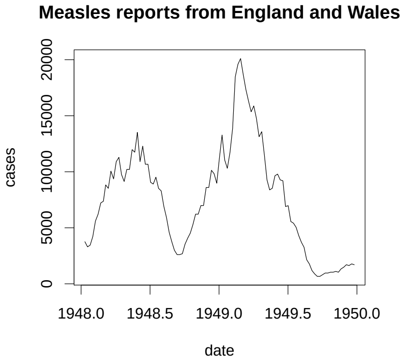
Approaches to dynamic fitting

Jonathan Dushoff, McMaster University
http://lalashan.mcmaster.ca/DushoffLab

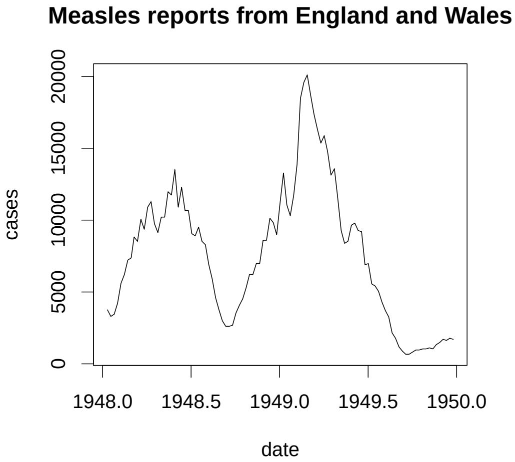
MMED 2016
http://www.ici3d.org/2016/

Measles data

- 

- Reconstruct the number of susceptibles
- Divide the data into generations
- Fit $\mathcal{R}_0$
- Predict

Why did I get the wrong answer?



Why did I get the wrong answer?

- Model structure may be wrong
- Population structure may be wrong
- Stochasticity in disease observation and recording
- Stochasticity in transmission
- Multi-parameter estimation
  - Generation intervals

Outline

Conceptual framework

- How do we assume our data relate to our model world?
  - **No error:** We could attempt to model everything we see, in exact detail
  - **Observation error:** we could assume that the world is perfectly deterministic, but our *observations* are imperfect
  - **Process error:** we could assume that we observe perfectly, but that the world is stochastic
  - **Both kinds of error:** the world is stochastic, and our observations are imperfect

No error

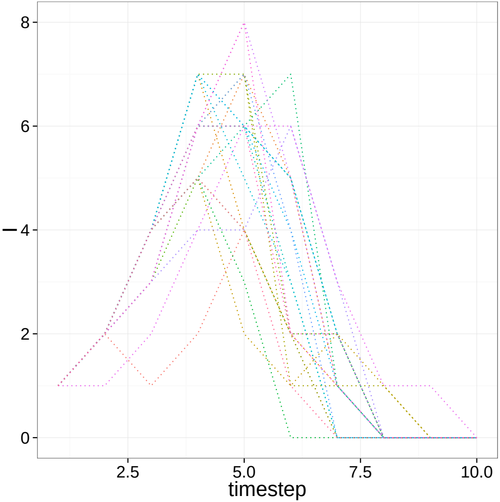
- Impossible
- Even if possible, not clear what we would learn

Observation error only

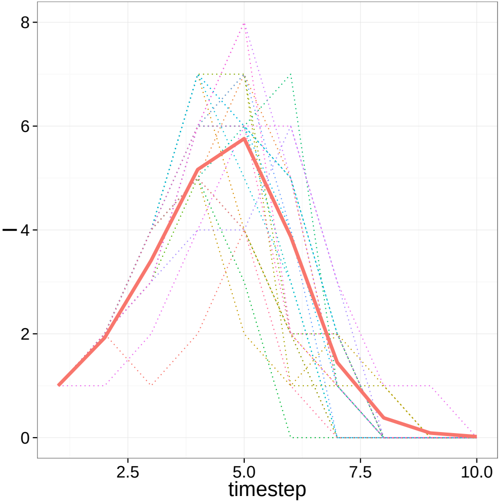
- Point your model at the target
- Give it starting conditions and parameters
- Let it go
- Compare final results to observations

