Tracy A. Heath

@trayc7
http://phyloworks.org

Introduction to some important Bayesian concepts

Bayes Rule

posterior probability of H_θ given the data

likelihood of the data given H_θ

prior probability of H_θ befor seeing any of the data

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\,\Pr(\theta)}{\sum_\theta \Pr(D \mid \theta)\,\Pr(\theta)}$$

marginal probability of the data

Bayesian Inference

Estimate the probability of a hypothesis (model) conditional on observed data

The probability represents a **researcher's degree of belief**

Bayes Rule (also called Bayes Theorem) specifies the conditional probability of the hypothesis given the data

Bayes Rule

the posterior probability of a discrete parameter δ conditional on the data *D* is

$$\Pr(\delta \mid D) = \frac{\Pr(D \mid \delta)\Pr(\delta)}{\sum_{\delta}\Pr(D \mid \delta)\Pr(\delta)}$$

the likelihood marginalized over all possible values of δ

Bayes Rule

the posterior probability of a discrete parameter θ conditional on the data D is

$$f(\theta \mid D) = \frac{f(D \mid \theta) f(\theta)}{\int_\theta f(D \mid \theta) f(\theta)}$$

the likelihood marginalized over all possible values of θ

Priors

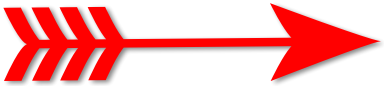
Prior distributions are an important part of Bayesian statistics

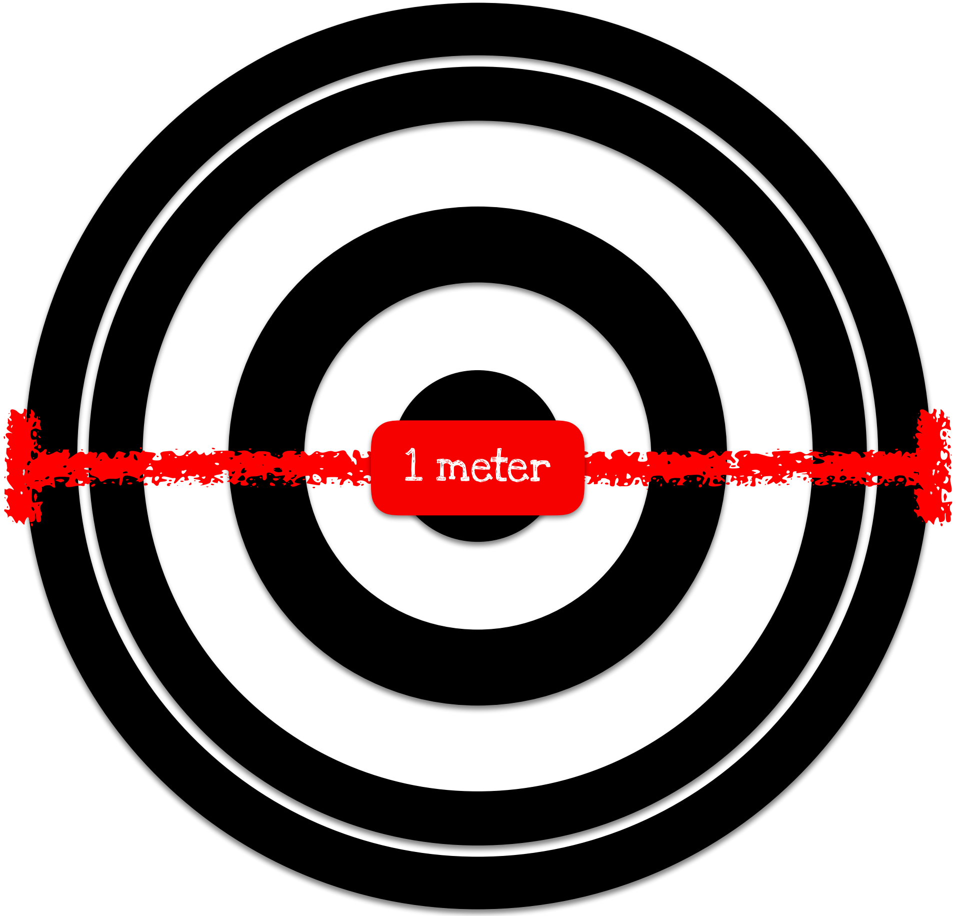
The distribution of θ before any data are collected is the prior

$$f(\theta)$$

The prior describes your uncertainty in the parameters of your model

Priors: Archery



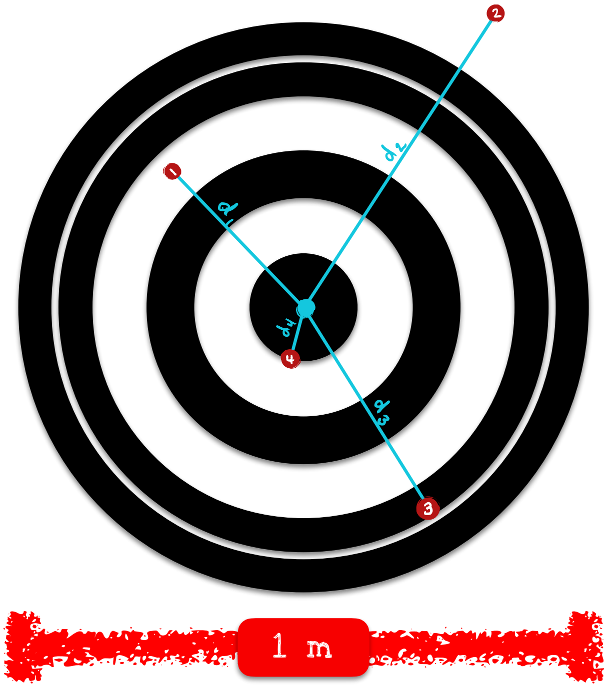


(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

In this example we want to assess an archer's accuracy at hitting the bullseye

To quantify this, we will measure the distance *d* from the center of the target (in centimeters)



* in this example, d is an absolute value

(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

Consider your prior knowledge about my archery abilities and draw a curve representing your view of the chances of my arrow landing a distance *d* centimeters from the bullseye

When formalizing your prior belief, also consider what you know about *d*

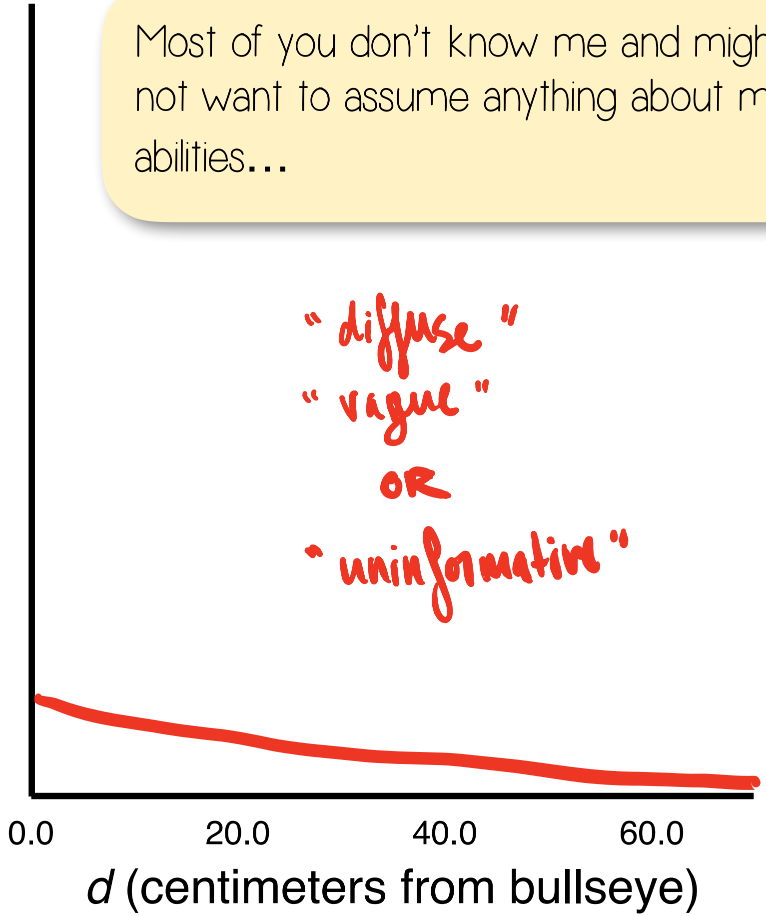
0.0 20.0 40.0 60.0

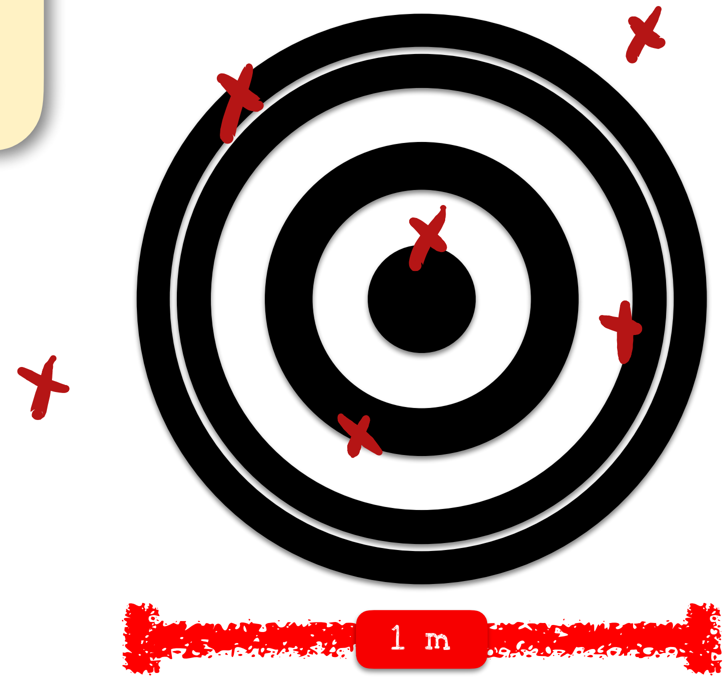
d (centimeters from bullseye)

(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

Most of you don't know me and might not want to assume anything about my abilities…

