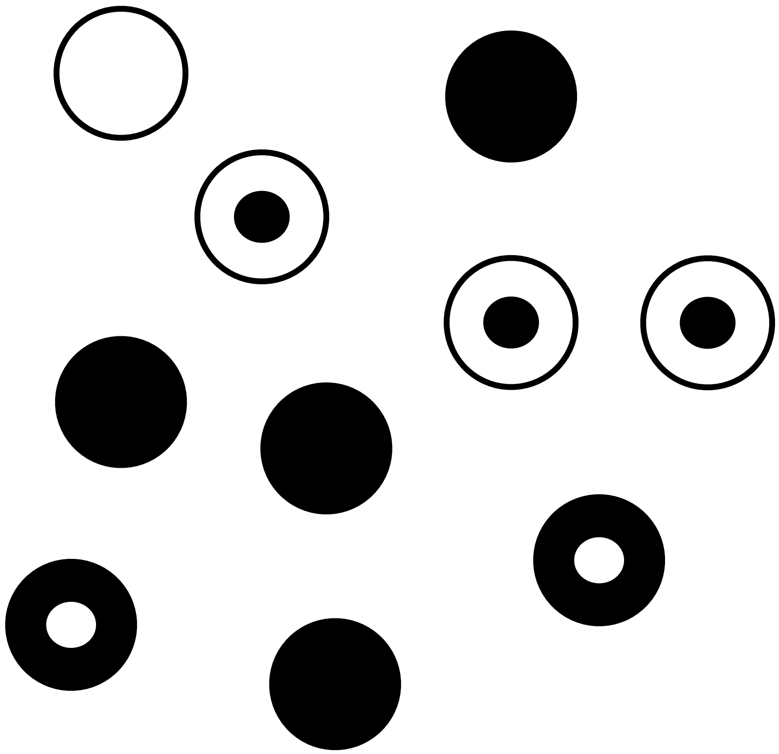
Bayes' rule



$$\Pr(B,D) = 1/5$$

$$p(B)\,p(D|B) = \left(\frac{\cancel{3}}{5}\right)\left(\frac{1}{\cancel{3}}\right) = \frac{1}{5}$$

$$p(D)\,p(B|D) = \left(\frac{1}{\cancel{2}}\right)\left(\frac{\cancel{2}}{5}\right) = \frac{1}{5}$$

$$p(D)\,p(B|D) = p(B)\,p(D|B)$$

$$\underline{p(B|D)} = \frac{p(B)\,p(D|B)}{p(D)}$$

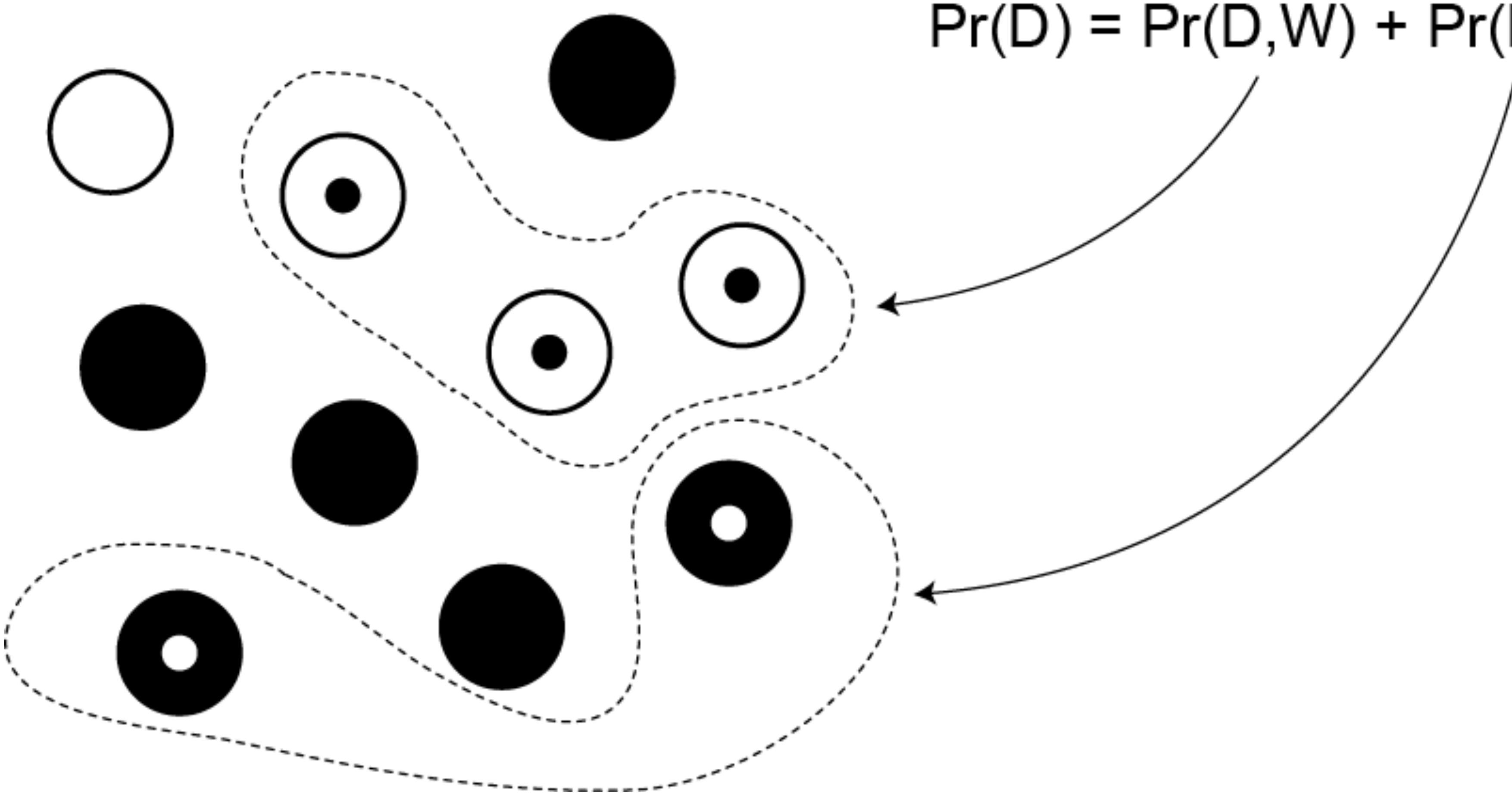
Bayes' Rule

{ total
marginal
unconditional

Probability of "Dotted"

$\frac{1}{2} = 3/10 + 2/10$

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



Bayes' rule (cont.)

