Taming the Beast Workshop

Priors and starting values

Veronika Bošková & Chi Zhang

June 28, 2016

What is a prior?

- Distribution of a parameter before the data is collected and analysed
- as opposed to POSTERIOR distribution which combines the information from the prior and the data

What is a prior?

- Using Bayes theorem, we can decompose the posterior:

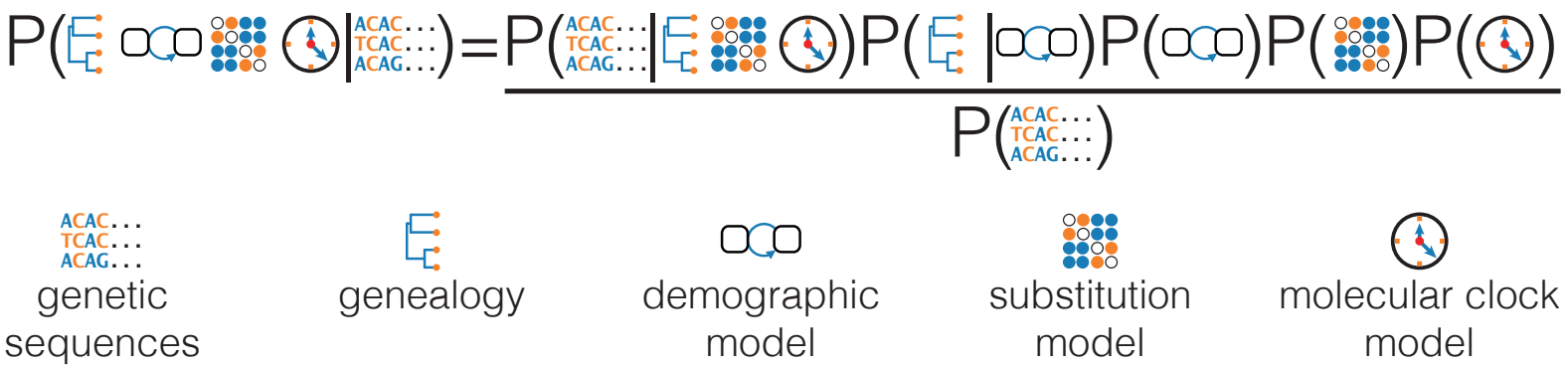


Figure adapted from [du Plessis and Stadler, 2015]

What is a prior?

- Using Bayes theorem, we can decompose the posterior:

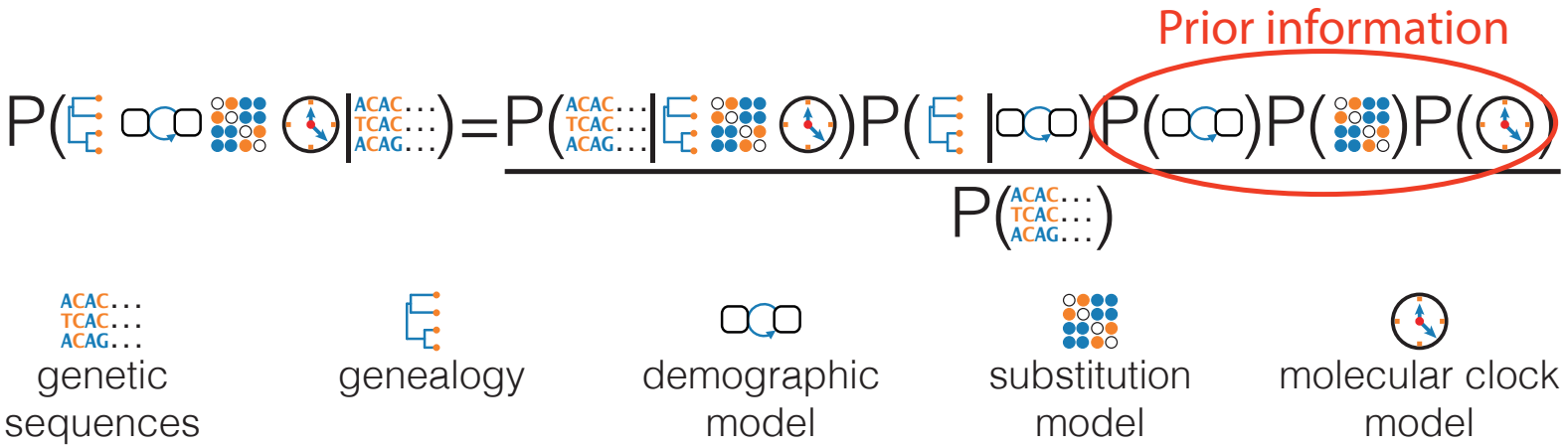


Figure adapted from [du Plessis and Stadler, 2015]

What is a prior?

- Using Bayes theorem, we can decompose the posterior:

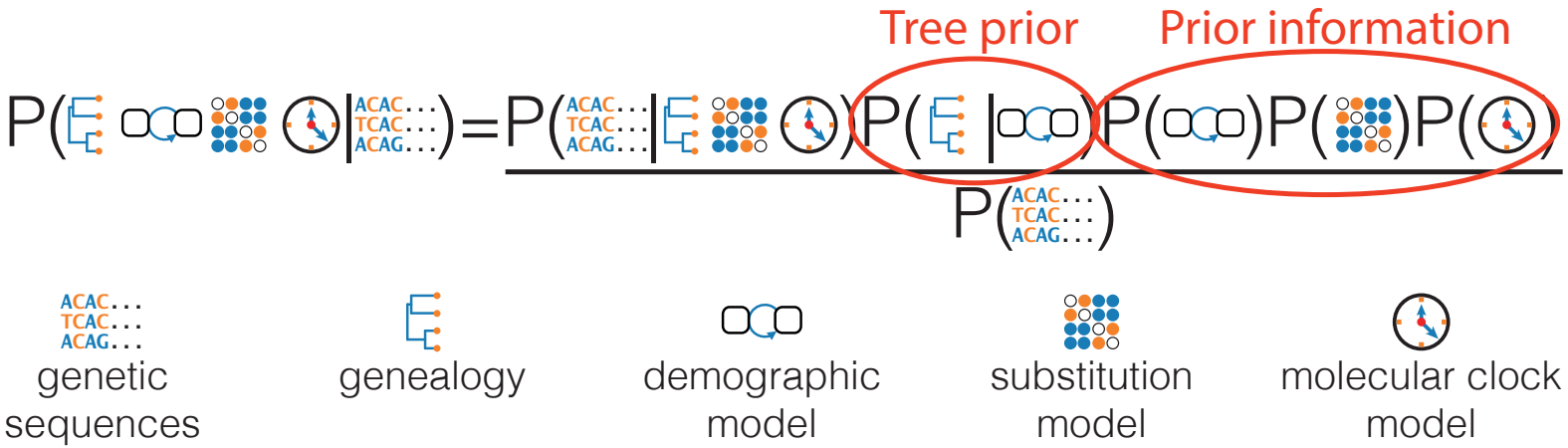


Figure adapted from [du Plessis and Stadler, 2015]

Prior

- Allows us to include any information we have on the process, before looking at the data
  - Do not be afraid of using it in the inference
- Prior distribution does not have to, and is not expected to, be exactly the same as the posterior

Prior

- Should not be and is not universal for all the analyses you will ever do in your research
- Should incorporate prior (before looking at the data) knowledge about the parameter/underlying process
  - use results of previous independent experiments
  - use other independent evidence
- Should not be too restrictive if prior knowledge/assumptions are weak
  - One can use diffuse priors
- May not be adjusted after the run, to give higher and higher posterior support

Prior

- Is a choice of
  - model
    - tree-generating models, nucleotide/AA/codon substitution models, ...

  and of
  - distribution of plausible values for a parameter of interest
    - Uniform, Normal, Beta,...

Tree prior (tree-generating model)

- Have to pick one from Coalescent or Birth-death process framework
- Have to put priors on parameters of the chosen model
  - e.g. growth-rate of the population, R0, extinction rate, ...

