Introduction to the Course

ESS 575: Models for Ecological Data

N. Thompson Hobbs

January 16, 2019

Graduate Degree Program in Ecology
Department of Ecosystem Science and Sustainability
Colorado State University

What is this course about?

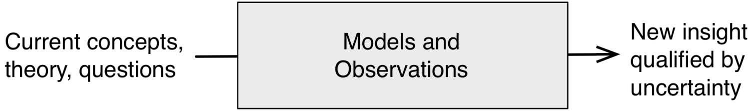
Gaining new insight about ecological processes using models and observations in the Bayesian framework.

$[z_i \vert \boldsymbol{\theta}_p]$	A model of a process
$[y_i \vert z_i, \boldsymbol{\theta}_d]$	A model of the data that arise from the process
$[\boldsymbol{\theta}_p][\boldsymbol{\theta}_d]$	Models of parameters

Exercise

What do statements made by journalists, attorneys, and scientists have in common? What sets the statements of scientists apart?

What is this course about?



What is this course about?

Probability model

Observations (data)

Deterministic model

Idea!

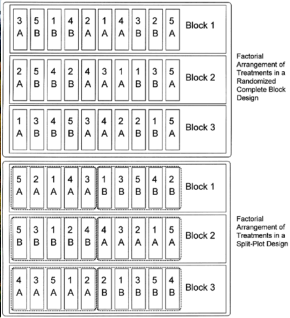
What is the probability that I would observe the data if my model is a faithful representation of the processes that gave rise to the data?

Why this course?

KEY TO STATISTICAL METHODS

	Design or Purpose	Measurement Variables	Ranked Variables	Attributes
1 variable 1 sample	Examination of a single sample	Procedure for grouping a frequency distribution, Box 2.1; stem-and-leaf display, Section 2.5; testing for outliers, Section 13.4 Computing median of frequency distribution, Box 4.1 Computing arithmetic mean: unordered sample, Box 4.2; frequency distribution, Box 4.3 Computing standard deviation: unordered sample, Box 4.2; frequency distribution, Box 4.3 Setting confidence limits: mean, Box 7.2; variance, Box 7.3 Computing g_1 and g_2, Box 6.2		Confidence limits for a percentage, Section 17.1 Runs test for randomness in dichotomized data, Box 18.3
	Comparison of a single sample with an expected frequency distribution	Normal expected frequencies, Box 6.1 Goodness of fit tests: parameters from an extrinsic hypothesis, Box 17.1; from an intrinsic hypothesis, Box 17.2 Kolmogorov–Smirnov test of goodness of fit, Box 17.3 Graphic "tests" for normality: large sample sizes, Box 6.3; small sample sizes (rankit test), Box 6.4 Test of sample statistic against expected value, Box 7.4		Binomial expected frequencies, Box 5.1 Poisson expected frequencies, Box 5.2 Goodness of fit tests: parameters from an extrinsic hypothesis, Box 17.1; from an intrinsic hypothesis, Box 17.2
1 variable $\geqslant 2$ samples	Single classification	Single classification anova: unequal sample sizes, Box 9.1; equal sample sizes, Box 9.4 Planned comparison of means in anova, Box 9.8; single degree of freedom comparisons of means, Box 14.10 Unplanned comparison of means: T-method, equal sample sizes, Box 9.9; T′, GT2, and Tukey–Kramer, unequal sample sizes, Box 9.10; Welsch step-up, Box 9.11; STP test, Section 9.7; contrasts using Scheffé, T, and GT2, Box 9.12; multiple confidence limits, Section 14.10 Estimate variance components: unequal sample sizes, Box 9.2; equal sample sizes, Box 9.3 Setting confidence limits to a variance component, Box 9.3 Tests of homogeneity of variances, Box 13.1 Tests of equality of means when variances are heterogeneous, Box 13.2	Kruskal–Wallis test, Box 13.5 Unplanned comparison of means by a nonparametric STP, Box 17.5	*G*-test for homogeneity of percentages, Boxes 17.5 and 17.8 Comparison of several samples with an expected frequency distribution, Box 17.4; unplanned analysis of replicated tests of goodness of fit, Box 17.5
	Nested classification	Two-level nested anova: equal sample sizes, Box 10.1; unequal sample sizes, Box 10.4 Three-level nested anova: equal sample sizes, Box 10.3; unequal sample sizes, Box 10.5		
	Two-way or multi-way classification	Two-way anova: with replication, Box 11.1; without replication, Box 11.2; unequal but proportional subclass sizes, Box 11.4; with a single missing observation, Box 11.5 Three-way anova, Box 12.1 More-than-three-way classification, Section 12.3 and Box 12.2 Test for nonadditivity in a two-way anova, Box 13.4	Friedman's method for randomized blocks, Box 13.9	Three-way log-linear model, Box 17.9 Randomized blocks for frequency data (repeated testing of the same individuals), Box 17.11

Why this course?



