MCMC

- We will discuss:
  - What it is
  - How it works (at a high level)
  - How to implement it
- We will not discuss:
  - Theory/why it works
  - Convergence details

Monte Carlo Simulation

- Idea: instead of computing a probability or distribution analytically, we can estimate it via simulation
- Generate a large number of random samples and estimate empirically
- Examples:
  - Is this coin fair?
  - Homework 6 question 1
  - Monte Carlo Monte Hall

Markov Chain Review

- Stochastic process X_n, $n = 0, 1, 2, \dots$
- $X_n \in \mathcal{S}$ (finite or not, discrete or not)
- $P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) =$ $P(X_{n+1} = j | X_n = i) = P_{ij}$
  - Transition probability P_{ij} only depends on the current state
  - P_{ij} does not change over time
- Equilibrium distribution

Markov Chain Monte Carlo (MCMC)

- Construct a Markov chain with stationary distribution equal to the posterior distribution of the parameter(s)
- Run the chain until it is in equilibrium, then get a Monte Carlo estimate of the posterior distribution, or any posterior quantity
- How do we do this?
  - Metropolis–Hastings Algorithm
  - Gibbs Sampling
- Irreducible, aperiodic, time-reversible
- Detailed balance

Metropolis–Hastings

- Conditional on the current state of the chain, θ (values of the parameters), propose a new candidate value(s), θ^*
- For this candidate, either accept with a certain probability and move to that state, or reject and remain in the current state
- The probability of acceptance is

$$\alpha = \min\left(1,\ \frac{L(\theta^*)\pi(\theta^*)/q(\theta^*|\theta)}{L(\theta)\pi(\theta)/q(\theta|\theta^*)}\right)$$

where l is the likelihood, π is the prior, and q is the density function for the proposal

Simple Example

Your mischievous brother owns a coin which you know to be loaded so that it comes up heads 70% of the time. He comes to you with a coin and wants to make a bet with you. You don't know if this is the loaded coin or a fair one. He lets you flip it 5 times to check it out, and you get 2 heads and 3 tails. Which coin do you think it is?

Parameter $\theta \in \{\text{fair}, \text{loaded}\}$

Likelihood:

$$
\begin{aligned}
P(X = x|\theta = \text{fair}) &= \binom{5}{x}\left(\frac{1}{2}\right)^5 \\
P(X = x|\theta = \text{loaded}) &= \binom{5}{x}(.7)^x(.3)^{5-x}
\end{aligned}
$$

MLE:

$$
f(\theta|X = 2) = \begin{cases} 0.3125 & \text{if } \theta = \text{fair} \\ 0.1323 & \text{if } \theta = \text{loaded} \end{cases}
$$

Since you know your brother reasonably well, you might think that there is a 60% chance that this is the loaded coin, before you get to flip it 5 times. You then get 2 heads and 3 tails. What is your posterior probability that the coin is loaded?

$$
\begin{aligned}
Pr(\text{loaded}|X) &= \\
&= \frac{Pr(X|\text{loaded})Pr(\text{loaded})}{Pr(X|\text{loaded})Pr(\text{loaded}) + Pr(X|\text{fair})Pr(\text{fair})} \\
&= \frac{\binom{5}{2}(.7)^2(.3)^3(.6)}{\binom{5}{2}(.7)^2(.3)^3(.6) + \binom{5}{2}(.5)^2(.5)^3(.4)} \\
&= \frac{0.07938}{0.07938 + 0.125} = 0.388
\end{aligned}
$$

We could use MCMC instead, and our MCMC estimate for the probability the coin is loaded is is the fraction of $\theta^{(i)} = \textit{loaded}$ in the MCMC sample.

1. Start the chain at an arbitrary value, say $\theta^{(0)}$=fair.
2. Let $\theta^* = \begin{cases} \text{fair} & \text{if } \theta^{(i)} = \text{loaded} \\ \text{loaded} & \text{if } \theta^{(i)} = \text{fair} \end{cases}$
3. $\alpha = \min\left\{\frac{L(\theta^*)P(\theta^*)q(\theta^{(i)}|\theta^*)}{L(\theta^{(i)})P(\theta^{(i)})q(\theta^*|\theta^{(i)})}, 1\right\}$.
   Note that $q = 1$.

If $\theta^{(i)}$ = fair, then θ^* = loaded and

$$\alpha = \frac{\left(\binom{5}{2}(.7)^2(.3)^3\right)(.6)(1)}{\left(\binom{5}{2}(.5)^2(.5)^3\right)(.4)(1)} = \frac{0.07938}{0.125} = 0.635$$

If $\theta^{(i)}$ = loaded, then θ^* = fair and

$$\alpha = \min\left(1, \frac{\left(\binom{5}{2}(.5)^2(.5)^3\right)(.4)(1)}{\left(\binom{5}{2}(.7)^2(.3)^3\right)(.6)(1)}\right) = 1$$

So $\alpha = \begin{cases} 0.635 & \text{if } \theta^* = \text{fair} \\ 1 & \text{if } \theta^* = \text{loaded} \end{cases}$

4. Thus if $\theta^{(i)}$ = loaded, then $\theta^{(i+1)}$ = fair.
   If $\theta^{(i)}$ = fair, then
   $$\theta^{(i+1)} = \begin{cases} \text{loaded} & \text{with probability } 0.635 \\ \text{fair} & \text{with probability } 1 - 0.635 \end{cases}$$
5. Repeat steps 2-4.

In equilibrium (which this chain is in immediately), the chain should spend $\frac{0.635}{1+0.635} = 0.388$ of its time in the *loaded* state, which is the same answer we got analytically. In most cases, we won't be able to do this computation, but we can still construct and sample from the chain, and do estimation using the MCMC sample.

