# Statistical Inference

Bob O'Hara

bob.ohara@ntnu.no

# Administration Matters

- Reference Group
- Blackboard

# Warning

This should be the most mathematical part of the course.

The purpose is to provide some of the theory that underlies statistical modelling

# All models are wrong. . .

Statistical models assume that the data are produced by a random process

- there is some probability of the data
- we take a sample from a population

## . . . but some are useful

Our model of the data is the probability of sampling it given the population

Different samples -> different data from the same population

We want to learn about the population from the data

# The general idea

Our data is $Y$. We sample it randomly from a population, according to some probability $P(Y|X, \theta)$

- $X$ are (fixed) covariates
- $\theta$ are parameters that describe the population

$P(Y|X, \theta)$ is a model that describes how we sample

We want an estimator of $\theta$

# An Example

What proportion of the earth is land?

If we have a globe, how can we estimate what proportion in land and what proportion sea?

(plant cover is a real example of this problem)

# Sampling The Earth

Toss the globe around When you catch it. put your finger on a point, and say whether it lands on the land or sea Then toss it to someone else

We will record the number of times we get Land or Sea, and use this as an estimate of the proportion of the glode that is land

# The Model

Each observation is a sample from the real world

- "Bernoulli trial"

We observe $N$ trials, of which $n$ are land, and $(N - n)$ are water

$n$ follows a binomial distribution, with an unknown $p$ (the population-level mean)

# The Model

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

We want an estimator of $p$, which is the “true” population-level mean

# Terminology

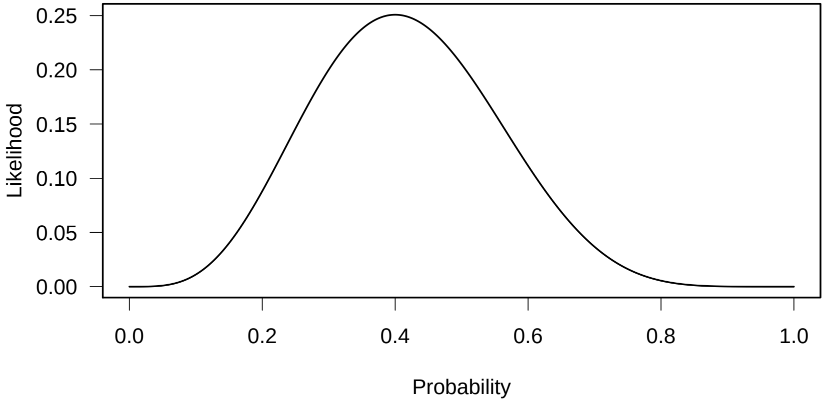
We call $p$ the *estimand*: this is what we want an estimator of

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).

# Likelihood

$Pr(n = r|N, p)$ as a function of $p$: call it the *likelihood*: $L(p|N, r)$




Find the value of $p$ that maximises $Pr(n = r|N, p)$

# 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: two tricks

The maximum of the likelihood is at the same value of $p$ as the maximum of the log-likelihood

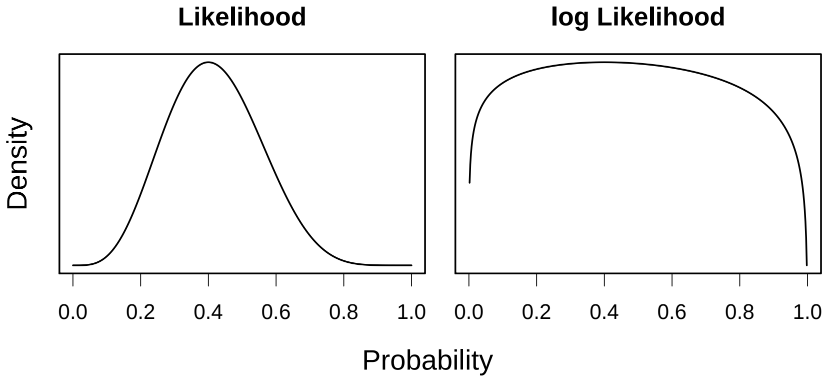
i.e. we work with $\log(L(p|N,r)) = l(p|N,r)$

