

# Bayesian Inference

Emily Jane McTavish

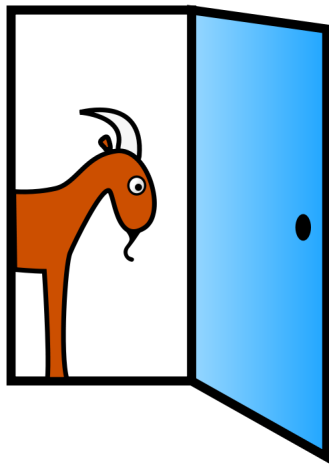
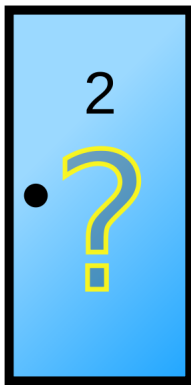
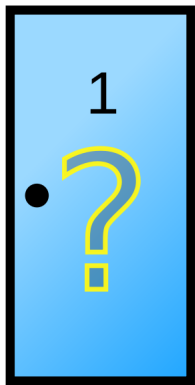
Life and Environmental Sciences  
University of California, Merced

`ejmctavish@ucmerced.edu`, `twitter:snacktavish`

(With thanks to Mark Holder, and Paul Lewis for slides!)

## Monte Hall problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 2, which has a goat. He then says to you, "Do you want to pick door No. 3?" Is it to your advantage to switch your choice?



## Monte Hall problem

Exercise:

- ▶ Find a partner.
- ▶ Take 3 playing cards - two red (goats), and one black (car).
- ▶ Play the Monte Hall game. (take turns as the host and the contestant)
- ▶ Record if you got the car, and if you chose to switch or not, over at least 12 games per tactic.
- ▶ Try out other variants! (what if the host doesn't know which door has the car? What if you have 3 goats and a car?)

Suppose that you are worried that you might have a rare disease.

You decide to get tested.

Test is correct 99 percent of the time.

(in other words, if you have the disease, it shows that you do with 99 percent probability, and if you don't have the disease, it shows that you do not with 99 percent probability).

Suppose that you are worried that you might have a rare disease.

You decide to get tested.

Test is correct 99 percent of the time.

(in other words, if you have the disease, it shows that you do with 99 percent probability, and if you don't have the disease, it shows that you do not with 99 percent probability).

This disease is actually quite rare, occurring randomly in the general population in only one of every 10,000 people.

**If your test results come back positive, what are your chances that you actually have the disease?**

Example from Su, Francis E., et al. "Medical Tests and Bayes' Theorem." Math Fun Facts.  
<<http://www.math.hmc.edu/funfacts>>.

If your test results come back positive, what are your chances that you actually have the disease?

- ▶ .99
- ▶ .90
- ▶ .10
- ▶ .01

If your test results come back positive, what are your chances that you actually have the disease?

- ▶ .99
- ▶ .90
- ▶ .10
- ▶ .01

Congratulations, there is less than 1 percent chance that you have the disease!

We can think about this problem in terms of conditional probabilities.



We can think about this problem in terms of conditional probabilities. Event A is the event you have this disease, and event B is the event that you test positive.

$P(B|notA)$  is the probability of a "false positive": that you test positive even though you don't have the disease.

We can think about this problem in terms of conditional probabilities. Event A is the event you have this disease, and event B is the event that you test positive.

$P(B|notA)$  is the probability of a "false positive": that you test positive even though you don't have the disease.

Here,  $P(B|A) = .99$

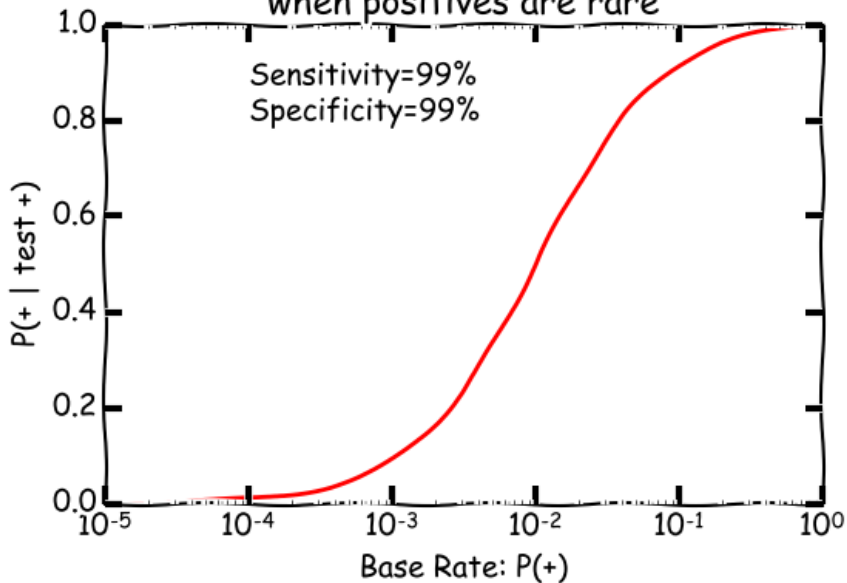
but  $P(A) = .0001$ .

