Session 2: Introduction to Probability

Foundations of Quantitative Ecology (EEOB 8896.11)

Paul J. Hurtado
(hurtado.10@mbi.osu.edu)

Mathematical Biosciences Institute (MBI)
The Ohio State University

Updated Lecture Slides from August 28, 2013.
Last compile: September 4, 2013

Why Probability?

One answer: **Statistics**.

Why Probability?

One answer: **Statistics**.

But more importantly...

- Biological processes are noisy! (See Jagers 2010)

Why Probability?

One answer: **Statistics**.

But more importantly...

- Biological processes are noisy! (See Jagers 2010)
- The fundamental units in biology are *individuals*.
  Thus, *Demographic (intrinsic)* noise is commonplace.

Why Probability?

One answer: **Statistics**.

But more importantly...

- Biological processes are noisy! (See Jagers 2010)
- The fundamental units in biology are *individuals*.
  Thus, *Demographic (intrinsic)* noise is commonplace.
- Environments are constantly changing!
  Thus, *Environmental (extrinsic)* noise is also ubiquitous.

Why Probability?

One answer: **Statistics**.

But more importantly...

- Biological processes are noisy! (See Jagers 2010)
- The fundamental units in biology are *individuals*.
  Thus, *Demographic (intrinsic)* noise is commonplace.
- Environments are constantly changing!
  Thus, *Environmental (extrinsic)* noise is also ubiquitous.

In short, **ALL models of living systems are, or are simplifications of, stochastic models.**

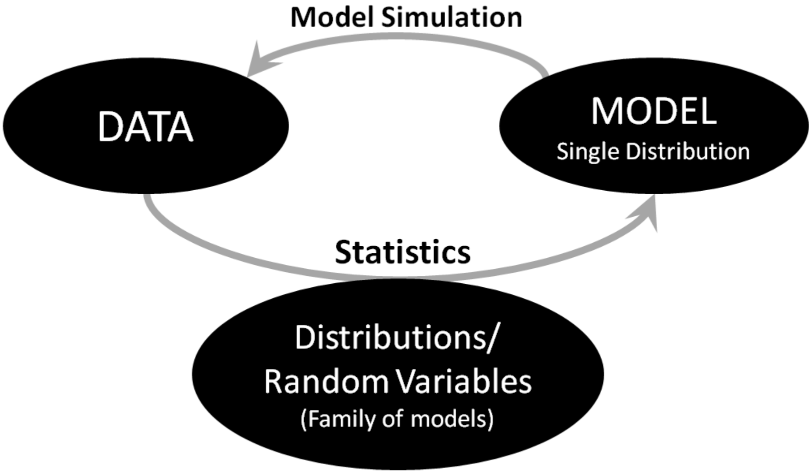
Why Probability?

So, *how/why* are probability concepts widely used in science?

A few examples...

- Simulation (sampling from distributions, e.g., to mimic data)
- Qualitative properties (expected values, expected deviations)
- Approximation (Law of Large Numbers)
- Deriving relationships (e.g., functional forms) and other models
- Statistics (e.g., Maximum Likelihood = Maximum density!)

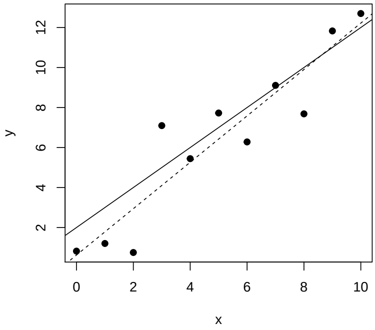
Conceptual Framework



Simulation

Example: Linear Regression $y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim N(0, \sigma)$.

```
set.seed(1492); ## ?set.seed or ask me :-)
x=0:10;
b0=2; b1=1; sig=2;   y=b0+b1*x+rnorm(length(x),0,sig);
plot(x,y,pch=19); abline(b0,b1);
abline(lm(y~x,data=data.frame(x,y)),lty=2)
```



Distribution Properties

Mean vs Expected value? Standard Deviation? Moment Generating Function? Conjugate Distributions (Baysian prior & posterior)?

```
x = rbinom(100, 20, p = 0.2)
mean(x)  ## Compare mean(x) vs. E(x)=n*p

## [1] 4.06

sd(x)  ## Compare sd(x)^2 vs. Var(x)=n*p*(1-p)

## [1] 1.699

sqrt(20 * 0.2 * (1 - 0.2))

## [1] 1.789
```

General mathematical results (aka Analytical results) are really powerful, *if* we can find them! They give general answers to our scientific questions, guide biological intuition, and speed up computations.

