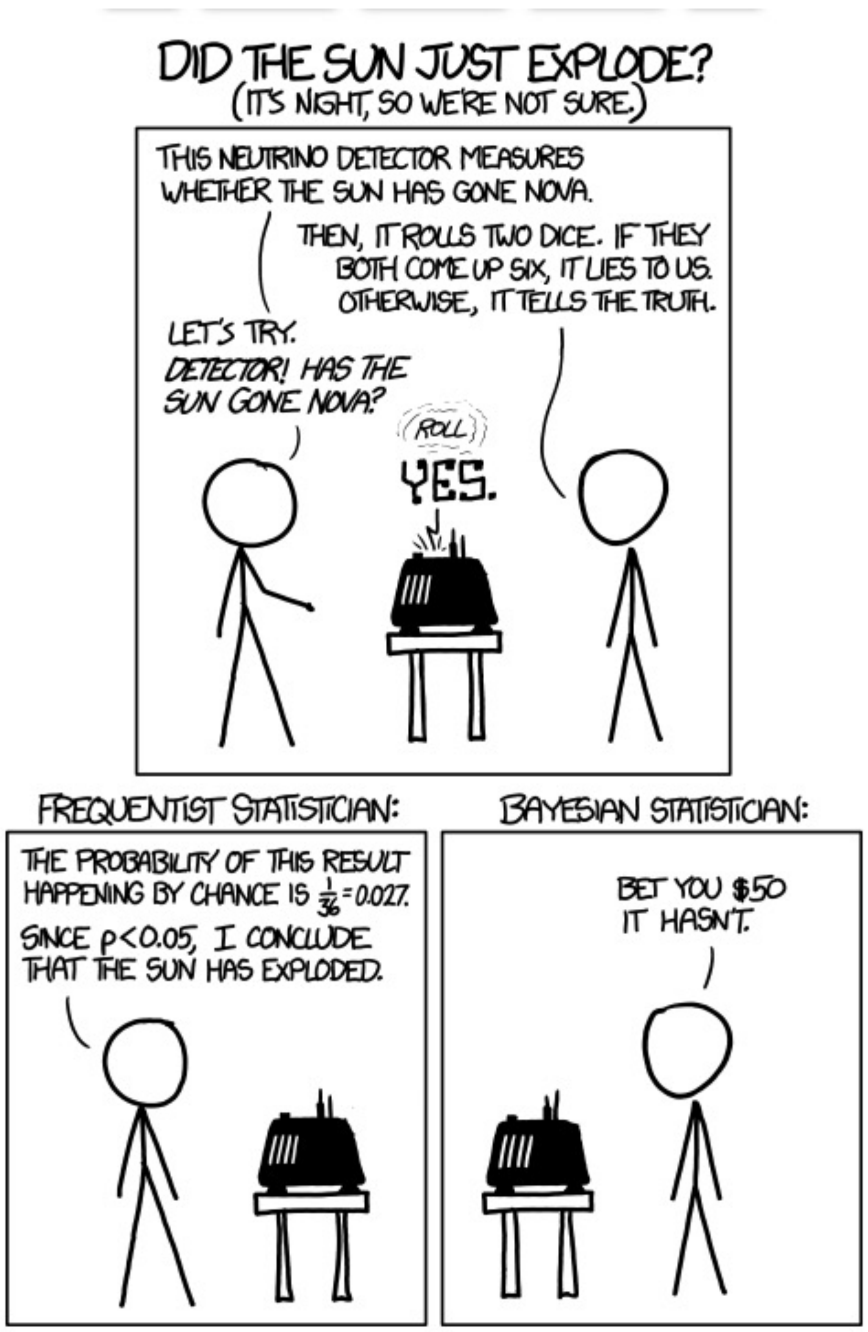
Bayesian Stats and MCMC



Frequentist approach

You have a null hypothesis and calculate the probability of observing your data under that hypothesis.

Null: The sun has not exploded

$$\Pr(det\ yes \mid sun\ is\ ok) = 0.027 = \frac{1}{6} \times \frac{1}{6}$$

this is less than the typical $\propto$ level of 0.05 so we reject the null that the sun is ok.

Bayesian approach

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$$

$$\Pr(nova|det\ yes) = \frac{\Pr(det\ yes|nova) \times \Pr(nova)}{\Pr(\det yes)}$$

$$\Pr(nova|det\ yes) << 0.00000000037\ \frac{1 \times 0.000000001}{0.027}$$

Bayesian approach

Imagine a disease present in 1 person per 100,000. If you take a test that correctly returns a positive result 99.9% of the time when someone is infected but has a false positive rate of 0.5%.

How likely are you to have the disease if you test positive?

Should you be concerned if you test positive?

Bayes' theorem provides a natural way to think about this.

Lets do this!

Bayesian approach

$$\Pr(inf|pos.test) = \frac{\Pr(pos.test|inf)\times\Pr(inf)}{\Pr(pos.test)}$$

$$\Pr(inf|pos.test) = 0.001 = \frac{0.999\times0.00001}{0.00501}$$

Which means that you have only a 0.1% chance of having the disease even if you test positive.

What if your doctor noticed a symptom that made them give you this test what changes?

