Lesson 6
Maximum Likelihood: Part III
How the likelihood is constructed

- $L = \Pr(\text{data}|\text{model parameters})$
- How is the data modeled?
  - What type of data?
  - What process generated this data?
  - What distributions are an appropriate description of the data?
- How is the process modeled?
- Each analysis should be approached individually
  - Problem solving, creativity
How is the data modeled?

• What type of data is it:
  – Continuous
  – Integer / Count
  – Boolean (0/1)
  – Factor / categorical

• Are there range restrictions on the data?
  – Are negative values allowed?
  – Is there an upper bound?
  – Are the observed data near the bounds?
How is the data modeled?

• What are the dominant sources of variability in the data?
  – Observation/measurement error
  – Process variability
    • Space
    • Time
    • Individual/Site/Species
    • Random
  – Missing data
How is the data modeled?

• Are there multiple processes involved?
  – Zero-inflated data
    • \( Pr(\text{abundance}|\text{present}) \cdot Pr(\text{present}) \)

• Are there multiple types or sources of data?
  – Tree growth: tree rings + DBH
  – Tree fecundity: cone counts + seed trap
  – Tree crown: remote sensing + model + crown class

• Is the process observed directly or inferred?
How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Nonlinear models
- Hierarchical models
- Mechanistic models

Note: Will shy away from “fishing” models except for EDA: regression trees, splines, neural networks, cluster analysis, etc.
A quick beastiary of functions

- Polynomials
- Piecewise polynomials
- Rational (ratio based)
- Exponential based
- Power-based
- Sometimes chosen for mechanistic reasons, sometimes because they “fit right”

Supplemental reading: Bolker Ch 3
Polynomial

- *Linear* with respect to the model parameters
- Taylor series: Can approximate any smooth continuous function

Figure 6: Taylor series expansion of a 4th-order polynomial.
Figure 7: Piecewise polynomial functions: the first three (threshold, hockey stick, general piecewise linear) are all piecewise linear. Splines are piecewise cubic; the equations are complicated and usually handled by software (see \texttt{spline} and \texttt{smooth.spline}).
Rational functions:

- **Hyperbolic:** $f(x) = \frac{a}{b+x}$
- **Michaelis–Menten:** $f(x) = \frac{ax}{b+x}$
- **Holling type III:** $f(x) = \frac{ax^2}{b^2 + x^2}$
- **Holling type IV (c<0):** $f(x) = \frac{ax^2}{b + cx + x^2}$

Figure 8: Rational functions.
negative exponential: $f(x) = ae^{-bx}$

monomolecular: $f(x) = a(1 - e^{-bx})$

Ricker: $f(x) = axe^{-bx}$

logistic: $f(x) = \frac{e^{a+bx}}{1 + e^{a+bx}}$

Figure 9: Exponential-based functions. “M-M” in the monomolecular figure is the Michaelis-Menten function with the same asymptote and initial slope.
Figure 10: Power-based functions. The lower left panel shows the Ricker function for comparison with the Shepherd and Hassell functions. The lower right shows the Michaelis-Menten function for comparison with the non-rectangular hyperbola.
<table>
<thead>
<tr>
<th>Function</th>
<th>Range</th>
<th>Left end</th>
<th>Right end</th>
<th>Middle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Polynomials</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td>([-\infty, \infty])</td>
<td>(y \rightarrow \pm \infty), constant slope</td>
<td>(y \rightarrow \pm \infty), constant slope</td>
<td>monotonic</td>
</tr>
<tr>
<td>Quadratic</td>
<td>([-\infty, \infty])</td>
<td>(y \rightarrow \pm \infty), accelerating</td>
<td>(y \rightarrow \pm \infty), accelerating</td>
<td>single max/min</td>
</tr>
<tr>
<td>Cubic</td>
<td>([-\infty, \infty])</td>
<td>(y \rightarrow \pm \infty), accelerating</td>
<td>(y \rightarrow \pm \infty), accelerating</td>
<td>up to 2 max/min</td>
</tr>
<tr>
<td><strong>Piecewise polynomials</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold</td>
<td>([-\infty, \infty])</td>
<td>flat</td>
<td>flat</td>
<td>breakpoint</td>
</tr>
<tr>
<td>Hockey stick</td>
<td>([-\infty, \infty])</td>
<td>flat or linear</td>
<td>flat or linear</td>
<td>breakpoint</td>
</tr>
<tr>
<td>Piecewise linear</td>
<td>([-\infty, \infty])</td>
<td>linear</td>
<td>linear</td>
<td>breakpoint</td>
</tr>
<tr>
<td><strong>Rational</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperbolic</td>
<td>([0, \infty])</td>
<td>(y \rightarrow \infty)</td>
<td>(y \rightarrow 0)</td>
<td>decreasing</td>
</tr>
<tr>
<td>Michaelis-Menten</td>
<td>([0, \infty])</td>
<td>(y = 0), linear</td>
<td>asymptote</td>
<td>saturating</td>
</tr>
<tr>
<td>Holling type III</td>
<td>([0, \infty])</td>
<td>(y = 0), accelerating</td>
<td>asymptote</td>
<td>sigmoid</td>
</tr>
<tr>
<td>Holling type IV ((c &lt; 0))</td>
<td>([0, \infty])</td>
<td>(y = 0), accelerating</td>
<td>asymptote</td>
<td>hump-shaped</td>
</tr>
<tr>
<td><strong>Exponential-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. exponential</td>
<td>([0, \infty])</td>
<td>(y) finite</td>
<td>(y \rightarrow 0)</td>
<td>decreasing</td>
</tr>
<tr>
<td>Monomolecular</td>
<td>([0, \infty])</td>
<td>(y = 0), linear</td>
<td>(y \rightarrow 0)</td>
<td>saturating</td>
</tr>
<tr>
<td>Ricker</td>
<td>([0, \infty])</td>
<td>(y = 0), linear</td>
<td>(y \rightarrow 0)</td>
<td>hump-shaped</td>
</tr>
<tr>
<td>logistic</td>
<td>([0, \infty])</td>
<td>(y) small, accelerating</td>
<td>asymptote</td>
<td>sigmoid</td>
</tr>
<tr>
<td><strong>Power-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power law</td>
<td>([0, \infty])</td>
<td>(y \rightarrow 0) or (\rightarrow \infty)</td>
<td>(y \rightarrow 0) or (\rightarrow \infty)</td>
<td>monotonic</td>
</tr>
<tr>
<td>von Bertalanffy</td>
<td></td>
<td>like logistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gompertz</td>
<td></td>
<td>ditto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shepherd</td>
<td></td>
<td>like Ricker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hassell</td>
<td></td>
<td>ditto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-rectangular hyperbola</td>
<td></td>
<td>like Michaelis-Menten</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Linear Regression

