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: $P(DS|TT+) = \dfrac{P(DS\&TT+)}{P(TT+)}$

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:

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

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

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

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 $\boxed{0.60.}$
- 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 $\boxed{\text{1 in 1000}}$ 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 $\boxed{0.60}$.
- 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 $\boxed{\text{1 in 1000}}$ 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 $\boxed{0.60}$.
- 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 $\boxed{\text{1 in 1000}}$ 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 $\boxed{0.60}$.
- 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

Likelihood **Prior**

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

Bayes Theorem

Likelihood **Prior**

Posterior

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

Bayes Theorem

Likelihood **Prior**

Posterior

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

Pesky lil' guy

Bayes Theorem

Likelihood

Prior

Posterior

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

Pesky lil' guy

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

Philosophical differences

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

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:

  murders = alpha + beta*(% 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:

murders = alpha + beta*(% 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:

murders = alpha + beta*(% visible moon)

- As a frequentist you find:

$$\hat{\beta} = 7.5$$
$$95\% Confidence Interval : (3.8, 10.2)$$

- As a bayesian you find:

$$Median\ P(\beta|data) = 7.2$$
$$95\% Credible Interval : (3.4, 10.0)$$

- 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

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

$$Posterior = \frac{Likelihood * Prior}{P(data)}$$

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

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

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

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

Okay, let's use Bayes' theorem

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

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

Priors

Likelihood

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

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:

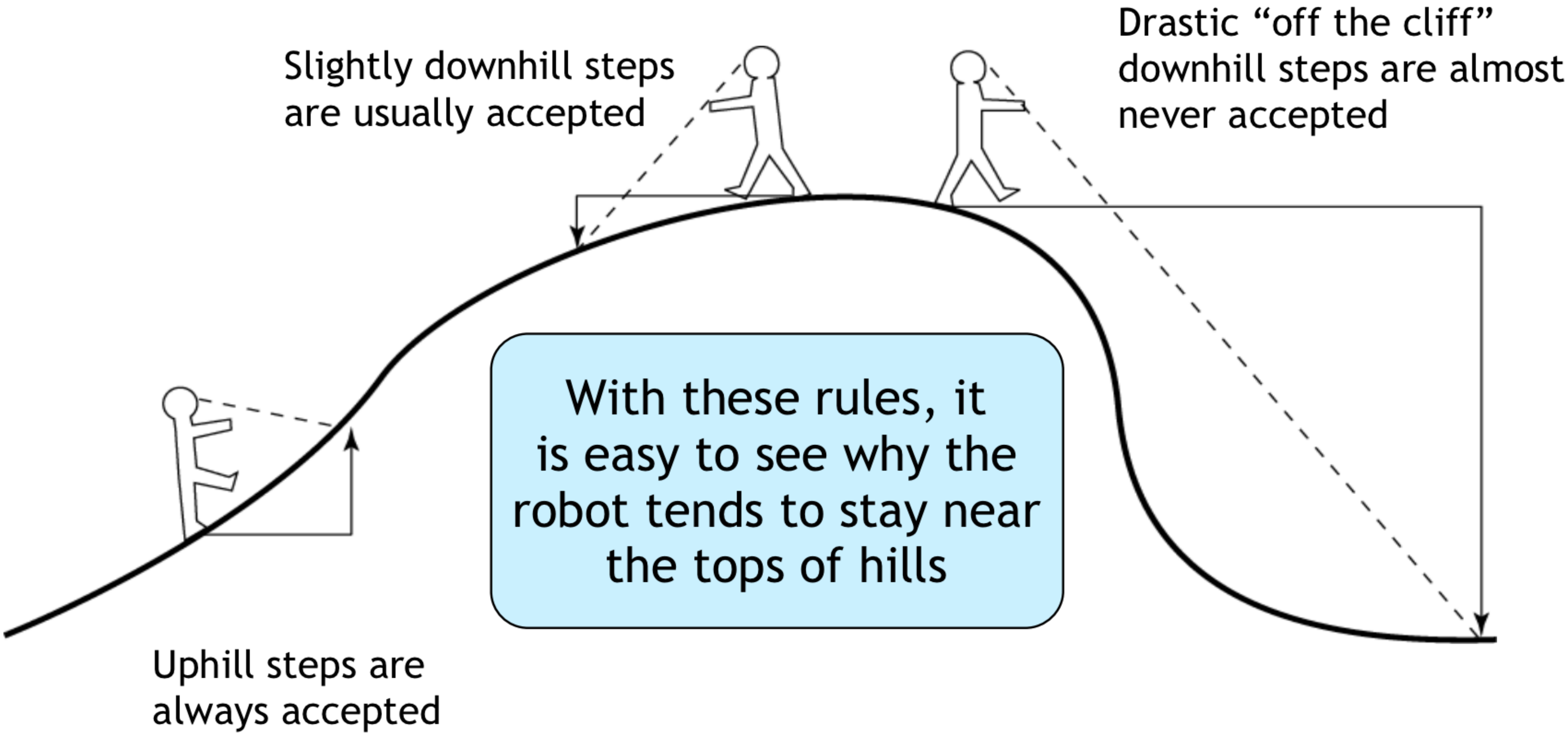
$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

