Introduction to Infectious Disease Modelling and its Applications, LSHTM, 18 - 29 June 2018

Fitting models to data
II. Numerical optimisation and sensitivity analysis

Mark Jit[1,2,3]
[1]London School of Hygiene & Tropical Medicine
[2]Modelling and Economics Unit, Public Health England
[3]School of Public Health, University of Hong Kong

25 June 2018

Improving health worldwide

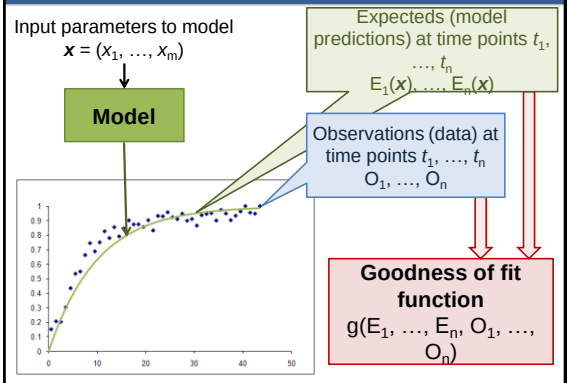
www.lshtm.ac.uk

Objectives

By the end of this lecture you should be able to:

1. Explain the purpose and some shortcomings of numerical optimisation algorithms, using gradient descent as an example.
2. Explain the need for sensitivity analysis to explore changes in results when input parameters are varied.
3. Conduct one-way sensitivity analysis using Berkeley Madonna.
4. Explain the purpose of multi-way sensitivity analysis and the principles behind different methods of doing this (grid search, random sampling, Latin hypercube sampling).
5. Use and interpret histograms and tornado graphs to show the results of sensitivity analyses.

RECAP ON MODEL FITTING



FITTING ALGORITHMS

Once we have a measure of goodness of fit, how do we find the best fitting parameters?

We need an algorithm to find $\boldsymbol{x}$ in order to:

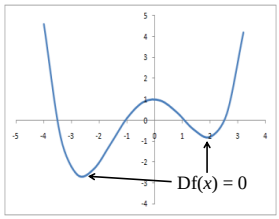
Maximise f($\boldsymbol{x}$) subject to $\boldsymbol{x} \in X$

Or equivalently

Minimise -f($\boldsymbol{x}$) subject to $\boldsymbol{x} \in X$

This is called **numerical optimisation**, and a very large body of mathematical theory exists about it in the field of numerical analysis.

FITTING ALGORITHMS



Consider a function f(x) of a single variable x.

NB. Df(x) = df/dx

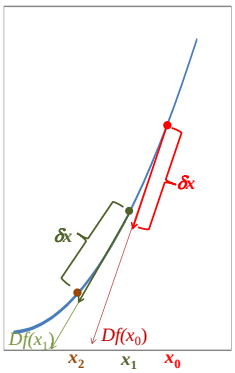
If we know the derivative of f(x), Df(x), then we know that when f(x) is minimised, then Df(x) = 0.

In practice, we rarely have f($\boldsymbol{x}$) in an explicit expression, let alone its derivative. However, we can estimate Df(x) numerically for a given $x=x_0$.

This idea gives rise to one of the simplest numerical optimisation algorithms …

GRADIENT DESCENT

1. Choose a starting point x_0.
2. Search in the direction that f is decreasing most rapidly (the downhill gradient -Df(x_0)).
3. Move in that direction a certain distance δx.
4. Get to a new point $x_1 = x_0 - \delta x$ Df(x_0).
5. Repeat until Df(x) is sufficiently small.

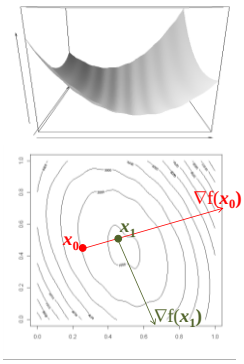


GRADIENT DESCENT

If f(***x***) is a function of two variables ***x***=(x_1, x_2), we look for ***x*** such that ∇f(***x***) = 0.

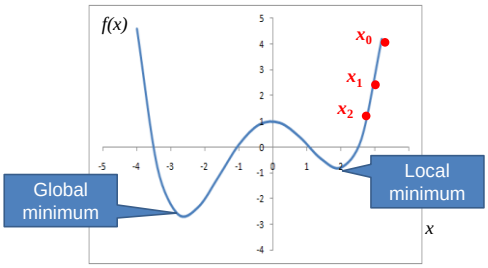
∇f(***x***) 'del f(***x***)' is a generalisation of Df(x) when ***x*** is a vector rather than a scalar (a point in more than one dimension).

