Markov chain Monte carlo

The Bayesian approach

- Given new data, we update our beliefs
- All parameters are treated as random variables
- Posterior is proportional to the likelihood x prior

$$
\begin{gathered}
[\text{posterior}] \propto [\text{likelihood}][\text{prior}] \\
[\theta|y] \propto [y|\theta][\theta]
\end{gathered}
$$

Bayes theorem

$$
[\theta|y] = \frac{[y|\theta][\theta]}{[y]}
$$

$[y]$ is a normalizing constant, with the goal of permitting probabilistic statements about the parameters in $[\theta]$

Example: Bayesian linear regression

$$
\begin{aligned}
y_i &\sim \text{Normal}(\text{mean} = \mu_i, \text{SD} = \sigma) && [\text{likelihood}] \\
\mu_i &= \alpha + \beta * x_i && [\text{linear model}]
\end{aligned}
$$

Parameters we need to estimate: α, β, σ

$$
\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{average likelihood}}
$$

$$
[\alpha, \beta, \sigma \mid y] = \frac{[y \mid \alpha, \beta, \sigma] \times [\alpha][\beta][\sigma]}{\iiint [y \mid \alpha, \beta, \sigma][\alpha][\beta][\sigma] \, d\alpha \, d\beta \, d\sigma}
$$

The denominator is a beast. In complex models it is impossible to calculate.

Enter MCMC. But first, let's review the components of a Bayesian analysis.

Step 1: Decide on a prior $[\theta]$

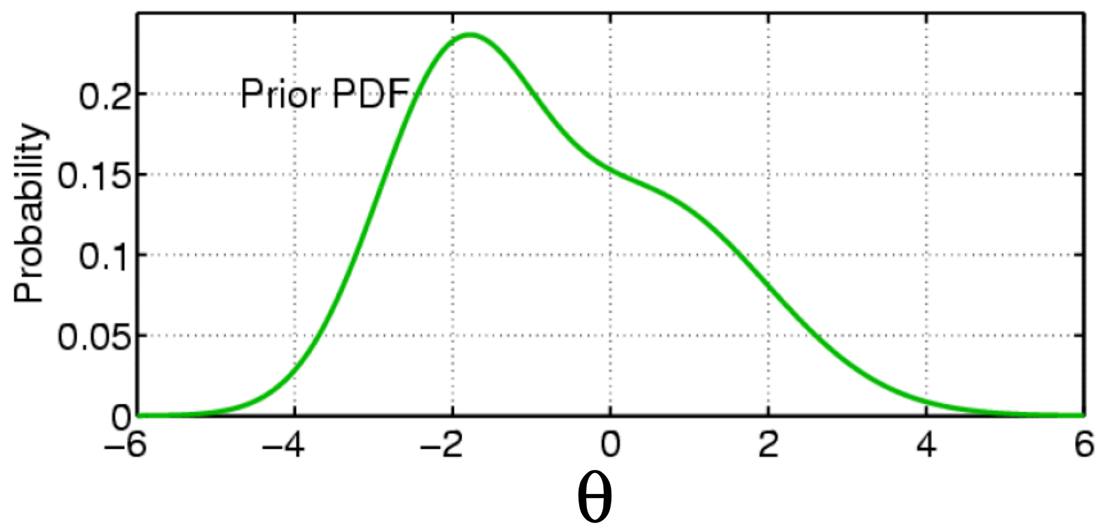

Figure 1: Image from a lecture by Che-Castaldo, Collins, Hobbs (2020)

