Parameter estimation, uncertainty, model fitting, model selection, and sensitivity and uncertainty analysis

Jamie Lloyd-Smith

Center for Infectious Disease Dynamics
Pennsylvania State University

with thanks to Matt Ferrari for sharing some slides,

and big thanks to Ben Bolker for making his wonderful book available online.

Outline

Estimating R_0

Parameter estimation

- Likelihood approaches
- Bayesian approaches

Fitting more complex models

Estimating uncertainties

- Likelihood profiles
- Quadratic approximations
- Bootstrapping

Model selection

Examples

Sensitivity and uncertainty analysis

Resources for further study

The Ecological Detective: Confronting Models with Data
Ray Hilborn and Marc Mangel
Princeton Monographs in Population Biology, 1997

Ecological Models and Data in R
Ben Bolker
Princeton Monograph… 2008?
unpublished, but PDF available at http://www.zoo.ufl.edu/bolker/emdbook/

Infectious Diseases of Humans: Dynamics and Control
Roy Anderson and Robert May
Oxford 1991

Estimating R_0: from individual parameters

In its simplest form, $R_0 = \beta/\gamma = c\ p\ D$ where

c = contact rate
p = probability of transmission given contact
D = duration of infectiousness

So why can't we just estimate it from individual-level parameters?
Problems:

- for many diseases we can't estimate the contact rate, since "contact" is not precisely defined. The exceptions are STDs and vector-borne diseases, where contacts are (in principle) countable, though heterogeneity complicates this.
- Estimates based on R_0 expressions are highly model-dependent.
- $E(c\ p\ D) \neq E(c)\ E(p)\ E(D)$ in general.

Estimating R_0: from epidemic data

Epidemic time series data are very useful in estimating R_0.

Simple analysis of the SIR model yields two useful approaches:

1) If the exponential growth rate of the initial phase of the epidemic is r, then $\boldsymbol{R_0 = 1 + rD}$
2) Equivalently, if t_d is the doubling time of the number infected, then
$$\boldsymbol{R_0 = 1 + \frac{D \ln 2}{t_d}}$$
3) If s_0 and s_∞ are the susceptible proportions before the epidemic and after it runs to completion, then
$$\boldsymbol{R_0 = \frac{\ln(s_0) - \ln(s_\infty)}{(s_0 - s_\infty)}}$$

Estimating R_0: from epidemic data

All of those estimates are based on simple ODE models, and hence assume exponentially distributed infectious periods.

Wallinga and Lipsitch (2007, Proc Roy Soc B 274: 599-604) analyze how the distribution of the serial interval influences the relationship between r and R_0.

They find $\boldsymbol{R_0 = \frac{1}{M(-r)}}$

where $\boldsymbol{M(z)}$ is the moment generating function for the distribution of the serial interval.

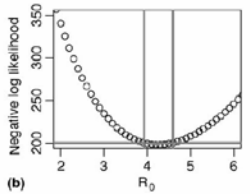
→ 1. Can calculate R_0 from r for any distribution of serial interval.

→ 2. Prove that the upper bound on R_0 is $R_0 = e^{rT}$ where T is the mean serial interval.

Estimating R_0: from epidemic data

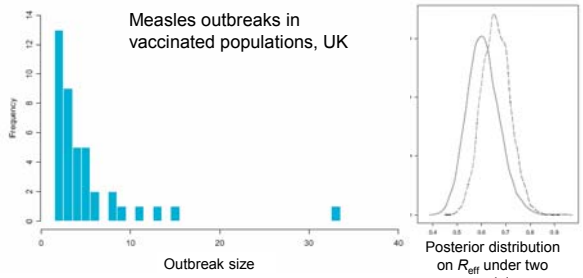
If case data are collected in discrete intervals, estimation from continuous-time models is difficult.

Ferrari et al (2005, Math Biosci 198: 14-26) derive an approach based on chain binomial models that provides a maximum-likelihood estimator for R_0 and the associated uncertainty.



But, like the s_∞ approaches, it requires that the epidemic runs to its natural completion.

Estimation from outbreaks when $R_0 < 1$



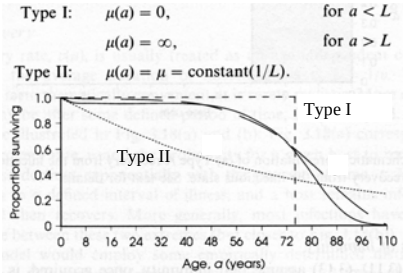
Branching process models allow analysis of outbreak size to make inference about the effective reproductive number when $R_{eff}<1$.

Farrington et al (2003) Biostatistics 4: 279-295.

Estimating R_0: from endemic data

Anderson & May derive a number of simple expressions for R_0 in the endemic setting.

Their results depend on the age-dependent rate of mortality in the population $\mu(a)$, which yields a "Type I" or "Type II" mortality curve.



Estimating R_0: from endemic data

Anderson & May derive a number of simple expressions for R_0 in the endemic setting.

For Type I mortality, $R_0 \approx L/A$

where L is the mean lifespan and A is the mean age at first infection.

For Type II mortality, $R_0 = L/A$ exactly.

Of course, these simple estimates depend on strong assumptions about random mixing, no heterogeneities, no age-dependence of the force of infection, and constant population size.

