# MH Logistics

Jacob M. Montgomery

2017

# Metropolis-Hastings Algorithm and Logistic Regression

# The setup

$$\log\left(\frac{p(\mathbf{x_i})}{1 - p(\mathbf{x_i})}\right) = \alpha + \beta \mathbf{x}_i$$

# The setup

$$\log\left(\frac{p(\mathbf{x_i})}{1-p(\mathbf{x_i})}\right) = \alpha + \boldsymbol{\beta}\mathbf{x}_i$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \boldsymbol{\beta}\mathbf{x}_i$$

# The setup

$$\log\left(\frac{p(\mathbf{x_i})}{1 - p(\mathbf{x_i})}\right) = \alpha + \boldsymbol{\beta}\mathbf{x}_i$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \alpha + \boldsymbol{\beta}\mathbf{x}_i$$

$$p_i = \frac{\exp(\alpha + \boldsymbol{\beta}\mathbf{x})}{1 + \exp(\alpha + \boldsymbol{\beta}\mathbf{x})}$$

# Overview

- We can calculate the likelihood, and we can put on a prior
- But it is difficult to turn this into a Gibbs sampler
- Instead we use the Metropolis-Hastings algorithm

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.
- The MH algorithm consists of:
  1. **Propose** a new value for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}'$ from some proposal density $f(\boldsymbol{\theta})$.
  2. **Accept** this proposal with probability:

$$\min\left\{\frac{L(\boldsymbol{\theta}')\pi(\theta')f(\boldsymbol{\theta})}{L(\boldsymbol{\theta})\pi(\theta)f(\boldsymbol{\theta}')},1\right\}$$

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.
- The MH algorithm consists of:
  1. **Propose** a new value for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}'$ from some proposal density $f(\boldsymbol{\theta})$.
  2. **Accept** this proposal with probability:

$$\min\left\{\frac{L(\boldsymbol{\theta}')\pi(\theta')f(\boldsymbol{\theta})}{L(\boldsymbol{\theta})\pi(\theta)f(\boldsymbol{\theta}')}, 1\right\}$$

Let's chalk about that for a minute.

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.
- The MH algorithm consists of:
  1. **Propose** a new value for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}'$ from some proposal density $f(\boldsymbol{\theta})$.
  2. **Accept** this proposal with probability:

$$\min\left\{\frac{L(\boldsymbol{\theta}')\pi(\theta')f(\boldsymbol{\theta})}{L(\boldsymbol{\theta})\pi(\theta)f(\boldsymbol{\theta}')}, 1\right\}$$

Let's chalk about that for a minute. We want to generate a stationary distribution

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.
- The MH algorithm consists of:
  1. **Propose** a new value for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}'$ from some proposal density $f(\boldsymbol{\theta})$.
  2. **Accept** this proposal with probability:

$$\min\left\{\frac{L(\boldsymbol{\theta}')\pi(\theta')f(\boldsymbol{\theta})}{L(\boldsymbol{\theta})\pi(\theta)f(\boldsymbol{\theta}')}, 1\right\}$$

Let's chalk about that for a minute. We want to generate a stationary distribution and this distribution must be unique.

# Overview of MH

- Let $\boldsymbol{\theta}$ be some parameter(s) we wish to to sample, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\theta)$ is the prior distribution.
- The MH algorithm consists of:
  1. **Propose** a new value for $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}'$ from some proposal density $f(\boldsymbol{\theta})$.
  2. **Accept** this proposal with probability:

$$\min\left\{\frac{L(\boldsymbol{\theta}')\pi(\theta')f(\boldsymbol{\theta})}{L(\boldsymbol{\theta})\pi(\theta)f(\boldsymbol{\theta}')}, 1\right\}$$

Let's chalk about that for a minute. We want to generate a stationary distribution and this distribution must be unique.

Note that we are going to want to calculate this in terms of logs.

## Back to the logistic distribution: Likelihood

- So let's figure out the likelihood together:

$$L(\alpha, \beta) \propto$$

## Back to the logistic distribution: Likelihood

- So let's figure out the likelihood together:

$$L(\alpha, \beta) \propto \prod_{i=1}^{n} \left( \frac{\exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)}{1 + \exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)} \right)^{y_i} \left( \frac{1}{1 + \exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)} \right)^{1 - y_i}$$

## Back to the logistic distribution: Likelihood

- So let's figure out the likelihood together:

$$L(\alpha, \beta) \propto \prod_{i=1}^{n} \left( \frac{\exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)}{1 + \exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)} \right)^{y_i} \left( \frac{1}{1 + \exp(\alpha + \boldsymbol{\beta} \mathbf{x}_i)} \right)^{1-y_i}$$

- Let $q_i = 1 - p_i = \frac{1}{1+\exp(\alpha+\boldsymbol{\beta}\mathbf{x}_i)}$
- $$\mathcal{L}(\alpha, \beta) \propto \sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i)(\log(q_i))$$

## Prior

- We are going to use something called an "Empirical Bayes" approach
- That is, we are going to construct our prior from the data itself.

