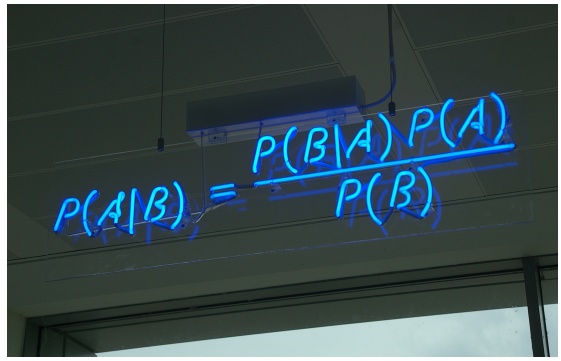
Bayesian Statistics

A Linear Progression

- Least Squares fitting with Linear Models
  - Estimate 'true' slope and intercept
  - State confidence in our estimate
  - Evaluate probabilty of obtaining data or more extreme data given a hypothesis
- Likelihood fitting with Linear Models
  - Estimate 'true' slope and intercept
  - State confidence in our estimate
  - Evaluate likelihood of data versus likelihood of alternate hypothesis

Bayesian Philosophy



- Estimate probability of a parameter
- State degree of believe in specific parameter values
- Evaluate probability of hypothesis given the data
- Incorporate prior knowledge
- Recognizes that Data is one realization of some parameter distribution

Let's Go Back to Probability!

Probability - The fraction of observations of an event given multiple repeated independent observations.

A Feeding Trial Example

Let's say you've offered 50 budworms a leaf to eat. 45 eat. P(eats) $= \frac{45}{50} = 0.9$

Two Events

We offer our budworms a leaf. 45 eat it. Then we offern them seconds. 20 of the original 45 each the second leaf.

P(eats twice) $= \frac{20}{50} = 0.4$

$= \frac{45}{50} * \frac{20}{45}$

P(A and B) = P(A)P(B)

Two Conditional Events

If we are interested in the probability of eating twice - i.e. the probability of eating a second time *given* that a budworm ate once, we phrase that somewhat differently.

$P(eats_2|eats_1)$

So, P(A and B) = $P(eats_2|eats_1)P(eats_1)$
Or, P(A and B) = $P(eats_1|eats_2)P(eats_2)$

Bayes Theorem

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

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

Rev. T. Bayes

GMO Collateral Damage

You had a rare but extremely harmful budworm munching your crops. You've developed a GMO tobacco leaf to help stop it. It has a 75% kill rate. Miraculously, it only has a 15% kill rate of non-budworms. Budworms make up about 10% of the insects in a field. What porportion of dead insects WON'T be budworms?

$$P(!W|D) = \frac{P(D|!W)P(!W)}{P(D)}$$

$$= \frac{P(D|!W)P(!W)}{P(D|W)P(W)+P(D|!W)P(!W)} = \frac{0.15*0.9}{0.75*0.1+0.15*0.9} = 0.643$$

The majority of the dead!

Denominator: The Marginal Distribution

$$p(A|B) = \frac{p(B|A)P(A)}{\sum_{i=0}^{j} p(B|A)p(A)}$$

Essentially, all alternate hypotheses?

Denominator - marginal distribution - becomes an integral of likelihoods if B is continuous - i.e. fitting a particular parameter value. It normalizes the equation to be between 0 and 1.

Why are we talking about this??

In a *frequentist* framework, we talk about the probability of observing data given that a hypothesis is true

$$P(Data|Hypothesis)$$

In a *Bayesian* framework,we talk about the probability of a hypothesis given the data

$$P(Hypothesis|Data)$$

.

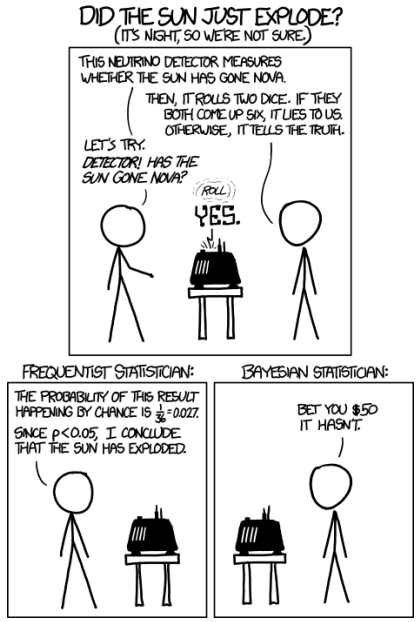
Bayes Theorem

$$p(Hypothesis|Data) = \frac{P(Data|Hypothesis)p(Hypothesis)}{p(Data)}$$

Bayes Theorem

$$p(\theta|D) = \frac{p(D|\theta)P(\theta)}{p(D)}$$

Bayes Theorem in Action



http://xkcd.com/1132/

Bayes Theorem in Action

$$p(SunExplodes|Yes) = \frac{p(Yes|SunExplodes)p(SunExplodes)}{p(Yes)}$$

We know/assume:
p(Sun Explodes) = 0.0001, P(Yes | Sun Explodes) = 35/36

We can calculate:
p(Yes) = P(Yes | Sun Explodes)p(Sun Explodes) + P(Yes | Sun Doesn't Explode)p(Sun Doesn't Explodes)
= 35/36 * 0.0001 + 1/36 * 0.9999 = 0.0277775

credit: Amelia Hoover

Bayes Theorem in Action

$$p(SunExplodes|Yes) = \frac{p(Yes|SunExplodes)p(SunExplodes)}{p(Yes)}$$

$$p(SunExplodes|Yes) = \frac{0.0001 * 35/36}{0.028} = 0.0035$$

Incorporating Prior Information about the Sun Exploding gives us a *very* different answer

