

# Gibbs sampling for Bayesian Inference

**Simon Wood**, University of Edinburgh, U.K.

## One at a time updating

- ▶ In MH sampling it can be difficult to find a distribution that makes good proposals for the whole parameter vector,  $\theta$ .
- ▶ But there is nothing in the MH scheme to prevent us using a proposal that updates the parameters in smaller blocks, or even one at a time, cycling through the elements of  $\theta$ .
- ▶ Such single element proposals are easier to tune.
- ▶ Another interesting fact is that if our proposal distribution is  $\pi(\theta|\mathbf{y})$  then the MH acceptance probability is exactly 1 (try it).
- ▶ This observation is not very practical for all at once updating of  $\theta$ , but it turns out to be very useful for one at a time updating, and is one way of motivating *Gibbs sampling*.

# Gibbs sampling

- ▶ Let  $\boldsymbol{\theta}_{-i}$  denote the elements of  $\boldsymbol{\theta}$  other than  $\theta_i$ .
- ▶ Suppose we want a random draw from  $\pi(\boldsymbol{\theta}|\mathbf{y})$ , and we already have a random draw,  $\boldsymbol{\theta}_{-i}^*$ , from  $\pi(\boldsymbol{\theta}_{-i}|\mathbf{y})$ .
- ▶ All we need to do is generate  $\theta_i^*$  from  $\pi(\theta_i|\boldsymbol{\theta}_{-i}^*, \mathbf{y})$  and append this to  $\boldsymbol{\theta}_{-i}^*$  to get  $\boldsymbol{\theta}^* \sim \pi(\boldsymbol{\theta}|\mathbf{y}) = \pi(\theta_i|\boldsymbol{\theta}_{-i}^*, \mathbf{y})\pi(\boldsymbol{\theta}_{-i}|\mathbf{y})$ .
- ▶ To get a slightly different draw from the same distribution we could discard  $\theta_i^*$  and draw a fresh value in the same way.
- ▶ We could repeatedly cycle through all elements of  $\boldsymbol{\theta}$ , dropping each and replacing with a draw from its conditional  $\pi(\theta_i|\boldsymbol{\theta}_{-i}^*, \mathbf{y})$ .
- ▶ This method is *Gibbs sampling*.
- ▶ To use it we need to identify the conditional densities.
- ▶ When identification of a conditional is not possible, we can use a Metropolis Hastings step for that parameter, or more efficient one dimensional alternatives.

## A simple example

- ▶ Consider  $n$  observations from model  $x_i \sim N(\mu, \sigma^2)$ , with priors
  - ▶  $\tau = 1/\sigma^2 \sim \text{gamma}(a, b)$ , i.e. prior  $\pi(\tau) = b^a \tau^{a-1} e^{-b\tau} / \Gamma(a)$
  - ▶ Independently,  $\mu \sim N(c, d)$ .
- ▶ The joint density of  $\mathbf{x}$ ,  $\mu$  and  $\tau$  is

$$\begin{aligned}\pi(\mathbf{x}, \tau, \mu) &= \pi(\mathbf{x}|\tau, \mu)\pi(\tau)\pi(\mu) \\ &\propto \tau^{n/2} e^{-\sum_i \tau(x_i - \mu)^2/2} e^{-(\mu - c)^2/(2d)} \tau^{a-1} e^{-b\tau}\end{aligned}$$

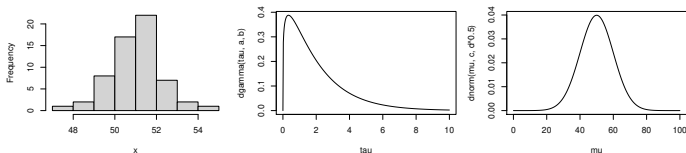
- ▶ By Bayes Theorem, the conditional for each parameter will be proportional to  $\pi(\mathbf{x}, \tau, \mu)$ . The trick is to recognize the conditional from its kernel - i.e from the terms actually involving the parameter.
- ▶ e.g.  $\pi(\tau|\mathbf{x}, \mu) \propto \tau^{n/2+a-1} e^{-\sum_i \tau(x_i - \mu)^2/2 - b\tau}$ , recognizable as a  $\text{gamma}(n/2 + a, \sum_i (x_i - \mu)^2/2 + b)$ .
- ▶ Similarly, but with a bit more effort,

$$\mu|\mathbf{x}, \tau \sim N\{(dn\bar{x}\tau + c)/(dn\tau + 1), d/(dn\tau + 1)\}$$

# Checking the priors

- ▶ Let's apply the preceding model to the `nhtemp` data. This will involve choosing the constants  $a$ ,  $b$ ,  $c$  and  $d$  defining the priors.
- ▶ We should look at the data, and check the priors...

```
x <- nhtemp ## just to save later typing
a <- 1.2; b <- .6 ## gamma prior shape and scale
c <- 50; d <- 100 ## normal prior mean and variance
## check data and priors...
par(mfrow=c(1,3),mar=c(4,4,1,1))
hist(x,main="")
tau <- seq(0,10,length=400)
plot(tau,dgamma(tau,a,b),type="l")
mu <- seq(0,100,length=400)
plot(mu,dnorm(mu,c,d^.5),type="l")
```