Shooting

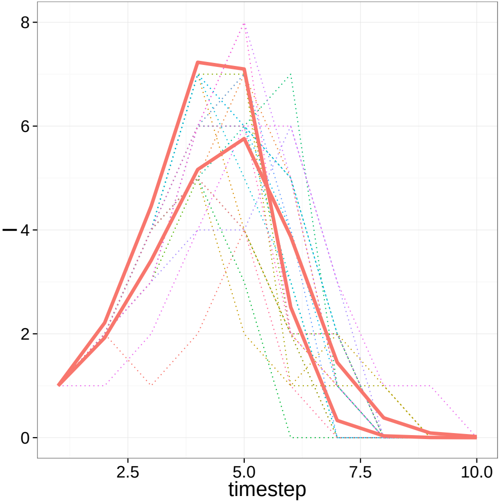
Shooting



Shooting



Shooting

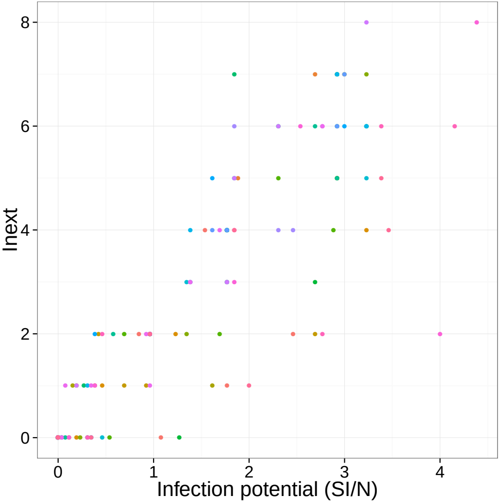


Process error only

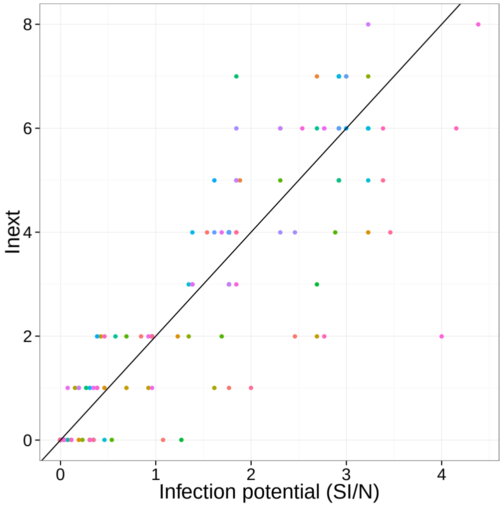
- Look at each step separately.
- See how the model is doing for that step.
- Reset based on observed data before taking the next step