Priors

$$\Pr(A|B) = \frac{\Pr(B|A) \times \boldsymbol{\Pr(A)}}{\Pr(B)}$$

Pr(A) this is the prior. It is powerful because it allows you to incorporate previous knowledge into your analysis, but it can lead to very bad inference if you are careless.

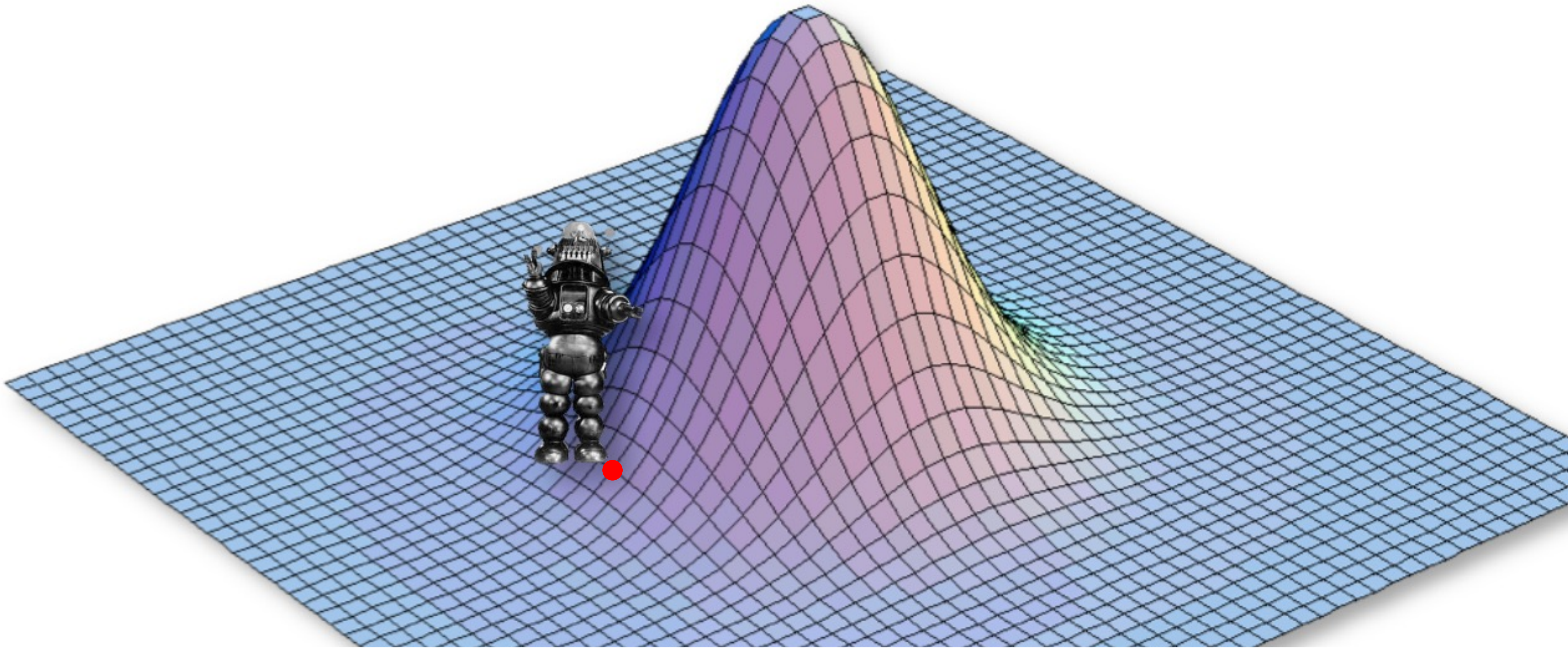
Bayesian methods in practice (MCMC)

In practice our models are quite complex with many parameters we would like to estimate. Each of these has its own prior distribution. The way that we explore the space of solutions is using an MCMC.