Simple Normal Example

$$X \sim N(\mu, 1)$$

$$\pi(\mu) \sim N(0, 1)$$

Thus we know analytically that

$$\mu | X \sim N\left(\frac{\sum x_i}{n+1}, \frac{1}{n+1}\right)$$

But as an example, we will use the Metropolis algorithm (using a symmetric proposal, so that $q(\mu^*|\mu) = q(\mu|\mu^*)$ and that term cancels out).

`mcmc.normal.1.c`

1. Start the chain at an arbitrary value, for example, drawing $\mu^{(0)} \sim U[-0.5, 0.5]$.

2. Sample $\mu^* \sim U\left[\mu^{(i)} - 0.1,\ \mu^{(i)} - 0.1\right]$

3.

$$
\begin{aligned}
\alpha &= \min\left\{\frac{L(\mu^*)P(\mu^*)}{L(\mu^{(i)})P(\mu^{(i)})}, 1\right\} \\
&= \min\left(\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x_i - \mu^*\right)^2\right)\exp\left(-\frac{1}{2}\mu^{*2}\right)}{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x_i - \mu^{(i)}\right)^2\right)\exp\left(-\frac{1}{2}\mu^{(i)2}\right)}, 1\right)
\end{aligned}
$$

4. Draw $U \sim U[0, 1]$.

$$
\mu^{(i+1)} = \begin{cases} \mu^* & \text{if } U \leq \alpha \\ \mu^{(i)} & \text{if } U > \alpha \end{cases}
$$

5. Repeat steps 2-4.

A Different Normal Example

$$X \sim N(\mu, 1)$$

$$\pi(\mu) \sim \Gamma(3, 1)$$

Now we can't get the posterior analytically, so we have to use MCMC

What proposal distribution should we use?

- Propose $\mu^* \sim U\left[\mu^{(i)} - 0.1,\ \mu^{(i)} - 0.1\right]$
  - Simpler proposal – don't have to worry about q
  - Need to make sure $\mu^* > 0$
  - `mcmc.normal.2.c`
- Propose $\mu^* \sim U\left[0.5 * \mu^{(i)},\ 2 * \mu^{(i)}\right]$
  - Now we need to include q
  - This may explore the space better if the parameter is multiplicative
  - `mcmc.normal.3.c`

Gibbs Sampling

- Draw proposals from their complete conditional distributions
- Thus $q = L * \pi$, so $\alpha = 1$.
- Usually more efficient, since nothing is rejected, but the step sizes can be large
- For multiple parameters, cycle through them
- Can mix Gibbs steps and Metropolis-Hastings steps
- Generally prefer Gibbs steps when possible

One Last Normal Example

$$X \sim N(\mu, 1)$$

$$\pi(\mu|\psi) \sim N(\psi, 1)$$

$$\pi(\psi) \sim \Gamma(3, 1)$$

$$\begin{aligned}
P(\mu, \psi|X) &\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \exp\left[-\frac{1}{2}(\mu - \psi)^2\right] \psi^2 \exp\left[-\psi\right] \\
P(\mu|\psi, X) &\propto \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \mu)^2 + (\mu - \psi)^2\right)\right] \\
\mu|\psi, X &\sim N\left(\frac{\sum x_i + \psi}{n+1}, \frac{1}{n+1}\right)
\end{aligned}$$

- Update μ with a Gibbs step
- Update ψ with a Metropolis-Hastings step

`mcmc.normal.4.c`

Fitting a Hierarchical Model for the VA Dataset

- Account for overdispersion with a separate θ_i for each hospital:

$$
\begin{aligned}
y_i|\theta_i &\sim Bin(n_i, \theta_i) \\
\theta_i|\alpha, \beta &\sim Beta(\alpha, \beta) \\
\alpha|a, b &\sim \Gamma(a, b) \\
\beta|a, b &\sim \Gamma(a, b)
\end{aligned}
$$

- Posterior: $P(\boldsymbol{\theta}, \alpha, \beta|\mathbf{y}) \propto$

$$\prod_{i=1}^{M} \left[\binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i-y_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1-\theta_i)^{\beta-1} \right] \alpha^{a-1} e^{-b\alpha} \beta^{a-1} e^{-b\beta}$$

- $\theta_i|\alpha, \beta, \mathbf{y} \sim Beta(y_i + \alpha, n_i - y_i + \beta)$
- $P(\alpha|\boldsymbol{\theta}, \beta, \mathbf{y}) \propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\right]^M \left[\prod_{i=1}^{M} \theta_i\right]^{\alpha-1} \alpha^{a-1} e^{-b\alpha}$

`mcmc.va.c`

MCMC Issues

- Convergence/burn-in
  - Plot output sequentially and observe
  - Some diagnostic tools exist, but the jury is still out on them
- Mixing and exploration
  - Don't want the chain to get stuck
  - Chain should visit all important regions of the space (e.g., all large local maxima)
- Efficiency
  - Want to explore the space quickly, want large enough steps
  - Don't want to reject too many candidates
  - Multivariate draws
  - Optimal acceptance rate in (0.25, 0.45), higher dimensional draws on the lower end of that interval

- Autocorrelation
  - Successive draws are usually highly correlated
  - Generally doesn't matter except for estimating variances
  - Can just take every fifth (tenth, etc.) draw
- Starting values
  - Starting value may affect where the chain ends up
  - Good strategy is to try several different starting points
- One chain or many?
  - For checking convergence/exploration, several chains are better
  - We tend not to worry about autocorrelation