Stepping

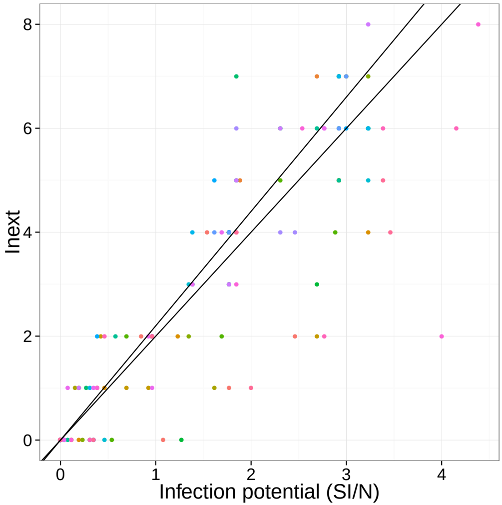
Stepping



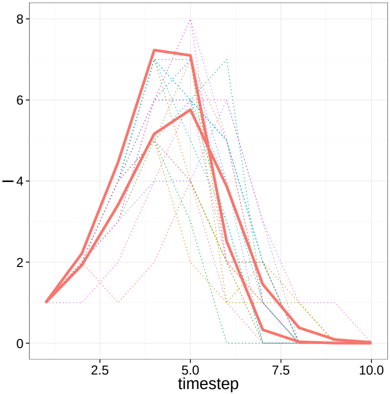
Stepping

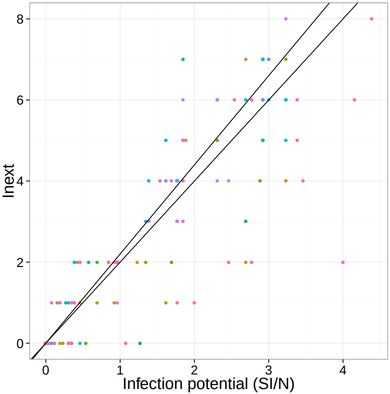


Stepping



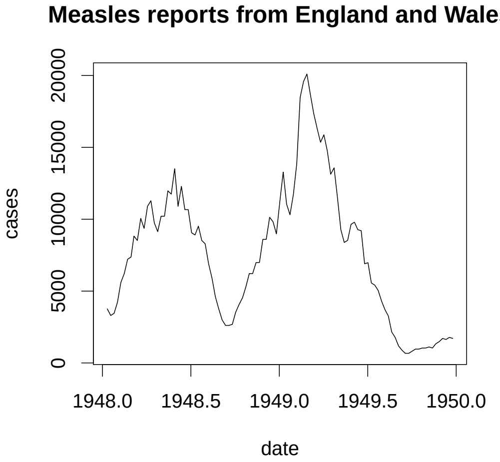
Comparing approaches





Observation and process error

- Latent variable models
  - We need to keep track of, and integrate over, things that we don't observe



Outline

How to fit?

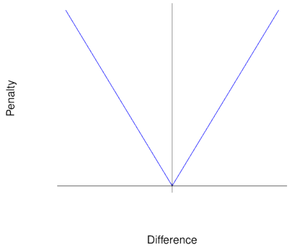
- Solving an equation
- By eye (fiddling with parameters)
- *Minimizing a distance function*
- Likelihood

Distance functions

$$D = \sum_i y_i - \hat{y}_i$$

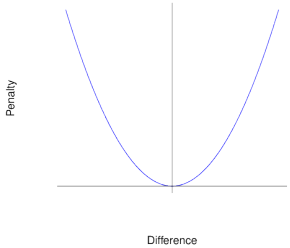
Distance functions

$$D = \sum_i |y_i - \hat{y}_i|$$



Distance functions

$$D = \sum_i (y_i - \hat{y}_i)^2$$



Outline

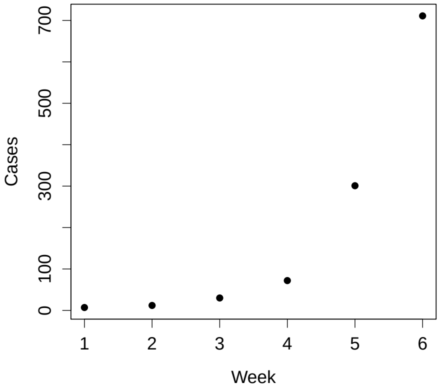
Likelihoods

- Assume that the difference between the estimate $\hat{y}_i$ and the data point y_i is normally distributed. What is the log likelihood?
- $$L = \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(\hat{y}_i - y_i)^2}{2\sigma^2}\right)$$
- $$\ell = \sum_i -\log(\sigma\sqrt{2\pi}) - \sum_i \frac{(\hat{y}_i - y_i)^2}{2\sigma^2}$$
- *We minimize the likelihood by minimizing the sum of squares*
  - and then solving for σ

Least squares $\rightarrow$ likelihood

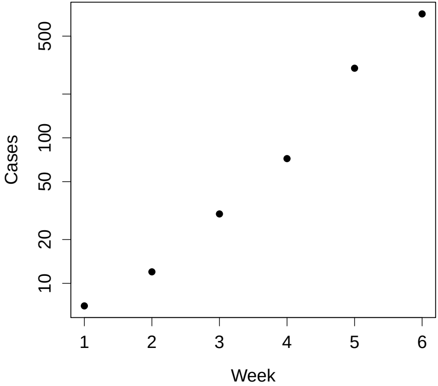
- Attaching your least squares fit to a likelihood means:
  - You can *use it* for statistical inference (LRT)
  - You can *challenge* the assumptions

