Day 3
Lecture 3: Assessing model uncertainty and calibration

Short course on modelling infectious disease dynamics in R

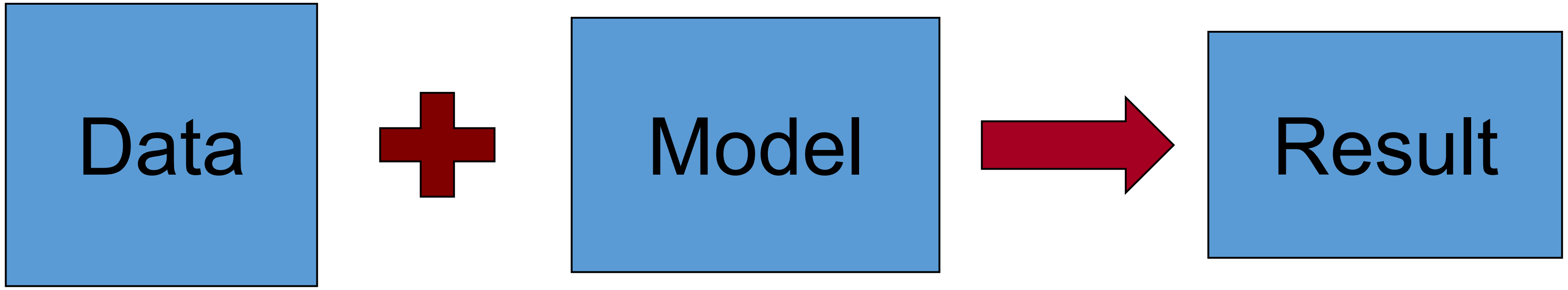
Ankara, Türkiye, September 2025

Dr Juan F Vesga

Aims of the session

- Understand the sources of uncertainty
- Understand the importance of communicating uncertainty
- Learn main methods for addressing model uncertainty

The nature of the problem

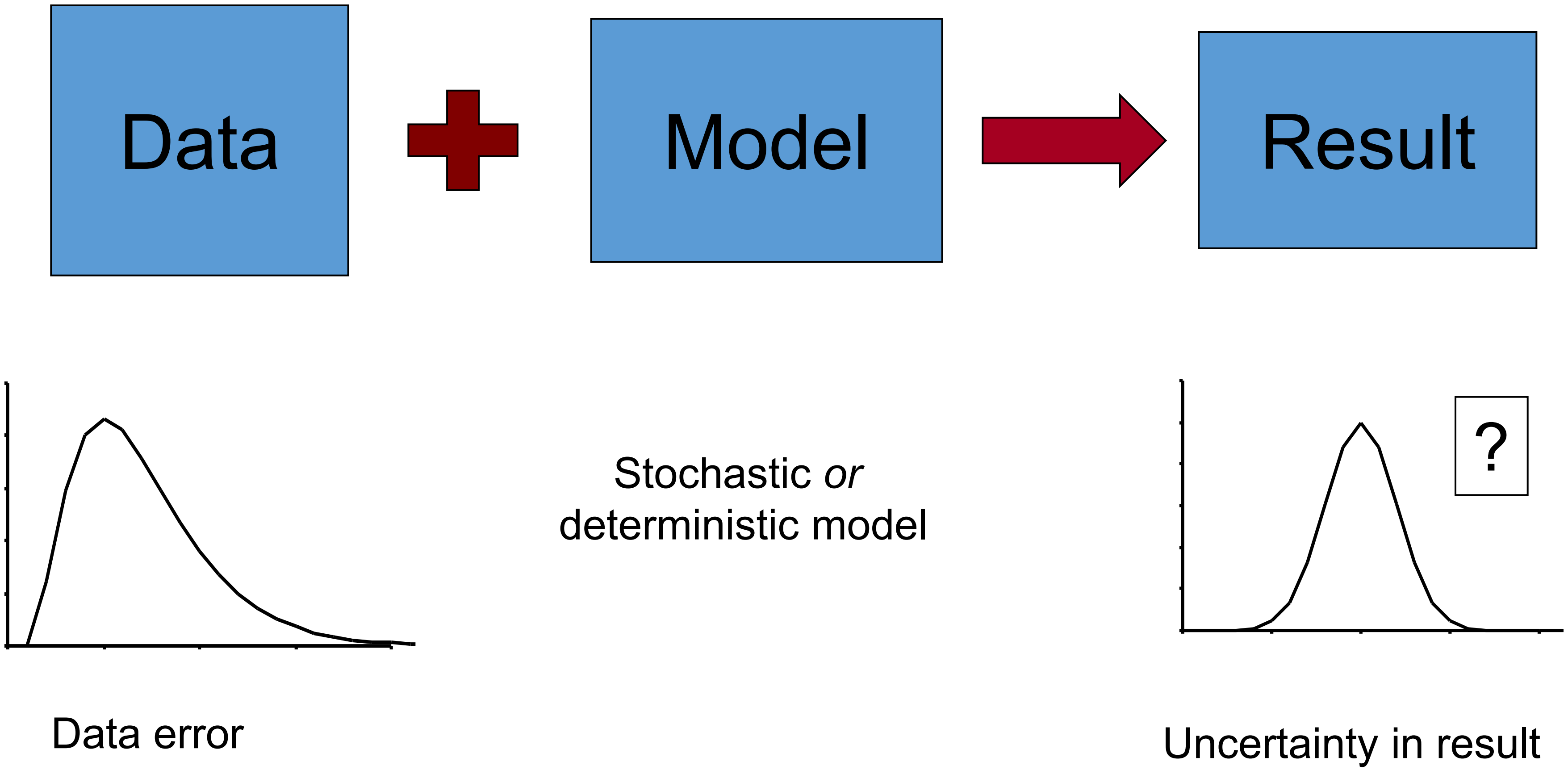


Raw data: incidence, serological data, etc. or parameter values from other studies.

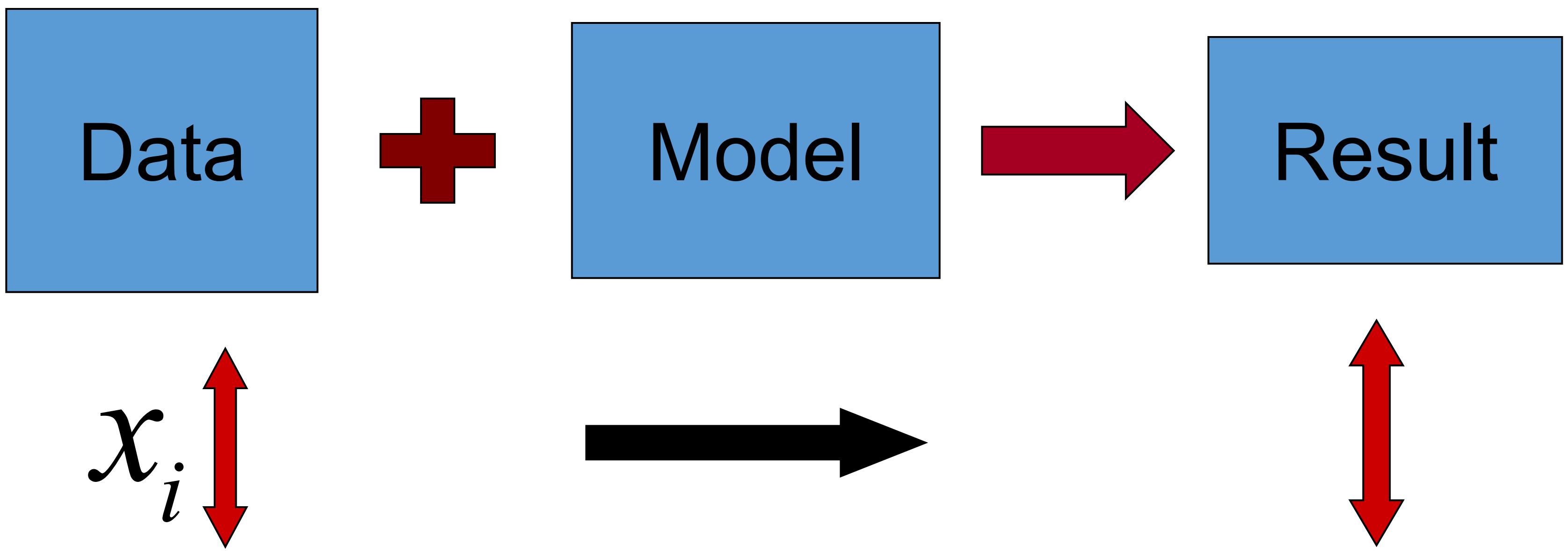
Deterministic or stochastic

Property or statistic of the system (Contact rate, R_0, etc.)

