Statistical Inference: One Parameter

Bob O'Hara

bob.ohara@ntnu.no

Administration Matters

- Reference Group
- Blackboard

The Simulation Function, from yesterday

(changed to avoid a possible error that coud mess things up)

```
simGlobe <- function(probability=0.5, NTrials=10) {
  sim <- rbinom(NTrials, 1, probability)
# this next line changes so the function will return
# zero counts, rather than leave them off
  Res <- factor(c("Sea", "Land")[1+sim],
                levels = c("Sea", "Land"))
  return(table(Res))
}

(reps <- replicate(3, simGlobe(probability = 0.1,
                               NTrials = 10)))
```

```
##
## Res    [,1] [,2] [,3]
##   Sea      9    8    9
##   Land     1    2    1
```

What to do

Use the `simGlobe()` function to simulate getting more data

- use the same `NTrials` as we used
- decide on a “good” value of `prob`

Compare your simulations with your guesses, and with what we actually got

Try the `hist()` function to plot the data:

- you will need to plot only the first rowm so use `hist(reps["Land",])` or `hist(reps[2,])`

Taking Stock

So far we have seen that there is a range of possible results we could get from the same parameters

We want to estimate the parameters, so will different parameters give us different similar results?

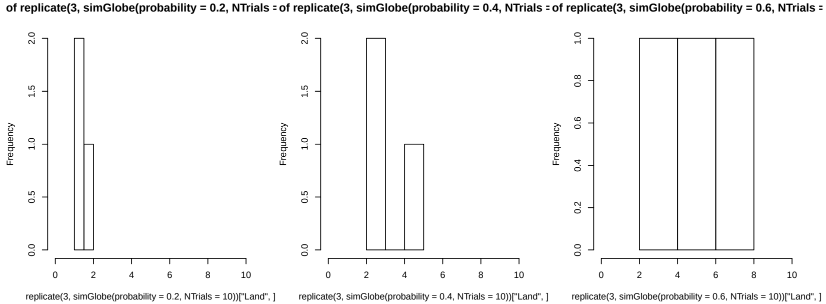
Different Probabilities

Now we have some idea about the variation in the results we could get from one parameter, what if there is another parameter?

Different Probabilities

Simulate data with a `probability` 0.2 higher, then 0.2 lower

```
par(mfrow=c(1,3)) # This gives 3 plots in one row
hist(replicate(3, simGlobe(probability = 0.2,
              NTrials = 10))["Land",], xlim=c(0,10))
hist(replicate(3, simGlobe(probability = 0.4,
              NTrials = 10))["Land",], xlim=c(0,10))
hist(replicate(3, simGlobe(probability = 0.6,
              NTrials = 10))["Land",], xlim=c(0,10))
```



The Inference Problem

We have seen that even with one probability, we can get a range of observations. And we can get the same observation with a range of probabilities

So how do we find a good probability?

How do we know what are reasonable probabilities?

The Inference Problem

We have seen that even with one probability, we can get a range of observations. And we can get the same observation with a range of probabilities

So how do we find a good probability?

How do we know what are reasonable probabilities?

One way: find the ‘probability‘ that makes the data most likely

What we need to do

Find out how to calculate the probability of the data

Find the maximum value

- this might need some maths, or we can do it numerically, or by simulation

Some Maths

If we have 1 trial, the probability of Land is `p`

If we have 2 trials, we could have Land-Land, Land-Sea, Sea-Land, Sea-Sea

So

$$
\begin{aligned}
Pr(2Land) &= p^2 \\
Pr(1Land) &= 2p(1-p) \\
Pr(1Land) &= (1-p)^2
\end{aligned}
$$

The Binomial Distribution

If we have N trials, and observe r "successes" then the probability of this is

$$Pr(n = r|N, p) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$$

Which has 2 parts. The important part is $p^r(1-p)^{N-r}$ which is

$$p^{successs}(1-p)^{failures}$$

The other part

$$\frac{N!}{r!(N-r)!}$$

this counts the number of combinations of r successes and $N-r$ failures

e.g. if $N=3$ and $r=1$ we have

- success - failure - failure
- failure - success - failure
- failure - failure - success

So $\frac{3!}{1!(3-1)!}=3$

Likelihood

If we know p (the probability of `Land`), we can calculate the probability of obtaining the data, given the parameter

- this is called the *likelihood*

But we don't know p: this is what we want to estimate

log-Likelihood

It is ofter easier if we use the log-likelihood,
$l(p|n = r, N) = \log(Pr(n = r|N, p))$

$$(l(p|n = r, N) = \log\left(\frac{N!}{r!(N-r)!}\right) + r\log p + (N-r)\log(1-p)$$

This is a function of p, so if we ignore constants we have

$$l(p|n = r, N) = r\log p + (N-r)\log(1-p) + C$$

Using the Likelihood

We can calculate the likelihood for different values of p

```
NLand <- 4; NSea <- 6; N <- NLand + NSea
dbinom(NLand, N, 0.4) # calculate the likelihood
```

```
## [1] 0.2508227
```

```
dbinom(NLand, N, 0.4, log=TRUE) # the log-likelihood
```

```
## [1] -1.383009
```

We can also calculate several values:

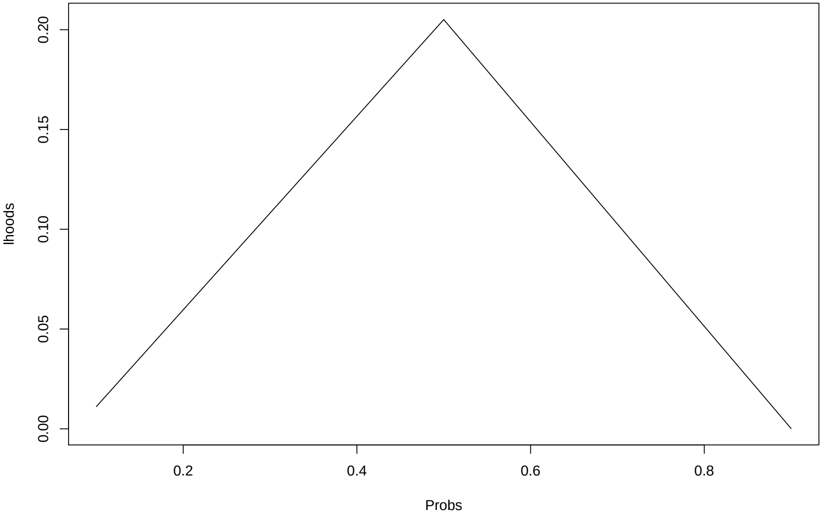
```
# seq() creates a sequence of numbers
Probs <- seq(from = 0.1, to = 0.9, length.out = 3)
(lhoods <- dbinom(NLand, size = N, prob = Probs))
```

```
## [1] 0.011160261 0.205078125 0.000137781
```

Your task: Finding a Good Likelihood

Calculate the likelihood for the data for different values of p

```
plot(Probs, lhoods, type="l")
```



The Philosophy

The likelihood is a data generating mechanism: it is a statistical model

We assume that the data are random, and the parameters (and model) are fixed

We want to find the parameters which are most likely to give rise to the data

- we maximise the likelihood

Maximising the likelihood

Poking around and trying values is not the best way to find the maximum.

Alternatives are:

- analytic: do the maths (works for this problem)
- numerical: use an algorithm thatfinds the maximum
- simulation: simulate the likelihood & find the best value

The maximum of the likelihood is the same as the maximum of the log-likelihood, so we usually work on the log scale

Maximising the log likelihood for the binomial

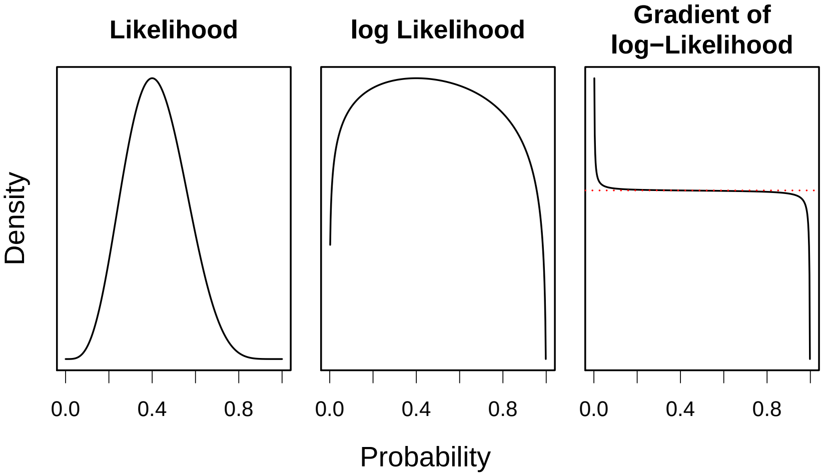
We can do this analytically. We want to get find an eqiation for the slope, then set this to zero. The likeilhood is

$$l(p|n) = r\log(p) + (N-r)\log(1-p) + C$$

and after we differnetiate (to get the slope) we have

$$\frac{dl(p|n)}{dp} = \frac{r}{p} - \frac{N-r}{1-p}$$

In Figures


Likelihood
log Likelihood
Gradient of
log−Likelihood
Density
0.0
0.4
0.8
0.0
0.4
0.8
0.0
0.4
0.8
Probability

Maximising

Set the gradient to 0:

$$0 = \frac{r}{p} - \frac{N-r}{1-p}$$

So

$$\frac{p}{1-p} = \frac{r}{N-r}$$

i.e. the *odds* of success are equal to the ratio of successes to failure.

We can re-arrange to get

$$\hat{p} = \frac{r}{N}$$

So. . .

We have maximised the likelihood to get an estimator of p

$$\hat{p} = \frac{r}{N}$$

In more complicated problems we do the same thing, but sometimes the maximisation is done numerically (or even through simulation)

But we always use the log-likelihood & ignore the normalising constants

Terminology

We call p the *estimand*: this is what we want an estimator of

We will call the *estimator* $\hat{p}$

Because we will get $\hat{p}$ by maximising the likelihood, we call it the *maximum likelihood estimator* (MLE).

What happens if we take another sample?

e.g. the second time we sampled the earth, we had 6 `Land`and 4 `Sea`

What happens if we take another sample?

Each sample gives us a different $\hat{p}$

p is fixed, and the data are random, so $\hat{p}$ is a property of the data

We can sample repeatedly many times, and each time get a different $\hat{p}$

The likelihood is the distribution of $\hat{p}$

More samples: your task

Look at the distribution of possible estimates of p.

- Assume the "true" value is 0.4.
- Simulate the data for 10 trials
- Calculate the maximum likelihood estimator
    - if you an do this all in the same function, it is easier

Repeat this many times, and plot a histogram of the estimates

Summary

We have seen that data vary

We can estimate the “best” parameters from the data

Different data will give different “best” parameters, even if the process is the same

Next Week

Adding confidence!