Mexican flu example



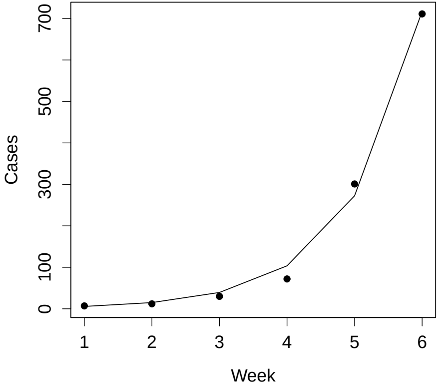
- How fast is it growing? r
- How hard will it be to control? $\mathcal{R}_0$

A different perspective

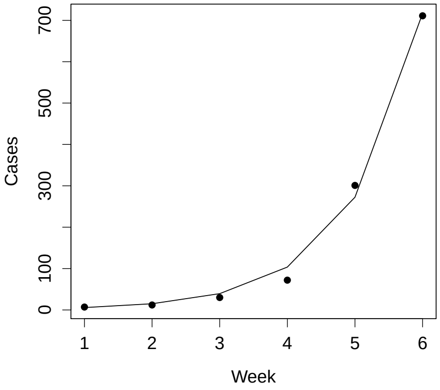


- We could make the normal assumption on either scale
- How much does it matter?

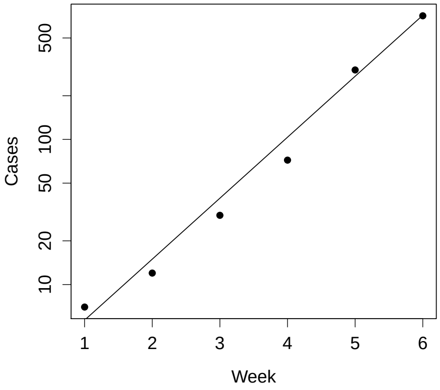
Normal assumption



- Least squares on the linear scale
- 10:50 :: 980:1020
- Gives relatively too much weight to large observations

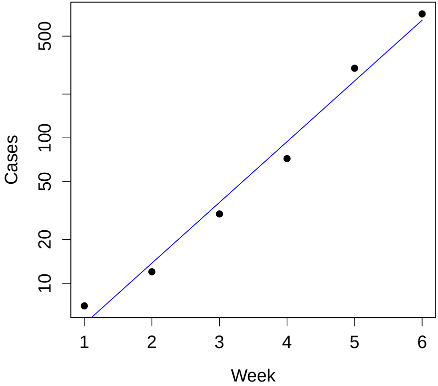


Cases
0
100
300
500
700
1
2
3
4
5
6
Week

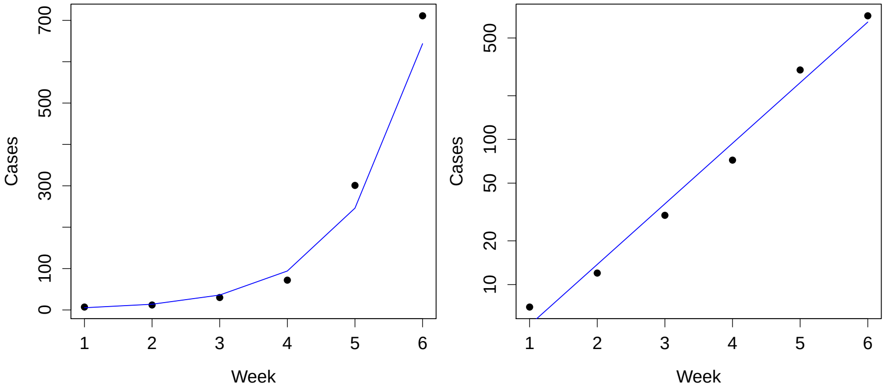


Cases
10
20
50
100
500
1
2
3
4
5
6
Week

Lognormal assumption



- Least squares on the log scale
- 3:5 :: 300:500
- Gives relatively too much weight to small observations