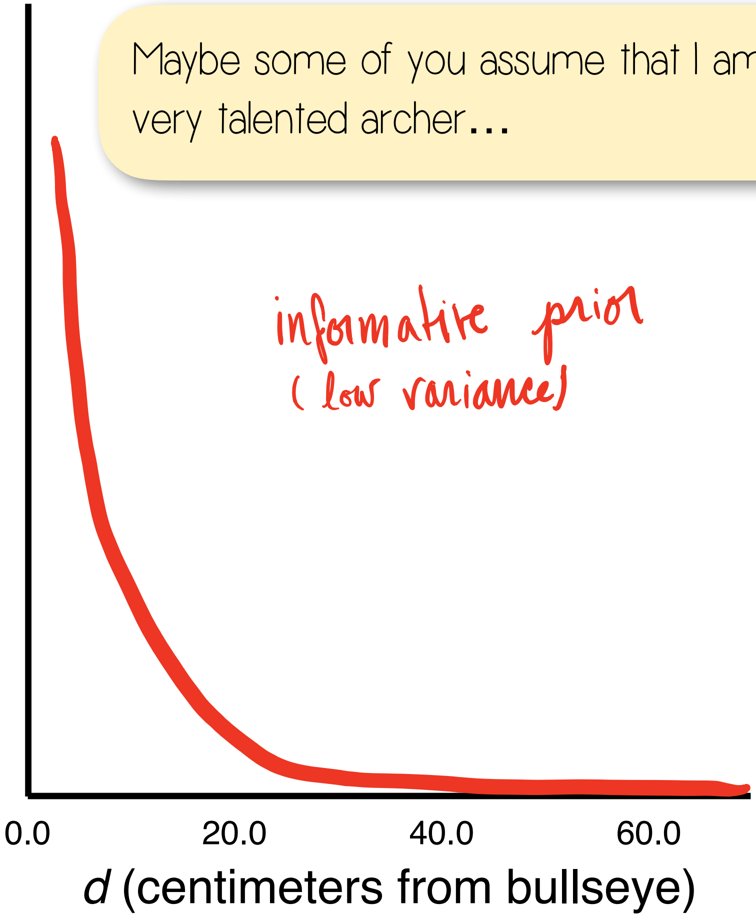




(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

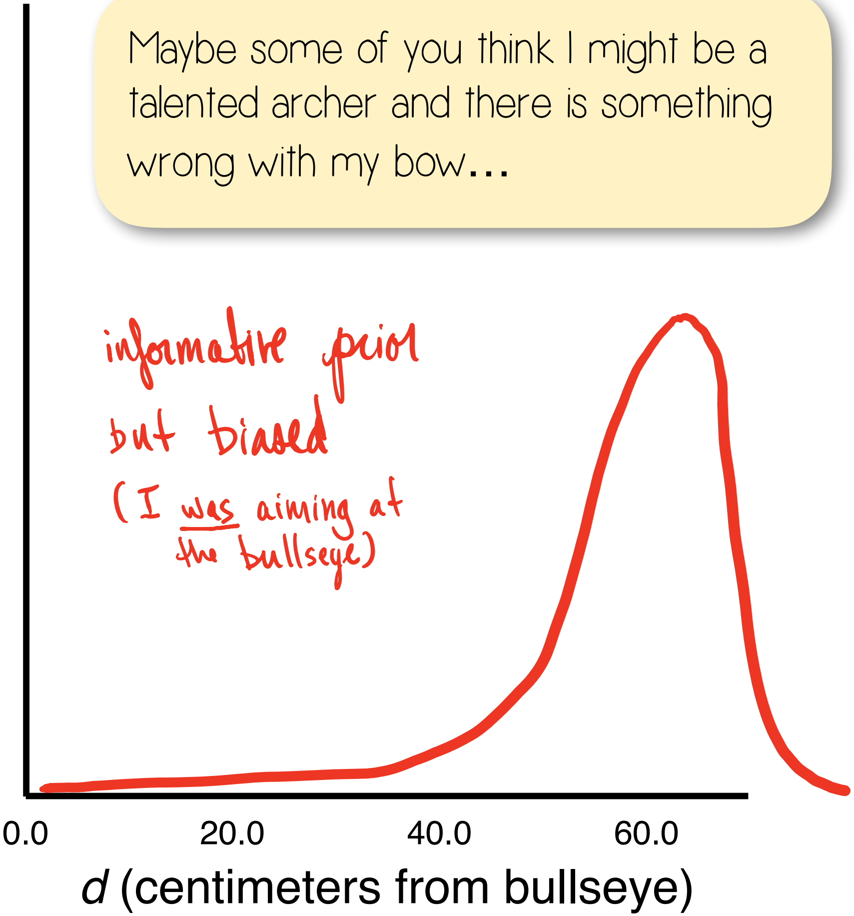
Maybe some of you assume that I am a very talented archer...



(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

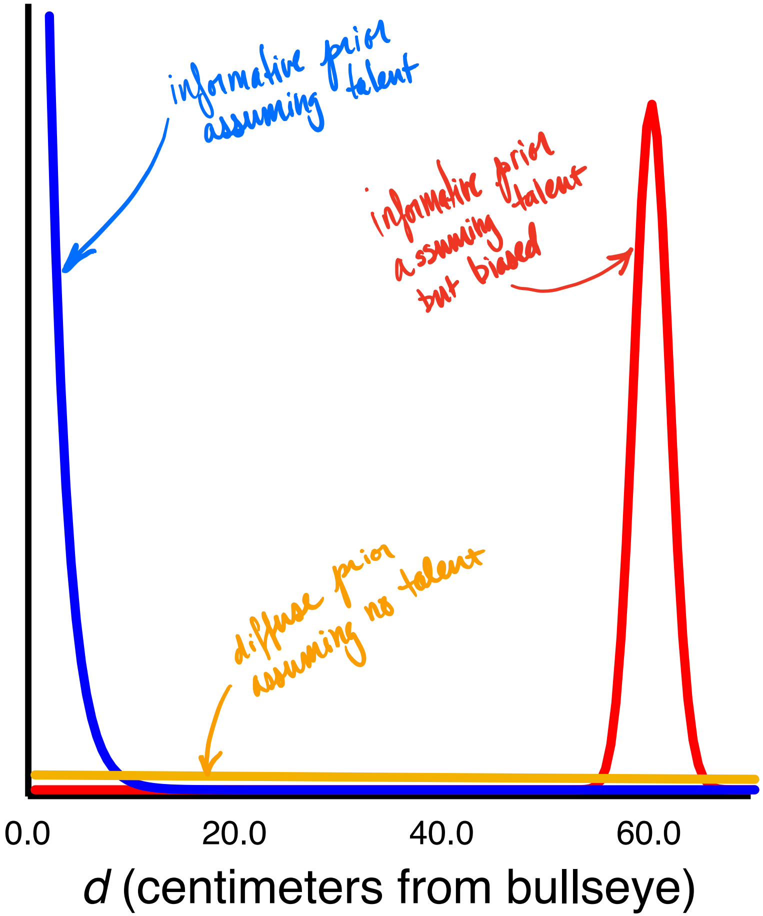
Priors: Archery

Maybe some of you think I might be a talented archer and there is something wrong with my bow…



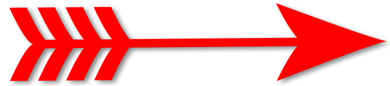
(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

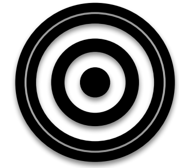
Priors: Archery



(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

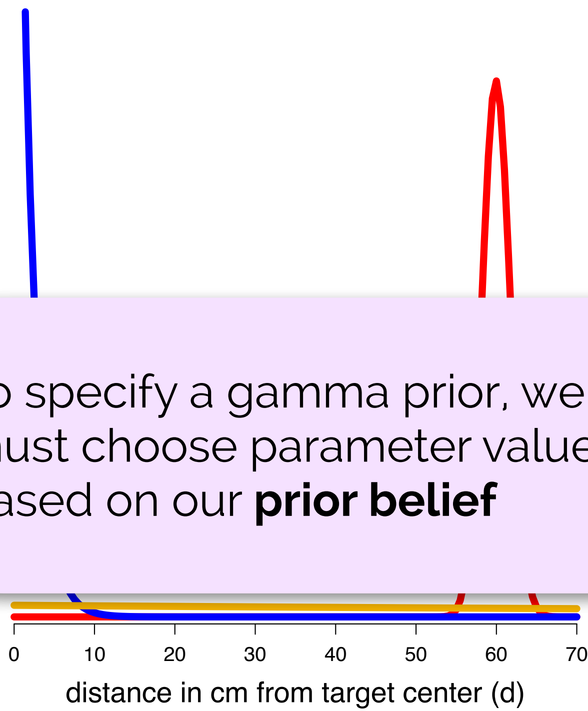




Each of these prior densities can be defined using a gamma distribution.

$$d \sim \text{Gamma}(\alpha, \beta)$$

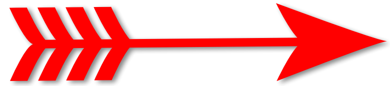
$$f(d \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} d^{\alpha-1} e^{-\frac{d}{\beta}}$$



To specify a gamma prior, we must choose parameter values based on our **prior belief**

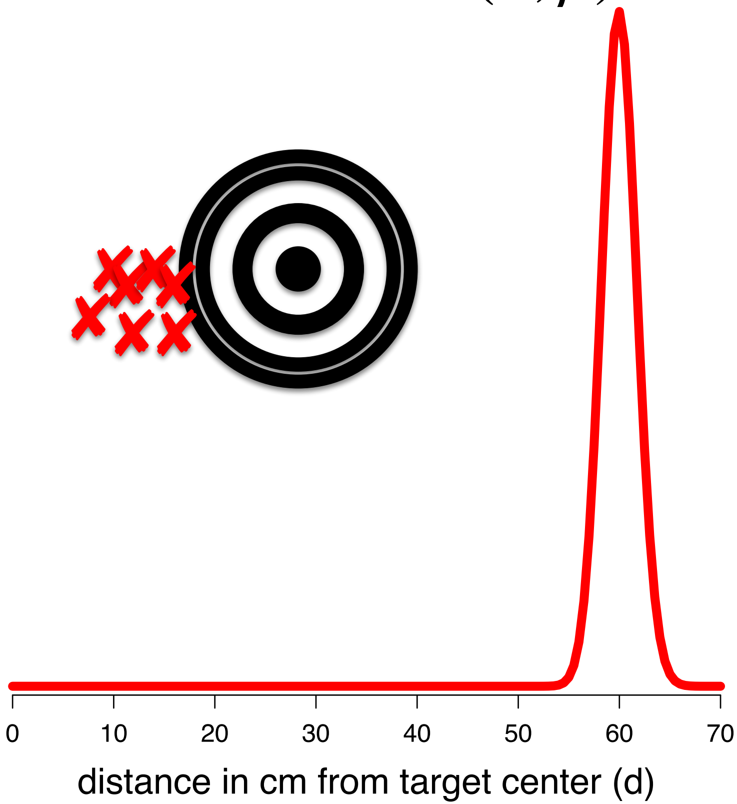
(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery



Let's assume that I will consistently miss the target

$d \sim \text{Gamma}(\alpha, \beta)$



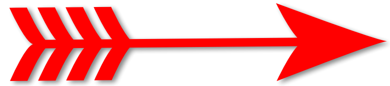
This is a gamma distribution with a mean (*m*) of 60 and a variance (*v*) of 3

mean = accuracy

variance = precision

(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

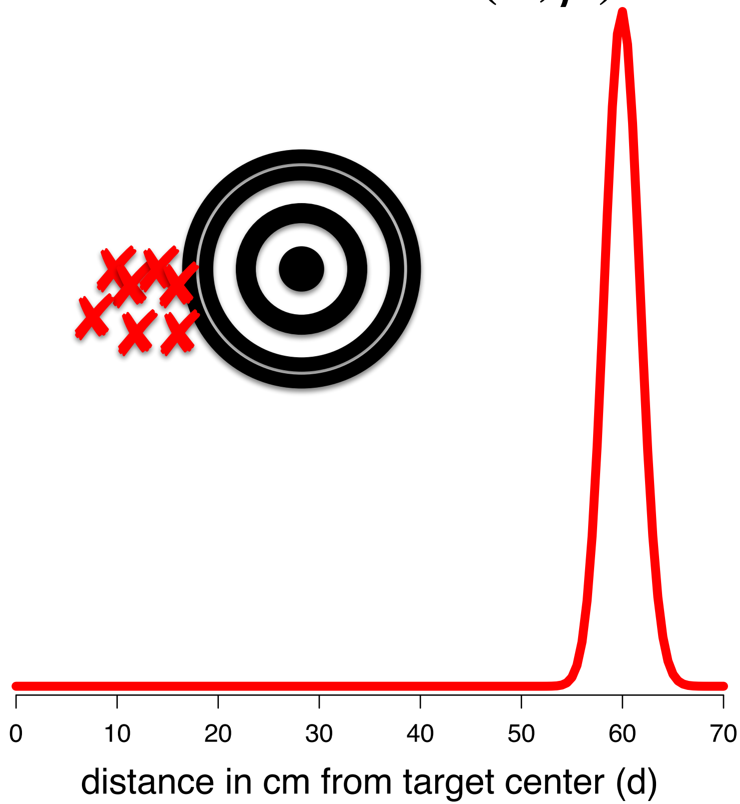


If we have prior knowledge of the mean and variance of the gamma distribution, we can compute the shape and rate parameters

$$m = \frac{\alpha}{\beta},\ \alpha = \frac{m^2}{v}$$

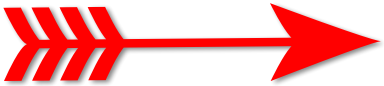
$$v = \frac{\alpha}{\beta^2},\ \beta = \frac{m}{v}$$

$$d \sim \text{Gamma}(\alpha, \beta)$$



(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

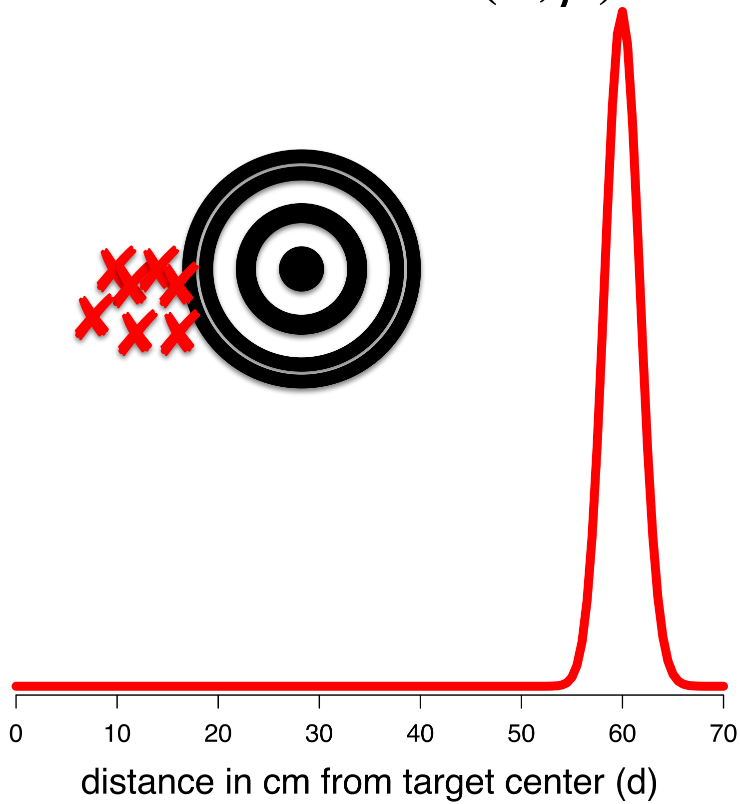


$$m = 60, \quad v = 3$$

$$\alpha = \frac{60^2}{3} = 1200$$

$$\beta = \frac{60}{3} = 20$$

$d \sim \text{Gamma}(\alpha, \beta)$

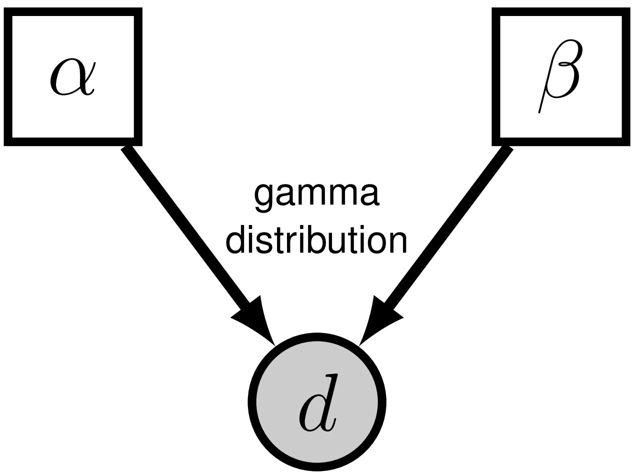


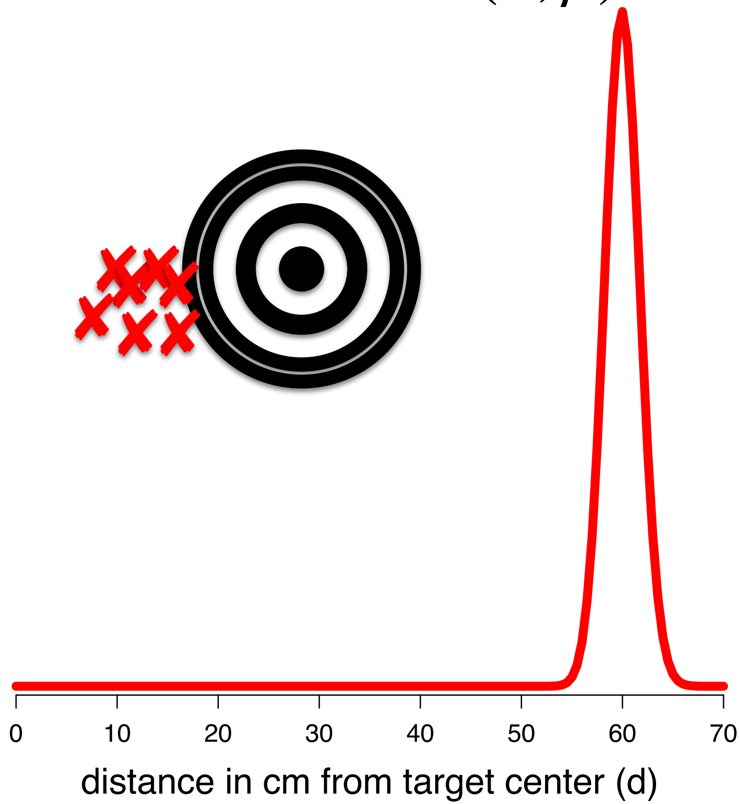
(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

Another way of expressing this distribution is with a probabilistic graphical model

$$d \sim \text{Gamma}(\alpha, \beta)$$



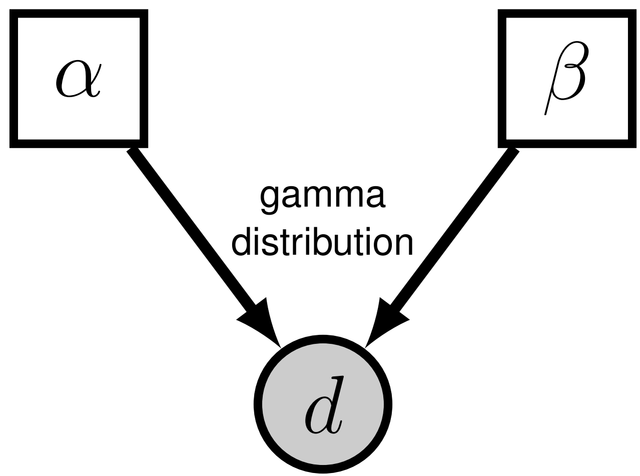


(Based on slides by Paul Lewis https://molevol.mbl.edu/index.php/Paul_Lewis)

Priors: Archery

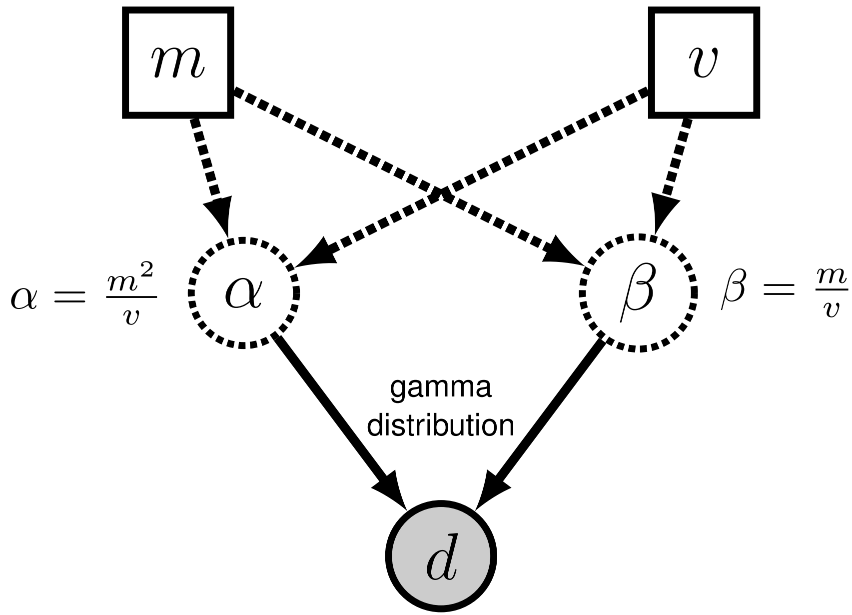
$$d \sim \text{Gamma}(\alpha, \beta)$$

This shows that our observed datum (d = a single observed shot) is conditionally dependent on the shape (α) and rate (β) of the gamma distribution



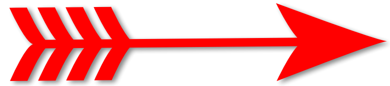
Priors: Archery

We can parameterize the model using the mean (*m*) and variance (*v*), where α and β are computed using *m* and *v*



We may have more intuition about the mean and variance than we do about the shape and rate.