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

Becuse this is a function of $p$, not $N$ or $r$, several terms are constants:

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

# The log-likelihood



# Maximising the log likelihood

Differentiate with respect to p

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

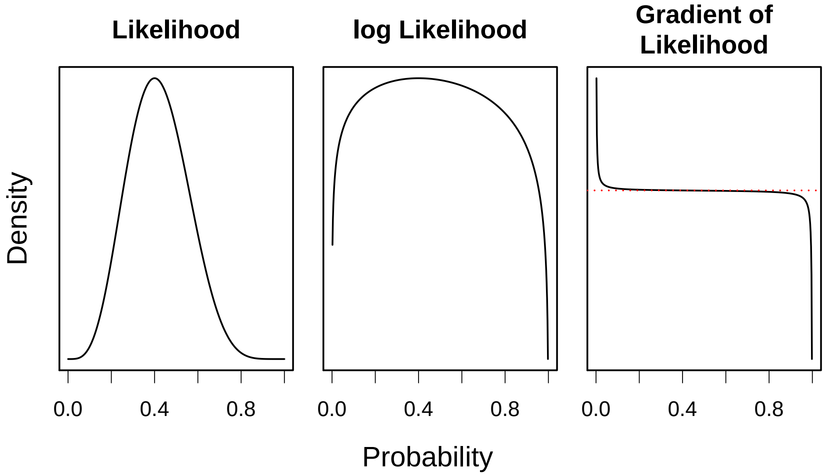
using the chain rule for the second term ($u = 1-p$):

$$\frac{d\log(1-p)}{dp} = \frac{d\log u}{du}\frac{du}{dp} = \frac{1}{u}(-1) = -\frac{1}{1-p}$$

we get

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

# In Figures


Likelihood
log Likelihood
Gradient of Likelihood
Density
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-write this as $\frac{1-p}{p} = \frac{N-r}{r}$, and re-arrange to get xtable

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

# So...

We have (analytically) 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

# What happens if we take another sample?

??

# What happens if we take another sample?

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

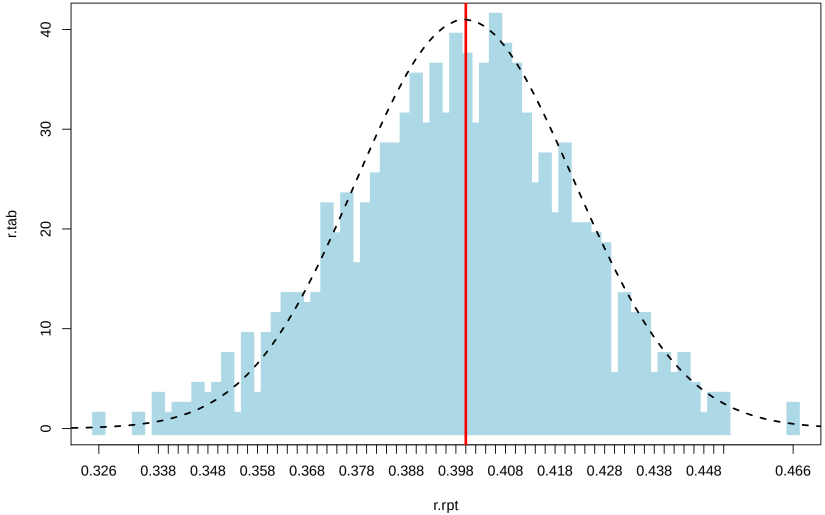
$p$ is fixed, and the data are random, so $\hat{p}$ is a property of the data

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

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

# Samples

$N = 500, \; p = 0.4$



# Uncertainty

Because different samples give different estimates, we want to quantify this - suggest plausible values

What summaries could we use?