See later chapters of Anderson & May, or Dietz (1993) Stat Meth Med Res 2: 23-41, for more advanced treatments.

Estimating R_0: from age-seroprevalence data

From age-seroprevalence data,

the age-dependent force of infection can be estimated directly.

To estimate R_0, need to make assumption about WAIFW matrix.

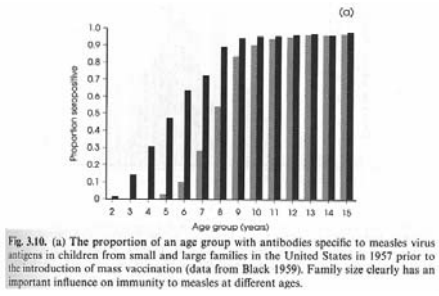
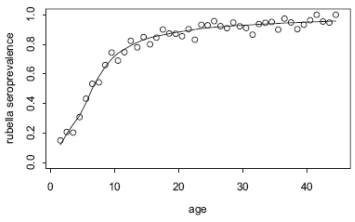


Fig. 3.10. (a) The proportion of an age group with antibodies specific to measles virus antigens in children from small and large families in the United States in 1957 prior to the introduction of mass vaccination (data from Black 1959). Family size clearly has an important influence on immunity to measles at different ages.

Estimation of effective reproduction numbers for infectious diseases using serological survey data

C. P. FARRINGTON*, H. J. WHITAKER

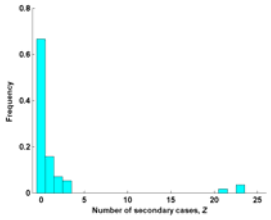
Biostatistics (2003), **4**, 4, pp. 621–632



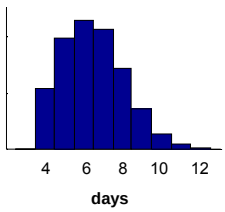
And see comprehensive review: "Estimation of the basic reproductive number for infectious diseases from age-stratified serological survey data" (2001) Appl Statist 50: 251-292.

Fitting single distributions

Observed offspring distribution for SARS in Singapore



Observed distribution of incubation periods for SARS



Fitting single distributions: method of moments

A fancy name for a simple idea:

For most standard probability distributions, the parameters can be expressed in terms of the moments of the distribution (e.g. the mean and variance).

e.g. exponential distribution, $f(x) = \lambda \exp(-\lambda x)$

$$\lambda = 1/\mu$$

Method of moments: calculate the sample moments from your data, and plug them into these expressions

$$\lambda \approx 1/\overline{\mu}$$

Estimates may be biased, but this is a good way to get a quick estimate.

Fitting single distributions: maximum likelihood

The likelihood is the probability of observing the data given the model (and parameter values for the model).

Y = data set, $\{Y_1, Y_2, \ldots Y_n\}$

p = model parameters

Then the likelihood is $\Lambda = \Pr(Y|p)$,

where the "model" (in this case the probability distribution we're fitting) will determine the form of the probability.

The basic idea of maximum likelihood estimation (MLE) is to find the parameter set that maximizes the likelihood of observing your data.

Example: the binomial distribution

The binomial distribution describes the number of successes out of N trials, if each trial has probability p of success.

For a single observation from a binomial distribution (say, the number of susceptibles infected in a day, out of a beginning total of N), the likelihood that k out of N are infected, if the per capita infection probability is p, is

$$\text{Prob}(k|p,N) = \binom{N}{k} p^k (1-p)^{N-k}$$

If we have n independent observations of this process, each with the same number of susceptibles N, and the number infected on the i^{th} observation is k_i, then the likelihood is

$$\mathcal{L} = \prod_{i=1}^{n} \binom{N}{k_i} p^{k_i} (1-p)^{N-k_i}$$

Likelihood example: binomial distribution

$$\mathcal{L} = \prod_{i=1}^{n} \binom{N}{k_i} p^{k_i} (1-p)^{N-k_i}$$

It is conventional to work with the log-likelihood, $L = \log(\mathcal{L})$, for two reasons:

(1) It turns the product (which arises from the joint probability of many independent observations) into a sum.

(2) The probabilities are often very small numbers (usually <<1!), and working with a product of small numbers causes numerical problems in computation.

The log-likelihood for our binomial problem is:

$$L = \sum_{i=1}^{n} \left(\log \binom{N}{k_i} + k_i \log p + (N - k_i) \log(1-p) \right)$$

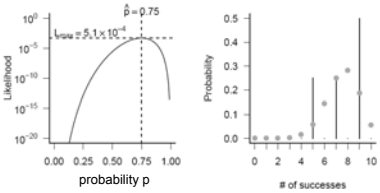
Likelihood example: binomial distribution

We now want to find the parameter p that maximizes L.

For this simple example, it can actually be calculated analytically, and yields a sensible answer:

$$\hat{p} = \frac{\sum_{i=1}^{n} k_i}{nN}$$

Example plot of binomial likelihood curve and fit to data, from Bolker 200X.



Maximum likelihood estimation

Other common distributions also have simple MLE parameters:

e.g. the ML estimate for the mean of the Poisson, normal, exponential, gamma, and negative binomial distributions are all equal to the mean of the data.