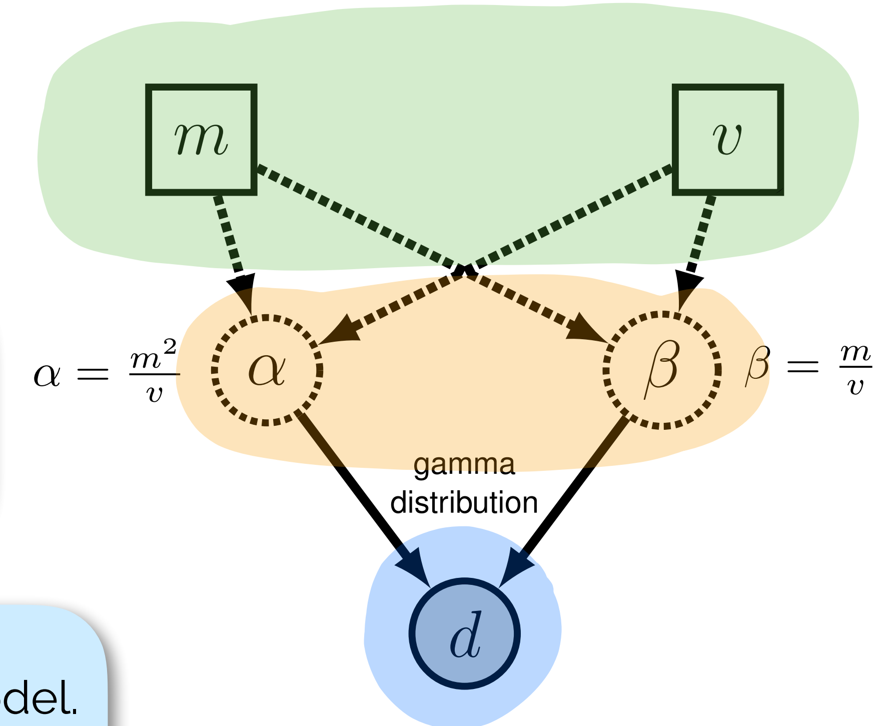
Priors: Archery



Constant nodes represent a fixed value that is asserted or known

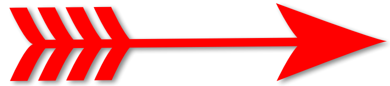
Deterministic nodes represent unknown random variable whose values are determined by other nodes

Stochastic nodes are random variables generated by the model. If we observe the value of a stochastic node, we fix it to that value

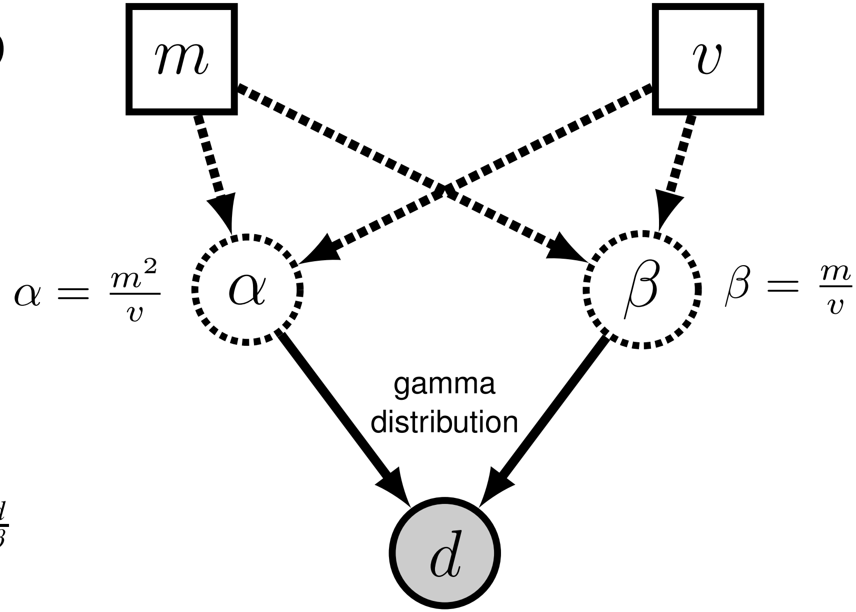


This graphical model has 3 types of nodes

Priors: Archery



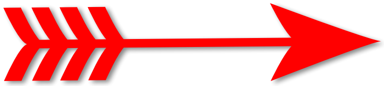
If we set *m* and *v* to values corresponding to our assumed model, then we can calculate the likelihood of any observed shot



$$f(d \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} d^{\alpha-1} e^{-\frac{d}{\beta}}$$

$$f(d = 39.76 \mid \alpha = 1200, \beta = 20) = 7.89916e - 40$$

Priors: Archery



What if we do not know *m* and *v*?

We can use maximum likelihood or Bayesian methods to estimate their values

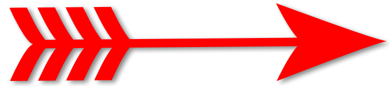
Maximum likelihood methods require us to find the values of *m* and *v* that maximize

$$f(d \mid m, v)$$

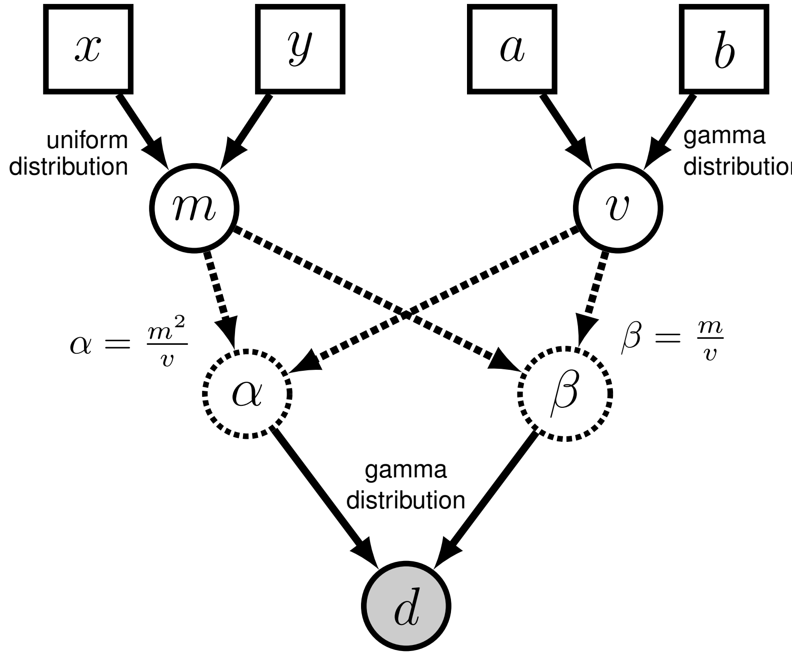
Bayesian methods use prior distributions to describe our uncertainty in *m* and *v* and estimate

$$f(m, v \mid d)$$

Priors: Archery



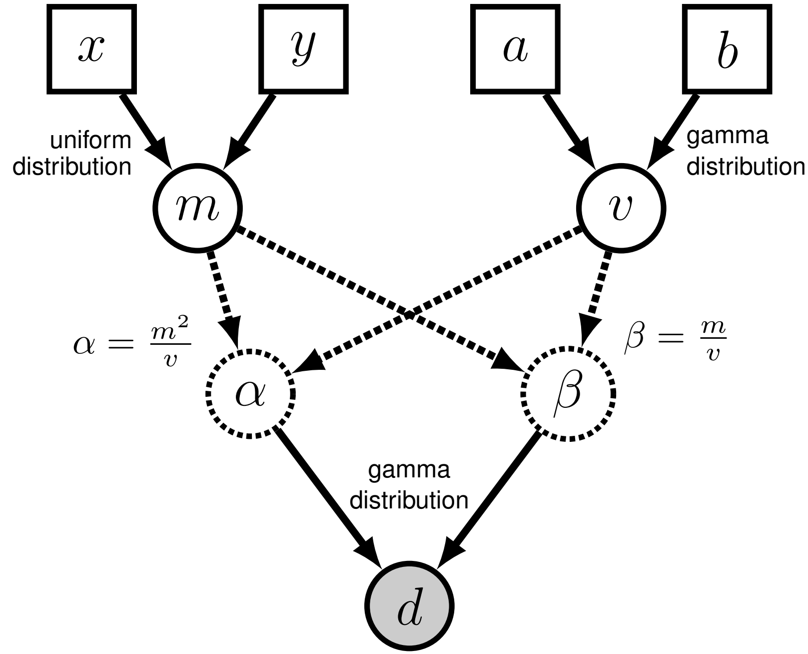
We must define prior distributions for *m* and *v* to account for uncertainty and estimate the posterior densities of those parameters



