From Probability to Likelihood

Eric Scott

2020-01-21

TUFTS ANALYTICS WITHOUT BORDERS CONFERENCE

FRIDAY & SATURDAY
FEBRUARY 28-29, 2020

https://sites.tufts.edu/naumovalabs/tawb2020/

SUBMIT YOUR ABSTRACTS BY FEB. 14TH!

JOIN US FOR 1.5 DAYS AT THE FRIEDMAN SCHOOL!

Interactive Workshops

Professional Networking Luncheon

Panel Discussions

Oral & Poster Presentations

Circulate Your Resume

GRADUATE $20
UNDERGRAD $15

CONTACT US AT:

RYAN SIMPSON,
RYAN.SIMPSON@TUFTS.EDU

ELIZABETH MARSH,
ELIZABETH.MARSH@TUFTS.EDU

Sponsored by:

ASA American Statistical Association — Promoting the Practice and Profession of Statistics

Tufts University Institute of the Environment

Tufts University Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Analytics Without Borders

nutrition data summit

Tufts University Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Bentley University

Bryant University

Exams

- Short answer/practical problems
- Calculations and coding (bring a laptop)
  - btw, you can check out laptops for 4 hours from the library
- Write out code (it will be minimal)
- formulae and function cheat-sheet provided.

Textbook Definition of Probabilty

"If an observation is made N times and event A occurs N_A times, then with a high degree of certainty, the relative frequency of N_A/N is close to $P(A)$, the probability of A in a single trial, $P(A) \approx N_A/N$, provided N is sufficiently large."

- "event" = something that does or does not occur
- "trial" = single observation or data point or "experiment" in which the event can occur or not

- P(A) = probability that event (A) will occur in a single trial
- N = total number of trials
- N_A = number of times A occurs

$$P(A) = \lim_{N \to \infty} \left(\frac{N_A}{N} \right)$$

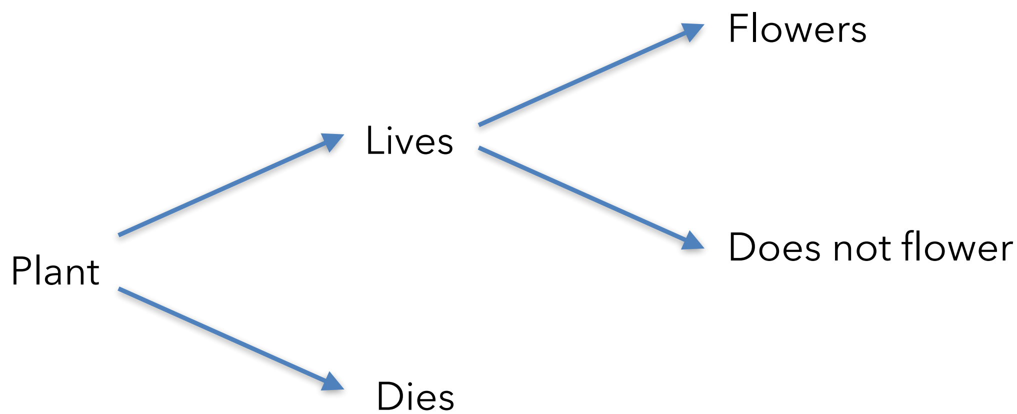
Example application: Demography of a perennial wildflower

