Model Selection
Outline

● Philosophy of science and multiple alternative models
● Complexity/identifiability trade-off
● Likelihood-based metrics
  – Likelihood Ratio Test
  – AIC
● Bayesian metrics
  – DIC
  – Predictive Loss
“The” Scientific Method?

• Popper
  – Falsification of hypotheses
    • Hypotheses can not be proved, only disproved
• Stats: “Hypothesis testing”
  – Single hypothesis is disproved by confrontation with the data
  – Likelihood the data would have been observed if the null hypothesis was true
  – If this probability (p-value) is small enough we reject the null
Alternative Philosophies of Science

• Kuhn – Scientific Paradigms
  – Dominant paradigm used until there is so much contradictory information that it is “overthrown”
  – Requires an alternate paradigm that is “better”

• Polanyi – Republic of science
  – Multiple views of the world by different scientists
  – Confrontation between views and data judged by plausibility, value, and interest

• Lakatos – Scientific research program
  – Confrontation of multiple hypothesis with data as arbitrator
Null models

• All these alternatives acknowledge
  – There may be multiple alternative models
  – Simple null models often scientifically trivial, uninteresting
  – Doesn't make sense to reject a model if there is not an alternative

• Likelihood and Bayesian stats both well suited to “judge” the contest between multiple competing hypotheses and data
Models vs Hypotheses

- Models usually more specific than hypothesis
- Hypoth: Birds forage more efficiently in flocks
- Models: Consumption vs Size
  - Consumption proportional
  - Consumption saturates
  - Increases then decreases

\[
C = aS \\
C = \frac{aS}{1 + bS} \\
C = aS e^{-bs}
\]

- “All models are wrong but some are useful”
  -- George Box
Model selection

- Focus on choosing between multiple competing models rather than refuting a single null model

- How do we judge models?
  - Complexity
    - Number of parameters
  - Uncertainty
    - Model residuals
    - Parameter error (identifiability)
  - Data as ultimate arbiter

- “Make everything as simple as possible, but not simpler.” - A. Einstein
Characteristics of model selection metrics

- Will select the “best fitting” model
  - Minimum Deviance
- Will penalize fit for the number of parameters
- Asymptotically (large sample size) should select the “true” model rather than the one with the most parameters

- Often the statistically best model is simpler than we would think (and sometimes simpler than the “true” model)
Likelihood Ratio Test

- LR = \( \frac{L(x|\theta_A)}{L(x|\theta_B)} \)

- \( D = -2\ln L(x|\theta_A) - 2\ln L(x|\theta_B) \)

- The test statistic \( D \) is known to be distributed with a \( \chi^2 \) distribution

- Degrees of freedom = Difference in # of param.
  - Overall, \( L \) increases (-lnL declines) with # of param.
  - Penalizes model with more parameters

- \( p\text{-val} = 1 - p\text{chisq}(D,df) \)
LRT pro/con

- Only applies to *nested* models
- Only PAIRWISE comparisons
- Asymptotically, slightly biased toward more complex models
- Provides a p-value
- Additional reminders:
  - **ALL** model selection criteria require application to the same data with same sample size
  - e.g. If adding covariate Z requires rows to be dropped because of missing values, have to drop from the model w/o Z as well
Nested Models

• The more complex model collapses to the simpler model when one or more of the parameters is FIXED

• Examples:
  – Weibull vs Exponential (Lab 3) (fix c=1)
  – Pine cone: combined vs AMB/ELEV (Lab 4)
  – Regression: Inclusion of additional covariates (fix slope = 0)
Example: Polynomial
Example: Polynomial

• Candidate models:
  – \( Y = b0 \)
  – \( Y = b0 + b1 \cdot x \)
  – \( Y = b0 + b1 \cdot x + b2 \cdot x^2 \)
  – \( Y = b0 + b1 \cdot x + b2 \cdot x^2 + b3 \cdot x^3 \)

• Comparisons
  – 0 vs 1
  – 1 vs 2
  – 2 vs 3
• 0 vs 1
  p=7.6e-10

• 1 vs 2
  p=0.00019

• 2 vs 3
  p=0.9238
Akaike Information Criterion

\[ AIC = -2 \ln L + 2p \]

• \( p \) = number of parameters in the model
• Based on information theory
• Lowest value “wins”
• No \( p \)-value
• Often expressed relative to best model, \( \Delta AIC \)
• “Rules of thumb”
  – 0-2 = similar  
  – 2-5 = weak support  
  – >5 = strong
- $\Delta AIC$
- 0  
  47.77
- 1  
  11.91
- 2  
  0.00
- 3  
  1.99
P-value

• Probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

• Not the probability that the null hypothesis is true
  – P-value can be close to zero when the posterior probability of the null is close to 1

• Not the probability of falsely rejecting the null hypothesis
Example: Southern Brown Frog

- Researcher surveys a pond for the frog
- From prior experience 80% detection | present
- No frogs observed
- If null hypothesis is frogs are absent
  - $P = 1.0$ -- Fail to reject
  - Further surveys that fail to find the frog, $p=1.0$
- If null hypothesis is frogs are present
  - $P = 0.2$ -- Fail to reject
Power

- Probability of correctly rejecting the null hypothesis
- Requires that some explicit alternative hypothesis is stated
  - Parameter values
  - Variance
  - Sample size
- Often calculated as a function of sample size
- For complex models, calculate through simulation
Generic Example

```
LnL.A = function(theta){
  -sum(dnorm(y,f(x,theta),sd))
}
lnL.0 = function(mu){
  -sum(dnorm(y,mu,sd))
}
for(i in 1:nsim){
  Ey = f(x,theta)       ## process model
  y = rnorm(N,Ey,sd)    ## data model
  outA = optim(ic,LnL.A) ##fit of alternative
  out0 = optim(ic,lnL.0) ##fit of null
  pval[i] = 1-pchisq(2*(outA$value-out0$value),df)
}
power = sum(pval < 0.05)/nsim
```
Identifiability

- Data may not provide information on all parameters in a model
- Often requires restructuring model
- Not fixed by collecting more data
- Parameters often “trade-off when fitting”
- Simple examples
  - \( N(\mu, \sigma^2 + \tau^2) \)
  - \( N(a/b, \sigma^2) \)
- Occur in both Likelihood and Bayes