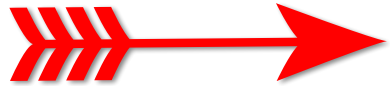
Priors: Archery

Now *x* and *y* are the parameters of the uniform prior on *m*

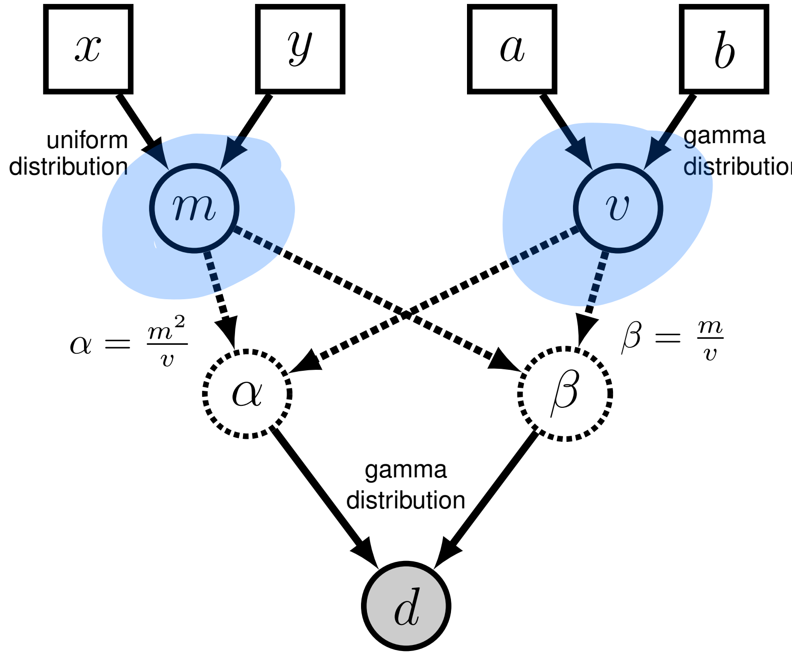
And *a* and *b* are the shoe and rate parameters of the gamma prior on *v*



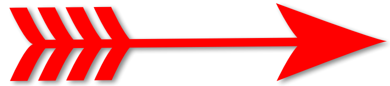
Priors: Archery



Stochastic nodes that are not observed are random variables that are unknown and estimated

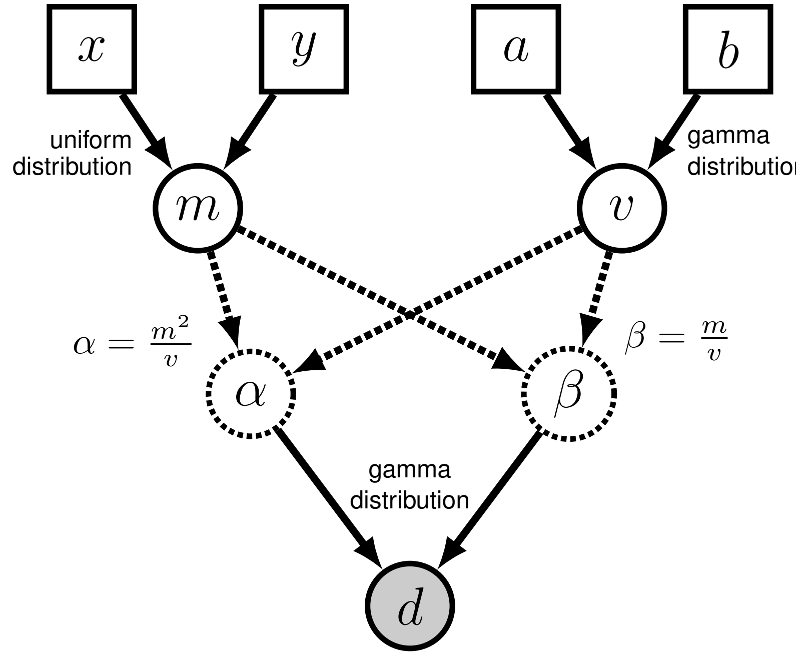


Priors: Archery



The values we choose for the parameters of these prior distributions should reflect our prior knowledge

If we observed a previous shot at 39.76 cm, the we can use this to parameterize our priors for analysis of future observations



Priors: Archery

$m \sim \text{Uniform}(x, y)$

$x = 10$

$y = 50$

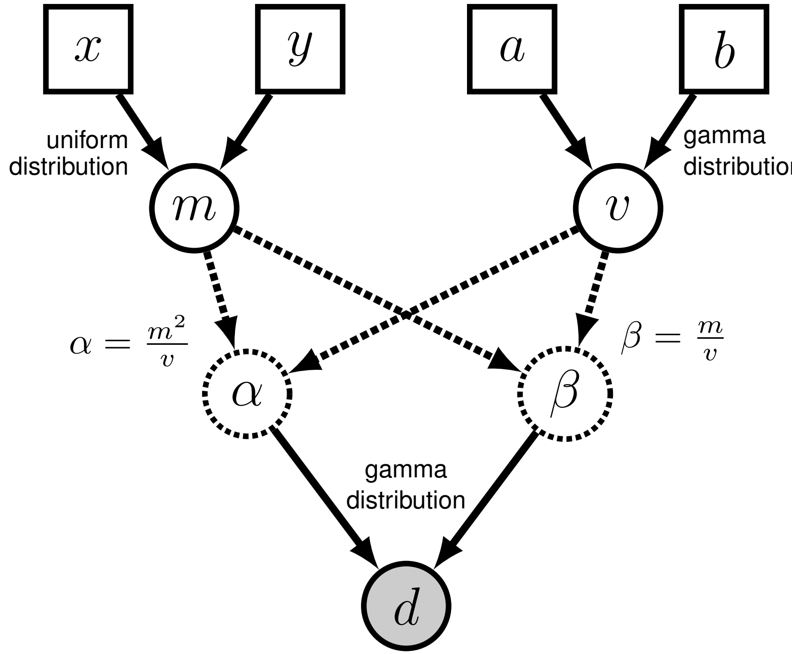
$\mathbb{E}(m) = 30$

$v \sim \text{Gamma}(a, b)$

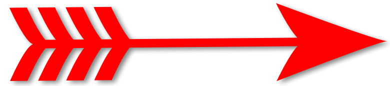
$a = 20$

$b = 2$

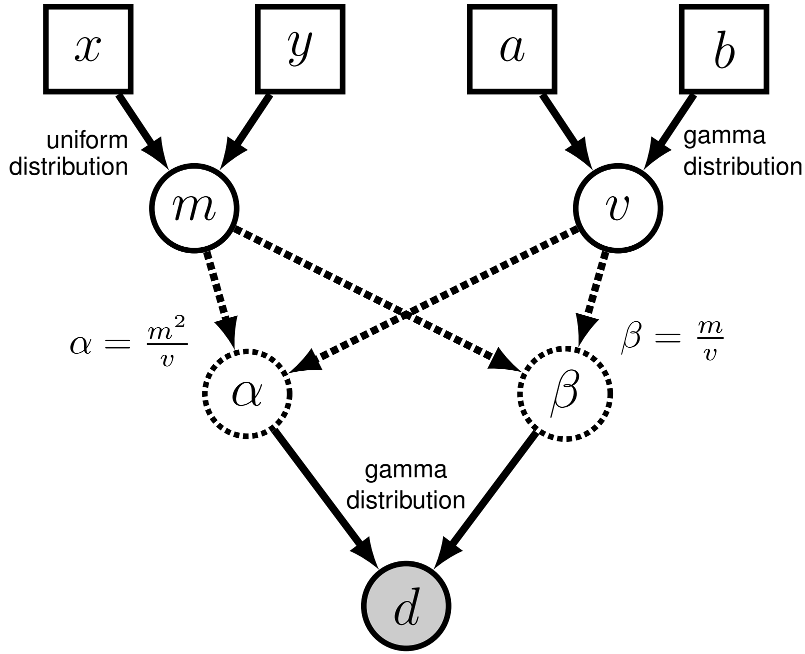
$\mathbb{E}(v) = 10$



Priors: Archery



Now that we have a defined model, how do we estimate the posterior probability density?

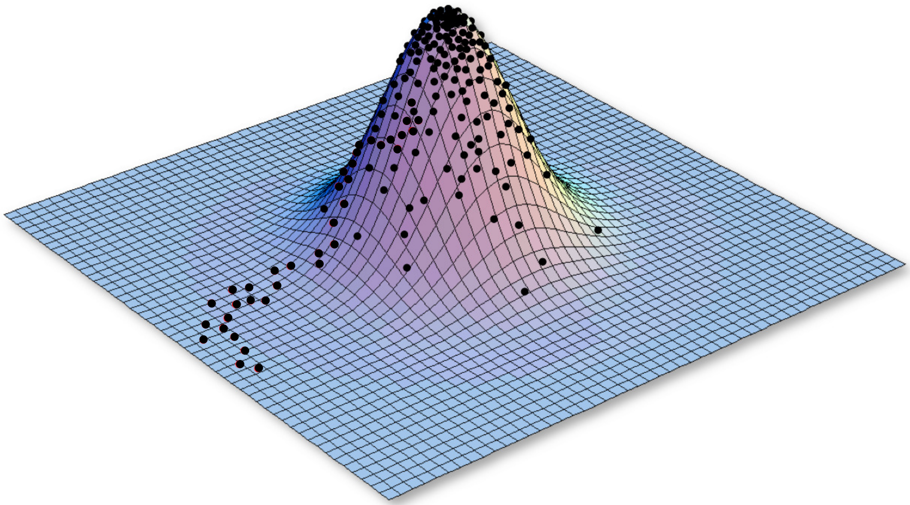


$$m \sim \text{Uniform}(x, y)$$
$$v \sim \text{Gamma}(a, b)$$
$$d \sim \text{Gamma}(\alpha, \beta)$$

$$f(m, v \mid d, a, b, x, y) \propto f(d \mid , \alpha = \frac{m^2}{v}, \beta = \frac{m}{v}) f(m \mid x, y) f(v \mid a, b)$$

Markov Chain Monte Carlo

An algorithm for approximating the posterior distribution



Metropolis, et al. 1953. Equations of state calculations by fast computing machines. J. Chem. Phys.

Hastings. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika.

