Probability Distributions

Leo Featherstone

University of Melbourne

2021

About me

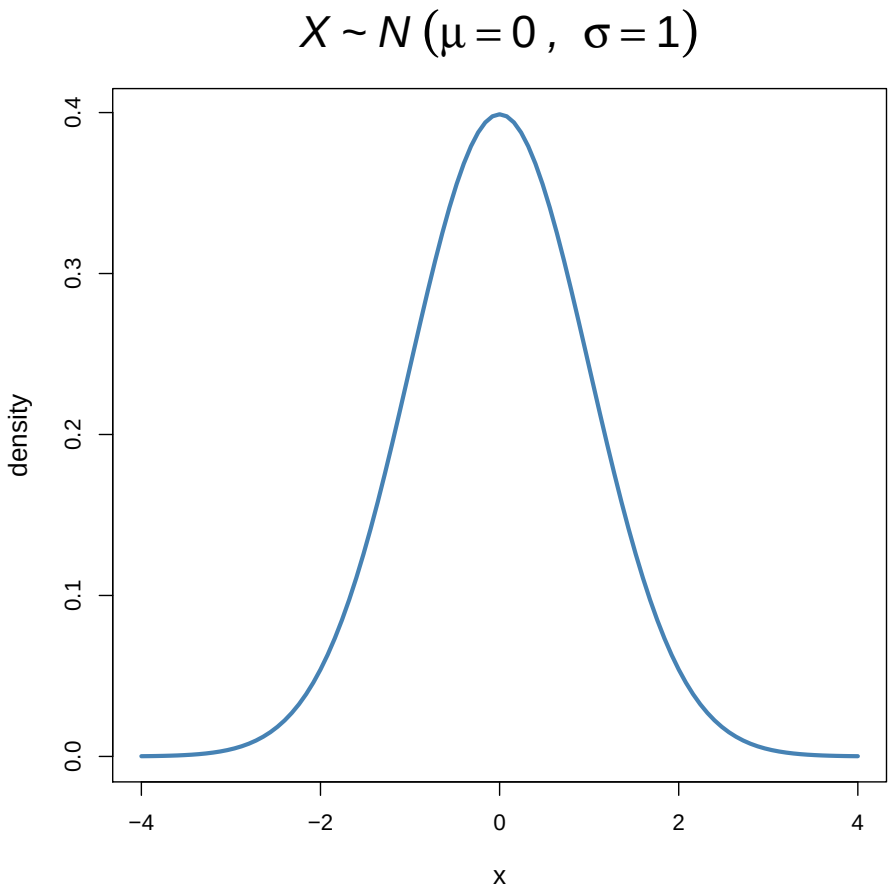
- PhD Student & RA in phylodynamics
- Fascinated by theory behind our models and optimising them for future public health challenges
- EEB & Mathematics at University of Melbourne with Honours in Holt group at CCS
- 

What's the value of this information?

- It's not enough just to use default priors in BEAST
- We need some commands over prior distributions to reflect prior knowledge
- We still Google this stuff all the time, but we know what we are looking for!

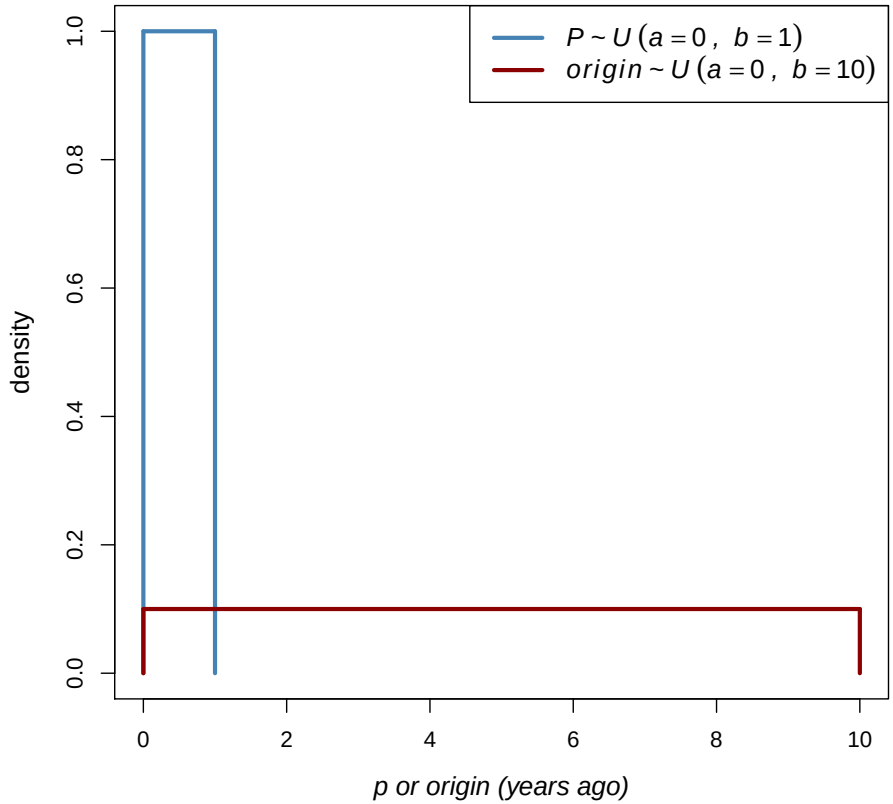
What is a distribution?

