Taming the Beast Workshop

Prior selection and Troubleshooting

Jana Huisman

Based on presentations by:
Timothy Vaughan; Veronika Bošková; Jūlija Pečerska

18.2.2018

Part I

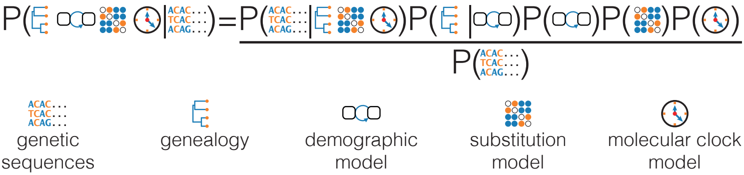
Choosing Priors

Bayesian Inference

In a traditional Bayesian inference we use Bayes theorem to learn about some parameters ϕ using some data D:

$$P(\phi|D) = \frac{P(D|\phi)P(\phi)}{P(D)}$$

In Bayesian phylodynamic inference, the details are more complex but the idea remains the same:



What do we mean by "prior"?

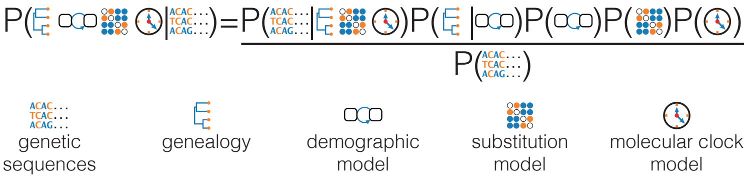
$$P(\phi|D) = \frac{P(D|\phi)P(\phi)}{P(D)}$$

- The prior quantifies what you knew prior to taking the data D into account.
- This can include both:
  - Aspects of the model (e.g. which phylodynamic model to use)
  - Knowledge of specific parameters (e.g. possible clock rates).
- When people talk about priors, they are *usually* referring to prior probability distributions over model parameters.
- Treatment of the model itself as part of the prior becomes important in the context of model selection, which Remco will focus on tomorrow.

General advice for choosing priors

- Every prior distribution should be chosen with the particular analysis in mind: no priors are universal.
- This is because we must incorporate relevant prior knowledge into each distribution, including:
  - What we have learned from previous data,
  - Constraints imposed by expert knowledge.
- Avoid making untenable assumptions!
  - E.g. using a prior which is zero for certain parameter values.
- Extremely important that the prior be chosen without recourse to the data to be analyzed.
  - Do not adjust after the run, to give higher posterior support
  - Breaking this rule is known as “double-dipping” and produces incorrect results.

Parametric priors in Bayesian phylodynamics

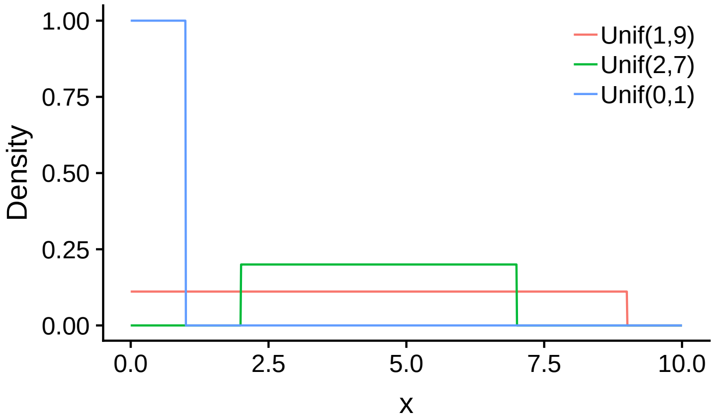


There are several parametric priors that can be specified in a phylodynamic analysis:

- Priors on clock rates.
- Priors on substitution model parameters (e.g. transition/transversion rate ratio).
- Priors on phylodynamic model parameters (birth rates / effective population sizes).

Most of these priors are specified using members of a small family of univariate probability densities defined on $x \in [0, \infty]$.

Univariate priors: Uniform distribution



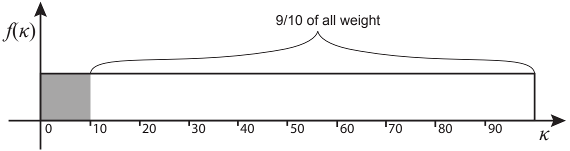
Probability density function for $\mathrm{Unif}(x_{\mathrm{min}}, x_{\mathrm{max}})$:

$$f(x) = \begin{cases} \frac{1}{x_{\mathrm{max}} - x_{\mathrm{min}}} & \text{if } x_{\mathrm{min}} < x < x_{\mathrm{max}}, \\ 0 & \text{otherwise} \end{cases}$$

Used to restrict parameter value to definite bounds.

