no unique solution to such problems, but they are typically addressed in three ways:

1. Design the simulation runs for monitoring of convergence, in particular running multiple sequences with starting points dispersed through parameter space. (Hard in multidim.)

2. Monitor the convergence of all quantities of interest by comparing variation between and within the sequences.

[This] 2

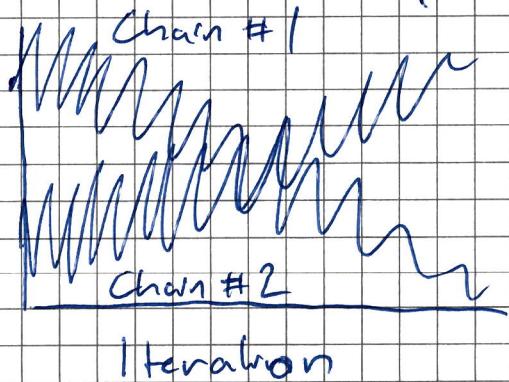
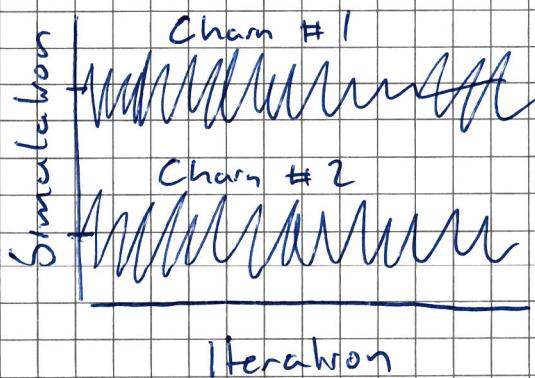
Here you are looking  
for

- mixing (they sample the same distribution)
- stationarity

3. If the simulation efficiency is unacceptably low, the algorithm might have to be altered.

### Monitoring - scalar - estimands

- All parameters of interest
  - The value of some predicted observable
- $$\langle O(\theta) \rangle = \frac{1}{N} \sum_{i=1}^N O(\theta_i)$$
- The log-posterior
- (Fig 11.3 from Gelman) a) No mixing  
b) Not stationary (but mixing)



## Assessing mixing

Take a number of sequences  
 (after discarding warmup) and  
 split in two

$m$  = # chains after splitting

$n$  = length of each chain

Gelman - Rubin (1992)

Label the simulations  $\gamma_{ij}$ , where  $i=1, \dots, n$   
 $j=1, \dots, m$

Within sequence variation:

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2, \text{ where } s_j^2 = \frac{1}{n-1} \left( \sum_{i=1}^n \gamma_{ij} - \bar{\gamma}_{\cdot j} \right)^2$$

$$\bar{\gamma}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \gamma_{ij} \quad (\text{mean of chain } j)$$

Between sequence variation:

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\gamma}_{\cdot j} - \bar{\gamma}_{\cdot \cdot})^2, \text{ where } \bar{\gamma}_{\cdot \cdot} = \frac{1}{m} \sum_{j=1}^m \bar{\gamma}_{\cdot j}$$

(mean of all chains)

The marginal posterior variance

$$\widehat{\text{var}}(\gamma | D, I) = \frac{n-1}{n} W + \frac{1}{n} B$$

and finally  $\widehat{R} = \frac{\widehat{\text{var}}(\gamma | D, I)}{W} \stackrel{\text{overestimates}}{\geq} 1 \stackrel{\text{underestimates}}{\leq} 1$

## False-data check

### Debugging technique

Choose some "true values" of your parameters and simulate data with your model.

Then run your MCMC algorithm and check that it converges.

Over many simulations:

- 50% should contain the true values within the 50% posterior intervals
- 95% should ... 95% interval

## Tempered sampling

The most popular methods to achieve mixing for multimodal distributions are:

- Nested Sampling  
(see Orsi's lecture MCMC 3)  
e.g. with multinest
- Simulated tempering  
(similar to simulated annealing in opt. problems)
- Parallel tempering

| Note: these algorithms are built on top of your sampler.

### Parallel tempering

$$\text{Let } p(\theta | D, I) = C p(\theta | I) p(D | \theta, I)$$

be the target posterior dist.

(where  $C$  is a normalization).

Let us introduce a temperature parameter  $T$  via  $\beta \equiv 1/T$

and construct distributions

$$\pi(\theta|D, \beta, I) = p(\theta|I) p(D|\theta, I)^\beta$$

or

$$\log(\pi) = \log[p(\theta|I)] + \beta \log[p(D|\theta, I)]$$

for  $0 < \beta < 1$

Note that  $\pi(\theta|D, \beta=1, I)$   
is the desired distribution.

How does  $\pi$  look like for  
increasing  $\beta$  (decreasing  $\beta$ )?

$\Rightarrow$  Effectively sample larger  
regions of space that are  
sequentially connected to  
the  $\beta=1$  one.

Use  $\beta \in \{1, \beta_2, \dots, \beta_m\}$

[Th 1b] 7

- Sample from these different distributions in parallel
- At intervals, pick a pair of adjacent chains at random  $\beta_i$  and  $\beta_{i+1}$
- Suggest a swap of their current positions at this time  $t$

$\theta_{t,i}$  and  $\theta_{t,i+1}$

The swap is accepted with

probability

$$r = \min \left\{ 1, \frac{\pi(\theta_{+,i+1} | D, \beta_i, I) \pi(\theta_{+,i} | D, \beta_{i+1}, I)}{\pi(\theta_{+,i} | D, \beta_i, I) \pi(\theta_{+,i+1} | D, \beta_{i+1}, I)} \right\}$$

Practicalities:

- Propose a swap every  $n_s$  iterations (achieved by drawing a random number  $U_i \sim \text{Uniform}[0, 1]$  every iteration and proposing a swap if  $U_i \leq 1/n_s$ )
- The choice of the length and intervals of the temperature ladder,