# Gibbs sampling loop

- Now we can run the Gibbs sampling loop...

```
ns <- 10000 ## number of samples
th <- matrix(0,2,ns) ## sample storage
mu <- 0;tau <- 0.1 ## initial states
n <- length(x)
## store constants needed repeatedly...
dn <- d*n; dnx <- dn*mean(x)

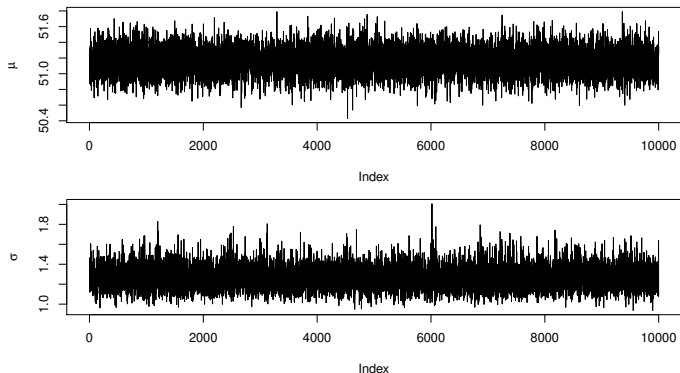
for (i in 1:ns) { ## Gibbs loop
  ## update mu | tau, x ...
  mu <- rnorm(1, (dnx*tau+c)/(dn*tau+1), sqrt(d/(dn*tau+1)))
  ## update tau | mu, x ...
  tau <- rgamma(1,n/2+a,sum((x-mu)^2)/2 + b)

  th[,i] <- c(mu,1/sqrt(tau)) ## store as mean and sd
}
```

# Checking the chains

- First look at the trace plots...

```
par(mfrow=c(2,1))  
## see ?plotmath for adding maths to plots  
plot(th[1,],type="l",ylab=expression(mu))  
plot(th[2,],type="l",ylab=expression(sigma))
```

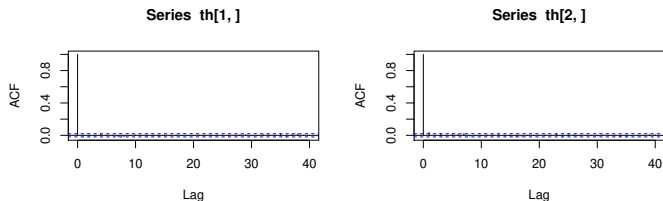


- Convergence appears to be very rapid, and mixing good.

# Checking autocorrelation

- Let's check the impression of rapid mixing with ACF plots...

```
par(mfrow=c(1,2))  
acf(th[1,]); acf(th[2,])
```



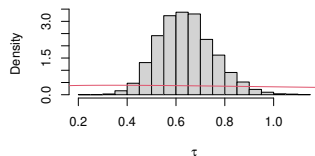
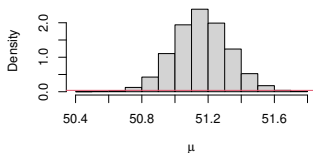
- The samples appear to be effectively independent of each other - actually this behaviour is special to this example - parameters with high posterior correlation would give chains with higher autocorrelation.



## Did the data matter?

- ▶ A common error is to believe that having updated your beliefs about the parameters using data, then *all* those beliefs are now validated by data. The Covid literature provides many examples.
- ▶ To avoid this it is a good idea to compare priors to posteriors, to see what effect the data had, if any.

```
hist(th[1,],xlab=expression(mu),main="",probability=TRUE)
mu <- seq(50,52,length=100)
lines(mu,dnorm(mu,c,d^.5),col=2)
hist(1/th[2,]^2,xlab=expression(tau),main="",probability=TRUE)
tau <- seq(0.1,1.2,length=100)
lines(tau,dgamma(tau,a,b),col=2)
```



- ▶ The posteriors are very different to the priors, which have little influence - here the posteriors are determined by the data.

## Posterior means and CIs

- Nothing is different here to in the MH sampling case (except that we only have 2 parameter in this example).

```
> pm <- rowMeans(th)
> names(pm) <- c("mu", "sigma")
> pm ## posterior mean
      mu      sigma
51.158764  1.263793
>
> ci <- apply(th, 1, quantile, probs=c(.025, .975))
> colnames(ci) <- c("mu", "sigma")
> ci ## confidence intervals
      mu      sigma
2.5%  50.83651  1.064045
97.5%  51.48939  1.511449
```

## **Temporary page!**

$\text{\LaTeX}$  was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will go away, because  $\text{\LaTeX}$  now knows how many pages to expect for the document.