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

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

← sum over all colors

Pr(*D*) is the **marginal probability** of being dotted

To compute it, we **marginalize over colors**

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

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

Marginal (total) probabilities

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

$p(D) = p(D,B) + p(D,W)$

$p(S)$

$p(B)$ $p(W)$

Bayes' rule (cont.)

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

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

$$= \frac{\Pr(B)\Pr(D|B)}{\sum_{\theta \in \{B,W\}} \Pr(\theta)\Pr(D|\theta)}$$

Bayes' rule in statistics

Likelihood of hypothesis θ

Prior probability of hypothesis θ

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

Posterior probability of hypothesis θ

aka marginal likelihood

Marginal probability of the data (marginalizing over hypotheses)

Simplest paternity example

child's genotype: **Aa**   mother's genotype: **aa**

possible fathers

Possibilities	θ_1	θ_2	Row sum
Genotypes	AA	Aa	---
Prior	$\frac{1}{2}$	$\frac{1}{2}$	1
Likelihood	1	$\frac{1}{2}$	—
Likelihood × Prior	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$
Posterior	$2/3$	$1/3$	1

$\frac{2}{4} / \frac{3}{4}$ $\frac{1}{4} / \frac{3}{4}$

Likelihood vs. Probability

Outcome	Fair coin model	Two-heads model
H	0.5	1
T	0.5	0
	1	1

likelihood of model given the data

probability of data given model (hypothesis)

The prior can be your friend

Suppose the test for a **rare** disease has the following true and false positive probabilities:

positive test

$$\Pr(+ \mid \text{disease}) = 1.00$$ true positive

$$\Pr(+ \mid \text{healthy}) = 0.01$$ false positive

datum hypothesis

(Note that we do not need to consider the case of a negative test result.)

Suppose further I **test positive** for the disease. How worried should I be?

It is very tempting to (mis)interpret the likelihood as a posterior probability and conclude "There is a 100% chance that I have the disease."

The prior can be your friend

prob. + given disease

prior prob. of having the disease

$$\Pr(\text{disease}|+) = \frac{(1.0)\left(\frac{1}{1000000}\right)}{(1.0)\left(\frac{1}{1000000}\right) + (0.01)\left(\frac{999999}{1000000}\right)}$$

1 person out of a million has a true positive result

10,000 people out a million will have a false positive result

Thus, the odds *against* having the disease are actually 10000 to 1!

Bayes' rule: continuous case

Likelihood

Prior probability density

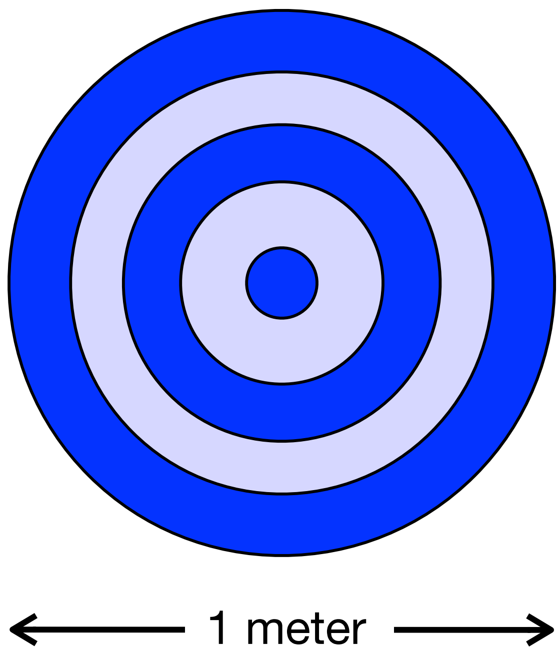
$$p(\theta|D) = \frac{p(D|\theta)\ p(\theta)}{\int p(D|\theta')\ p(\theta')d\theta'}$$

Posterior probability density

Marginal probability of the data

(a.k.a. marginal likelihood)

If you had to guess...