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

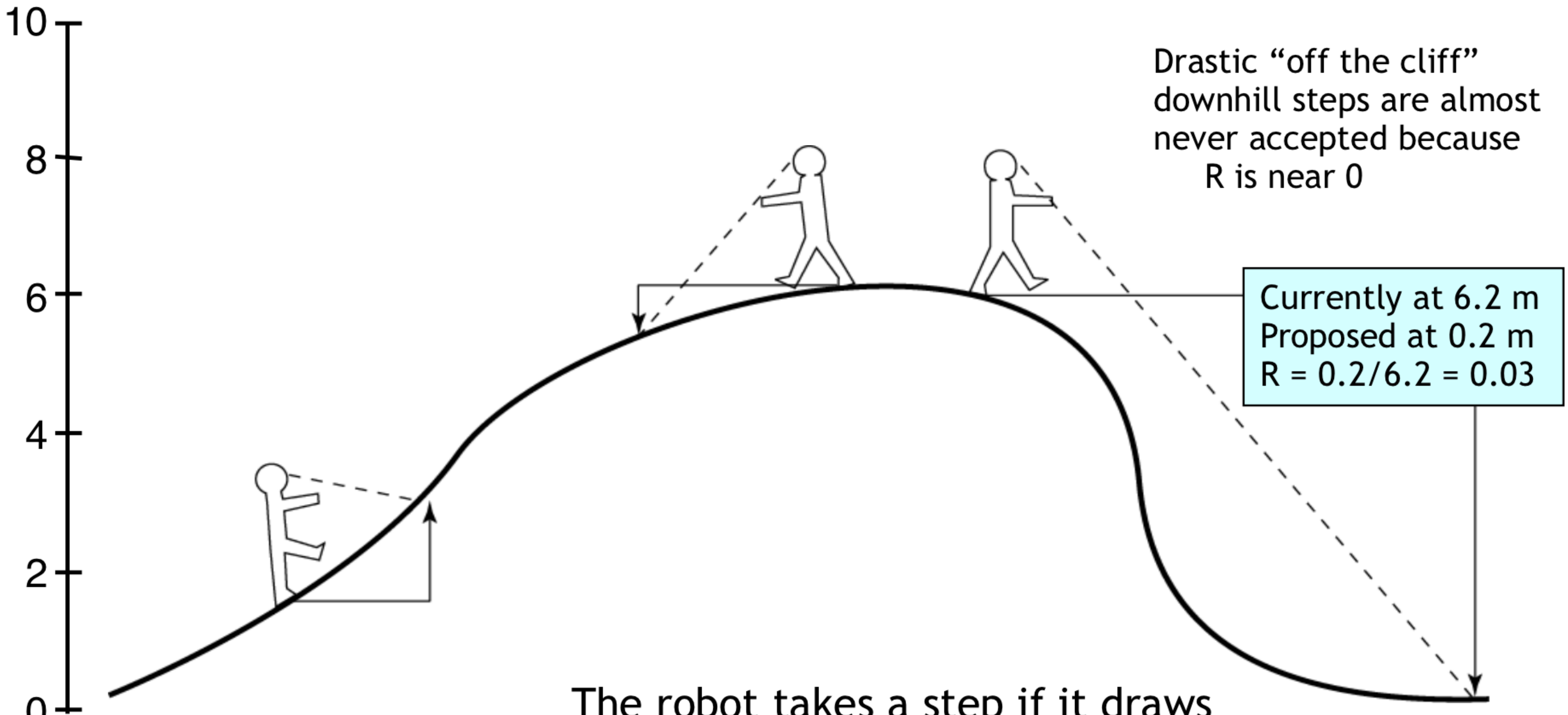
- MCMC uses simulation to numerically approximate the posterior