$P(B)$  may be derived by conditioning on whether event A does or does not occur:

$$P(B) = P(B|A)P(A) + P(B|notA)P(notA)$$

or  $.99*.0001 + .01*.9999$ . Thus the probability that you have the disease is less than 1 percent.

Even a *\*very good\** test  
will be wrong most of the time  
when positives are rare



**Bayes' theorem** (aka Bayes' rule or Bayes law) helps us find the probability of event A given event B,  $P(A|B)$ , in terms of the probability of B given A,  $P(B|A)$ , and the probabilities of A and B:

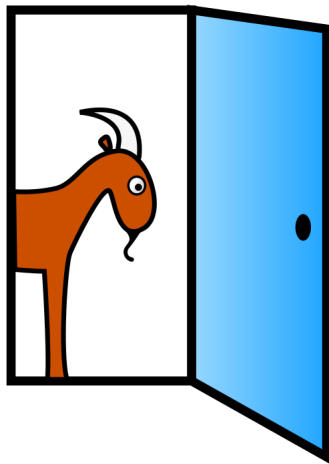
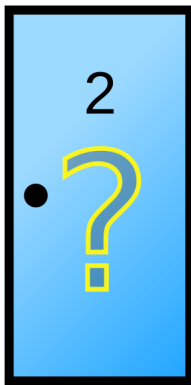
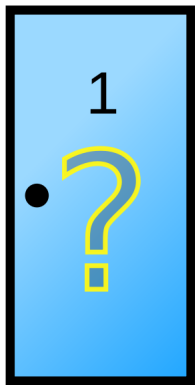
$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$



Reverend Bayes

## Monte Hall problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 2, which has a goat. He then says to you, "Do you want to pick door No. 3?" Is it to your advantage to switch your choice?



# Applying Bayes theorem to the Monty Hall problem:

Assume we pick Door 1 and then Monty shows us a goat behind Door 2.

Now let A be the event that the car is behind Door 1.

B be the event that Monty shows us a goat behind Door 2.

$$\begin{aligned}\Pr(A | B) &= \frac{\Pr(B | A) \times \Pr(A)}{\Pr(B)} \\ &= \frac{1/2 \times 1/3}{1/3 \times 1/2 + 1/3 \times 0 + 1/3 \times 1} \\ &= 1/3.\end{aligned}$$

[http://angrystatistician.blogspot.com/2012/06/  
bayes-solution-to-monty-hall.html](http://angrystatistician.blogspot.com/2012/06/bayes-solution-to-monty-hall.html)

The tricky calculation is  $\Pr(B)$ .

We initially chose Door 1.

if the car is behind Door 1, Monty will show us a goat behind Door 2 half the time,  $1/3 \times 1/2$

The tricky calculation is  $\Pr(B)$ .

We initially chose Door 1.

if the car is behind Door 1, Monty will show us a goat behind Door 2 half the time,  $1/3 \times 1/2$

If the car is behind Door 2, Monty never shows us a goat behind Door 2.  
 $1/3 \times 0$

Finally, if the car is behind Door 3, Monty shows us a goat behind Door 2 every time.  $1/3 \times 1$  Thus,  $\Pr(B) = 1/3 \times 1/2 + 1/3 \times 0 + 1/3 \times 1 = 1/2$ .



Key aspect of this problem: The added information is not THAT Monty Hall revealed a goat, it is that he picked a door specifically because it had a goat.

We saw a goat behind door 2. The car is either behind Door 1 or Door 3, since the probability that it is behind Door 1 is  $1/3$  and the sum of the two probabilities must equal 1, the probability the car is behind Door 3 is  $1 - 1/3 = 2/3$ .