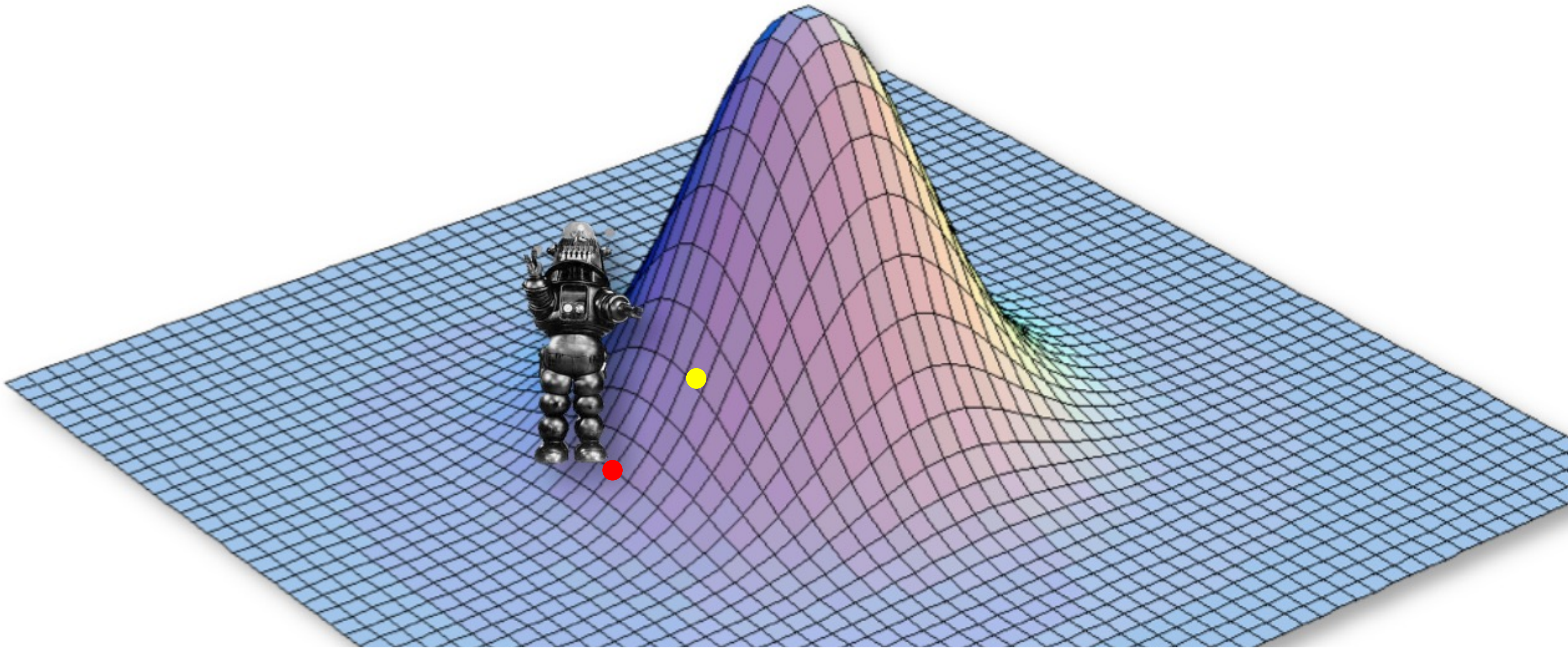
$$\Pr(A|B) = \frac{Pr(B|A) \times \Pr(A)}{Pr(B)}$$

1) Pick an arbitrary starting point
2) Calculate the current probability p_t
3) Generate a change to one of our parameters
4) Calculate the new probability p_{t+1}
5) Accept new parameter value with probability $\frac{p_{t+1}}{p_t}$
6) Return to step 3 (repeat for a really long time)

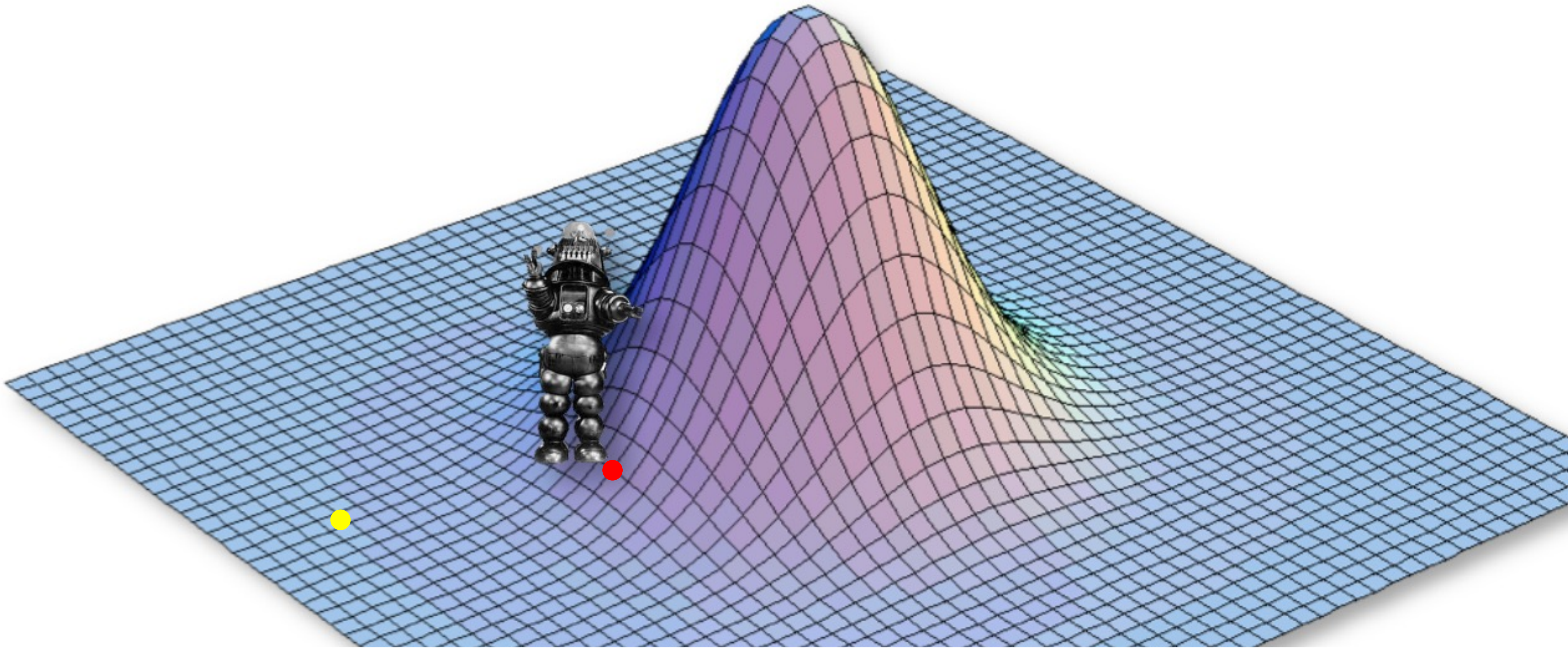
Bayesian methods in practice (MCMC)