For most problems, though, this maximization cannot be solved analytically, so we optimize numerically.

It is conventional to minimize the negative log-likelihood (NLL).

For more complex distributions, such as the gamma or negative binomial, this can be a multi-dimensional optimization problem.

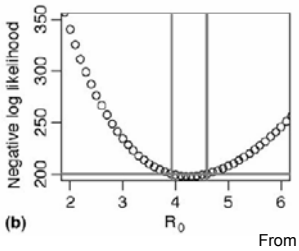
This can be handled in various software packages:

fminsearch in Matlab (not Octave?), optim in R, ____ in Python (?)

Maximum likelihood estimation

In addition to finding the optimal value (the maximum likelihood estimate) it is very useful to examine the likelihood curve in the area near the MLE.

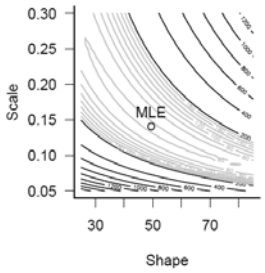
This gives information how the likelihood changes with the parameter value, and tells you about uncertainty.



From Ferrari et al (2005)

Maximum likelihood estimation

Example plot of likelihood surface from MLE fit to a gamma distribution (Bolker 200X)



For a multivariate problem, this will be a likelihood surface, and gives information about correlation between parameter estimates as well as uncertainty.

Bayesian vs frequentist statistics

The maximum-likelihood approach we just saw is an example of frequentist statistics.

In **frequentist statistics**, parameters are assumed to have fixed values that we are trying to estimate as precisely as possible.

In **Bayesian statistics**, in contrast, parameters are treated as random variables, with probabilities assigned to particular values of a parameter to reflect the degree of evidence for that value.

Fitting single distributions: Bayesian analysis

Bayesian estimation of distribution parameters is also based on the likelihood, but there are two major differences from MLE:

- The likelihood is combined with a prior probability distribution, which represents information from other sources regarding the values of the parameters.

  These elements are combined to yield a posterior probability distribution, which represents the our best estimate of the probability that the parameter takes certain values.

2. The Bayesian parameter estimates are usually given as the mean of the posterior distribution rather than the mode (as for MLE), because the mean encapsulates more information about the shape of the distribution.

Fitting single distributions: Bayesian analysis

In the setting of parameter estimation, if we have a dataset Y and model parameters θ, then Bayes rule states that the posterior distribution on θ is

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

$P(Y|\theta)$ is the likelihood.

$P(\theta)$ is the prior distribution, which we define.

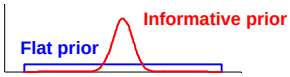
$P(Y)$ is the probability of the data.

Fitting single distributions: Bayesian analysis

How do we choose a prior distribution?

Many opinions on this:

- It is either a useful way to incorporate other information about your system or a "necessary evil".
- If a useful tool, then choose a distribution that reflects your prior information about the parameter.
- If a necessary evil, then
  - choose a "flat prior" that doesn't include much information about particular values (e.g. a uniform distribution)
  - or choose a prior that simplifies the computation



Fitting single distributions: Bayesian analysis

How do we find the probability of the data?
(and what does it even mean?)

There are two important facts about the P(Y) term

- It is a constant. This is very useful for some numerical techniques where we're interested in the ratio of posterior probabilities.
- The posterior probability distribution must be normalized. So we can write:

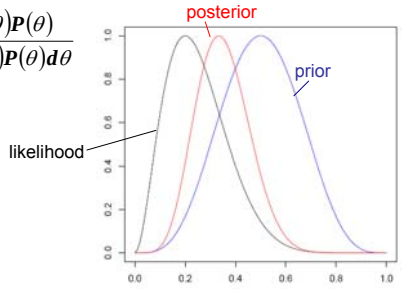
$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}$$

For simple problems this integral can be calculated numerically.
For high-dimensional problems we need other tricks (MCMC).

Fitting single distributions: Bayesian analysis

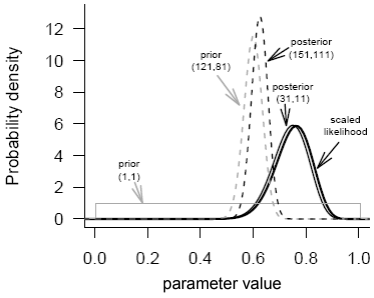
Simple rule to remember is that the posterior is proportional to the product of the likelihood and the prior.

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}$$



Fitting single distributions: Bayesian analysis

The more "informative" your prior, the more it will influence the shape of your posterior.



Example:
Bayesian estimates of binomial probability with different prior distributions
(Bolker 200X)

Fitting single distributions: Bayesian analysis

For multi-parameter distributions, you get a multivariate posterior.
e.g. for a gamma distribution with parameters {*a*,*s*}

$$\text{Posterior}(a,s) = \frac{\text{Prior}(a,s) \times L(a,s)}{\iint \text{Prior}(a,s)L(a,s)\,da\,ds}$$

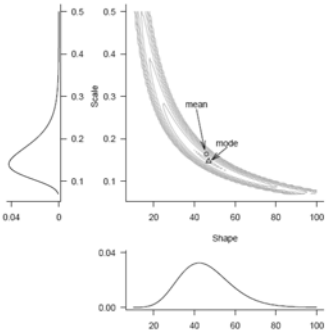
To learn about parameters individually, look at marginal distributions:

$$\text{Posterior}(a) = \int \text{Posterior}(a,s)s\,ds$$

Or take mean values:

$$\bar{a} = \int \text{Posterior}(a) \cdot a\,da$$

Fitting single distributions: Bayesian analysis



Example:
Bivariate and marginal distributions for a Bayesian fit with a gamma distribution
(Bolker 200X)

Fitting more complex models

So far we've been estimating the parameters of probability distributions, but we'll often want to estimate parameters from more complex models – sometimes even from our whole dynamic model.