In fact we can generalise this to f(***x***) being a function of n variables ***x***=(x_1, …, x_n).

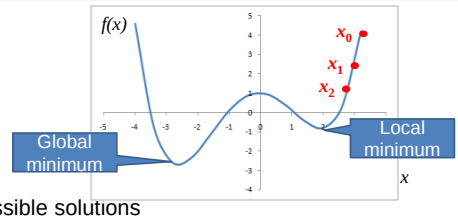


GRADIENT DESCENT

Gradient descent runs into problems when you have multiple local minima. (There's an example of this in the CJD practical next week.)



GRADIENT DESCENT



Possible solutions

- Try to start reasonably close to the global minimum if you have an idea where it is
- Take multiple starting points.
- Use a probabilistic algorithm eg. simulated annealing.

You should also make sure the final solution is physically and biologically plausible.

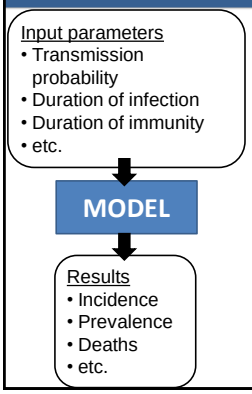
OTHER FITTING ALGORITHMS

In practice, gradient descent is not used for problems of significant complexity because it is inefficient. Instead, we tend to use algorithms such as:

- Levenberg-Marquardt (Berkeley Madonna, MATLAB and Mathematica use this)
- Generalised reduced gradient (used by Excel Solver)
- Nelder-Mead (downhill simplex)

Technical details of these algorithms aren't really necessary, but they all suffer from similar problems as gradient descent (though to a lesser extent). This is partly why Excel Solver and Berkeley Madonna have problems with fitting (you need to start them off at the right place).

SENSITIVITY ANALYSIS



Input parameters to a model are uncertain. To what extent does uncertainty in these parameters affect results?

This depends on two things:

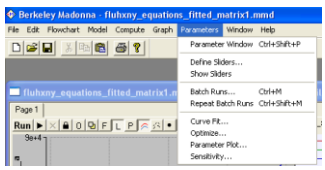
- The magnitude of uncertainty around each parameter
- How important each parameter is (some parameters are more influential than others in a non-linear model).

(Parametric) sensitivity analysis explores the change in results when input parameters are varied.

ONE-WAY SENSITIVITY ANALYSIS

One-way sensitivity analysis

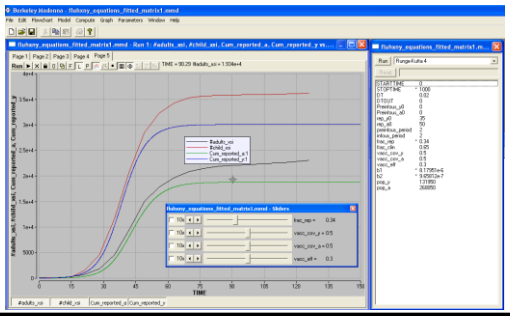
- Change the value of individual parameters (keeping the remaining parameters fixed), and see what effect this has on outcomes of interest.
- Berkeley Madonna has options under the "Parameters" menu that can do this.