$$P(A) = \lim_{N \to \infty} \left(\frac{N_A}{N} \right)$$

P(survival) = 4/5 = 0.8

P(flowering) = 1/4 = 0.25

Plant #	fate
1	Flowers
2	Vegetative
3	Vegetative
4	Vegetative
5	Dead

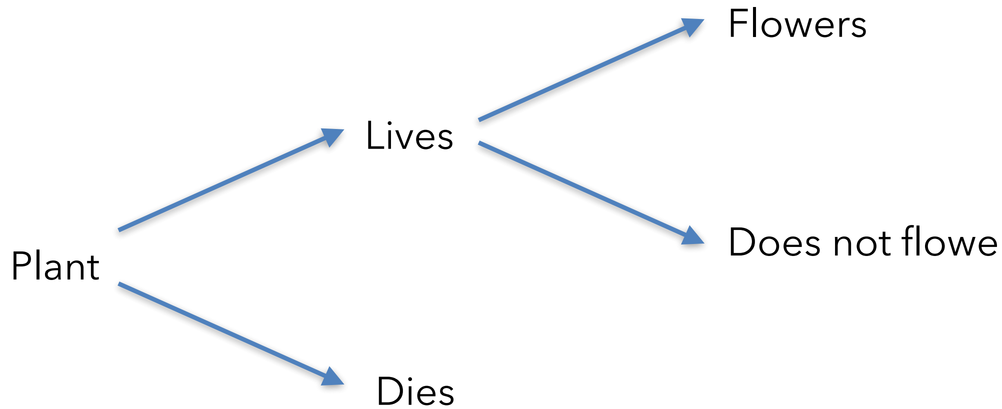


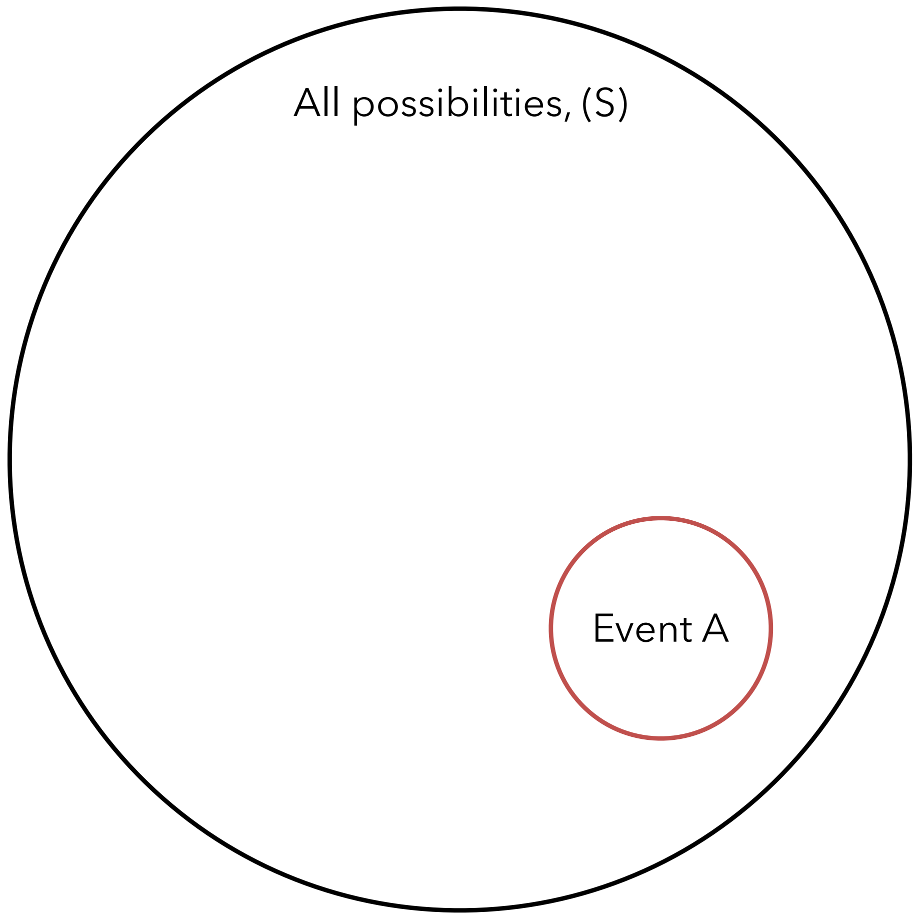
Example application: Demography of a perennial wildflower

What is the probability that a plant survives AND flowers?

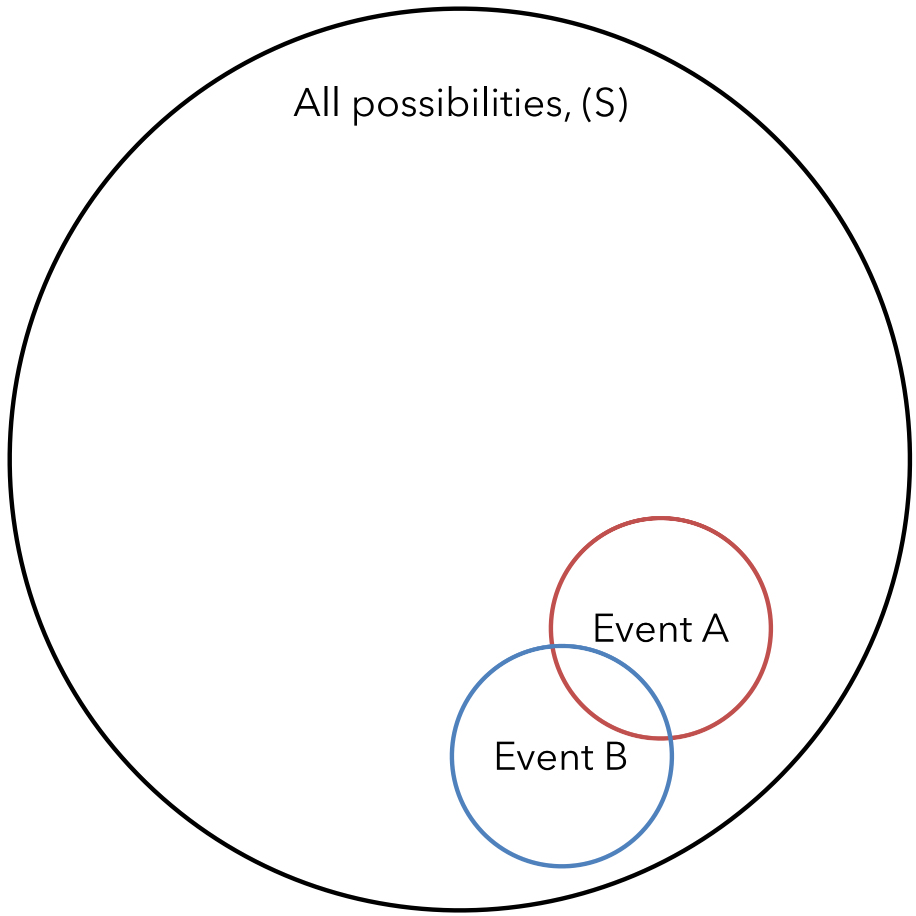
Need to know about the rules of probability…