# Standard Errors

Standard Deviations of Statistics: $s$

Binomial variance of $n$: $Var(n|N, p) = Np(1-p)$

Our statistic: $n/N$

$Var(n/N) = 1/N^2 Var(n) = p(1-p)/N$

Standard error:

$s = \sqrt{p(1-p)/N}$

# Confidence Intervals

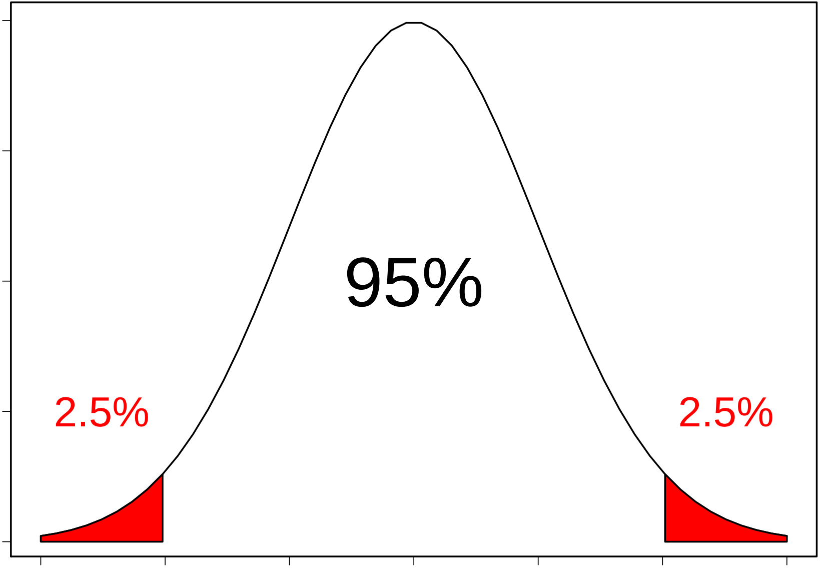
Standard errors are useful, but what values of $p$ are likely?

Confidence intervals give the range of values in which the statistic is likely

- usually we use 95%

A 95% confidence interval is one that has a 95% probability of containing the estimate of $p$ if the estimate $\hat{p}$ is the true value.