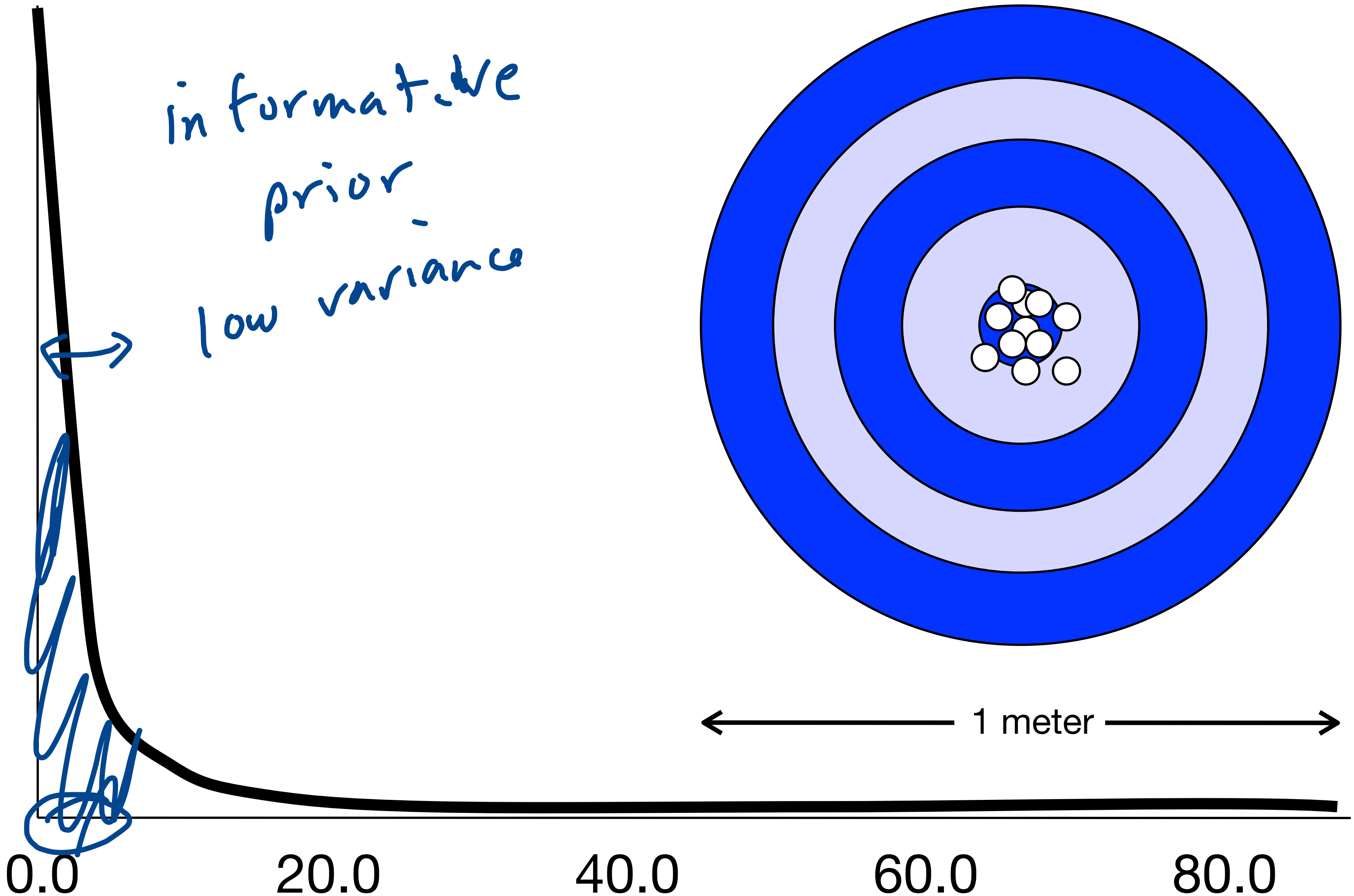


Not knowing anything about my archery abilities, draw a curve representing your view of the chances of my arrow landing a distance *d* from the center of the target (if it helps, I'm standing 50 meters away from the target)

