Bayesian Perspective on Catch Comparison Analysis

Jon Brodziak
Northeast Fisheries Science Center

Bayesian Perspective on Catch Comparison Analysis

- Bayesian and Frequentist Perspectives
- Prior Distribution
  - Uninformative
  - Informative
- Likelihood
- Posterior Distribution and Measuring Effect Size
  - Paired Two Sample T-test
  - Winbugs Program Paired_T_test.odc

What is Bayesian Inference?

- Information available before study is conducted is summarized in a quantitative model (hypothesis): the prior probability distribution
- Use Bayes' rule to construct a posterior probability distribution based on the prior and the likelihood of the data collected in the study
- The posterior probability distributions provides a direct measure of the degree of probability (belief) that can be placed on parameters (estimates, hypotheses, or models).

Classical Frequentist Inference

- Estimate probability of the data given a hypothesis Pr(D|H)
- Probability is defined as the long-run relative frequency of events
- Inference is based solely on the likelihood of the sample data
- Model parameters are estimates of fixed “true” quantities

Bayesian Inference

- Estimate probability that hypothesis is true given the data $Pr(H|D)$
- Probability is defined as an individual's degree of belief in the likelihood of an event
- Inference is based on prior knowledge and the likelihood of the sample data
- Model parameters are random variables (e.g., probability distributions)

Joint probability distribution for model parameters (θ) and data (D), p(θ,D), depends on the prior distribution of model parameters, p(θ), and the likelihood of observing the data, p(D|θ)📪L(D|θ)

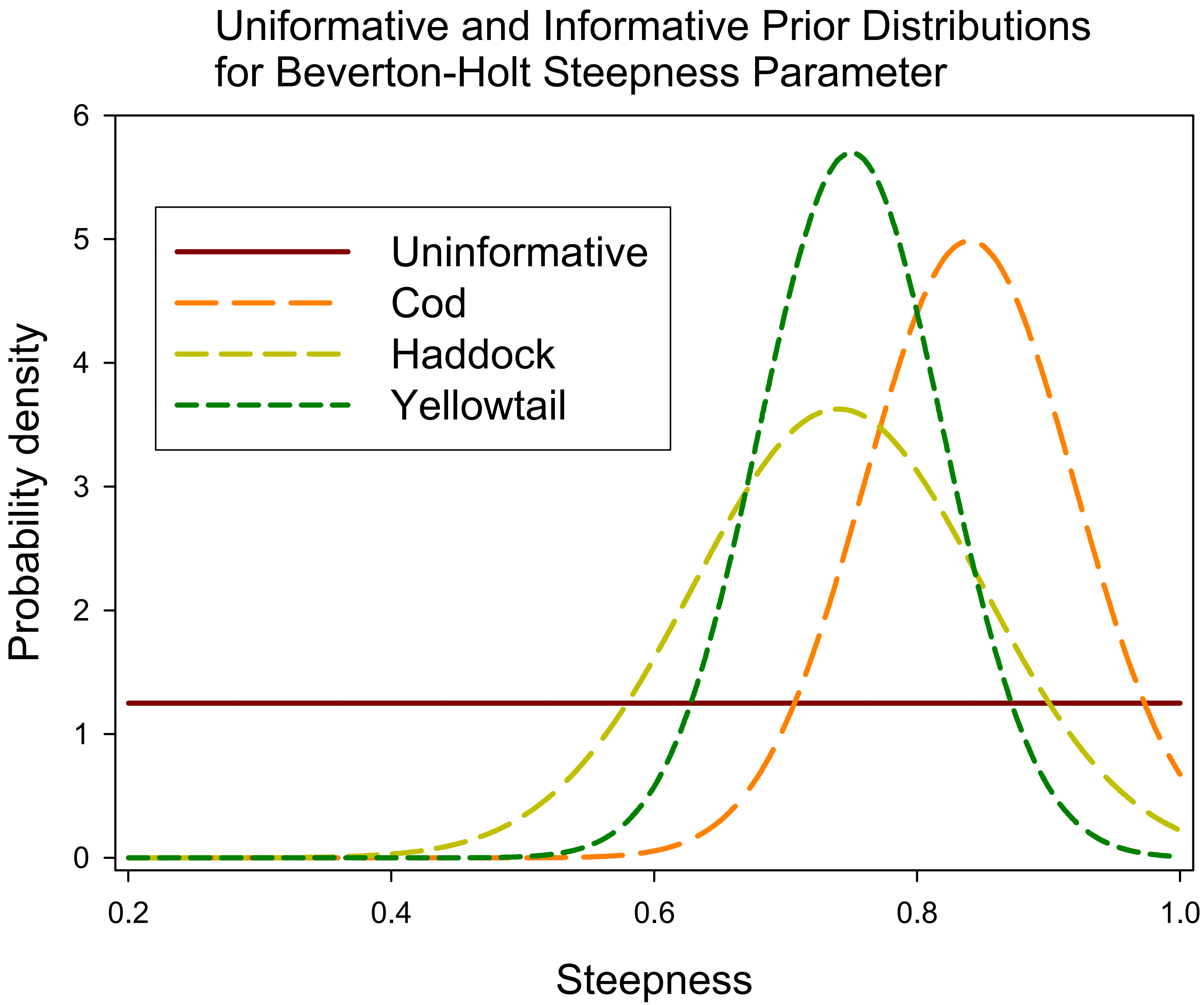
$$p(\theta, D) = p(\theta)\, p(D|\theta)$$

