Eruptions of volcanoes in Iceland: Katla and Eyjafjallojökull

P.M.E.Altham, University of Cambridge.

May 4, 2010

The following data on the 'Eruptions of the Katla volcanic centre in historical times' was published by G.Larsen in 1993: see

```
www.earthice.hi.is
```

This gives the Duration of the eruption in days, and the Repose time in years: the latter being the time since the previous eruption. Some readings are only approximate, and are correspondingly marked, for example

```
~120
```

The Duration times are not known for eruptions before 1625.
What can you make of this dataset? For example, you might like to try

```
 library(MASS)
fitdistr(Duration, ...)
```

The next Katla eruption was predicted to be before 2010 (see for example the 2007 Lonely Planet Guidebook to Iceland). Comment.
Does the Duration of an eruption increase as the corresponding ReposeTime increases? (We haven't really got enough data to tell?)

```
Year Duration ReposeTime
 1918   24     58
 1860   20     37
 1823   28     68
 1755 ~120     34
1721  ~100     61
1660  ~ 60     35
1625    13     13
1612    NA     32
1580    NA     ~80
~1500   NA
15thC
1416
~1357
1262
1245
~1179
12thC
~934
~920
~900
```

In the past 1000 years Eyjafjallojökull has erupted in the years 920, 1612, 1821 and 2010. Discuss.

Afterword It now comes to me that I have dealt with a version of this problem before, in giving statistical advice to Dr C.S.M.Doake in 1977. The resulting paper is

'Climatic change and geomagnetic field reversals: a statistical correlation' by C.S.M.Doake, published in *Earth and Planetary Science Letters*,**38**(1978) 313–318.
In this paper 18 'magnetic reversal events' are linked to each of 5 climate change indicators, of which 3 are obtained from records of planktonic foraminifera. Thus, for example Magnetic reversal events are known to have occurred at 18 distinct ages (given in Myr) and the first palaeclimatic indicator has 14 different 'events' during the same time period, some being very close in time to the magnetic reversal events.
In the paper, the methods applied are given on pp 247–248 of the 1966 book by D.R.Cox and P.A.W.Lewis 'The statistical analysis of series of events'.

In the notation of the paper, let θ be the probability that a CE (climatic event) causes a MR (magnetic reversal) in an interval of length t, and that other MR's occur randomly. We condition on R, N, the total number of MR, CE events respectively, seen in the total period of time T. Then with n as the total number of MR's which 'coincide' with CE's (in the sense of being in an interval of width t around a CE) our model is

$$n = U + V,$$

where $U \sim Bin(N, \theta)$ (these are the MR's 'caused' by a CE), and
$V|U = u \sim Po(tN(R-u)/T)$ (these are the 'randomly-occurring' MR's).
Hence

$$E(n) = E(U) + E(V),$$

giving

$$E(n) = N\theta + tN(R - N\theta)/T.$$

If we now replace $E(n)$ by n in this equation, and solve the resulting equation for θ, we find our estimator $\hat{\theta}$ as

$$\hat{\theta} = \frac{nT - NRt}{N(T - Nt)}.$$

I just hope that the formula I derived for the variance of $\hat{\theta}$ on p316 is correct.
In any case, with this probability model, one could now write down the probability distribution for n, from the assumption $n = U + V$, given above, and use this to maximise directly the resulting log-likelihood function for θ, finding the se of the mle in the usual way.

Looking at this problem from a more modern perspective, here is my preferred approach. Let us consider the data given in Table 2 on p315 of Doake (1978). We will take the $\delta^{18}O$ series, for which we see that there were 14 CE's in the period of observation of length 2.8 Myr, and during this period there were 18 MR's. Of these 18 MR's, exactly 5 took place within $+/-$ .01 Myr of a CE. Let $t_1 = 14 * .02 = .28$, the total length of the intervals containing a CE, and let $t_2 = 2.8 - 0.28$, the total length of the remaining time under observation.
Hence, $y_1 = 5$ can be considered to have a $Po(\lambda\mu t_1)$ distribution, and
$y_2 = 9$ can be considered to have a $Po(\mu t_2)$ distribution, and
we will assume y_1, y_2 are independent. Our question of interest is whether $\lambda > 1$, equivalently whether $\log\lambda > 0$, and this is easily addressed in R as follows.

```
y = c( 5,9) ; Time = c(.28, 2.52); CE = c(1,0) ; CE = as.factor(CE)
first.glm = glm(y ~ CE + offset(log(Time)), poisson) ; summary(first.glm)
```

This gives the following output

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.2730     0.3333   3.819 0.000134 ***
CE1            1.6094     0.5578   2.885 0.003908 **
```

showing that our estimate of $\log\lambda$ is $1.6094(se = 0.5578)$. Hence our estimate of λ is 4.99981, and the corresponding 95% confidence interval for λ is $(2.00, 12.48)$.
The rate of MR events 'close to' a CE event is increased by a factor of about 5.00,

Never mind that the deviance is 0, with 0 degrees of freedom: with the dataset compressed as it is here, we are fitting 2 parameters to 2 observations, so must get a perfect fit.
This is obviously rather a simple-minded approach, and presumably others have researched more sophisticated modelling, but it is hoped that the approach I give here is helpful in setting out basic notation and an easy method of calculation.
Note that the distribution of $y_1|y_1 + y_2$ will be Binomial, parameters $y_1 + y_2$ and p say, where

$$\log(p/(1-p)) = \log(\lambda) + \log(t_1/t_2)$$

which provides an alternative (and equivalent) way of estimating λ, and also suggests the exact test based on the binomial distribution that we used in the 1978 paper.