The challenge here is to define the likelihood for the model. Because the likelihood is based on probabilities, it requires that we think about the stochastic components of the processes that generated the data – including both the underlying mechanisms and the observation process.

This is not a simple problem, but there are two main approaches:

1) Consider whether the basic mechanism of the model corresponds to a clearly defined stochastic process.
2) Do a rough fit of the model to the data, and examine the residuals to look for systematic patterns that correspond to basic distributions.

Fitting more complex models

Some rules of thumb in defining likelihoods:

When the quantities in the data are:

- Proportions → consider a binomial distribution
- Rare events → consider a Poisson distribution, or negative binomial if there seems to be over-dispersion
- Sums of many contributions → consider a normal distribution
- Products of multiplicative probabilities → consider a log-normal distribution

Look for the corresponding patterns in the model residuals.

These are not definitive rules, just starting points. See Hilborn & Mangel (1997) for an excellent discussion of this problem.

Fitting more complex models

Examples of clear stochastic mechanisms:

1) In an epidemic model, if we know the number of susceptible and infectious individuals at each point in time, then the likelihood is binomial:

   $S(t)$ = the number of susceptible individuals
   $\lambda(t)$ = the force of infection = $\beta I(t)/N(t)$

   $p(t)$ = Pr(susc. becomes infected in time Δt) = $1-\exp(-\lambda(t)\Delta t)$

   and the number of new infections generated in $(t,t+\Delta t)$ is
   NewCases ~ Binomial ($S(t)$,$p(t)$)

2) If data are available in (close to) continuous time, then individual infection can be modelled as an exponential process with force of infection varying in time.
   (*example later in lecture*)

Fitting the whole model: sum-of-squares (χ^2)

If you simply can't see how to define a likelihood for your model, don't despair!

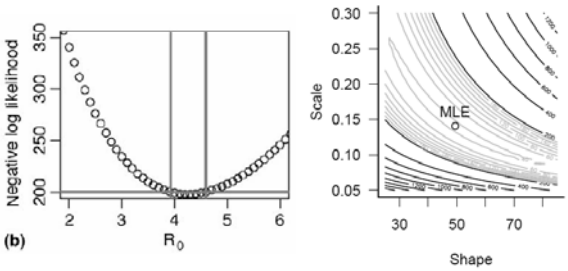
Many studies are published based on simpler fitting procedures, most frequently the method of least squares, or its close relative the χ^2 goodness-of-fit, which is based on minimizing the statistic

$$Y = \Sigma_i (O_i - E_i)^2 / E_i$$

where O_i = Observed value at point i (from real data)
E_i = Expected value at point i (from the model output)

Minimizing this quantity usually yields a decent fit to the data.

Estimating uncertainties: MLE approach

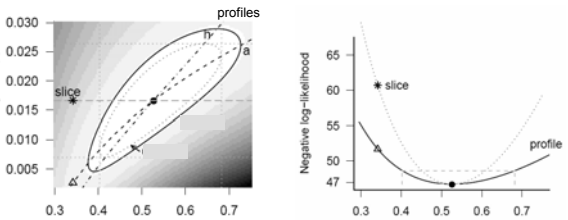


Negative log-likelihood curves and surfaces map out the "badness-of-fit" of different parameter values to the data.

We can analyze the curvature of these surfaces to get confidence intervals for our estimates.

Likelihood slices and profiles

(Figures from Bolker 200X)



Likelihood slice: fix the values of all but one parameter, and calculate the likelihood for a range of values of that parameter.

Likelihood profile: choose a range of values for the focal parameter, and for each value maximize the likelihood with respect to all other parameters.

Estimating uncertainties: MLE approach

Now we can calculate confidence intervals for our ML parameter estimates based on likelihood curves (for 1-d problems) or likelihood profiles (for higher-dimensional problems).

** Don't use likelihood slices!! **

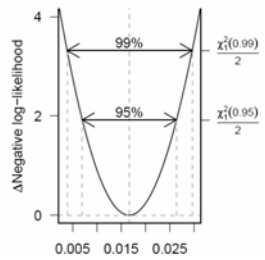
The difference in NLL values between the MLE and other points on a likelihood profile is asymptotically χ^2-distributed with one degree of freedom.