An R script to simulate the Monte Hall problem by Corey Chivers,  
<https://bayesianbiologist.com/>  
[https://raw.githubusercontent.com/McTavishLab/GradPhylo/  
master/docs/scripts/MonteHall.R](https://raw.githubusercontent.com/McTavishLab/GradPhylo/master/docs/scripts/MonteHall.R)

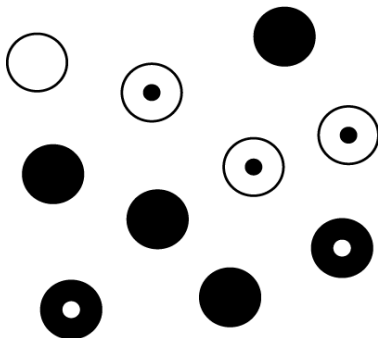
HW: What is the probability you get the car after switching if there are 10 doors? (1 car, 9 goats, everything else the same.)

Advantages of Bayesian inference:

- ▶ integrating prior information
- ▶ marginalizing inferences over parameter values

# Joint probabilities

B = Black      S = Solid  
W = White    D = Dotted



$$\begin{aligned}\Pr(B) &= 0.6 & \Pr(S) &= 0.5 \\ \Pr(W) &= 0.4 & \Pr(D) &= 0.5\end{aligned}$$

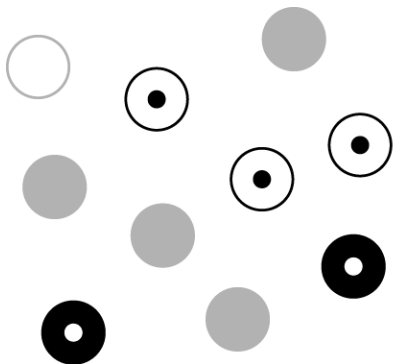
$$\Pr(\bullet\odot) = \Pr(B, D) = 0.2$$

$$\Pr(\bullet\bullet) = \Pr(B, S) = 0.4$$

$$\Pr(\odot\odot) = \Pr(W, D) = 0.3$$

$$\Pr(\odot\bullet) = \Pr(W, S) = 0.1$$

# Conditional probabilities



$$\Pr(B|D) = \frac{2}{5} = 0.4$$

Hide all solid marbles  
(leaving 5 with dot)

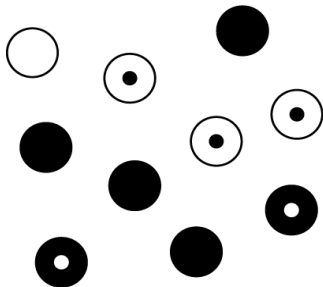
Of those left, 2 are black

# Bayes' rule

$\Pr(B, D)$

$$\Pr(D) \Pr(B|D) = \Pr(B) \Pr(D|B)$$

$$\frac{1}{2} \times \frac{2}{5} = \frac{3}{5} \times \frac{1}{3}$$

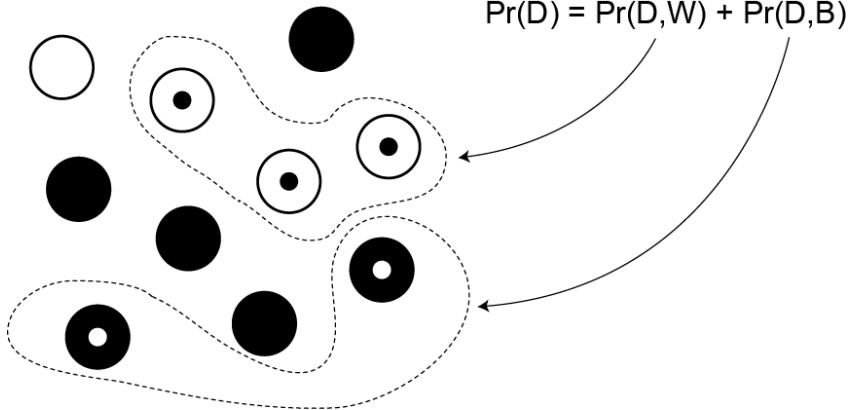


$$\begin{aligned} \Pr(B|D) &= \frac{\Pr(B) \Pr(D|B)}{\Pr(D)} \\ &= \frac{\frac{3}{5} \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{5} \end{aligned}$$



# Probability of "Dotted"

$$\Pr(D) = \Pr(D,W) + \Pr(D,B)$$



## Bayes' rule (cont.)

$$\begin{aligned}\Pr(B|D) &= \frac{\Pr(B) \Pr(D|B)}{\Pr(D)} \\ &= \frac{\Pr(D, B)}{\Pr(D, B) + \Pr(D, W)}\end{aligned}$$

$\Pr(D)$  is the **marginal probability** of being dotted  
To compute it, we **marginalize over colors**

