

# How to do (some) Bayesian stuff in R

Spencer Fox\*

\* Disclaimer: I'm not a statistician

# Optimistic Goals

# Optimistic Goals

1. Know when (if) you should use bayesian statistics

# Optimistic Goals

1. Know when (if) you should use bayesian statistics
2. Gain intuition about the theory

# Optimistic Goals

1. Know when (if) you should use bayesian statistics
2. Gain intuition about the theory
3. Implement Bayesian stats in R

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

goal:  $P(DS|TT+)$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

goal:  $P(DS|TT+)$

general multiplication rule:

- 1  $P(DS|TT+)P(TT+) = P(DS \& TT+)$
- 2  $P(TT+|DS)P(DS) = P(DS \& TT+)$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

goal:  $P(DS|TT+)$

general multiplication rule:

- 1  $P(DS|TT+)P(TT+) = P(DS \& TT+)$
- 2  $P(TT+|DS)P(DS) = P(DS \& TT+)$

rearrange #1:  $P(DS|TT+) = \frac{P(DS \& TT+)}{P(TT+)}$

# Example

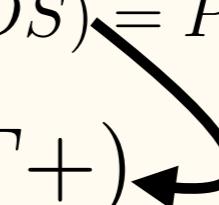
Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

goal:  $P(DS|TT+)$

general multiplication rule:

$$\begin{aligned} 1 \quad P(DS|TT+)P(TT+) &= P(DS \& TT+) \\ 2 \quad P(TT+|DS)P(DS) &= P(DS \& TT+) \end{aligned}$$


rearrange #1:  $P(DS|TT+) = \frac{P(DS \& TT+)}{P(TT+)}$

add in #2:  $P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

Law of Total Probability:

$$P(TT+) = P(DS \& TT+) + P(!DS \& TT+)$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

Law of Total Probability:

$$P(TT+) = P(DS \& TT+) + P(!DS \& TT+)$$

$$P(TT+) = P(TT+|DS)P(DS) + P(TT+|!DS)P(!DS)$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{P(TT+|DS)P(DS)}{P(TT+)}$$

Law of Total Probability:

$$P(TT+) = P(DS \& TT+) + P(!DS \& TT+)$$

$$P(TT+) = P(TT+|DS)P(DS) + P(TT+|!DS)P(!DS)$$

$$P(TT+) = (0.6)(1/1000) + (0.05)(9999/1000) = .05006$$

# Example

Down syndrome (DS) occurs in about 1 in 1000 pregnancies. A new pre-natal test, called the triple test, is safer than amniocentesis, but is more error prone.

- True positives: The probability that a DS fetus will correctly test positive is 0.60.
- False positives: The probability that a fetus without DS will incorrectly test positive is 0.05

What is the probability that a fetus with a positive test result actually has DS?

$$P(DS|TT+) = \frac{(0.6)(1/1000)}{(0.05)} = 0.012$$

Law of Total Probability:

$$P(TT+) = P(DS \& TT+) + P(!DS \& TT+)$$

$$P(TT+) = P(TT+ | DS)P(DS) + P(TT+ | !DS)P(!DS)$$

$$P(TT+) = (0.6)(1/1000) + (0.05)(9999/1000) = .05006$$

# Bayes Theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

# Bayes Theorem

**Likelihood**



$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

# Bayes Theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

**Likelihood**                            **Prior**

The diagram illustrates the components of Bayes' Theorem. The equation  $P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$  is shown. Two arrows point downwards from the text labels "Likelihood" and "Prior" to the terms  $P(y|\theta)P(\theta)$  in the numerator. The "Likelihood" arrow points to the term  $P(y|\theta)$ , and the "Prior" arrow points to the term  $P(\theta)$ .

# Bayes Theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Posterior

Likelihood

Prior

The diagram illustrates the components of Bayes' Theorem. On the left, the term 'Posterior' is written above an arrow pointing towards the equation. Above the equation, the word 'Likelihood' is centered above a downward-pointing arrow. To the right of the equation, the word 'Prior' is centered above an arrow pointing towards the denominator. The equation itself is  $P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$ .

# Bayes Theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Posterior

Likelihood

Prior

Pesky lil' guy

The diagram illustrates the components of Bayes' Theorem. The formula  $P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$  is centered. Arrows point from the words "Likelihood" and "Prior" to the terms  $P(y|\theta)$  and  $P(\theta)$  respectively in the numerator. An arrow points from the word "Posterior" to the final result  $P(\theta|y)$ .

