Canopy Light: Synthesizing multiple data sources

- Tree growth depends upon light (previous example, lab 7)
- Hard to measure how much light an ADULT tree receives
- Multiple sources of proxy data
  - Exposed Canopy Area
    - aerial photography, Quickbird
  - Canopy status
    - suppressed, intermediate, dominant (ex 8.2.2)
  - Light models
    - Allometries, stand map
Mechanistic Light Model

• Estimate light levels based on a 3D ray-tracing light model

• Parameterized based on canopy photos, tree allometries
Linear models
Logistic
Multinomial

Non-zero ECA observations
$\lambda^{(e)} > 0$

Light availability

Linear scale

Model light estimate
$\lambda^{(m)}$

Posterior light estimate

$\beta_0\beta_1$
$c_0$
$v_l$
$c_1v_e$
$a_0a_1v_m$
Exposed Canopy Area

- Error in relationship between “true” light $\lambda$ and observations $\lambda^e$

$$p(\lambda_{i}^{(e)}) = \begin{cases} 
1 - p_i & \lambda_i^{(e)} = 0 \\
p_i \mathcal{N}(\ln(\lambda_i^{(e)})|\ln(\lambda_i), \nu_e) & \lambda_i^{(e)} > 0
\end{cases}$$

- Probability of observing the tree in imagery increases with “true” light availability

$$\text{logit}(p_i) = c_0 + c_1 \lambda_i$$
Mechanistic Light Model

- Assume a log-log linear relationship between “true” light and modeled light
- Provides a continuous estimate of light availability for understory trees
  - ECA = 0
  - Status = 1

\[ p(\lambda_i^{(m)}) = N(\ln(\lambda_i^{(m)})|a_0 + a_1 \cdot \ln(\lambda_i), \nu_m) \]
Model Fitting

- Model fit all at once
- Find the conditional probabilities for each parameter (i.e. those expressions that contain that parameter)
  - Always at least 2 – likelihood and prior
  - Can be multiple likelihoods
- MCMC iteratively updates each parameter conditioned on the current value of all others
Recap and fill in the blanks
Course Overview

• Sections 1 & 2: Basics
  – Probability, distribution, theory/methods for point estimation in Likelihood and Bayes, Interval estimation

• Section 3: Core
  – Model comparison
  – Generalized Linear Mixed Models (GLMM)
  – Relaxing the assumptions of linear models

• Section 4: Advanced applications

Third Exam on Monday April 2nd
Course Goals

- Acquire a “toolbox” of basic flexible techniques
- Gain experience applying these tools (LAB)
- Be able to read and evaluate statistics used in the biological literature
- Be able to understand and evaluate new statistics
- Be able to relax the assumptions of existing methods or devise new models tailored to the problem at hand
SOME OF THEIR RULES CAN BE BENT, OTHERS CAN BE BROKEN
Experimental Design

• Traditional design
  – Minimize sources of variability
  – Balanced Replication
  – Simple treatments
  – Classical ANOVA and Regression stats
  – Popperian falsification
  – High power
  – Limited generality
    • “Does this matter in the real world”
Experimental Design

• “Real world” experiments
  – Real world is variable
  – Gain a broader scope of reference
  – Loose power
  – Multiple alternative hypotheses
  – Need a statistical framework that can account for the complexity and variability of the real world
Probability Theory

- Basis of both Likelihood and Bayesian perspectives
- Conditional probability
- Random variables – have a PDF
- Likelihood $P(\text{data} | \text{model})$
  - Parameters fixed, data random
- Bayes' posterior $P(\text{model} | \text{data})$
  - Data fixed, parameters random
Probability Density Functions

\[ X \sim f(\theta) \]

- Must integrate to 1
- Choose based on
  - Type of X:
    - continuous or discrete
    - Numeric, ordinal, categorical
  - Range of X
    - Lower bound, upper bound
  - Shape of distribution
  - Conjugacy
Adding flexibility

- Truncation (don't forget to re-normalize)

- Mixtures
  - Based on conditionals
  - Zero-inflation
    - $P(X \mid \text{event}) \times P(\text{event})$
  - Hierarchical structure
    - $P(X \mid \theta_1) \times P(\theta_1 \mid \theta_2)$
  - Each stage can incorporate process models
    - $P(\text{cone}) = P(\text{cone} = f(\text{size}) \mid \text{fecund}) \times p(\text{fecund} = f(\text{size}))$
Bayes' Theorem

\[ P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \]

\[ = \frac{P(y|\theta)P(\theta)}{\int_{-\infty}^{\infty} P(y|\theta)P(\theta)\,d\theta} \]
Think Conditionally
Data model

- Choose PDF that is appropriate for the data
- Can accommodate more complex structures to observation errors