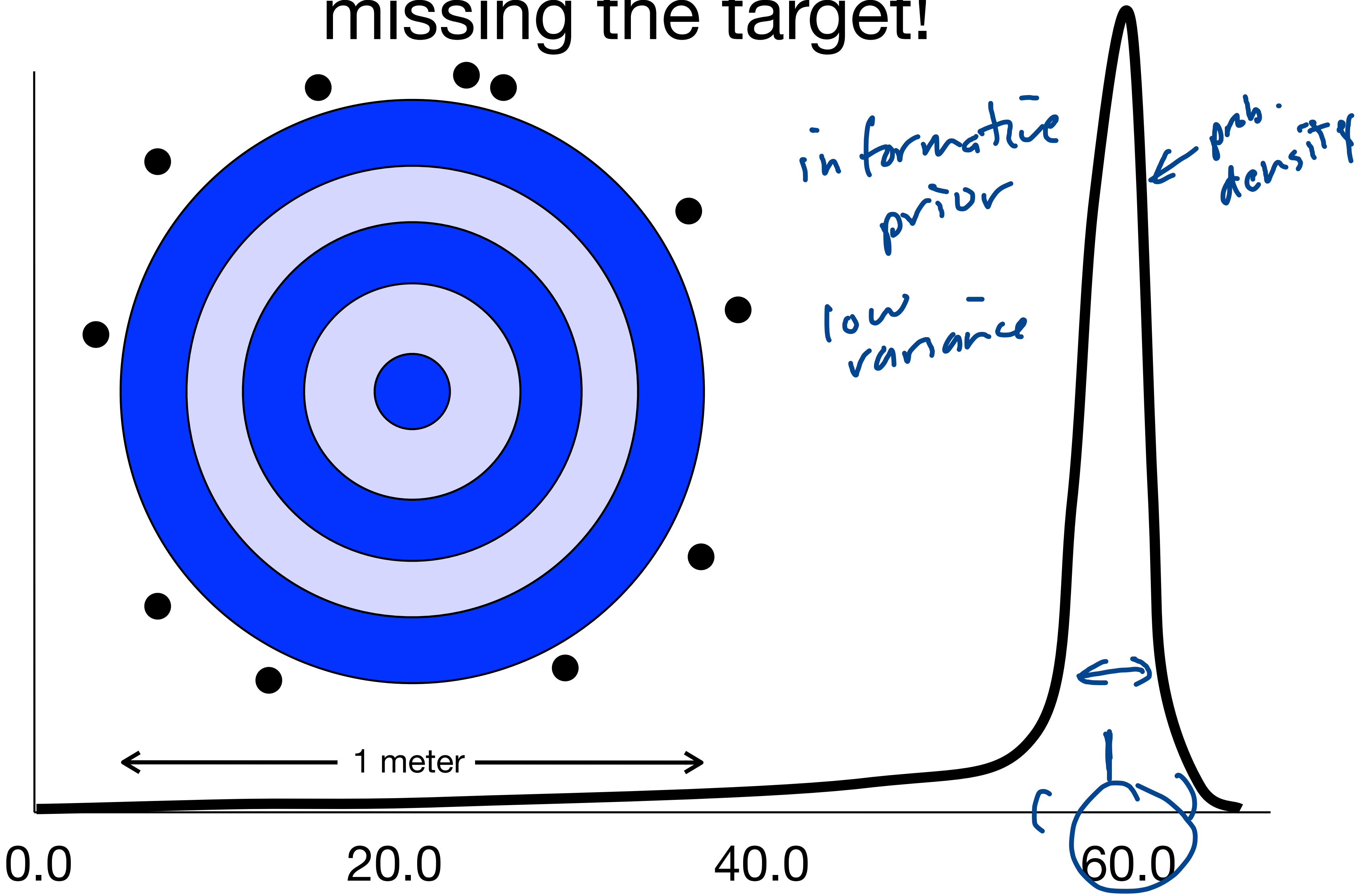
0.0 *d* ∞

distance in centimeters from target center

Case 1: assume I have talent



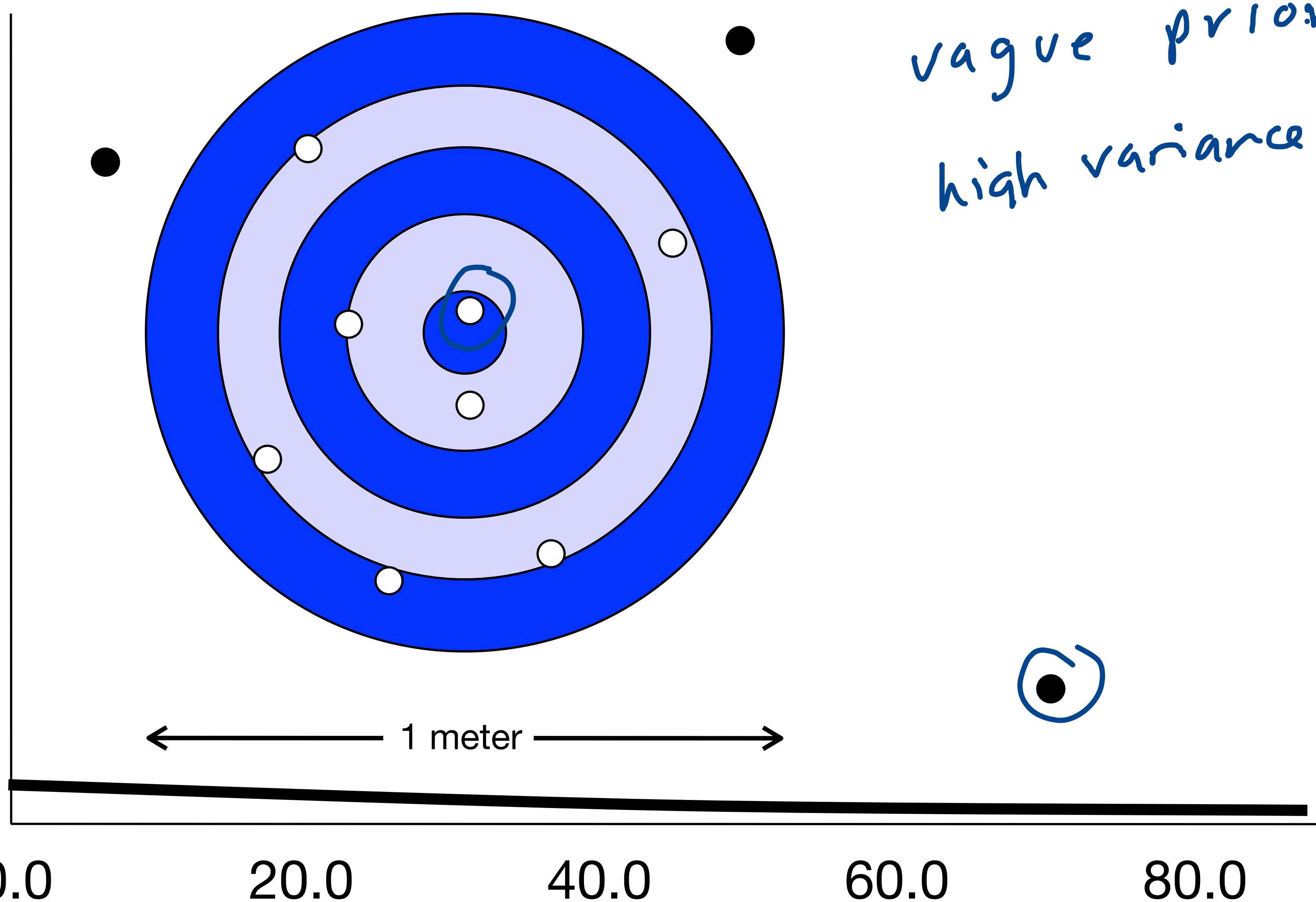
Case 2: assume I have a talent for missing the target!