# Bayes Theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Posterior

Likelihood

Prior

Pesky lil' guy

The diagram illustrates the components of Bayes' Theorem. At the top, the words "Likelihood" and "Prior" are positioned above their respective terms in the formula. Arrows point from "Likelihood" to the term  $P(y|\theta)$  and from "Prior" to the term  $P(\theta)$ . Below the formula, the word "Posterior" has an arrow pointing to the left side of the equation. At the bottom, the phrase "Pesky lil' guy" has an arrow pointing upwards towards the denominator  $P(y)$ .

$$P(y) = \int P(y|\theta)P(\theta)d\theta$$

# Philosophical differences

- Bayesian
  - Data are observations from sample
  - Parameters are unknown and described probabilistically
  - Data are Fixed
- Frequentist
  - Data are repeatable random sample
  - Underlying parameters constant during repeatable process
  - Parameters are Fixed

# Example

- You think there's a relationship between the percentage of visible moon and the number of murders in a city, so you collect lunar data and murder rates. You fit a linear model to understand the relationship:

$$\text{murders} = \alpha + \beta * (\% \text{ visible moon})$$

- As a frequentist you find:

- As a bayesian you find:

- What's the difference?

# Example

- You think there's a relationship between the percentage of visible moon and the number of murders in a city, so you collect lunar data and murder rates. You fit a linear model to understand the relationship:

$$\text{murders} = \alpha + \beta * (\% \text{ visible moon})$$

- As a frequentist you find:

$$\hat{\beta} = 7.5$$

*95% Confidence Interval : (3.8, 10.2)*

- As a bayesian you find:

- What's the difference?

# Example

- You think there's a relationship between the percentage of visible moon and the number of murders in a city, so you collect lunar data and murder rates. You fit a linear model to understand the relationship:

$$\text{murders} = \alpha + \beta * (\% \text{ visible moon})$$

- As a frequentist you find:

$$\hat{\beta} = 7.5$$

*95% Confidence Interval : (3.8, 10.2)*

- As a bayesian you find:

$$\text{Median } P(\beta | \text{data}) = 7.2$$

*95% Credible Interval : (3.4, 10.0)*

- What's the difference?

So when do they differ?  
(Not an exhaustive list & somewhat my opinion)

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data
  - Can favor running a bayesian or frequentist, depending on question

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data
  - Can favor running a bayesian or frequentist, depending on question
- Have excellent prior information to inform model

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data
  - Can favor running a bayesian or frequentist, depending on question
- Have excellent prior information to inform model
  - Bayesian

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data
  - Can favor running a bayesian or frequentist, depending on question
- Have excellent prior information to inform model
  - Bayesian
- Parameter interpretation

# So when do they differ? (Not an exhaustive list & somewhat my opinion)

- Not much data
  - Can favor running a bayesian or frequentist, depending on question
- Have excellent prior information to inform model
  - Bayesian
- Parameter interpretation
  - Bayesian interpretation is much more intuitive

Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad Posterior = \frac{Likelihood * Prior}{P(data)}$$

# Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad Posterior = \frac{Likelihood * Prior}{P(data)}$$

How to think  
about it:

$$P(model|data) = \frac{P(data|model)P(model)}{P(data)}$$

# Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad Posterior = \frac{Likelihood * Prior}{P(data)}$$

How to think  
about it:

$$P(model|data) = \frac{P(data|model)P(model)}{P(data)}$$

What we use:  $P(model|data) \propto P(data|model)P(model)$

# Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad Posterior = \frac{Likelihood * Prior}{P(data)}$$

How to think  
about it:

$$P(model|data) = \frac{P(data|model)P(model)}{P(data)}$$

What we use:  $P(model|data) \propto P(data|model)P(model)$

Parameters define our model

# Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad \text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{P(\text{data})}$$

How to think  
about it:

$$P(\text{model}|data) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$$

What we use:  $P(\text{model}|data) \propto P(\text{data}|\text{model})P(\text{model})$

Parameters define our model

$$P(\theta, \phi|y) = \frac{P(y|\theta, \phi)P(\theta)P(\phi)}{\int \int P(y|\theta, \phi)P(\theta)P(\phi)d\theta d\phi}$$

# Okay, let's use Bayes' theorem

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad \text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{P(\text{data})}$$

How to think  
about it:

$$P(\text{model}|data) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$$

What we use:  $P(\text{model}|data) \propto P(\text{data}|\text{model})P(\text{model})$