Uncertainty



Sensitivity



For a deterministic model, two types:

•Local sensitivity: change in result for given change in parameter.

•Global sensitivity: how uncertainty in result is related to that of any input parameter.

Model for illustration of parameter sensitivity

Outbreak of an SIR type in finite population, e.g. influenza, measles, Rubella, Hepatitis A.

Scenario: introduce I_0 infectives into a population of N-I_0 susceptibles.

$$S \to I \quad \text{: rate } \lambda = \frac{\beta I}{N} \text{ /day,}$$

$$I \to R \quad \text{: rate } \mu \text{ /day}$$

$$R_0 = \beta / \mu$$

Model for illustration of parameter sensitivity

The model can be described by the following system of equations:

$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$
$I_0 = 100$

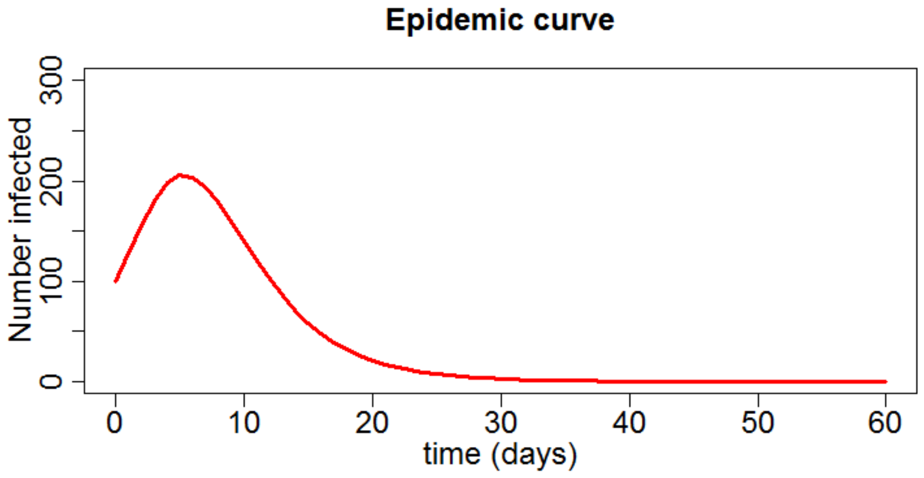
Model for illustration of parameter sensitivity

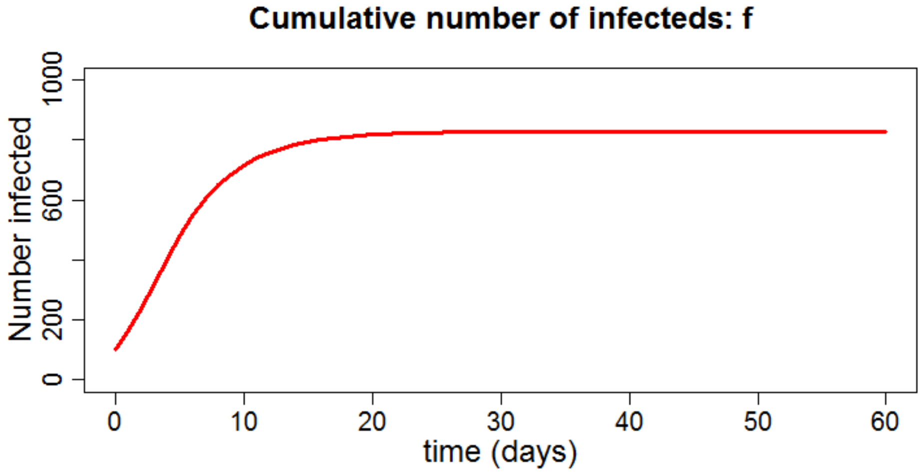
$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$
$I_0 = 100$





Model for illustration of parameter sensitivity

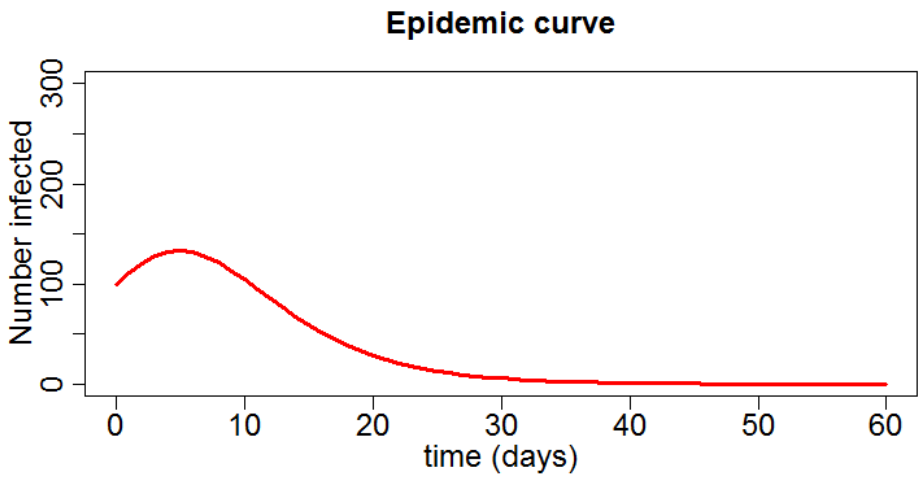
$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

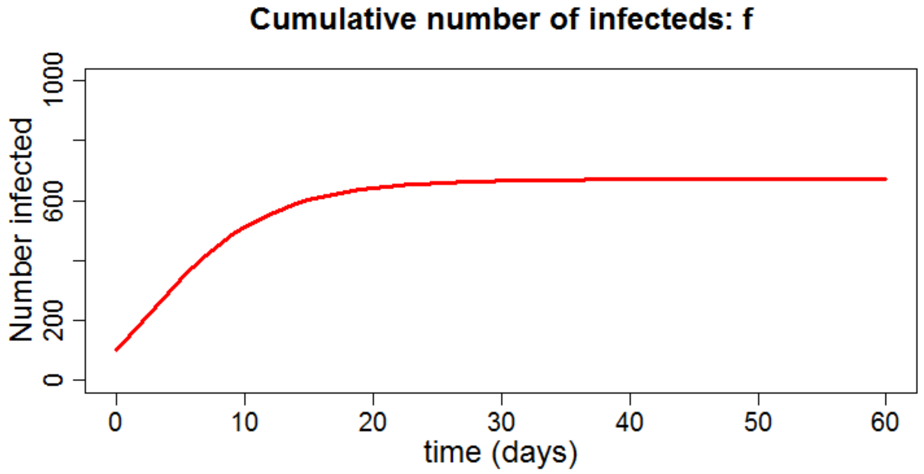
$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