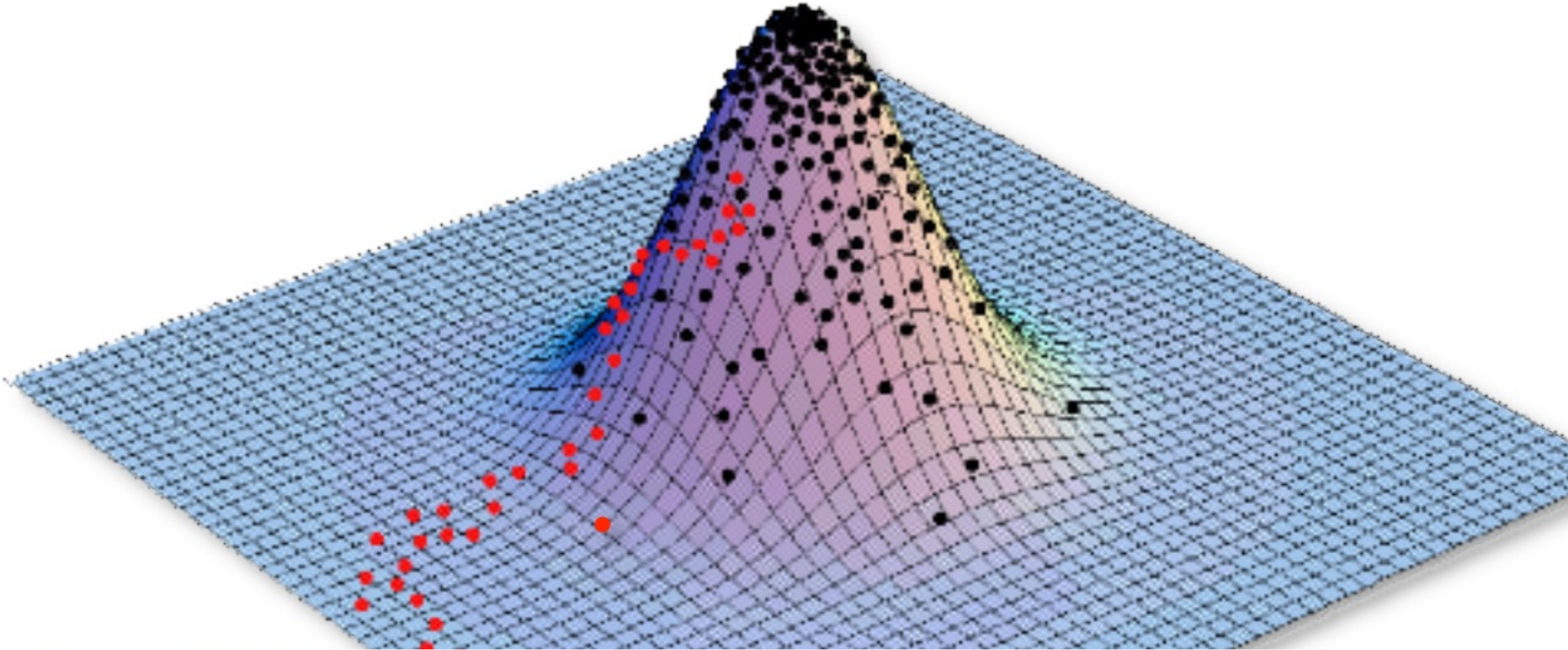
Bayesian methods in practice (MCMC)



Bayesian methods in practice (MCMC)



Bayesian methods in practice (MCMC)



Steps in a Bayesian analysis

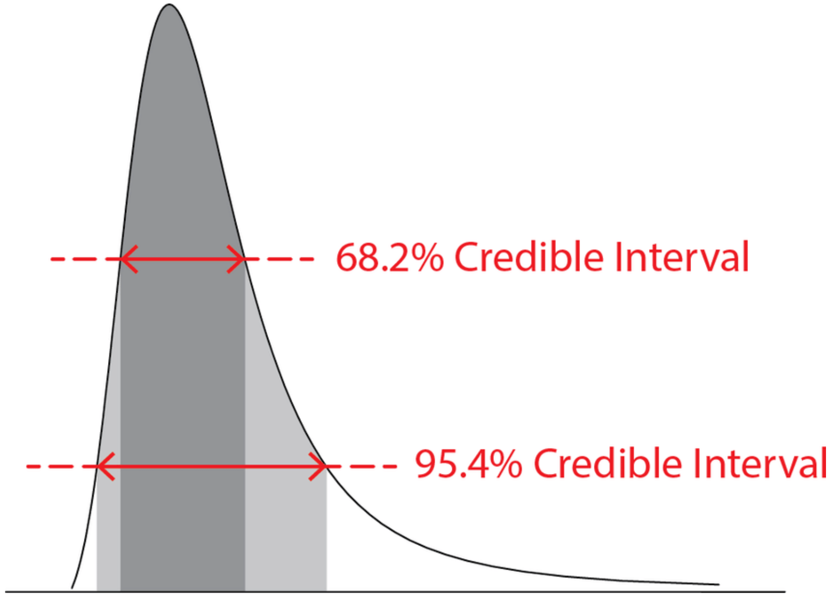
1. Choose appropriate priors to use for your parameters
2. Run your mcmc (How long? How many times?)
3. Once you are confident in your run remove burn-in
4. Check to make sure priors didn’t determine outcome.
5. Make estimates of parameters based on the post burn-in results
   - Use summaries appropriate for Bayesian methods.

