Exercise 16 – Time series of events. Testing for randomness, patterns, and trends.

Analyze the volcano Aso time series (VolcanoAso.txt) for trends and randomness. The series contains years in which the Japanese volcano Aso erupted. Investigate whether the volcano becomes more or less active with time. Analyze whether the risk of eruption depends on when the previous eruption occurred.

For many of these tasks you can use the code EventsTimSer.m. Make sure that you understand what the code is doing. Before applying the code, you will need to prepare the time variable 'time' and the interval variable 'int' from the original data.

```
% Create time and interval variables
time=VolcanoAso;
d=max(size(time));   %number of points
d1=d-1;                   % number of intervals
int=time(2:d)-time(1:d-1);  % series of intervals between events
event=1:d;            % event number

plot(time,event);    % cumulative plot of the number of events vs. time
```

From the figure menu, select Tools/Basic Fitting. Fit a linear line. (If events occurred at regular intervals, their cumulative plot would have been a straight line. Deviations from linearity therefore indicate periods when events were less frequent than average or more frequent than average. Check the checkbox to plot residuals. Observed different periods of activity.

```
plot(time,0,'o')       % dot-plot of events

nperiods=50;
hist(time,nperiods); % the number of events in each successive time period
```

Notice that events seem to have been occurring more frequently in the recent past. Let's check if this apparent increase is significant. For example, let's plot the last histogram in polytool and verify if there is a significant trend towards increasing the number of eruptions per period:

```
[n nout]=hist(time,nperiods);   % n=number of events, nout = positions of bins
polytool(nout,n)
```

In the figure menu, select Bounds/Curve. This shows the possible bounds for the regression line. Although the graph shows an upward–sloping line, the bounds indicate that a zero-slope line is also possible within 95% confidence interval. So the trend is not significant, at least as far the as the complete series is concerned.

Check if the periods between subsequent eruptions are related:

```
plot(int(1:d-1),int(2:d),'o')           % return map
```

Long intervals between the eruptions seem to be more likely to be followed by a short interval, rather than by another long interval. However, there are too few of the long intervals to draw solid conclusions. Short intervals seem to be uncorrelated.

Let's do another test for randomness. If the eruptions occur at random intervals, the number of eruptions per a fixed interval should follow the Poisson statistics. (Just like the statistics of cosmic particles, or the number of radioactive atoms decaying per second). There is a section of the code in 'EventsTimSer.m' which does that. First, it calculates the number (n) of eruptions per each period of time:

```
n=hist(time,nperiods);
```

and then bins those in a histogram, i.e. it calculates how many times there were 1 eruprion per period, 2 eruptions, 3 eruptions, and so on.

```
nbins=15;
[Observed nout1]=hist(n,nbins);
```

The Observed histogram is then compared to the Expected histogram from a Poisson distribution. The code then calculates chi-square, which is a measure of how the expected matches the observed. The result, unfortunately, does depend on the number of bins, and the number of data points is not terribly large to have a decent statistics.

Investicate how your results are affected by the choice of the number of counting periods (nperiods) and the number of bins in the histogram (nbins).

Can you answer these questions:

- Is the interval between eruptions random?

There does not seem to be any pattern, at least for short intervals.

- Regardless of whether random or not, is the probability of having a long interval the same as for a short interval?

No. To see this, plot the histogram of intervals:

```
hist(int)
```

The histogram shows that short intervals are more likely than long intervals.

- What is the typical (expected) interval between eruptions?

```
mean(int)
```

- If you were to forecast the risk of an eruption occurring within the next 2 years, how would you do that?

The just produced histogram of intervals can be thought of as representing the p.d.f., i.e. it reflects the probability with which the interval of a given length can occur. To obtain a normalized p.d.f., the histogram values need to be divided by the total number of intervals (d1). Suppose, for the sake of concreteness, that an eruption has just occurred. The probability that an eruption will occur within the next 2 years then is equivalent to the probability that the interval will be shorter than 2 years. This probability is represented by the area under the p.d.f. curve, for abscissa (interval) values between 0 and 2.
This approach, of course, assumes that the interval between eruptions is random, i.e. it does not depend on previous eruptions, and that there is no trend, i.e. that the p.d.f. does not change with time. (I have no idea if these approximations are reasonable – what do I know about volcanoes?)

- Is there any trend in the volcano activity? Is the trend significant?

We saw that the trend is not significant.

- Did you have to subtract the trend from the data when analyzing for randomness?

The non-significant trend does not need to be subtracted, although it wouldn't hurt.