Plant #	fate
1	Flowers
2	Vegetative
3	Vegetative
4	Vegetative
5	Dead

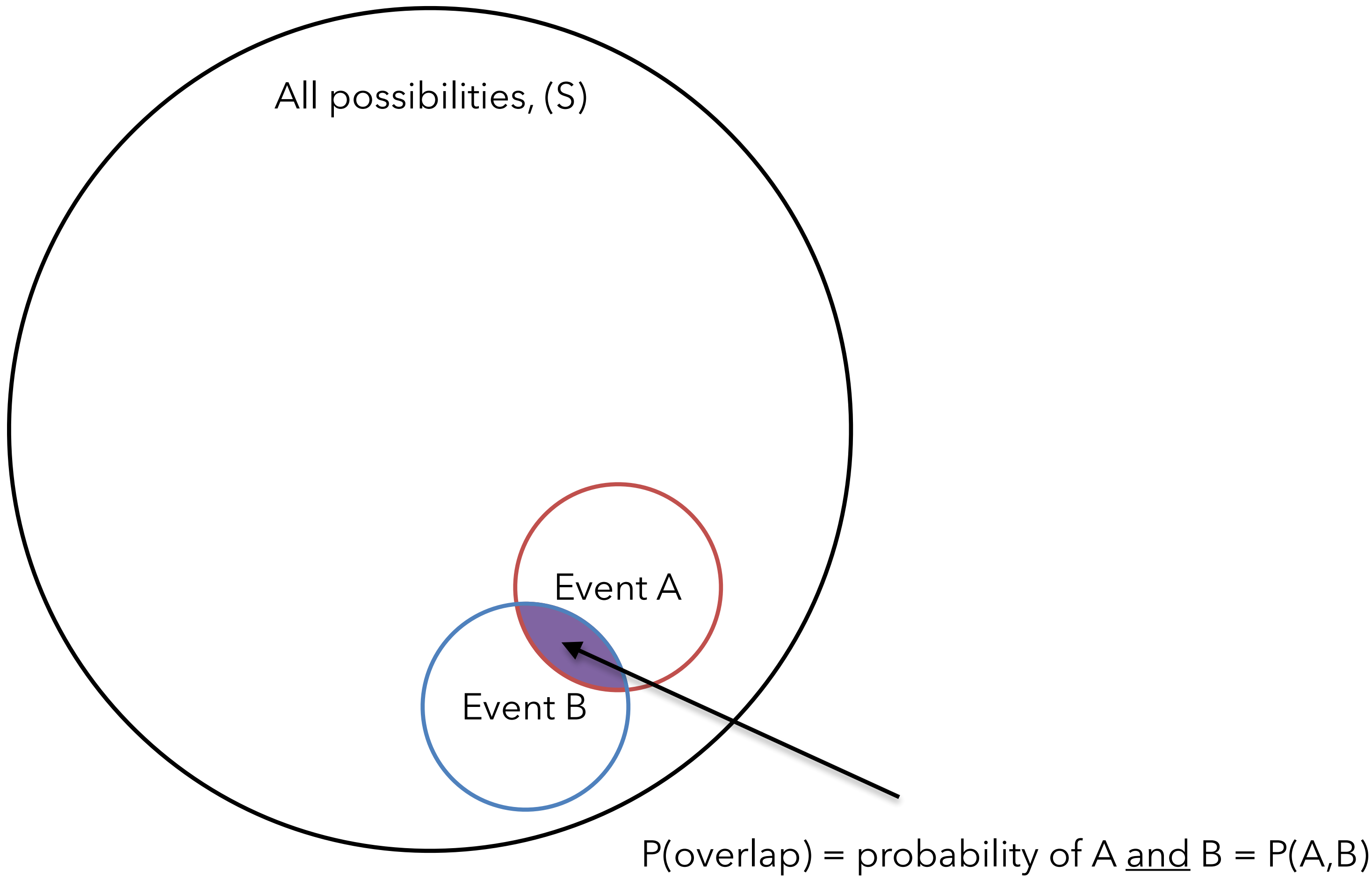




$$P(A) = \frac{\text{area A}}{\text{area S}}$$