Approximation & Deriving Other Models

Computation: Gillespie's Stochastic Simulation Algorithm is driven by "coin tosses" (aka *Bernoulli random variables*) – to speed up computations, approximate multiple coin tosses with a single *binomial distribution*.

Approximation & Deriving Other Models

Computation: Gillespie's Stochastic Simulation Algorithm is driven by "coin tosses" (aka *Bernoulli random variables*) – to speed up computations, approximate multiple coin tosses with a single *binomial distribution*.

Statistical assumptions: Error distributions may be known (for mechanistic reasons) to be, e.g., Binomial, which can sometimes be approximated by a Normal distribution.

Approximation & Deriving Other Models

Computation: Gillespie's Stochastic Simulation Algorithm is driven by "coin tosses" (aka *Bernoulli random variables*) – to speed up computations, approximate multiple coin tosses with a single *binomial distribution*.

Statistical assumptions: Error distributions may be known (for mechanistic reasons) to be, e.g., Binomial, which can sometimes be approximated by a Normal distribution.

Dynamics: We might want to ignore the noise, and just look at averages. Ex: Lotka-Volterra-type foodweb models are really useful!

Approximation & Deriving Other Models

Computation: Gillespie's Stochastic Simulation Algorithm is driven by "coin tosses" (aka *Bernoulli random variables*) – to speed up computations, approximate multiple coin tosses with a single *binomial distribution*.

Statistical assumptions: Error distributions may be known (for mechanistic reasons) to be, e.g., Binomial, which can sometimes be approximated by a Normal distribution.

Dynamics: We might want to ignore the noise, and just look at averages. Ex: Lotka-Volterra-type foodweb models are really useful!

General Results: Model approximation (or considering special cases of a model) can yield well understood (approximate) models for which useful, general results already exist!

Statistics: Maximum Likelihood

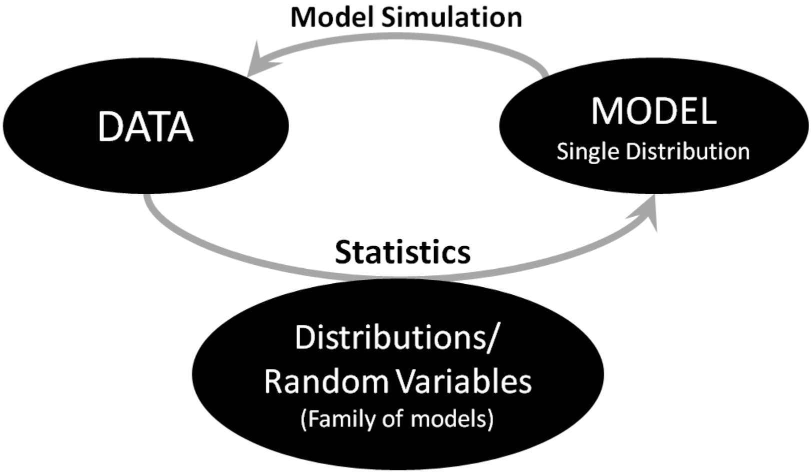
Up to this point, we think of density functions as having fixed parameters $\theta = (\theta_1, ..., \theta_k)$, with arbitrary input value x. *Likelihood functions* are **the exact same functions** except the "inputs" are fixed data values x_1, ... x_n and our parameters are the arbitrary inputs of interest. Specifically, we want the parameters that maximize our likelihood function value for this particular data set.