Prior Distribution, p(🗎)

- Often prior information or knowledge are available about a parameter before an experiment is conducted
- Uninformative priors are used when there is little or no information available beforehand
- Informative priors are used when there is existing information on a parameter value before the experiment is conducted
- Using uninformative priors implies that virtually all information about a parameter is contained in the likelihood of the sample data

An Example of Alternative Priors

$$a = \frac{4 \cdot z}{(1 - z) \cdot SPR(0)}$$



Likelihood, p(D|🗎)

- When confronting models with data we usually want to know how well the data support alternative hypotheses
- After the experiment, the data are known, but the best-supported hypothesis is not known
- We want to know the answer to “Given these data, how likely are the possible outcomes?”
- To do this, we introduce the concept of likelihood, L, to measure the likelihood of observing the data given the hypothesis

Distinction between Probability and Likelihood

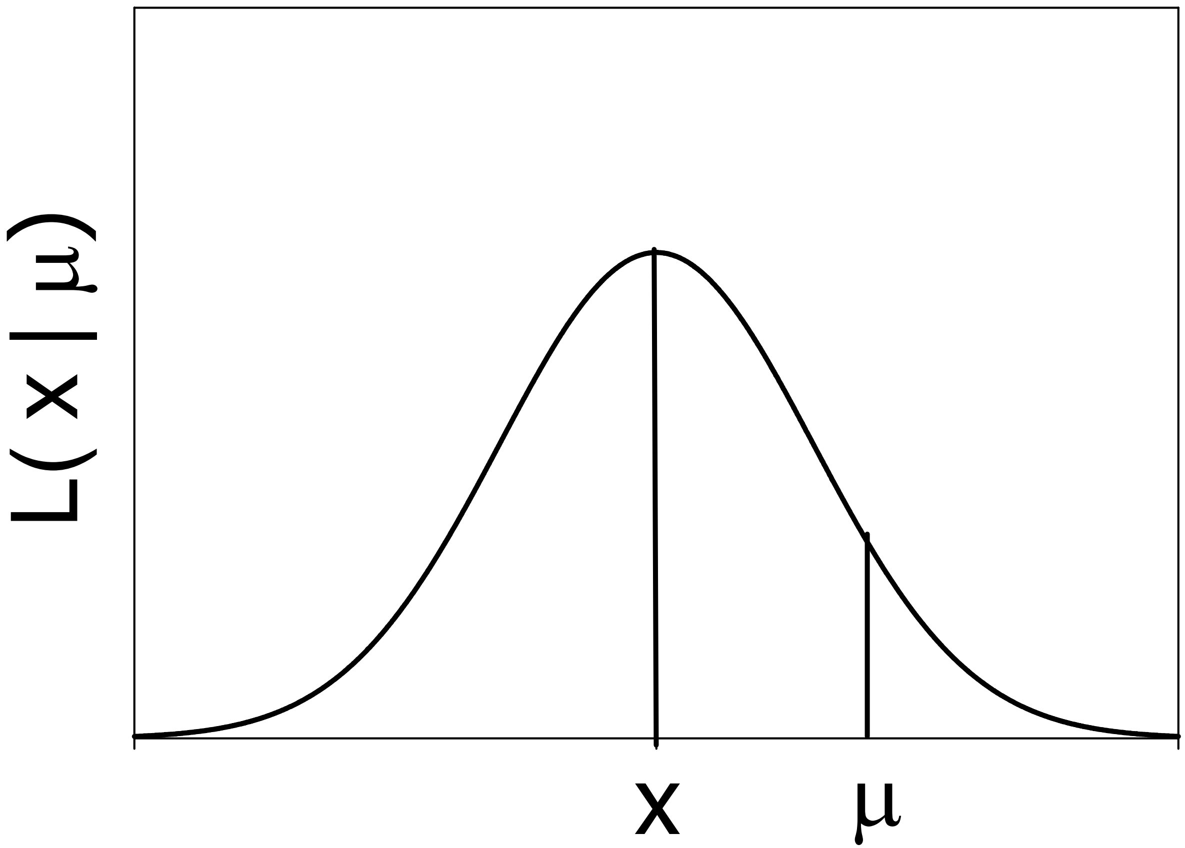
- Probability: Hypothesis is known, but the data are not known
- Likelihood: Data are known, but the (best) hypothesis is not known

