Wrapping Up:
The End at Last

The Story So far...

- We have a subjective view of probability
  - summarises our uncertainty in what is happening
- If we do experiments or make field observations, we hope to reduce our uncertainty
  - i.e. to learn about the system
- The Bayesian approach allows us to do this rationally

The Story So far...

- Quantify our uncertainty about the data (*X*)
- Need a model of how the data is generated
  - has parameters (θ)
- Update our beliefs about the parameters with Bayes' Theorem:

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$

  - $P(\theta \mid X)$: Posterior (after the data)
  - $P(\theta)$: Prior (before the data)
  - $P(X|\theta)$: Likelihood (effect of the data)

Of Marginal Interest...

- Working with the posterior is just using probability theory
- e.g. we can work with *marginal distributions*:

$$
\begin{aligned}
P(\theta_1|X) &= \int P(\theta_1 \theta_2|X)\, d\theta_2 \\
&= \int P(\theta_1|\theta_2 X)\, P(\theta_2|X)\, d\theta_2
\end{aligned}
$$

- Includes the uncertainty in the other parameters
- Gives us a way of dealing with nuisance parameters

Predictions

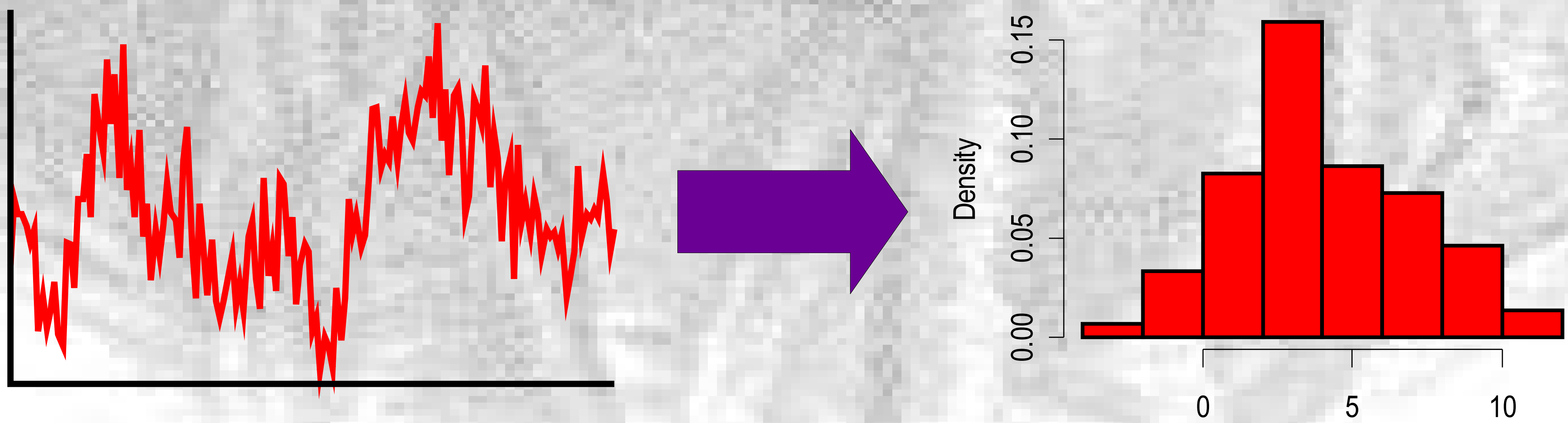
- We deal with predictions in the same way
- If we want to do predict X_{new}, then we use:

$$P(X_{new}|X) = \int P(X_{new}, \theta|X)\, d\theta$$
$$= \int P(X_{new}|\theta X)\, P(\theta|X)\, d\theta$$

- Which is essentially the same trick

In Practice

- In practice most of these calculations are too difficult to do on paper
- Instead, we simulate the posterior distribution
  - posterior simulation
- Usually use MCMC