R_0 **= 1.5**

$I_0 = 100$





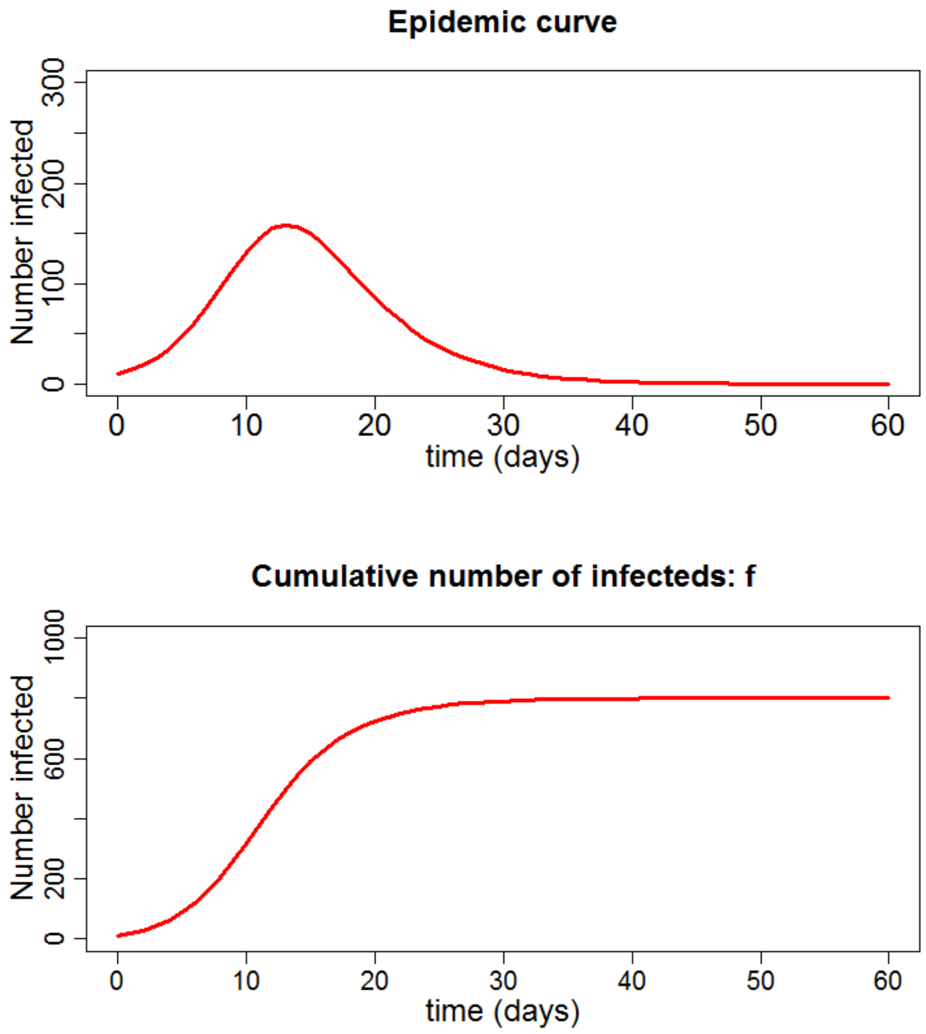
Model for illustration of parameter sensitivity

$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$
$\boldsymbol{I_0 = 10}$



Model for illustration of parameter sensitivity

We are interested in the number of individuals infected over the course of the epidemic, a.k.a. the final size, f, or attack rate, as it depends on I_0 and R_0.

For a deterministic model,

$$1 - f = (1 - I_0 / N) \exp(-R_0 f)$$

Model for illustration of parameter sensitivity

We are interested in the number of individuals infected over the course of the epidemic, a.k.a. the final size, f, or attack rate, as it depends on I_0 and R_0.
For a deterministic model,

$$1 - f = (1 - I_0 / N)\exp(-R_0 f)$$

Sensitivity to variation in *R_0*:

I_0	R_0	f
100	1.25	544
100	1.5	671
100	**2**	**828**
100	3	948
100	4	982

Model for illustration of parameter sensitivity

We are interested in the number of individuals infected over the course of the epidemic, a.k.a. the final size, f, or attack rate, as it depends on I_0 and R_0.
For a deterministic model,

$$1 - f = (1 - I_0 / N)\exp(-R_0 f)$$

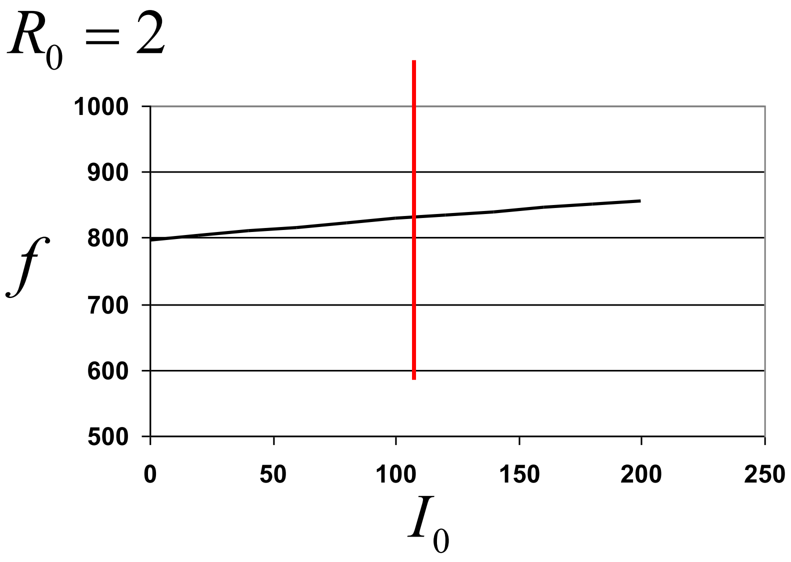
Sensitivity to variation in I_0:

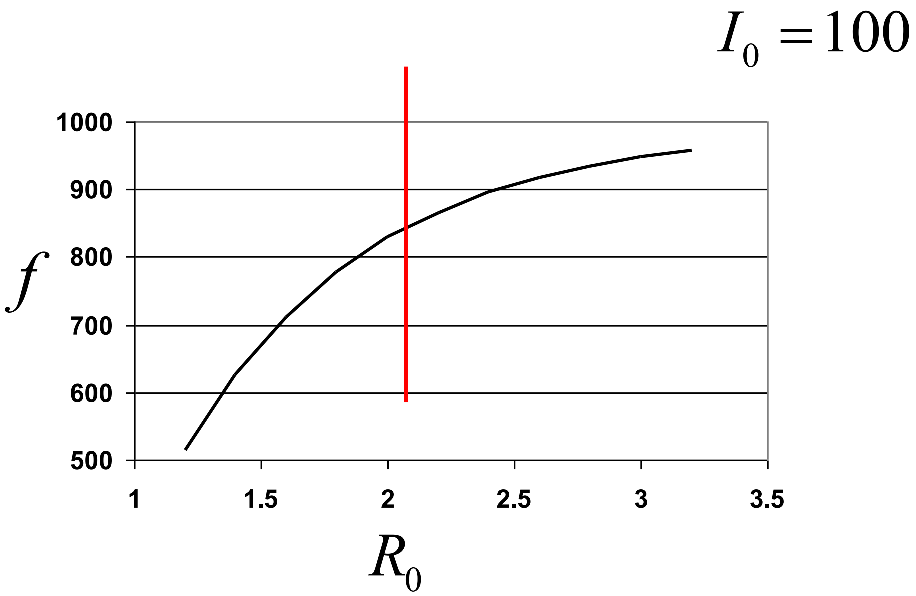
I_0	R_0	f
1	2	797
10	2	800
100	**2**	**828**
200	2	855
300	2	879

Local uncertainty

- Local sensitivity looks at the change in the outcome of interest as params are varied one at a time.
- In our model, final size depends on reproduction number, R_0, and initial infectives, I_0.