- Many methods to do this, I'm going to only talk about the metropolis-hastings method

MCMC robot's rules

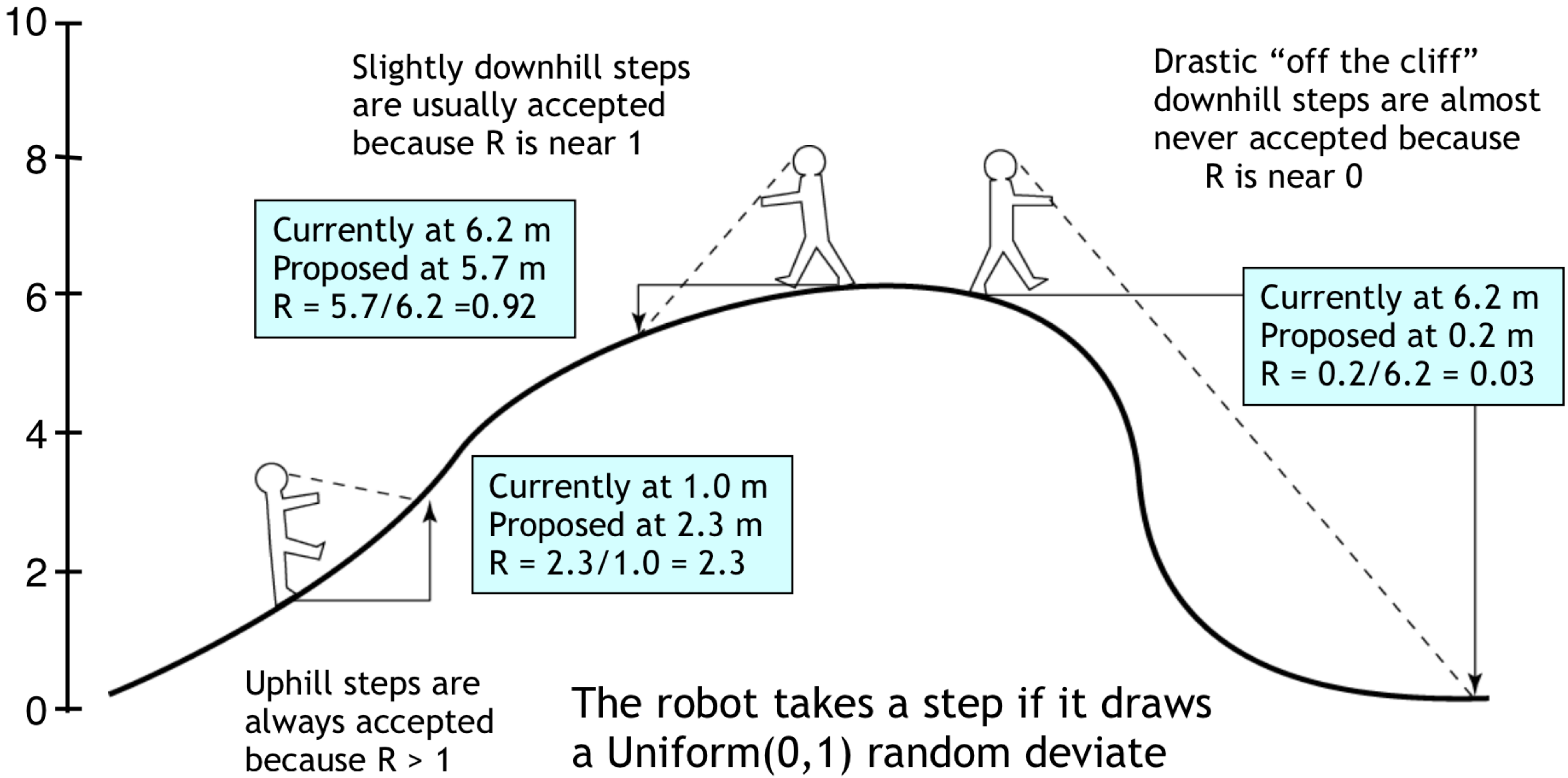


(Actual) MCMC robot rules



The robot takes a step if it draws a Uniform(0,1) random deviate that is less than or equal to R

