Monday Lecture:

Outline:

I. Opening
   a. Introductions
   b. Wiki location
   c. Software
   d. Resources/facilities
   e. Plan for the week
II. What is statistics? Why do we do statistics? <interactive>
   a. The study of variation? Data?
   b. The science of separating noise from pattern (stochastic from deterministic). One person's noise is another person's dissertation.
   c. Testing hypotheses/drawing inferences
   d. Making decisions, predictions

"The application of mathematics to understanding variation, testing hypotheses, generating predictions, and making decisions." ?

III. Think of data as consisting of pattern/process + variation
   a. We will attempt to describe pattern/process using deterministic relationships.
   b. And whatever we consider to be variation, drawn from our understanding of stochasticity or probability theory.
   c. Introduce notation? Data ~ variation(covariates + parameters)
IV. Estimation/Goodness of Fit
   a. To be useful this concept of data as consisting of stochastic and deterministic processes requires that we come up with some way of estimating the value of parameters such that these processes fit the observed data as closely as possible.
   b. Many different estimation methods exist
      i. Least-squares
      ii. Maximum likelihood
      iii. Bayesian methods (MCMC, gibbs sampling, metropolis hastings algorithms)
V. Philosophies of inference:
   a. To the extent that our models of data are really hypotheses about the data and underlying processes, to make progress in our analyses (and in science), we need some framework that allows us to draw conclusions (or inferences) about our models/hypotheses.
      i. Strong/weak inference.
      ii. Parsimony
   b. Long-standing, contentious debates have raged over what this framework should be. Major camps include
      i. Frequentist approaches/null hypothesis testing
      ii. Multiple hypothesis testing/MLE

iii. Bayesians

c. Estimation methods have traditionally been strongly associated with particular inference frameworks, but it doesn't have to be this way.

d. This week, we'll be focusing on Maximum likelihood estimation:

i. What it is

1. In the language of conditional probabilities, p(theta|data), rather than p(data|theta)…

ii. Why use MLE?

iii. How to design models and estimate parameters using MLE

iv. How to draw inferences based on MLE results

<Break>

VI. Probability theory

a. Introduce notation,

b. Joint probability

c. Conditional probability

i. Mention that these concepts will show up next week.

VII. Distributions/derivations

a. Binomial

b. Poisson

c. Normal

d. Point out distribution tools/resources available

VIII. Likelihood

a. How to calculate likelihoods, given our probability theory intro

b. Joint likelihood of multiple data values

c. Lab activity

IX. Maximum Likelihood

a. Takes likelihood and allows us to start drawing inferences.

b. How to maximize a function (likelihood or otherwise)

i. Analytically (demo with poisson in lecture)

ii. Numerically (demo with sim_data_lab.R)

c. Lab activity

i. Students solve binomial analytically

ii. Students solve normal distribution analytically (provide hints)

Loose ends: Model comparison?

Afternoon Lab:

I. Spill-over from morning lecture (Intro to likelihood activity)

II. Species diversity and fire

Use collection of stats quotes?

Chpts 1-2 of ED book

Philosophies of Inference

- really @ the heart of science / scientific Method: How do we evaluate competing ideas given the evidence we can assemble (from observation + experiment)?
    - long standing, contentious debates ensue

Popperian view (associated w/ "frequentist" paradigm)

Is the classic / standard approach to science, and is reflected in our standard statistics.

"Hypotheses can only be falsified, never proven"
Confrontation of null hypotheses with a single alternative hypothesis

↳ repeated over and over; hypothesis that is repeatedly not falsified gains respect, and is often treated as truth.

Standard approach:

focus on a single null hypothesis (assume value of certain parameters), and calculate the probability that the data would have been observed if the null Hyp. were true

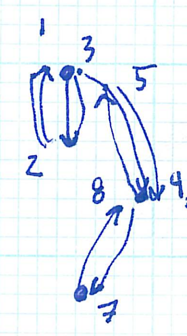
↳ decision is made based on this conclusion

Eg, Normal distribution, $\mu \neq 0$ vs $\mu = 0$

$$P(x|\theta) \quad \text{or} \quad P(x|\mu=0) \sim N(\mu, \sigma^2)$$

P(data given theta)

① Confrontation btwn single hypothesis + data



② Critical Expt's

③ Falsification as only truth

Lakatosian View (1978)

: Confrontation of Multiple hypotheses w/ data as Arbitrator
- don't reject hypth's unless we come up with something better.

