Model building
and
Modes of Inference:

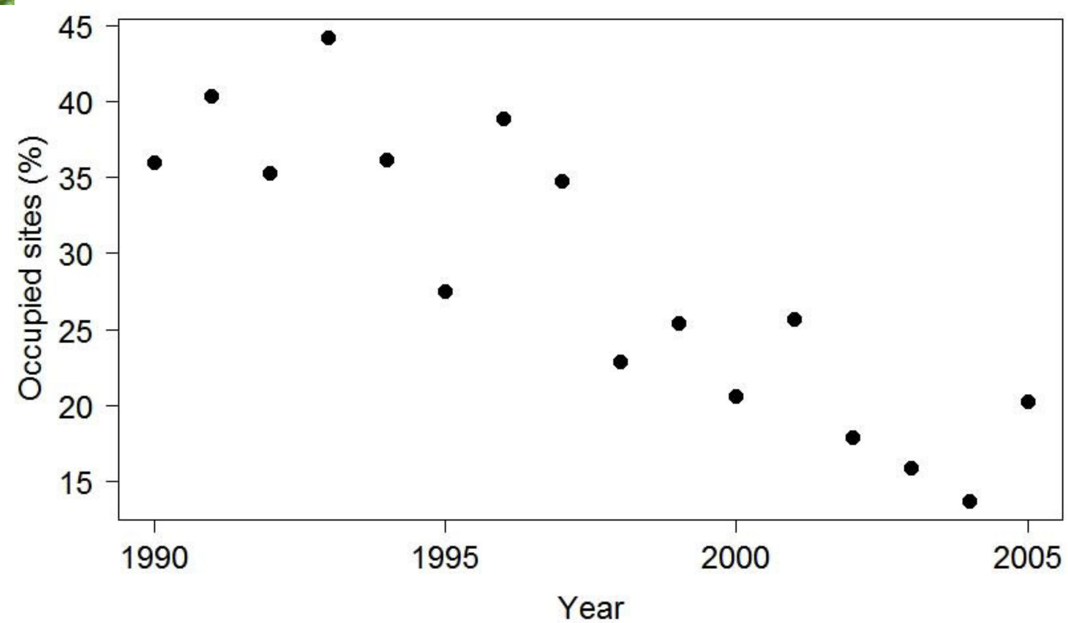
A brief review

Suggested reading: Bolker Chapter 1

Outline

- Role of models in science
  - Mathematical vs statistical models
- Analysis of models
  - Frequentist approaches
  - Bayesian approaches