As Easy As MCMC

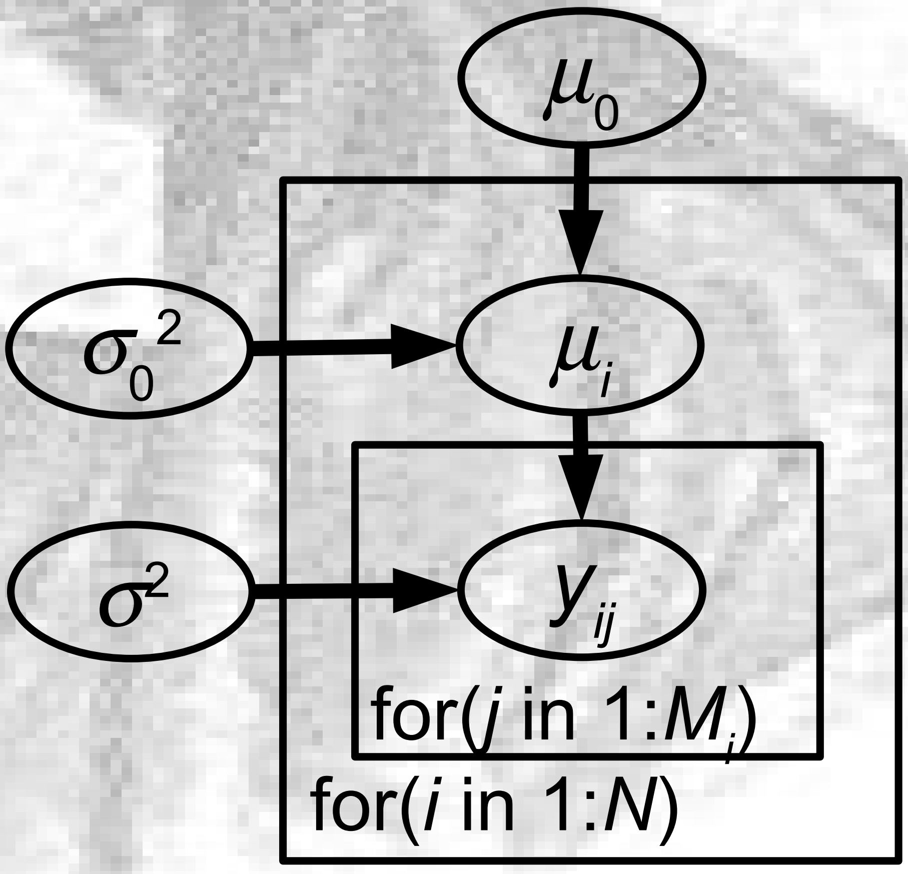
- With MCMC we take marginal distributions by ignoring the parameters we are not interested in
- Prediction: for each iteration, make a prediction
  - the distribution is then the posterior predictive distribution we want

Types of Models

- The development of MCMC has made it easier to develop models
- One type: hierarchical models
- The likelihood says that the data is drawn from a distribution
- The distribution is controlled by some parameters
- The parameters are determined by the prior
  - another distribution

Models As Art

- We can use the prior like a likelihood
  - determine the parameters of the distribution with hyper parameters
- e.g. random effects
  - $y_{ij} \sim N(\mu_i, \sigma^2)$
  - $\mu_i \sim N(\mu_0, \sigma_0^2)$
- Provides a flexible framework for developing models



Now...

- Ask some general questions:
- What is the difference between the Bayesian and “traditional” approaches?
- Should we always take the Bayesian approach?
- What should we look out for in Bayesian analyses?

Alternative Approaches

- Statistical analyses are attempts to quantify our statements about data
- Three general approaches to data analysis
  - non-parametric
  - frequentist (Maximum Likelihood)
  - Bayesian
- All have their adherents

Non-parametric Inference

- Try not to make any distributional assumptions
  - actually assume exchangeability
- Examples include the Wilcoxon sign rank test
  - a.k.a. Mann-Whitney
- Problem: looses power
- Difficult to analyse complex situations
- Some ideas appear in other areas
  - e.g. smoothing of curves

Likelihood

- Both the frequentist and the Bayesian approaches use likelihood
- Frequentist: maximise the likelihood
  - Find the parameters which maximise $P(X|\theta)$
  - find the parameters most likely to give the data
- Assume these parameters are “correct”
- Talk about the probabilities of getting the data given these parameters
  - summarise the data with statistics

Frequentists

- Frequentists assume the ML estimate is correct
- Inference is done using statistics
  - functions of the data
  - strictly, not the parameters
  - talk about the probability of the parameter given the data, with the ML estimate being true
- Standard error is an estimate of the variation in the parameter

The Arguments

- Frequentists claim that they are objective
  - they do not have priors
- Their interpretation of probability is based on
- repeated trials
  - rolling the dice 1669024 times
- Makes the interpretation complicated
  - e.g. hypothesis tests – “what is the probability of getting this value of the statistic or greater if the null hypothesis is correct”
    - NOT the probability that the null hypothesis is correct

The Bayesian Approach

- Assumes a subjective interpretation of probability
  - need priors
- Once we have this, the interpretations of the
- probability statements are simpler
  - a significance test is the probability that the null hypothesis is true
    - or rather the probability we would assign

Do These Differences Matter?

- Do we get the same answers in practice?
- With a uniform prior, then we have $P(\theta) \propto 1$
- So:

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$
$$\propto P(X|\theta)$$

- In other words, the posterior will be the same as the likelihood
- So estimates should be the same

But...

- Often there is not a natural uniform prior
- However, we only need the prior to be uniform in the region with high a likelihood
  - only locally uniform
  - so, vague priors should work almost as well
- If the data is informative, the likelihood should dominate, so the prior does not matter

When It Is Different

- With complex models, there may still be a differences
- Look at one parameter, treat others as nuisance parameters
- Bayesian: marginal distributions
- Frequentist: more difficult. Often use a
- *profile likelihood*

Differences in Style

- Most classical analyses are carried out in a different way to Bayesian analyses
  - at least in biology!
- The question is framed in terms of “is factor
- X significant”
- This then emphasises significance testing
- Parameters are reported as point estimates with standard errors

Significance Tests are Evil

- In statistics, significance has a special meaning
  - the data is unlikely to have been obtained if the null hypothesis were true
- In reality the null hypothesis is never true
  - with enough data, a test will be significant
- What scientists need is a measure of practical significance
  - is an hypothesis important?

