Essential pre-requisite (August 5, 2015)

How to activate solver function in Excel?

Hiroshi Nishiura

1. Excel 2003 or older

We will use ‘solver’ function of Microsoft Excel to find optimal solutions. Select the “solver” option from the “Tools” menu if it’s available. If it’s not available, select the “add-ins” option in the Tools menu and click on the “Solver add-in” option, click on OK and then select the “solver” option.

The “Solver” option enables you to minimize a function you specify and find some estimate of the force of infection. It requires you to specify

a) the location of the cell to minimize

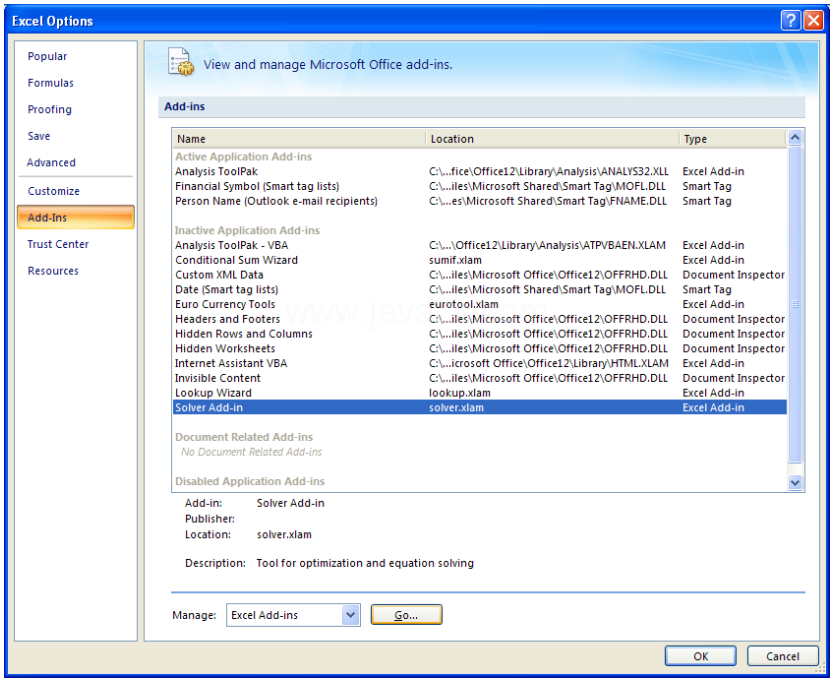
b) the cells which you want to change

In this way, please set up the “Target cells” and the “By changing cells” options. Select the “Min” option under the “Equal to” option and click on the “Solve” button.

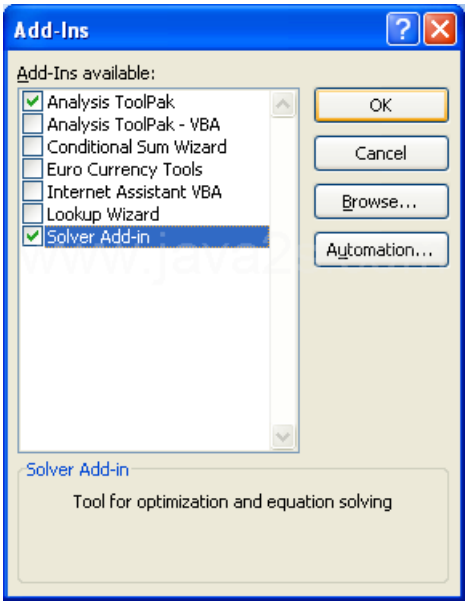
2. Excel 2007 and later

Click on the Office Button, the big round decoration in the top left of the Excel window. Then click the Excel Option at the bottom. Alternatively, you click ‘File’ tab and then click ‘Options’ which is second from the bottom on your left.

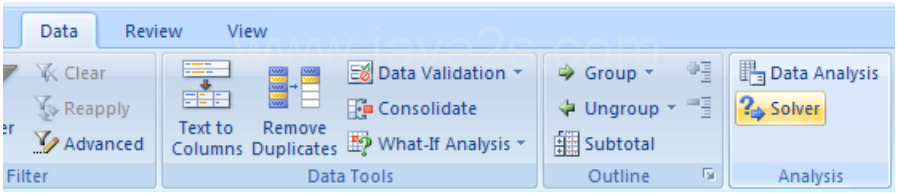
Please select ‘Add-Ins’ on your left. At the bottom, select ‘Excel Add-ins’ and click ‘Go’.



Select Solver add-in in the list. Click OK.



Click OK to install Solver. Solver button is added to Data tab.



