Likelihood

Eric Scott

2020-01-21

Alternative presentation: Bolker, pp 103-104

Probability: textbook explanation #2:

- big circle = all possibilities, S

- little circle #1 = event A

    - probability of event A, $P(A) = \frac{area\ of\ little\ circle}{area\ of\ big\ circle}$

- add 2nd little circle = event B

    - probability of event B, $P(B) = \frac{area\ of\ 2nd\ circle}{area\ of\ big\ circle}$

- A and B are **mutually exclusive**. Impossible for them to both occurr.

- Probability of A or B, $P(A + B) = P(A) + P(B)$ **OR rule**

- add 3nd little circle = event C (partly overlaps event A)

- A and C are **not** mutually exclusive.

    - probability of event C, $P(C) = \frac{area\ of\ 3rd\ circle}{area\ of\ big\ circle}$
    - probability of A and C, $P(A, C) = \frac{area\ of\ overalp}{area\ of\ big\ circle}$
    - probability of A or C, $P(A + C) = P(A) + P(C) - P(A, C)$. **need to correct for possibility of both ocurring**

Notation:

- $P(A)$: probability of event A
- $P(B)$: probability of event B
- $P(A + B)$: probability of event A or B
- $P(A, B)$: probability of events A and B
- $P(A|B)$: probability that event A occurs, conditioned on the fact that event B has occurred

Definitions

1. Mutually exclusive events: $P(A, B) = 0$ [i.e., it is impossible for both to occur]

- if A and B are mutually exclusive, $P(A, B) = 0$
- in this case, $P(A + B) = P(A) + P(B)$

2. Conditional probability, $P(C|A)$ [the probability that one event occurs, given that we know the other has]

- probability of C given A, $P(C|A) = \frac{area\ of\ overlap}{area\ of\ C}$

- this is called conditional probability of C
- $P(C|A) = P(A, C)/P(A)$

3. Independent events: whether one occurs does not affect whether the other does

- So, $P(C|A) = P(A, C)/P(A)$
- if A and C are independent then $P(C|A) = P(C)$
- So $P(C) = P(A, C)/P(A)$, and $P(A, C) = P(A) \times P(C)$
- this is the **AND rule**

Two MAJOR axioms of probability

1. Probability that one of two mutually exclusive events occurs:

- $P(A + B) = P(A) + P(B)$

2. Probability that (both of) two independent events occur:

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

Using these two axioms allow us to do fairly sophisticated statistical analysis: FOR EXAMPLE. . . Returning to our (very) small orchid population:

We have 5 plants, 4 live, and 1 of these flowers

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

Survival (s): $P(A) \approx 0.8$

Flowering, for plants that survive (f): $P(A) \approx 0.25$

QUESTION: Is N "large enough" to get a good estimate?

One way to answer this question is to compare different values of s and f, to find out what range of values are consistent with our data.

STEP 1

FOR A PARTICULAR PAIR OF VALUES OF s and f, we can calculate the probabilities of our getting our demographic data set using the axioms of probability:

FOR EXAMPLE: ASSUME estimates of $P(A)$ are correct: $s = 0.8$, $f = 0.25$

Plant#	fate	Events leading to fate
1	Flowering	Survive, flower
2	Vegetative	Survive, not flower

Plant#	fate	Events leading to fate
3	Vegetative	Survive, not flower
4	Vegetative	Survive, not flower
5	Dead	Don't survive

Which axiom allows us to calculate the probability of each fate? (and rule, axiom 2)

Plant#	fate	Events leading to fate	probability
1	Flowering	Survive, flower	$0.8 \times 0.25 = 0.2$
2	Vegetative	Survive, not flower	$0.8 \times (1 - 0.25) = 0.6$
3	Vegetative	Survive, not flower	$0.8 \times (1 - 0.25) = 0.6$
4	Vegetative	Survive, not flower	$0.8 \times (1 - 0.25) = 0.6$
5	Dead	Don't survive	$(1 - 0.8) = 0.2$

