# APTS Statistical Inference, Assessment

Jonathan Rougier, 2013.

The volcano Campi Flegrei near Naples has had 10 large eruptions in the last forty thousand years. The mayor of Naples is worried about a large eruption happening in the next fifty years. He would like a statistician to assess the probability of this happening, denoted $\phi$.

Treat the large eruptions as a homogeneous Poisson process with rate $\lambda \,\text{kyr}^{-1}$, so that $X$, the number of eruptions in $t\,\text{kyr}$, has the conditional distribution

$$X \mid \lambda \sim \text{Poisson}(t\lambda).$$

1. Show that the score function is

$$u(x, \ell) = -t + \frac{x}{\ell}$$

   and confirm that $E\{u(X, \ell) \mid \lambda = \ell\} = 0$.

2. Show that the expected and observed Fisher Information are

$$I(\ell) = \frac{t}{\ell} \quad \text{and} \quad \hat{J}(x) = \frac{t^2}{x}$$

   respectively. Comment on the relationship between these two functions.

3. Explain why

$$\frac{(X - t\lambda_0)^2}{t\lambda_0}$$

   has approximately a $\chi^2_1$ distribution under the null hypothesis $H_0 : \lambda = \lambda_0$. (Hint: Normal approximation to the Poisson.)

4. Use the result from Q3 to show that an approximate 95% confidence interval for $\lambda$ is $(0.14, 0.46)\,\text{kyr}^{-1}$. What is the corresponding 95% confidence interval for $\phi$? (Hint: construct a confidence interval for $\lambda$ from an approximate $p$-value.)

5. Assuming that you got the second part of the previous question correct, you tell the Mayor that an approximate 95% confidence interval for the probability of a large eruption in the next 50 yr is

$(0.68\%, 2.30\%)$. He responds, "Holy cow! Are you telling me that the probability of a large eruption in the next 50 yr may be as high as 2.3%?" How do you respond?

6. Thoroughly bemused by your answer (if you were honest!), the Mayor, who is no slouch when it comes to statistics, asks you to compute the actual coverage of the interval you provided. What is it? (Hint: bootstrap.)

7. Consider a Bayesian assessment of the posterior distribution of $\lambda$, based on a prior distribution of the general form

$$\pi_\lambda(\ell) \propto \ell^{a-1} e^{-b\ell}$$

for $\ell > 0$ and zero otherwise, the hyperparameters $a$ and $b$, both positive, being specified—this is a Gamma$(a, b)$ distribution. Show that the posterior distribution has the general form

$$\pi^*_\lambda(\ell) \propto \ell^{a+x-1} e^{-(b+t)\ell}$$

and provide an interpretation for $a$ and $b$.

8. You decide to adopt the minimally-informative values $a = b = 0$. Although $\pi_\lambda$ is then improper, $\pi^*_\lambda$ is proper, because both $x$ and $t$ are positive. What is your 95% equi-tailed posterior credible interval for $\lambda$?

9. What is your probability for a large event in the next 50 yr? How does it compare with the confidence interval from Q4/Q5? (Hint: you should be able to compute $\phi$ directly, without simulation or numerical integration.)

10. Reflect on the fact that the 'Frequentist' answer provides an interval for the probability $\phi$ while the 'Bayesian' answer provides a single value. Which do you think is more helpful to the Mayor?