- Bayesian interpretation: Both data and parameters are random variables

Likelihood Principle

- Likelihood of the data given the hypothesis is assumed to be proportional to the probability of the data given the hypothesis
- L(x| μ, ☞^{2}) = c·Pr(DATA | HYPOTHESIS)
- All relevant information about the hypothesis (parameter value) is contained in the likelihood function

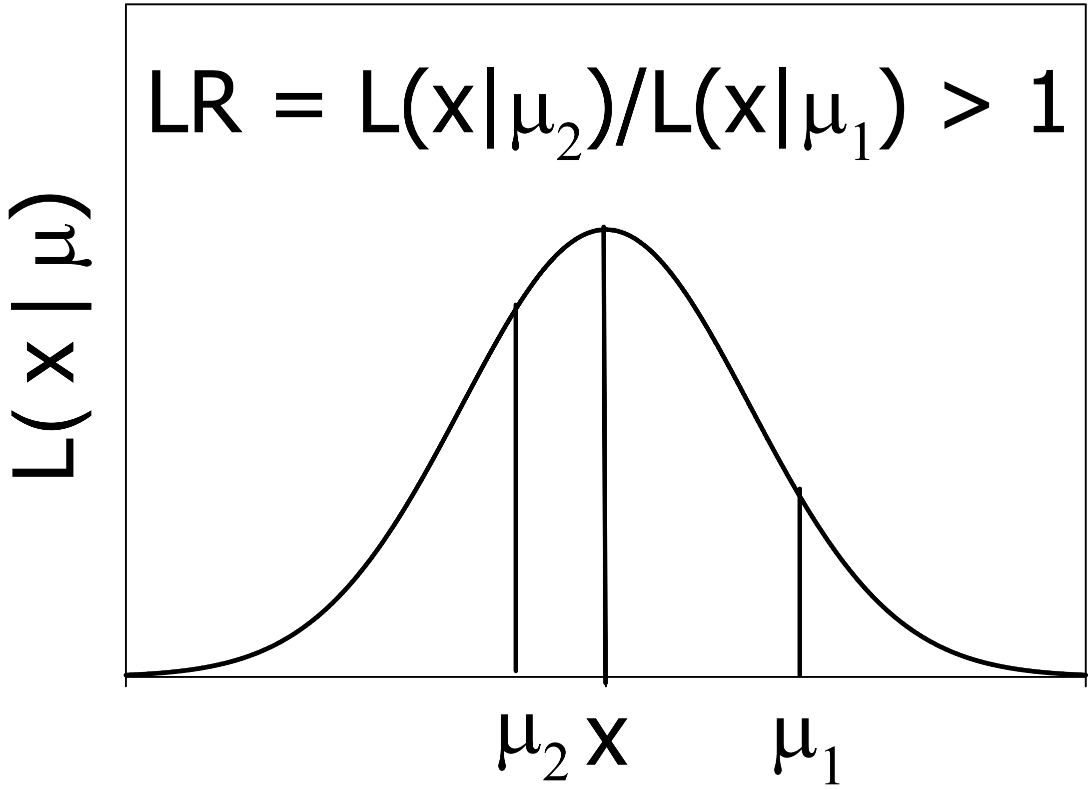
Likelihood of observed data x given hypothesis (μ,σ^2)