Which axiom allows us to calculate the probability of whole data set? (and rule, axiom 2)

$0.2 \times 0.6 \times 0.6 \times 0.6 \times 0.2 = 0.00864$

[assumption: "independent events", at least over your scope of inference]

Step 2

Compare this to a different set of parameters. ASSUME $s = 0.6$, $f = 0.4$

Plant#	fate	Events leading to fate	probability
1	Flowering	Survive, flower	$0.6 \times 0.4 = 0.24$
2	Vegetative	Survive, not flower	$0.6 \times (1 - 0.4) = 0.36$
3	Vegetative	Survive, not flower	$0.6 \times (1 - 0.4) = 0.36$
4	Vegetative	Survive, not flower	$0.6 \times (1 - 0.4) = 0.36$
5	Dead	Don't survive	$(1 - 0.6) = 0.4$

$0.24 \times 0.36 \times 0.36 \times 0.36 \times 0.4 \approx 0.00448$

SO: The probability of getting our data is about twice as high with our first set of parameters as with the second.

BUT: What does that tell us about the parameter estimates?

Using probabilities, we can compute the **likelihoods** of different parameters (and, later, alternative competing, models that differ in other ways), given our data set.

Edwards (1972) DEFINITION:

> "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]

Edwards' definition means that the likelihood of the 1st set of parameters is proportional to 0.00864

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

and the likelihood of the second set of parameters is proportional to 0.00448

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

If the two models, H_1 and H_2 are fit to EXACTLY the same set of data, $k_1 = k_2 = k$ and we can say that the first set of parameters is about twice as likely as the second set, given our small data set.

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

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

which means our first set of parameters, $H_1(s = 0.8, f = 0.2)$ is about twice as likely as our second, $H_2(s = 0.6, f = 0.4)$, given this small data set.

THIS IS MAGIC. 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, GIVEN A DATA SET.

FACT: We can use the ratio of two likelihoods to assess whether one set of parameters is "significantly" better than the other. This is known as a likelihood ratio test. We will do this calculation next week.

BUT FIRST, more about likelihoods

To understand the somewhat mysterious data-set-specific constants, compare these results to those of a luckier botanist who has a data set of 10 orchids, with exactly the same (proportional) distribution of fates: 2 floweing, 6 vegetative, 2 deaths

Likelihood of model #1: $L(s = 0.8, f = 0.25|R_2) = k_3 \times 0.2^2 \times 0.6^6 \times 0.2^2 = 0.00007465$

Likelihood of model #2: $L(s = 0.6, f = 0.4|R_2) = k_3 \times 0.24^2 \times 0.36^6 \times 0.4^2 = 0.00002006$

Likelihood ratio: $\frac{0.00007465 \times k_3}{0.00002006 k_3} = 3.72$

NOTE: 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. That is why the constant of proportionality is specific to a particular data set. This is also why we cannot use likelihoods to compare models fitted to different data sets

2. Adding more data makes the **likelihood ratio** between two models get larger (all else being equal; i.e., if the 2nd data set were identical to the first) because of the **AND** rule. In other words, larger sample sizes allow you to detect smaller differences among parameters or models.

3. To avoid dealing with tiny numbers for typical (large) data sets, we often use log likelihoods:

- $\ln(0.00007465) = -9.50$
- $\ln(0.00002006) = -10.82$

Recall some 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

- $\ln(AB) = \ln(A) + \ln(B)$
- $\ln(A/B) = \ln(A)^{\vee} \ln(B)$
- Therefore, $\ln(A^C) = C \times \ln(A)$

Implications for the and axiom of probability:

$$P(A,B) = P(A)P(B) = \ln[P(A,B)] = \ln[P(A)] + \ln[P(B)]$$

Log-likelihoods of H1 and H2:

H1:

$$ln(k_3 \times 0.2^2 \times 0.6^6 \times 0.2^2) =$$
$$\ln(k_3) + \ln(0.2^2) + \ln(0.6^6) + \ln(0.2^2) =$$
$$\ln(k^3) + 2\ln(0.2) + 6\ln(0.6) + 2\ln(0.2) =$$
$$-9.503 + \ln(k_3)$$

H2:

$$\ln(k_3 \times 0.24^2 \times 0.36^6 \times 0.4^2) =$$
$$\ln(k_3) + 2\ln(0.24) + 6\ln(0.36) + 2\ln(0.4) =$$
$$-10.817 + \ln(k_3)$$

log-transformed likelihood ratio become differences in likelihoods:

$$\ln\left(\frac{0.00007465}{0.00002006}\right) = \ln(0.00007465) - \ln(0.00002006) = -9.50 - (-10.82) = 1.32$$

This is the *log-likelihood ratio*. Can be backtransformed to calcualte relative support in an absolute sense: -$\exp(1.32) = 3.72$

AS TIME PERMITS:

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

   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.
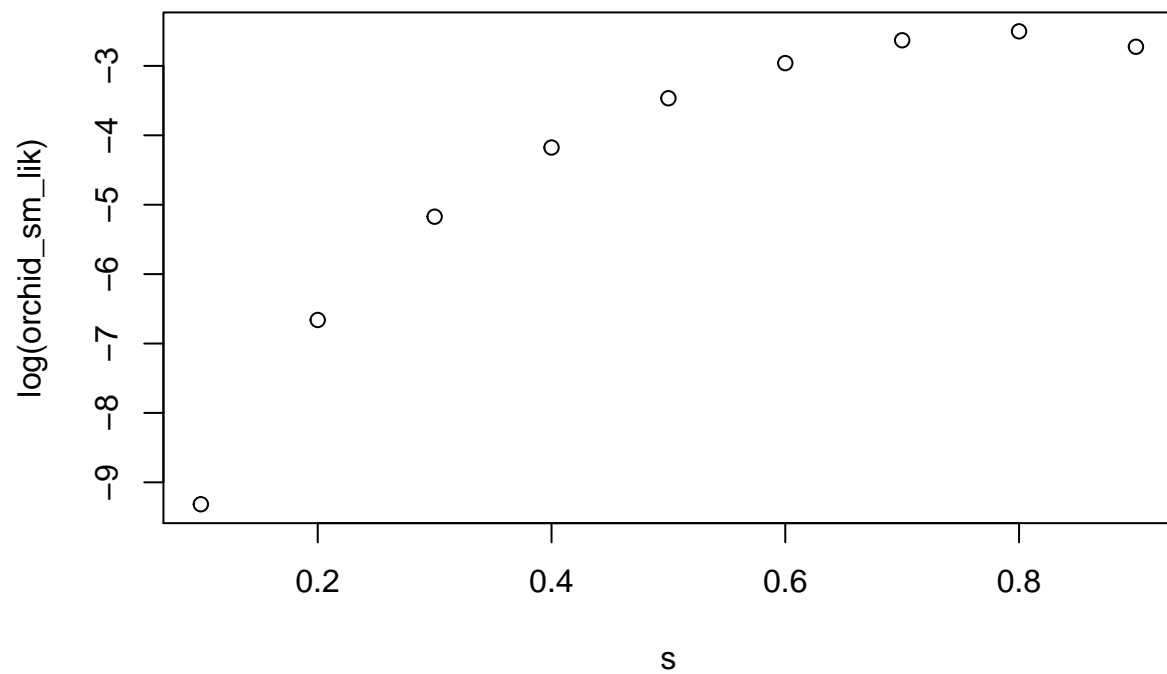
```
s <- seq(0.1, 0.9, 0.1)
s
```

```
## [1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
```

```
orchid_sm_lik <- s^4 * (1-s) #vectorized over all values of s
orchid_sm_lik
```

```
## [1] 0.00009 0.00128 0.00567 0.01536 0.03125 0.05184 0.07203 0.08192 0.06561
```