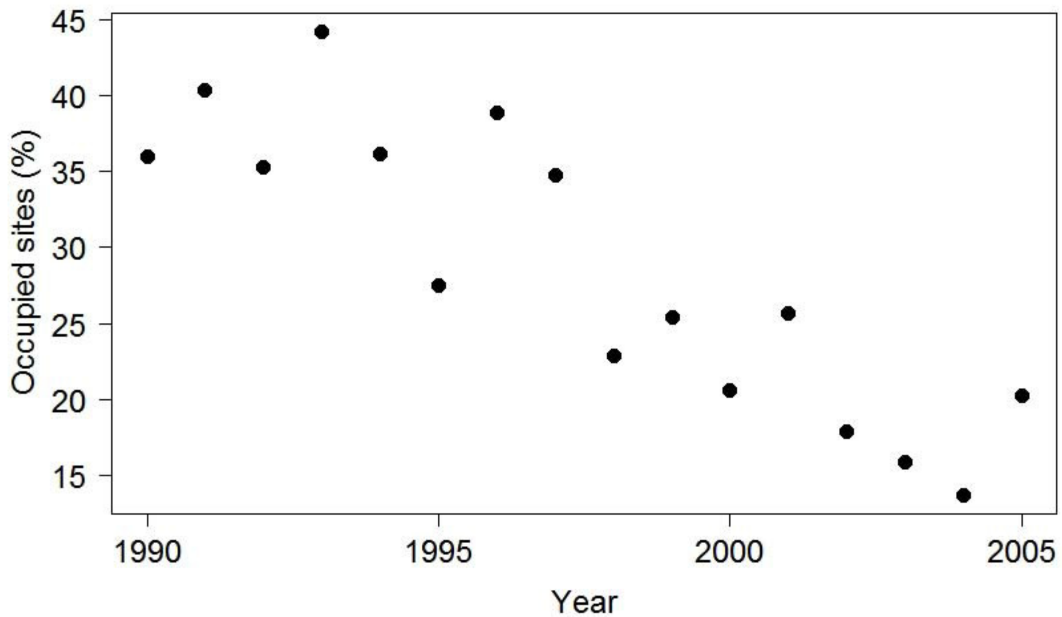
A simple example of a model



A simple example of a model

$$y = mx + b + \varepsilon$$

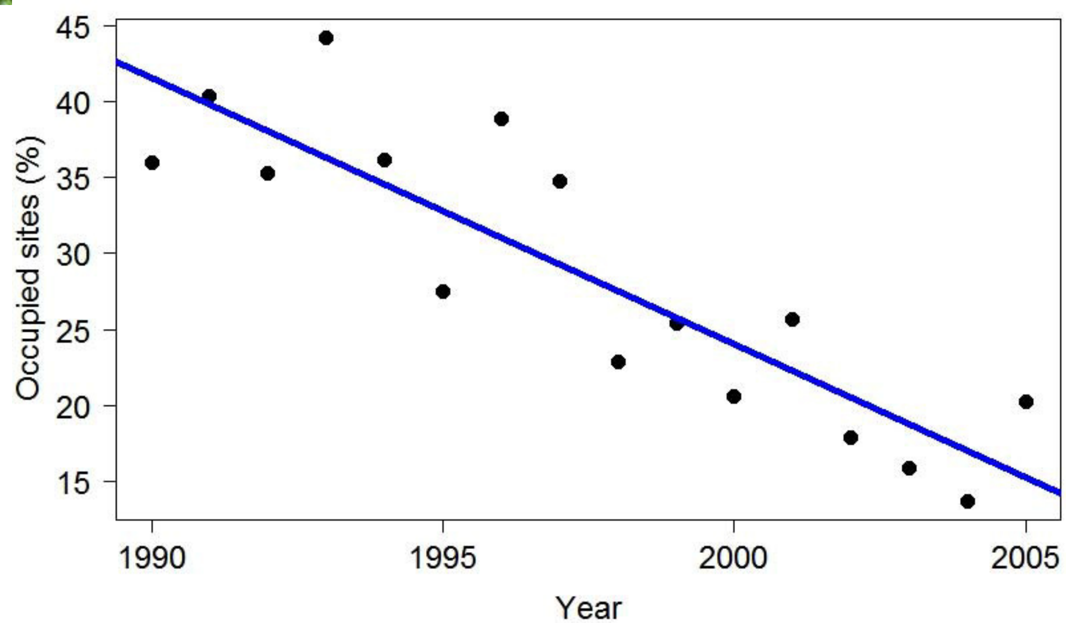
$$\varepsilon \sim Norm(0, \sigma^2)$$



A simple example of a model

Trend estimate

m = -1.754



$$y = mx + b + \varepsilon$$

$$\varepsilon \sim Norm(0, \sigma^2)$$

A simple example of a model

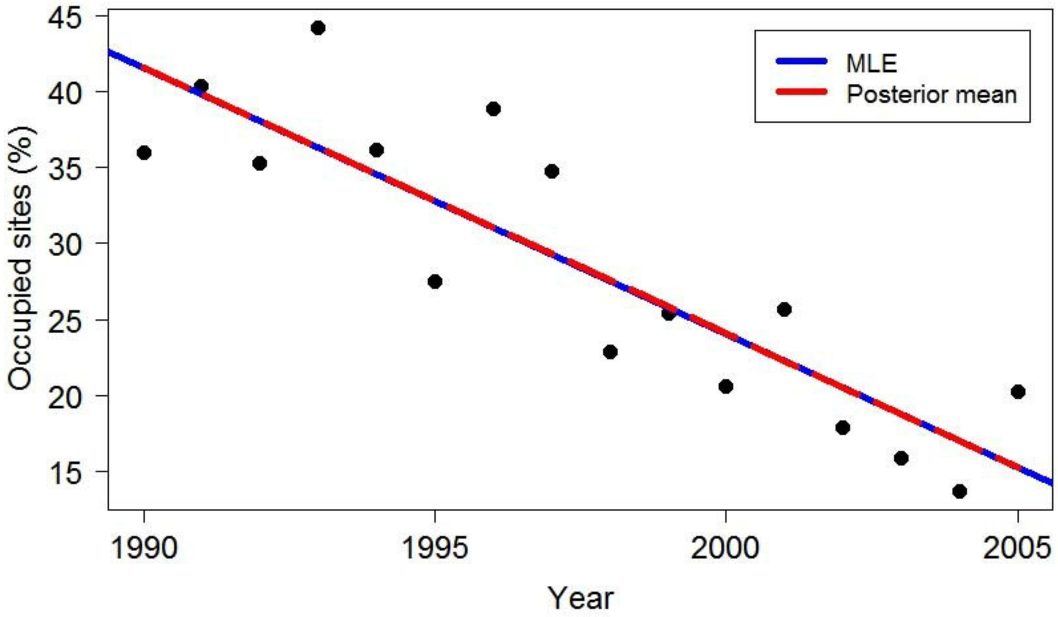
Trend estimate

m = -1.754

m = -1.754

$$y = mx + b + \varepsilon$$

$$\varepsilon \sim Norm(0, \sigma^2)$$



A model is a model no matter how it is analyzed!

- Statistical models exist independently from method of statistical analysis!
- There are no “Bayesian models” or “frequentist models”
- May choose to analyze a model (e.g., linear regression) in a Bayesian way
- Typically, Bayesian and frequentist analyses yield numerically very similar estimates

The role of models in science

- Science: explain nature, so you can better **understand** and/or **predict**
- Nature is too complex to understand must reduce complexity
- A model (broadly): greatly simplified version of nature, should help understand/predict
- Every model has an objective:

  e.g. understanding ≈ mechanism

  e.g. predicting ≈ description

Everybody is a modeler!

- Model = set of assumptions
- Description of model: words, graphs, algebra, ...
- Any explanation is based on a model, stated or unstated:

  To make sense of an observation, everybody needs a model ... Whether they knows it or not!

- Interpretation of data without a model is impossible
- Explicit models are better than implicit models (e.g., assumptions more transparent, can test them)

Mathematical vs. Statistical models