- A curve of probability of values of a variable
  - Area underneath adds to 1
- Shape governed by parameters - We must understand these



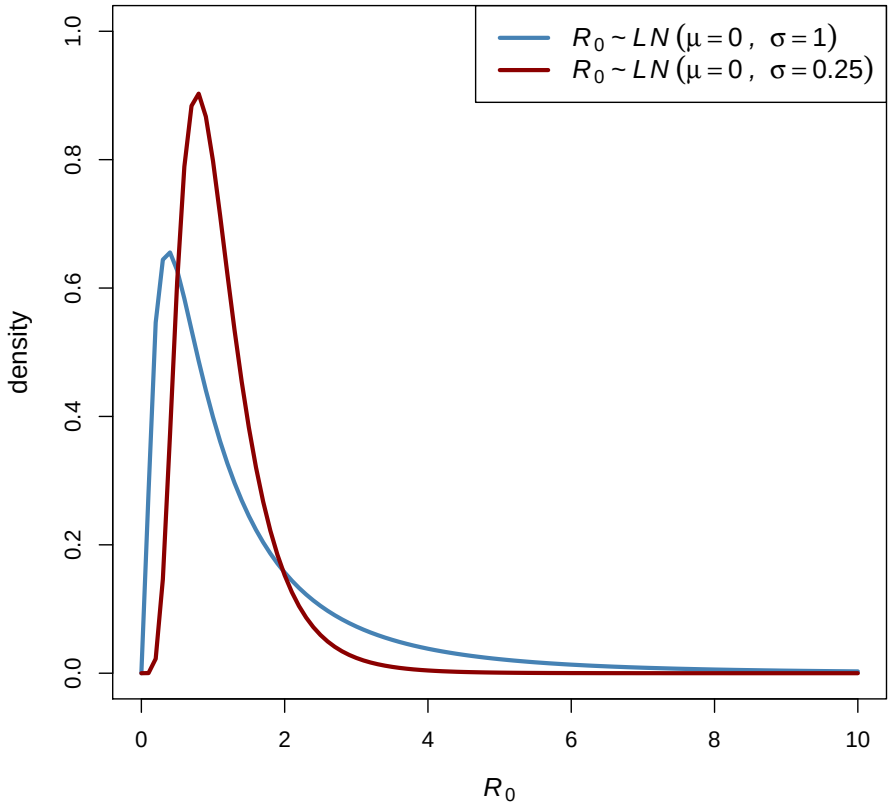
Uniform distribution: $U[a, b]$

- Relevant to sampling probability p or origin of tree
- Parameters are a and b, the lower and upper values