- Errors in Variables: $x^{(o)} \sim g(x \mid \theta_x)$
  - e.g. Lab 9: TDR as a proxy for soil moisture

- Observation error in y: $y^{(o)} \sim h(y \mid \theta_y)$
Process Model

• Mathematical statements of our hypotheses
• What most people think of as “modeling”
• Usually deterministic
  – Conditioned on the value for the parameters, computation will always give the same answer
• Usually used to give $E[y]$
• There is nothing sacred about linear models
  – You will rarely find a theory that looks like linear regression
  – Consider asymptotics
Parameter Model

- In the Bayesian perspective, **all parameters** need priors
  - Don't forget the data model! (e.g. $\sigma^2$)
- In latent variable problems, state variables are unknowns and thus need priors
- In hierarchical models, the parameter models are more complex and have free parameters
  - Require hyperpriors
Priors

- Should be proper (finite positive integral), cannot make inference from improper posteriors

- Considerations for PDFs
  - Range, shape, etc.
  - Mixtures/truncation are permissible
  - Logical relationships are permissible (e.g. order)

- When applied, the prior parameters must be specified and stated explicitly

- Must be specified independent of DATA

- Updatable
Latent variables

- Sometimes the state variables in a system are
  - Unobservable
  - Observed with error
  - Inferred from proxy data
- Need to be estimated
- Need to integrate over their uncertainty
  - Don't ignore the uncertainty in calibration curves
  - Don't just interpolate, average, bin, smooth, transform, etc.
Linear model: \( y = Xb + \varepsilon \)

- **Data model:**
  - Depends on characteristics of \( y \)
    - Continuous, discrete, bound over a range, boolean, categorical, circular, etc.

- **Process model:** \( Xb \) or \( \text{link}(Xb) \)

- **Parameter model:**
  - \( b \) is almost always continuous
  - Regardless of whether \( X \) is cont/disc, bound, etc.
When X isn't continuous

- **Discrete:** Usually treated the same
- **Categorical:**
  - **Nominal:**
    - Design matrix of indicator variables (0,1)
    - Equivalent to ANOVA
    - Identifiability
      - n-1 columns with one group as the REFERENCE group
      - OR drop intercept term
  - **Ordinal:**
    - Similar to Nominal but may build ORDER restrictions into MLE / prior for $\beta$
Assumptions of Linear Model

- Homoskedasticity
- No error in X variables
- No missing data
- Normally distributed error
- Residual error in Y variables is measurement error
- Observations are independent
- Linearity

Model variance
Errors in variables
Missing data model
GLM

Hierarchical Models
Nonlinear models
Heteroskedasticity

1) Transform the data
   1) Pro: No additional parameters
   2) Cons: No longer modeling the original data, likelihood & process model have different meaning, backtransformation non-trivial (Jensen's Inequality)

2) Model the variance
   1) Pro: working with original data and model, no transf.
   2) Con: additional process model and parameters (and priors)
$y \sim N(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2)$
Errors in Variables

\[ \tilde{y} \sim N( \mathbf{X} \tilde{\beta}, \sigma^2 ) \]

\[ x^{(o)} \sim N( x, \tau^2 ) \]
Missing Data Model

\[ \tilde{y} \sim N(\mathbf{X} \tilde{\beta}, \sigma^2) \]
ASSUMPTION!!

• Missing data models assume that the data is *missing at random*

• If data is missing SYSTEMATICALLY it can not be estimated
Generalized Linear Models

- Allows for alternate PDFs to be used in likelihood
- Typically a link function is used to relate linear model to PDF

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link Name</th>
<th>Link Function</th>
<th>Mean Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Identity</td>
<td>$X_b = \mu$</td>
<td>$\mu = X_b$</td>
</tr>
<tr>
<td>Poisson</td>
<td>Log</td>
<td>$X_b = \ln(\mu)$</td>
<td>$\mu = \exp(X_b)$</td>
</tr>
<tr>
<td>Binomial</td>
<td>Logit</td>
<td>$X_b = \ln\left(\frac{\mu}{1-\mu}\right)$</td>
<td>$\mu = \frac{\exp(X_b)}{1+\exp(X_b)}$</td>
</tr>
<tr>
<td>Multinomial</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Can use most any function as a link function but may only be valid over a restricted range
If you're **response** variable is

- **Boolean** (0/1, True/False, Present/Absent)
  - Bernoulli likelihood (logistic regression)
- **Count data**
  - Poisson or Negative Binomial regression
- **Categorical data**
  - Multinomial likelihood (cumulative logistic)
- **Continuous but > 0**
  - Lognormal, exponential, or gamma likelihood
- **Proportion** (0,1 continuous)
  - Beta, Logit-normal, truncated normal
Hierarchical Models

