Linking Epidemiology and Phylogenetics

Leo Featherstone
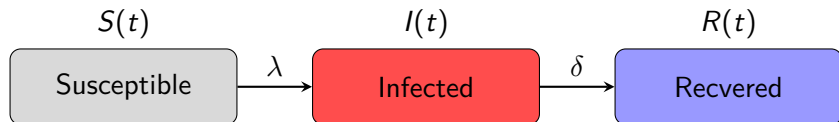
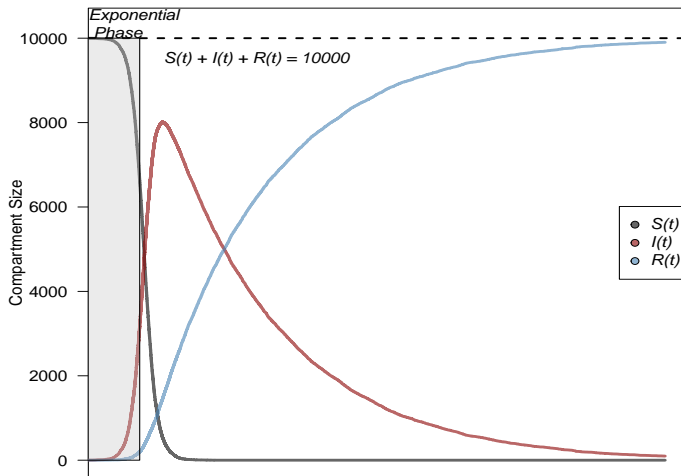University of Melbourne

2021

Breakdown

- Introduce core epidemiology - SIR
- Two core models: Birth Death and Coalescent Exponential
- Coalescent Exponential detail
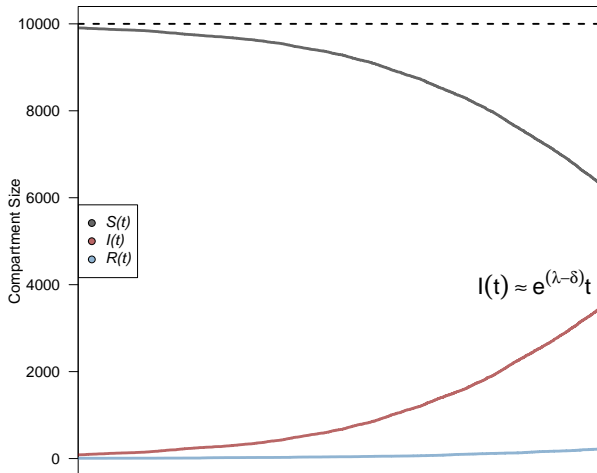- Birth Death detail
- Compare and contrast the two

SIR Model

SIR Trajectory

SIR Exponential Phase

A closer look at the exponential Phase

$$I(t) \approx e^{(\lambda - \delta)}t$$

Constant Rates in Exponential Phase

The logic:
$$S(t) \approx N \text{ (recalling that N=total population size)}$$
$$\Rightarrow \text{constant } \lambda \text{ and } \delta$$
$$\Rightarrow r = \lambda - \delta$$
$$\Rightarrow R_0 = \frac{\lambda}{\delta}$$

Coalescent Exponential

- Comes from population genetics
- Assumes an exponentially growing population
- Two parameters; exponential growth rate (r) and scaled effective population size (ϕ)
    - ϕ is not an easily explained number, but it reflects incidence prevalence
    - Under exponential growth, $\phi \propto N_e \propto I(t)$
    - If you want number of missed cases, use $N = e^{rt}$ instead!
- Offers R_0 through $R_0 = r + 1/D$ where D so $\delta = \frac{1}{D}$ 🐸