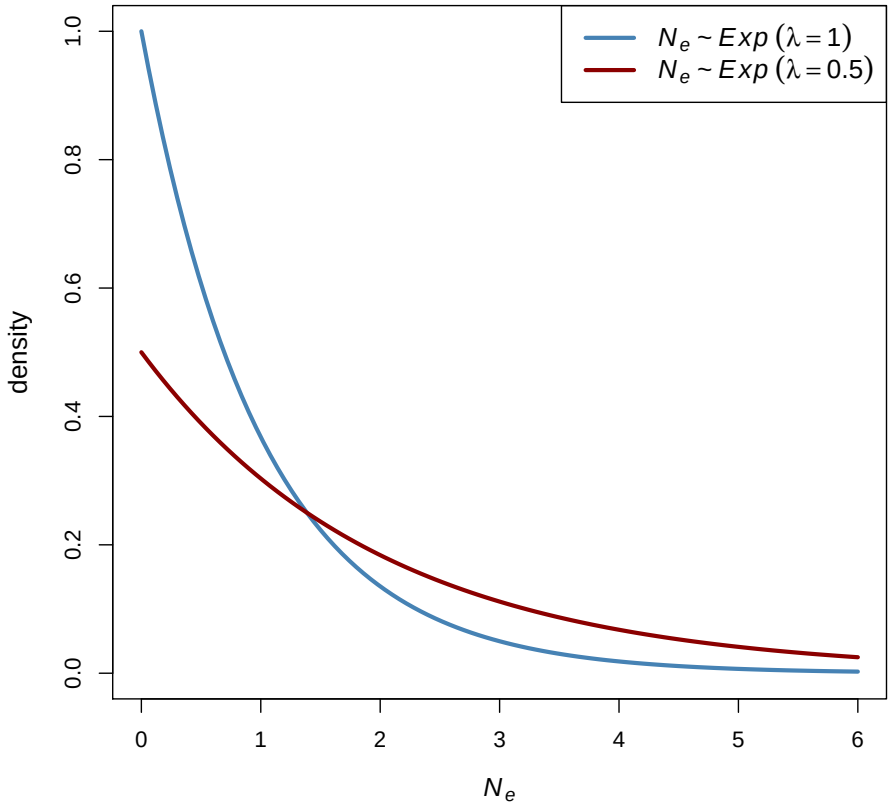
Lognormal

- Relevant to sampling probability $R_0 \in [0, \infty)$
- Parameters are μ and σ: mean and standard deviation of the normal distribution achieved if log taken



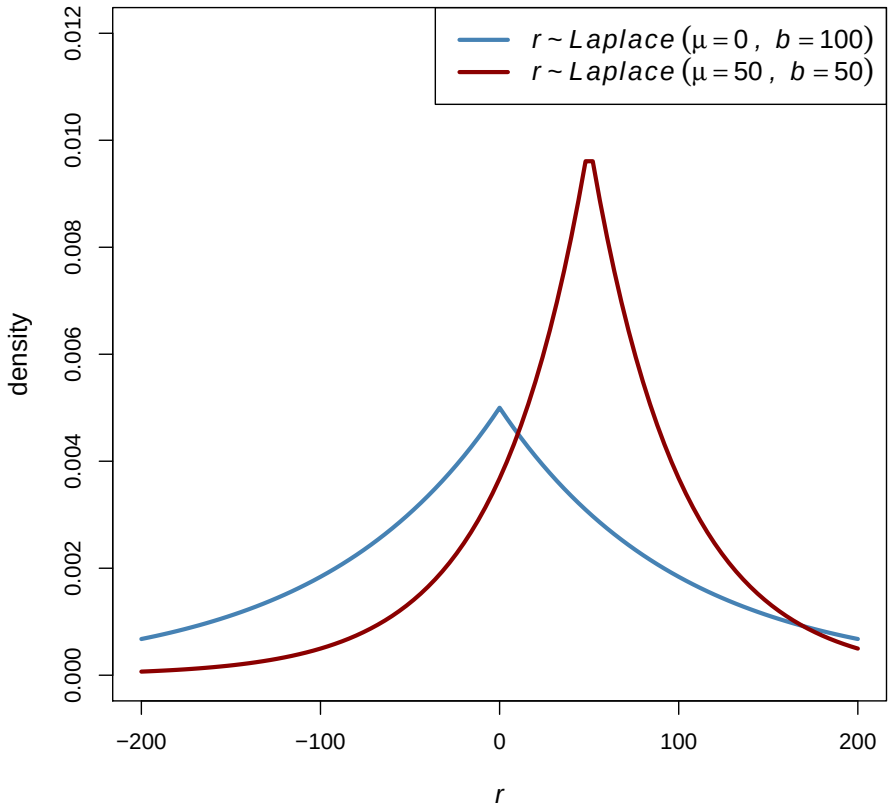
Exponential Distribution

- Relevant to effective population size: $N_e \in [0, \infty)$
- Also models waiting times. Hence λ also called the 'rate'
- Is actually what we use to model time between branching or sampling events!