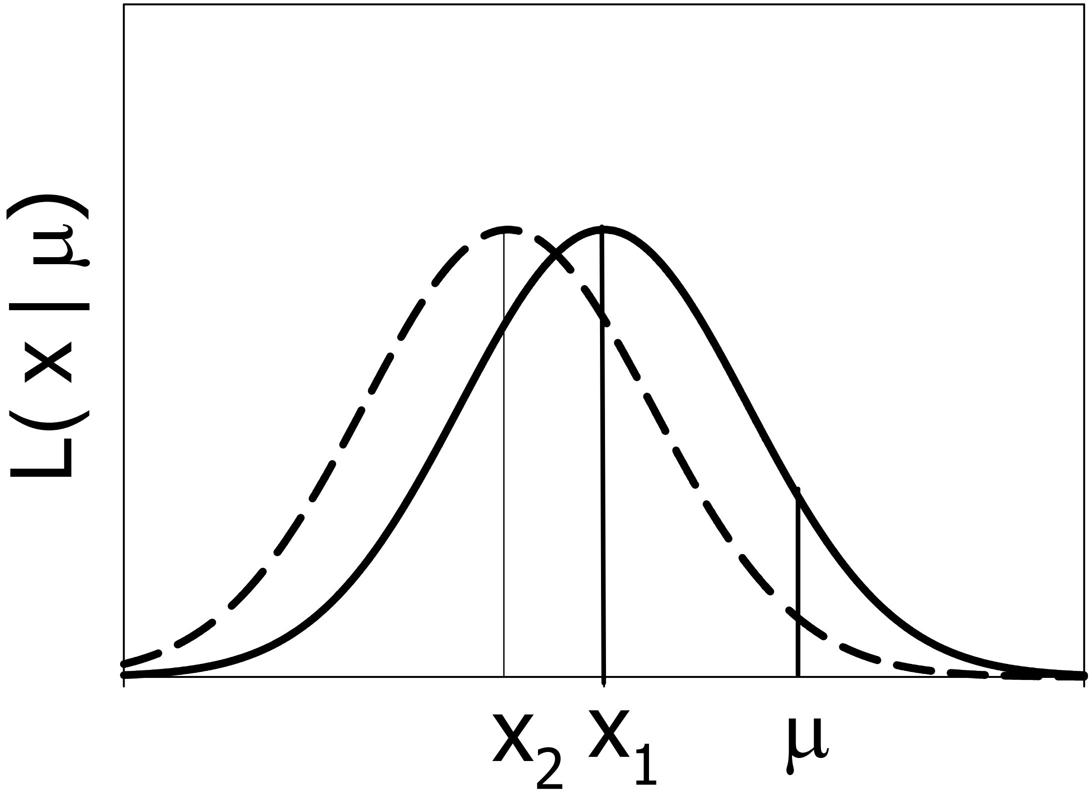
Likelihood is Relative

- How likely is one hypothesis (H_0) versus another (H_1)?
- Use the ratio of likelihoods (LR) given hypothesis to measure the relative likelihood of the hypotheses
- $LR = L(x| H_1)/L(x| H_0)$

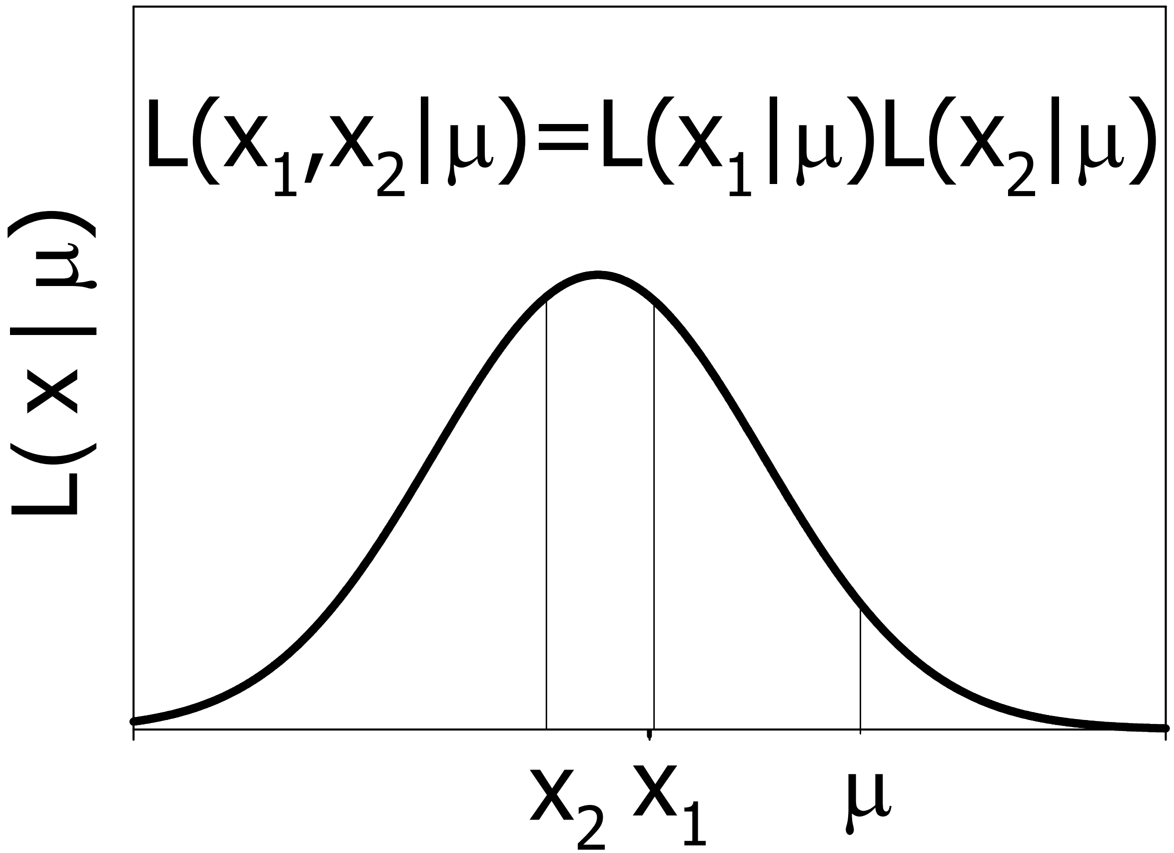
Likelihood ratio of observed data x comparing hypothesis (μ_1,σ^2) with hypothesis (μ_2,σ^2)