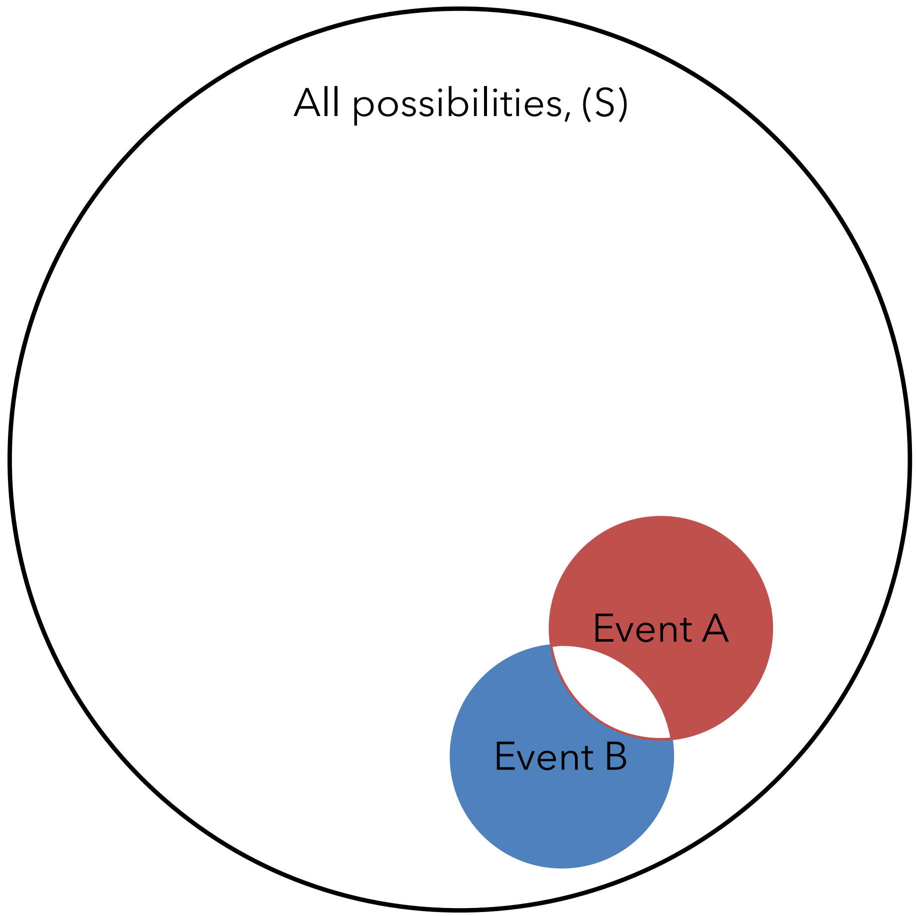


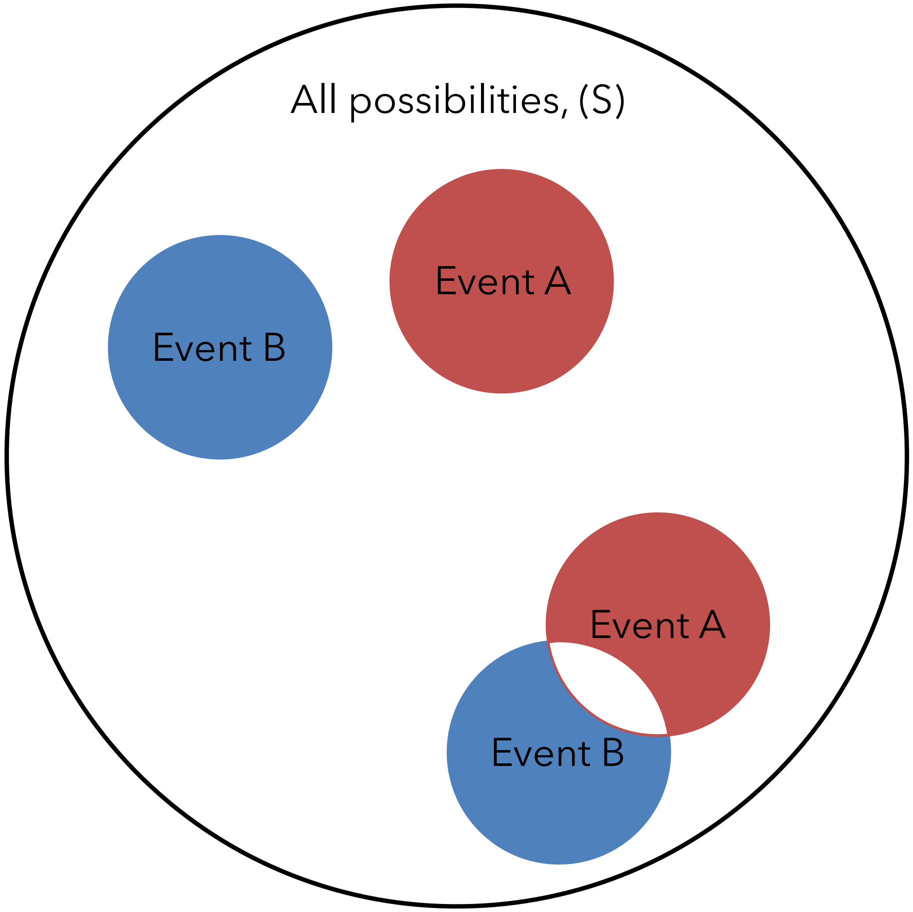
$$P(B) = \frac{\text{area A}}{\text{area S}}$$


All possibilities, (S)
Event A
Event B