ONE-WAY SENSITIVITY ANALYSIS

Berkeley Madonna

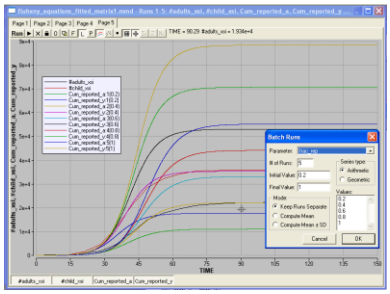
Using sliders or the parameter window



ONE-WAY SENSITIVITY ANALYSIS

Berkeley Madonna

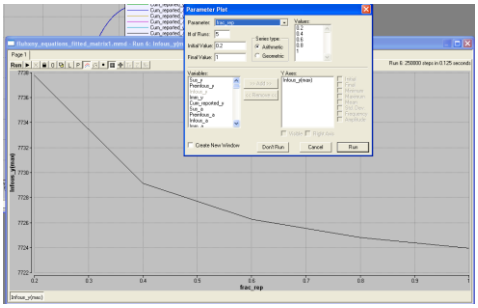
Using batch runs



ONE-WAY SENSITIVITY ANALYSIS

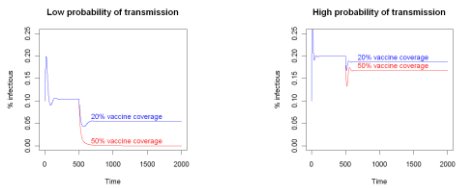
Berkeley Madonna

Using parameter plots



MULTI-WAY SENSITIVITY ANALYSIS

- Varying parameters one at a time while holding others at base case values does not give a complete description of the sensitivity of the model to each parameter.
- Consider an SIR model with vaccination introduced at



- Increasing vaccine coverage from 20% to 50% makes a much bigger difference on prevalence when the probability of transmission is low.

GRID SEARCH

How can we explore the joint effect of more than one parameter at the same time?

One method is **grid search** – systematically search the joint parameter space of all relevant parameters.

For two parameters (*A* and *B*) with 3 values each, we need to sample $3^2 = 9$ parameter sets.

If number of parameters or number of values for each parameter is large then grid search becomes unfeasible.

With 5 parameters and 10 values each, we need 100,000 sets.

		A		
		a_1	a_2	a_3
B	b_1	X	X	X
	b_2	X	X	X
	b_3	X	X	X

RANDOM SAMPLING

Instead of systematic exploration of the parameter space, we can use a technique called **Monte Carlo sampling**[1]:

1. Pick a value for each parameter we are uncertain about from some distribution. (For example, we may pick *A* and *B* uniformly from the range 0.01 – 0.5).
2. Evaluate the outcome measure by solving the model for that set of parameters.
3. Repeat this process many times (e.g. 100,000).

This form of sensitivity analysis is called **probabilistic sensitivity analysis**.

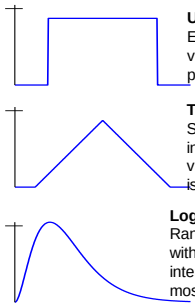
[1]Named after the Monte Carlo casinos. Nick Metropolis, one of its inventors, thought of the

RANDOM SAMPLING

How do we choose appropriate probability distributions for each parameter?

We can use:

- Sampling distributions in epidemiological studies.
- Expert elicitation.
- Some form of evidence synthesis (eg. meta-analysis) of available data.



Uniform
Equally likely values within a plausible range.

Triangular
Some intermediate value in a range is most likely.

Lognormal
Range in (0,∞), with an intermediate value most likely.

RANDOM SAMPLING

Important note: The probability distribution around a parameter should represent the **uncertainty** around that parameter rather than the **variability.**

Uncertainty – Lack of knowledge about a quantity, which can be reduced by further study (eg. taking a larger sample).

Variability – Heterogeneity between individuals, which is inherent in the population and will not be reduced by further study.

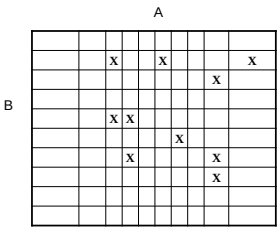
For instance, suppose we wanted to estimate the mean height of people in this room.

If we measured the height of everyone in this room, there would be low uncertainty in our mean estimate, but a lot of variability.

On the other hand, if we sampled two people in this class and they were both 1.7 metres tall, there would be no variability in our estimate but a great deal of uncertainty.

RANDOM SAMPLING

Random sample for two parameters

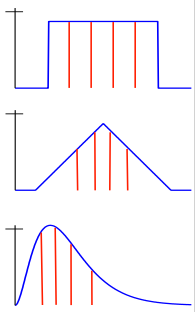


If we sample completely at random from the probability distribution of each parameter, the method is inefficient as it doesn't ensure full coverage of parameter space. Instead we may get clusters of points (i.e. over-sample certain parts of the parameter space and under-sample others).

LATIN HYPERCUBE SAMPLING

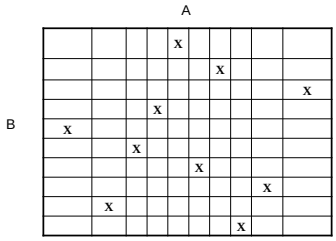
A more efficient way to sample is **Latin hypercube sampling**. This uses the following method:

1. Divide the probability distribution of each parameter into N equal probability sections.
2. Sample each parameter from one of its available sections (without replacement).
3. Repeat until you have built up N equal parameter sets that together encompass all sections of all parameters.