The “Solver” option enables you to minimize a function you specify and find some estimate of the force of infection. It requires you to specify

a) the location of the cell to minimize

b) the cells which you want to change

In this way, please set up the “Target cells” and the “By changing cells” options.

Select the “Min” option under the “Equal to” option and click on the “Solve” button.

References and useful external links:

http://peltiertech.com/WordPress/installing-an-add-in-in-excel-2007/

http://www.java2s.com/Tutorial/Microsoft-Office-Excel-2007/0200__Data-Analysis/AddInstallSolverAddin.htm

http://ulearnoffice.com/excel/analysis.htm

[Practical] Estimating transmission potential of an infectious disease (August 5, 2015)

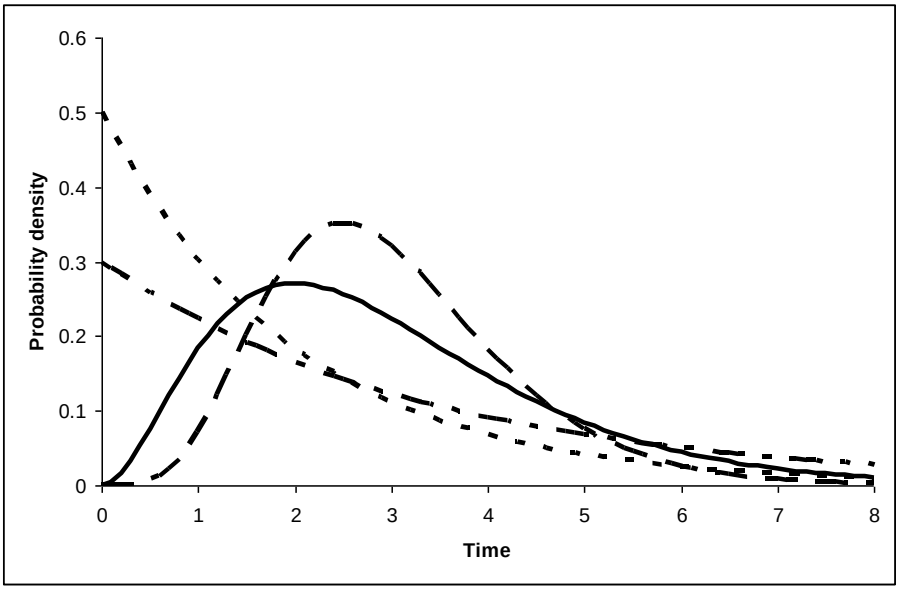
Hiroshi Nishiura
E-mail: nishiurah@m.u-tokyo.ac.jp

1. Estimation of R_0 from exponential growth of cases

Theory

When an epidemic is in its initial phase, the number of infected individuals (prevalence) increases (more or less) exponentially, as does the number of new cases per unit of time (incidence). Because during the exponential growth phase the proportion of susceptibles in the population remains approximately constant ($S \approx N \Leftrightarrow s \approx 1$, if everyone was susceptible, as was the case with SARS), it is possible to estimate the exponential growth rate r from incidence data, and then use the exponential growth rate r to estimate R_0.

What we need to link r and R_0 is the mean *generation time* T_g, which is the mean time at which an infected person infects new cases, measured from the moment she was infected herself. Besides the mean generation time, also the distribution determines how r and R_0 are related. Examples of generation time distributions may be:

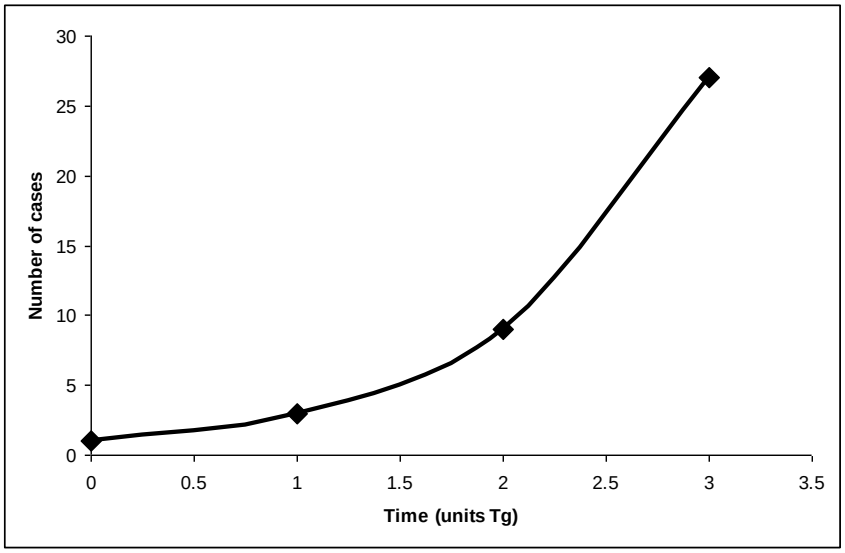


It may be clear that if the generation time distribution is complicated, it is more difficult to estimate R_0 from the exponential growth rate r. There are two relatively simple models to link real time growth (r) and growth in generations (R_0), and the first one is derived from the *SIR* model with exponentially distributed infectious period.