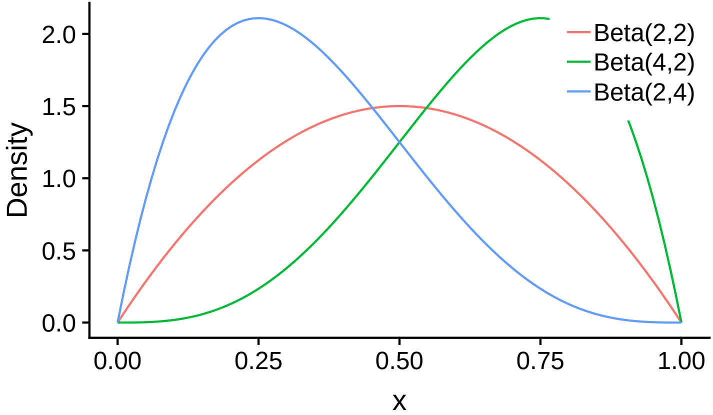
Is uniform distribution a non-informative prior?

- 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, ∞)

Univariate priors: Beta distribution

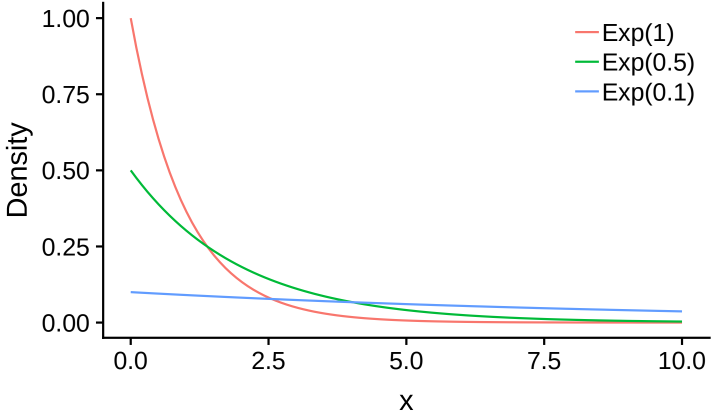


Probability density function for $\mathrm{Beta}(a, b)$ when $0 \leqslant x \leqslant 1$:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Flexible prior for quantities defined on the $[0, 1]$ interval.

Univariate priors: Exponential distribution

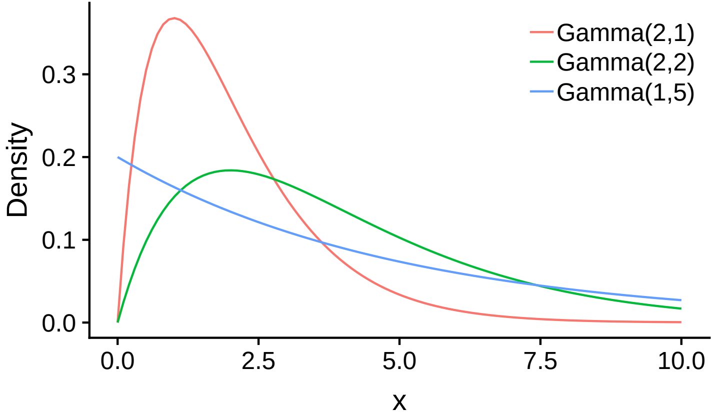


Probability density function for $\mathrm{Exp}(r)$ for $x \geqslant 0$:

$$f(x) = e^{-rx} r$$

Mean and variance are $1/r$ and $1/r^2$ respectively. Quite an informative prior for the parameter, so use with care.

Univariate priors: Gamma distribution

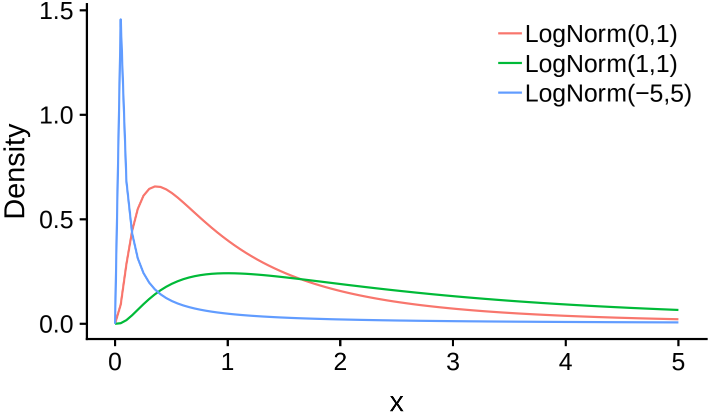


Probability density function for $\mathrm{Gamma}(\alpha, \beta)$ where α and β are the "shape" and "scale" parameters:

$$f(x) = \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

A flexible generalization of the exponential distribution.