Parameters define our model

**Likelihood** →  $P(\theta, \phi|y) = \frac{P(y|\theta, \phi)P(\theta)P(\phi)}{\int \int P(y|\theta, \phi)P(\theta)P(\phi)d\theta d\phi}$

↑  
**Priors**

# MCMC

- What we know:

# MCMC

- What we know:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

# MCMC

- What we know:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

$$\int P(\theta|y)d\theta = 1$$

# MCMC

- What we know:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

$$\int P(\theta|y)d\theta = 1$$

- MCMC uses simulation to numerically approximate the posterior

# MCMC

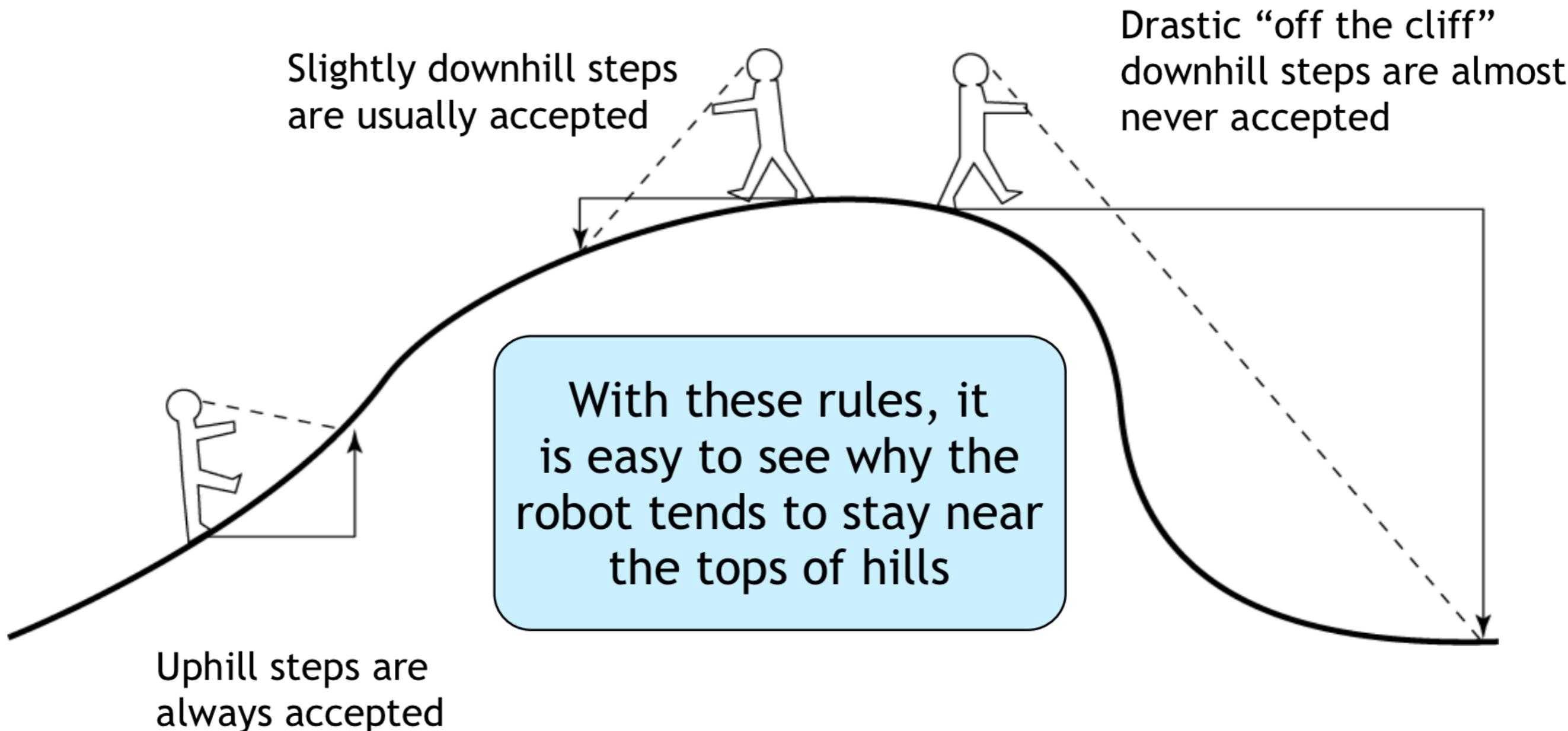
- What we know:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