- Mathematical models: a description of a system composed of variables, typically written with algebra
- A simple example is the equation for a line:

$$y = mx + b$$

- Advantage of using algebra: transparency greatly increased over description in words and forces clarity of thought

Mathematical vs. Statistical models

- Statistical models: a description of a system composed of variables but where one or more ***random variables*** are related to other variables
- Explicitly acknowledge stochasticity in systems
- A simple example is the equation for a line:

$$y = mx + b + \varepsilon \qquad \varepsilon \sim Norm(0, \sigma^2)$$

- Response = systematic part + random part

Statistical models

Three essential kinds of random variability:

- Measurement error is the variability imposed by our imperfect observation of the world. It is often modeled by adding normally distributed variability around a mean value.
- Demographic stochasticity is the innate variability in outcomes due to random processes even among otherwise identical units. For example, the number of tadpoles out of an initial cohort eaten by predators in a set amount of time will vary between experiments even if we controlled everything about the environment.
- Environmental stochasticity is variability imposed from "outside" the ecological system, such as climatic, seasonal, or topographic variation.

Statistical models

- Parametric statistical model: Description of the processes using probability distributions thought to have produced the data

  (in contrast to *non-parametric* models which do not assume that data belong to a particular distribution)

- Generalized linear model (GLM):

  quintessential statistical model

Statistical models

Two frequently used GLMs in EEB:

- **Normal response**

  Random part: $y \sim Norm(\mu, \sigma^2)$

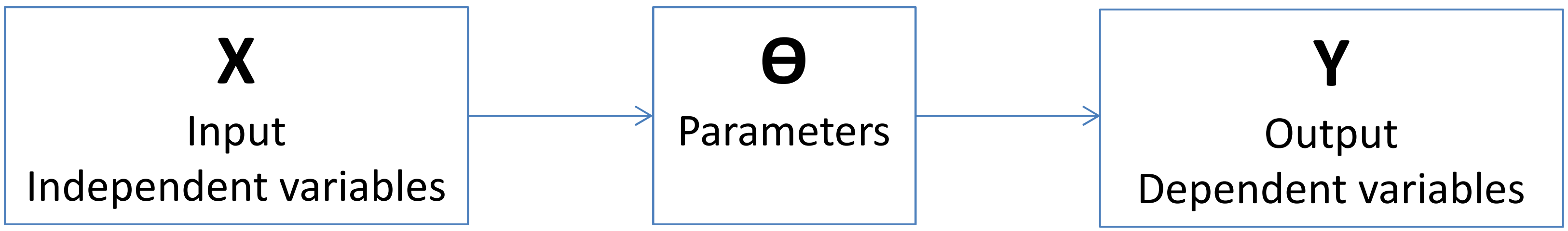
  Systematic part: $\mu = a + b \cdot x$

- **Binomial response**

  Random part: $y \sim Bin(p, N)$

  Systematic part: $logit(p) = a + b \cdot x$

Analysis of a statistical model



- Data viewed as result of random process(es)
- Parameters (θ) are **unknown** variables (of interest)
- How should we guess at value(s) of θ ?

  ...at missing covariates (x) ? ... at missing response (y) ?

--> Statisticians devise many procedures for guessing

- method of moments
- least-squares
- maximum likelihood, maximum partial likelihood, pseudolikelihood, penalized likelihood, ...
- Bayesian analysis

Analysis of a statistical model

X Input Independent variables	→	**θ** Parameters	→	**Y** Output Dependent variables

- Data viewed as result of random process(es)
- Parameters (θ) are **unknown** variables (of interest)
- How should we guess at value(s) of θ ?

  ...at missing covariates (x) ? ... at missing response (y) ?

--> Statisticians devise many procedures for guessing

- method of moments
- least-squares
- maximum likelihood, maximum partial likelihood, pseudolikelihood, penalized likelihood, ...
- Bayesian analysis

Frequentist analysis of a model

- Example: Estimate probability of detection (θ) of tadpoles -> Release n=50 in artificial pond, later resight y=20

Frequentist analysis of a model

- (One) Frequentist way of guessing at θ: maximum likelihood
- Parametric model describes data-generating probabilistic mechanism: sampling distribution p(y|θ)

  "probability of observing data y, given fixed param value θ*"*

- **Note:** probability statement about the data, **not** about θ
- Probability defined as long-run frequency in hypothetical replicate data sets
- E.g., binomial sampling distribution (y $\sim Bin(\Theta, n)$) with PMF:

$$p(y|\Theta) = \frac{n!}{y!\ (n-y)!}\Theta^{y}(1-\Theta)^{n-y}$$

Frequentist analysis of a model

Reminder! What is a PMF?

- PMF = Probability Mass Function
- A PMF is a function that gives the probability that a discrete random variable is exactly equal to some value

- Similarly PDF = Probability Density Function
- A PDF is the function that gives the probability that a continuous random variable falls within some range

Frequentist analysis of a model

- Maximum likelihood
- **Idea:** good choice of θ is that which maximizes function value of sampling distribution for the data set

- **Likelihood function:** reading the sampling distribution "in reverse" as a function of θ. $p(y|\Theta)$= (L$(\Theta|y)$

$$\mathrm{L}(\Theta|y) = \frac{n!}{y!\,(n-y)!}\Theta^{y}(1-\Theta)^{n-y}$$

- *Probability* describes a function of the outcome given a fixed parameter value. i.e., coin is flipped 10 times and it is a fair coin, what is the *probability* of it landing heads-up every time?

- *Likelihood* is used when describing a function of a parameter given an *outcome.* i.e., if a coin is flipped 10 times and it has landed heads-up 10 times, what is the *likelihood* that the coin is fair?

Frequentist analysis of a model

- Maximum likelihood
- **Idea:** good choice of θ is that which maximizes function value of sampling distribution for the data set

- **Likelihood function:** reading the sampling distribution "in reverse" as a function of θ. $p(y|\Theta)$= $(\mathrm{L}(\Theta|y)$

$$\mathrm{L}(\Theta|y) = \frac{n!}{y!\,(n-y)!}\Theta^{y}(1-\Theta)^{n-y}$$

- Call the value of θ that maximizes L the Maximum Likelihood estimate (MLE)

$$L(\Theta|20) = \frac{50!}{20!\ (50-20)!}\Theta^{20}(1-\Theta)^{50-20}$$

Frequentist analysis of a model

Maximum likelihood

- How to find the MLE ?