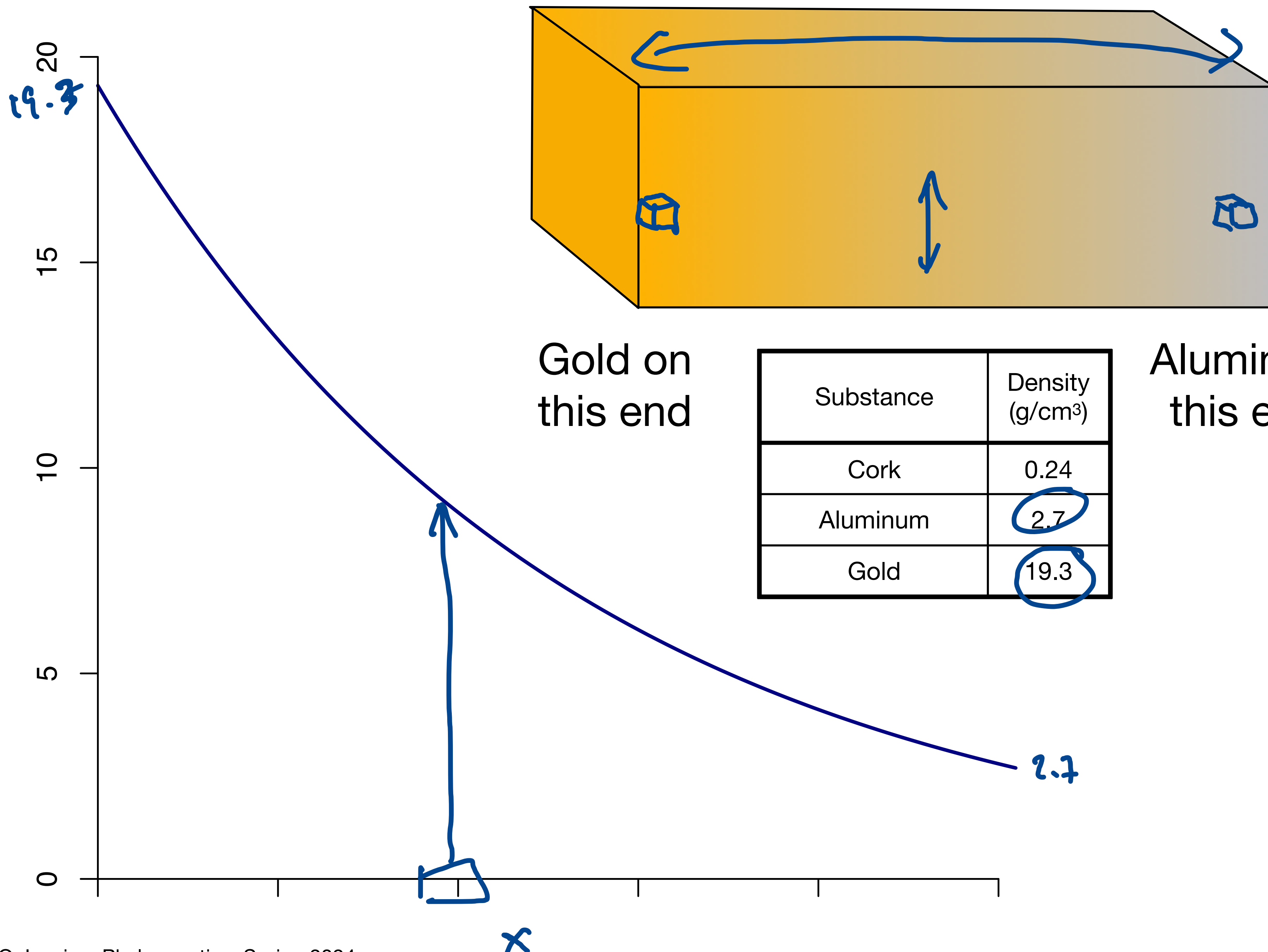


Case 3: assume I have no talent

vague prior

high variance



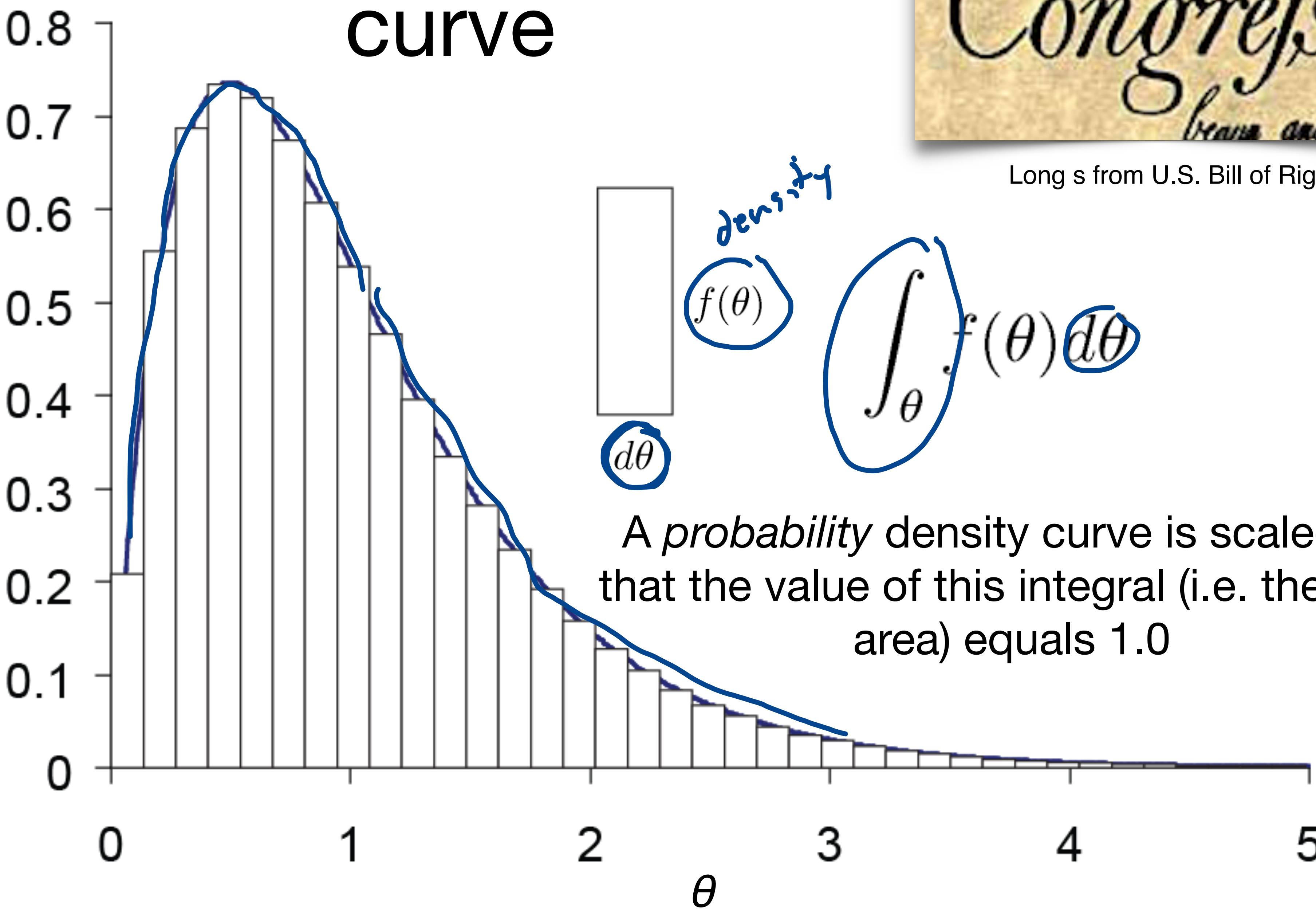


Gold on this end

Aluminum this end

Substance	Density (g/cm³)
Cork	0.24
Aluminum	2.7
Gold	19.3

Probability density curve



Long s from U.S. Bill of Rights

density