## Bayes' rule (cont.)

It is easy to see that  $\Pr(D)$  serves as a *normalization constant*, ensuring that  $\Pr(B|D) + \Pr(W|D) = 1.0$

$$\Pr(B|D) = \frac{\Pr(D, B)}{\Pr(D, B) + \Pr(D, W)} \longleftarrow \Pr(D)$$

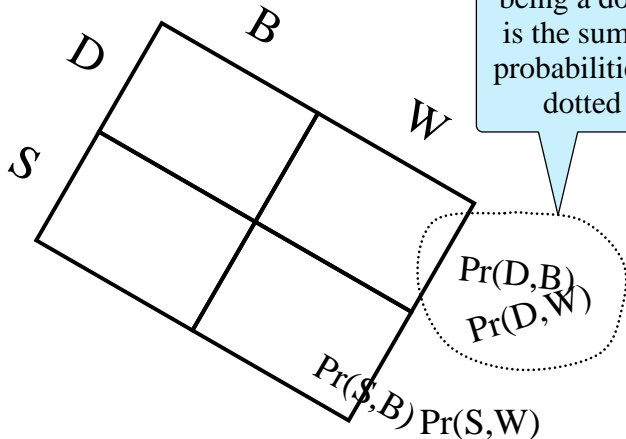
$$\Pr(W|D) = \frac{\Pr(D, W)}{\Pr(D, B) + \Pr(D, W)} \longleftarrow \Pr(D)$$

$$\Pr(B|D) + \Pr(W|D) = \frac{\cancel{\Pr(D, B)} + \cancel{\Pr(D, W)}}{\cancel{\Pr(D, B)} + \cancel{\Pr(D, W)}} = 1$$

# Joint probabilities

	B	W
D	$\Pr(D,B)$	$\Pr(D,W)$
S	$\Pr(S,B)$	$\Pr(S,W)$

# Marginalizing over colors



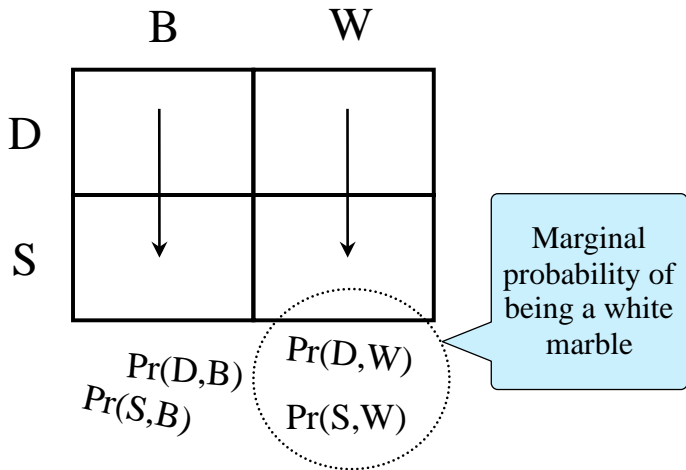
# Marginal probabilities

	B	W	
D			$\Pr(D) = \text{marginal probability of being dotted}$
			$\Pr(D,B) + \Pr(D,W)$
S			
			$\Pr(S,B) + \Pr(S,W)$
			$\Pr(S) = \text{marginal probability of being solid}$

# Joint probabilities

	B	W
D	$\Pr(D,B)$	$\Pr(D,W)$
S	$\Pr(S,B)$	$\Pr(S,W)$

# Marginalizing over "dottedness"





## Bayes' rule (cont.)

$$\begin{aligned}\Pr(B|D) &= \frac{\Pr(B) \Pr(D|B)}{\Pr(D, B) + \Pr(D, W)} \\&= \frac{\Pr(B) \Pr(D|B)}{\Pr(B) \Pr(D|B) + \Pr(W) \Pr(D|W)} \\&= \frac{\Pr(B) \Pr(D|B)}{\sum_{\theta \in \{B, W\}} \Pr(\theta) \Pr(D|\theta)}\end{aligned}$$

# Bayes' rule in Statistics

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

$D$  refers to the "observables" (i.e. the **Data**)

$\theta$  refers to one or more "unobservables"

(i.e. **parameters** of a model, or the **model itself**):

- *tree model* (i.e. tree topology)
- *substitution model* (e.g. JC, F84, GTR, etc.)
- *parameter* of a substitution model (e.g. a branch length, a base frequency, transition/transversion rate ratio, etc.)
- *hypothesis* (i.e. a special case of a model)
- a *latent variable* (e.g. ancestral state)

# Bayes' rule in statistics

The diagram illustrates Bayes' rule with the following components and annotations:

- Likelihood of hypothesis  $\theta$** : An arrow points from this text to the blue box containing  $\Pr(D|\theta)$ .
- Prior probability of hypothesis  $\theta$** : An arrow points from this text to the orange box containing  $\Pr(\theta)$ .
- Posterior probability of hypothesis  $\theta$** : An arrow points from this text to the purple box containing  $\Pr(\theta|D)$ .
- Marginal probability of the data (marginalizing over hypotheses)**: An arrow points from this text to the green box containing the denominator  $\sum_{\theta} \Pr(D|\theta) \Pr(\theta)$ .