Step 2: Compute the likelihood $[y \mid \theta]$

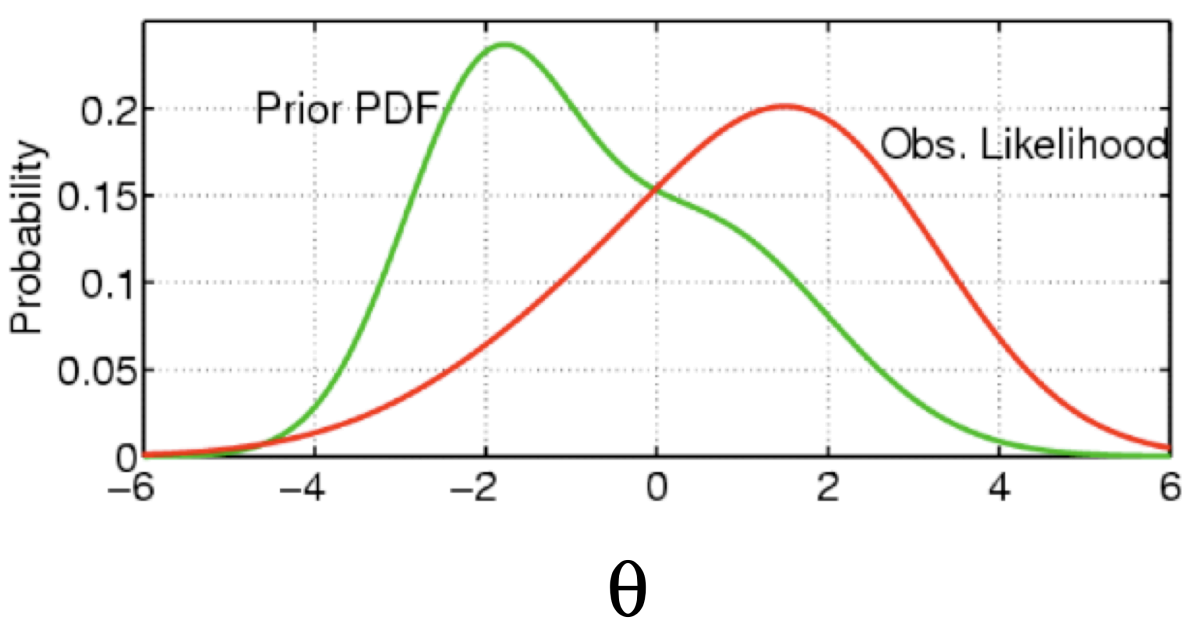

Figure 2: Image from a lecture by Che-Castaldo, Collins, Hobbs (2020)

Step 3: Calculate the numerator $[y \mid \theta]\,[\theta]$

$[y \mid \theta]\,[\theta] = [y, \theta]$

$[y, \theta]$ is the *joint distribution*

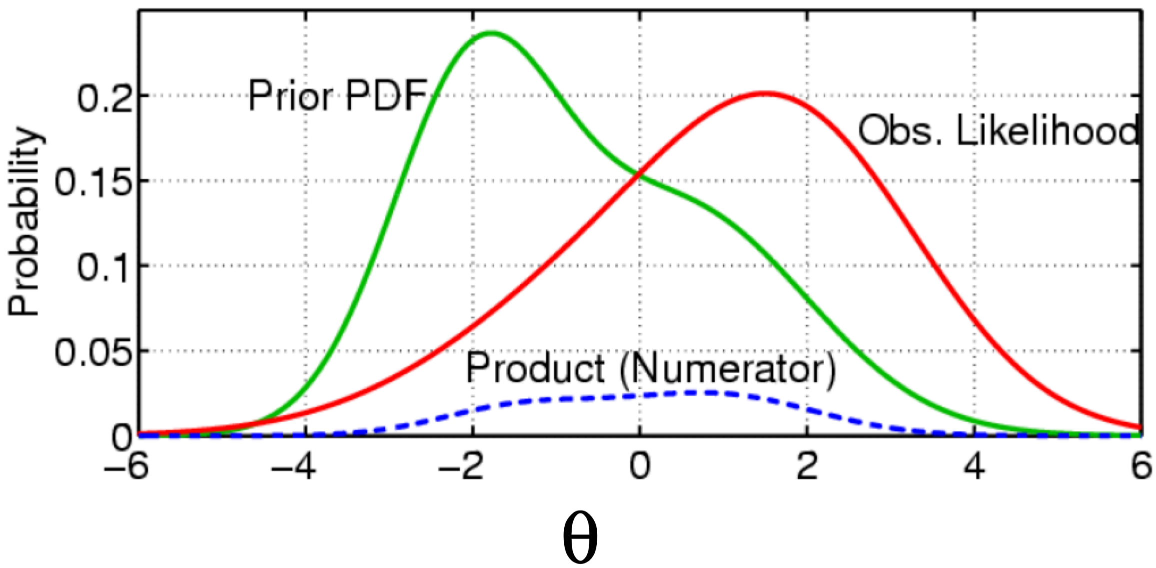

Figure 3: Image from a lecture by Che-Castaldo, Collins, Hobbs (2020)

Step 4: Integrate the joint distribution

The denominator is the area under the joint distribution:

$\int_\theta [y \mid \theta]\,[\theta]\,d\theta$

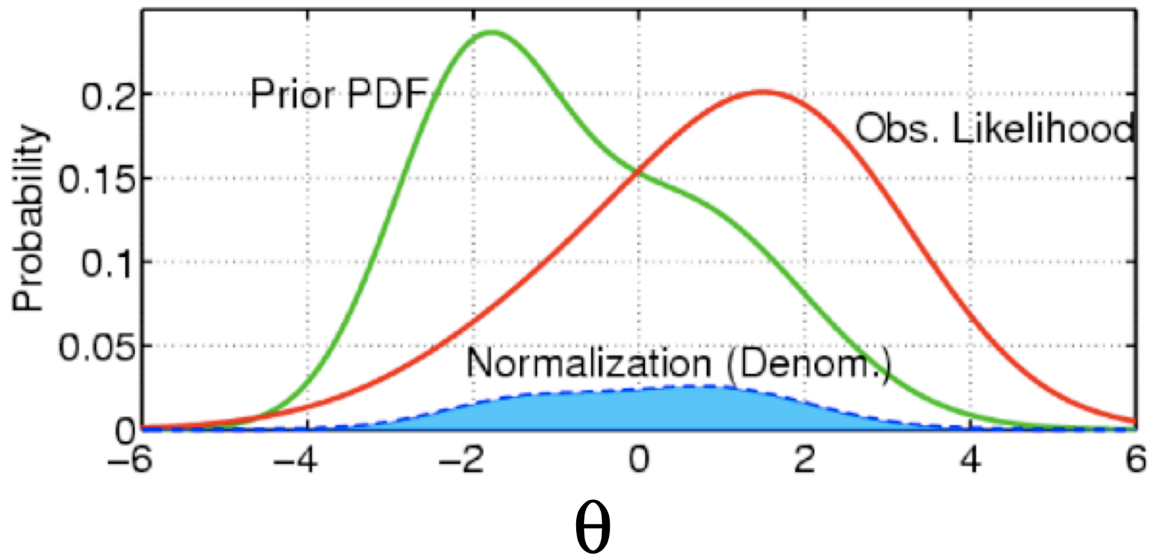

Figure 4: Image from a lecture by Che-Castaldo, Collins, Hobbs (2020)

Note that we are dividing each point on the dashed line by the area under the dashed line to obtain a probability density function.

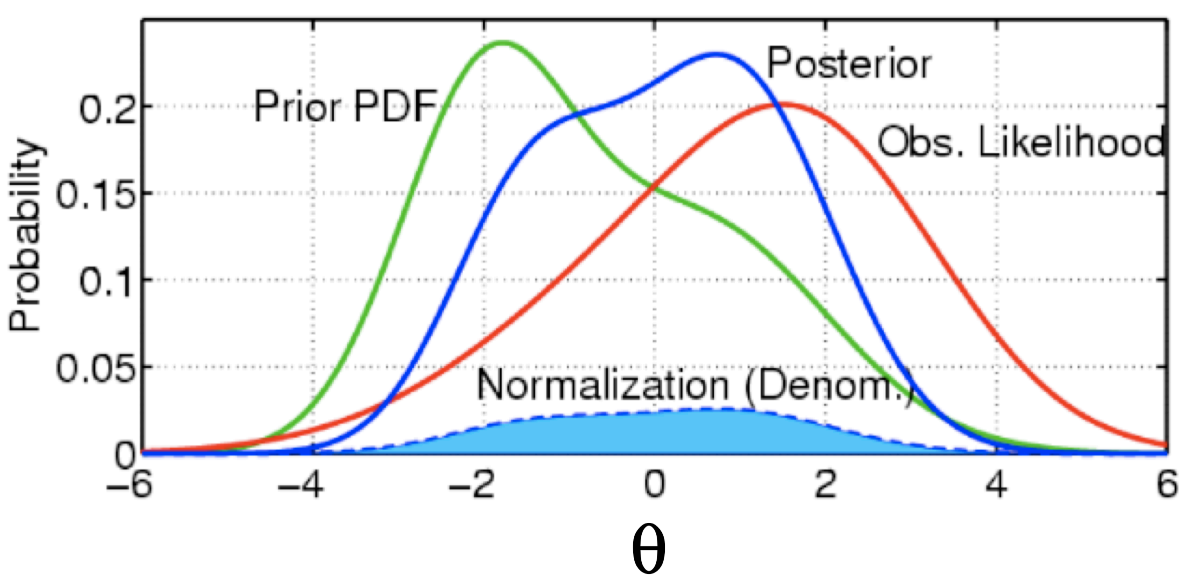

Figure 5: Image from a lecture by Che-Castaldo, Collins, Hobbs (2020)

How do we compute the posterior probability?

