Fitting models

Forrest W. Crawford

Calibration and fitting

Calibration and fitting

Fitting deterministic models by least squares

Consider the general "loss" function

$$L(\theta) = \sum_{i=1}^{n} (y(t_i|\theta) - \hat{y}_i)^2$$

This is a sum of differences between the model output $y(t_i|\theta)$ and the observation $\hat{y}_i$, squared. Why squared? Because squaring this difference makes the result positive, and gives larger values to larger deviations.

We want to make this loss function $L(\theta)$ as small as possible.

Example

Example: Suppose we have a SI system characterized by

$$\frac{dI}{dt} = \beta S(t) I(t)$$

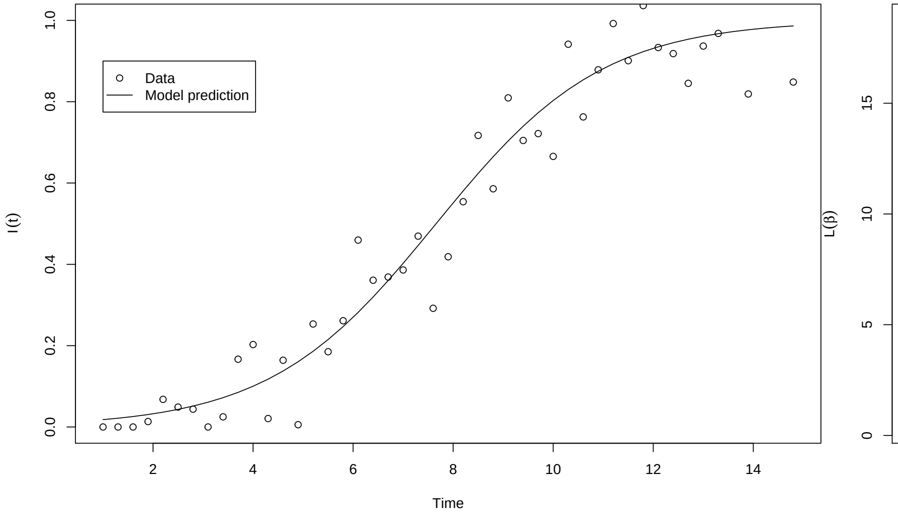
and $N = S(t) + I(t)$ is known. We observe the number of infectives $I(t)$ at times $0, t_1, \ldots, t_n$. Call these values $\hat{I}_1, \ldots, \hat{I}_n$. We wish to estimate β. The loss function is

$$L(\beta) = \sum_{j=1}^{n} (\hat{I}_j - I(t_j, \beta))^2$$

where $I(t_j, \beta)$ is the values of $I(t)$ at t_j using parameter value β obtained by numerical solution. The value of β that minimizes $L(\beta)$ is our estimate of β:

$$\beta = \underset{\beta}{\operatorname{argmin}}\ L(\beta)$$

the first thing to notice is that this loss function is a sum of convex functions [define]. This means it has a unique minimum on a closed set. This means it is “U”-shaped, and we want to find the value of θ at the “bottom” of the “U”.



Data
Model prediction
I(t)
Time
L(β)

Fitting deterministic models with statistical error distributions

Deterministic models produce a single trajectory for model compartments. But real data are noisy and bumpy. We can hypothesize a model for how error enters the model output. For example, some researchers assume that the observed count is a Poisson random variable whose mean is the true compartment value in the deterministic model. For the I compartment,

$$\hat{I}(t) \sim \text{Poisson}(I(t, \beta))$$

Other error distributions can be used as well:

$$\hat{I}(t) = I(t) + \epsilon_t$$

where

$$\epsilon_t \sim \text{Normal}(0, \sigma_t^2)$$

In these cases, we can use the implied likelihood of the statistical error distribution as our (negative) loss function.

If the errors are assumed independent, then the likelihood is a product over the observations:

$$\text{Likelihood}(\theta) = \prod_{i=1}^{n} f(\hat{y}_i|\theta)$$

Fitting stochastic compartmental models using likelihood methods

How do we fit? We want to minimize these functions with respect to the unknown parameters θ.

When the loss is convex, or the (log-)likelihood is concave, there is a unique value of θ tha minimizes (maximizes) the objective function.

When θ has high dimension, this can be difficult to find even with a computer, but when it's low-dimensional, there are some easy solutions.

nlm

optim

sampling-based solutions

Not covered: Bayesian methods

Crash course in gradient descent and Newton-Raphson iteration

Alternatively just use optim or nlm.

References