DNA Identification: Bayesian Belief Update

Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA

TrueAllele® Lectures
Fall, 2010

Cybergenetics

Cybergenetics © 2003-2010

Prior Probability

$$\Pr\{X = x\}$$

What do we believe before we see any data?

Likelihood Function

$$\Pr\{data|X = x\}$$

How well does each hypothesis explain the data?

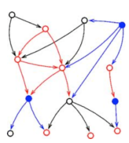
Posterior Probability

$$\Pr\left\{X = x \middle| data\right\}$$

What do we believe after we see the data?

Bayes Theorem

Rev Bayes, 1763



Computers, 1992

Bayes Theorem

Our belief in a hypothesis
after seeing data
is proportional to
how well that hypothesis explains the data
times our initial belief.

All hypotheses must be considered.
Need computers to do this properly.

Find the probability of causes by examining effects.

Bayesian Update

$$\Pr\left\{X = x \middle| data\right\} \propto \Pr\left\{data \middle| X = x\right\} \cdot \Pr\left\{X = x\right\}$$

Posterior probability — Likelihood function — Prior probability

Bayesian Update

$$\Pr\left\{X = x \middle| data\right\} \propto \Pr\left\{data \middle| X = x\right\} \cdot \Pr\left\{X = x\right\}$$

Posterior probability   Likelihood function   Prior probability

$$\Pr\left\{X = x \middle| data\right\} = \frac{\Pr\left\{data \middle| X = x\right\} \cdot \Pr\left\{X = x\right\}}{\sum_{x' \in X} \Pr\left\{data \middle| X = x'\right\} \cdot \Pr\left\{X = x'\right\}}$$

Consider all possibilities

Parameter Update

Bayes original example

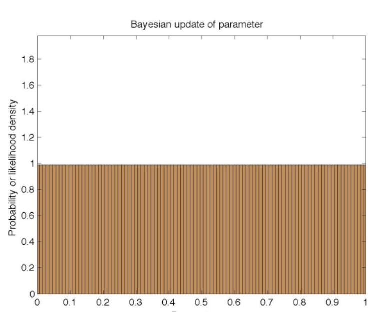
$$\Pr\left(X = x\right) \propto x^{a-1} \cdot (1-x)^{b-1}$$ Beta distribution

$$\Pr\left(k \middle| X = x\right) \propto x^{k} \cdot (1-x)^{n-k}$$ Binomial distribution

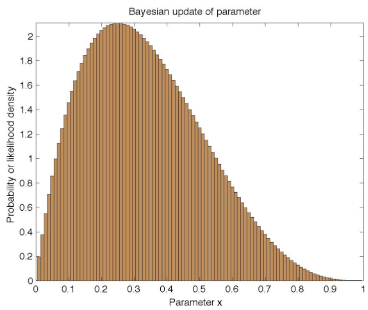
$$\Pr\left(X = x \middle| k\right) \propto \Pr\left(k \middle| X = x\right) \cdot \Pr\left(X = x\right)$$