(the argument closely parallels that for the likelihood ratio test)

→ To find the (1-α)% confidence limits on our estimate, we find the parameter values corresponding to NLL values of

$$\mathbf{NNL_{MLE}} + \chi^2_1(1-\alpha)/2$$

Estimating uncertainties: MLE approach

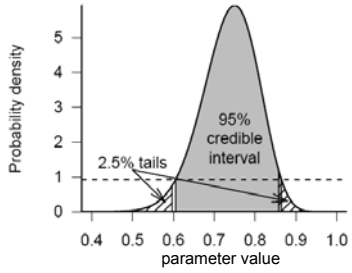


(Figure from Bolker 200X)

(Hopefully that's enough info that you can calculate profile confidence intervals yourself. If not, then R has in-built functions `profile` and `confint` that will do it for you (in the `bbmle` or `emdbook` packages).)

Estimating uncertainties: Bayesian approach

Instead of confidence intervals, Bayesians calculate so-called credible intervals which are the region in the center of the posterior distribution containing 95% of the density.



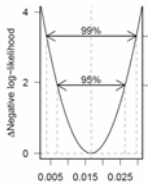
(Figure from Bolker 200X)

Estimating uncertainties: Quadratic approximation

The likelihood profile approach is great when you have a small number of parameters (i.e. 2 or 3),

but becomes computationally impractical for models with more parameters, since for an *n*-parameter model, you have to optimize over $n-1$ parameters for each point on your likelihood profile (never mind doing a 2-D profile!).

Luckily, classical likelihood theory tells us that we can learn about the variance of our estimate by considering the second derivative of the likelihood curve – essentially by using a quadratic approximation to the region around the minimum.



Estimating uncertainties: Quadratic approximation

It turns out that, asymptotically (i.e. if the data set is large enough), the sampling distribution for the parameter is asymptotically normal with standard deviation $\left(\frac{d^2(\log \mathcal{L})}{dp^2}\right)^{-1/2}$

→ the width of the interval that gives (1-α) confidence is

$$N(\alpha)\left(\frac{d^2(\log \mathcal{L})}{dp^2}\right)^{-1/2}$$

where $N(\alpha)$ is the appropriate quantile from the standard normal distribution.

To compute the second derivative numerically, use

$$\left.\frac{d^2 f}{dp^2}\right|_{p=m} \approx \frac{f(m+2\Delta p) - 2f(m+\Delta p) + f(m)}{(\Delta p)^2}$$

Estimating uncertainties: Quadratic approximation

For multi-parameter models, the same idea applies, but we need to work with the matrix of second derivatives (the Hessian).

$$\begin{pmatrix} \frac{\partial^2 L}{\partial \mu^2} & \frac{\partial^2 L}{\partial \mu \partial \sigma} \\ \frac{\partial^2 L}{\partial \mu \partial \sigma} & \frac{\partial^2 L}{\partial \sigma^2}. \end{pmatrix}$$

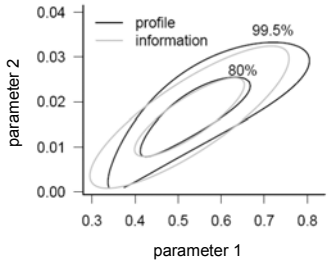
(For stats buffs: the Hessian matrix = −Fisher information matrix)

If we evaluate the Hessian at the MLE and invert it, we obtain the variance-covariance matrix for the parameters:

$$\mathbf{V} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

Estimating uncertainties: Quadratic approximation

For multi-parameter models, the same idea applies, but we need to work with the matrix of second derivatives (the Hessian).

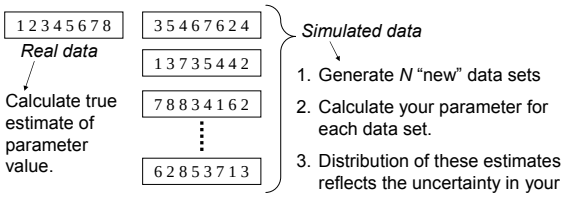


(Figure from Bolker 200X)

Estimating uncertainties: Bootstrapping

Bootstrapping is a completely different approach to estimating uncertainties. It is completely non-parametric (i.e. it doesn't depend on any assumptions about the distribution that underlies your data) and relies on heavy computation.

Basic idea: simulate new data sets by randomly re-sampling with replacement from the observed data.



Model selection

So far we've focused entirely on how to fit the parameters of a single model to data.

But how do we know that we're using the best model?

Or that we need all those parameters?

Luckily, there is a set of formal tools for comparing models in the context of data.

What all of these tools have in common is:

1. Models that fit the data better is preferred.
2. Parsimonious models are preferred (i.e. models are penalized for having more parameters).

The core text in this field is Burnham & Anderson, *Model Selection and Multi-Model Inference.*

Model selection: The likelihood ratio test

Model A is *nested* in model B if it corresponds to some special case of model B where one or more parameters have particular values.

e.g. $f(x) = ax^2+c$ is nested in $g(x) = ax^2+bx+c$ for $b=0$.

An epidemiological example might be whether an additional parameter is justified to describe the possible effect of male circumcision on male→female transmission of HIV.

The likelihood ratio test provides a pair-wise comparison between two models when one is nested within the other.

The test computes a statistic that compares the log-likelihoods calculated from the two models, and determines whether the additional complexity is justified by the data.

Model selection: Akaike information criterion

The Akaike information criterion (AIC) provides a more flexible framework for model selection, that does not require models to be nested and can compare many models at once.

An AIC value can be calculated for any model which has been fit by MLE, and takes the value

$$AIC = -2L + 2k$$

where L is the log-likelihood of the MLE and k is the number of free parameters in the model.