We can look at the sensitivity around R_0=2, I_0=100, say.





Doesn’t tell us how different parameters *combine* to affect the result.

Latin Hypercube Sampling

•If evaluation expensive, don't want to go through all values.

•LHC is a method of picking a subset of points that 'span' the region.

From previous example:

Increasing size

(I_{01}) → (I_{01}, R_{07})

(I_{02}) → (I_{02}, R_{08})

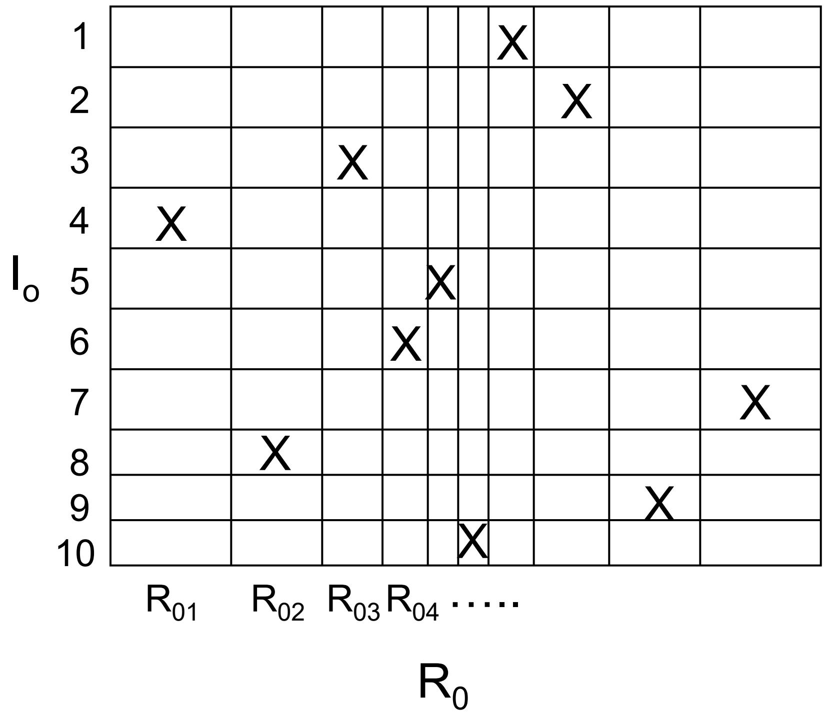
(I_{03}) → (I_{03}, R_{03})

(I_{04}) → (I_{04}, R_{01})

(I_{05}) → (I_{05}, R_{05})

$\ldots$ → $\ldots$

choose an R_0 for each I_0



10X10 cube: mean = 780, s.d. = 86

•Each parameter value appears once and only once.
•Can repeat the process to get further sets.

Monte Carlo Markov Chain

•MCMC is a way of generating a sequence of parameter sets which appear to have come from distribution.

•Each new set generated from previous one, hence 'chain', by a rule.

$$(x_1, x_2)_n \Rightarrow (x_1, x_2)_{n+1}$$

<u>Metropolis Algorithm</u>

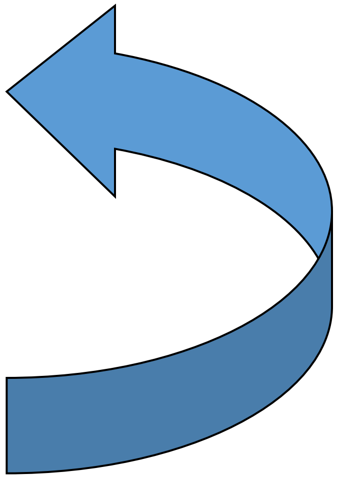
First, create a way of randomly picking a new set of params from the old. (must be symmetrical. E.g. Add normal deviate to each).

•From last point, (x_1,x_2), generate new point, (x_1^*, x_2^*)

•Calculate A, where $A = p(x_1^*, x_2^*) / p(x_2, x_2)$

•Get a random number, u, between [0,1].

•If $u<A$, accept new point. Otherwise use old one again.

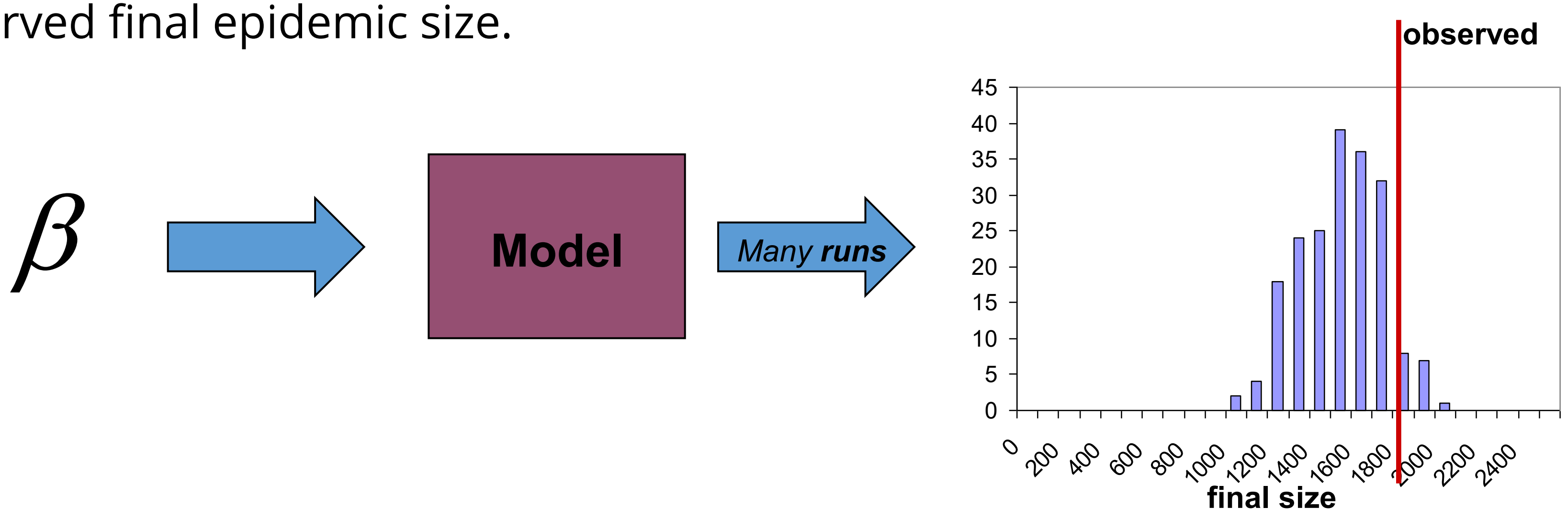


N.B. successive points are *<u>not</u>* independent!

Stochastic and probabilistic models: likelihood

•Stochastic models incorporate the randomness of the processes they represent, e.g. infection, recovery, death, etc.
•Hence each run of the model will produce different results from the same parameters.