For that model, the mean generation time is equal to the mean infectious period, so the relation between r and R_0 is:

$$\left.\begin{aligned} R_0 &= \frac{\beta}{\gamma} \\ r &= \beta - \gamma \end{aligned}\right\} \Rightarrow R_0 = 1 + \frac{r}{\gamma} = 1 + rT_g$$

For the second one the assumption is made that the generation time T_g is fixed, which means that every infected person infects new cases exactly T_g days (weeks,…) after she was infected herself:



In that case, at $t = T_g$ there are R_0 times as many cases as at $t = 0$, so that:

$$R_0 = e^{rT_g}$$

(Exercise A) The dataset and analysis

In the Excel file '5Aug2015_1.xls' there is a dataset, derived from the Hong Kong SARS epidemic, but changed by some random alterations. The dataset contains the days of symptom onset for 655 cases that were infected until the 55th day of the epidemic, and that were not linked to some particular clusters, so that mixing can be assumed to have been more or less random.

(A-1) make a graph of the epidemic curve up to day 55 (time vs incidence).

(A-2) determine by visual inspection when the exponential growth phase ends

(A-3) use the data of the exponential growth phase to estimate the daily growth rate r

(A-4) the mean generation time for SARS is in the range 10-16 days. What is R_0?

2. Estimation of R_0 from "final size" of an epidemic

Theory

The final size of an epidemic represents the fraction of the population that has experienced infection by the end of an epidemic. Comparing final sizes of 30% and 90%, we feel that a disease with 90% has the higher transmissibility than that with 30%. Indeed, the final size is directly relevant to the transmission potential of an infectious disease, and is determined by the next-generation matrix and also by R_0. In a homogeneously mixing population, we have the so-called final size equation:

$$1 - z = \exp(-zR_0) \quad (1)$$

where z is the final size ($0 < z < 1$).

(Exercise B) The dataset and analysis

In the Excel file '5Aug2015_2.xls' there is a dataset, derived from an epidemic of influenza A virus (H3N8) among racehorses in Japan.

(B-1) There are epidemic data for 5 different racecourses. Please calculate the final size and its corresponding 95% confidence interval by means of normal approximation to a binomial.

[hint] Supposing that the observed proportion infected (i.e. sample proportion) is p, the approximate confidence interval is given by

$$p \pm 1.96\sqrt{\frac{p(1-p)}{N}}$$

(B-2) Using the final size, please estimate R_0 of equine influenza for each racecourse.

3. Estimation of R_0 from age-stratified "final size" of an epidemic

Theory

In a heterogeneously mixing population, final size of type i host (e.g. final size of age-group i) is described by

$$1 - z_i = \exp(-\sum_j z_j R_{ij}) \quad (2)$$

where R_{ij} is an element of the next-generation matrix, representing the average number of secondary cases arising in host i caused by an average infectious host j. In other words, if we have final size data in our hand, we can estimate transmission potential using equations (1) and (2). We use these relationships to estimate the next-generation matrix as well as R_0.

The next-generation matrix is

$$K = \begin{pmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{pmatrix} \quad (3)$$

and R_0 is given by the largest (dominant) eigenvalue of K.

(Exercise C) The dataset and analysis

Please open the file "5Aug2015_3.xls" and look at the sheet "data". In the Excel file there is a dataset from an epidemic of seasonal influenza in a rural population in a country in Southeast Asia.

(C-1) Please calculate the final size for each age-group as well as the total population.

Please move on to the sheet "calc". Using this sheet, we are going to estimate the next-generation matrix K and R_0.

(C-2) Please check that cells B2 and B3 contain the final size of two age-groups.

(C-3) Please check the formulae in cells C2 and C3. These are the right-hand side of the final size equation (1). Two parameters to be estimated are cells G2 and G3, and the next-generation matrix is assumed to be given by

$$K = \begin{pmatrix} aa & ab \\ ba & bb \end{pmatrix} \quad (4)$$

which can be seen from cells C8, C9, D8 and D9.

(C-4) Cells D2 and D3 are the sum of squared errors. And D5 is the sum of D2 and D3, which we are going to minimize. Select solver function, select D5 as the cell to minimize, and select G2 and G3 as the parameters to vary. What are the estimates of G2 and G3?