Frequentist analysis of a model

Maximum likelihood

- How to find the MLE ?

    - Analytically (sometimes)

        Take the derivative of the likelihood and find the maximum value

Frequentist analysis of a model

Maximum likelihood

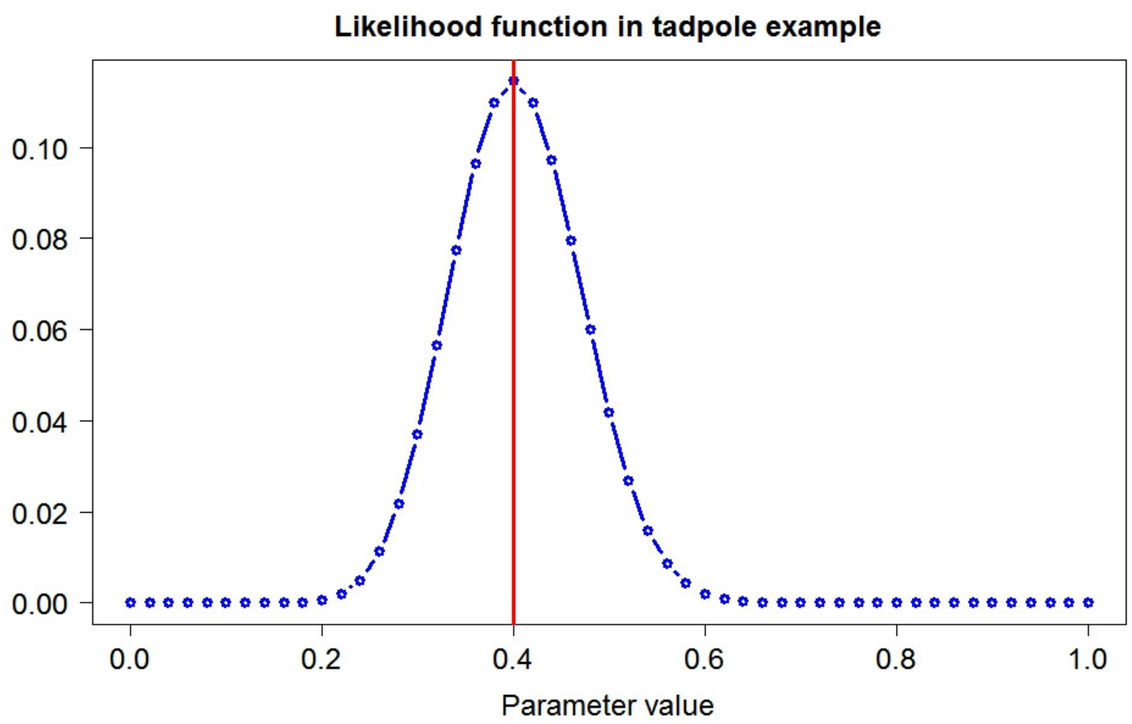
- How to find the MLE ?

    - Numerically (most of the times)

Frequentist analysis of a model

Maximum likelihood

Numerical estimation by brute force: try out and plot large number of values for θ



$$L(\theta|20) = \frac{50!}{20!\ (50-20)!}\theta^{20}(1-\theta)^{50-20}$$

Frequentist analysis of a model

Maximum likelihood – numerical approach
approximating derivation

- Numerical estimation by function minimization: e.g. optim() in R (also nlm() and others)
  - -> Specify the likelihood function
  - -> "Take the derivative" through numerical approximation
  - -> The value for which the negative (log)likelihood is minimized is the MLE
- Numerical estimation using special functions: R glm()

Frequentist analysis of a model

"Gold standard" in ecological and evolutionary statistics

- Well developed, consistent, and reliable – work for a large set of problems
- A number of desirable mathematical properties
  - Unbiased at large sample sizes (asymptotically unbiased)
  - Have approximate normal distributions and sample variances -> easily generate confidence bounds for hypothesis tests
- Lots of (easy to use) software packages offer a number of canned functions -> fast estimation process, minimal coding, reduced computational burden