More pragmatic view of science
↳ work with what we've got
↳ make decisions w/ the best understanding we have

Often we don't have just one hypothesis

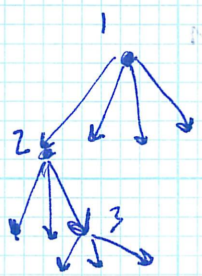
Many competing and sometimes comparable ideas can exist

Likelihood + Bayesian approaches are well suited to this sort of paradigm.
↳ they both produce measures of the likelihood or chance of the model (expressed as its parameters) given the data,

$P(\theta | x)$ or $\mathcal{L}(\theta | x)$

↳ innate appeal to me b/c x is "true", "observed"

we can compare likelihoods of models to compete models against each other, mediated by the data observed



Parsimony - simpler models are better

we can't explain all variation all the time, or our models get too complex

AE Quote about simple models

We'll focus on MLE this week - Thing's we'll cover

1. what is it?
   - estimation / Goodness of fit method allowing us to calculate the most likely values of parameters, given observed data
   - we can generate multiple models, determine their maximum likelihood (given best fit pars) and compare them with each other given these likelihoods.
2. why use it?
3. How to design models + estimate pars using MLE
4. How to draw inferences about models describing variation in biological systems given MLE results.

Break

After which we'll cover

1. Prob. theory
2. stochastic distrib's / derivations
3. Likelihood
4. Maximizing likelihood

Lunch

Lab

Probability Theory

- body of mathematical ideas / tools that enable us to study variability through the lense of probability.

Probability theory starts by thinking about the set of all possible things (events) that could happen (called the sample space), and then considers the probability with which a particular thing (or event), or combination of events occurs.

Review of some basic probability rules

- imagine we're sampling chiclid fish in lake Malawi
- Our sample space consists of the population we're sampling from, and the properties of the fish we focus on.

Fish can be either:

1. Red or Blue (R or B)
2. Male or female (M or F)

Rules

1. $P(R)$ is equivalent notation for "probability fish is Red"
   $P(\text{event})$ is how we'll denote probability.

2. Events are mutually exclusive if only 1 of the set of events can happen at once. Eg, a fish can't be both M and F.
   When events are mutually exclusive, the probability with which either event occurs equals the sum of their individual probabilities.

$$P(A \cup B) = P(A) + P(B)$$
$$P(\text{male or female}) = P(\text{male}) + P(\text{female})$$

2b Sum of all mutually exclusive events $= 1$

3. If two events are not mutually exclusive, they have some probability of occuring at the same time. We call this probability of co-occurence the Joint Probability of the events and denote it as such: $P(A \cap B)$; $P(\text{male and blue})$

④ Joint probabilities allow us to modify our earlier rule describing the probability of mutually exclusive events.

previously, $P(M \cup F) = P(M) + P(F)$

what about $P(M \cup B) = ?$ not mutually exclusive.

$$P(M \cup B) = P(M) + P(B) - \underbrace{P(M \cap B)}_{\text{accounts for double Counting.}}$$

⑤ The <u>conditional probability</u> of 2 events, or the probability of event A given that event B has happened, is denoted

$P(A|B)$ or $P(\text{male given Blue})$

And can be calculated

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad P(M|B) = \frac{P(M \cap B)}{P(B)}$$

The "|" notation will show up in our definitions of likelihood and this rule/concept of conditional probabilities is essential for understanding the foundation of Bayes Rule / Bayesian analysis and will show up again next week.

⑥ Two events are <u>independent</u> if the conditional prob. of one event given the other is equal to the prob. of the event itself. (ie, knowledge of the second event doesn't change our understanding of how probable the first event is)

<u>Or</u>, $P(A|B) = P(A)$

This implies that $P(A \cap B) = P(A)P(B)$

$$\left[\begin{array}{l} P(A|B) = \dfrac{P(A \cap B)}{P(B)} \\ P(A) = \dfrac{P(A \cap B)}{P(B)} \\ P(A)P(B) = P(A \cap B) \end{array}\right.$$

<u>So</u>, multiplying probabilities of independent events is how we can determine their <u>joint probability</u> of occurence.

This holds for >2 events...

$$P(A \cap B \cap C \cap \ldots) = P(A)P(B)P(C)\ldots$$

<u>Or</u> $\log(P(A \cap B)) = \log(P(A) \cdot P(B)) = \log P(A) + \log P(B)$

From these rules, we can derive a number of

Probability distributions

"functions that describe the probability of an event with a certain value occuring"

Collectively, the probability of all events described by a distrib. must sum (or integrate) to 1.

Events can be:

Discrete	Continuous
- Heads vs Tails	- how long it takes to drink a cup of coffee
- roll of a dice	- The above ground biomass of a plant
- number of surviving individuals	

We will classify distributions based on whether or not they describe discrete or continuous events (or data) as well as the number of parameters, or variables, it takes to completely specify the distribution, and the range of events / data possible.

<project EMD book>
Table 4.1

Binomial Distribution:

Describes the probability of observing X number of successes out of a total of N attempts, given that the probability of success in a single attempt is p.