- Analytical approach

- Grid approximation
- Quadratic approximation
- Problems of high dimension will require Markov chain Monte Carlo

Markov chain Monte Carlo

- Markov: Russian mathematician (1856-1922)
- chain: sequence of random samples drawn from a probability distribution
- Monte Carlo: a famous casino

MCMC finds the posterior distribution by sampling from it

- Wait. How is it possible to draw samples from something that is unknown?
- Well, the posterior distribution is not entirely unknown
- $[\theta|y] \propto [y|\theta][\theta]$

What are we doing in MCMC?

- The posterior distribution is unknown, but we know the likelihood and the priors (i.e., the joint probability, $[\theta, y]$)
- We want to accumulate *many* values that represent random samples in proportion to their density in the posterior distribution
- MCMC generates these samples using $[\theta, y]$ to decide which samples to keep and which to throw away
- We can then use these samples to calculate statistics describing the distribution: means, medians, credible intervals, etc

MCMC algorithms

Accept-reject methods

- Metropolis (symmetric proposals)
- Metropolis-Hastings (asymmetric proposals)
- Gibbs sampling (adaptive proposals)
  - BUGS (Bayesian inferences using Gibbs sampling)
  - JAGS (just another Gibbs sampler)

Gradient methods

- Hamiltonian MC
  - Stan (named after Stanislaw Ulam)

Metropolis - one parameter

- θ: vector of K draws
- $\theta^{(k)}$: current value in the chain
- $\theta^{(*)}$: proposed value

1. Choose starting value for θ^1
2. Choose a new value, $\theta^{(*)}$, the *proposal* (can be independent of, or dependent on, θ^1)
3. Compute a probability of accepting the proposal
4. Accept the proposed value $\theta^{(*)}$ with the probability computed in step 3
5. Rinse and repeat