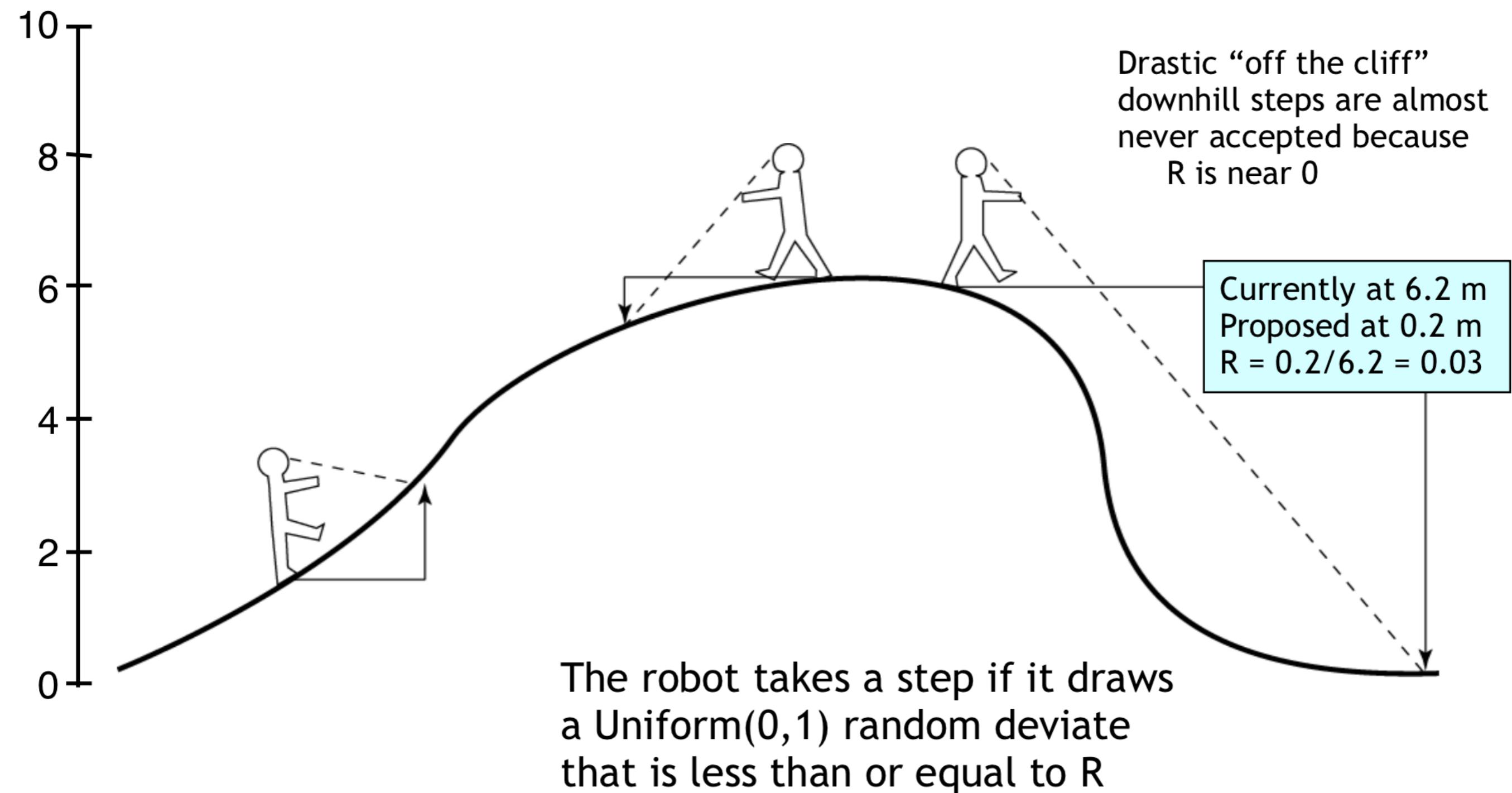
$$\int P(\theta|y)d\theta = 1$$

- MCMC uses simulation to numerically approximate the posterior
- Many methods to do this, I'm going to only talk about the metropolis-hastings method