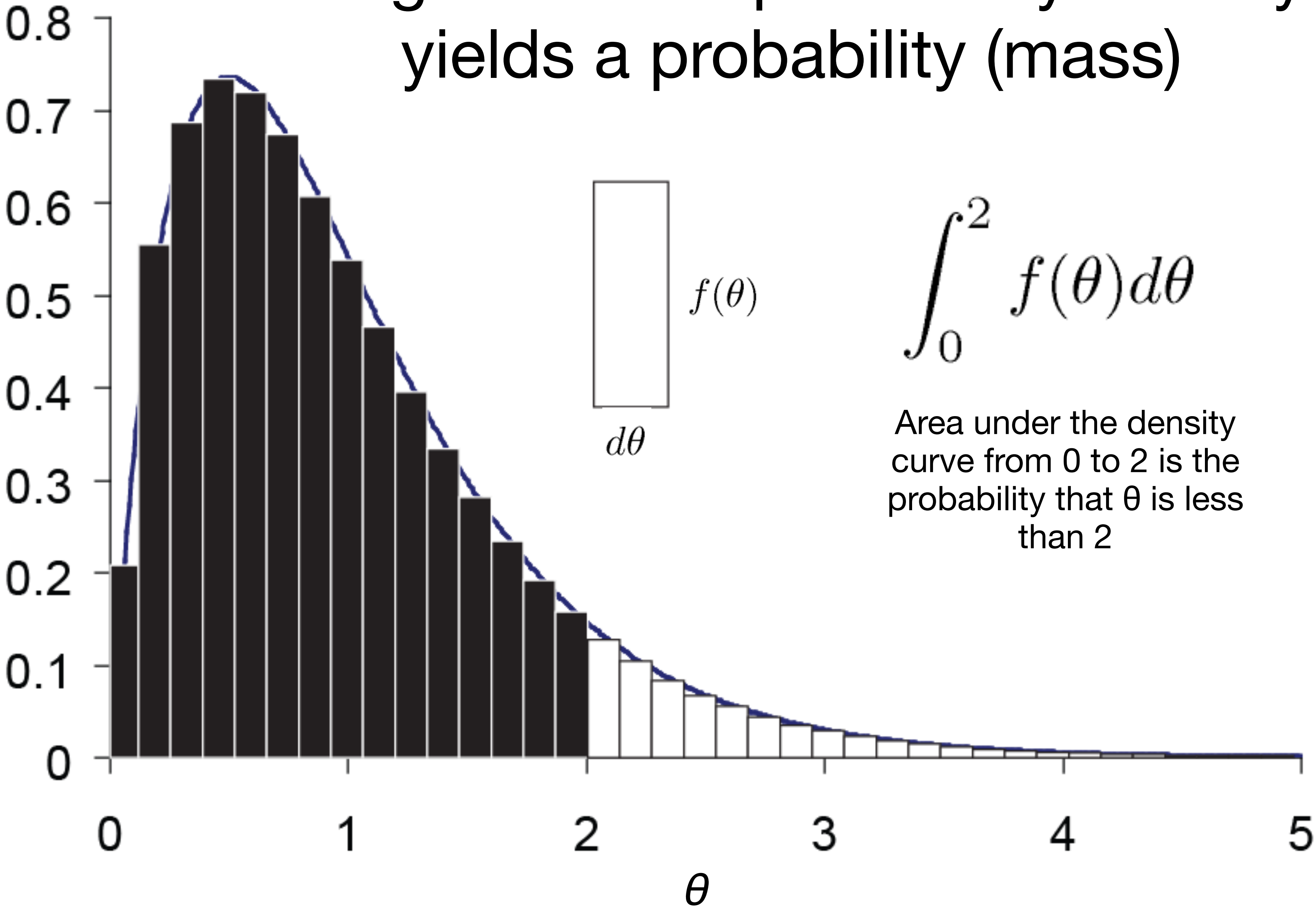
$f(\theta)$

$d\theta$

$$\int_\theta f(\theta)\, d\theta$$

A *probability* density curve is scaled so that the value of this integral (i.e. the total area) equals 1.0

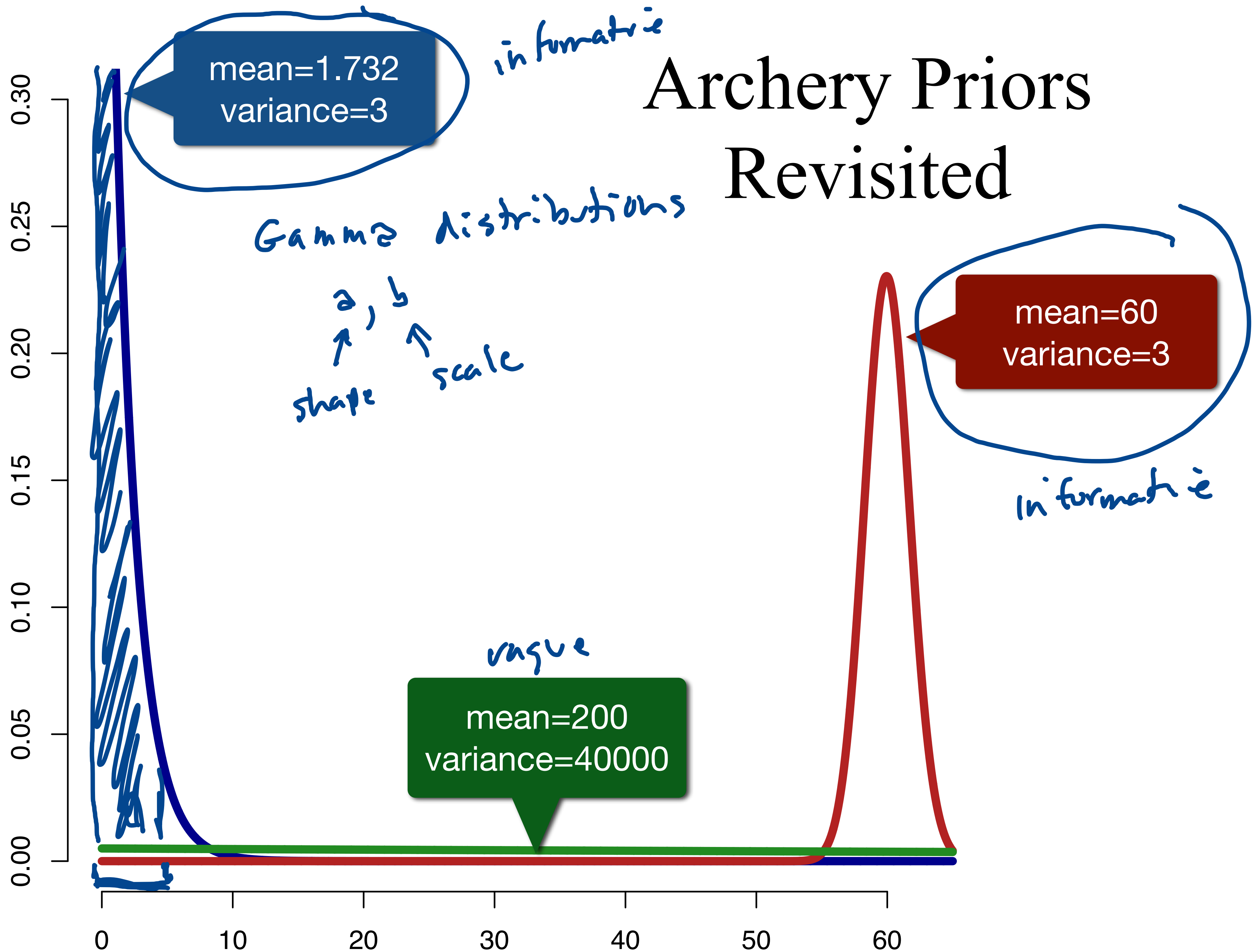
Integration of a probability density yields a probability (mass)