For small sample sizes, i.e. when the number of data points n is such that $n/k < 40$, a corrected AIC should be used:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

Model selection: Akaike information criterion

Many models can be compared by simply comparing their AIC values. The model with the lowest AIC value is preferred.

Because only the relative values of AICs matter, they are often reduced to differences from the lowest value obtained:

$$\Delta AIC_i = AIC_i - AIC_{min}$$

As a rule of thumb:

- models with AIC <2 units apart have roughly equivalent support
- models with AIC 4-7 units apart are clearly distinguishable
- models with AIC >10 units apart are definitely different

Model selection: Akaike weights and model averaging

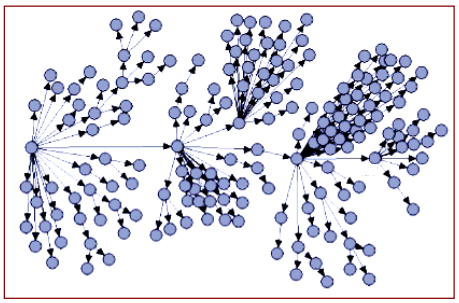
The ΔAIC values can also be used to calculate the Akaike weight associated with each model *i*,

$$w_i = \frac{e^{-\Delta AIC_i/2}}{\sum_j e^{-\Delta AIC_j/2}}$$

These weights can be used for model averaging, i.e. to generate an "average" output from several models that is weighted by the support for each model from the data.

Example: Parameter estimation and model selection

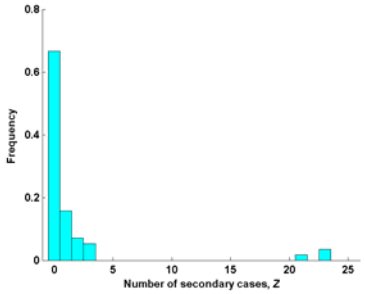
SARS transmission chain, Singapore 2003



Morbidity & Mortality Weekly Report (2003)

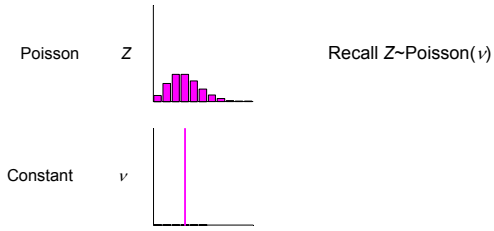
Example: parameter estimation and model selection

Observed offspring distribution, SARS in Singapore

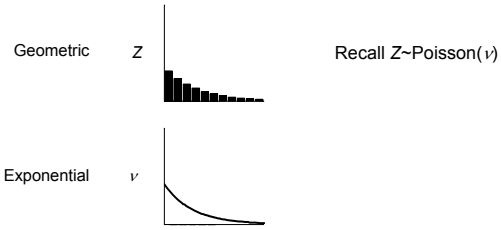


For details on this example, see Lloyd-Smith et al (2005) Nature 438: 355-359 (and especially the online Supplementary Information).

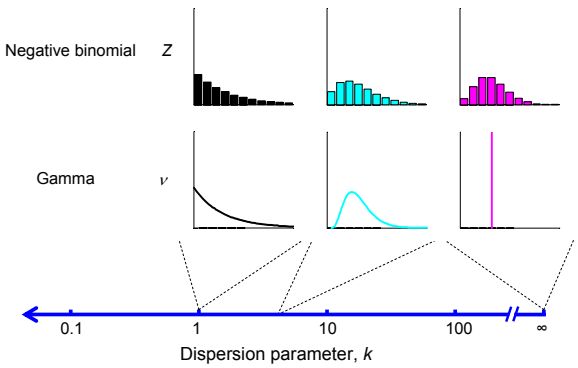
Candidate model 1 Completely homogeneous population, all $\nu = R_0$