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

# Bayes' rule: continuous case

The diagram illustrates Bayes' rule for the continuous case. It features the equation  $f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$ . The components are highlighted with colored boxes: the numerator terms  $f(D|\theta)$  and  $f(\theta)$  are in blue and orange respectively, the denominator integral is in a green box, and the entire left side  $f(\theta|D)$  is in a purple box. Arrows point from descriptive labels to these components: 'Likelihood' points to  $f(D|\theta)$ , 'Prior probability density' points to  $f(\theta)$ , 'Posterior probability density' points to  $f(\theta|D)$ , and 'Marginal probability of the data' points to the denominator integral.

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

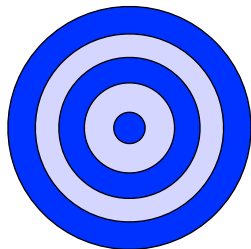
Likelihood

Prior probability *density*

Posterior probability *density*

Marginal probability of the data

# If you had to guess...



← 1 meter →

0.0

$d$

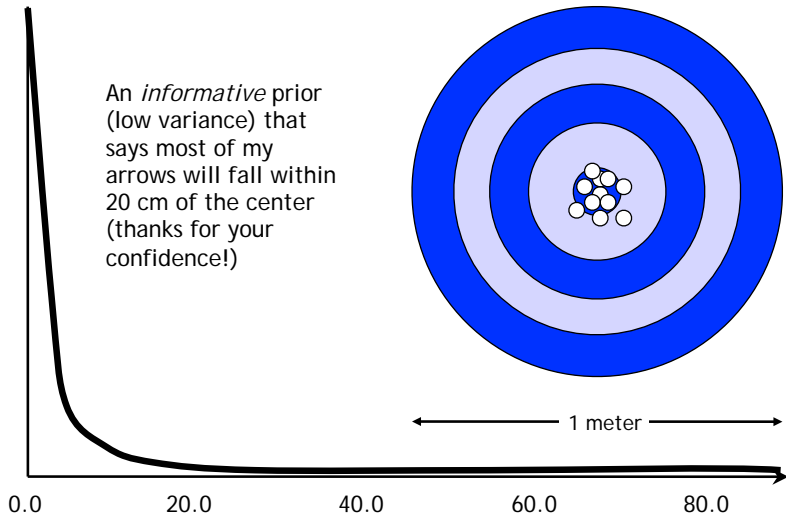
$\infty$

*Not knowing anything about my archery abilities, draw a curve representing your view of the chances of my arrow landing a distance  $d$  centimeters from the center of the target (if it helps, I'm standing 50 meters away from the target)*