Example: Flipping a coin. Heads = a success. The binomial distribution helps us calculate your probability of getting, say, 2 heads out of 3 tosses. (of a fair coin)

Prob. 1 Head:

$$P(H) = p = 0.5 = \frac{1}{2}$$

Prob. 2 Heads: $P(H) \cdot P(H) = \frac{1}{2} \cdot \frac{1}{2} = p^2$

Prob. 1 Tail: $1 - P(H) = P(T) = (1-p)^1$ (mutually exclusive events sum to 1)

Prob 2 Heads and 1 Tail: $P(H)\,P(H)\,P(T) = p^2 (1-p)^1$

Normalizing factor: $\binom{3}{2} = \frac{3!}{2!\,(3-2)!}$

accounts for the # of different ways you can draw 2H+1T
$\begin{pmatrix} HHT \\ HTH \\ THH \end{pmatrix}$ relative to other combos

- Don't sweat this part, it just makes sure the distribution sums to 1.

So

$$P(X=2) = \binom{3}{2} p^2 (1-p)^1$$
$$= \frac{6}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(1-\frac{1}{2}\right)^1$$
$$= 3 \cdot \left(\frac{1}{2}\right)^3 = 3/8$$

We can generalize this to N number of ~~exam~~ attempts / trials

$$P(x|p) = P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

← prob. of failure raised to # of failures, $(N-x)$

Normalizing factor

Probability of single success raised to the # of successes total

Mean: $N \cdot p$

Variance: $Np(1-p)$

Range: discrete, $0 \le x \le N$

Parameters: $p \in [0,1]$, N

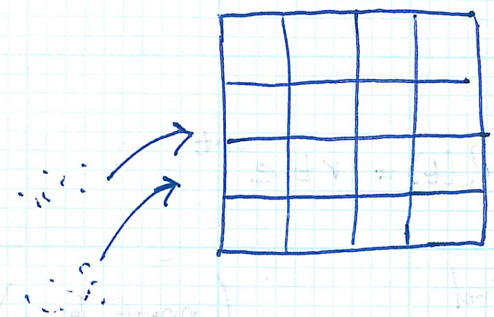
Poisson Distribution

The poisson distrib. models the probability distrib. of the number of individuals/arrivals/events/counts/etc that occur within a given interval of time/space/unit of observation effort, if:

- each event is independent of all others
- the rate at which events occur is constant

Maps a continuous process (arrivals in time) onto a discrete outcome (# of individuals).

Think about space:



- each square is a sampling "unit"
- ~~seeds~~ dispersing seeds rain down on this surface, landing at a random location, at some rate r
- If we look at the number of squares with $0, 1, 2, \ldots, n$ number of seeds relative to the number of squares observed, we have a poisson distribution.

⊛ **Derivation** (if time allows)

let $P_0 =$ fraction of squares w/ 0 seeds
$P_1 =$ " " " " 1 seed
...

Set this up like a differential equation model

$$\boxed{P_0} \xrightarrow{r} \boxed{P_1} \xrightarrow{r} \boxed{P_2} \longrightarrow \ldots$$

Initially: - no seeds fallen

$1 \qquad 0 \qquad 0$ at $t=0$

$$\begin{cases} \dfrac{dP_0}{dt} = -rP_0 \\ \dfrac{dP_1}{dt} = rP_0 - rP_1 \\ \dfrac{dP_2}{dt} = rP_1 - rP_2 \\ \ldots = \ldots \end{cases}$$

We can solve this

Start with:

$$\frac{dP_0}{dt} = -rP_0$$

$$\frac{dP_0}{P_0} = -r\,dt$$

$$\int \frac{dP_0}{P_0} = \int -r\,dt$$

integration constant

$$\ln(P_0(t)) = -rt + C$$

What is C?

- Given $P_0(0) = 1$,

$$\ln(P_0(0)) = -r \cdot 0 + C$$

$$\ln(1) = C$$

$$\hookrightarrow C = 0$$

exponentiating, $P_0(t) = e^{-rt+C}$

$$P_0(t) = e^{-rt}$$

Next:

$$\frac{dP_1}{dt} = rP_0 - rP_1$$

Substitute: $\frac{dP_1}{dt} = re^{-rt} - rP_1$] Solve this (omitted) $\Rightarrow$ $P_1(t) = rte^{-rt}$

Repeat this process over and over, and you find (proof by induction)

P_0	P_1	P_2	P_3	...	P_n
e^{-rt}	rte^{-rt}	$\frac{(rt)^2 e^{-rt}}{2!}$	$\frac{(rt)^3 e^{-rt}}{3!}$		$\frac{(rt)^n e^{-rt}}{n!}$

Poisson Distribution:

$$P(n|\lambda) = P(n) = \frac{e^{-rt}(rt)^n}{n!} \quad \text{or} \quad \frac{e^{-\lambda}(\lambda)^n}{n!} \quad \text{given } rt = \lambda$$

Mean: λ

Variance: λ

Range: discrete, $0 \le n$

Parameters: λ, expected # / sample