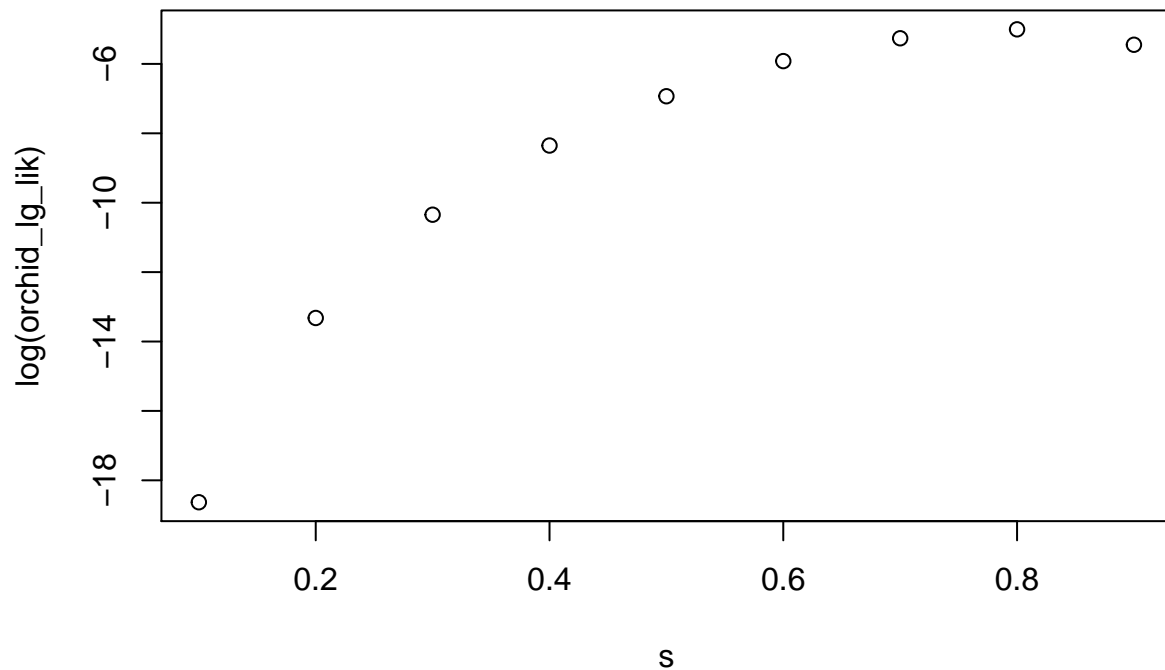   b. Make a graph of the log-likelihood (y-axis) vs. value of survival (x-axis).

```
#base R plotting
plot(s, log(orchid_sm_lik))
```