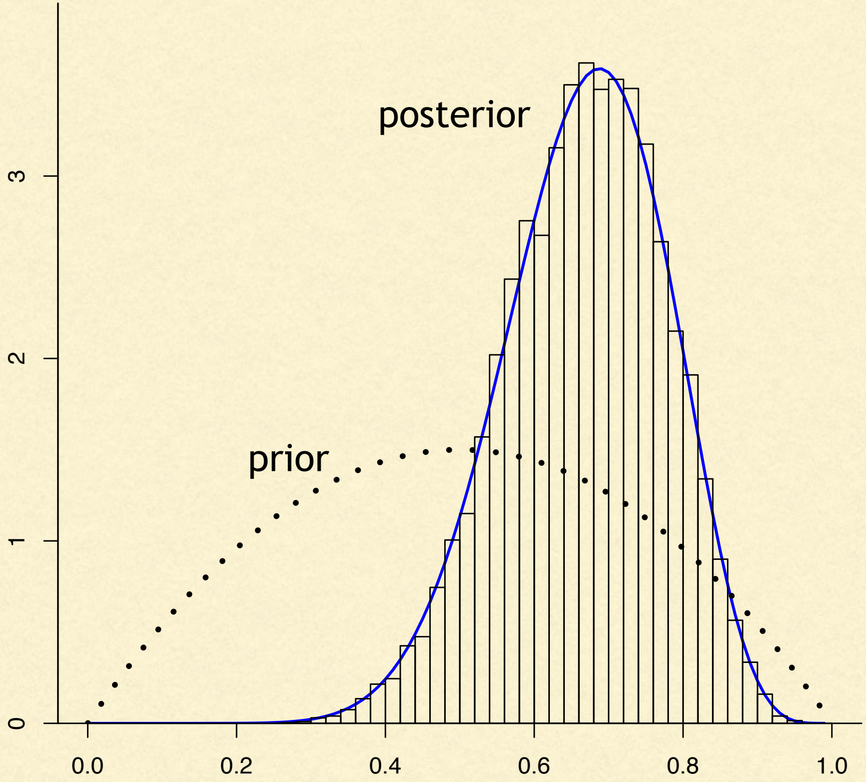
Markov Chain Monte Carlo

More on MCMC from Paul Lewis and his lecture on Bayesian phylogenetics

Slides source: https://molevol.mbl.edu/index.php/Paul_Lewis

Also see: https://www.youtube.com/watch?v=4PWlnNsfz90

Markov chain Monte Carlo (MCMC)

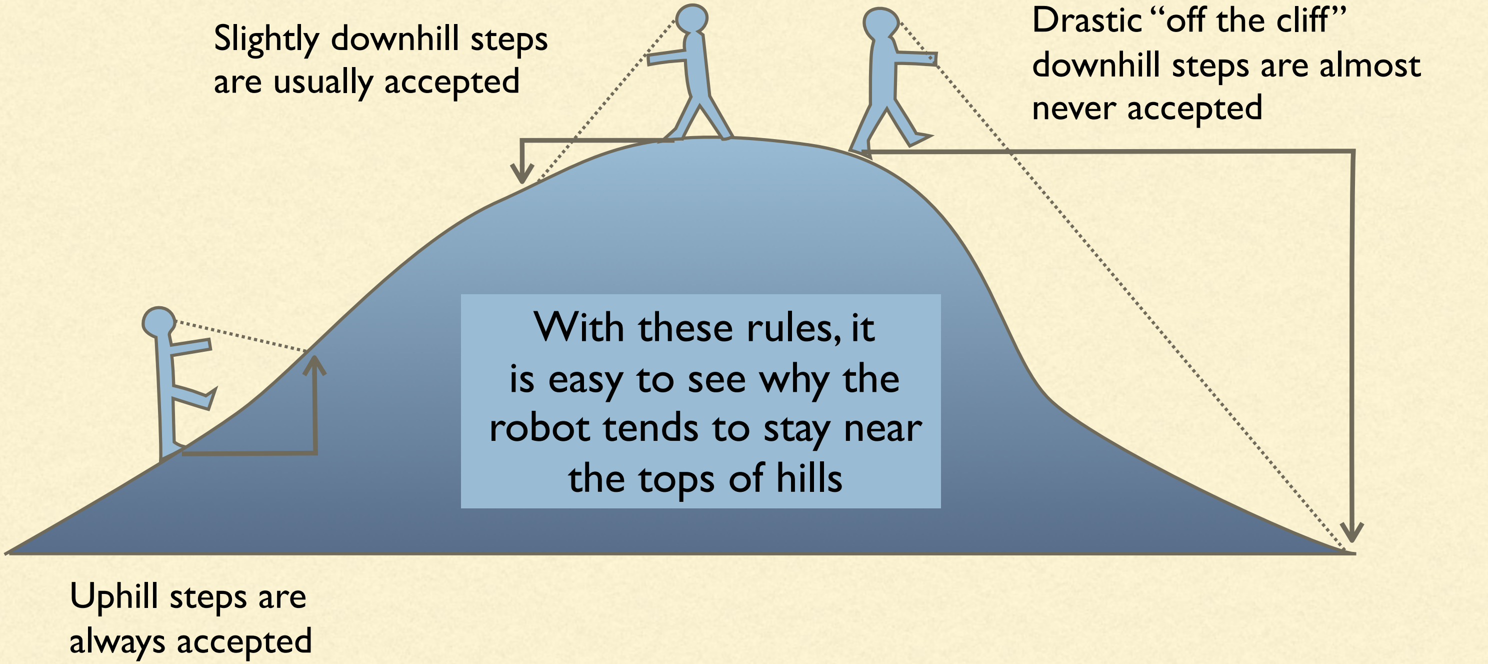


For more complex problems, we might settle for a

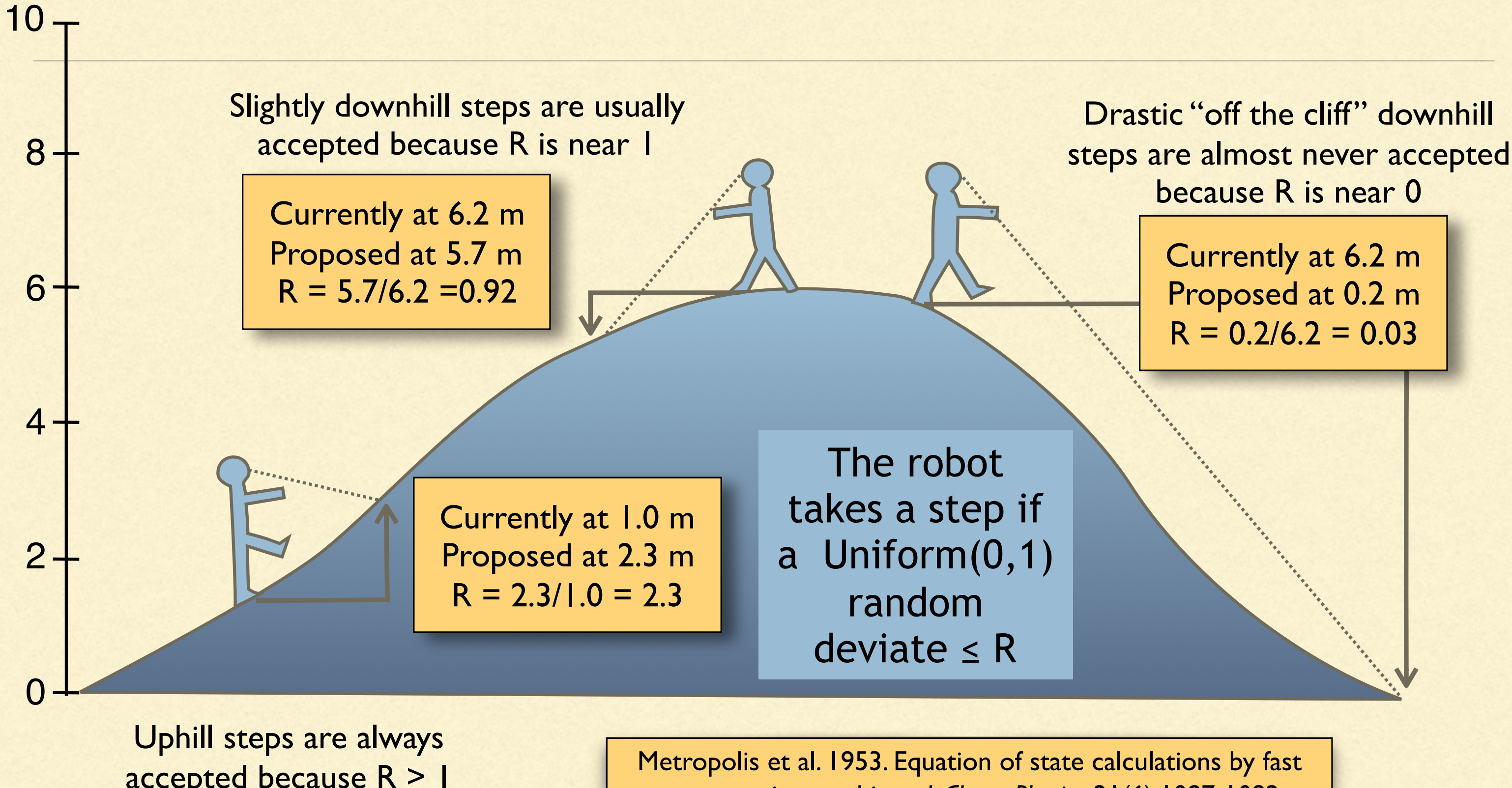
good approximation

to the posterior distribution

MCMC robot's rules



Actual rules (Metropolis algorithm)



Metropolis et al. 1953. Equation of state calculations by fast computing machines. *J. Chem. Physics* 21(6):1087-1092.