Likelihoods of observed data points x_1 and x_2 given hypothesis (μ,σ^2)



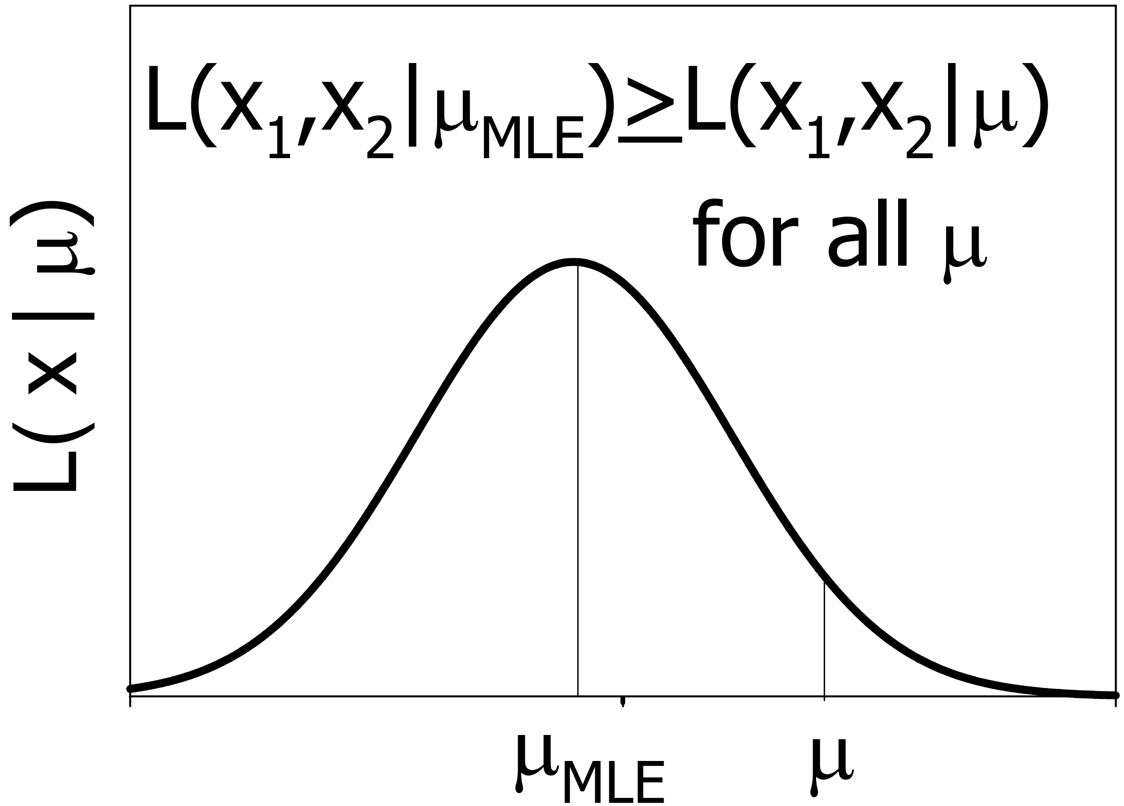
Joint likelihood of observed data x_1 and x_2 given hypothesis (μ,σ^2)



What is the best hypothesis?

- Use likelihood to measure degree of agreement between hypotheses and data
- Maximization of the likelihood function is a way to estimate the best parameters
- Maximizing the likelihood function optimizes the use of information included in data

Maximum likelihood estimate μ_{MLE} of μ maximizes the joint likelihood of observing x_1 and x_2



Posterior Distribution, $p(\theta|D)$ and Measuring Effect Size

Bayes' rule gives the conditional probability of model parameters given the data; the result is $p(\theta|D)$, the posterior distribution

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

If the parameter has a continuous density, p(D) is an integral over all possible θs

$$p(D) = \int p(\theta)p\left(D|\theta\right)d\theta$$

If the parameter has a discrete distribution, p(D) is the sum over all possible θs

$$p(D) = \sum_{\theta} p(\theta)p\left(D|\theta\right)$$

Since p(D) is a constant, the posterior distribution is proportional to the prior times the likelihood

$$p(\theta|D) \propto p(\theta)\, p(D|\theta)$$

How can we calculate the posterior distribution?

