Lecture 6

Outline:

1) Bayesian posteriors on ΛCDM parameters
2) Metropolis-Hasting Markov chain sampling.
3) Affine invariant ensemble MCMC.

Posterior $p(\theta | d_{\ell m}\text{'s})$ on the ΛCDM parameters $\theta = (\Omega_c h^2, \Omega_b h^2, \dots, n_s)$

Recall: our idealized data $\vec{d}$ is a vector of $d_{\ell m}$, $\ell = 0, 1, \dots, \ell_{max}$, $-\ell \le m \le \ell$ where

$$d_{\ell m} = T_{\ell m} + \varepsilon_{\ell m}$$

($T_{\ell m}$ depends on ΛCDM parameters)

s.t.

$$E(T_{\ell m}\overline{T_{\ell' m'}}) = \delta_{\ell\ell'}\delta_{mm'} C_\ell^{TT,\theta}$$

$$E(\varepsilon_{\ell m}\overline{\varepsilon_{\ell' m'}}) = \delta_{\ell\ell'}\delta_{mm'} \sigma^2 e^{\frac{b^2}{8\log 2}\ell(\ell+1)}$$

Given a prior $\pi(\theta)$ on the ΛCDM parameters the posterior $p(\theta|\vec{d})$ has the form

$$p(\theta|\vec{d}) \propto p(\vec{d}|\theta)\,\pi(\theta)$$

$$\therefore p(\theta|\vec{d}) \propto \pi(\theta) \prod_{\ell=0}^{\ell_{max}} \left(C_\ell^{TT,\theta} + C_\ell^{\varepsilon\varepsilon}\right)^{-\frac{2\ell+1}{2}} \exp\left(-\frac{2\ell+1}{2}\frac{\hat{C}_\ell}{\left(C_\ell^{TT,\theta} + C_\ell^{\varepsilon\varepsilon}\right)}\right)$$

Sampling $p(\theta|\vec{d})$ with Metropolis-Hastings (MH for short)

MH is one of the easiest & most basic techniques for setting up a Markov chain $\theta_1, \theta_2, \theta_3, \dots$ s.t. in the limit as $N \to \infty$, $\theta_N \sim p(\theta|\vec{d})$.

* Note: Many of these sampling algorithms are designed to circumvent having to compute the normalization factor in $p(\theta|\vec{d})$.

To run a MH chain you need a proposal density: $p_{prop}(\theta|\theta_0)$ ← depends on a previously sampled θ_0.

Algorithm: Basic MH

Input: initial start θ_0, proposal density $p_{prop}(\cdot|\cdot)$

Output: $\theta_1, \theta_2, \theta_3, \dots$

for $i = 1, 2, \dots$

* Set $\theta_{curr} = \theta_{i-1}$
* Sample $\theta_{prop} \sim p_{prop}(\theta | \theta_{curr})$
* Draw $u \sim Uniform(0,1)$ & define

$$\alpha := \min\left(\frac{p(\theta_{prop}|\vec{d})}{p(\theta_{curr}|\vec{d})}\frac{p_{prop}(\theta_{curr}|\theta_{prop})}{p_{prop}(\theta_{prop}|\theta_{curr})}, 1\right)$$

* if $u \le \alpha$
  - set $\theta_i = \theta_{prop}$
  - else
  - set $\theta_i = \theta_{curr}$
  - end

end

Fact 1:

If $p_{prop}(\theta \mid \theta_{curr}) = p_{prop}(\theta_{curr} \mid \theta)$

then α can be simplified to

$$\alpha := \min\left(\frac{p_{prop}(\theta_{prop} \mid \vec{d})}{p_{prop}(\theta_{curr} \mid \vec{d})}, 1\right)$$

Fact 2:

MH does not require evaluation of the normalizing constant in $p(\theta \mid \vec{d})$

Fact 3:

The hard part about MH is choosing a p_{prop} which gives good acceptance rates. This is difficult when degeneracies are present.

Why does MH work?

Claim: Let $x_1, x_2, \ldots$ be a markov chain [i.e. that $p(x_{i+1} \mid x_1, x_2, \ldots, x_i) = p(x_{i+1} \mid x_i)$] with invariant density f [i.e. $x_i \sim f \Rightarrow x_{i+1} \sim f$]

Then $x_n \xrightarrow[n \to \infty]{T.V.} f$.

see p. 163 in Robert & Casellas book "Monte Carlo Statistical Methods"

∴ All we need to show is that MH has the posterior $p(\theta \mid \vec{d})$ has an invariant density.

Claim: Let $x_1, x_2, \ldots$ be a MH markov chain for sampling from some density f where the support of $p_{prop}(x \mid y)$ contains the support of f. Let $K(x \mid y)$ denote the transition density. Then

i) $K(x \mid y) f(y) = K(y \mid x) f(x)$
(detailed balance ... same thing as reversibility)

which implies

ii) f is an invariant density.

Proof: Let

$$\alpha(x \mid y) = \min\left(\frac{f(x)}{f(y)} \frac{p_{prop}(y \mid x)}{p_{prop}(x \mid y)}, 1\right).$$

Fix y.