$$\propto x^{(k+a)-1} \cdot (1-x)^{(b+n-k)-1}$$ Beta distribution

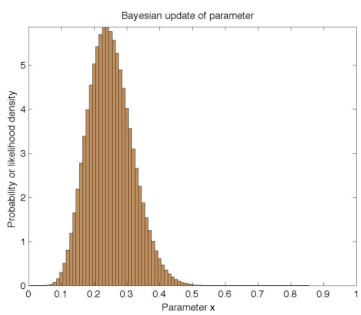
Beta distribution: uniform prior



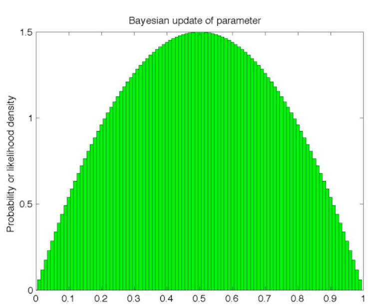
Weak prior probability of 1/3



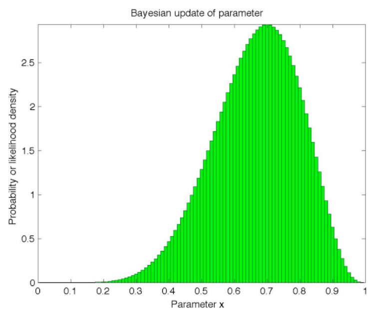
Strong prior around 1/4



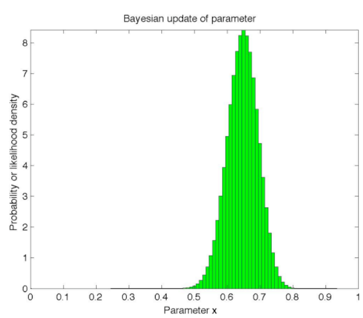
Binomial likelihood: 1 in 2



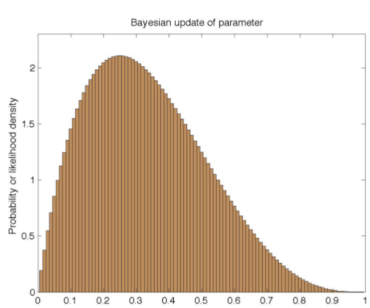
Observe: 7 counts in 10 trials



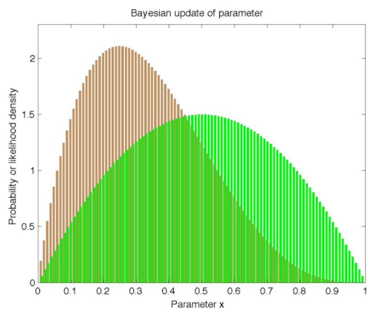
65 counts in 100 trials



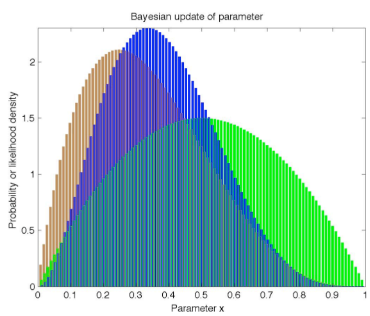
Initial belief around 1/3



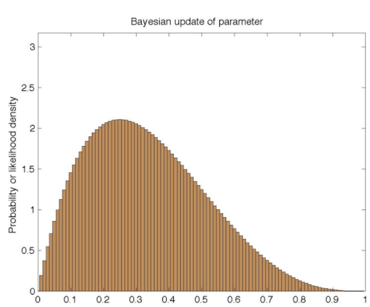
Observe 1 count in 2 trials



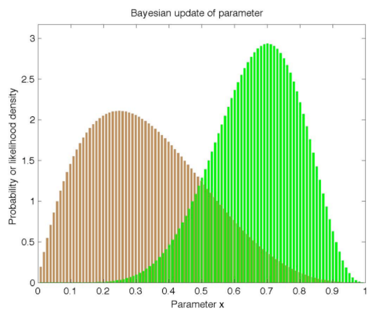
Final belief: posterior probability



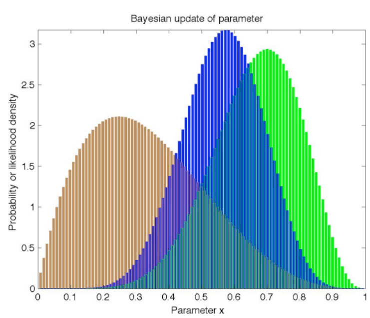
Initial belief (prior probability)



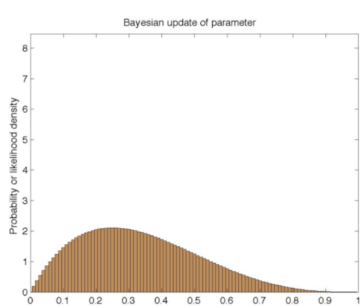
Data: 7 counts in 10 trials



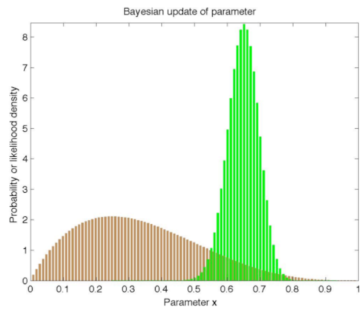
Final belief (posterior probability)



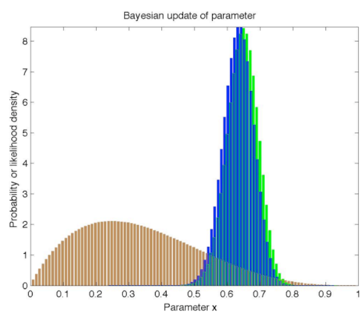
Weak prior probability (1/3)



Strong likelihood (65 in 100)



Strong posterior probability



Disease Prevalence

Free of Disease	Got the Disease
Pr(Free) 99.9% (= 100% - 0.1%)	Pr(Got) 0.1% 1 in a thousand

With a positive medical test result,
what is the chance of having this rare disease?

