Fitting the SIR model to data

Ben Shaby

5/18/2010

A review of the SIR model

$$\frac{dS}{dt} = -\beta IS$$
$$\frac{dI}{dt} = \beta IS - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

- Describes the movement between susceptible, infected, and recovered pools.
- β and γ are unknown parameters. Can we estimate them from data?

ODEs in matlab

Look in your folders for a file called `sir_ode.m`. This encodes the model for matlab.

```
function dydt = sir_ode(t,y,p)

  beta = p(1);
  gamma = p(2);

  S = y(1);
  I = y(2);
  R = y(3);

  dydt = [-beta*I*S; beta*I*S - gamma*I; gamma*I];
```

Let's load the data

Let's set up some `matlab` variables so we can play with the SIR system.

```
% Input the data and the time window.  Data is the number
% of infected individuals at each time step.
data = [3 6 25 73 222 294 258 237 191 125 69 27 11 4];
time = 1:14;

% Initial conditions: 760 susceptible, 3 infected,
% 0 recovered
y0 = [760 3 0];

% tspan will hold the times for which to solve the system
% of ODEs.  We want to solve it at the times that
% correspond to our data
tspan = time;
```

The matlab ODE solver

Now we can ask matlab to solve the SIR system for a given set of parameters.

```
% Try solving the system at a guess for good parameter value
p0 = [.01 .1];
[t,y] = ode45(@sir_ode,tspan,y0,[],p0);
```

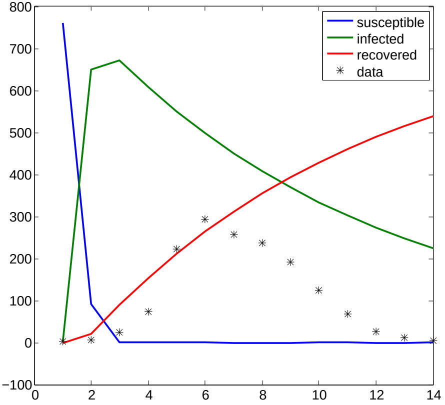
What is the result that matlab passes back (stored in the variables y and t)?

Plotting the solution

Let's plot the data on the same axes as the solution we just obtained for the parameter values $(\beta = .01, \gamma = .1)$.

```
figure(1), clf;
plot(t, y, 'linewidth', 2);
hold on;
plot(time, data, 'k*', 'markersize', 10);
legend('susceptible','infected','recovered','data')
set(gca, 'FontSize', 15)
```

Plotting the solution



Do these parameter values seem to match the data well?

How do we know how well we're doing?

Can you think of a good way to quantify how well the model fits the data, given the parameters you've chosen?

Can you think of a good way to quantify how well the model fits the data, given the parameters you've chosen?

Your task: Write a function that

- Takes as input `p`, `data`, `tspan`, and `y0`(for some particular value of `p`)
- Solves the SIR system
- Outputs a number that describes the quality of the model fit.

How do we know how well we're doing?

One approach is "sums of squares".

How do we know how well we're doing?

One approach is "sums of squares".

Here's my function:

```
function disc = sir_discrepancy(p, data, tspan, y0)

[t,y] = ode45(@sir_ode,tspan,y0,[],p);
I = y(:,2);
disc = sum((I-data').^2);
```

Try running your descrepancy (aka loss, regret, etc.) function for different values of `p`. How small can you make it?

Numerical optimization of the parameters

- `matlab` can do optimization using the function `fminsearch`.

```
>> help fminsearch
 FMINSEARCH Multidimensional unconstrained nonlinear
 minimization (Nelder-Mead).
    X = FMINSEARCH(FUN,X0) starts at X0 and attempts to
    find a local minimizer X of the function FUN
```

- Our descrepancy function function takes several inputs, but `fminsearch` only optimized functions with one input.
- Can you figure out a way around this by using the `matlab` help functions?

Numerical optimization of the parameters

My solution:

```
function p_opt = sir_optimize(data, tspan, y0, p0)

p_opt = fminsearch(@sir_disc_nested, p0);
    function disc = sir_disc_nested(p)
      disc = sir_discrepancy(p, data, tspan, y0);
    end
end
```

Now we can run

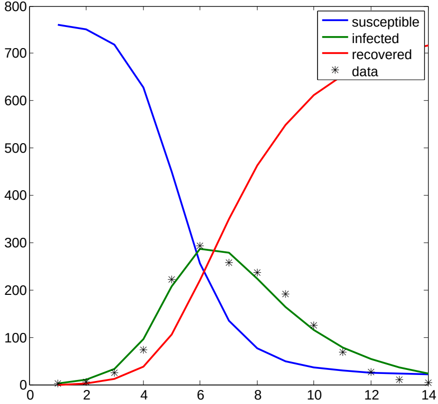
```
>> p_opt = sir_optimize(data, tspan, y0, p0);
```

Numerical optimization of the parameters

- What were your optimal values of β and γ?
- Now use the ODE solver to solve the SIR system with the optimal parameter values. Plot the results.

Numerical optimization of the parameters

- What were your optimal values of β and γ?
- Now use the ODE solver to solve the SIR system with the optimal parameter values. Plot the results.



Better than before?