Assessing Convergence

R example

Uncertainty

Credible Interval vs Confidence Interval



$$CI = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$$

Standard deviation

Mean value

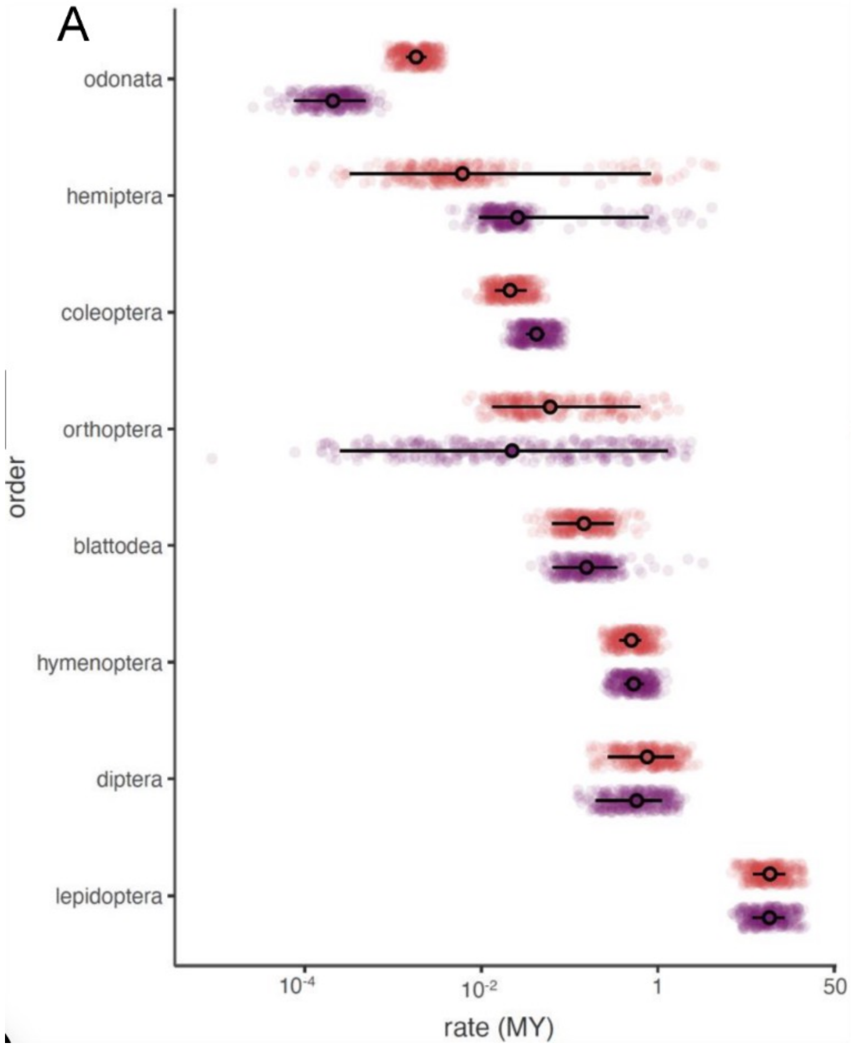