$$\int_0^2 f(\theta)d\theta$$

Area under the density curve from 0 to 2 is the probability that θ is less than 2

Archery Priors Revisited



Usually there are many parameters...

A 2-parameter example

Likelihood

Prior probability density

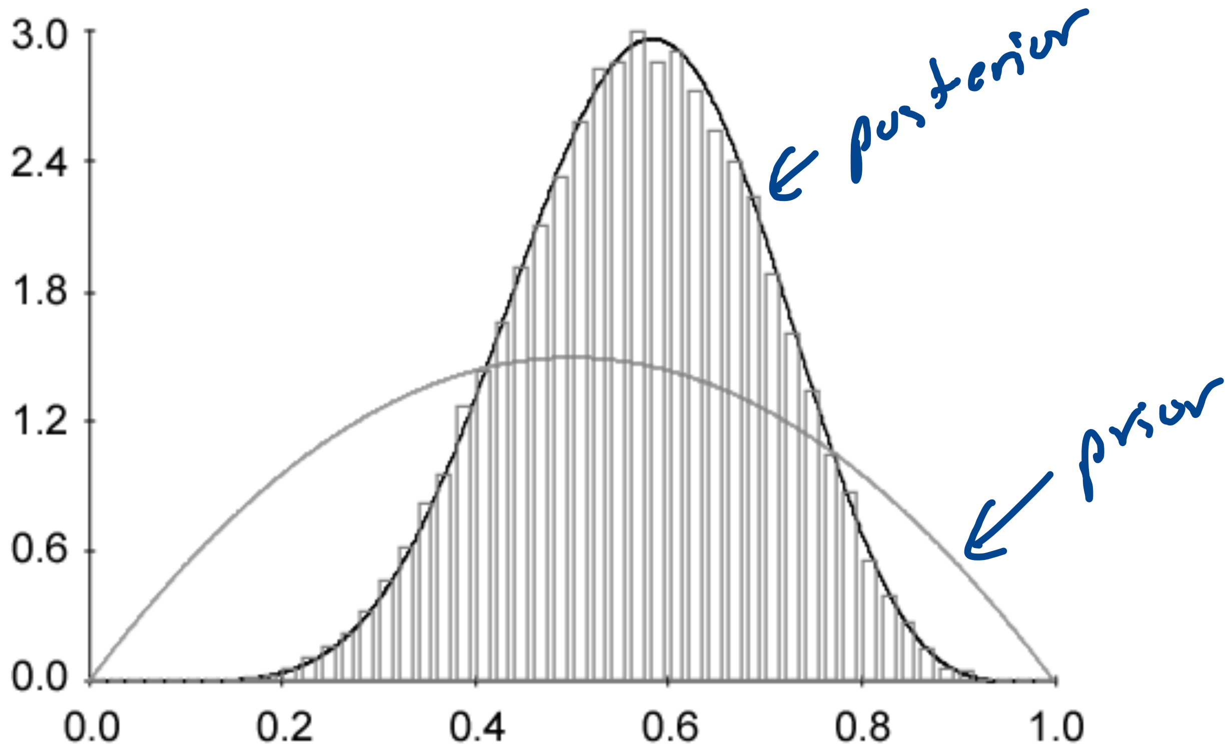
$$f(\theta, \phi | D) = \frac{f(D|\theta, \phi)\ f(\theta) f(\phi)}{\int_\theta \int_\phi f(D|\theta)\ f(\theta) f(\phi)\ d\theta d\phi}$$

Posterior 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...

Markov chain Monte Carlo (MCMC)