Common mean

Hierarchical

Independent

\[ \mu \]

\[ \theta \theta \theta \]

\[ \mu_1 \mu_2 \mu_3 \]

\[ Y_1 Y_2 Y_3 \]
Hierarchical Models

- Model variability in the parameters of a model
- Partition variability more explicitly into multiple terms
- Borrow strength across data sets
Hierarchical Models

Data Model

\[ Y_1, \ldots, Y_k, \ldots, Y_n \]

Process Model

\[ \mu_1, \mu_k, \mu_n, \sigma^2 \]

Parameter Model

\[ \mu, \tau^2, S_1, S_2 \]

Hyperparameters

\[ m_0, V_\mu, t_1, t_2 \]
Mixed effects models

- Can rearrange a linear hierarchical model in terms of a **fixed effects** linear model and one or more **random effects** with mean 0

\[
Y_k \sim N \left( X \beta + \alpha_k, \sigma^2 \right)
\]

\[
\alpha_k \sim N \left( 0, \tau^2 \right)
\]

- This works for non-normal likelihoods and is referred to as Generalized Linear Mixed Models (GLMM)
The Devil's in the Details
The Details are in the Subscripts

• Example: Measured the growth of seedling $i$ in plot $j$ in year $t$ for species $s = g_{i,j,t,s}$

• $g_i$'s are NOT independent

• Would want to consider how $i$, $j$, $t$, and $s$ affect $g$
  – Which are fixed effects? Individual-level covariates?
  – Which are random effects? Hierarchical covariates?
  – Spatial, temporal, or phylogenetic autocorrelation?
Assumptions of Linear Model

- Homoskedasticity
- No error in X variables
- No missing data
- Normally distributed error
- Residual error in Y variables is measurement error
- Observations are independent
- Linearity

Model variance
Errors in variables
Missing data model
GLM

Hierarchical Models
Nonlinear models
Which techniques work where

• All work in Bayesian

• Straightforward in Likelihood
  – Heteroskedasticity
  – GLM & simple mixed models (GLMM)
  – Nonlinear

• Difficult to impossible in Likelihood
  – Complex Hierarchical models
  – Errors in variables / Latent variables
  – Missing data
Other Common Errors

• Forgetting about Jensen's Inequality
  – Transformations, non-linear models

• Treating “random” quantities like fixed numbers
  – Treating regression parameters like data
  – Ratios
  – Prediction without uncertainty statements
  – Work as close to the raw data as possible

• Log + 1 transform of zero-count data

• Having same covariate on both sides of eqn.
Notational equivalence

\[ Y \sim N(X \beta, \sigma^2) \]

\[ N(Y | X \beta, \sigma^2) \]

\[ Y = X \beta + \epsilon \]
\[ \epsilon \sim N(0, \sigma^2) \]

\[ \mu = X \beta \]
\[ Y \sim N(\mu, \sigma^2) \]
Linking graphs, equations, and code

\[ \tilde{\mathbf{y}} \sim N(\mathbf{X} \tilde{\mathbf{\beta}}, \sigma^2) \]
\[ \tilde{\mathbf{\beta}} \sim N(\mathbf{B}_0, \mathbf{V}_\beta) \]
\[ \sigma^2 \sim IG(s_1, s_2) \]

\[
\text{model} \\
\begin{align*}
    \text{mu} & \sim \text{dnorm}(0, 0.001) \\
    \text{sigma} & \sim \text{dgamma}(0.001, 0.001) \\
    \text{for}(i \text{ in } 1:n) & \\
    \quad \text{x}[i] & \sim \text{dnorm}(\text{mu}[i], \text{sigma}) \\
\end{align*}
\]

Diagram:

- **Data Model**: \( \mathbf{X} \rightarrow \mathbf{Y} \)
- **Process Model**: \( \tilde{\mathbf{\beta}}, \sigma^2 \)
- **Parameter Model**: \( \mathbf{B}_0, \mathbf{V}_\beta, s_1, s_2 \)
Linking Full Posterior & Conditionals

\[ \tilde{y} \sim N(X \tilde{\beta}, \sigma^2) \]
\[ \tilde{\beta} \sim N(B_0, V_{\beta}) \]
\[ \sigma^2 \sim IG(s_1, s_2) \]

\[
p(\tilde{\beta}, \sigma^2|X, Y) \propto N(Y|X \tilde{\beta}, \sigma^2) \times N(\tilde{\beta}|B_0, V_{\beta}) IG(\sigma^2|s_1, s_2)
\]

\[ \tilde{\beta} \sim N(\tilde{y}|X \tilde{\beta}, \sigma^2) N(B_0, V_{\beta}) \]
\[ \sigma^2 \sim N(\tilde{y}|X \tilde{\beta}, \sigma^2) IG(\sigma^2|s_1, s_2) \]
Linking Full Posterior & Conditionals

\[ \tilde{y} \sim N(X \tilde{\beta}, \sigma^2) \]
\[ \tilde{\beta} \sim N(B_0, V_\beta) \]
\[ \sigma^2 \sim IG(s_1, s_2) \]

\[ P(\beta, \sigma^2) \]
\[ P(\beta | \sigma^2) \]
\[ P(\sigma^2 | \beta) \]
Evaluating Analyses

Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest

Kyle E. Harms, S. Joseph Wright, Osvaldo Calderón, Andrés Hernández & Edward Allen Herre

- Errors?
- Alternatives?