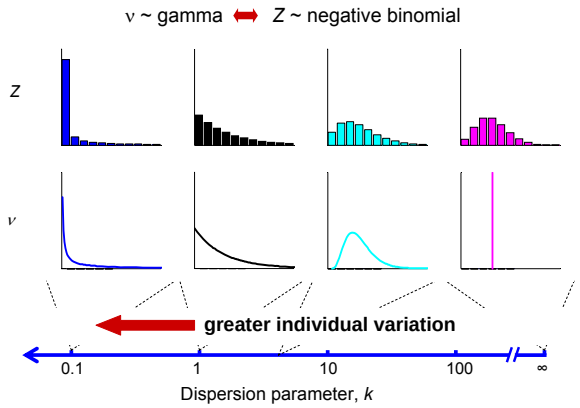


Candidate model 2 (ODE models)
Constant recovery rate, homogeneous transmission

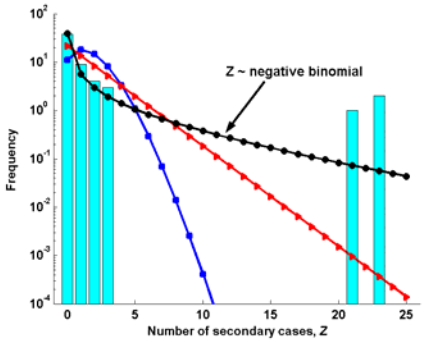


Candidate model 3
A more flexible family of distributions





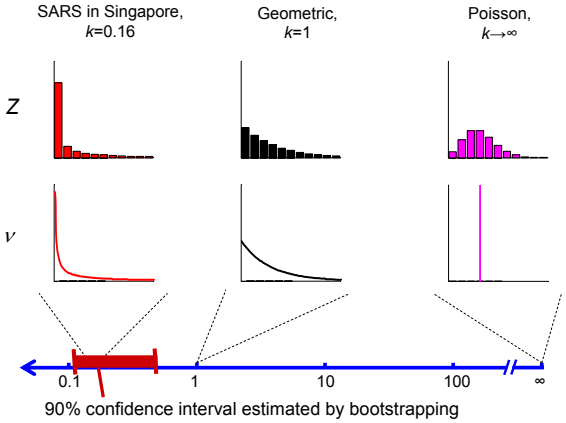
Singapore SARS outbreak, 2003



Singapore SARS outbreak, 2003

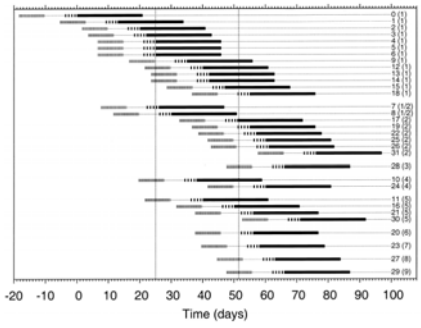
ν Distribution	Z distribution	ΔAIC_c	Akaike weight
ν ~ constant	Z ~ Poisson	250.4	< 0.0001
ν ~ exponential	Z ~ Geometric	41.2	< 0.0001
ν ~ gamma	Z ~ Negative binomial	0	>0.9999

Model selection strongly favours negative binomial distribution.



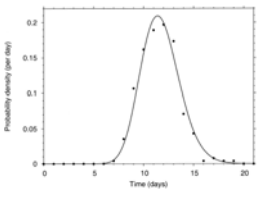
Example: MLE parameters for an observed outbreak

Smallpox outbreak in Abakaliki, Nigeria in 1967 (32 cases)

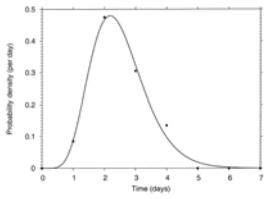


Eichner & Dietz (2003) Am J Epi 158: 110-117

Example: MLE parameters for an observed outbreak



Distribution of period from infection to onset of fever (gamma distribution)



Distribution of period from onset of fever to onset of disease (gamma distribution)

TABLE 3. Durations of the period from infection to onset of fever, from onset of fever to onset of rash, and from onset of rash to recovery, respectively, Abakaliki, Nigeria, 1967

	Mean (days)	Standard deviation (days)	Percentile 5%	Percentile 95%	Coefficient of variation (%)
Period before fever	μ_I = 11.6	σ_I = 1.90	8.62	14.9	16.4
Period from fever to rash	μ_F = 2.49	σ_F = 0.88	1.32	3.8	35.3
Period with rash	μ_R = 16.0	σ_R = 2.83	11.70	20.9	17.7
Duration until isolation	μ_Q = 2.0	σ_Q = 2.00	0.10	6.0	100.0

Example: MLE parameters for an observed outbreak

Likelihood of an individual being infected was modelled as an exponential process with time-varying force of infection:

$$L_{\text{case}}(t_k, c_k, f_k, s_k) = (1-v)^{\delta_v(k)} \int_{t_{k-1}}^{t_k} \lambda(t, c_k, f_k) e^{-\int_{-\infty}^{t} \lambda(\tau, c_k, f_k) d\tau} I(t_k, t) dt$$

TABLE 4. Parameter estimates and 95% confidence intervals (based on the profile likelihood with $\ln(L_{\max}) = -332.08$) for a smallpox epidemic in Abakaliki, Nigeria, in 1967

		Estimate	95% confidence interval
Vaccine efficacy	v	0.816	0.644, 0.922
Contacts in compound (per day)	κ_h	0.335	0.192, 0.527
Contacts in Faith Tabernacle Church (per day)	κ_f	0.0562	0.0187, 0.127
Contacts in total population (per day)	κ_e	0.0281	0.00447, 0.101
Factor for early infectivity	b	0.157	0, 1.89
Onset of isolation measures (days)	t_0	51.5	44.7, 59.6
Basic reproduction number	$R_0 = (\beta_H + b\beta_H)(\kappa_h + \kappa_f + \kappa_e)$	6.87	4.52, 10.1
Reproduction number before rash	$R_P = b\beta_P(\kappa_h + \kappa_f + \kappa_e)$	0.164	0, 1.31
Fraction of compound contacts	$H = \kappa_h/(\kappa_h + \kappa_f + \kappa_e)$	0.799	0.636, 0.879
Fraction of close contacts	$C = (\kappa_h + \kappa_f)/(\kappa_h + \kappa_f + \kappa_e)$	0.933	0.806, 0.988

Estimated all relevant parameter values (R_0, vaccine efficacy, contact rates, etc) and used profile likelihood to estimate 95% confidence intervals.