Substitution model prior

- The selection is big: JC69, HKY85, ..., GTR
- Use model which has been previously identified to be best for your type of data
  - e.g. HKY85
    - Prior for transition/transversion rate ratio (κ)
    - Prior for base frequencies
- To choose the best model
  - Use model comparison to choose the one best fitting the data
  - Use rjMCMC directly in BEAST2 to sample from the posterior distribution including different substitution models. The model where rjMCMC spends the most time (samples the most from), is the best fitting model.

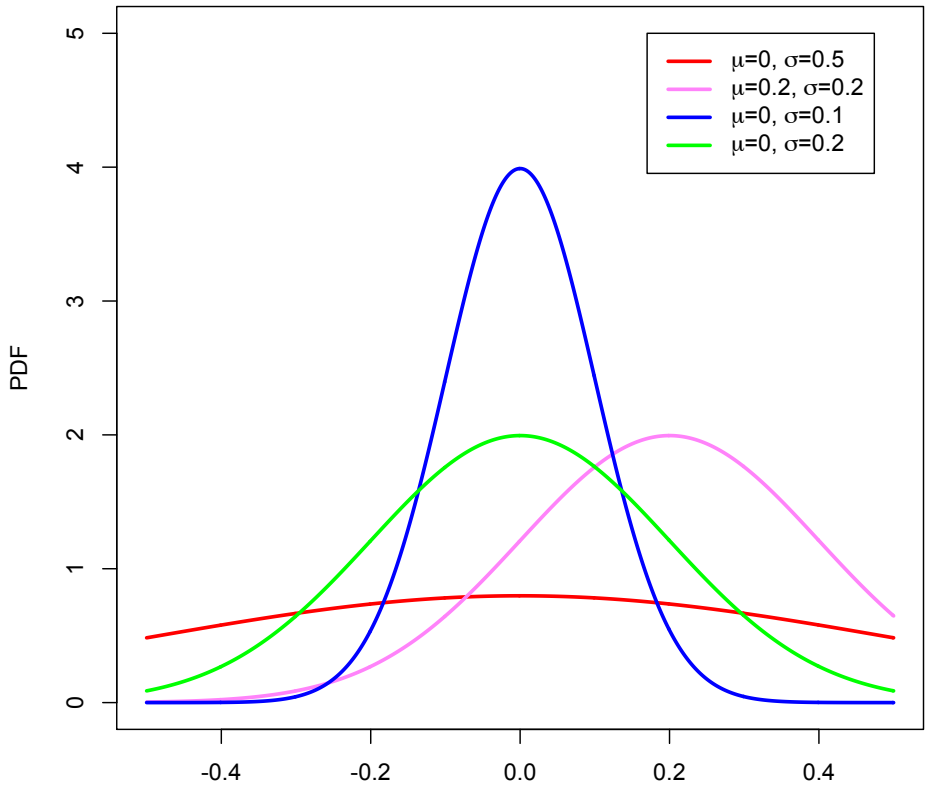
Clock prior (molecular clock model)

- Strict clock: all branches have the same clock rate
- Relaxed clock
  - Uncorrelated: branches have independent clock rate distributions
  - Correlated: child branch has clock rate distribution correlated to distribution of the parent branch