Apply Markov Chain Monte Carlo (MCMC) simulation to draw samples θ^t for t=1, 2, 3, ..., from the posterior distribution

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

MCMC simulates a random walk in parameter space that converges to a stationary distribution that is exactly the posterior distribution

Paired Two Sample T-Test

Collect n paired samples from populations X and Y $\{(X_1,Y_1),..,(X_n,Y_n)\}$

Compute the mean difference between the paired observations.

$$\overline{D} = \frac{1}{n}\sum_i \left(X_i - Y_i\right)$$

Paired Two Sample T-Test

Compute the standard deviation of the difference.

$$S_D = \sqrt{\frac{n\sum_i D_i^2 - \left(\sum_i D_i^2\right)^2}{n(n-1)}}$$

Compute the standard error of the mean difference

$$S_{\overline{D}} = \frac{S_D}{\sqrt{n}}$$

Paired Two Sample T-Test

Compute the t-statistic

$$t = \frac{\overline{D}}{S_{\overline{D}}}$$

Reject H_0 if $t > T_{n-1,\,\alpha/2}$ or $t < -T_{n-1,\,\alpha/2}$

Expectation of Posterior Distribution

Posterior expectation of t-statistic

$$E\left[t \,|\, D,\ M_j\right] = \int t(\theta)\cdot \Pr\left(\theta | D\right) d\theta$$

Generate samples 🖹^{1}, 🖹^{2}, 🖹^{3} … from Pr(🖹 | D) using Markov Chain Monte Carlo simulation to compute posterior expectations of t-statistic and difference

$$E\left[t \,|\, D\right] \approx \frac{1}{N - M} \sum_{T=M+1}^{N} t\left(\theta^T\right)$$

Anatomy of a Winbugs Program

- Specify prior(s)
- Specify likelihood
- Calculate parameters of interest

- Example Winbugs 1.4 program: Normal_Paired_T_test.odc

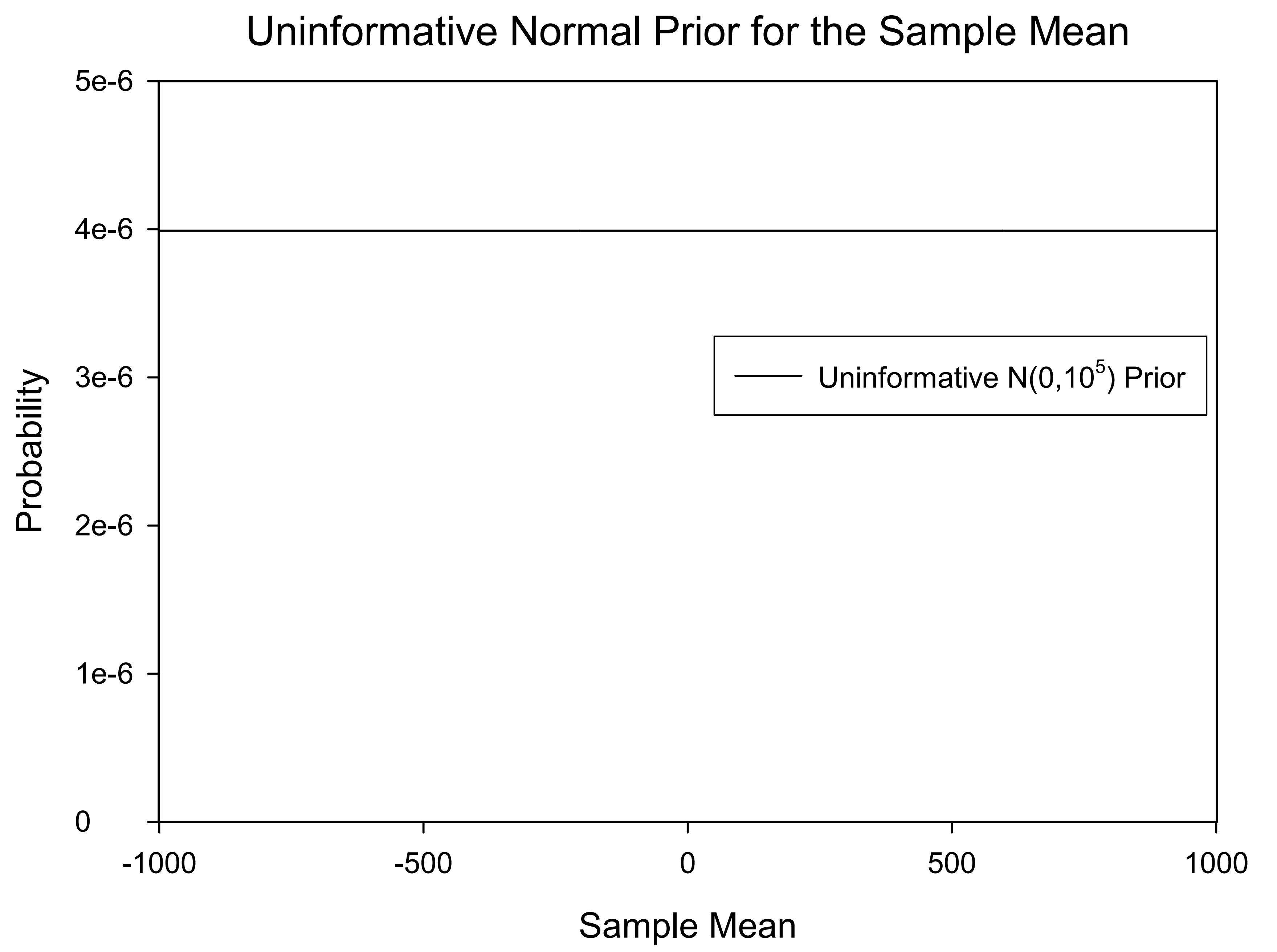
- Winbugs can be downloaded from:

http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

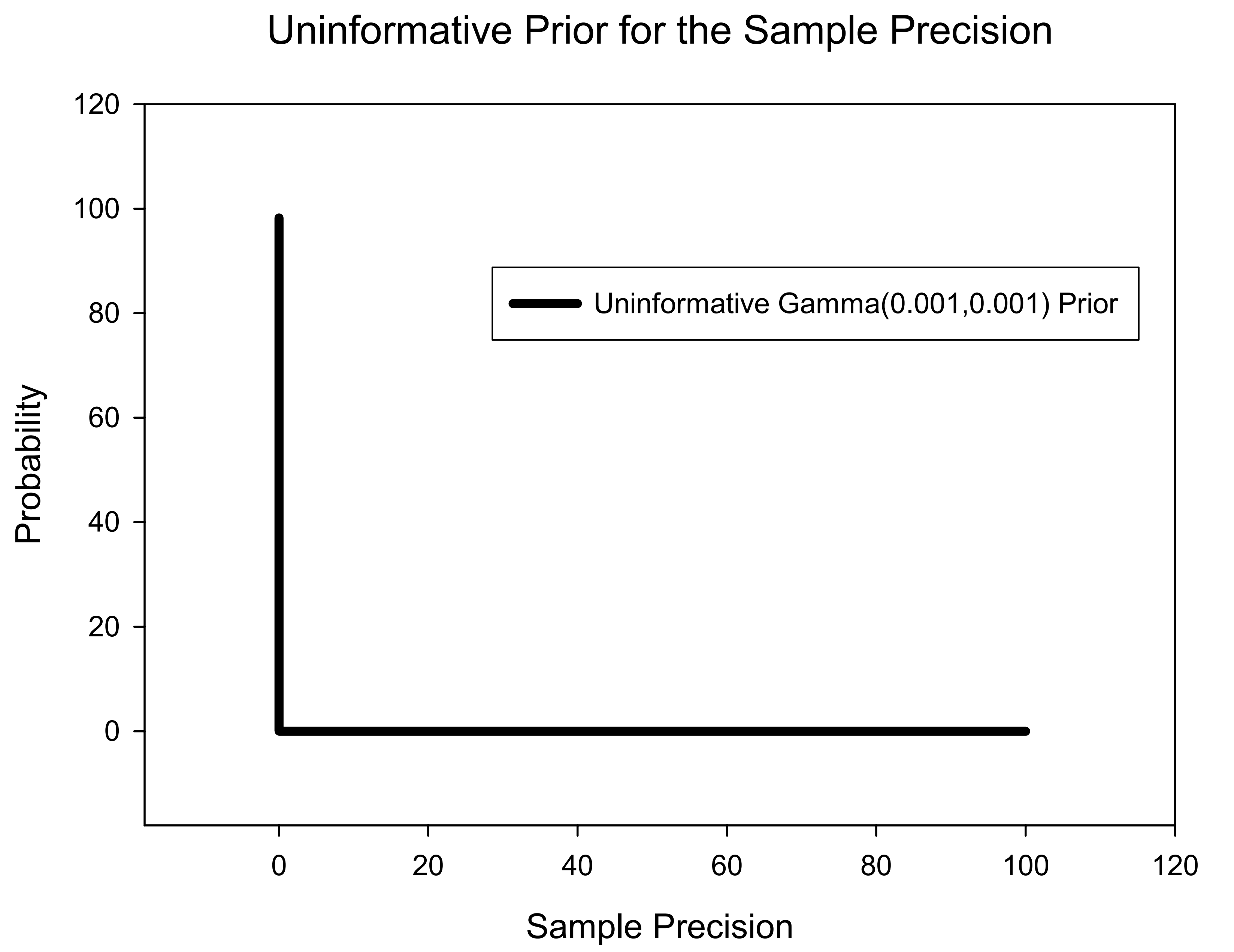
Running a Winbugs program

- Specify model
- Load data
- Compile program for one or more chains (parallel simulations)
- Load initial parameter values for each chain
- Configure model update tool
- Configure sampling tool
  - Choose variables to monitor
  - Choose percentiles to output