Cancellation of marginal likelihood

When calculating the ratio (R) of posterior densities, the marginal probability of the data cancels.

$$\frac{p(\theta^* \mid D)}{p(\theta \mid D)} = \frac{\dfrac{p(D \mid \theta^*)\, p(\theta^*)}{\cancel{p(D)}}}{\dfrac{p(D \mid \theta)\, p(\theta)}{\cancel{p(D)}}} = \frac{p(D \mid \theta^*)\, p(\theta^*)}{p(D \mid \theta)\, p(\theta)}$$

Posterior odds

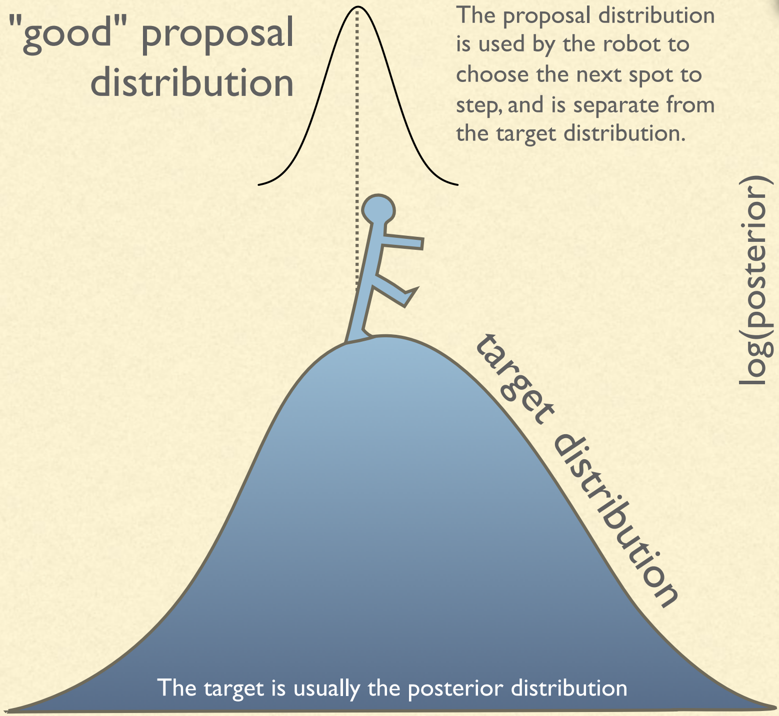
Apply Bayes' rule to both top and bottom

Likelihood ratio

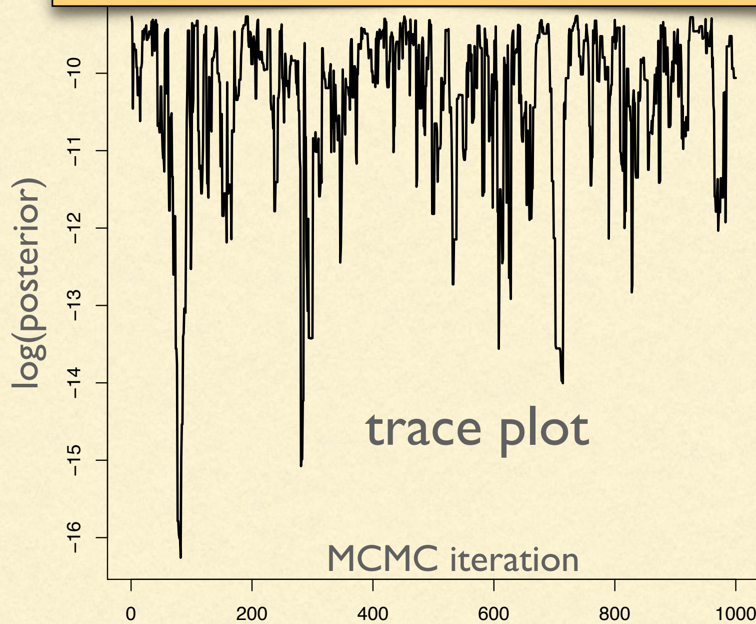
Prior odds

Target vs. Proposal Distributions

The proposal distribution is used by the robot to choose the next spot to step, and is separate from the target distribution.

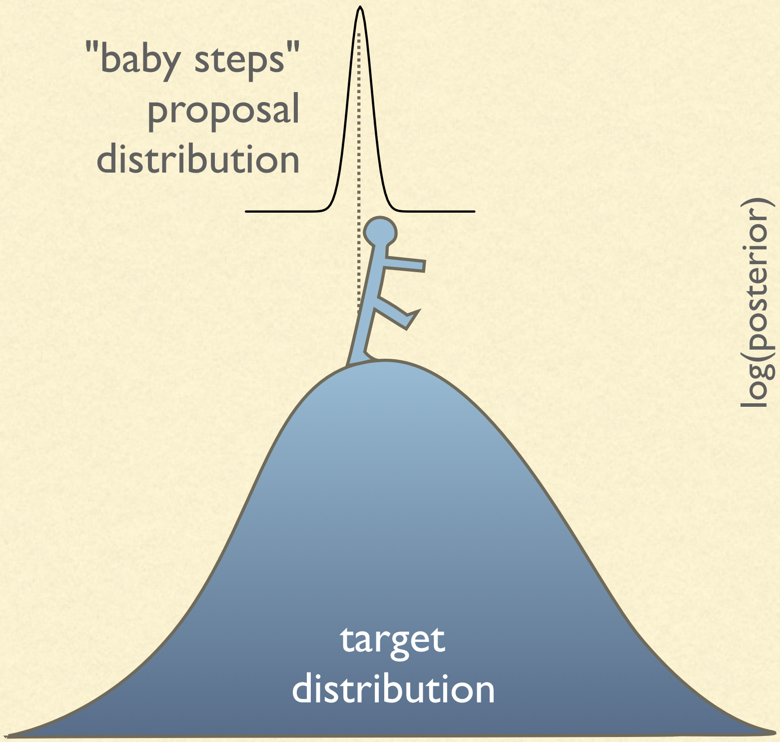


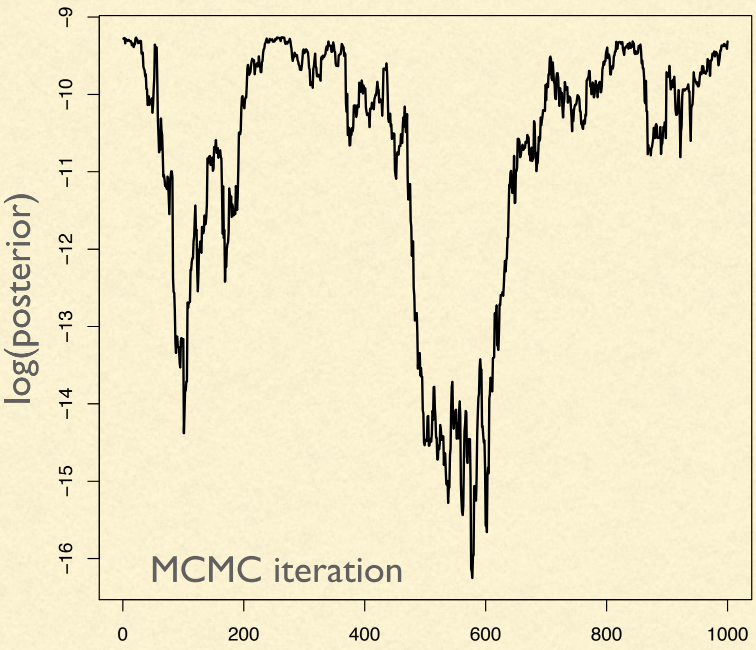
Tracer (app for generating trace plots from MCMC output):
https://github.com/beast-dev/tracer/releases/tag/v1.7.1



White noise appearance is a sign of good mixing

Target vs. Proposal Distributions

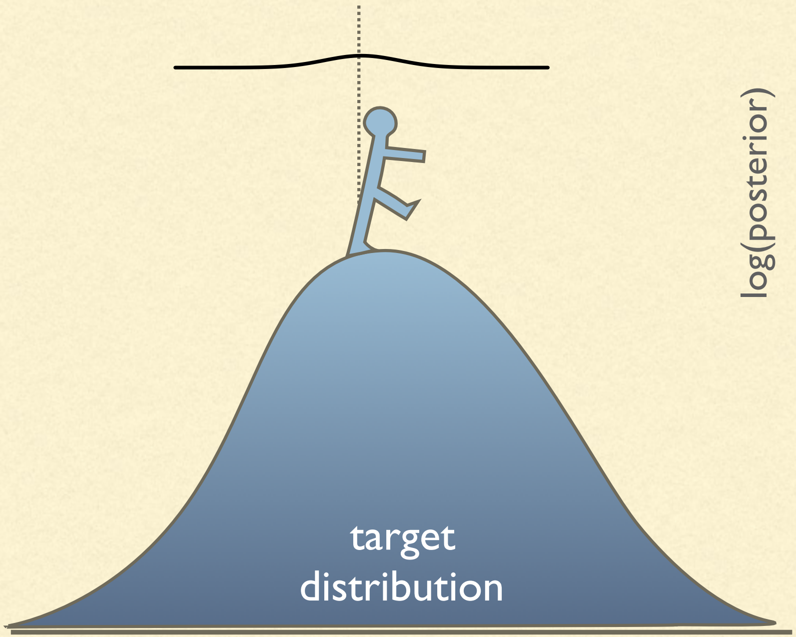


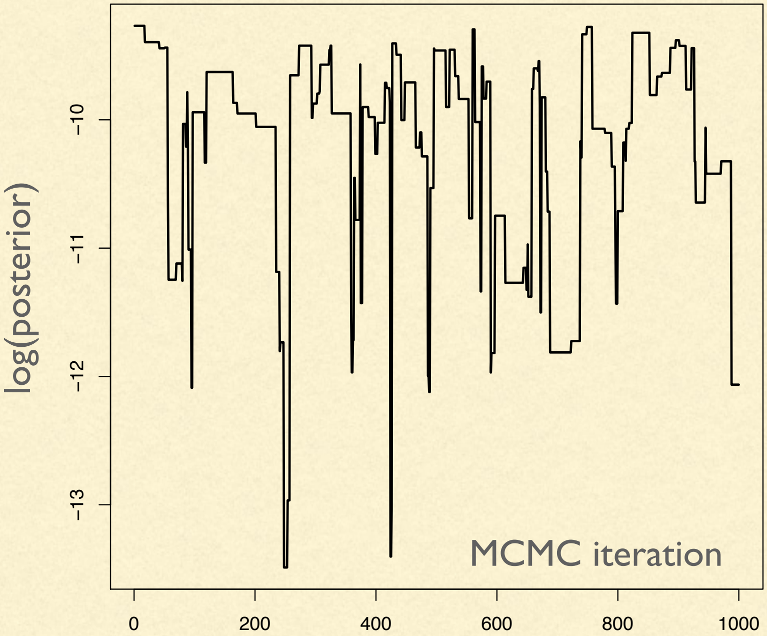


Big waves in trace plot indicate robot is crawling around

Target vs. Proposal Distributions

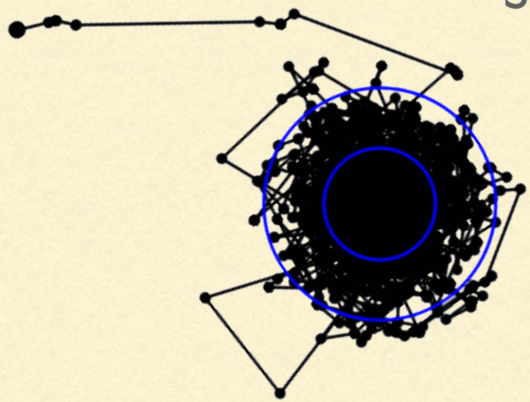
"overly bold" proposal distribution



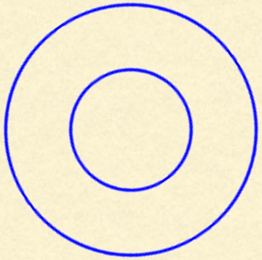


Plateaus in trace plot indicate robot is often stuck in one place

Metropolis-coupled Markov chain Monte Carlo (MCMCMC)