5

c. Repeat a&b for the 10-plant data set. How does the shape of the graph change?

```
orchid_lg_lik <- s^8 * (1-s)^2
```

```
#base R plotting
plot(s, log(orchid_lg_lik))
```

```
#using ggplot2 to put both lines on the same plot
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------
```

```
## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  2.1.3      v dplyr   0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------- tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# make a 'tidy' dataframe/tibble for plotting
df <-
  tibble(
    dataset = c(rep("small", 9), rep("large", 9)), #column indicating which dataset
    s = c(s,s), #x-axis values
    likelihood = c(orchid_sm_lik, orchid_lg_lik),
    log_lik = log(likelihood)
  ) #you can do math while creating a tibble

head(df)
```

```
## # A tibble: 6 x 4
##   dataset      s likelihood log_lik
##   <chr>    <dbl>      <dbl>   <dbl>
## 1 small      0.1    0.00009   -9.32
## 2 small      0.2    0.00128   -6.66
## 3 small      0.3    0.00567   -5.17
## 4 small      0.4    0.0154    -4.18
## 5 small      0.5    0.0312    -3.47
## 6 small      0.6    0.0518    -2.96
```

```
ggplot(df, aes(x = s, y = log_lik, color = dataset)) +
  geom_point() +
  geom_line() #add connecting lines
```
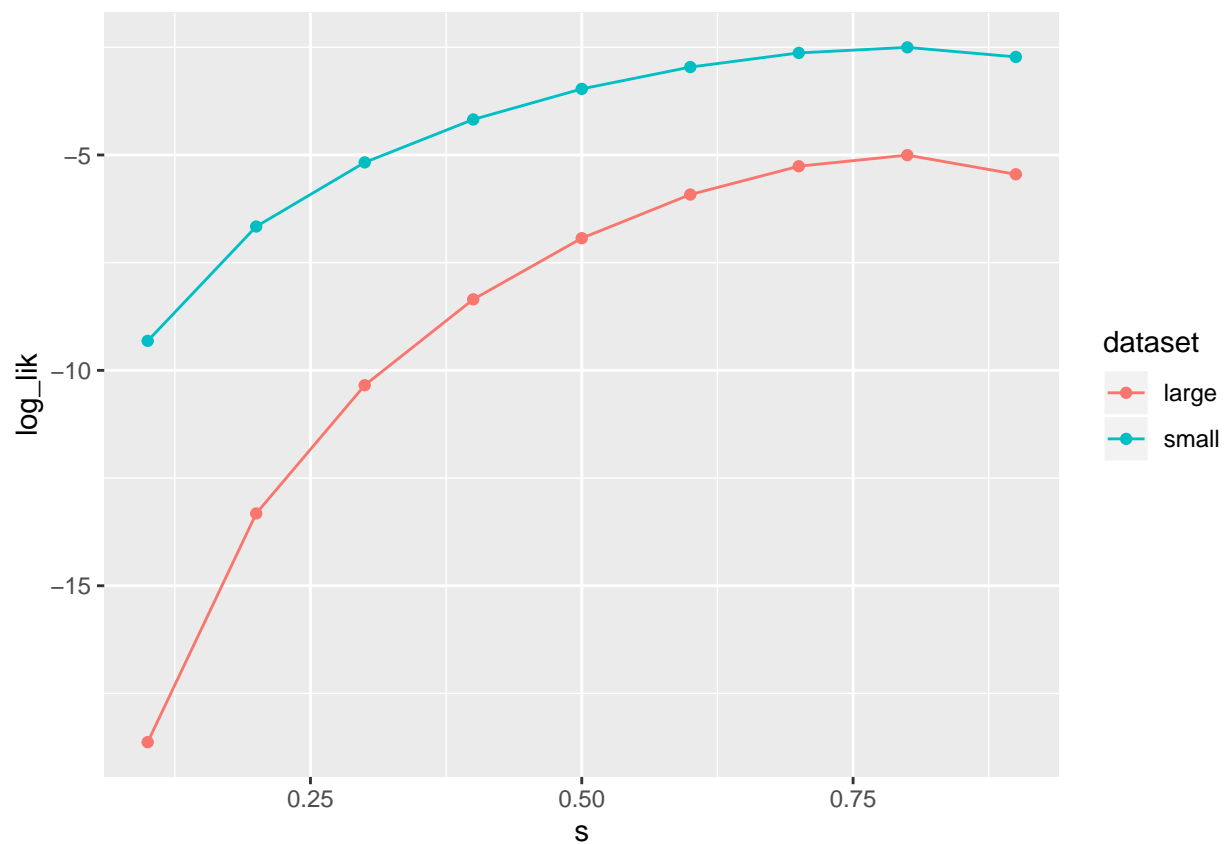

d. What is the likelihood that survival is 0? Or 1? What happens to the log-likelihood at these values?

```
0^8 * (1-0)^2
```

```
## [1] 0
```

```
log(0)
```

```
## [1] -Inf
```

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?

$P(B|A) = 0$ because there is no overlap. Not possible for events to be independent and mutually exclusive. E.g. heads & tails aren't independent.

b. What is the probability that A **or** B occurs if they are **not** mutually exclusive?

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