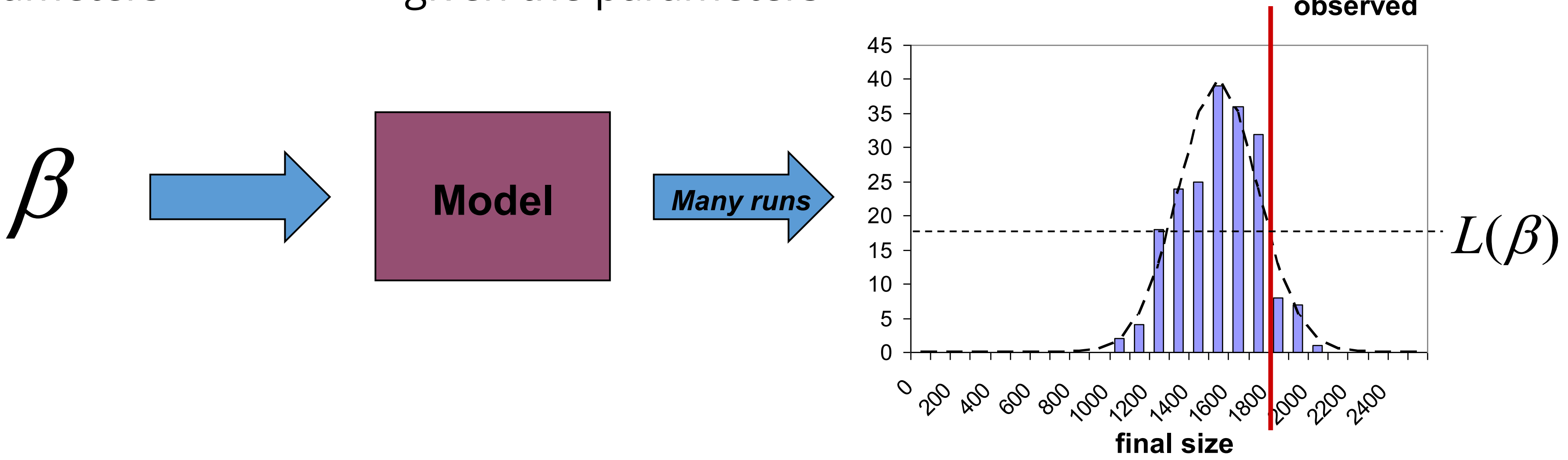
Consider an epidemic model: we want to know what values of β produce an observed final epidemic size.



Questions:
•What is the 'best' value for β?
•What range of values of β are acceptable?

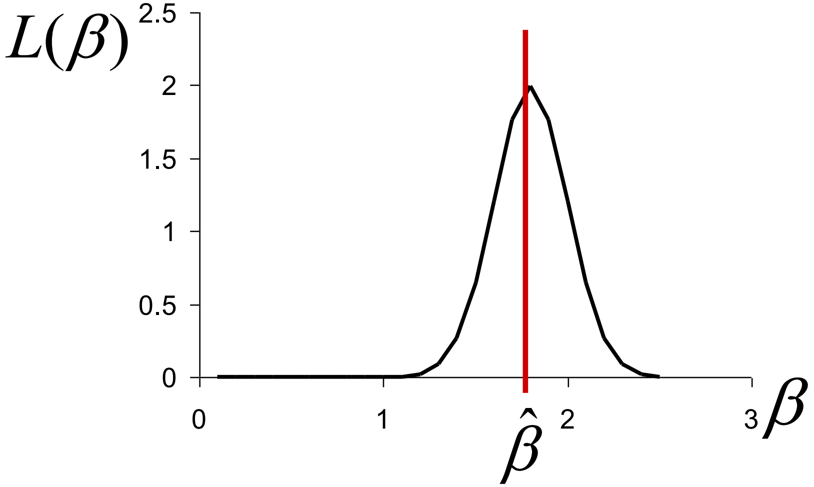
Likelihood

Likelihood of data, given parameters = Probability that the observed data is generated by the model, given the parameters



By repeating for a range of β's, we can generate a graph of likelihood against β.

The peak indicates the 'best' value, called the maximum likelihood estimator (MLE).



Log-Likelihood and confidence intervals

The shape of the log of the likelihood curve gives us the range of values of b that fit well.
If we measure down Q from top of the curve,

$$Q = \frac{1}{2}\chi^2_{0.05}(n)$$

This defines the 95% confidence for the parameter values.

Uncertainty arising from data

Consider sampling a population to estimate the proportion seropositive for a malaria antigen. Assume that:

Individuals sampled = **N**

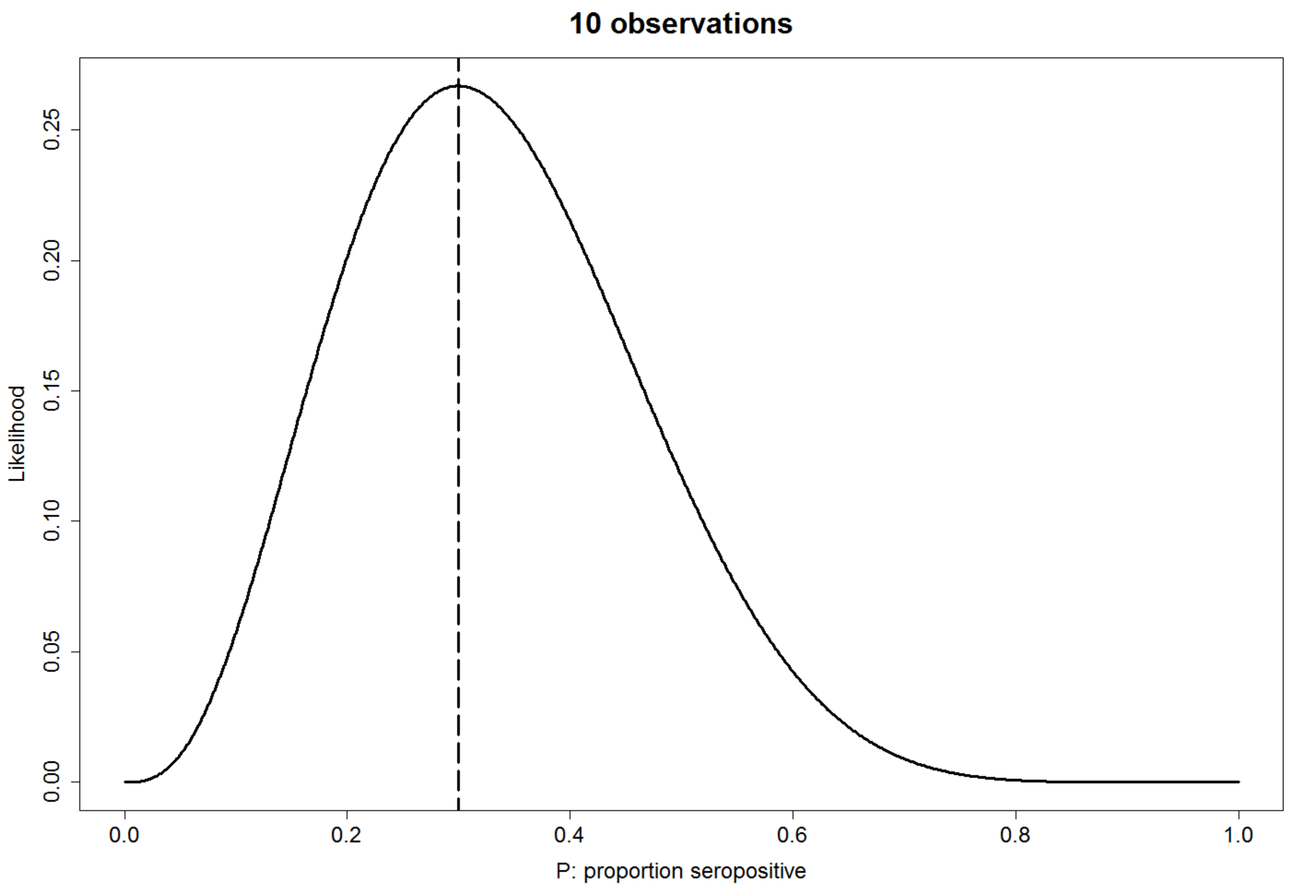
Individuals seropositive = **k**

Need to estimate the proportion seropositive **p**.