Frequentist analysis of a model

But there are some drawbacks…

- Working out the likelihoods can be difficult, especially when model structure is complicated
- Biased at small sample sizes (and it may be difficult to determine what's big enough)
- Interpretation can be tricky
  - Confidence interval: If this procedure was repeated on multiple samples, the 95% confidence interval (which would differ for each sample) would encompass the true population parameter 95% of the time.
  - p-value: Probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming that the hypothesis under consideration is true
- Frequentist statistics cannot attach a probability to a hypothesis

Bayesian analysis of a model

- How should we guess at values of θ ? … or missing x ? … or predict y ?
- **Bayesian approach:** in the face of uncertainty about magnitude of θ use conditional probability, $p(\theta|y)$
- "Guess" at θ conditioning on what is *certain* or what we *know* (i.e., data x and y)

Bayesian analysis of a model

Recipe of every Bayesian analysis:

1. What is known?	The data ($y=20$, $n=50$)
2. What is unknown?	Prob. of detection (θ)
3. What to do?	Calculate $p(\theta \mid y)$

- "*Probability of parameter, given data*"
- **Note:** probability statement about the parameter
- Data, once collected, are fixed
- Degree-of-belief concept of probability: Express imperfect knowledge (about θ) using probability distribution
- Hence, parameters treated as if they were random variables
- How should $p(\theta \mid y)$ be computed?

Bayesian analysis of a model

Bayes rule

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

- Mathematical fact of probability
- Can be deduced from p(A,B) = p(B | A) * p(A)

  (joint prob. = conditional prob. *marginal unconditional prob.)
- Thomas Bayes, English minister/mathematician (1702-1761)
- Thomas Bayes applied the rule to unobservables such as parameters, i.e., for parameter estimation

Bayesian analysis of a model

Bayes theorem:

- Basic tool of Bayesian analysis
- Provides the means by which we can learn from data
- Given a prior state of knowledge, it tells us how to update this belief based on observations

Bayesian analysis of a model

Bayes rule for statistical inference:

$$P\,(\Theta|y) = \frac{P(y|\Theta)P(\Theta)}{P(y)} = \frac{P(\Theta, y)}{P(y)}$$

Posterior distribution:	$P(\Theta|y)$
Likelihood function:	$P(y|\Theta)$
Prior distribution:	$P(\Theta)$
Prob. of data:	$P(y) = \int P(y|\Theta)P(\Theta)d\Theta$

- **NOTE:** Use probability to express imperfect knowledge
- Direct probability statements about unknown quantities: *Can* say *"… I am 95% certain that prob of detection > 0.2"*

Bayesian analysis of a model

Formal steps in Bayesian analysis

- Use probability as a universal measure of uncertainty about unknown quantities (here: θ)
- Treat all statistical inference (estimation, testing, ...) as a simple probability calculation
- Express your knowledge about parameter θ (excluding information contained in y) by a probability distribution: the prior p(θ)
- Use Bayes rule to *update* that knowledge with the information contained in the data y and embodied by the likelihood function, p(y|θ)
- **Result is probability distribution, p(θ|y), for every unknown; unlike ML, where result is single value**

Bayesian analysis of a model

Heuristic appeal of Bayes rule as model for inference

- “Human” concept of probability (“*I am 95% certain that...*”)
- Like human learning:

  - Conclusion is combination of experience and new information (e.g., 10 ft tall man)