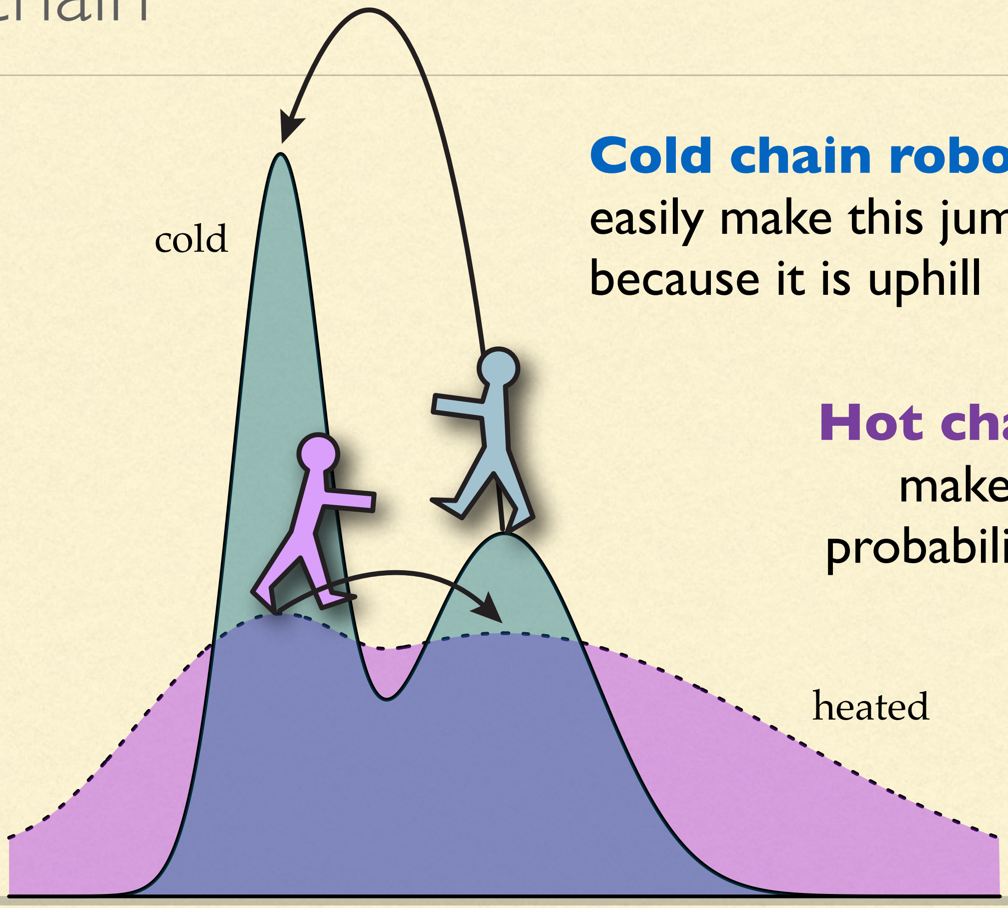
Sometimes the robot needs some help,



MCMCMC introduces helpers in the form of "heated chain" robots that can act as scouts.

Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood for dependent data. Pages 156-163 *in* Computing Science and Statistics (E. Keramidas, ed.).

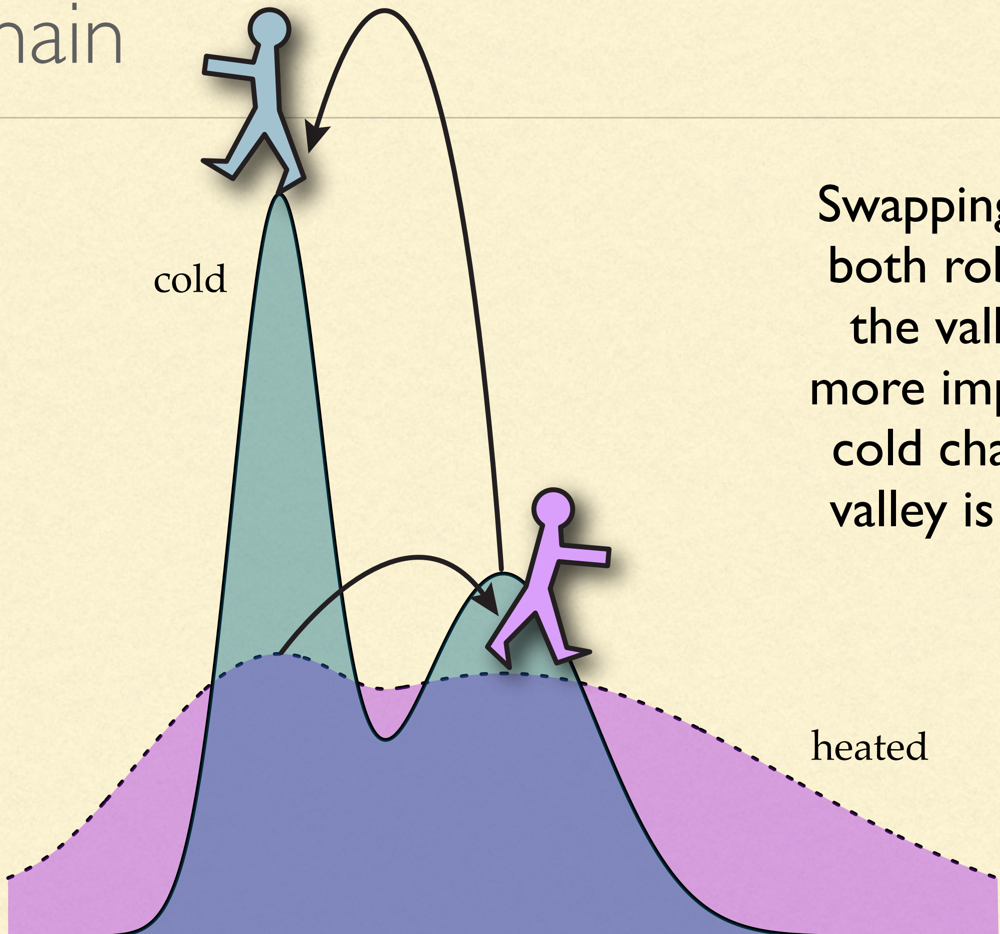
Heated chains act as scouts for the cold chain



Cold chain robot can easily make this jump because it is uphill

Hot chain robot can also make this jump with high probability because it is only slightly downhill

Heated chains act as scouts for the cold chain



Swapping places means both robots can cross the valley, but this is more important for the cold chain because its valley is much deeper.

Markov Chain Monte Carlo

Learn more about MCMC!

REVIEW ARTICLE
DOI: 10.1038/s41559-017-0280-x

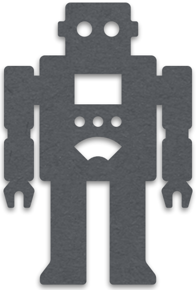
nature ecology & evolution

A biologist's guide to Bayesian phylogenetic analysis

Fabrícia F. Nascimento[1,4*], Mario dos Reis[2] and Ziheng Yang[3*]

https://thednainus.wordpress.com/2017/03/03/tutorial-bayesian-mcmc-phylogenetics-using-r/

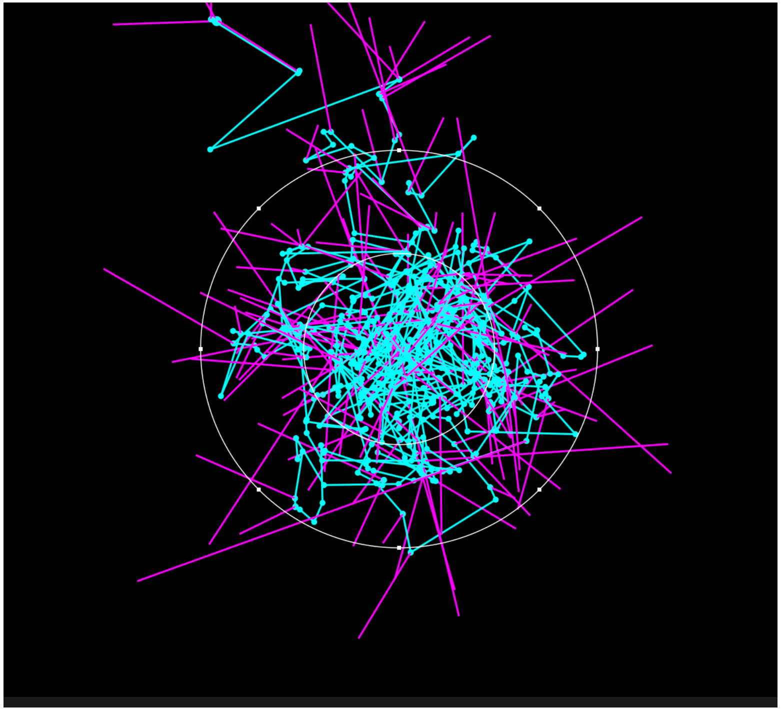
Markov Chain Monte Carlo



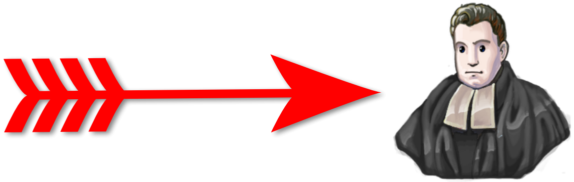
Learn more about MCMC!

https://phylogeny.uconn.edu/mcmc-robot/

MCMCRobot, a helpful tool for learning MCMC by Paul Lewis



RevBayes Demo



https://github.com/phyloworks/revbayes-workshop2017/blob/master/archery-model/archery-mcmc.ipynb