Tutorial 9b: frequentist inference for R_0

Required files: `Tutorial9b.R`, `stochasticSIRmodel.R`, `MLEforSIR.R`, `cplot.R`

1 Comparing maximum likelihood estimates of R_0

Suppose that a herd of 25 animals underwent a disease epidemic caused by a virus in a farm (hereafter called farm A). The epidemic lasted one year, and during this period the counts of susceptible, infectious and recovered individuals were recorded on a weekly basis. Also assume that the disease under study and the resulting epidemic data follow an SIR model.

Question 1.1: Given a 95% confidence interval for R_0 of $[0.67, 4.33]$ using the disease data from farm A, what can be said about the behaviour of the epidemic observed in this farm? Also, how is this interval interpreted?

An epidemic similar to the one observed in farm A was observed in a larger herd of size 50 in another farm B, also during one year. Using weekly counts of susceptible, infectious and recovered individuals for this epidemic, a confidence interval of $[1.35, 3.65]$ was estimated.

Question 1.2: How is this estimate compared to the one obtained using the data from farm A?

Question 1.3: Can you think of potential factors which might have contributed to the different estimates obtained in farms A and B?

2 Evaluating the effects of sample size and individual heterogeneity on R_0 estimates

Novel statistical models are often evaluated by checking the properties of their parameter estimates using simulated data. In this practical we will use this idea to study the properties of maximum likelihood estimates of R_0 based on data simulated from stochastic SIR models.

Assume the same virus disease considered in Section 1 and also that each replicate from the simulation is equivalent to a farm where an epidemic occurred. It will be also considered that there is variation among animals with respect to their ability to acquire and transmit the disease, but R_0 estimates to be obtained will ignore this variation. For now you don't need to understand how individual heterogeneity was included in the simulation, this will be described on Friday.

Use the script `Tutorial9b.R` to run the code required to answer the questions of this Section.

Take some time to familiarise yourself with the code. In summary, this code generates stochastic SIR epidemics assuming sample sizes 25, 50 and 100 and $R_0 = 2$ (we will compare all the estimates obtained from the simulated data with the true

value of R_0). Then, maximum likelihood estimates are obtained and plotted for each of the simulated replicates.

Generate 30 replicates of stochastic SIR data assuming sample size 25 and no individual variation. Then, calculate and plot maximum likelihood estimates and 90% confidence intervals for R_0. (*Use step 1 of the R script*)

Question 2.1: How close the point estimates of R_0 from the replicates are to the true R_0 value?

Question 2.2: Are the confidence intervals in line with their frequentist interpretation. Why?

Question 2.3: Based on these simulations, do you think that 25 animals are enough to use confidence intervals to infer whether the epidemic under study doesn't die out quickly? Why?

Question 2.4: Generate additional replicates of stochastic SIR data assuming sample sizes 50 and 100. Then, calculate and plot maximum likelihood estimates and 90% confidence intervals for R_0. (*Use step 3 of the R script*)

What's the impact of sample size on the R_0's estimates?

Question 2.5: Based on these simulations, what would be the minimum number of animals to infer whether the epidemic under study doesn't die out quickly? Why?

Now generate stochastic SIR data for sample sizes 25, 50 and 100, but this time considering **individual variation among animals with respect to their ability to acquire and transmit the disease**. Then, calculate and plot maximum likelihood estimates and 90% confidence intervals for R_0 (obtained *ignoring* individual heterogeneity) for all replicates (*Use step 4 of the R script*).

Question 2.6: What's the impact of individual heterogeneity on R_0's estimates?

3 Designing a transmission experiment for estimating epidemiological parameters

Suppose you are a researcher interested in understanding the transmission of IPN in Atlantic Salmon (described in the *Modelling Epidemics* lecture on Tuesday). For your research you have available a very generous lab with available resources to carry out infection experiments using an arbitrary number of fish and tanks in order to obtain epidemic data on IPN. Consider the following epidemiological assumptions:

- IPN is transmitted by direct contact.

- once fish is infected with IPN, it remains infected during its lifetime, with no latency period. So it can assumed an stochastic SI model for IPN.

- infection events in IPN epidemic are **independent.**

Question 3.2: Now assume that fish can recover from IPN. How R_0 could be estimated in this case?