Instructions for the program “Outbreak Detection P”

About the program

The program “Outbreak Detection” computes a non-parametric alarm statistic for detection of an outbreak from a constant level to increasing incidences. No baseline estimation is necessary and the method is thus robust. The alarm statistic is based on maximum likelihood estimates of two different regressions (monotonically increasing and a constant level) and theory for optimal surveillance as described in [1-3]. The “P” stands for Poisson which is suitable for low incidences. There is also a “N” version which is suitable for Normal distributions.

The program is written as VBA (Visual Basic for Applications) extensions to an Excel workbook; therefore the input data and results are saved in the same way as any regular Excel-file, and can be saved in multiple copies under different names. No installation is necessary; the files just need to be extracted to a directory with write permissions. It's important to keep the help subdirectory in the same directory. The macro security level in Excel should be set to medium to allow the program to be executed.

Running the program

Start the program by opening the Outbreak DetectionP workbook in Excel. Input data are entered in the Data-sheet, time order and y-value (observations) should be entered.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The program adds a custom menu to the standard menu bar. In Office 2007 this menu will be in the Add-Ins tab. To execute the program select Run Outbreak Detection from the Outbreak Detection menu.

Enter Alarm Limit and Warning Limit in the boxes, as shown below.

Click Execute to start the calculations
Choice of alarm limit

The choice of the alarm limit can be done by examining the results of some limits when applied to a model, historical data or simulated data. Some criteria will now be discussed.

Predictive value

The predictive value is the probability that the outbreak has occurred given that an alarm was triggered. In [2] the warning limit and alarm limit for influenza like illness (ILI) was chosen to 150 and 9000 respectively. The alarm limit for laboratory diagnosed influenza (LDI) in [1] was 5000. A warning limit of 100 is suitable for LDI. The warning and alarm limits gives a predictive value of about 0.95 and 0.99 respectively, for both ILI and LDI. To set the limit to give a fixed predictive value requires that information about the disease is analyzed as in [1] and [2]. For a quick choice of limits for other diagnosis for which less information is available, simpler measures can be used. To set the predictive value you need information on what happens at an outbreak and how often an outbreak can be expected. The following criteria is based only on false alarm properties and do thus not require any information besides on the non epidemic state.

Average time to false alarm

A criteria which only uses information on the distribution of the incidence before the outbreak is the average run length ARL0 or the median run length MRL0. For incidences which are Poisson distributed, the parameter $\lambda$ determines the MRL0. For LDI we used $\lambda=1$ and for ILI we used $\lambda=20$. In the figure below you can find some values of MRL0 for different limits and different values of $\lambda$.

Significance level

A simple criterion is the significance level. This measure is not recommended for surveillance but only for hypotheses testing since in surveillance it is the properties of the system of repeated decisions which is important. Anyhow, it can be used as a rough indicator on the properties of a certain level. The probability of a false alarm at a window of five observations is given in the table below.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Limit 100</th>
<th>Limit 5000</th>
<th>$\lambda$</th>
<th>Limit 150</th>
<th>Limit 9000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0026</td>
<td>&lt; 0.0001</td>
<td>10</td>
<td>0.0049</td>
<td>0.0001</td>
</tr>
<tr>
<td>1</td>
<td>0.0047</td>
<td>&lt; 0.0001</td>
<td>20</td>
<td>0.0047</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.0073</td>
<td>0.0001</td>
<td>30</td>
<td>0.0046</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>0.0096</td>
<td>0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Alarm statistics are displayed for each time point in the *Alarm* sheet as shown below. The time value indicates that the alarm statistic always is calculated from the first time point, i.e. the last alarm statistic below is calculated between the first and the fourteenth time point. If a warning or alarm has occurred it is showed by a 1 in the *Warning* and *Alarm* columns respectively, otherwise a 0 is shown.

The alarm statistic is also shown in the *Alarm Graph* sheet, as shown below. The alarm statistic is plotted against time and lines indicate the alarm and warning limits respectively.
The *Outbreak Regression* graph show the fitted non parametric regression curve and observations. A steep curve indicates that an outbreak has occurred The *Constant* graph shows the observations and the mean. If the process is still in the non outbreak phase the deviations from the mean should be relatively small. The graph below clearly shows that an outbreak has occurred; the early observations are below the constant line (the average of all observations) and the later above.
References:


Contacts

Linus.Schioler@statistics.gu.se
Marianne.Frisen@statistics.gu.se