P(overlap) = probability of A and B = P(A,B)

Probability of A <u>or</u> B

$P(A+B) = P(A) + P(B)$ - area of overlap

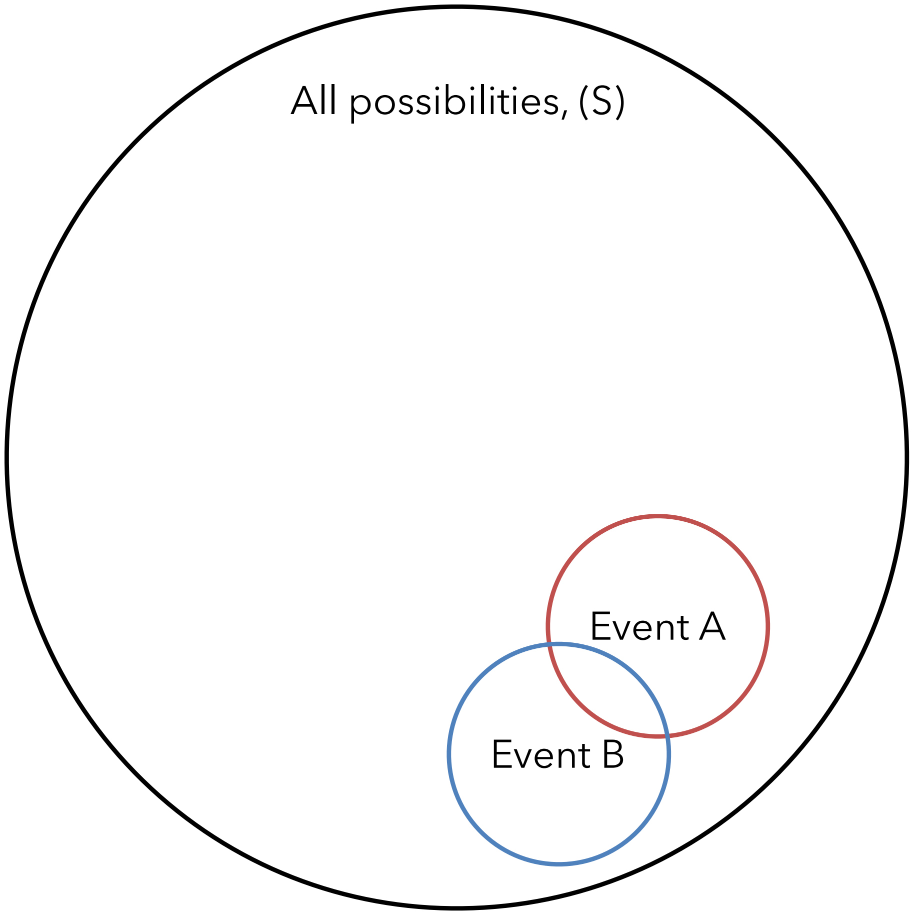




If A and B are mutually exclusive:

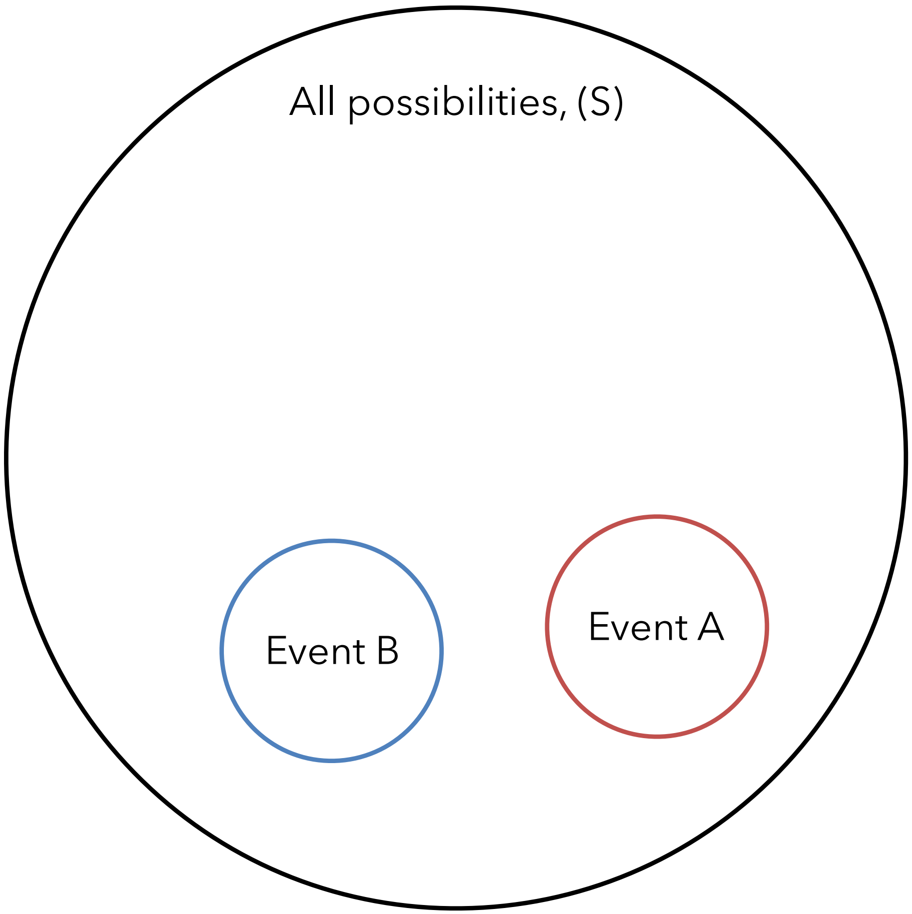
- P(A,B) = 0
- P(A + B) = P(A) + P(B)

P(B given A) = Probability of B and A out of probability of A



$$P(B \mid A) = \frac{P(A, B)}{P(A)}$$

Conditional probability of B



If A and B are independent, then one occurring has no effect on whether or not the other occurs.

P(B|A) = P(B)

Substitute this in to our conditional probability equation

$$P(B\,|\,A) = \frac{P(A,B)}{P(A)}$$

$$P(B) = \frac{P(A,B)}{P(A)}$$

$$P(B) \times P(A) = P(A,B)$$

Two MAJOR axioms of probability

1. Probability that one of two mutually exclusive events occurs:
    - P(A+B) = P(A) + P(B) (OR rule)

2. Probability that (both of) two independent events occur:
    - P(A,B) = P(A)*P(B) (AND rule)

Using these two axioms allow us to do fairly sophisticated statistical analysis:

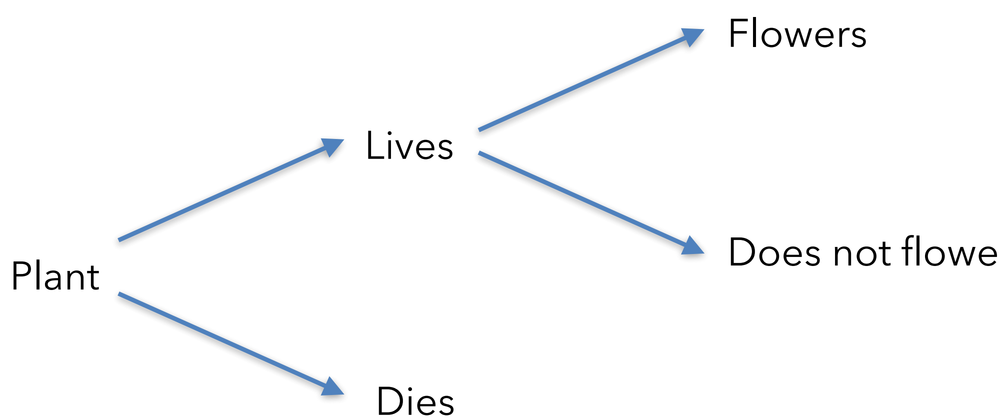
Example application: Demography of a perennial wildflower

How do we know if N is large enough to get a good estimate of our parameters?