LATIN HYPERCUBE SAMPLING

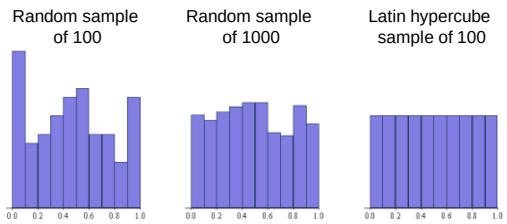
A Latin hypercube sample



The values chosen are more evenly distributed than in a simple random Monte Carlo sample.

COMPARISON OF SAMPLES

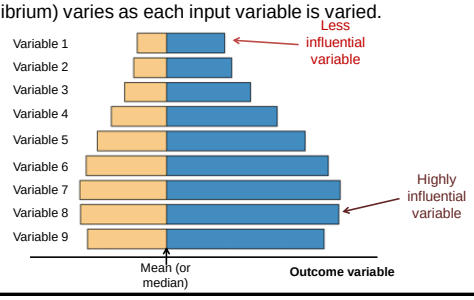
Samples from uniform distribution on (0, 1)



DISPLAYING RESULTS

How do we display the results?

Simplest way is to use a **tornado graph** – this shows how the outcome variable (eg. prevalence of infection at equilibrium) varies as each input variable is varied.



DISPLAYING RESULTS

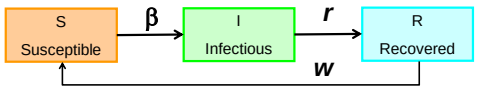
How do we construct a tornado plot?

One-way sensitivity analysis. Vary each input parameter within a certain range, and record how the output variable changes.

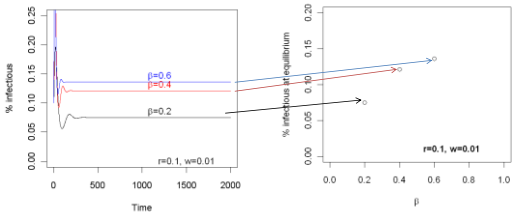
Multi-way sensitivity analysis. Construct a statistical (eg. linear) model of the association between the outcome variable and each of the input parameters, parameterised using the sampled parameter sets. Now vary each input parameter within its range in the model.

AN EXAMPLE

Consider an SIRS model (eg. pertussis)

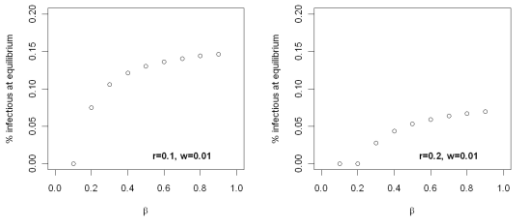


Vary β while keeping *r* and *w* fixed (one-way sensitivity analysis)



AN EXAMPLE

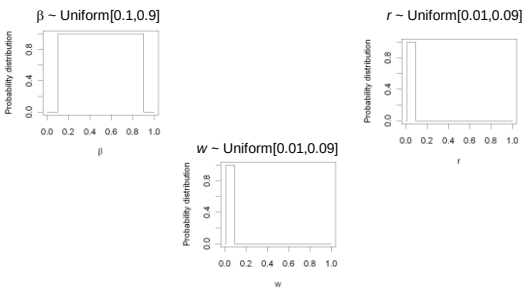
Comparing the effect of β for two different values of *r*:



Clearly the effect that β has on the % infectious at equilibrium changes for different values of *r*. (As expected – a short duration of infectiousness dampens the importance of the probability of transmission.)

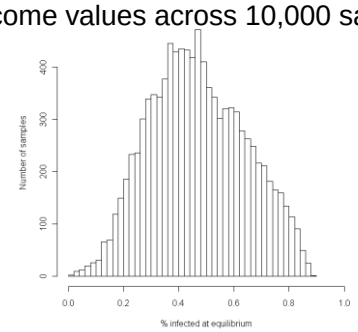
AN EXAMPLE

To investigate the joint effect of varying the parameters, sample β, *r*, *w* from the following uniform distributions.