Let A, B be the events

$$A := \{x_{prop} \in x + dx\}$$

$$B := \{U \le \alpha(x_{prop} \mid y)\}$$

Notice we are working with densities so we can suppose w.l.g that $x \neq y$

$$\therefore K(x \mid y)dx = P(A \cap B)$$

$$= \int P(A \cap B \mid x_{prop}) p_{prop}(x_{prop} \mid y) dx_{prop}$$

↓ (this factors to $P(A \mid x_{prop}) P(B \mid x_{prop})$)

$$= I_{(x, x+dx)} \alpha(x_{prop} \mid y)$$

$$= \alpha(x \mid y) p_{prop}(x \mid y) dx$$

Now we show detailed balance for K

$$K(x|y)f(y) = \alpha(x|y)p_{prop}(x|y)f(y)$$
$$= \min\left(f(x)p_{prop}(y|x), p_{prop}(x|y)f(y)\right)$$
$$= \alpha(y|x)p_{prop}(y|x)f(x).$$
$$= K(y|x)f(x).$$

This shows i). For ii) just notice

$$x_i \sim f(x) \Rightarrow x_{i+1} \sim \int K(x|y)f(y)dy$$
$$= \int K(y|x)f(x)dy$$
$$= f(x).$$

□

Choosing the proposal density p_{prop}

- If you have an idea of the posterior cov Σ and the parameters are not too nonlinear a common choice is

$$p_{prop}(\theta|\theta_{curr}) \sim N(\theta_{curr}, g\Sigma)$$

↑ tune this to get a good acceptance rate...

- If the data model is $\vec{d} \simeq \underbrace{X}_{\nabla_\theta \vec{e}^{T,\theta}}(\theta-\theta_0) + \vec{\varepsilon}$ try

$$p_{prop}(\cdot|\theta_{curr}) \sim N(\theta_{curr}, g(X^TN^{-1}X)^{-1})$$

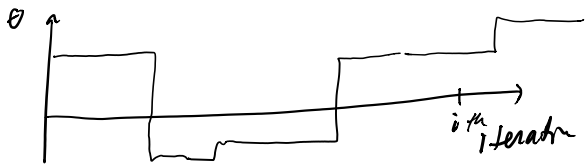
where $N = E(\vec{\varepsilon}\vec{\varepsilon}^T)$. ($g(X^TN^{-1}X)^{-1}$ encompasses linear dependence)

Note: Non informative priors of the form

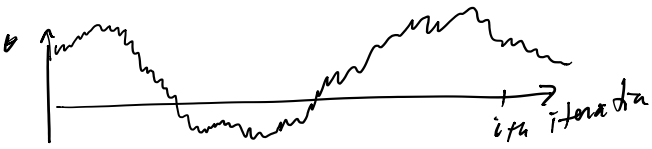
$$\pi(\beta) \sim N(\beta_0, g(X^TN^{-1}X)^{-1})$$

for the regression model $\vec{d} = X\beta + \varepsilon$ are called "g-priors".

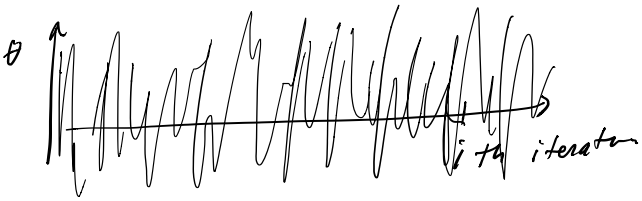
If the acceptance rate is too small the chain will look like



If too large...



A good rule of thumb is an acceptance of 20%



Burn in, thinning & initial start.

- For LCDM it can help to start at a numerical approximation to

$$\theta_{max} = \arg\max_\theta p(\theta|\vec{d}).$$

- Discard initial steps to avoid dependence on initial starting point
- thin the chain to reduce within chain dependence.

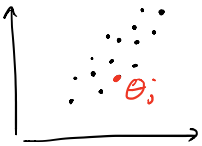
Affine-Invariant ensemble MCMC

Ref: Goodman & Weare (2010).

A MCMC chain whose performance is invariant under linear transformations ... useful for degenerate posteriors.

Algorithm

Input: $a > 0$, $n \in \mathbb{Z}^+$ and an ensemble of "walkers" $\vec{\theta} := (\theta_1, \dots, \theta_n)$ where each θ_i is in the parameter space (e.g. $a = 2$, $n = 100$).



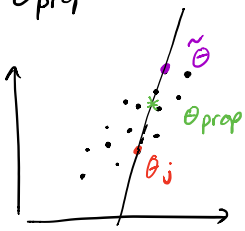
Output: $\vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3, \dots$

for $i = 1, 2, \dots$ ← i indexes iteration

* set $\vec{\theta}_{i+1} = \vec{\theta}_i$

for $j = 1, \dots, n$ ← j indexes walkers

* draw z with density proportional to $\frac{1}{\sqrt{z}} I_{[\frac{1}{a}, a]}(z)$
* Randomly choose a walker $\tilde{\theta}$ from $\{\theta_1, \dots, \theta_n\} \setminus \{\theta_j\}$
* set $\theta_{prop} = \tilde{\theta} + z(\theta_j - \tilde{\theta})$



* Draw $u \sim \text{Uniform}(0,1)$ & define

$$\alpha := \min\left(z^{d-1} \frac{P(\theta_{prop} \mid \vec{d})}{P(\theta_j \mid \vec{d})}, 1 \right)$$

$d :=$ dim of the parameter space

* if $u \le \alpha$
  set $(\vec{\theta}_{i+1})_j = \theta_{prop}$
  end

end

end

Note: the stationary distribution $\vec{\theta}$ is

$$\vec{\theta} \sim \prod_{j=1}^{N} P(\theta_j \mid \vec{d}) = \text{iid copies of } P(\theta \mid \vec{d}).$$