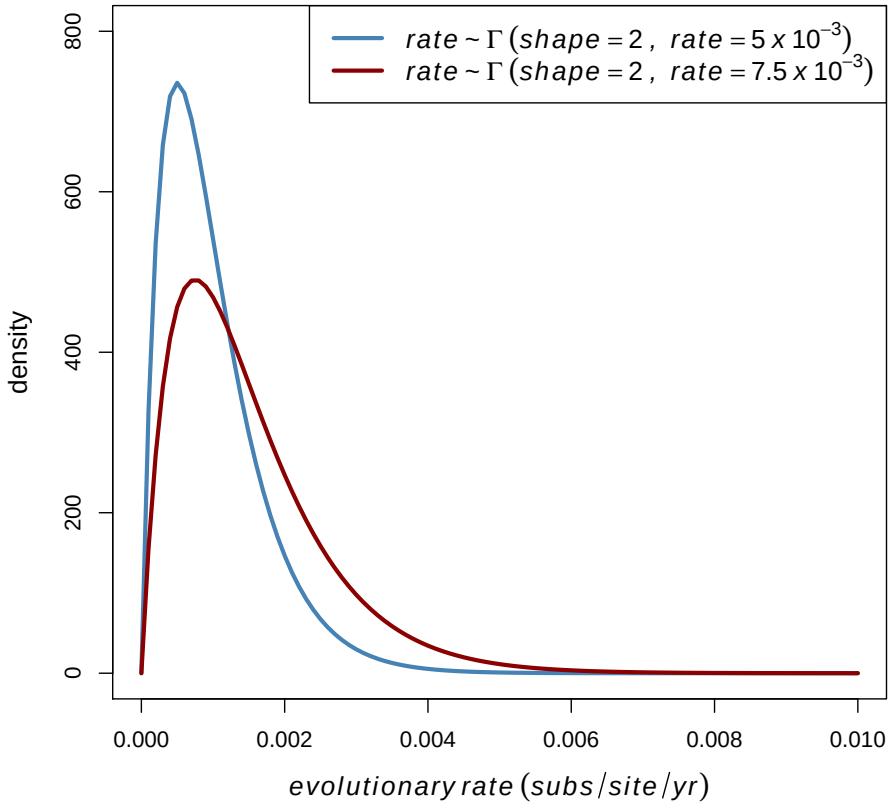
Laplace Distribution

- Relevant to population growth and decay rates $r \in (-\infty, \infty)$
- Basically back to back exponential distributions!
- μ is the 'location' or peak
- b is the scale. Bigger $\rightarrow$ wider distribution



Gamma Distribution

- Relevant to many parameters, including evolutionary rate $evoRate(subs/site/yr) > 0$
- Shape and scale parameters α, β in BEAUTI and k, θ on wikipedia
  - Annoying, I know...
- mean is $\alpha\beta$, variance is $\alpha\beta^2$



Gamma Example: setting an evolutionary rate prior

- You're doing a study on the hypothetical 'Furphy-Virus', a new member of the beer-virus family. You have some new sequences and want to analyse their epidemiology with BEAST. You will need an informative prior on the evolutionary rate to do this.
- *Human et al.* recently estimated it's evolutionary rate to be roughly $10^{-3} subs/site/year$ and literature on the Stella Artois and now extinct Fosters Viruses produce a variance in mean estimates of $0.5x10^{-5}$

- Suppose you choose a gamma-shaped prior and you bust out some high school mathematics!

$$\text{mean} = \alpha\beta = 10^{-3}$$

$$\text{variance} = \alpha\beta^2 = 0.5x10^{-3}$$

$$\implies \alpha = \frac{10^{-3}}{\beta}$$

$$\implies \frac{10^{-3}}{\beta}\beta^2 = 0.5x10^{-3}$$

$$\implies \beta = \frac{0.5x10^{-3}}{10^{-3}} = 0.5$$

$$\implies \alpha = \frac{10^{-3}}{0.5} = 2x10^{-3}$$

  - So you choose a $\Gamma(\alpha = 0.002, \beta = 0.5)$ prior