Statistics: Maximum Likelihood

Up to this point, we think of density functions as having fixed parameters $\theta = (\theta_1, ..., \theta_k)$, with arbitrary input value x. *Likelihood functions* are **the exact same functions** except the "inputs" are fixed data values x_1, ... x_n and our parameters are the arbitrary inputs of interest. Specifically, we want the parameters that maximize our likelihood function value for this particular data set.

To see how all this works, we need to start looking at probability distributions in detail.

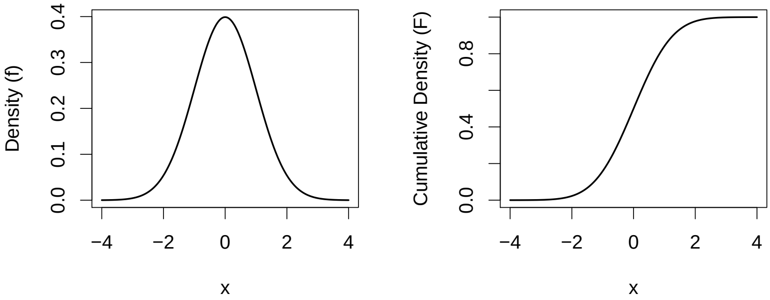
Conceptual Framework



Distribution & Density Functions

Two ways to think about the Normal distribution (a *continuous* distribution) from the relationship: $F(x) = \int_{-\infty}^{x} f(s)ds$

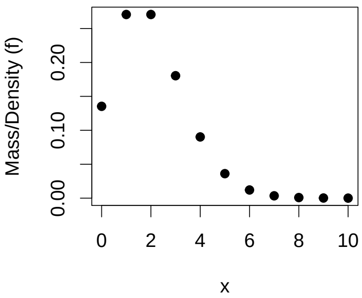
```
## Standard normal density function f(x) and distribution function F(x)
par(cex = 1.4)
x = seq(-4, 4, length = 200)
plot(x, dnorm(x, mean = 0, sd = 1), type = "l", lwd = 2, ylab = "Density (f)")
plot(x, pnorm(x), type = "l", lwd = 2, ylab = "Cumulative Density (F)")
```

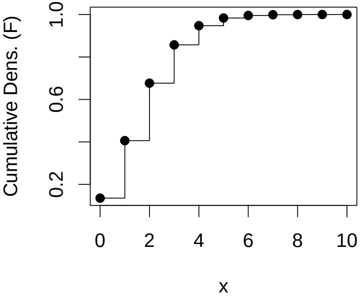


Distribution & Density Functions

Discrete distributions: replace integrals with sums. $F(x) = \sum_{i=0}^{x} f(i)$

```
## Poisson (mean 2) density and distribution functions
x = 0:10
par(cex = 1.4)
plot(x, dpois(x, lambda = 2), pch = 19, ylab = "Mass/Density (f)")
plot(x, ppois(x, lambda = 2), type = "s", ylab = "Cumulative Dens. (F)")
points(x, ppois(x, lambda = 2), pch = 19)
```





Exercises

1. Which distributions are approximately Normal? For which parameter constraints?

   TIP: Type `?Distributions` in the R console. Focus on familiar distributions. How do parameters affect distribution shape? Confirm your intuition by varying parameters and plotting distributions. See http://www.math.wm.edu/~leemis/chart/UDR/UDR.html

2. Plot the density and distribution functions for these distributions (see slides for code). Do this in a 2x2 figure: In the top row, compare to a Normal with the same mean and variance. In the bottom row, do the same but with parameters where the normal approximation fails.

   TIP: Google "R par mfrow example" for multi-panel plot examples. See `?par`.

3. Advanced option: Automate this with a for loop, or use the `lattice` or `ggplot2` packages for the graphics.

Density & Likelihood Functions

Density & Likelihood functions are the same!

Key distinction: As a *density function*, parameters are constants. As a *Likelihood function*, those parameters become the independent variables (inputs). Why do this? With fixed x values (data), the parameter set that yields the largest likelihood function value is the desired *Maximum Likelihood Estimate* (MLE).