Compare "competing models" with different values for *s* and *f*.

$\boxed{\begin{array}{l} s = 0.8 \\ f = 0.25 \end{array}}$ vs. $\boxed{\begin{array}{l} s = 0.6 \\ f = 0.4 \end{array}}$

Plant #	fate
1	Flowers
2	Vegetative
3	Vegetative
4	Vegetative
5	Dead



$s = 0.8$
$f = 0.25$

Plant #	fate	Events leading to fate?	Probability?
1	Flowers		
2	Vegetative		
3	Vegetative		
4	Vegetative		
5	Dead		

$s = 0.8$
$f = 0.25$

Plant #	fate	Events leading to fate?	Probability?
1	Flowers	survives AND flowers	
2	Vegetative	survives AND no flower	
3	Vegetative	survives AND no flower	
4	Vegetative	survives AND no flower	
5	Dead	dies	

s = 0.8
f = 0.25

Plant #	fate	Events leading to fate?	Probability?
1	Flowers	survives AND flowers	0.8 x 0.25 = 0.2
2	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
3	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
4	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
5	Dead	dies	1 - 0.8 = 0.2

Which axiom allows us to calculate the probability of the entire data set?
[assuming flowers live/die and flower independently]

$s = 0.8$
$f = 0.25$

Plant #	fate	Events leading to fate?	Probability?
1	Flowers	survives AND flowers	0.8 x 0.25 = 0.2
2	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
3	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
4	Vegetative	survives AND no flower	0.8 x (1 - 0.25) = 0.6
5	Dead	dies	1 - 0.8 = 0.2

Which axiom allows us to calculate the probability of the entire data set?
[assuming flowers live/die and flower independently]

0.2 x 0.6 x 0.6 x 0.6 x 0.2 = **0.00864**

Repeat with second model

s = 0.6
f = 0.4

Plant #	fate	Events leading to fate?	Probability?
1	Flowers	survives AND flowers	
2	Vegetative	survives AND no flower	
3	Vegetative	survives AND no flower	
4	Vegetative	survives AND no flower	
5	Dead	dies	

Probability of dataset:

$s = 0.6$
$f = 0.4$

Repeat with second model

Plant #	fate	Events leading to fate?	Probability?
1	Flowers	survives AND flowers	0.6 x 0.4 = 0.24
2	Vegetative	survives AND no flower	0.6 x (1 - 0.4) = 0.36
3	Vegetative	survives AND no flower	0.6 x (1 - 0.4) = 0.36
4	Vegetative	survives AND no flower	0.6 x (1 - 0.4) = 0.36
5	Dead	dies	1 - 0.6 = 0.4

Probability of dataset:

0.24 x 0.36 x 0.36 x 0.36 x 0.4 = **0.00448**

Likelihood

Using probabilities (of data), we can compute the **likelihoods** of different parameters given our dataset.

Likelihood (Edwards 1972) :

> "The likelihood L, of the hypothesis H, given data R, is proportional to $P(R|H)$, the constant of proportionality being arbitrary" [but specific to a particular data set]
>
> – (Note: "Hypothesis" = "model" = "set of parameters")

- Likelihood of first set of parameters (s = 0.8, f = 0.25) is proportional to 0.00864

$$L_1 = L(s = 0.8, f = 0.2 \,|\, R) = k_1 \times 0.00864$$

- Likelihood of 2nd set of parameters (s = 0.6, f = 0.4) is proportional to 0.00448