Beyond *P* and Evil

- For frequentists, significance tests are easy
- If two models are nested, compare the likelihoods
  - Likelihood Ratio Test
  - ANOVA table
- Because it is easy, it gets used
- The frequentist “style” has evolved
- Significance tests now get overused

Making Significance Insignificant

- Significance tests are more difficult for Bayesians in practice
  - we can use Bayes Factors or probabilities to express our belief in an hypothesis
  - but they do not come straight from the analysis
    - unless you build the model to put them in
- There has also been a move away from significance tests in the 90s
  - big arguments between psychologists

The Good Bayesians

- The Bayesian approach focuses more on parameters
- We can make statements like P(θ>0)=0.22
  - if θ is continuous, P(θ<0)=0.78 and P(θ=0)=0
- But these statements are less meaningful
- Instead we focus on what the parameters mean
  - which is what most statisticians say we should do anyway

Why Does the Bayesian Approach Help?

- The Bayesian approach focusses more on the model
  - we need to put priors on the parameters
  - the major software is designed around writing models
    - hierarchical models have great flexibility
- Get a deeper understanding of the model
  - may change when people write automatic software!

Should We Always Be Bayesian?

- If we are ideologues, then we have to be Bayesians if we believe in subjectivist probability
- If we are more interested in getting things done, then the answer is more difficult
- Opinions can vary, this is my view...

A Subjective View

- The Bayesian approach scales well with complexity
  - deals with complex models well
  - hierarchical models help a lot
- Simpler situations are not so easy
  - e.g. regression is straightforward for frequentists
  - Bayesians still need to use priors, and this makes the fitting more difficult
    - still need MCMC
- When the problem gets complex, go Bayesian!

Suggestions for Bayesian Analysis

- Most biologists will probably be using pre-written packages
  - e.g. MrBayes, Hickory
- These restrict your ability to develop models
- But there are still some problems that you need to be aware of

Prior Elicitation

- Two approaches to prior: informative, uninformative
- In principle we should all use informative priors
- In practice, most people don't want to think too hard, so use vague, uninformative priors
- Analyses can (and should?) be checked for sensitivity to priors

Informative Priors

- Deciding what you think you know, and designing your priors based on that
  - do thinks like decide what the most likely value is, and what sort of confidence limits to set
    - most people tend to be overly certain
  - With prediction, can lead to better results
- Can be a problem with hypothesis testing
  - you can be accused of fiddling the priors to get what you want
- Can also help the estimation

Uninformative Priors

- Expression of ignorance
- In principle, should give results closer to frequentist analyses
- One problem: there may be several uninformative priors for any parameter
  - e.g. the binomial distribution: uninformative on p or $\log(p/(1-p))$?
- Sometimes can imply strong assumptions
  - so not as easy as you might hope!

MCMC Problems

- MCMC is not guaranteed to be successful
- It does need to be checked for convergence and mixing
- At the very least, run two chains and see if they are similar
- Often worth plotting pairs of parameters
  - correlations cause a lot of problems

MCMC Solutions

- Convergence:
  - improve initial estimates
    - so the chain doesn't have to move as far
- Mixing
  - changing priors might help
    - but can be seen as cheating!
  - or change the proposal distribution
  - or the model
    - can just change the way it is written
-

Reading Bayesian Papers

- Dodgy things to spot
- Is the model described
  - or is there a suitable reference?
- What are the priors?
  - need to be stated. If they are informative, why were they chosen?
- Are the priors checked for sensitivity?
  - especially a problem with informative priors
- If MCMC is used, is convergence checked?

Bayesian Results

- Not just point values should be reported
  - some form of posterior standard deviation or confidence interval needs to be added
- CIs called either Credible Intervals or Bayesian Confidence Intervals
- Density plots are sometimes given
- Prior is often shown
  - if not, can you work out what it would look like?
    - it might be flat over the plotting region

Other Tricks

- Because the Bayesian approach gives us a full posterior distribution, any statistics we calculate will have the correct posterior
  - So, we can report these too
- Gives us a useful tool
- e.g. if we are comparing variance components, we can estimate them in a model and then calculate a correlation later

Where Next?

- There is a large literature on Bayesian methods
- For software, look at OpenBUGS
  - http://www.math.helsinki.fi/openbugs/
- Alas there is no general textbook for biologists
  - but in any area, there will be some literature

Further Reading

- Data Analysis Using Regression and Multilevel/Hierarchical Models
  - Gelman and Hill (2007)
    - Good non-specialist introduction to hierarchical models
- Bayesian Data Analysis
  - Gelman, Carlin, Stern, and Rubin (2004)
    - Aimed at statisticians, comprehensive
- Bayesian Methods for Ecology
  - McCarthy (2007)
    - introduction to the ideas

Will Bayesians Rule The World?

I hope not