AN EXAMPLE

Histogram showing the distribution of outcome values across 10,000 samples:



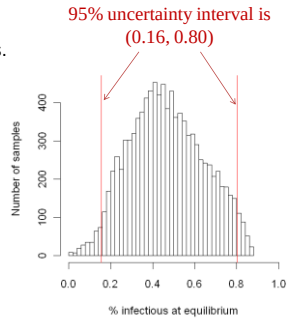
AN EXAMPLE

Uncertainty intervals

We can make probability statements about outcomes.

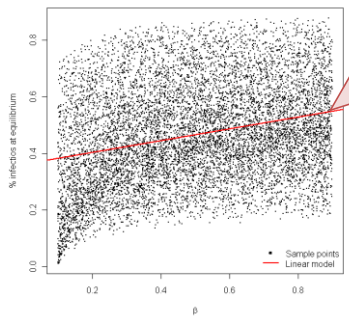
For instance, suppose for 2.5% of simulations the outcome Y is below Y_1 and for 2.5% it is above Y_2. Then (Y_1, Y_2) is a 95% uncertainty interval for Y.

This depends on the probability distributions chosen for the parameters.



AN EXAMPLE

Effect of varying β



The linear model predicts the average % infectious at equilibrium for different values of β when r, w are sampled over their entire distributions.

AN EXAMPLE

How did we construct the model?

$$y_i \sim c_1\beta_i + c_2 r_i + c_3 w_i + \varepsilon_i \; (i=1,\ldots,10000)$$

% infected at equilibrium

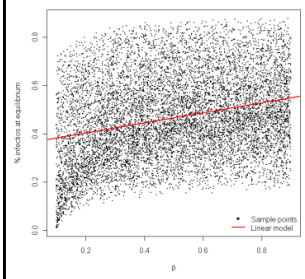
input parameters

error term

Notice that the model has no interaction terms between variables. This isn't realistic, but is helpful for the purposes of sensitivity analysis.

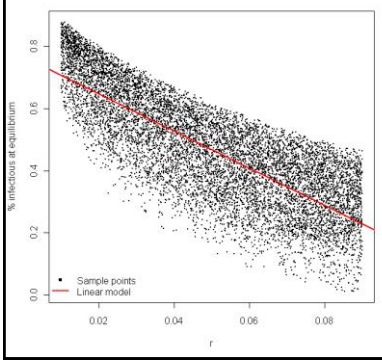
Coefficient	Estimate	Standard Error	t-value	P-value
Intercept	0.486526	0.001859	261.7	<0.001
c_1	0.208670	0.002075	100.6	<0.001
c_2	-5.940816	0.020759	-286.2	<0.001
c_3	3.485241	0.020783	167.7	<0.001

AN EXAMPLE



- We want to find the **marginal** effect of β (the effect of β regardless of what values the other parameters take).
- To do this, use the linear regression model, and assume that r and w are held constant and take their mean values.
- This is okay because there are no interactions between variables in our linear model.
- It is obviously an approximation, but it allows us to get a better handle on the influence of β than varying β in the full model.

AN EXAMPLE



- The relationship between r and the equilibrium level of infectiousness is negative.
- This is as we might expect – when the duration of infectiousness is short (rate of losing infectiousness is high), the endemic equilibrium level of infectiousness is lower.

OTHER SOURCES OF UNCERTAINTY

So far, we have only considered sensitivity of results due to input parameters. Other forms of sensitivity analyses take into account uncertainty in:

- **Model structure.** For example, you could build SIS, SIR and SIRS models and see how outcomes change across different structural choices.
- **Type of model.** You may need to consider how outcomes may change if for example you used a stochastic model instead of a deterministic model.
- **Initial conditions.** This should always be checked, but is unlikely to be important for deterministic models that are run to endemic equilibrium. It is more important for stochastic low-prevalence models and models of an unfolding epidemic.

If you are interested in this topic see the following paper:

Bilcke J *et al*. Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide. Med Decis Making 2011; 31(4): 675-92.

RECAP

You should now be able to do the following:

1. Explain the purpose and some shortcomings of numerical optimisation algorithms, using gradient descent as an example.
2. Explain the need for sensitivity analysis to explore changes in results when input parameters are varied.
3. Conduct one-way sensitivity analysis using Berkeley Madonna.
4. Explain the purpose of multi-way sensitivity analysis and the principles behind different methods of doing this (grid search, random sampling, Latin hypercube sampling).
5. Use and interpret histograms and tornado graphs to show the results of sensitivity analyses.