Medical Test: Likelihood

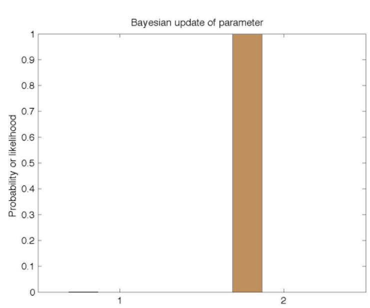
Data	Free of Disease	Got the Disease
Positive Test	False positive Pr(Pos | Free) 5% 5 in a hundred	True positive Pr(Pos | Got) 99% (= 100% - 1%)
Negative Test	True negative Pr(Neg | Free) 95% (= 100% - 5%)	False negative Pr(Neg | Got) 1% 1 in a hundred

Probability of Disease

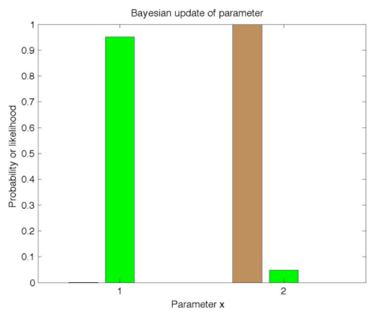
With a positive test (Pos),
what is the probability of having the disease (Got)?

$$\begin{aligned}\Pr(Got \mid Pos) &= \frac{\Pr(Pos \mid Got)\cdot\Pr(Got)}{\Pr(Pos \mid Got)\cdot\Pr(Got)+\Pr(Pos \mid Free)\cdot\Pr(Free)}\\ &= \frac{(99\%)\cdot(0.1\%)}{(99\%)\cdot(0.1\%)+(5\%)\cdot(99.9\%)}\\ &= \frac{0.1\%}{0.1\%+5\%}\\ &= \frac{0.1\%}{5.1\%}=\frac{1}{50}\\ &= 2\%\end{aligned}$$

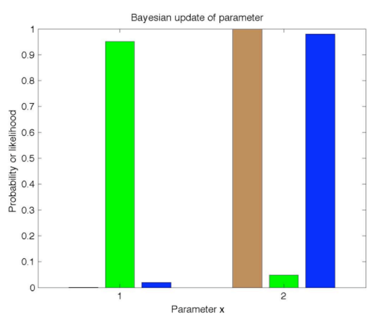
Prior: 99.9% disease free



Positive test: 99% true, 5% false



Posterior: 98% disease free



Odds of Disease

With a positive test (Pos),
what are the odds of having the disease (Got vs. Free)?

$$Posterior\ Odds = \frac{\Pr(Got|Pos)}{\Pr(Free|Pos)} = \frac{\Pr(Pos|Got)\cdot\Pr(Got)}{\Pr(Pos|Free)\cdot\Pr(Free)}$$

$$= \boxed{\frac{\Pr(Pos|Got)}{\Pr(Pos|Free)}} \cdot \frac{\Pr(Got)}{\Pr(Free)}$$

$$= \frac{99\%}{5\%} \cdot \frac{0.1\%}{99.9\%}$$

$$\boxed{Likelihood\ Ratio} \cdot Prior\ Odds$$

$$= 20 \cdot \frac{1}{1000}$$

$$= \frac{1}{50} = 2\%$$

Likelihood Ratio

Data	Free of Disease	Got the Disease
Positive Test	False positive Pr(Pos | Free) 5% 5 in a hundred	True positive Pr(Pos | Got) 99% (= 100% - 1%)
Negative Test	True negative Pr(Neg | Free) 95% (= 100% - 5%)	False negative Pr(Neg | Got) 1% 1 in a hundred