Lower/Upper limit

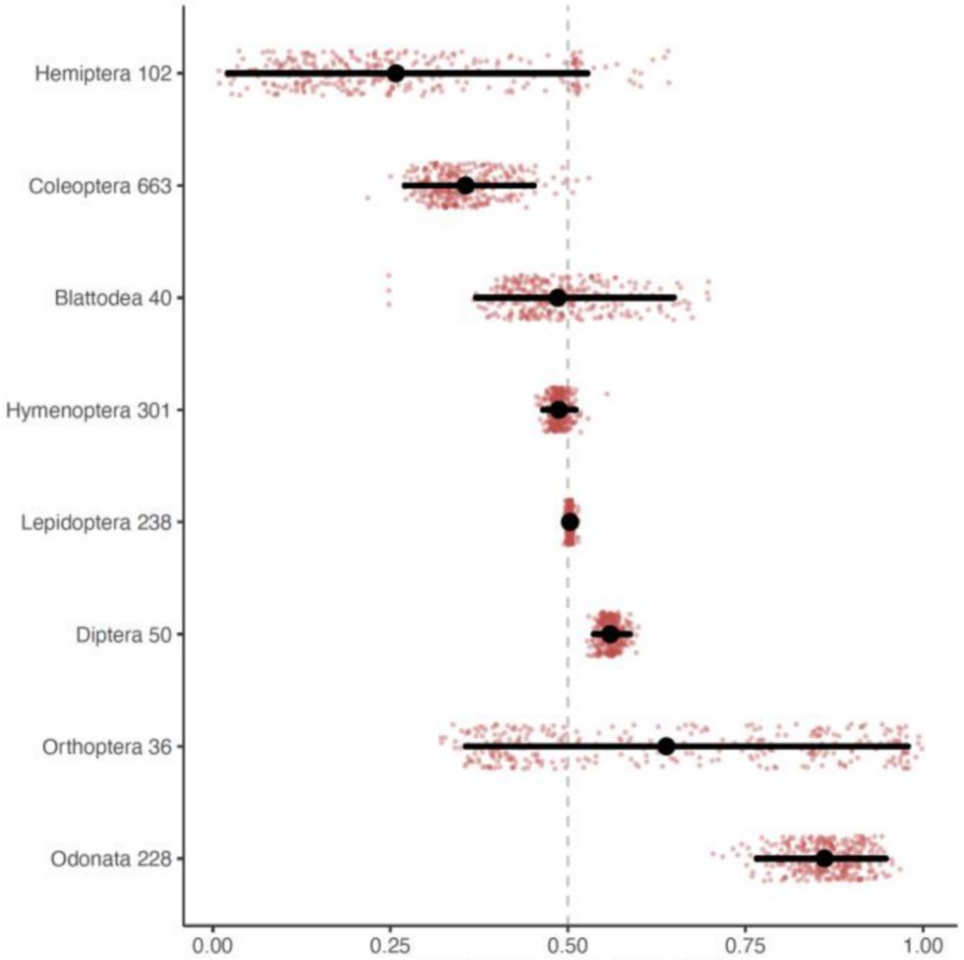
z-value for the confidence level

Sample size

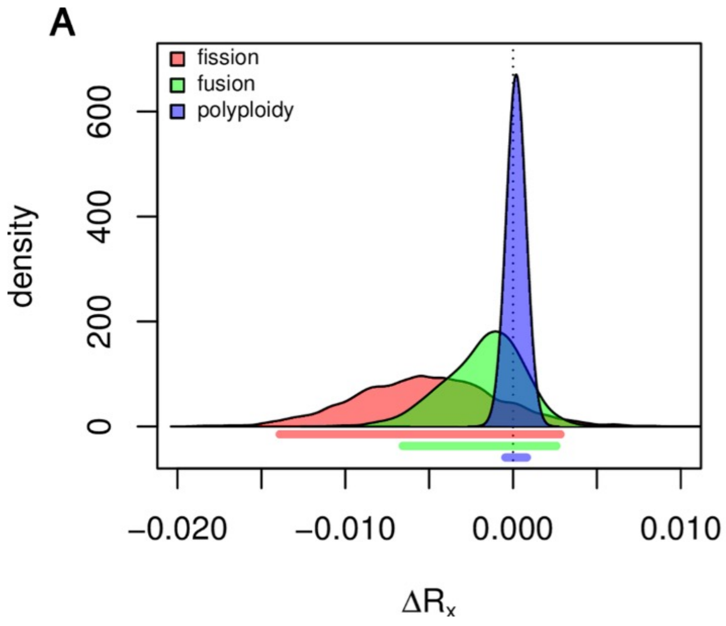
R example

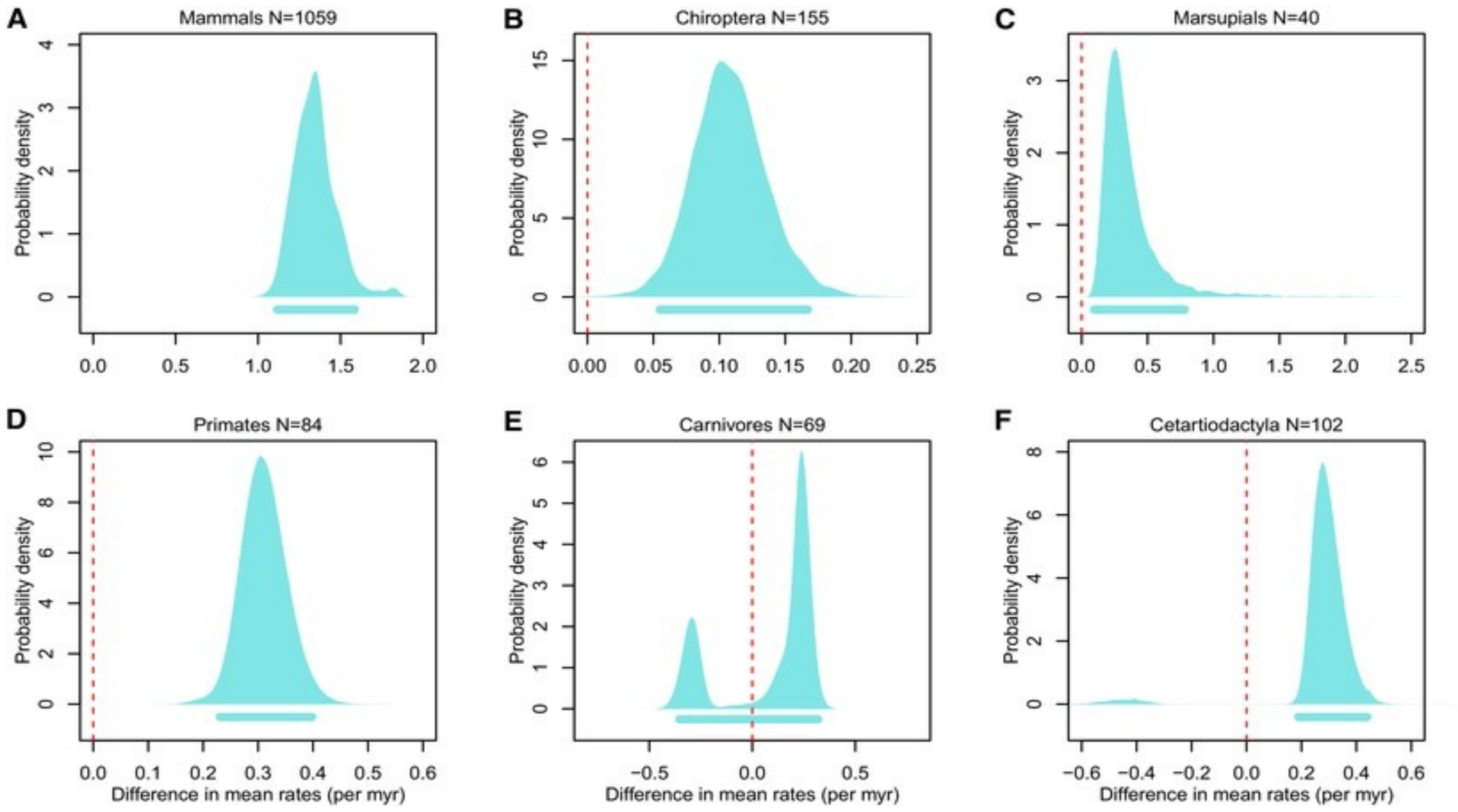
Inferences from posterior distributions