Metropolis

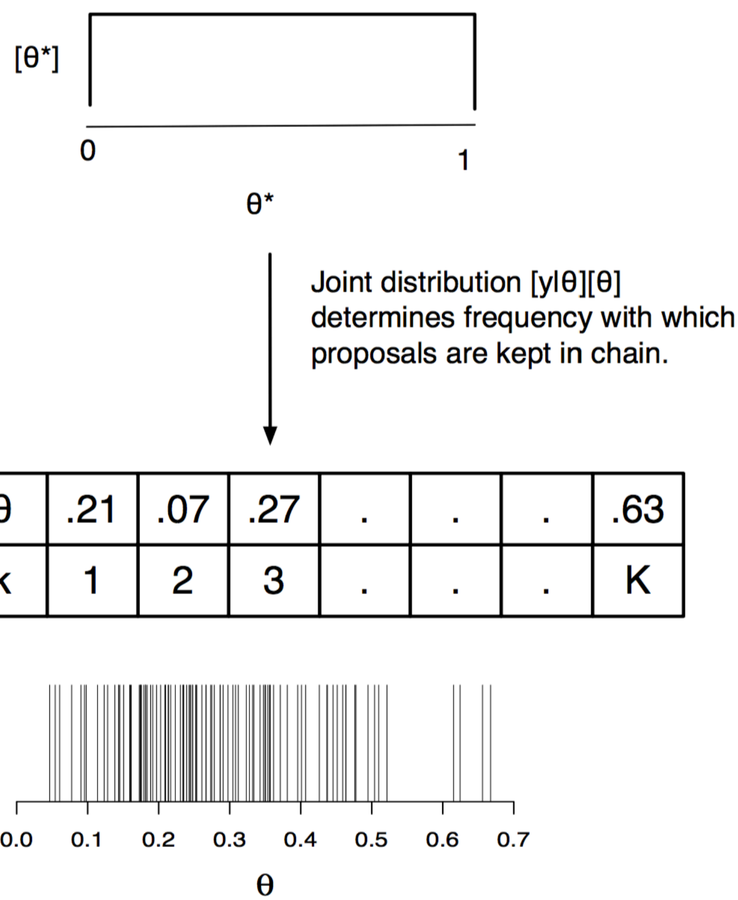

Figure 6: Hobbs and Hooten 2015 Fig 7.2

Samples from the posterior distribution

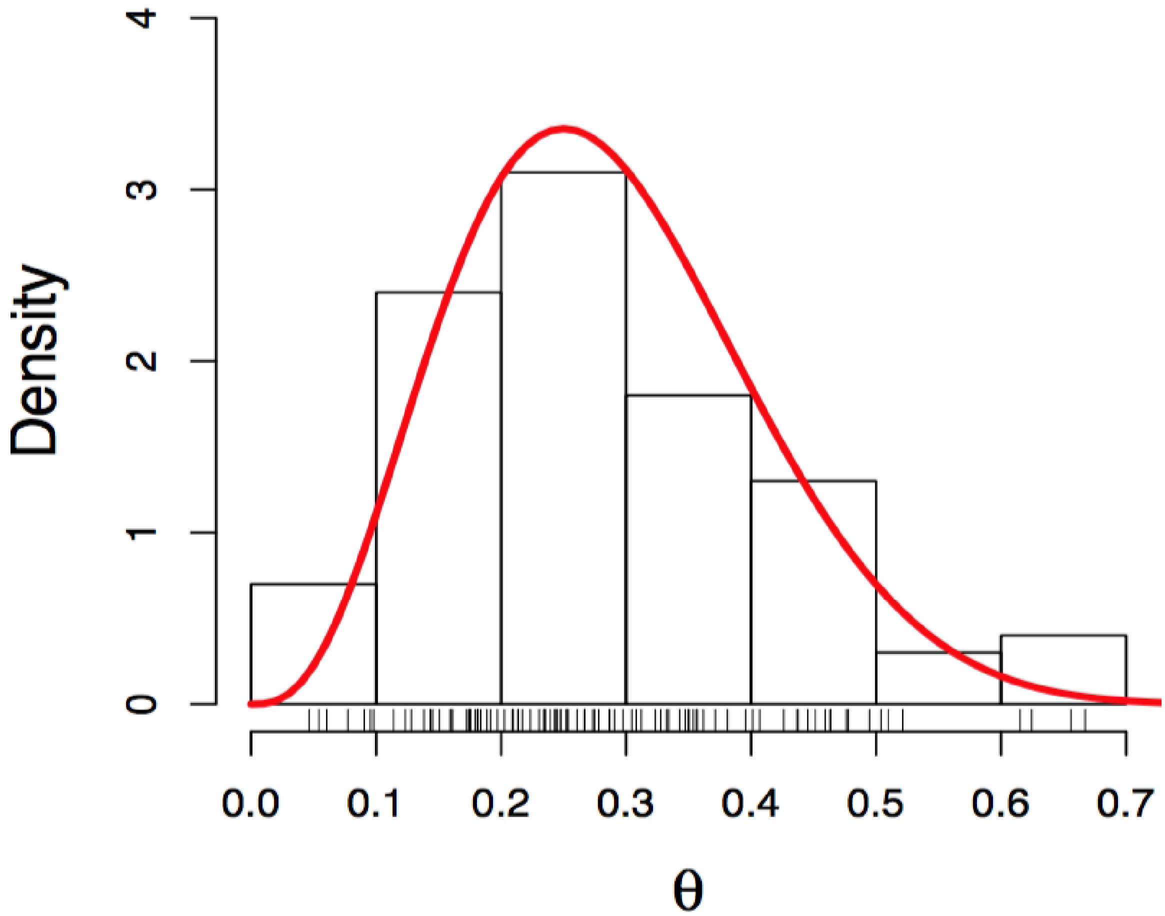

Figure 7: Hobbs and Hooten 2015 Fig 7.2

MCMC - multiple parameters

- For m parameters: each of the m unknowns has its own chain (i.e., θ_1, θ_2, θ_3,… θ_m).
- Assign an initial value to all chains.
- MCMC algorithm cycles over each parameter, treating it as if it were the *only* unknown - while the other parameters are treated (temporarily) as if they were known
- This decomposes a multivariate problem into a series of univariate problems

Homework

Read chapter 9 in *Statistical Rethinking* (McElreath 2020), and work through the code in 9.4 and 9.5 to practice the mechanics of the algorithm (Hamiltonian MCMC) we will use with `rethinking` and `Stan`.

Optional readings:

- chapter 7 in *Bayesian Models* (Hobbs and Hooten 2015)
- chapter 7 of *Bayes Rules* (Johnson, Ott, and Dogucu 2022); includes code for a Metropolis-Hastings algorithm
- a blog post by Thomas Wiecki (especially useful if you are partial to Python)

References

These notes are based on chapter 7 of *Bayesian Models* (Hobbs and Hooten 2015), a lecture by Che-Castaldo, Collins, and Hobbs as part of their Bayesian short course at SESYNC, and chapter 7 of *Bayes Rules!* (Johnson, Ott, and Dogucu 2022).

Hobbs, N Thompson, and Mevin B Hooten. 2015. *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press.

Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. 2022. *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. CRC Press.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.