Why this course?

Problems poorly suited to traditional approaches

- Multiple sources of data
- Multiple sources of uncertainty
- Inference across scales
- Unobservable quantities
- Missing data
- Derived quantities
- Forecasting

Why this course?

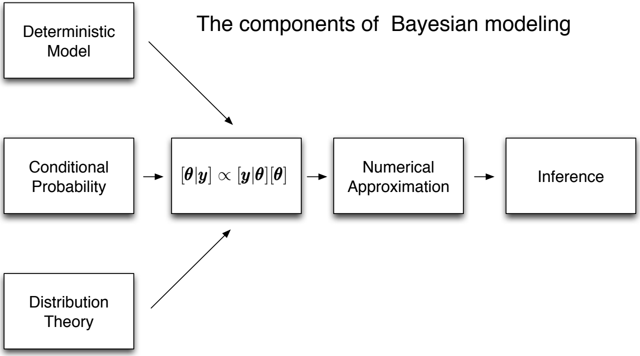
Recent ESS 575 alumni

Student	Position
Ann Raiho	Ph.D., Notre Dame
Megan Vahsen	Ph.D., Notre Dame
Nathan Galloway	Biologist, National Park Service
Nell Campbell	Research Scientist, Univ. New Hampshire
Katie Renwick	Post-doc, Univ. Montana
Alison Ketz	Post-doc, USGS
Zhongqi Miao	Ph.D., Berkeley
Greg Wann	Post-doc, USGS

Goals

- Provide *principles* based understanding
- Enhance intellectual satisfaction
- Foster collaboration
- Build a foundation for self-teaching

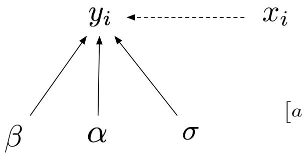
Learning outcomes



Learning outcomes

1. Explain basic principles of Bayesian inference.
2. Diagram and write out mathematically correct posterior and joint distributions for Bayesian models.
3. Explain basics of the Markov chain Monte Carlo (MCMC) algorithm.
4. Use software for implementing MCMC.
5. Develop and implement hierarchical models.
6. Evaluate model fit.
7. Appreciate possibilities for model selection.
8. Understand papers and proposals using Bayesian methods.

Cross cutting theme



$$\mu_i = \frac{m x_i^a}{h^a + x_i^a}$$

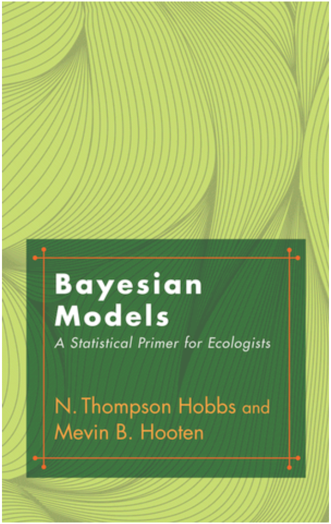
$$[a, h, m, \sigma^2 \mid y] \propto \prod_{i=1}^{n} [y_i | \mu_i, \sigma^2][a][h][m][\sigma^2]$$

```
model{

     for(i in 1:length(y)){

          mu[i] <- (m*x[i]^a)/(h^a+x[i]^a)
          y[i] ~ dgamma(mu[i]^2/sigma^2,mu[i]/sigma^2)

     }

a ~ dnorm(0,.0001)
m ~ dgamma(.01,.01)
h ~ dgamma(.01,.01)
sigma ~ dunif(0,5)
}
```

Teaching philosophy

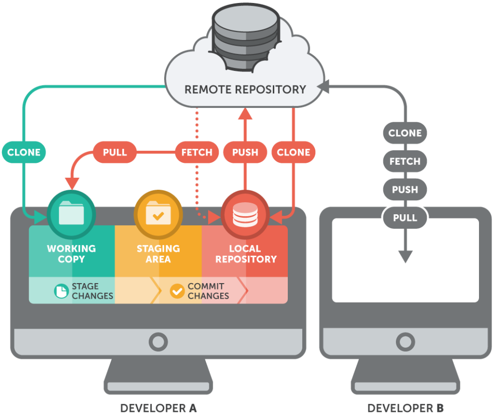
- Principles are primary
- Everyone learns, everyone teaches
- Teaching trumps evaluation.
- The best learning comes from solving problems.
- Whenever possible, I teach in the first person voice.

Text



Errata: http://warnercnr.colostate.edu/~hooten/papers/pdf/Hobbs_Hooten_Bayesian_Models_2015_errata.pdf

Accessing course materials on GitHub



Learn Version Control with Git: A step-by-step course for the complete beginner

https://www.git-tower.com/learn/

Accessing course materials on GitHub

Show possible file structure for course materials on board.