(Actual) MCMC robot rules



The robot takes a step if it draws a Uniform(0,1) random deviate that is less than or equal to R

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

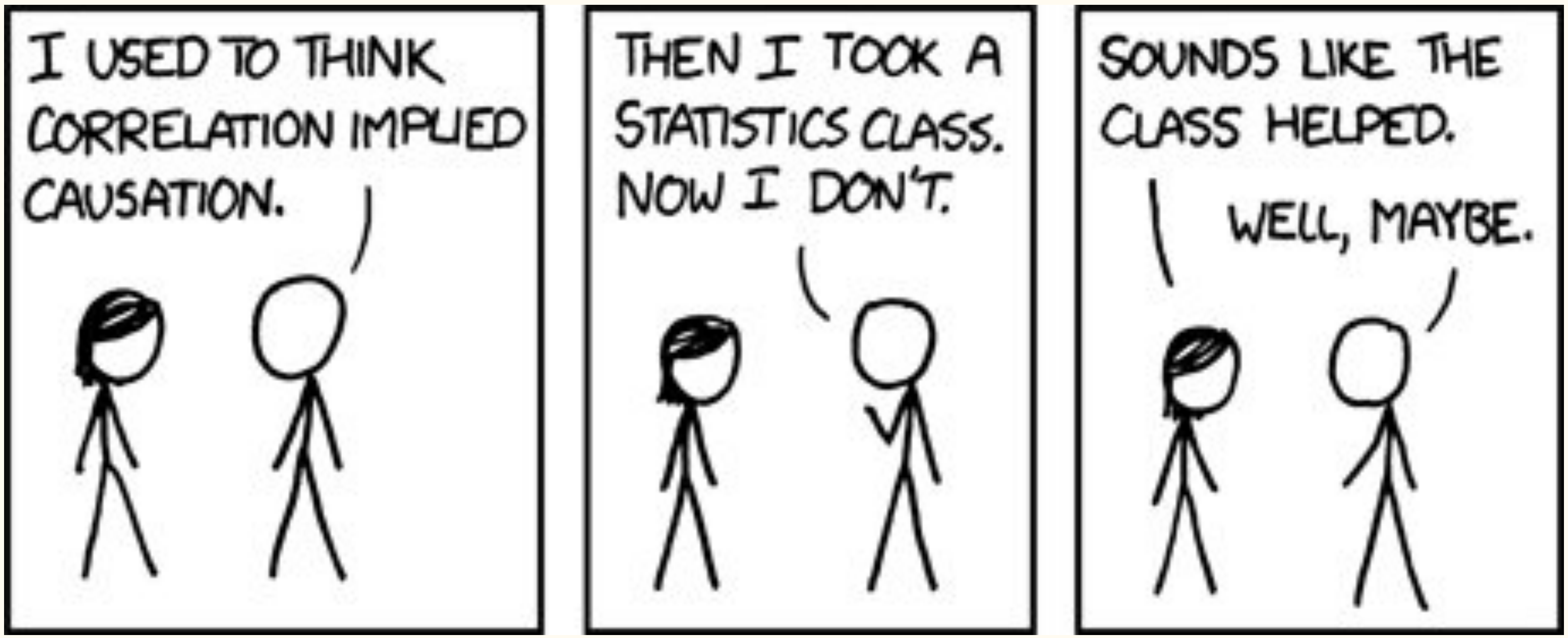
1. Start at random (reasonable) parameter value(s)
2. Propose new random parameter value(s)
3. Calculate proposal probability
4. Accept (change states) or reject (remain in same state) proposed value(s) with probability of alpha
5. Repeat 2-5 comparing proposal to most recent state

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

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

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