Parameter prior

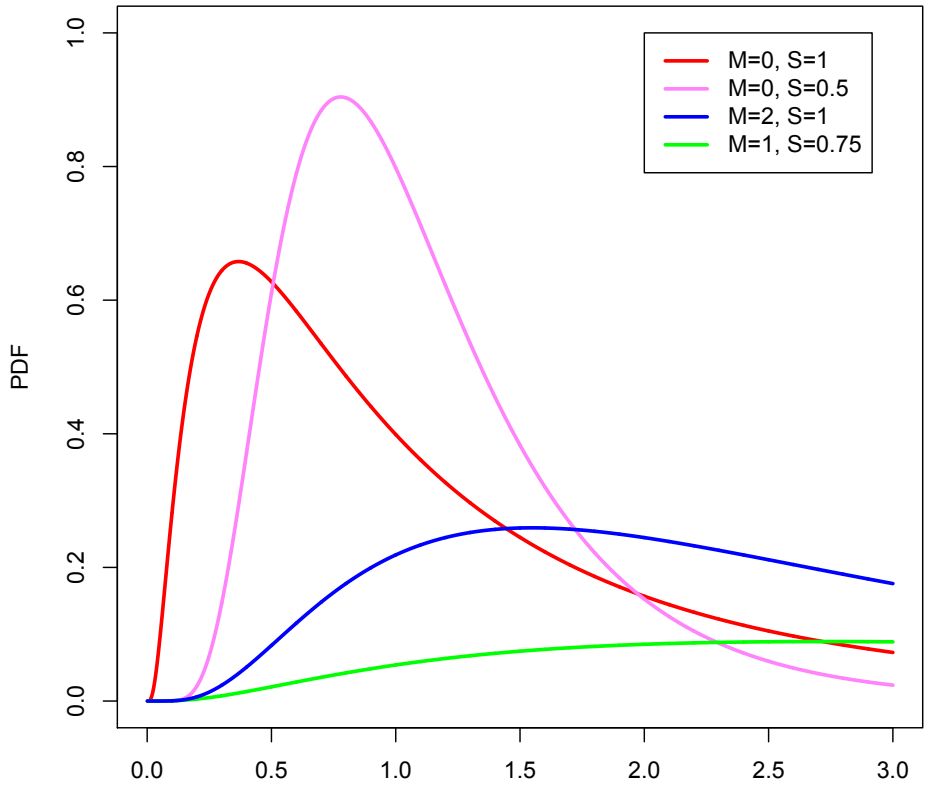
- Can be fixed to a given value
  (though this is generally not recommended)
- Can have upper and lower limits
  - If we know that any infected individual recovers after 5-10 days, we can set the distribution of infectious period to be e.g. min 4 days and max 11 days
- If specified by a parametric distribution, the parameters of this distribution can also be assigned a prior (hyperprior)
- You can visualise the distribution in BEAUti

Examples - Normal distribution



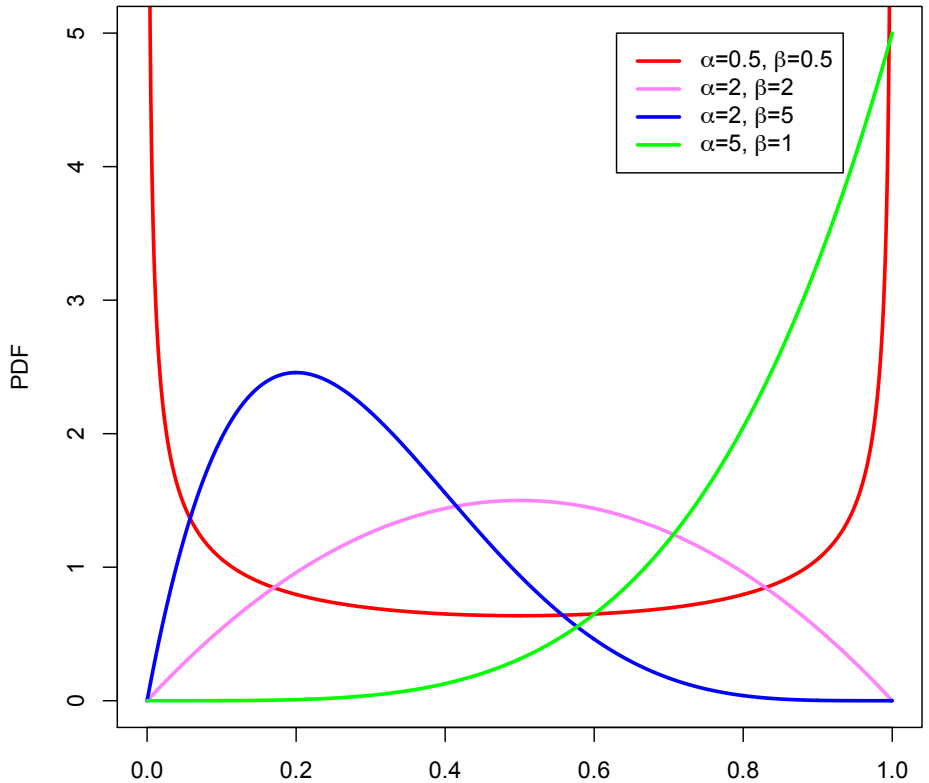
- Parameters: mean $\mu \in \mathbb{R}$, standard deviation $\sigma > 0$
- Range of values: $(-\infty,\infty)$