Note, we can also explicitly evaluate the probability of an alternate hypothesis - p(Sun Doesn't Explode | Yes)

But what about for a parameter?

Fitting a Parameter

Let's assume:

- Data from a normal distribution with mean 2, and SD 1.
- A flat prior N(0,1000)
- We're only estimating a mean

$$P(Mean = \theta | Data) = \frac{P(Data|Mean = \theta)P(\theta)}{\int P(Data|Mean = \theta)P(\theta)}$$

Note: Numerator = Likelihood * Prior!

Fitting a Parameter

```
#the data
x <- rnorm(30, mean=2)

#mean possibilities between -6 and 6
meanParam <- seq(-6,6,.001)
```

```
#A function to get a prior given a choice of a parameter
#Defaults to flat prior
getPrior <- function(x, mean=0, sd=10000000) dnorm(x, mean, sd)
```

```
#a function to get a p(data | mean) p(mean)
getProd <- function(data, meanParam, sdParam=1)
  prod(dnorm(data, meanParam, sdParam)) * getPrior(meanParam)
```

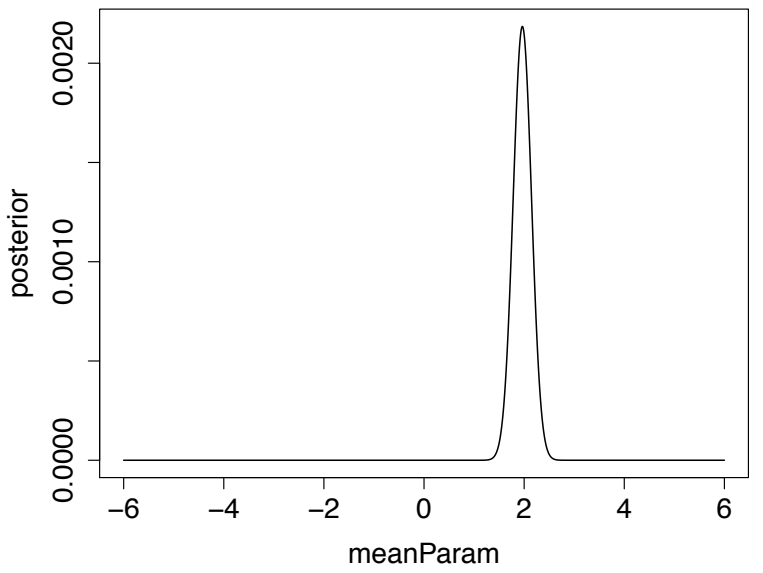
Fitting a Parameter

```
#Make the numerator of Bayes Rule
numerators <- sapply(meanParam, function(m) getProd(x,m))
```

```
#summed, that's your denominator
marginalDist <- sum(numerators)
```

```
#Get the posterior probability
posterior <- numerators/marginalDist
```

Fitting a Parameter



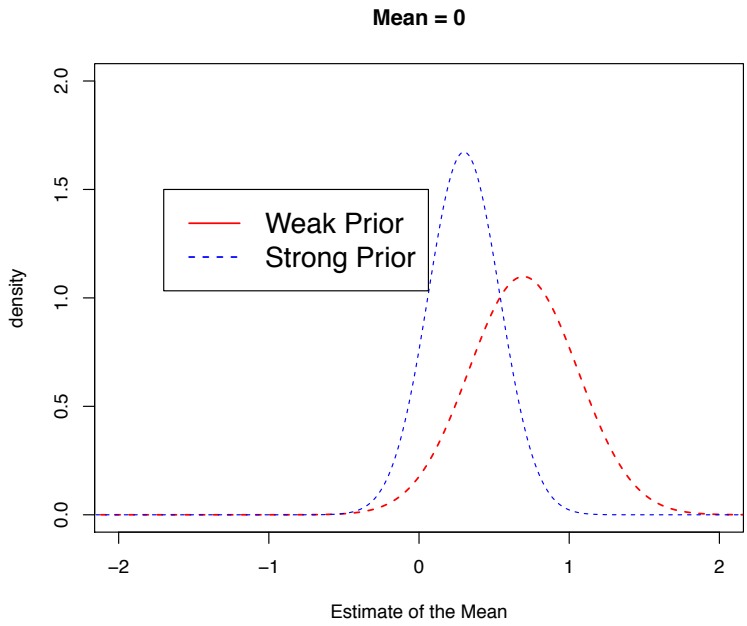
Priors & Posterior Credible Intervals

How do we Choose a Prior?