Figure 1 The relationship between the seed density and recruit seedling density for *Trichilia tuberculata* (Meliaceae). The slope of the log–log relationship is less than 1, indicating that recruitment is negatively density dependent. Nonetheless, recruit density increases with seed density. Each symbol represents a census station(s) consisting of one 0.5-m$^2$ seed trap and three 1-m$^2$ seedling plots. Symbol size is proportional to the number of stations (ranging from one to six) with identical counts of seeds and recruits.
Methods

• Likelihood
  – Finding MLE
  – Finding parameter and model CI/PI
  – Comparing models

• Bayes
  – Finding posterior – gives parameter CI, var, etc.
  – Model CI/PI
  – Comparing models
Finding The MLE

- Write Likelihood
- Express as negative log likelihood
- **Analytical**
  - Take derivative wrt each parameter
  - Set each to 0, solve for full set of parameters
- **Numerical**
  - Use numerical algorithm to find minimum
Likelihood CI

- **Parameter**
  - Likelihood profile: Deviance ~ chisq
  - Fisher Information: 
    \[ I = \left. \frac{-d^2 \ln L(\theta)}{d \theta^2} \right|_{\theta_{\text{ML}}} \]
  - Bootstrap
    - Parametric
    - Nonparametric

- **Model CI and PI**
  - Bootstrap
  - Variance Decomposition
    \[ \text{var} [f(x)] \approx \sum \left( \frac{\partial f}{\partial \theta_i} \right) \left( \frac{\partial f}{\partial \theta_j} \right) \text{cov} [\theta_i, \theta_j] \]
Bootstrap

- Monte Carlo method (numerical)
- Based on idea of generating parameter distribution based on large number of replicate data sets that are the same size as original (data random)
- Two variants
  - Parametric: pseudodata
  - Nonparametric: resample data
Non-parametric bootstrap

- Draw a replicate data set by resampling from the original data
- Fit parameters to resample
- Repeat procedure n times
- Estimate parameter CI based on sample quantiles
- Estimate parameter std error as sample s.d.
Parametric bootstrap

- Based on parameters fit to original data set generate pseudodata with same dist'n
- Fit parameters to resample
- Repeat procedure n times
- Estimate parameter CI based on sample quantiles
- Estimate parameter std error as sample s.d.
Numerical Methods for Bayes: Random samples from the posterior

- Approximate PDF with the histogram
- Performs *Monte Carlo Integration*
- Allows all quantities of interest to be calculated from the sample (mean, quantiles, var, etc)

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>5.000</td>
<td>5.000</td>
</tr>
<tr>
<td>median</td>
<td>5.000</td>
<td>5.004</td>
</tr>
<tr>
<td>var</td>
<td>9.000</td>
<td>9.006</td>
</tr>
<tr>
<td>Lower CI</td>
<td>-0.880</td>
<td>-0.881</td>
</tr>
<tr>
<td>Upper CI</td>
<td>10.880</td>
<td>10.872</td>
</tr>
</tbody>
</table>
Model Intervals

• Bayesian model CI and PI were generated from quantiles of model predictions from MCMC
  – CI: parameter uncertainty
  – PI: parameter + data uncertainty (pseudodata)
• The simplest Frequentist CI and PI is based on the bootstrap
  – Implementation is identical except use Bootstrap parameter sample rather than MCMC sample
Likelihood: Model Selection

- Deviance = -2 ln(L)
- Likelihood Ratio Test
  - Nested Models
  - $\Delta$Deviance $\sim$ chisq($\Delta$number of parameters)
  - Gives a p-value
- AIC
  - $\text{AIC} = \text{Deviance} + 2*p$
  - Lowest score wins
Bayesian Model Selection

- \( DIC = 2 \overline{D(\theta)} - D(\bar{\theta}) \)
  \[
  \overline{D(\theta)} = \sum D(\theta_i)/n_g
  \]
  \[
  D(\bar{\theta}) = -2\ln L(y|\bar{\theta})
  \]

- Predictive Loss = P + G
  \[
  \sum \text{var}[y_{rep}] + \sum (E[y_{rep}] - y_{obs})^2
  \]

- Lowest score “wins”