\[ y_i = a_0 + a_1 x_i + \epsilon_i \]
\[ \epsilon_i \sim N(0, \sigma^2) \]

- Step 1: Likelihood

\[
L = \prod_{i=1}^{n} N(y_i | a_0 + a_1 x_i, \epsilon_i)
= \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[ \frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]
\]
Intercept

\[
L = \left( \frac{1}{\sqrt{2\pi \sigma}} \right)^n \exp \left[ \frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]
\]

\[
\ln L = -n \ln \sigma - n \ln (2\pi) - \frac{1}{\gamma \sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2
\]

\[
\frac{\partial \ln L}{\partial a_0} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)
\]

\[
0 = \sum_{t=1}^{T} y_i - n a_0 - \sum_{t=1}^{T} a_1 x_i
\]

\[
0 = \bar{y} - a_0 - a_1 \bar{x}
\]

\[
a_0 = \bar{y} - a_1 \bar{x}
\]
Slope

\[
\ln L = -n \ln \sigma - n \ln (\pi \pi) - \frac{1}{2 \sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2
\]

\[
\frac{\partial \ln L}{\partial a_1} = \frac{1}{\sigma^2} \sum_{t=1}^{T} x_i (y_i - a_0 - a_1 x_i)
\]

\[
0 = \sum_{t=1}^{T} x_i y_i - \sum_{t=1}^{T} a_0 x_i - \sum_{t=1}^{T} a_1 x_i^2
\]

\[
0 = \overline{xy} - a_0 \overline{x} - a_1 \overline{x}^2
\]

\[
a_1 = \frac{\overline{xy} - a_0 \overline{x}}{\overline{x}^2}
\]
Combining slope and intercept

\[ a_0 = \bar{y} - a_1 \bar{x} \]

\[ a_1 = \frac{\bar{xy} - a_0 \bar{x}}{\bar{x}^2} \]

\[ a_1 = \frac{\bar{xy} - \bar{x} \bar{y} + a_1 \bar{x}^2}{\bar{x}^2} \]

\[ a_1 = \frac{\bar{xy} - \bar{x} \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\text{cov}[x, y]}{\text{var}[x]} \]

\[ a_0 = \frac{x^2 \bar{y} - \bar{x} \bar{xy}}{\text{var}[x]} \]
Variance

\[ \ln L = -n \ln \sigma - n \ln (2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \]

\[ \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = 0 \]

\[ \frac{n}{\sigma} = \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \]

\[ \sigma_{ML}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \]
Matrix notation

\[ y_i = \beta_1 + \beta_2 x_i \]

\[ \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1 x_{i1} + \beta_2 x_{i2} \]

Where \( x_{i1} = 1 \)

\[ \vec{y} = X \vec{\beta} \]

\[ X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \]

Design Matrix
Design Matrices

Multiple linear regression

\[
X = \begin{bmatrix}
1 & x_{12} & \cdots & x_{1k} \\
1 & x_{22} & \cdots & x_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{n2} & \cdots & x_{nk}
\end{bmatrix}
\]

\[
a_1 = \frac{\bar{xy} - \bar{x} \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\bar{x}^2 \bar{y} - \bar{x} \bar{xy}}{\text{var}[x]}
\]

\[
a_0 = \frac{\bar{x}^2 \bar{y} - \bar{x} \bar{xy}}{\text{var}[x]}
\]

\[
\hat{\beta} = (X^T X)^{-1} X^T y
\]
# ANOVA Design Matrices

## One Way Anova
3 levels, n=6

\[
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}
\]

## Two Way Anova
3 levels x 2 levels
2 reps each, n=12

\[
\begin{bmatrix}
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 0 & x_{13} \\
1 & 0 & x_{23} \\
1 & 0 & x_{33} \\
1 & 1 & x_{43} \\
1 & 1 & x_{53} \\
1 & 1 & x_{63} \\
\end{bmatrix}
\]

## ANCOVA
2 levels, 1 covariate
n=6