**Practical: Fit an epidemic**

Q1: Which of the following epidemiologic parameter could be estimated from the data in the file “data\_interval\_between\_symptom\_onset.csv”, incubation period, duration of infectiousness, serial interval or generation time? Why? [10 marks]

A: Serial interval.

Serial interval is generally defined as the interval between symptoms onsets of infector and his/her infectees, which could be directly estimated from the given data. It is often estimated as an alternative to generation time (i.e. the interval between the time of infections of infectors and his/her infectees) because the symptoms onsets are observable among clinical cases. However, the mean generation time is the same as the mean serial interval only when they have the same distribution, which is not necessarily true in real life.

*Reference 1: Paul Fine, The interval between successive cases of an infectious disease, American Journal of Epidemiology, 2003*

*Reference 2: Ake Svensson, A note on generation times in epidemic models, Mathematical Biosciences, 2008*

Q2: Using AIC, determine which of the following parametric models is most suitable for estimating the distribution of the epidemiologic parameter in Q1: exponential, binomial, or lognormal? Justify your answer and quote the parameter estimate(s) for the model you select. What is the mean of the epidemiologic parameter in Q1 based on the model you selected?

A: Binomial distribution is not widely used in estimating serial intervals. It is often used when the number of “trials” is known. Due to rounding, the data “data\_interval\_between\_symptom\_onset.csv” contain zeros. We could add one day to each interval to make the estimation easier. The estimated mean serial interval would not change after the one day is deducted. Please see details from the R script. We fit exponential, lognormal, gamma and Weibull distribution using the function “fitdistr” in R. Based on the AICs, we selected lognormal distribution and estimated mean serial interval is 3.0 days.

Q3: Simulate a deterministic SIR model for the Country Q. We know that Country Q has a population size of 100,000 and there were 90 infectious individuals on 3 Jan 2019. The disease surveillance department estimated that reporting rate was about 15%, i.e. 15% of the infections had been reported. Assume that the mean duration of infectiousness in the SIR model is the same as the mean of the epidemiologic parameter in Q1. How can you estimate the basic reproductive number *R*0? If so, what’s your best estimate of *R*0?

A: Assuming that the mean duration of infectiousness is 3.0 days, we simulated a deterministic SIR model for different values of *R*0 from 1 to 3 (i.e. 1.01, 1.02, …, 3.00). With each *R*0 value, we obtained a series of daily incidences.

To measure how “close” the epidemic curve generated by a certain *R*0 is to the observed epidemic curve, we could make use of the maximum likelihood method. Suppose that the observed incidence *Yt* on day *t* is a Poisson random sample with expectation of the true observed incidence *pCt* :



The probability of observing *Yt* if the true observed incidence on day *t* is *pCt* (we can use the Poisson probability mass function “dpois” in R)*:*



Under a particular *R*0 value, from the SIR model we could have a series of “true” incidence *Ct* (*t* = 1, 2, 3, …) and the probability of observing the incidence series of *Yt*is:



We want to look for a particular *R*0 under which it is most likely that we can observe the given epidemic curve. Therefore, our best estimate of *R*0 is the one which could maximize the above probability (i.e. the likelihood function). To find out the best estimate of *R*0 , we could calculate the above likelihood under different *R*0 values and see which *R*0 value has the maximum likelihood.

The best estimate of *R*0 is 1.74. Please see details from the R script.