## Prior

- We are going to use something called an "Empirical Bayes" approach
- That is, we are going to construct our prior from the data itself.
- Specifically, we are going to say that $\pi(\alpha, \beta) = \pi(\alpha|\hat{b})\pi(\beta)$ where $\pi(\alpha|\hat{b})$ is an exponential prior over $\exp(\alpha)$ with hyperparameter $\hat{b}$ estimated from the data:

$$\frac{1}{b} e^{\alpha} e^{-\exp(\alpha)/\hat{b}}$$

- In so doing, we are "centering" the prior near the MLE for $\alpha$ and a flat (uninformative) prior over $\beta$

## Proposal function

- We need to balance to different needs:
  1. We need to propose numbers that are “different” enough that we move around the parameter space.
  2. The proposals need to be reasonable enough that we don’t reject them all

# Proposal function

- We need to balance to different needs:
  1. We need to propose numbers that are "different" enough that we move around the parameter space.
  2. The proposals need to be reasonable enough that we don't reject them all (think about what happens with a proposal that is very unlikely)

## Proposal function

- We need to balance to different needs:
  1. We need to propose numbers that are "different" enough that we move around the parameter space.
  2. The proposals need to be reasonable enough that we don't reject them all (think about what happens with a proposal that is very unlikely)

- In this case, we are going to use the proposal density

$$f(\alpha, \beta) = \pi_\alpha(\alpha|\hat{b})\psi(\beta)$$

where $\pi_\alpha$ is our prior and $\psi(\beta)$ is a normal density with mean and posterior covariance coming from the MLE.

## Proposal function

- We need to balance to different needs:
    1. We need to propose numbers that are "different" enough that we move around the parameter space.
    2. The proposals need to be reasonable enough that we don't reject them all (think about what happens with a proposal that is very unlikely)

- In this case, we are going to use the proposal density

$$f(\alpha, \beta) = \pi_\alpha(\alpha|\hat{b})\psi(\beta)$$

  where $\pi_\alpha$ is our prior and $\psi(\beta)$ is a normal density with mean and posterior covariance coming from the MLE.
- Note that we often propose next steps based on where we are right now in the space, but any proposal density will do.

## MH implementation

- So in this case, we are going to propose new values of $\alpha$ and $\beta$ and then accept with probability

$$\min\left\{\frac{L(\alpha',\beta')\pi_\alpha(\alpha'|\hat{b})\pi_\alpha(\alpha|\hat{b})\psi(\beta)}{L(\alpha,\beta)\pi_\alpha(\alpha|\hat{b})\pi_\alpha(\alpha'|\hat{b})\psi(\beta')},1\right\}$$

## MH implementation

- So in this case, we are going to propose new values of $\alpha$ and $\beta$ and then accept with probability

$$\min\left\{\frac{L(\alpha',\beta')\pi_\alpha(\alpha'|\hat{b})\pi_\alpha(\alpha|\hat{b})\psi(\beta)}{L(\alpha,\beta)\pi_\alpha(\alpha|\hat{b})\pi_\alpha(\alpha'|\hat{b})\psi(\beta')},1\right\}$$

- $$\min\left\{\frac{L(\alpha',\beta')\psi(\beta)}{L(\alpha,\beta)\psi(\beta')},1\right\}$$

## MH implementation

- So in this case, we are going to propose new values of $\alpha$ and $\beta$ and then accept with probability

$$\min\left\{\frac{L(\alpha',\beta')\pi_\alpha(\alpha'|\hat{b})\pi_\alpha(\alpha|\hat{b})\psi(\beta)}{L(\alpha,\beta)\pi_\alpha(\alpha|\hat{b})\pi_\alpha(\alpha'|\hat{b})\psi(\beta')},1\right\}$$

- $$\min\left\{\frac{L(\alpha',\beta')\psi(\beta)}{L(\alpha,\beta)\psi(\beta')},1\right\}$$

- Sounds easy, right? Let's try it