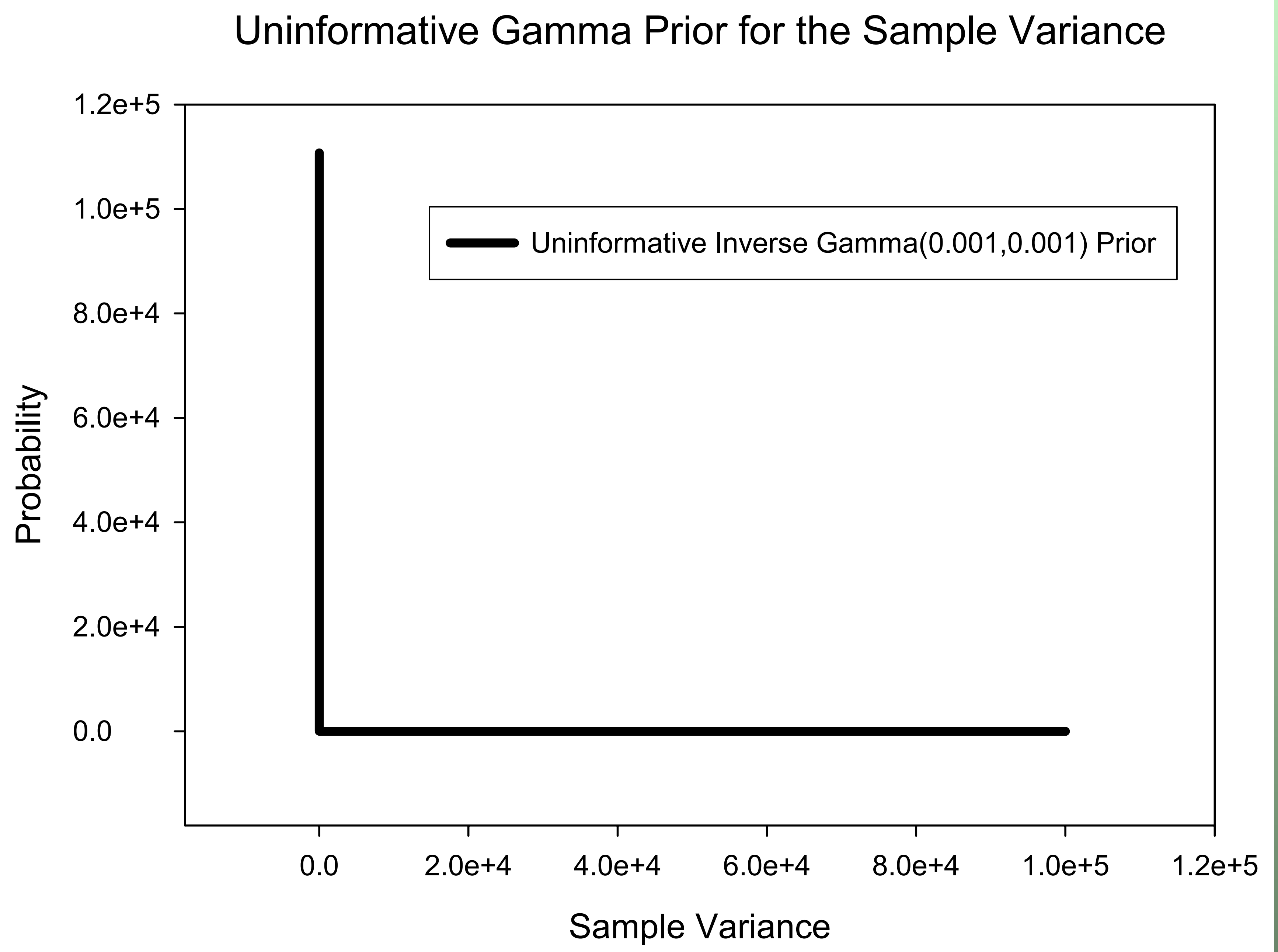
An Uninformative Prior



Uninformative Prior for Precision



Uninformative Prior for Variance



Uniform Prior for Sample Variance