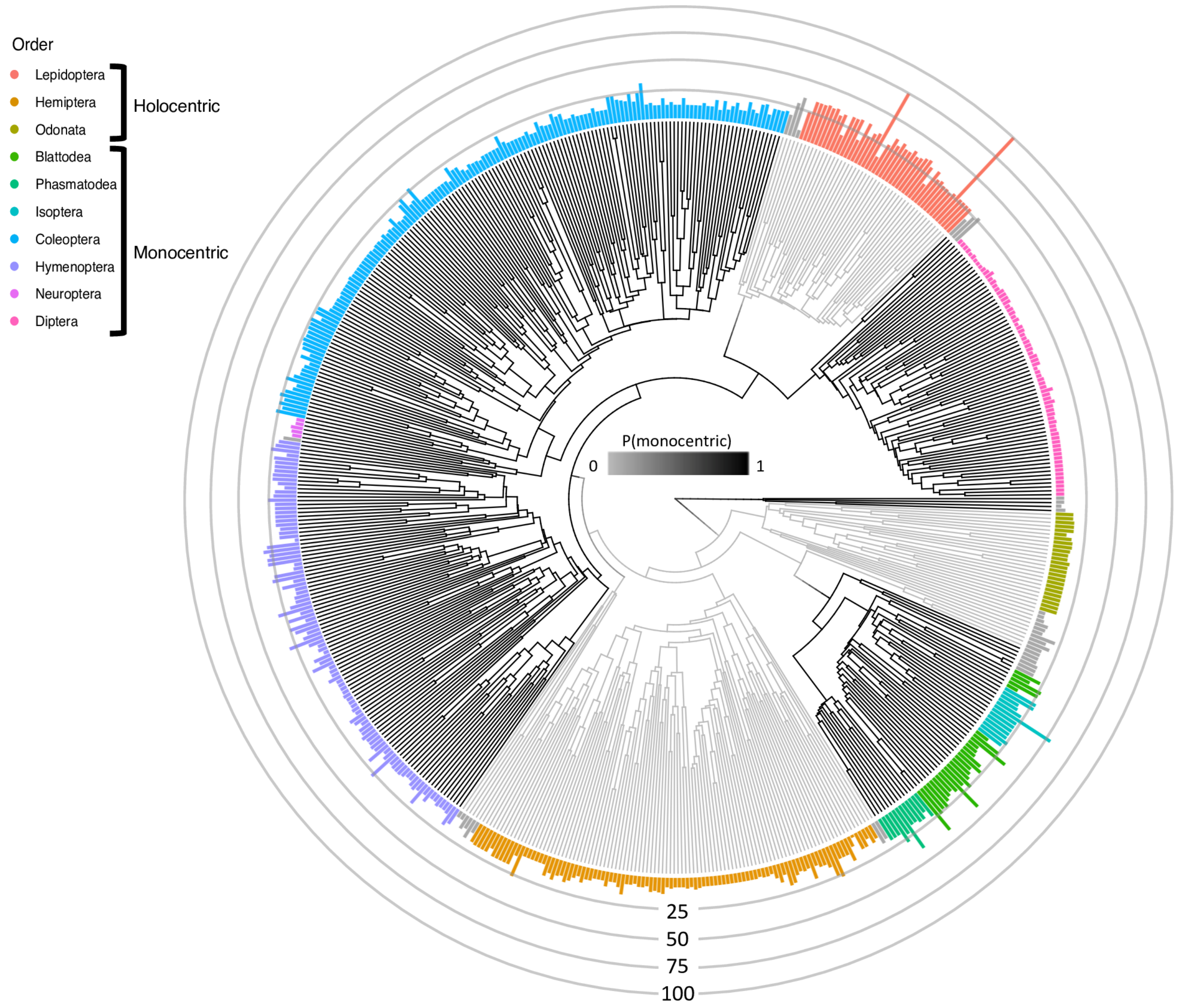


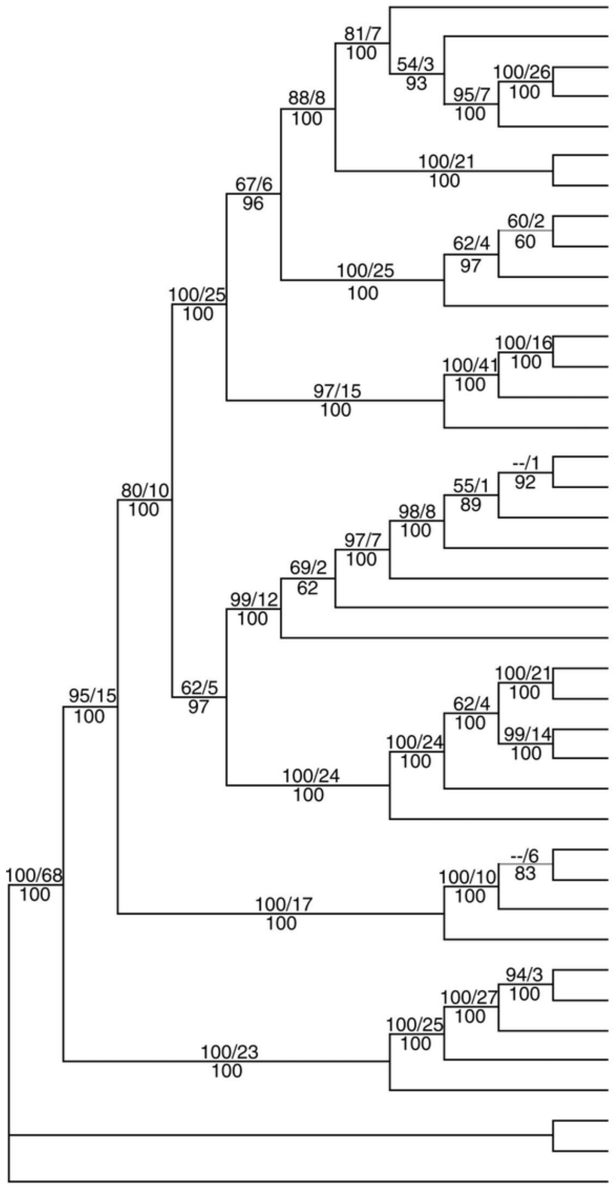
Inferences from posterior distributions



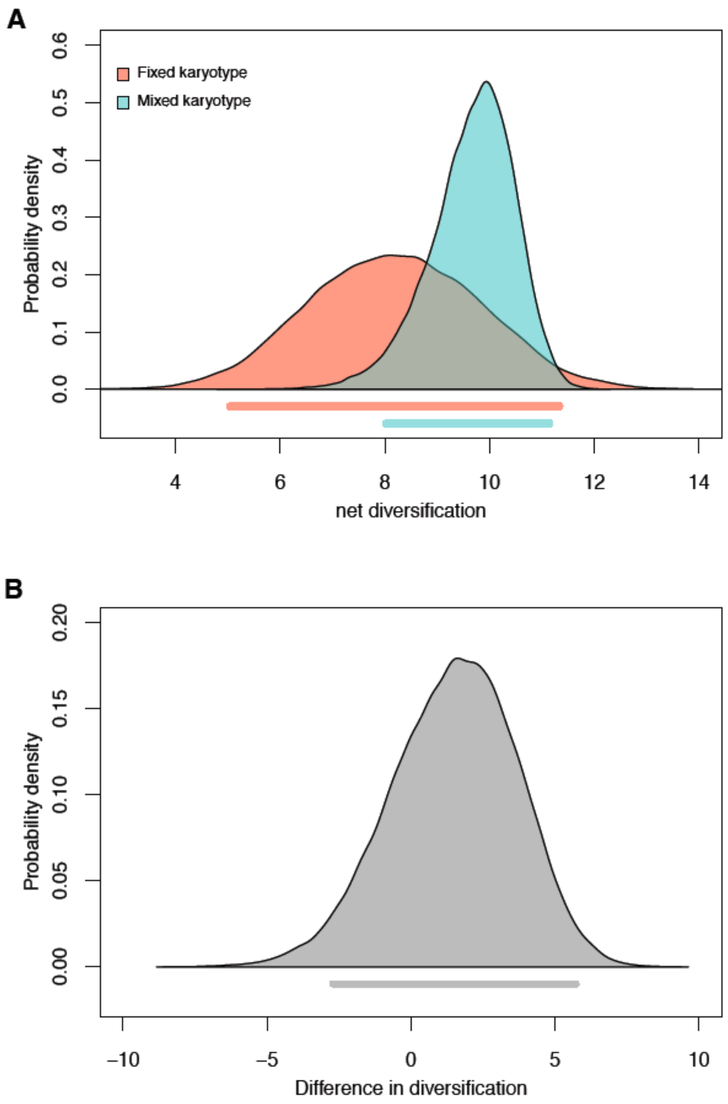


Inferences from posterior distributions





The power of Bayesian approaches



With a Bayesian approach we can take into account all types of uncertainty and be more conservative.

- Phylogenetic
- Model selection uncertainty
- Parameter value uncertainty
- Uncertainty in measurements