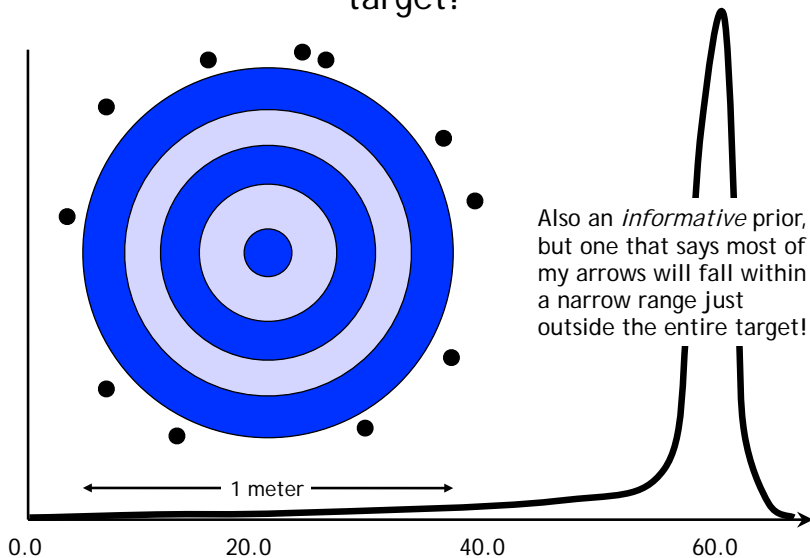


Photo by Tracy Heath

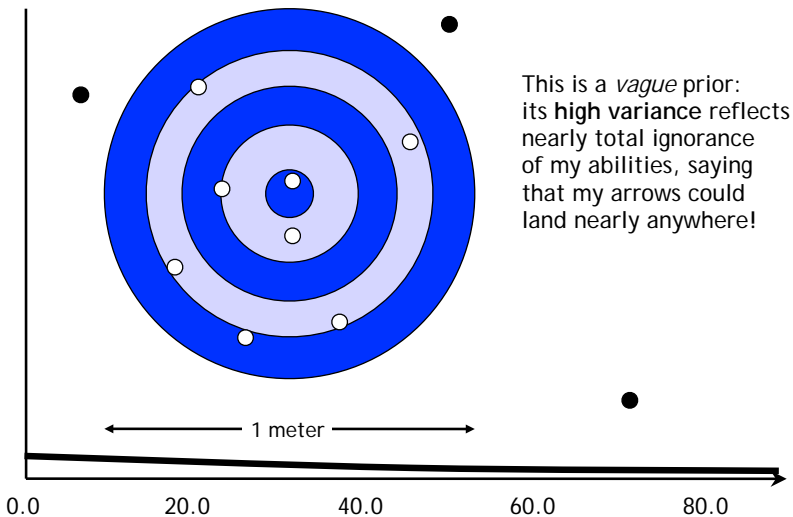
## Case 1: assume I have talent



## Case 2: assume I have a talent for missing the target!



### Case 3: assume I have no talent





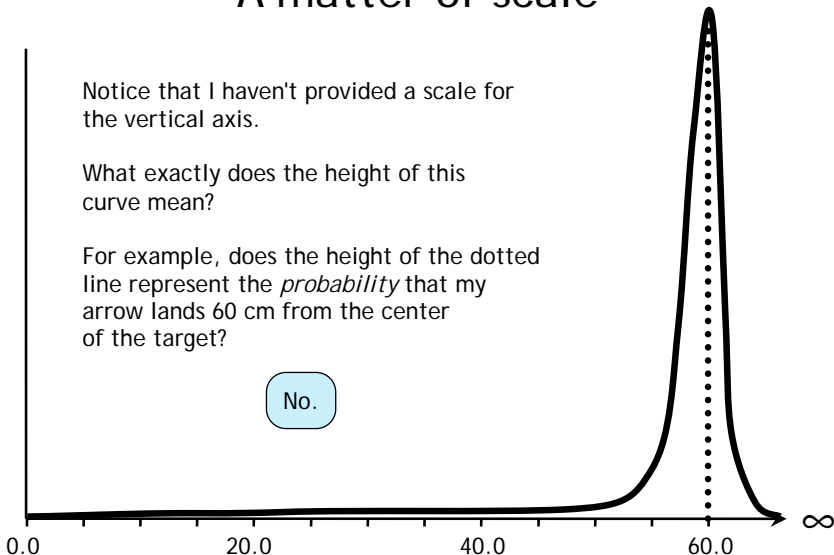
# A matter of scale

Notice that I haven't provided a scale for the vertical axis.

What exactly does the height of this curve mean?

For example, does the height of the dotted line represent the *probability* that my arrow lands 60 cm from the center of the target?

No.