Examples - LogNormal distribution



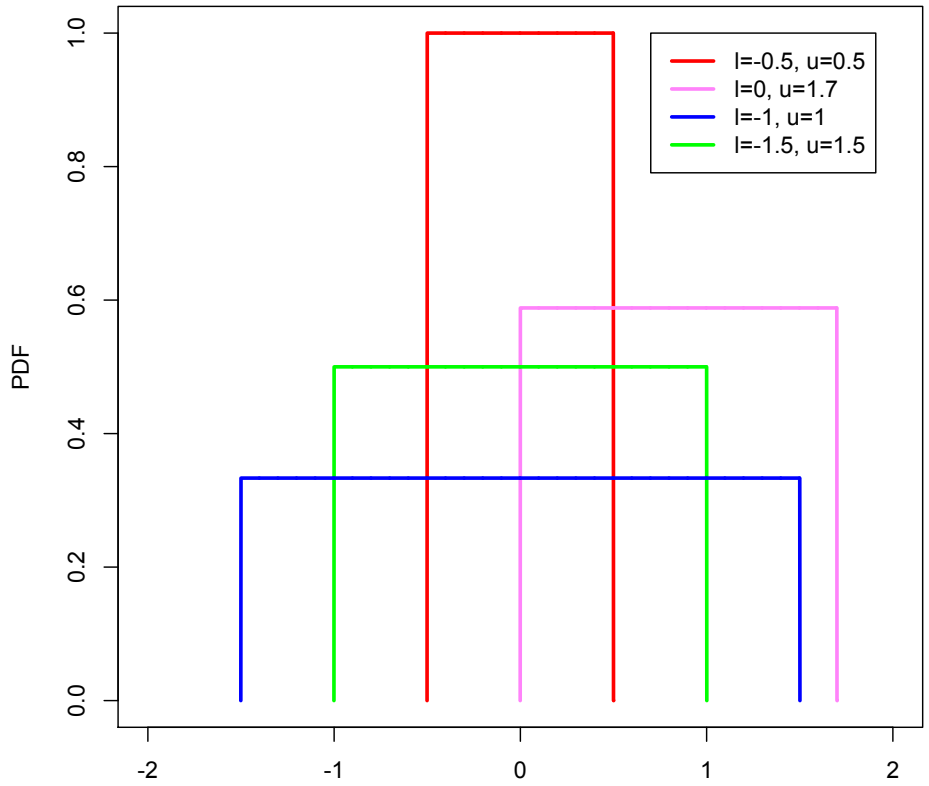
- Parameters: mean M $\in \mathbb{R}$, standard deviation S > 0
- Range of values: $[0,\infty)$
- Long tail, always positive

Examples - Beta distribution



- Parameters: shape $\alpha > 0$, shape $\beta > 0$
- Range of values: [0,1]
- Good for e.g. sampling probability prior

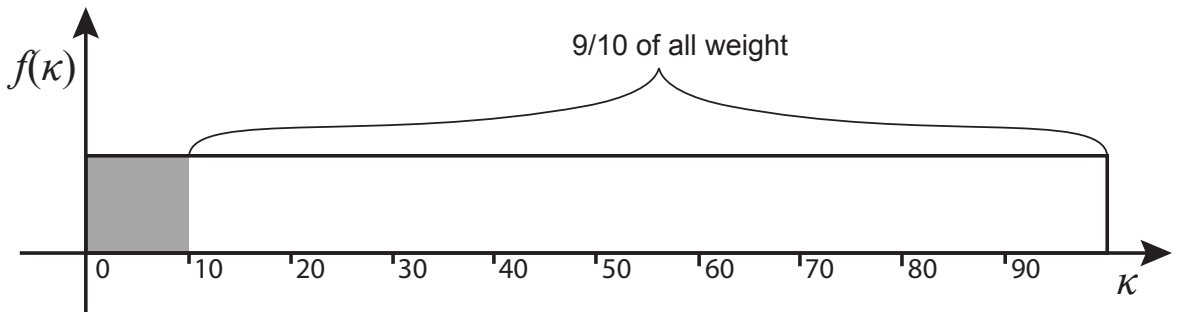
Examples - Uniform distribution



- Parameters: lower, upper bound
- Range of values: (-∞,∞)

Is uniform distribution a non-informative prior?