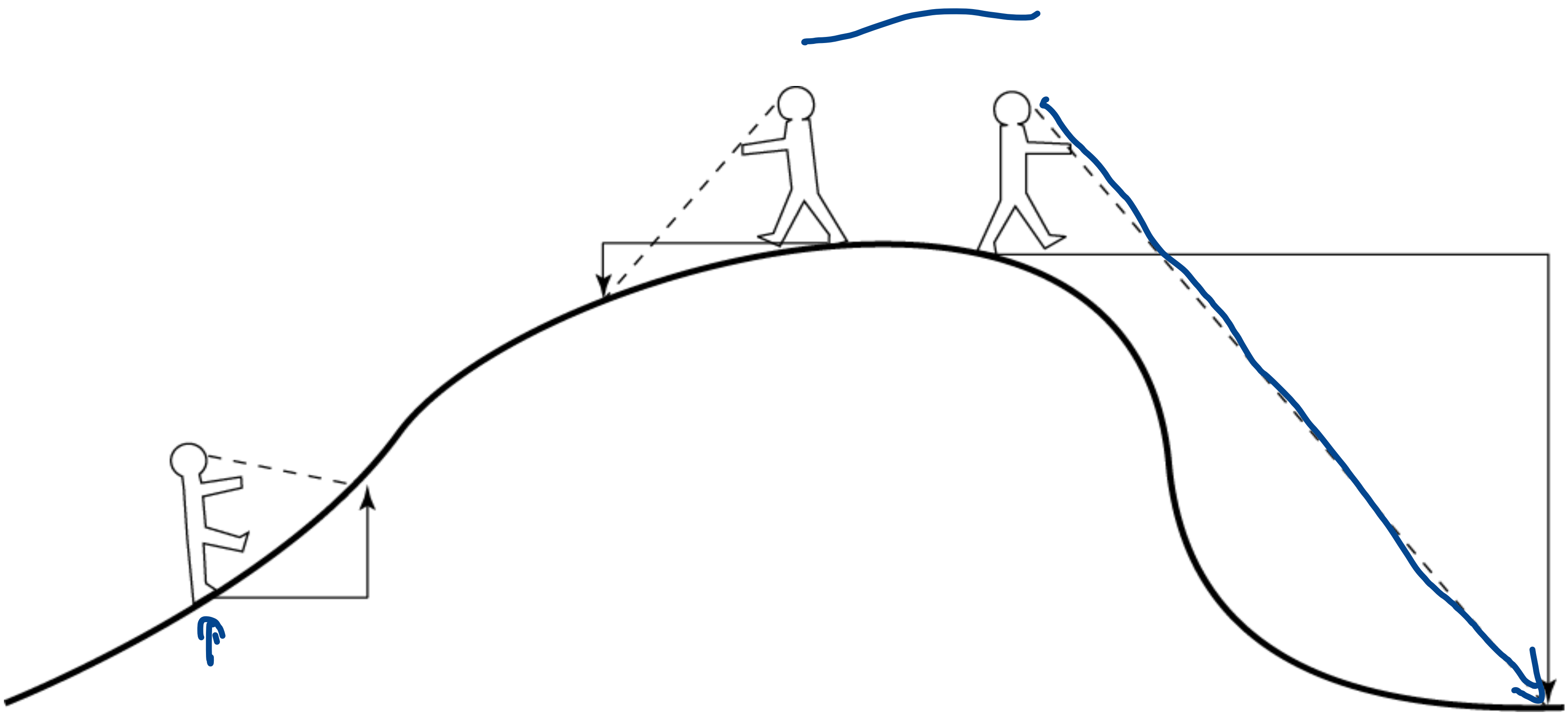
For more complex problems, we might settle for a

good approximation

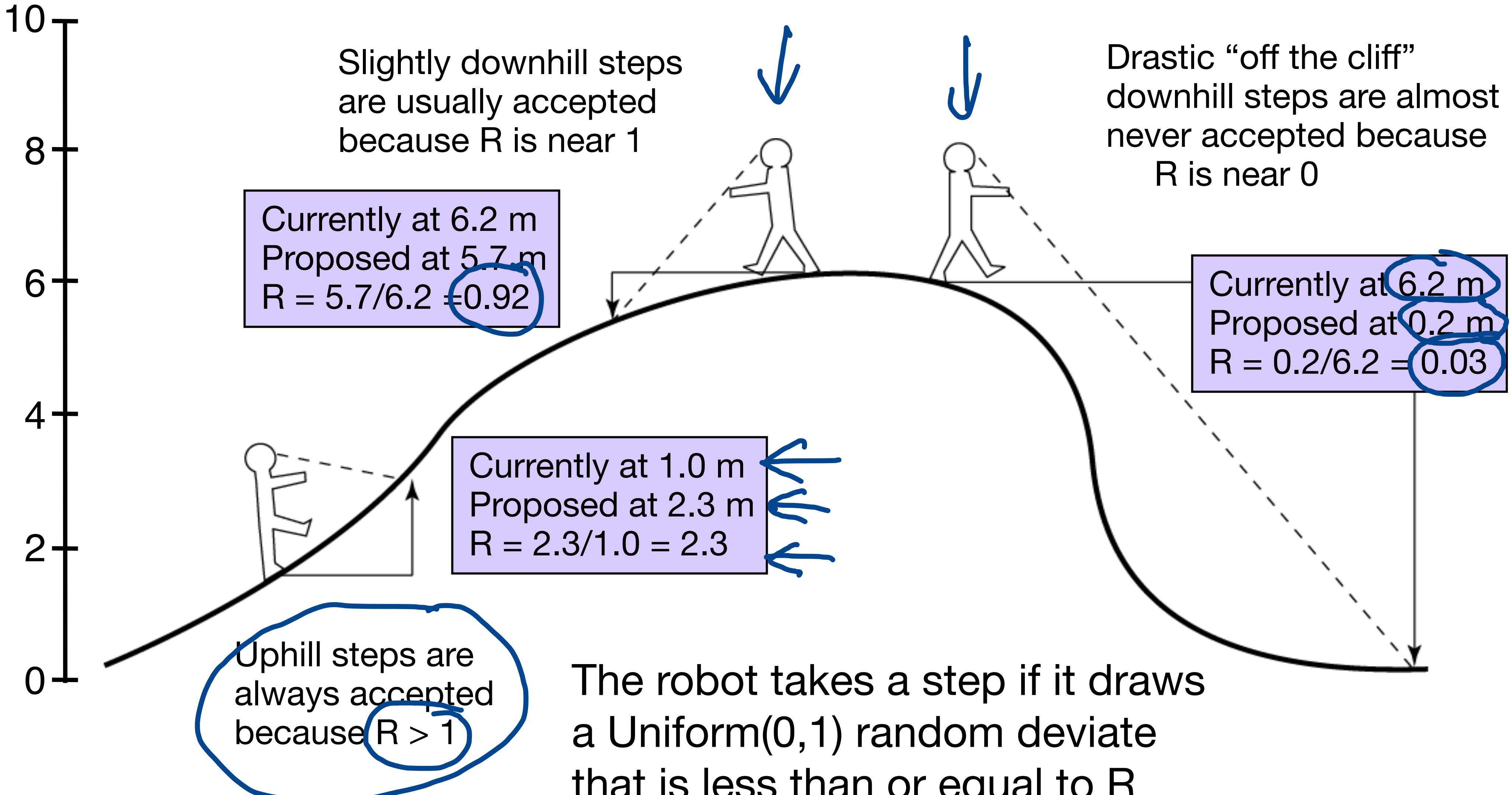
to the posterior distribution

STOPPED HERE 2024-02-20

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