Cases
700
500
300
100
0
1
2
3
4
5
6
Week
Cases
500
100
50
20
10
1
2
3
4
5
6
Week

A more realistic error distribution

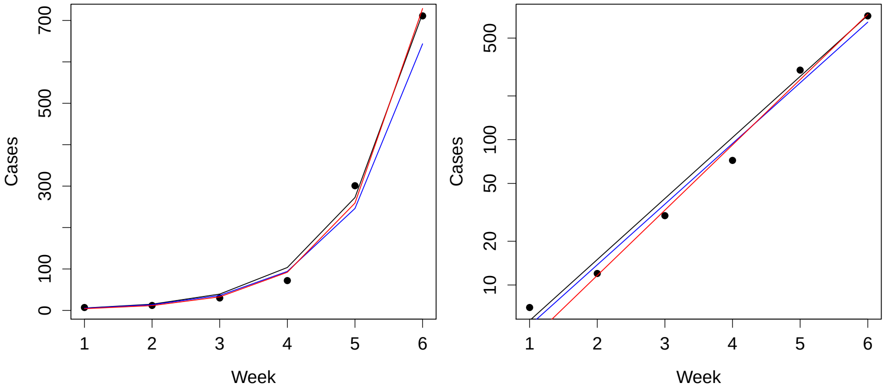
- My case counts are *individuals*
- What distributions can I use to reflect that?
- WRONG!
- *Sorry:*
  - OK, technically it's right, but you shouldn't do it.

Reality is complicated

- Poisson and binomial reflect *only* individual-level variation
  - No temporal variation
  - No clustered sampling
  - …

Distribution diagram

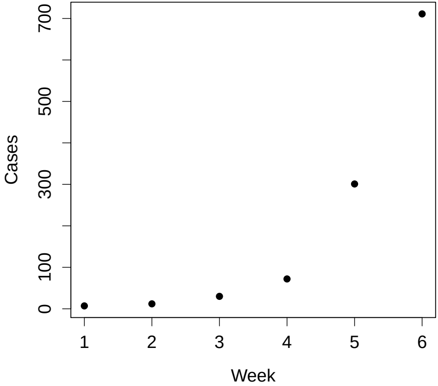
Negative binomial fits



Comparison

- Realistic error distribution provides (apparently) better fits
- Confidence intervals
  - Normal: r = 0.96–0.97/wk
  - Lognormal: r = 0.64–1.29/wk
  - Negative binomial: r = 0.90–1.14/wk
- How would you test these methods?

Identifiability



- What if we tried to estimate $\mathcal{R}_0$ from data like these?

Outline

Modern approaches

- Why are people using model worlds with no observation error?
  - or no process error?
- Sometimes they are good enough (model validation)
- Combining both is *hard*

Filtering

- Filtering is a little like shooting
  - Simulate from beginning to end, but use *stochastic* simulations
- You need a lot of simulations, and often ways of selecting and refining them
- A popular, state-of-the-art method is implemented in the R package pomp

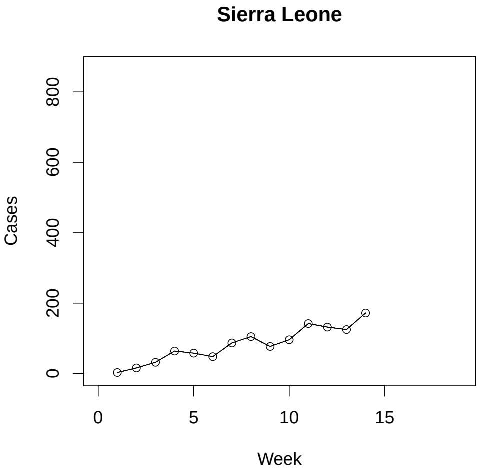
Latent variable methods

- Latent variable methods are a little like stepping
  - But we step to and from unknown values (our latent variables), so we need a way of exploring many possibilities
- Popular, state-of-the-art methods are available in the R packages rjags and rstan

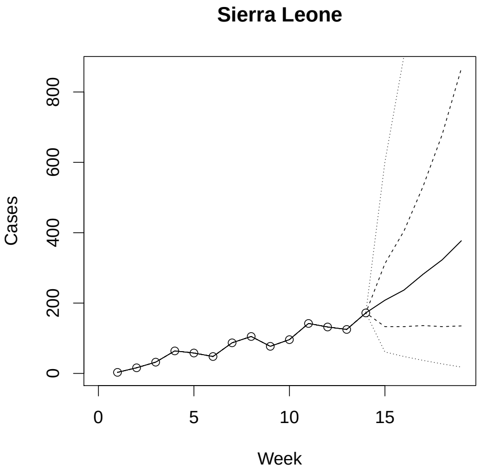
Multi-parameter inference

- Modern methods are already hard, and when you consider various sources of uncertainty, you're really on the bleeding edge
- Many high-profile models for Ebola, for example failed to consider process error.
- The biggest paper talking about process error neglected uncertainty in generation intervals
- Once you do multi-parameter inference, you may find that confidence intervals are very large – this may reflect the reality of knowledge, but may not make you look good