# Constructing a Confidence interval

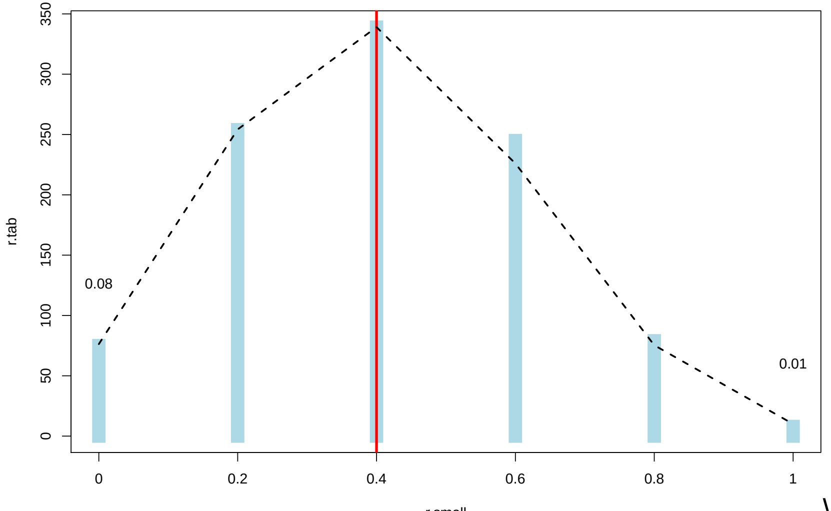


# Confidence Interval for a Binomial

More difficult, because data are discrete

e.g. N = 5

$N = 5$, $p = 0.4$




With

# Approximations

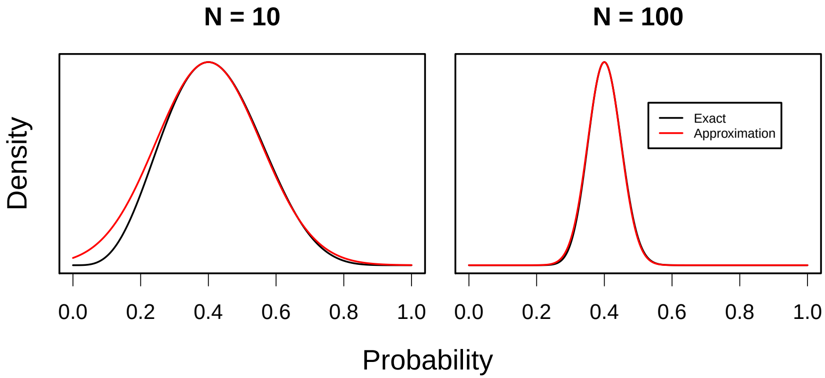
In general, when we have a lot of data the likelihood looks like a normal distribution

So we can use use that to make an approximation

# Approximating the Likelihood

The likelihood is approximately Normal with mean equal to the MLE, and standard deviation equal to the standard error

$$l(p|N,n) \sim N(\hat{p}, \sqrt{\hat{p}(1-\hat{p})/N})$$



# Approximating the Confidence Interval

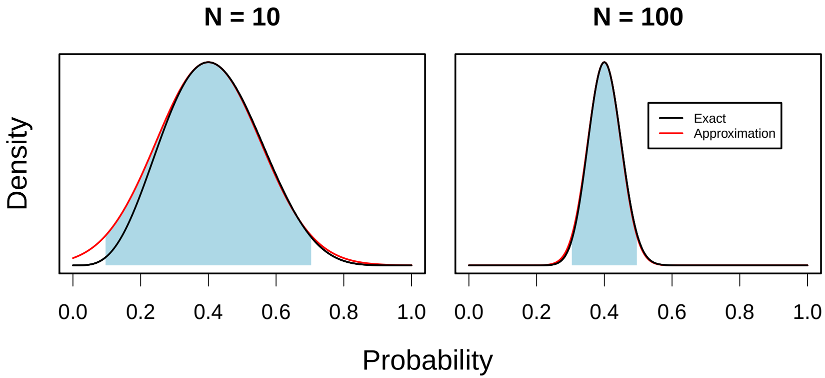
We can use athis approximation, so the interval is

$$\left(\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})/N}, \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})/N}\right)$$

So, for $N = 10$, $n = 4$ the interval is (0.4 - 0.15, 0.4 + 0.15), i.e. (0.25, 0.55).

With $N = 100$, $n = 40$ the interval is smaller: (0.35, 0.45).

# Approximating the Confidence Interval

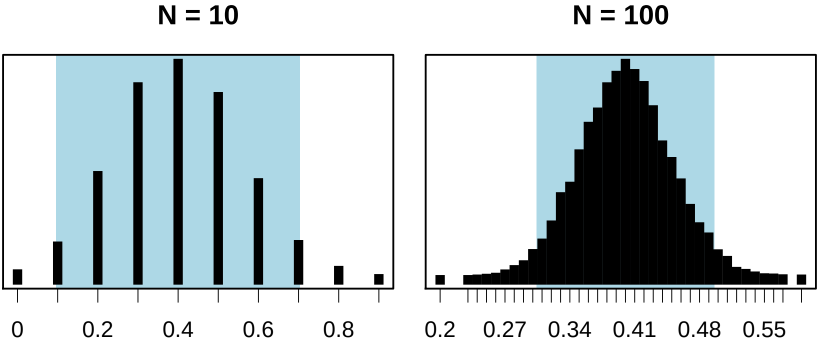


# How well are we doing?

The confidence interval is an interval where, if we repeat the same experiment many thimes, we have a 95% probability of getting an estimate inside the interval.

We can check this!

# How well are we doing?




For $N = 10$ we have a proportion 0.9811 within the CI,

For $N = 100$ the proportion is 0.9481 within the CI,

# Summary

We want to estimate parameters - summaries statistics

Estiamtors are functions of the data: the estimand is fixed

Likelihood gives the probability of the data (and thus estimators)

We can summarise the uncertainty with standard errors and confidence intervals

Our example was the binomial distribution

# Next Week

Regression: fitting models with straight lines

- multiple parameters
- normal likelihood
- model fit