Birth Death

- Three parameters: λ, δ, and ψ
  - note $\delta = \mu + \psi$ and μ is the natural death rate. I.e. End of infection without sampling
  - $p = \frac{\psi}{\mu + \psi}$ We work in terms of p rather than ψ
- Allows for stochastic population growth
- Incorporates sampling time data
- Offers R_0 by allowing us to estimate λ and δ such that $R_0 = \frac{\lambda}{\delta}$
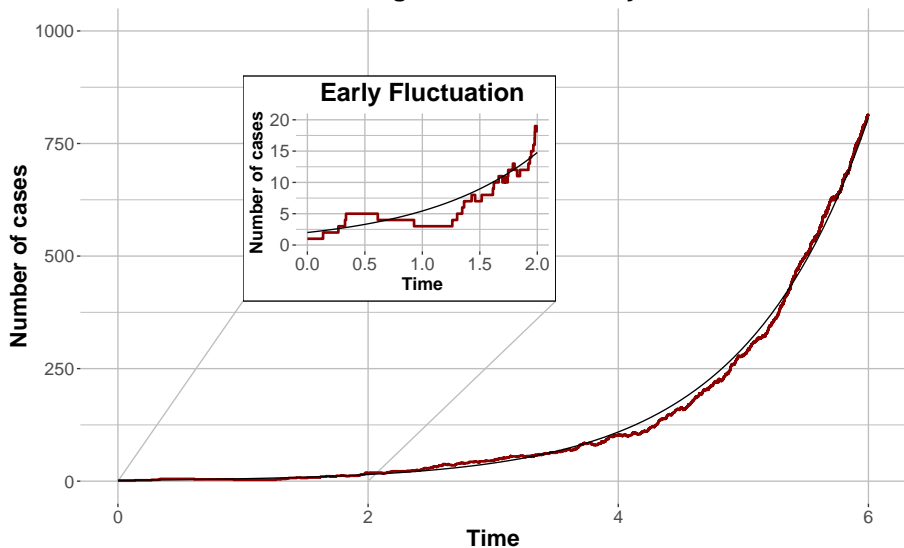
Coalescent Exponential vs Birth Death

cell1	Birth-Death	Coalescent Exponential
Stochastic Population Growth	yes	no
Uses Sampling information	yes	no

- BD more suited to smaller samples with a less clear signature of population growth
- CE A more conservative choice for larger exponentially growing datasets

Coalescent Exponential vs Birth Death

Long Term Predictability

Early Fluctuation
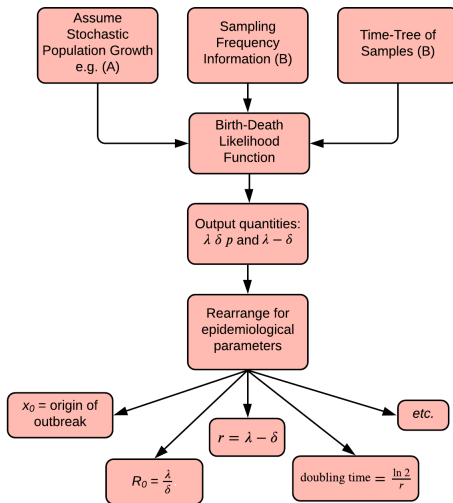
Coalescent Exponential vs Birth Death

- BD also uses information from sampling time. Use it if your times are informative, like if sampling a constant proportion of cases as epidemic grows!
- CE conditions on sampling times, so use it when you need to be robust to these

But Leo, where are the ACTUAL EQUATIONS??

- These models give us likelihood equations, expressing the probability of particular trees
  - You do not need to know them! Look here if interested.
- Our mcmc then explores the posterior distribution based on these likelihood equations, and others (substitution model, etc)

MCMC Algorithm

Likelihood → Posterior
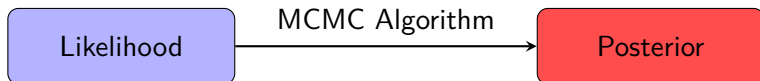
Birth Death Summary

Coalescent Exponential Summary
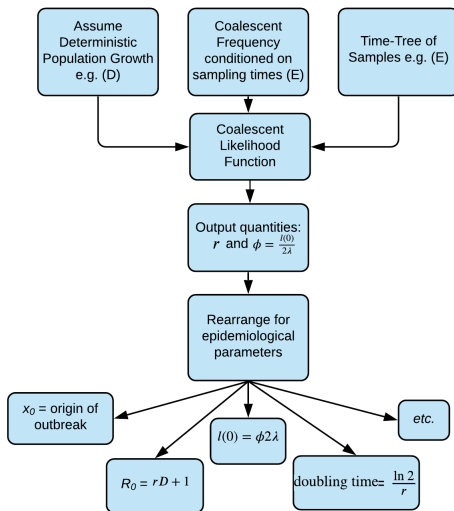
Not to worry! There are many other models that can accommodate more complex data. They are all based on the BD and CE though, so we needed to start here.

From here, you can understand the Bayesian Skyline, Birth Death SIR, Coalescent Constant, 🐸, and many more!