Univariate priors: Log-normal distribution

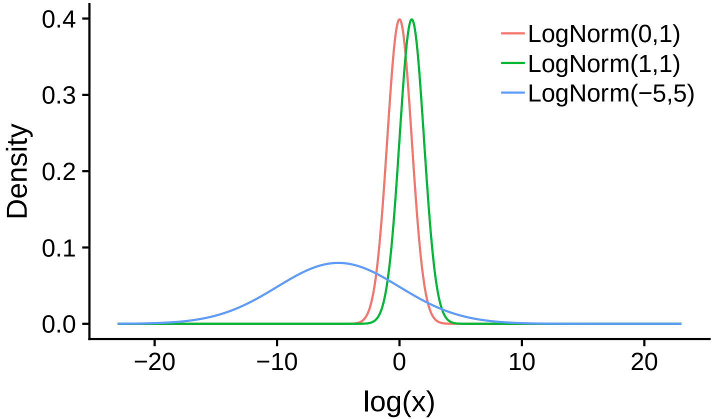


Probability density function for $\mathrm{LogNorm}(M, S)$:

$$f(x) = \frac{1}{\sqrt{2\pi S x}} e^{-(\log(x)-M)^2/(2S^2)}$$

Note that $\log(M)$ is the median while S determines the standard deviation in log space.

Univariate priors: Log-normal distribution

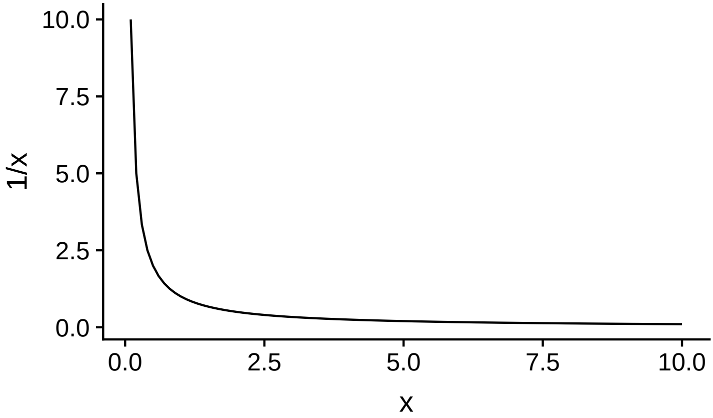


Probability density function for $\mathrm{LogNorm}(M, S)$:

$$f(x) = \frac{1}{\sqrt{2\pi S x}} e^{-(\log(x)-M)^2/(2S^2)}$$

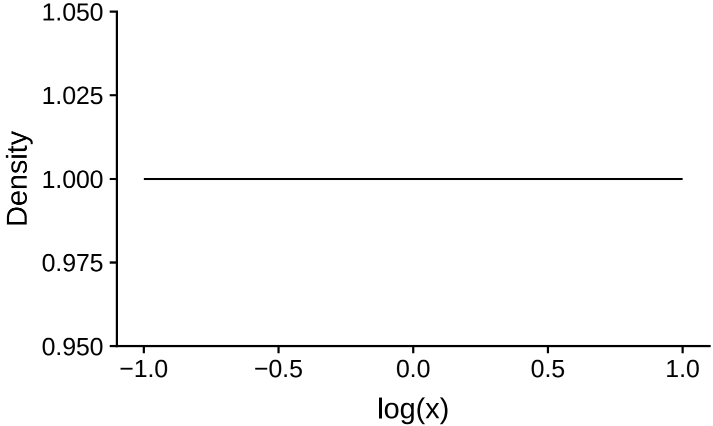
Note that $\log(M)$ is the median while S determines the standard deviation in log space.

Univariate priors: 1/X distribution



- Equivalent of a uniform distribution on the logarithm of the parameter. Without bounds, this distribution cannot be normalized. (The area under the curve is infinite.)
- May be a natural "non-informative" prior for rate parameters: uncertainty equally distributed amongst orders of magnitude.

Univariate priors: 1/X distribution



- Equivalent of a uniform distribution on the logarithm of the parameter. Without bounds, this distribution cannot be normalized. (The area under the curve is infinite.)
- May be a natural "non-informative" prior for rate parameters: uncertainty equally distributed amongst orders of magnitude.