$$L_2 = L(s = 0.6, f = 0.4 \,|\, R) = k_2 \times 0.00448$$

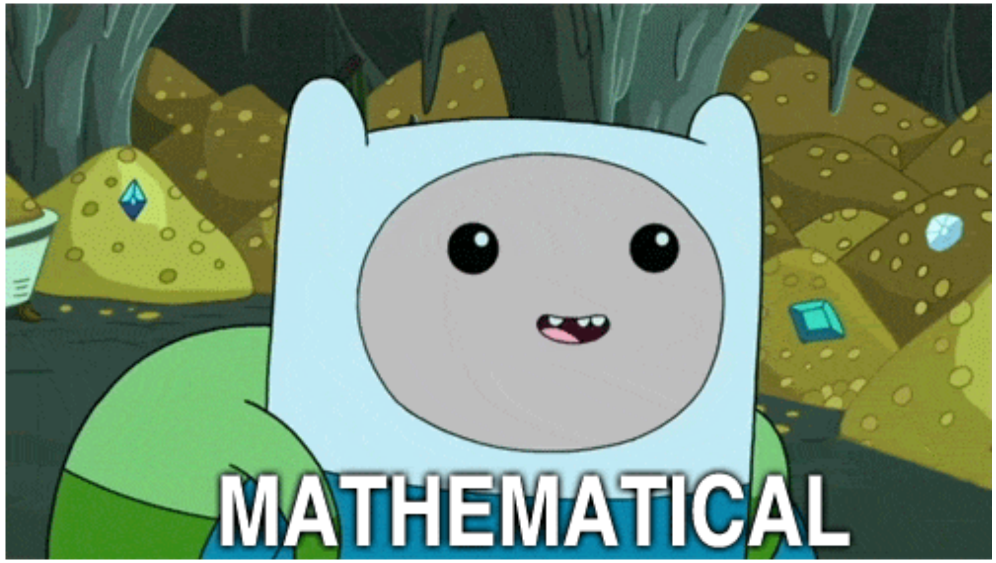
- Because our hypotheses are fit to EXACTLY the same dataset, k1 = k2 = k

Likelihood ratio

The **likelihood ratio** is our primary method of comparing models:

$$\frac{L_1}{L_2} = \frac{0.00864 \times k}{0.00448 \times k} = 1.93$$

First set of parameters is about twice as likely as our second, given this small data set.



We have gone from calculating the probability of our **data**, given fixed parameters, to making inferences about support for different possible values of our **parameters** (AKA models, hypotheses), given a dataset.

Repeat for a luckier ecologist who has 10 plants with the same distribution of fates: 2 flowering, 6 vegetative, 2 dead

- Likelihood of model #1:

$$L(s = 0.8, f = 0.25 \mid R_2) = k_3 \times 0.22 \times 0.66 \times 0.22 = 0.00007465$$

- Likelihood of model #2:

$$L(s = 0.6, f = 0.4 \mid R_2) = k_3 \times 0.242 \times 0.366 \times 0.42 = 0.00002006$$

- Likelihood ratio:

$$\frac{0.00007465 \times k_3}{0.00002006 \times k_3} = 3.72$$

- Same models, different likelihoods AND different likelihood ratio.

3 Features of Likelihood

1. Because it's the product of probabilities (and probabilities are all < 1), adding more data **reduces** the likelihood of any particular model.
    - Why the constant (k) is specific to a particular dataset
    - Also why we can't use likelihoods to compare models fit to different data

2. Adding more data makes the **likelihood ratio** between two models get larger because of the **AND** rule.
    - Larger sample size = greater ability to detect small differences
3. To avoid dealing with tiny numbers for typical (large) data sets, we often use log likelihoods:

$\ln(0.00007465) = -9.50$

$\ln(0.00002006) = -10.82$

Rules of working with exponents and logarithms:

1. Exponential transformation “undoes” (natural) log transformation, and vice versa:
   - ln(exp(A)) = A
   - exp(ln(A)) = A
2. Log transformation converts multiplication & division into addition & subtraction
   - ln(AB) = ln(A) + ln(B)
   - ln(A/B) = ln(A) - ln(B)
   - ln(A^C) = C x ln(A)
   - Therefore ln[P(A,B)] = ln[P(A)] + ln[P(B)]

Logarithms in R

```
A <- 5
log(exp(A))
#> [1] 5
exp(log(A))
#> [1] 5
B <- 6
log(A*B)
#> [1] 3.401197
log(A) + log(B)
#> [1] 3.401197
C <- 3
log(A^C)
#> [1] 4.828314
C * log(A)
#> [1] 4.828314
```

- R uses natural log as default.

Implications for the and axiom of probability

- Log-likelihoods of H1 and H2 [on board]
- log-transformed likelihood ratio become differences in likelihoods
- *log-likelihood ratio* can be back-transformed to calculate relative support in an absolute sense:

Ignoring flowering (for now), explore the likelihood of different values of survival, given the 5-plant data set (1 dies, 4 survive).

a. Calculate the likelihood of survival having each of the following values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.
b. Make a graph of the log-likelihood (y-axis) vs. value of survival (x-axis).
c. Repeat a&b for the 10-plant data set (2 die, 8 survive). How does the shape of the graph change?
d. What is the likelihood that survival is 0? Or 1? What happens to the log-likelihood at these values?

2. Thinking about probabilities…

a. What is P(B|A) for two mutually exclusive events? Is it possible for events to be independent and mutually exclusive?
b. What is P(A+B), the probability that A <u>or</u> B occurs, if they are independent events?