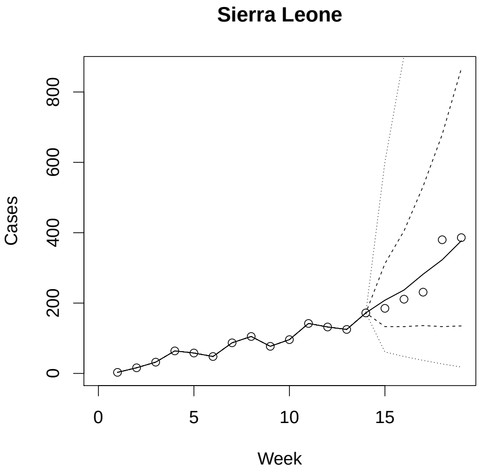
Assessing and reporting uncertainty



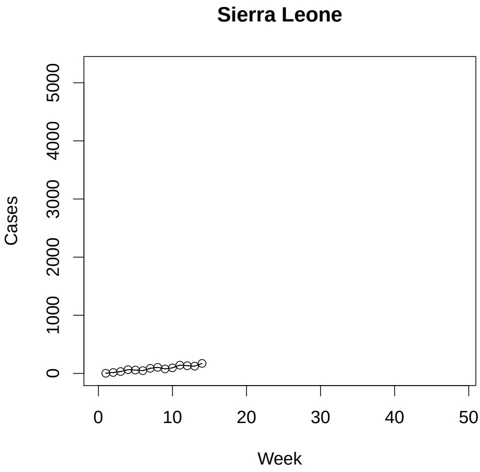
Assessing and reporting uncertainty



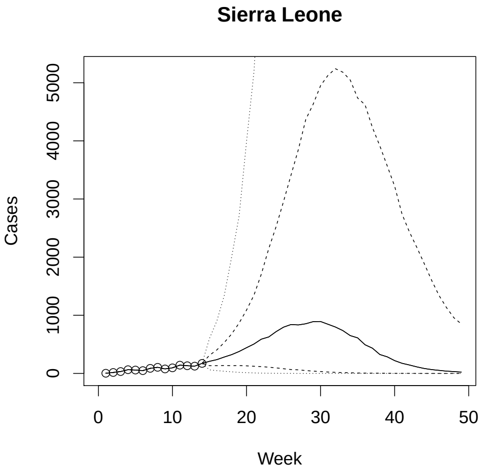
Assessing and reporting uncertainty



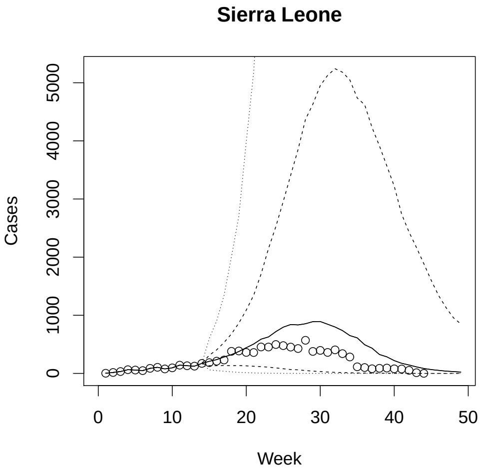
Assessing and reporting uncertainty



Assessing and reporting uncertainty



Assessing and reporting uncertainty



Outline

Likelihood

- Maximum likelihood and likelihood are not the same thing
- Bayesian approaches and frequentist approaches (including maximum likelihood) *both* depend on calculating (or approximating) likelihood

Frequentist inference

- To do frequentist inference on these complicated likelihoods, we need to:
  - estimate likelihoods
  - find the maximum likelihood
  - use the likelihood ratio test to find confidence intervals
- This is hard

Bayesian inference

- To do Bayesian inference on these complicated likelihoods, we need to:
  - construct prior distributions
  - estimate likelihoods
  - estimate the posterior
- Also hard, but sometimes easier than the frequentist approach