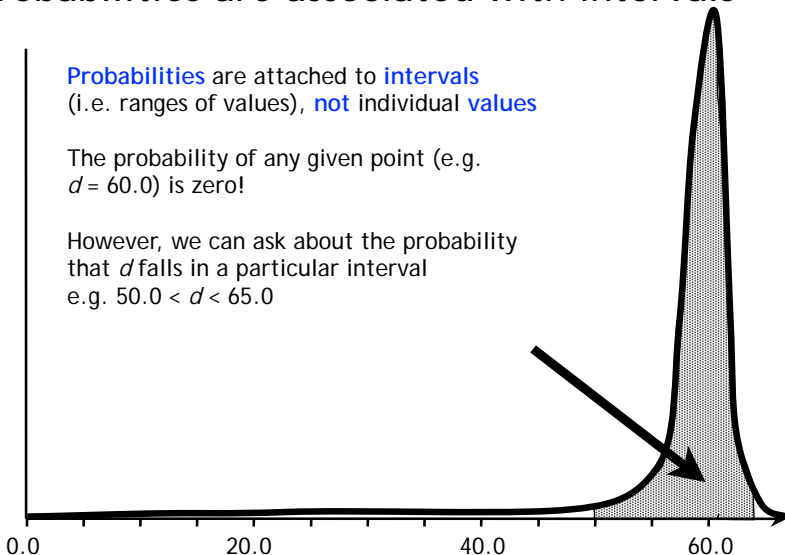


# Probabilities are associated with intervals

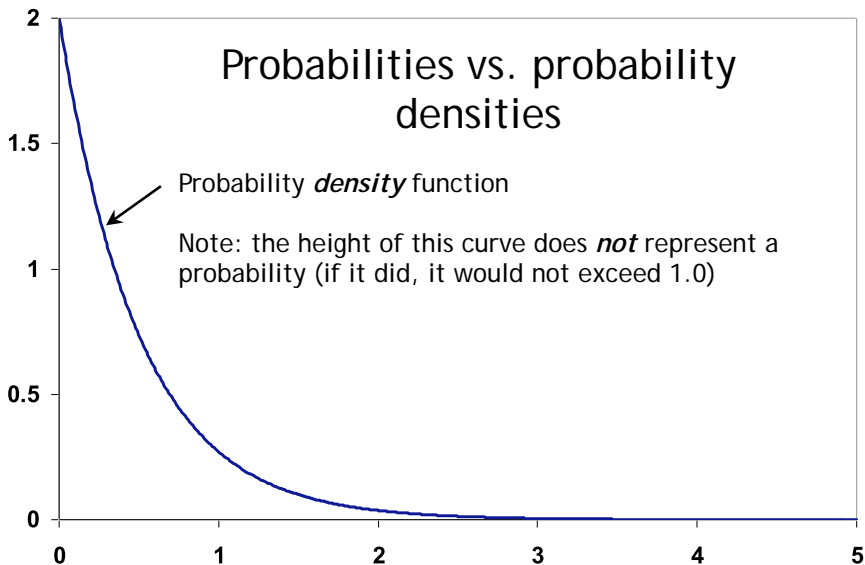
**Probabilities** are attached to **intervals**  
(i.e. ranges of values), **not** individual **values**

The probability of any given point (e.g.  
 $d = 60.0$ ) is zero!

However, we can ask about the probability  
that  $d$  falls in a particular interval  
e.g.  $50.0 < d < 65.0$