- Not really
  - Imagine setting a Uniform(0, 100) prior for the transition/transversion rate ratio (κ). You also know that the most likely values for κ are between 0 and 10. But you now put 9/10 of the weight to values > 10.



- In fact there is nothing such as an non-informative prior
- If little or no information on the parameter is available, use diffuse priors
- Try to avoid Uniform(-∞, ∞) or Uniform(0, ∞)

Proper vs improper priors

- Sometimes the prior distribution is such that the sum or the integral of the prior values does not converge, this is called an IMPROPER prior
- Examples
  - $1/x$
  - Uniform$(-\infty,\infty)$

Are my priors what I set them to be?

- Not always
  - Induced priors may change the picture, i.e. if the parameters interact, the marginal prior distribution for each individual parameter may be different from the originally specified prior
- Use sampling from the prior, to see what your 'real' prior is

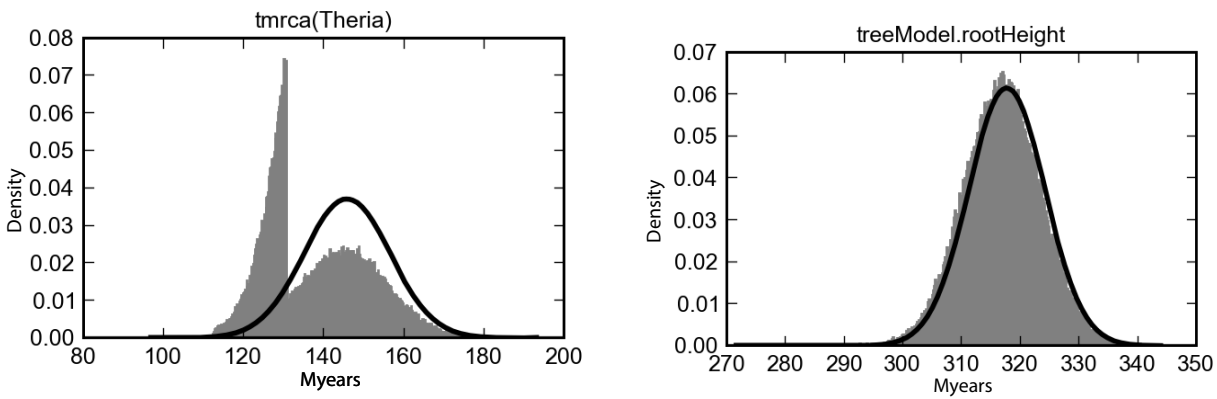


Figure adapted from [Heled and Drummond, 2012]

The marginal prior distributions that result from the multiplicative construction (gray) versus calibration densities (black line) specified for the calibrated nodes.

How to choose priors?

- Use all the prior knowledge you have to choose models and set appropriate parameter priors
- Sample from the prior distribution before using your data to check you really have the priors you want
- Check your posterior distribution against the prior

Word of caution

- In practice, it is important to evaluate the impact of the prior on the posterior in a Bayesian robustness analysis
- Ideally, the posterior should be dominated by your data, such that the choice of the prior has little influence on the result
- If this is not the case, the choice of prior is very important, and should be reported

Starting values

- Are just starting values
- Have to be within the prior distribution, and its upper and lower limits, you chose for the parameter
- Use your best guess
  - BEAST2 attempts 10 times at most (can be changed) to initialize the run, but if the starting values are unreasonable, the runs may keep failing
- Start from different starting values to make sure the chains converge to the same distribution

References I

- du Plessis, L. and Stadler, T. (2015). Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends in microbiology*, 23(7):383–386.
- Heled, J. and Drummond, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, 61(1):138–149.