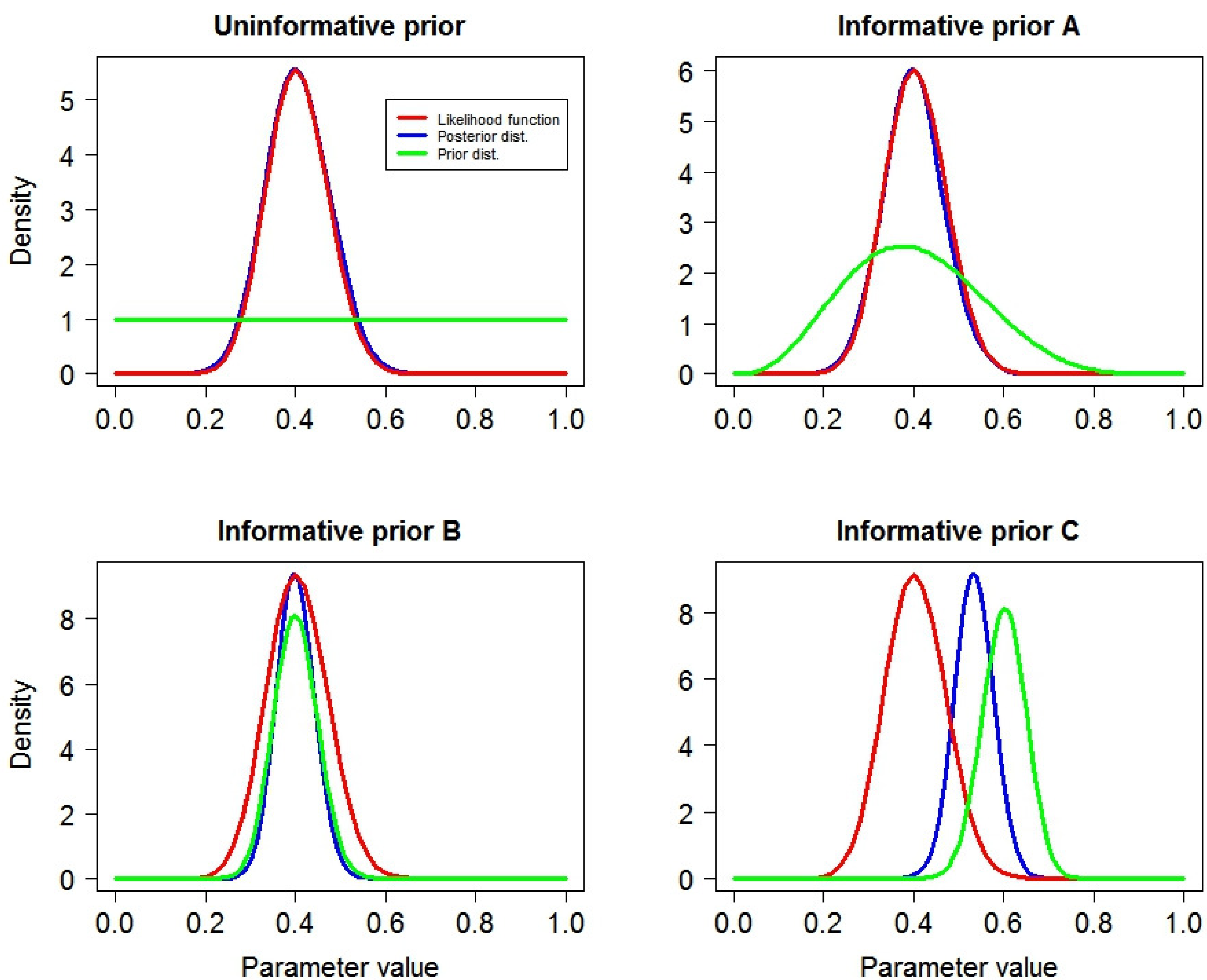
  - New information changes (“updates”) my previous state of knowledge to my current state of knowledge

  - Every analysis could be a meta-analysis: synthesizes *all* existing knowledge

Bayesian analysis of a model

- Think back to tadpole example: Estimate probability of detection (θ) of tadpoles -> Release n=50 in artificial pond, later resight y=20
- Recall: MLE = 0.4