## Probabilities vs. probability densities



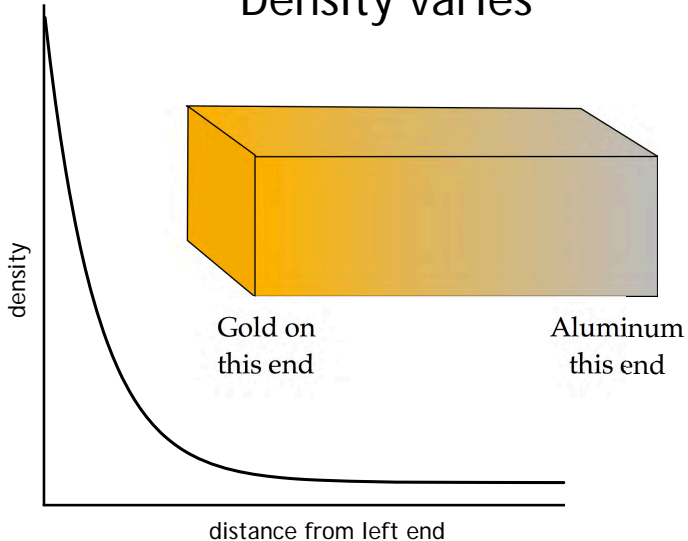
# Densities of various substances

Substance	Density (g/cm <sup>3</sup> )
Cork	0.24
Aluminum	2.7
Gold	19.3

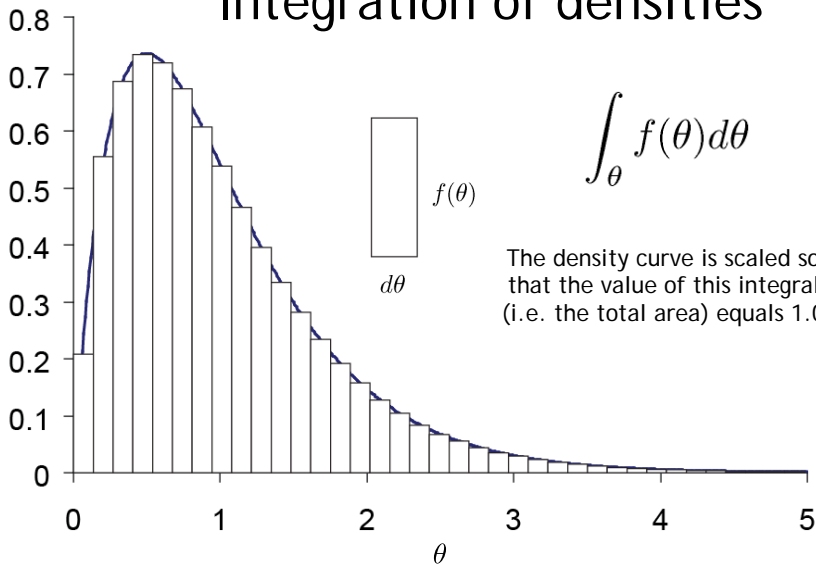
*Density does not equal mass*

$$\text{mass} = \text{density} \times \text{volume}$$

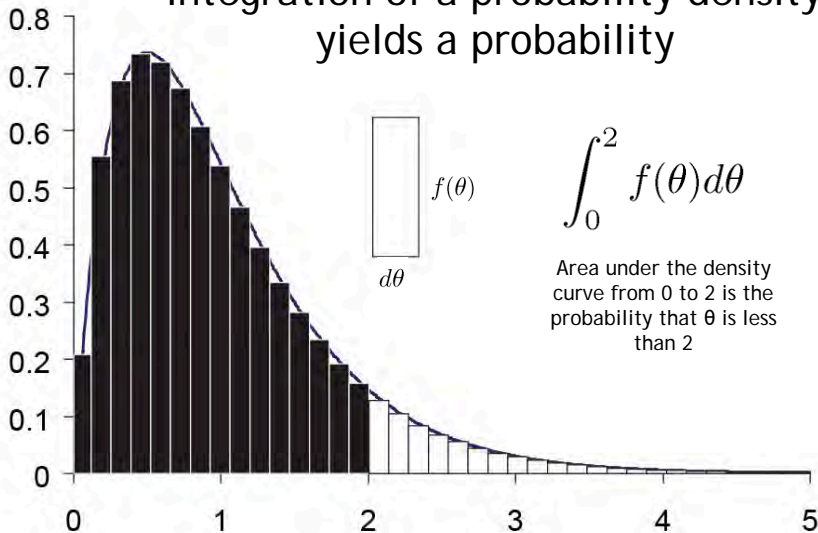
# Density varies



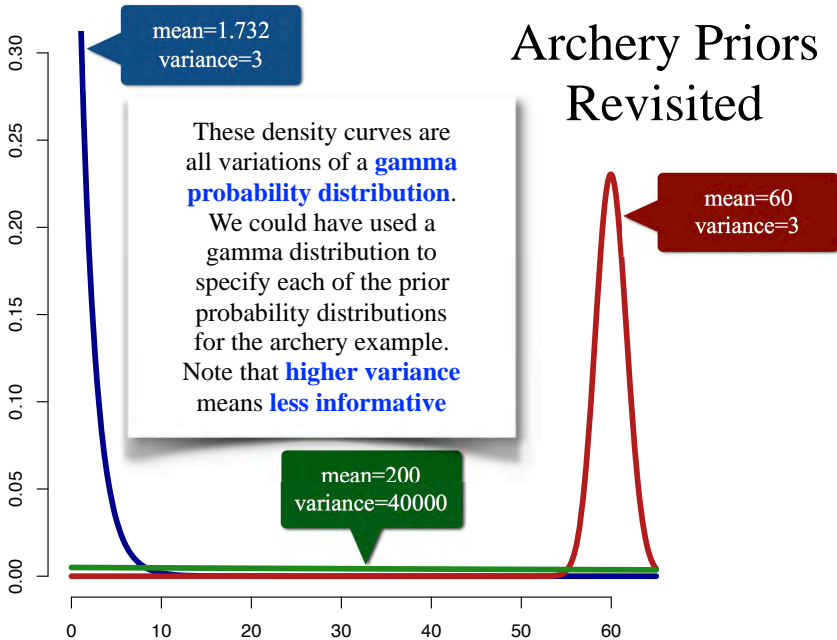
# Integration of densities



# Integration of a probability density yields a probability



# Archery Priors Revisited





# Usually there are many parameters...

A 2-parameter example

$$f(\theta, \phi | D) = \frac{f(D|\theta, \phi) f(\theta) f(\phi)}{\int_{\theta} \int_{\phi} f(D|\theta, \phi) f(\theta) f(\phi) d\theta d\phi}$$

Posterior probability density

Likelihood

Prior probability density

Marginal probability of data

An analysis of **100 sequences** under the simplest model (JC69) requires 197 branch length parameters. The denominator is a **197-fold integral** in this case! Now consider summing over **all possible tree topologies**! It would thus be nice to avoid having to calculate the marginal probability of the data...