(C-5) Now, from cells C8, C9, D8 and D9, you see the next-generation matrix. Please estimate R0.

(C-6) Rather than the assumption (3), one may assume

$$K = \begin{pmatrix} a & b \\ b & b \end{pmatrix} \quad (5)$$

as the next-generation matrix. Following this assumption, please first revise the equations in C2 and C3. Then, please revise the next-generation matrix in cells C8, C9, D8 and D9. Please estimate parameters a and b. Also, what's the estimate of R0?

[Practical] Estimating transmission potential of an infectious disease (August 5, 2015)

(Exercise A) A case study of exponential growth of SARS

(A-1) make a graph of the epidemic curve up to day 55 (time vs incidence).

This is done by making a column time and a column incidence, eg in columns C and D. The column time runs from 1 to 55, and for the column incidence you use the formula COUNTIF. Then you make a plot by selecting the two columns and using the graph wizard (use XY (scatter)).

(A-2) determine by visual inspection when the exponential growth phase ends

This can best be done by changing the Y-axis to a logarithmic axis (log incidence). During the exponential growth phase, in a logarithmic figure the graph will be approximately linear. For our data this seems to be until day 32 or 33.

(A-3) use the data of the exponential growth phase to estimate the daily growth rate *r*

This is done is five steps:

Step 1. Choose a single cell (eg cell H1) to contain the value for *r*, and enter an initial value, e.g. 0.1 in that cell. Also choose a single cell (cell H2) to contain the value for the expected incidence on day 0, and enter an initial value, e.g. 0.5 in that cell.

Step 2. Enter values for the expected incidence in a third column (column E), containing the following formula: “=H2*EXP(H1*C2)” in cell E2, and so on for all 33 cells of the exponential phase.

Step 3. Enter the individual contributions to the log-likelihood in a fourth column (F), containing the following formula: “=-E2+D2*LN(E2)” in cell F2, and so on for all 33 cells of the exponential phase.

Step 4. Choose a single cell (eg cell H4) to contain the log-likelihood, by entering the following formula: “=SUM(F:F)”.

Step 5. Maximize the log-likelihood by using the solver functionality. Select

Solver from the Tools menu. In the window that appears select as Target Cell the cell containing the log-likelihood (cell H4). Then select Equal to “max”. Finally select the cells containing the values for r and inc(0) in the part “By changing cells”. Then press Solve.
This results in an estimate of r = 0.142.

(A-4) the mean generation time for SARS is in the range 10-16 days. What is R0?

If the generation time is exponentially distributed, R0 ranges from 1 + 0.142*10 = 2.4 to 1 + 0.142*16 = 3.3
If the generation time is the same for everyone, R0 ranges from e0.142*10 = 4.1 to e0.142*16 = 9.7.
Because for SARS the generation time is highly variable, the first estimates are probably better.)

(Exercise B) A case study of final size of equine influenza
(B-1) There are epidemic data for 5 different racecourses. Please calculate the final size and its corresponding 95% confidence interval by means of normal approximation to a binomial.

The final size reads as follows:
Niigata 90.6% (95% CI: 88.4, 92.9)
Haruki 81.9% (95% CI: 79.2, 84.7)
Fukuyama 97.5% (95% CI: 96.3, 98.7)
Kawasaki 91.5% (95% CI: 89.4, 93.6)
Urawa 77.7% (95%CI: 74.5, 80.9)

(B-2) Using the final size, please estimate R0 of equine influenza for each racecourse.

R0 is estimated as follows:
Niigata 2.6
Haruki 2.1
Fukuyama 3.8
Kawasaki 2.7
Urawa 1.9

(Exercise C) A case study of final size of seasonal influenza

(C-1) Please calculate the final size for each age-group.

Total population 33.7%
Children 62.0%
Adults 25.5%

(C-4, C-5 and C-6) Cells D2 and D3 are the sum of squared errors. And D5 is the sum of D2 and D3, which we are going to minimize. Select solver function, select D5 as the cell to minimize, and select G2 and G3 as the parameters to vary. What are the estimates of G2 and G3?

a = 1.177927, b = 0.358156. Accordingly, the next-generation matrix reads

$K = \begin{pmatrix} 1.39 & 0.42 \\ 0.42 & 0.13 \end{pmatrix}$. R0 is estimated as 1.52.

Wrt an alternative assumption in C-6, we get a = 1.422683 and b = 0.336338.

Accordingly, the next-generation matrix reads $K = \begin{pmatrix} 1.42 & 0.34 \\ 0.34 & 0.34 \end{pmatrix}$. R0 is

estimated at 1.52.