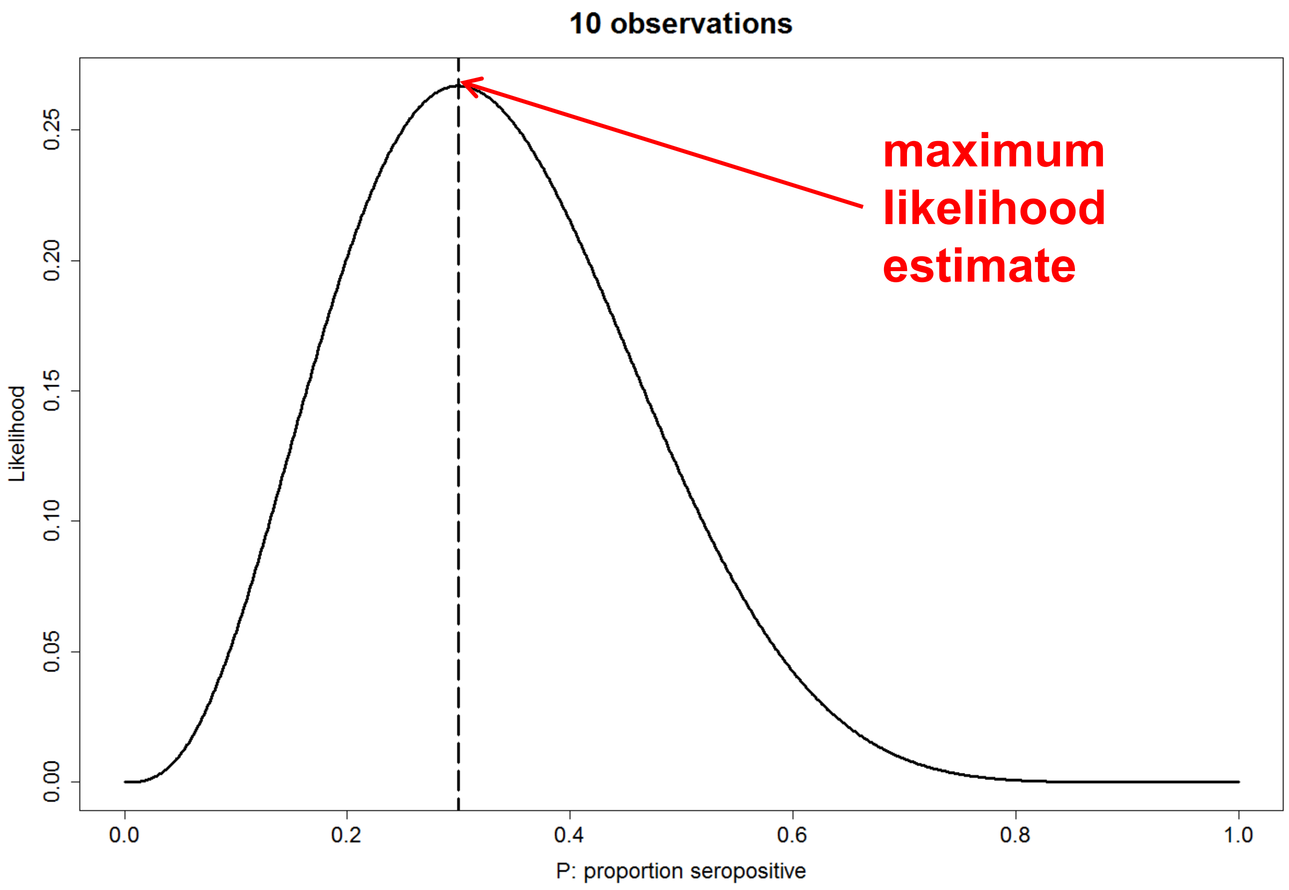
We use a binomial likelihood function:

$$L = \binom{N}{k} p^{k} \left(1 - p\right)^{N-k}$$

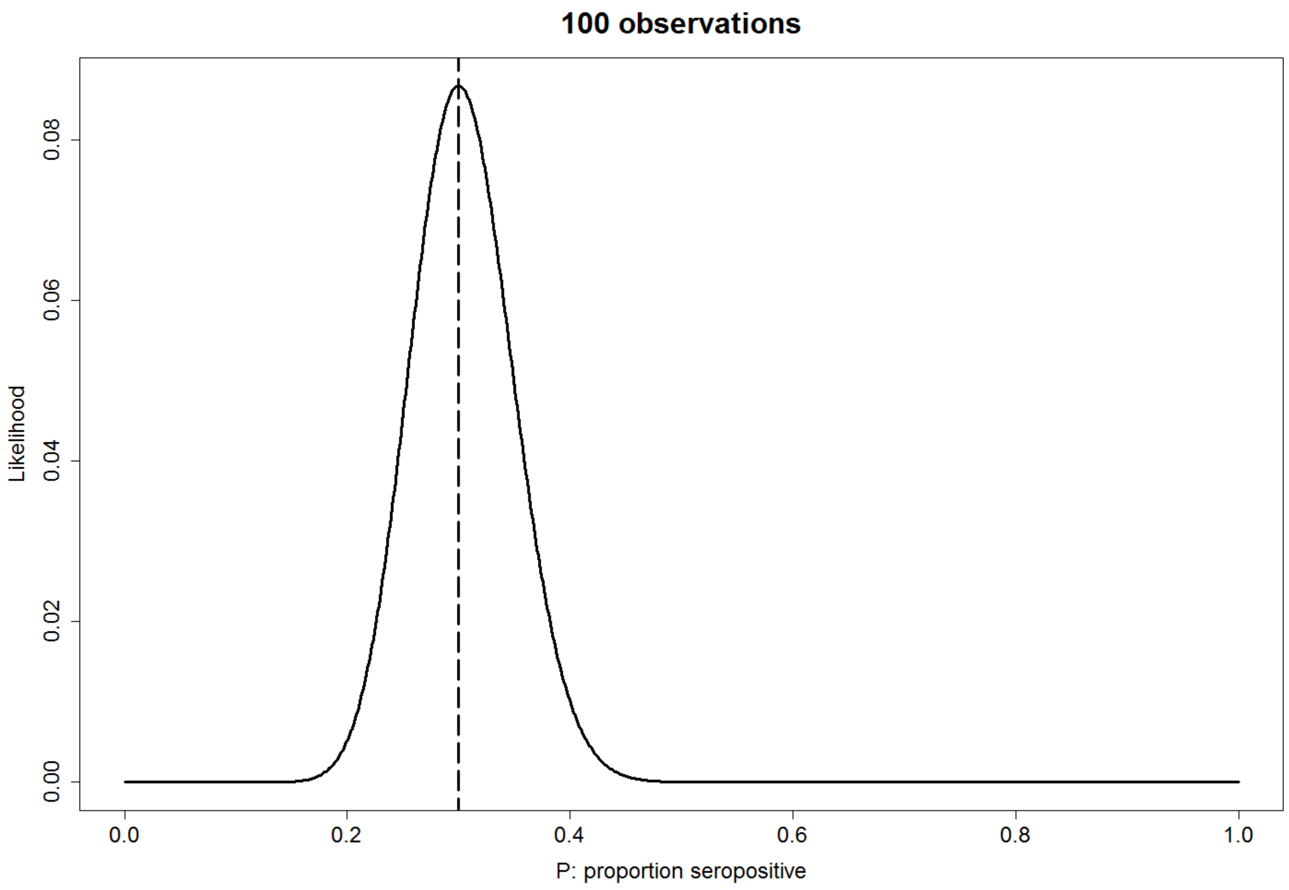
Uncertainty arising from data



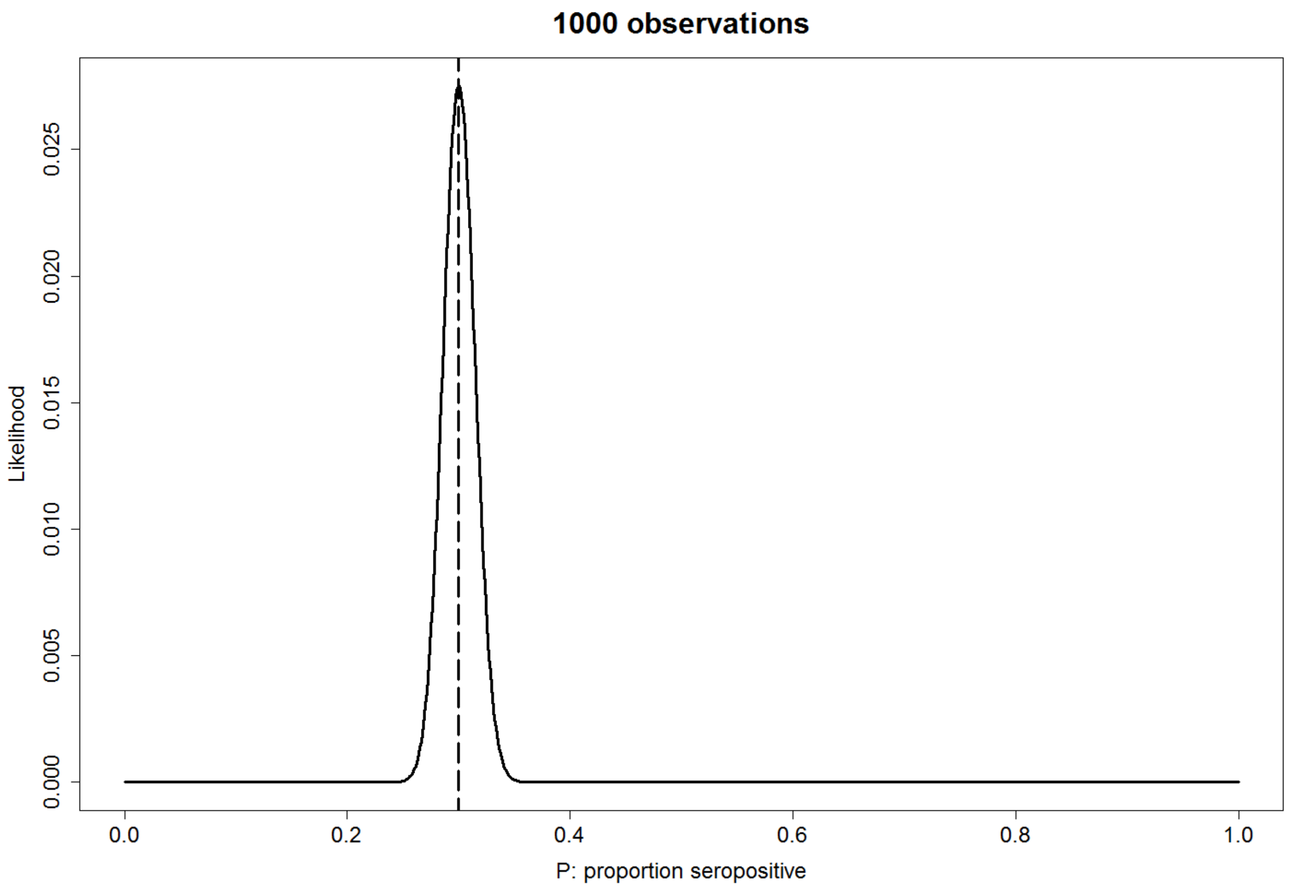
Uncertainty arising from data



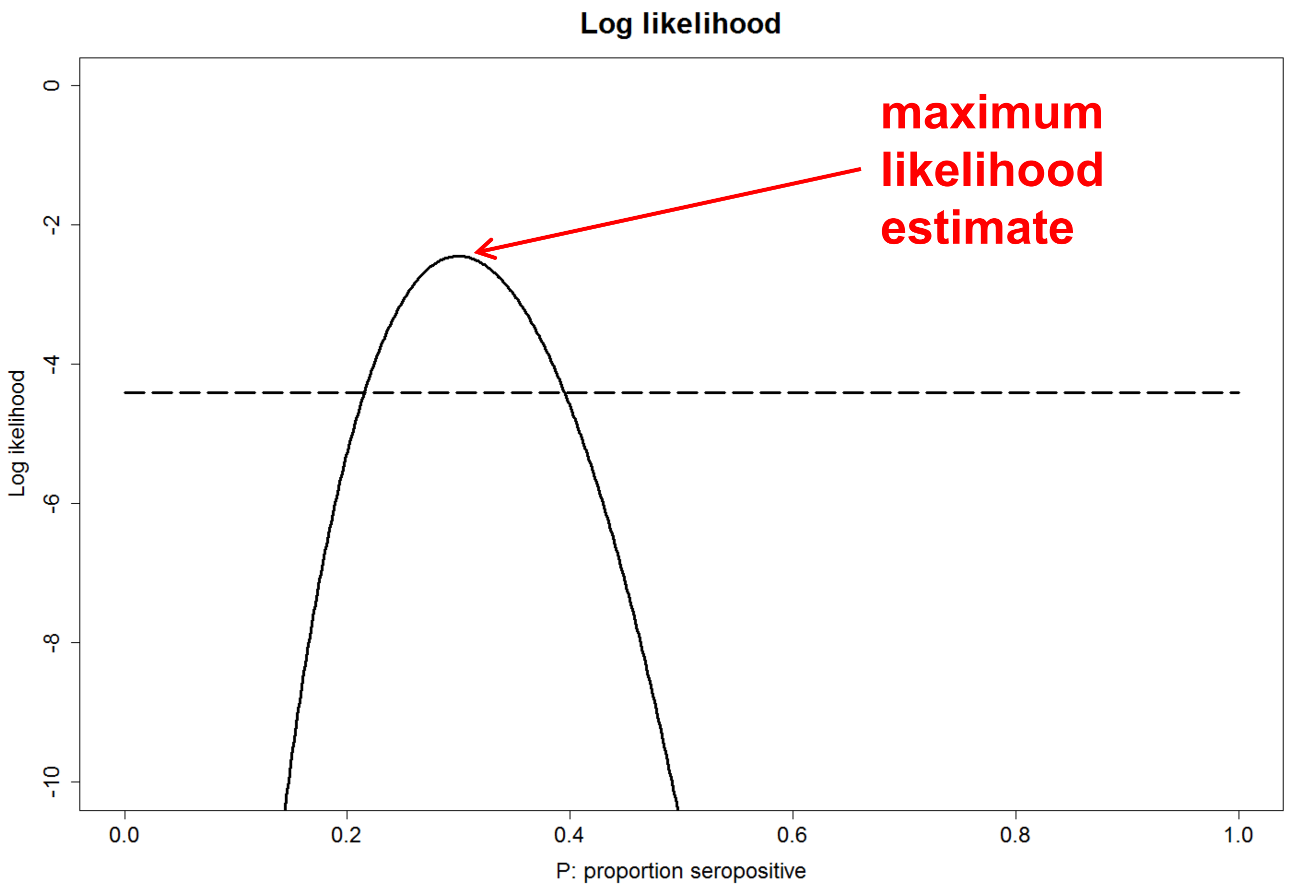
Uncertainty arising from data



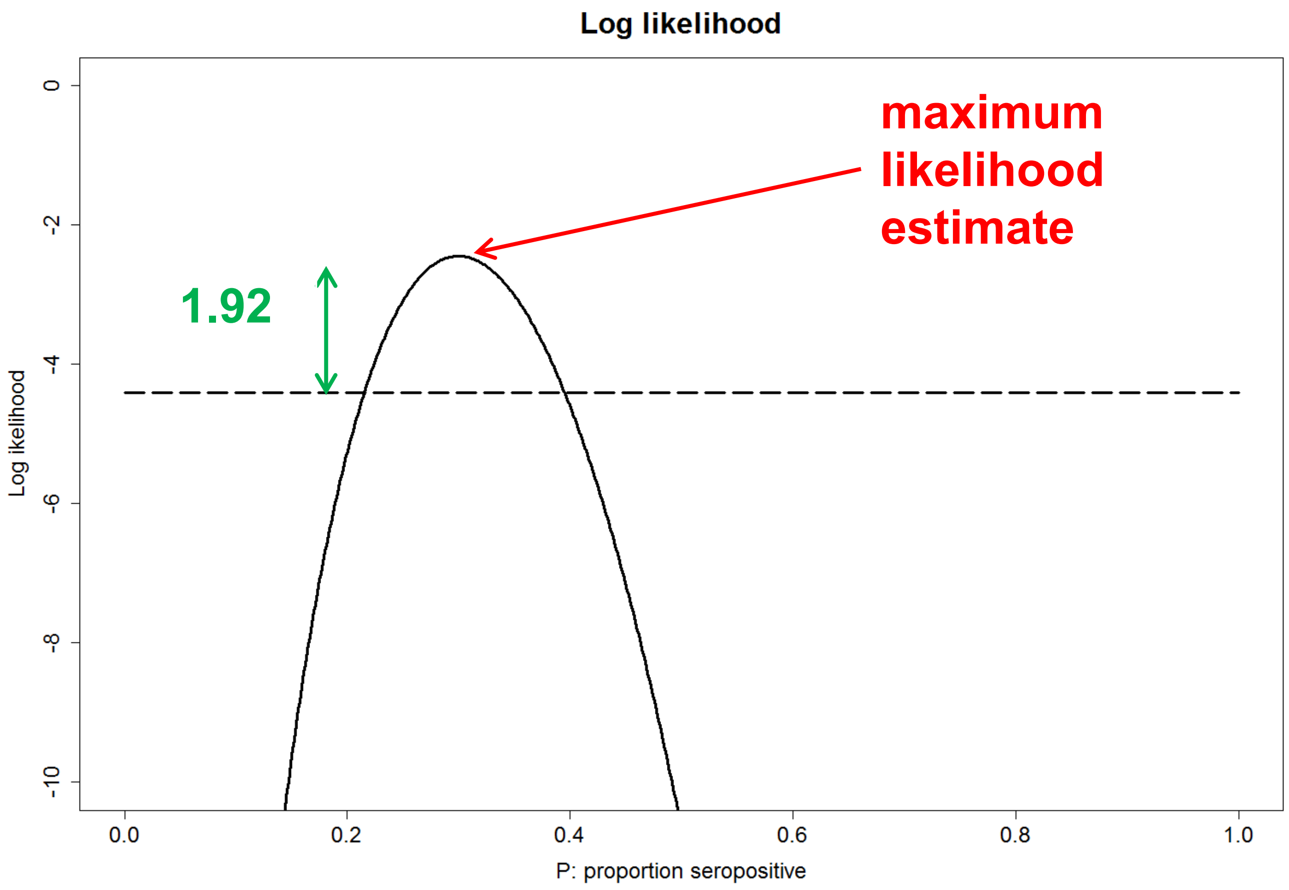
Uncertainty arising from data



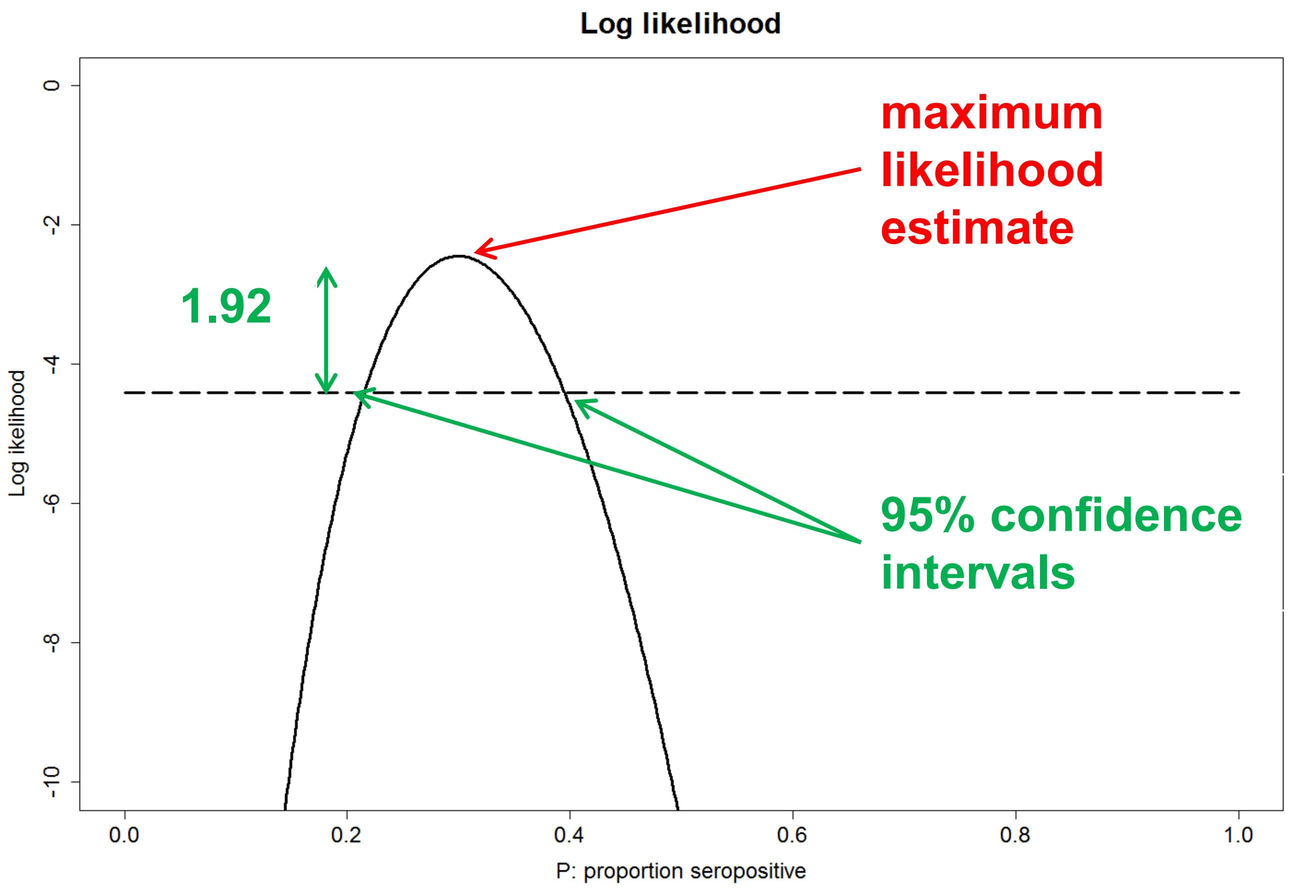
Uncertainty arising from data



Uncertainty arising from data



Uncertainty arising from data



Uncertainty arising from data

In general:

- more data => less uncertainty
- more data => narrower confidence intervals

Can depend on the choice of model, e.g. a badly chosen model will not be helped by more data.

Summary

- Looked at a number of ways to explore possible parameter values and
- Estimate the distribution and sensitivity of outcome variables.
- Another method for generating parameter set for analysis is through
- Monte Carlo Markov Chain algorithms. Very simple and robust algorithm in common use.
- Likelihood methods very powerful for finding parameter ranges corresponding to observed data.
- Calculating the likelihood becomes more difficult with more complicated data sets, e.g. incidence curves.
- The more data you have, the more accurate your parameter estimate.
- If the model's no good, you can still get a best fit, so check it's not nonsense.