Bayesian analysis of a model

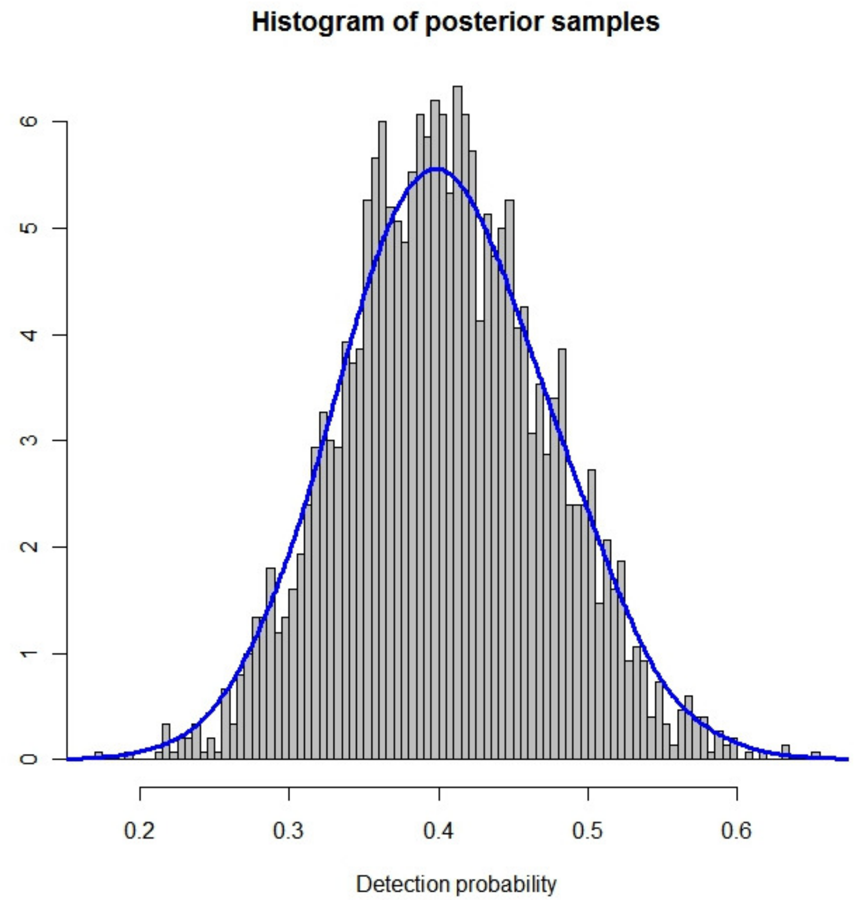


Bayesian computation

Tadpole example

```
.P
[1] 0.5265 0.4088 0.3885 0.348
[7] 0.4042 0.3593 0.3580 0.388
[13] 0.4935 0.2831 0.4827 0.46
[19] 0.4579 0.3605 0.4488 0.39
...
[2983] 0.3866 0.3265 0.3121 0.
[2989] 0.3446 0.3584 0.3839 0.
[2995] 0.3844 0.5067 0.4212 0.

> mean(p)
[1] 0.4047
> sd(p)
[1] 0.0674
> quantile(p, probs = c(0.025,0.975))
2.5% 97.5%
0.2771 0.5375
```



Bayesian analysis of a model

Why would you want to use a Bayesian analysis

- Sometimes finding the joint likelihood is really hard
  - Interacting parameters
  - Lots of random effects in a hierarchical structure
  - Integrating hundreds (or thousands!) of likelihoods
- Absence of asymptotics – unbiased regardless of sample size
- Ease of error computation – directly compute SE or confidence intervals (as opposed to using the delta method with MLE)
- Intuitive interpretation of parameters (e.g., I am 99% certain that…)

Bayesian analysis of a model

But there are drawbacks too…

- Specification of a prior distribution means that results are always dependent on that prior
  - Can specify ‘non-informative’ (vague) priors (though may be difficult to specify “non-information” in some cases)
  - Must report priors for every analysis
  - Justify choice of informative priors
- High computational cost, usually takes (much) longer to estimate parameters compared to MLE
- Usually must program yourself – less canned software available.

MLE vs Bayesian analysis: summary

- Pros and cons to each analysis framework
- Analysis framework should be chosen after considering each question and model
- Clearly delineate your assumptions and ensure that your data meet the criteria

Steps of the modeling process

1. Identify the ecological question – Know this at a general, conceptual level and at a specific level.
2. Choose the deterministic model(s) – This can be phenomenological or mechanistic.
3. Choose the stochastic model(s) – Need to know about the variability around the expected pattern.
4. Fit parameters – Decide on an analysis framework to estimate values for models.
5. Estimate confidence intervals / test hypotheses / select models – Quantify uncertainty in your estimates.
6. Put results together to answer question in part 1 – Modeling is an iterative process!

Programming the tadpole example

Super straightforward example estimating the detection probability using:

- Brute force with the binomial PMF
- Explicitly minimizing the negative log-likelihood
- The ‘glm’ function built in R