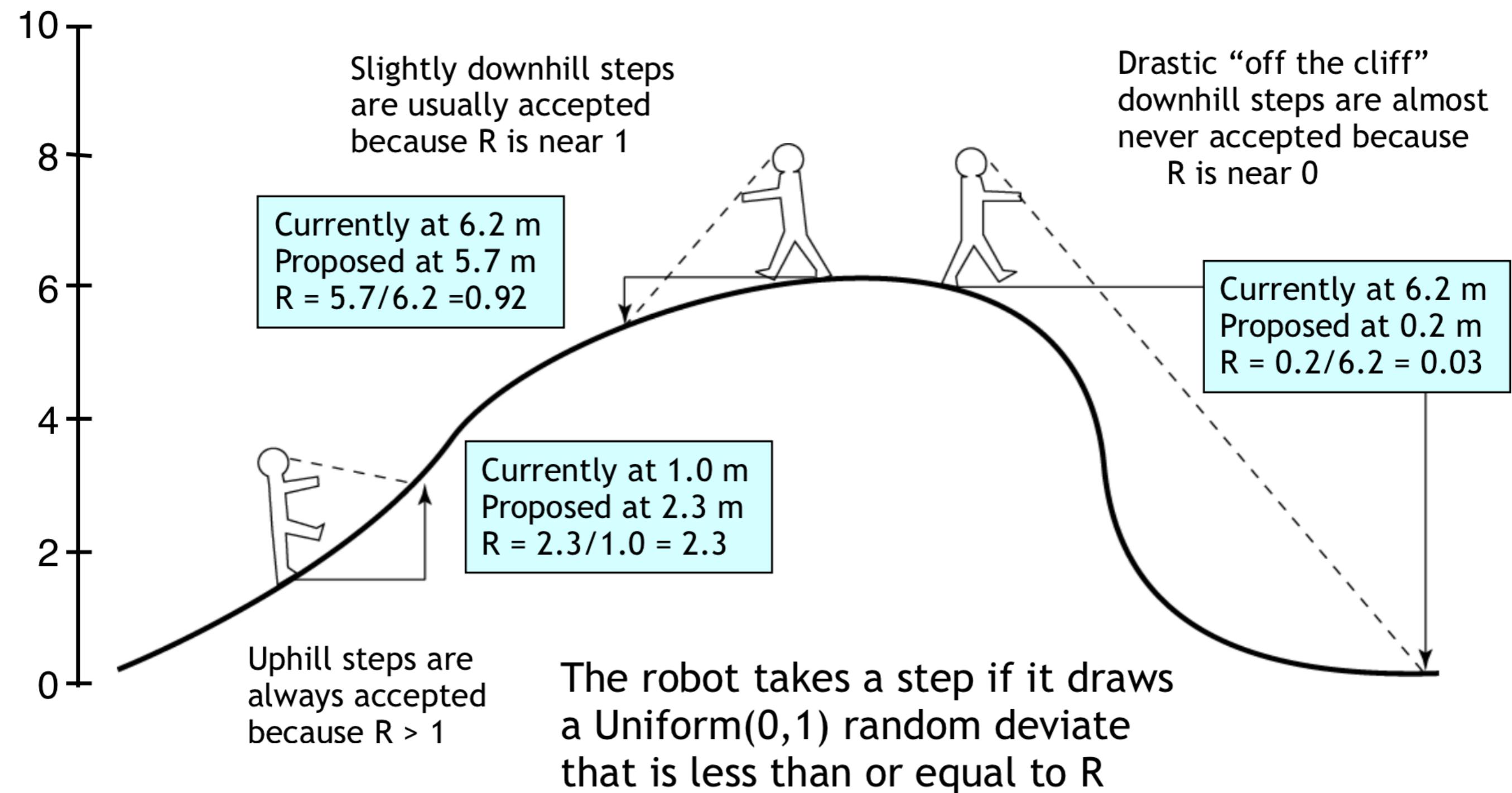
# MCMC robot's rules



# (Actual) MCMC robot rules



# (Actual) MCMC robot rules



# Metropolis-Hastings MCMC

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

$$\alpha_t = \min(1, \rho)$$

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

$$\alpha_t = \min(1, \rho)$$

$$\rho = \frac{P(\theta_{Proposed} | y)}{P(\theta_{Current} | y)} * \frac{G(\theta_{Current} | \theta_{Proposed})}{G(\theta_{Proposed} | \theta_{Current})}$$

Proposal Distribution = G(X|Y)

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

$$\alpha_t = \min(1, \rho)$$

$$\rho = \frac{\frac{P(y|\theta_{Proposed})P(\theta_{Proposed})}{P(y)}}{\frac{P(y|\theta_{Current})P(\theta_{Current})}{P(y)}} * \frac{G(\theta_{Current}|\theta_{Proposed})}{G(\theta_{Proposed}|\theta_{Current})}$$