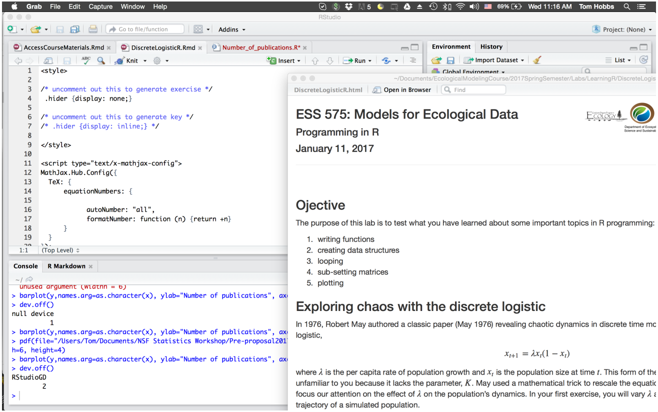
Housekeeping

- Lab in Natural Resources 254
  - You will need a laptop
  - Bring long power cords.
- Lecture in NESB A302 starting promptly at 9:30

Housekeeping

- R primer for first laboratory available on GitHub
- Lecture notes: download morning of class (after 8:30)
- Some board work, so be prepared to take notes.

R markdown


RStudio
AccessCourseMaterials.Rmd
DiscreteLogisticR.Rmd
Number_of_publications.R*
Environment
History
Import Dataset
Global Environment
<style>
/* uncomment out this to generate exercise */
.hider {display: none;}
/* uncomment out this to generate key */
/* .hider {display: inline;} */
</style>
<script type="text/x-mathjax-config">
MathJax.Hub.Config({
TeX: {
equationNumbers: {
autoNumber: "all",
formatNumber: function (n) {return +n}
Console
R Markdown
> barplot(y,names.arg=as.character(x), ylab="Number of publications",
> dev.off()
null device
1
> pdf(file="/Users/Tom/Documents/NSF Statistics Workshop/Pre-proposal201
h=6, height=4)
RStudioGD
2
DiscreteLogisticR.html
Open in Browser
ESS 575: Models for Ecological Data
Programming in R
January 11, 2017
Ojective
The purpose of this lab is to test what you have learned about some important topics in R programming:
1. writing functions
2. creating data structures
3. looping
4. sub-setting matrices
5. plotting
Exploring chaos with the discrete logistic
In 1976, Robert May authored a classic paper (May 1976) revealing chaotic dynamics in discrete time mc
logistic,
$x_{t+1} = \lambda x_t(1 - x_t)$
where λ is the per capita rate of population growth and x_t is the population size at time t. This form of the
unfamiliar to you because it lacks the parameter, K. May used a mathematical trick to rescale the equatic
focus our attention on the effect of λ on the population's dynamics. In your first exercise, you will vary λ a
trajectory of a simulated population.

Evaluation

- Eleven laboratory exercises worth 50 - 100 points each. (65% of grade)
- A mid-semester challenge (20% of grade)
- A capstone problem done individually (25% of grade)
- You are graded relative to material, not relative to each other.
- Relax. You will get an A if you do the assignments carefully and thoughtfully.
- See syllabus for details.

Individual projects

- Purpose
- Process
- Product

Getting help

- From me: Tuesday-Thursday 11:00 - 12:00 or by appointment, NESB B227 or by email (tom.hobbs@colostate.edu). Please put ESS 575 in subject line.
- From TA, Brian Avila: Fridaya 12:30-2:30 or by appointment, Wagar 203, mlvahsen@gmail.com.

Chores

- Fill out Google doc spreadsheet if you have not already done so.
- Get account on GitHub and pull repository ESS_575_2019 to your local machine. See instructions in `Accessing course material.html`.
- Install R and R studio before lab tomorrow. See instructions in R primer.
- Install the R package `ESS575` containing course data library. See instructions in `Accessing course material.html`.
- Print R primer for first laboratory.
- Read materials in `Admin` folder of ESS_575_2019.

First assignment

- Read the syllabus.
- Prepare ≤ 2 minute presentation about yourself: background, what are you studying, who is your major professor, why you are taking this class.
- Prepare a 1-2 paragraph description of an important non-linear, static, deterministic model in your field of ecology. See `FirstAssignment.pdf` in Admin folder of ESS_575_2019. Due Friday.
- Dust off your calculus book. Review the definite integral and how it is derived.

Discussion topic (if time)

What do you think of when someone is described as an "ecological modeler?" Mevin and I say in our book that all ecological researchers are modelers. Why do you suppose we say that? Describe your ideas about the relationships among observations, mathematical models, and statistical models in ecology.