Proper and improper priors

- Any true probability distribution can be used as a “proper” prior.
- Occasionally practitioners use priors that are not true normalizable probability distributions, e.g. $\mathtt{Unif}(0, \infty)$ or a $1/X$ prior with no upper bound. Such distributions are called “improper” priors.
- When using an improper prior, there is always a danger that the resulting posterior will also become improper.
- You should avoid using improper priors in practice: you will always be able to place some loose bounds on the values it can reasonably take.

Part II

Trouble Shooting

What could possibly go wrong?

Many things! But we will focus on the following two possibilities:

- MCMC starting state has zero probability.
- The chain runs, but one or more parameters mix slowly compared to the rest.

Zero probability starting state


```
Start likelihood: -Infinity after 1000 initialisation attempts
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

P(posterior) = -Infinity (was -Infinity) P(prior) = -Infinity (was -Infinity)
  P(BDMM) = -Infinity (was -Infinity)
    P(R0Prior) = -0.5586849541070393 (was -0.5586849541070393)
    P(rPrior) = -11.46042136866474 (was -11.46042136866474)
    P(rateMatrixPrior) = -0.14088025499381485 (was -0.14088025499381485)
    P(samplingProportionPrior) = -10.049507225748343 (was -10.049507225748343)
    P(becomeUninfectiousRatePrior) = -0.7811241751317991 (was -0.7811241751317991)
java.lang.RuntimeException: Could not find a proper state to initialise. Perhaps try another
seed.
at beast.core.MCMC.run(Unknown Source)
at beast.app.BeastMCMC.run(Unknown Source)
at beast.app.beastapp.BeastMain.<init>(Unknown Source)
at beast.app.beastapp.BeastMain.main(Unknown Source)
at beast.app.beastapp.BeastLauncher.main(Unknown Source)

Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

BEAST has terminated with an error. Please select QUIT from the menu.
```

Parameter prior is -Infinity

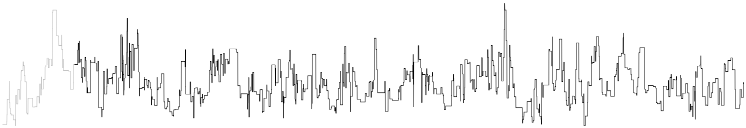
Example:

`P(rateMatrixPrior) = -Infinity (was -Infinity)`

Possible solutions:

- Do not change the seed!
  Instead: increase initialisation attempt number;
- Adjust initial conditions;
- Use excludable priors;
- Check for silly/incompatible priors;

One or more parameters mix slowly



Nothing mixed

Possible solutions:

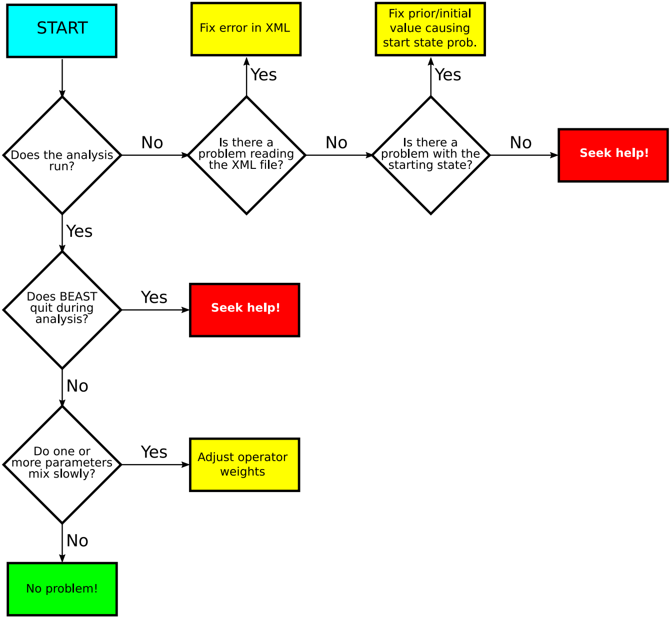
- Increase chain length;
- Run multiple independent chains;
- Increase sampling frequency (if ACT permits);
- Check identifiability;
- Check if model is misspecified.

One parameter did not mix

Possible solutions:

- Tweak the operator weights:
  - Increase weight for low ESS parameters;
  - Use UpDown operator for correlated parameters;
- Run longer (or combine several independent chains).

General problem-solving flow chart



Part III

Tutorials

Prior selection:
https://taming-the-beast.org/tutorials/Prior-selection/

Trouble-shooting:
https://taming-the-beast.org/tutorials/Troubleshooting/