Density & Likelihood Functions

Density & Likelihood functions are the same!

Key distinction: As a *density function*, parameters are constants. As a *Likelihood function*, those parameters become the independent variables (inputs). Why do this? With fixed x values (data), the parameter set that yields the largest likelihood function value is the desired *Maximum Likelihood Estimate* (MLE).

Ex: (Binomial) Suppose *random variable* x is Binomial with $size = 10$, $prob = 0.2$. That is, x is the number of heads in 10 coin tosses with an unfair coin that lands heads with probability 0.2. It has density f and likelihood $\mathcal{L}$ (n is known), shown here with their independent variables in red:

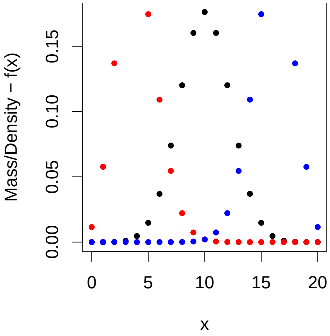
$$f(x = \color{red}{k} | size = n, prob = p) \equiv \binom{n}{\color{red}{k}} p^{\color{red}{k}} (1-p)^{n-\color{red}{k}}$$

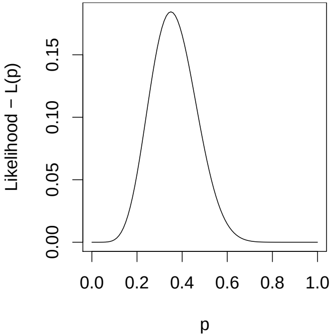
$$\mathcal{L}(\color{red}{p} | size = n, x = k) \equiv \binom{n}{k} \color{red}{p}^{k} (1-\color{red}{p})^{n-k}$$

Density & Likelihood Functions

Generally, data x from a distribution with density $f(x|\theta)$ has likelihood function $\mathcal{L}(\theta|x) \equiv f(x|\theta)$. Compare the horizontal axis labels below:

```
## If a binomial experiment yields x=7 heads out of n=20 trials,
## what value of p maximizes the likelihood function?
plot(x, dbinom(x,size=20,prob=0.5), pch=20, ylab="Mass/Density - f(x)")
points(x, dbinom(x,size=20,prob=0.2), pch=20, col="red")
points(x, dbinom(x,size=20,prob=0.8), pch=20, col="blue")
plot(p, dbinom(k,size=20,prob=p), type="l", ylab="Likelihood - L(p)")
```

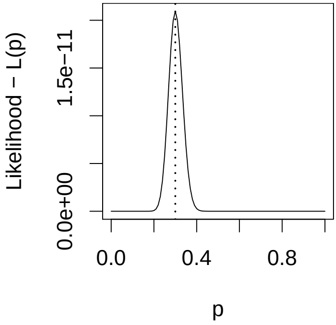




Density & Likelihood Functions

Real data are often *independent & identically distributed* (iid) replicates. The joint distribution of an iid data set $x = (x_1, ..., x_N)$, each x_i with distribution f, is the product of their density functions. Thus, the joint density $f_N(x|\theta)$ (& likelihood $\mathcal{L}$) is $\prod_i f(x_i|\theta)$.

```
## Same experiment (p=0.3), 10 replicates yield k=8,6,7,7,6,2,7,10,1,6
plot(p, sapply(p,function(x) prod(dbinom(k,size=20,prob=x))),
        type="l", ylab="Likelihood - L(p)")
```



With more observations, the likelihood function narrows and gives a more confident estimate. (See previous)

Note that density $f(x|n,p)$ takes `length(x)=10` inputs, while likelihood $\mathcal{L}(p|n,x)$ is a univariate function (one input; p).

For more on Maximum Likelihood and statistical probability for ecology and evolutionary biology, see the relevant chapters in Ben Bolker's book (draft chapters online as PDFs) Ecological Models and Data in R at:

http://ms.mcmaster.ca/~bolker/emdbook/