Sensitivity and uncertainty analysis

Uncertainty arises from two main sources in epidemic models:

- parameter values: often unknown or imprecise
- model structure: does it capture the right mechanisms?

Uncertainty analysis aims to assess the variability in model outputs that arises from uncertainty in model inputs.

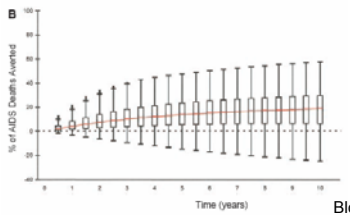
Sensitivity analysis extends this to determine which parameters (or changes in model structure) are most important in determining the model output, and to quantify the influence of each parameter on particular outputs.

Why do uncertainty analysis?

- Determine how much confidence should be placed in quantitative projections generated by models.

  What "error bars" should be placed on output quantities?

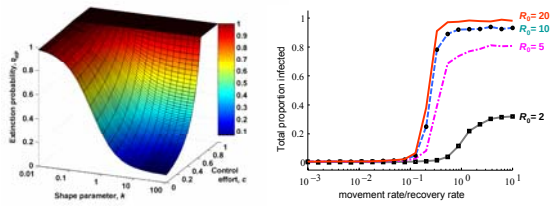
- Understand whether differences between model outputs (or between model outputs and data) are significant.



Blower et al (2000)

Why do sensitivity analysis?

- Understand the relative importance of different mechanisms in generating observed patterns.
- Determine which points in the system are good targets for intervention efforts.
- Guide collection of further data – gather more information on those parameters that are most influential.



Sensitivity and elasticity: formal definitions

The **sensitivity** of outcome λ to the value of parameter θ is

$$S = \frac{\partial \lambda}{\partial \theta}$$

A *partial* derivative since we're holding all other parameters constant.

But parameters are measured on many different scales, making sensitivity values difficult to compare.

Elasticity is the proportional response to a proportional perturbation.

The elasticity of outcome λ to the value of parameter θ is

$$E = \frac{\theta}{\lambda}\frac{\partial \lambda}{\partial \theta} = \frac{\partial \log \lambda}{\partial \log \theta}$$

Structural sensitivity

Structural sensitivity describes how changes in the design of a model influence its output.

There are many subjective decisions, and many assumptions, involved in making a model – but very few studies take the time to test these explicitly and show their effect on model predictions.

Why not?

a) it's a lot of work!

b) lack of established methods

But they should, because we don't want our assumptions to bias the conclusions we draw from our models.

The ultimate structural sensitivity analysis is to have several independent groups of researchers work on the same problem.

e.g. avian flu emergence, foot and mouth disease in the UK, SARS

Parameter sensitivity: Formal approaches

Formal methods exist to investigate the influence of parameter values on model outputs.

Univariate: vary one parameter while holding all others constant.

Bivariate: vary two parameters while holding all others constant.

Complete: vary all parameters at once.

- Full-factorial design: use every value of every parameter, and examine output from every possible combination.
- Efficient sampling design: use fewer parameter values, chosen carefully to avoid bias.

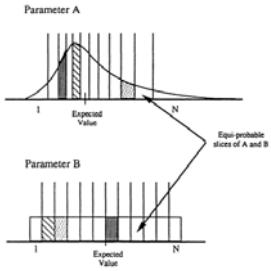
x	x	x	x
x	x	x	x
x	x	x	x
x	x	x	x

		x	
			x
x			
	x		

Latin Hypercube Sampling (LHS) is a popular approach for epidemic models.

Latin Hypercube Sampling (Blower & Dowlatabadi, 1994)

1. Define probability distribution functions for parameters based on uncertainty.
2. Calculate the necessary number of simulations ($N > 4/3\ K$, where N is the number of simulations and K the number of parameters).
3. Divide the range of each parameter into N equi-probable intervals.
4. Create a LHS table of parameter sets to simulate.



X	X	X	X
X	X	X	X
X	X	X	X
X	X	X	X

		X	
			X
X			
	X		

Latin Hypercube Sampling (Blower & Dowlatabadi, 1994)

5. Perform the N simulations using parameters from the tables. Collect the values of model outputs of interest.
6. Uncertainty analysis: the distribution of values of model outputs will be representative of the range of probable outcomes for the parameter distributions chosen.
7. Sensitivity analysis: use Partial Rank Correlation Coefficients (PRCC) to establish the statistical relationship between each parameter and the model output.

PRCCs measure the degree of correlation between one parameter and one model output, while keeping all other parameter values fixed.

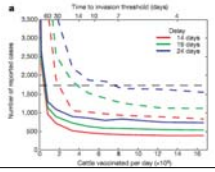
Software for LHS (and other sensitivity/uncertainty analyses) is free online at http://simlab.jrc.cec.eu.int/

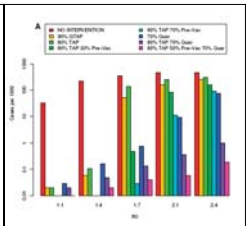
Parameter sensitivity: *ad hoc* approaches

In practice, many (most?) disease modelling studies use less formal approaches to sensitivity analysis.

Common approaches include:

Identify parameters of interest for applied questions (e.g. parameters describing control measures) and perform univariate or bivariate analysis of sensitivity to them.





Construct "scenarios" depicting possible courses of action, and study model outputs that result.