Proposal Distribution = G(X|Y)

# Metropolis-Hastings MCMC

1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

$$\alpha_t = \min(1, \rho)$$

$$\rho = \frac{P(y|\theta_{Proposed})P(\theta_{Proposed})}{P(y|\theta_{Current})P(\theta_{Current})} * \frac{G(\theta_{Current}|\theta_{Proposed})}{G(\theta_{Proposed}|\theta_{Current})}$$

Proposal Distribution = G(X|Y)

# Metropolis-Hastings MCMC

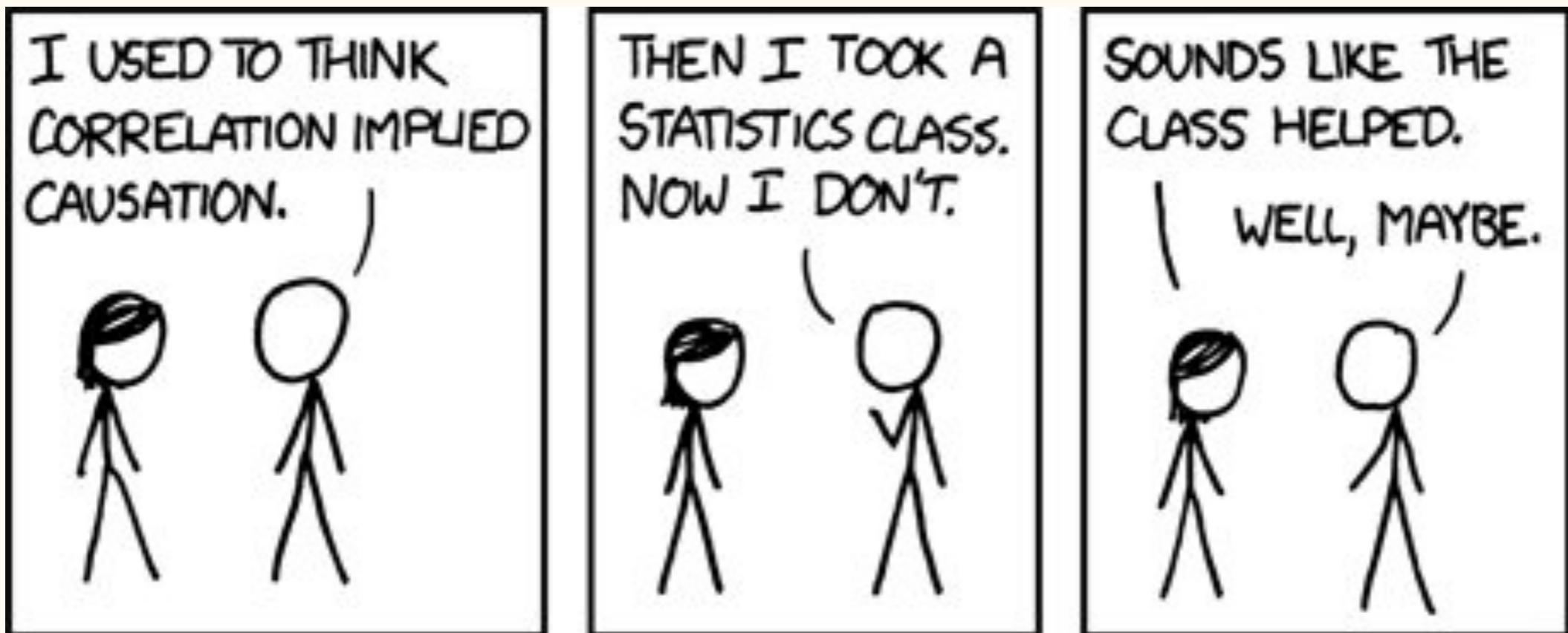
1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

$$\alpha_t = \min(1, \rho)$$

$$\rho = \frac{P(y|\theta_{Proposed})P(\theta_{Proposed})}{P(y|\theta_{Current})P(\theta_{Current})} * \frac{\cancel{G(\theta_{Current}|\theta_{Proposed})}}{\cancel{G(\theta_{Proposed}|\theta_{Current})}}$$

Proposal Distribution = G(X|Y)

# Conclusions



# Examples in R



# Sources

- <http://www-users.york.ac.uk/~pml1/bayes/cartoons/welcome.htm>
- <http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf>
- Paul Lewis Molecular Evolution
- International Clinics on Infectious Disease Dynamics (ICI3D) Program: <https://www.ici3d.org>