- A prior is a powerful tool, but it can also influence our results of chosen poorly. This is a highly debated topic.
- Conjugate priors make some forms of Bayes Theorem analytically solveable
- If we have objective prior information - from pilot studies or the literature - we can use it to obtain a more informative posterior distribution
- If we do not, we can use a weak or flat prior (e.g., N(0,1000)). Note: constraining the range of possible values can still be weakly informative - and in some cases beneficial

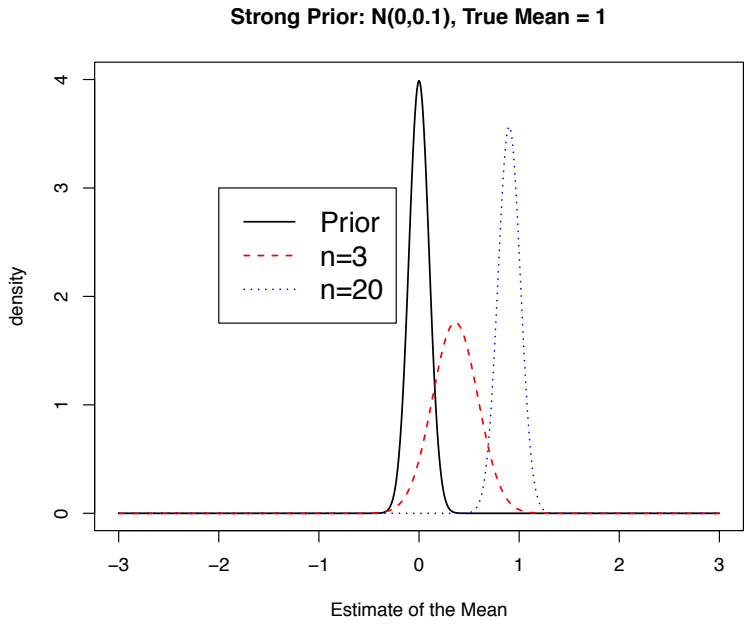
The Influence of Priors

Here's the posterior distribution drawn using the same sample - but in one case with a weak prior, and one a strong prior.



Priors and Sample Size

The influence of priors decreases with same size. A large sample size 'overwhelms' the prior.



Evaluation of a Posterior: Frequentist Confidence Intervals

In Frequentist analyses, the **95% Confidence Interval** of a parameter is the region in which, were we to repeat the experiment an infinite number of times, the *true value* would occur 95% of the time. For normal distributions of parameters:

$$\hat{\beta} - t(\alpha, df)SE_\beta \leq \beta \leq \hat{\beta} + t(\alpha, df)SE_\beta$$

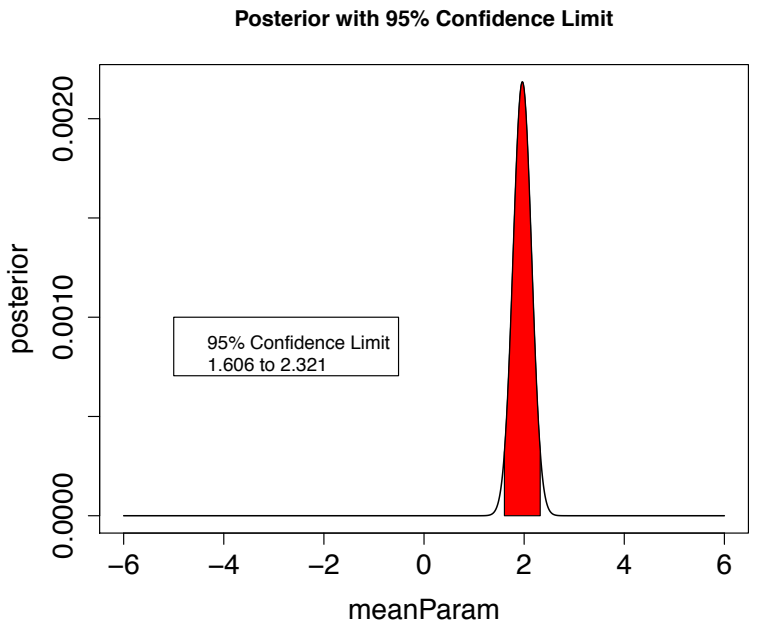
Evaluation of a Posterior: Bayesian Credible Intervals

In Bayesian analyses, the **95% Credible Interval** is the region in which we find 95% of the possible parameter values. The observed parameter is drawn from this distribution. For normally distributed parameters:

$$\hat{\beta} - 2 * \hat{SD} \leq \hat{\beta} \leq \hat{\beta} + 2 * \hat{SD}$$

where $\hat{SD}$ is the SD of the posterior distribution of the parameter β. Note, for other types of parameters, the distribution may be different.

Evaluation of a Posterior: Bayesian Credible Intervals



Discussion

So, Bayes?

Priors, Credible Intervals, $P(H|D)$ v. $P(D|H)$