Model Selection

UofSC Python STEM workshop

AIC / BIC

Bayesian Evidence

Some questions we should ask when fitting a model to data

Does this model **match** my data ?

Can this model help to **explain** my data ?

Can this model **predict** future / different data ?

Is this the **best** model for my data?

Some questions we should ask when fitting a model to data

Does this model **match** my data ?

Goodness of fit tests (e.g. chi^2, p-values)

Can this model help to **explain** my data ?

Model validity : apply Physics / Chemistry / Biology / Sociology / ...

Can this model **predict** future / different data ?

Cross-validation ; New experimentation

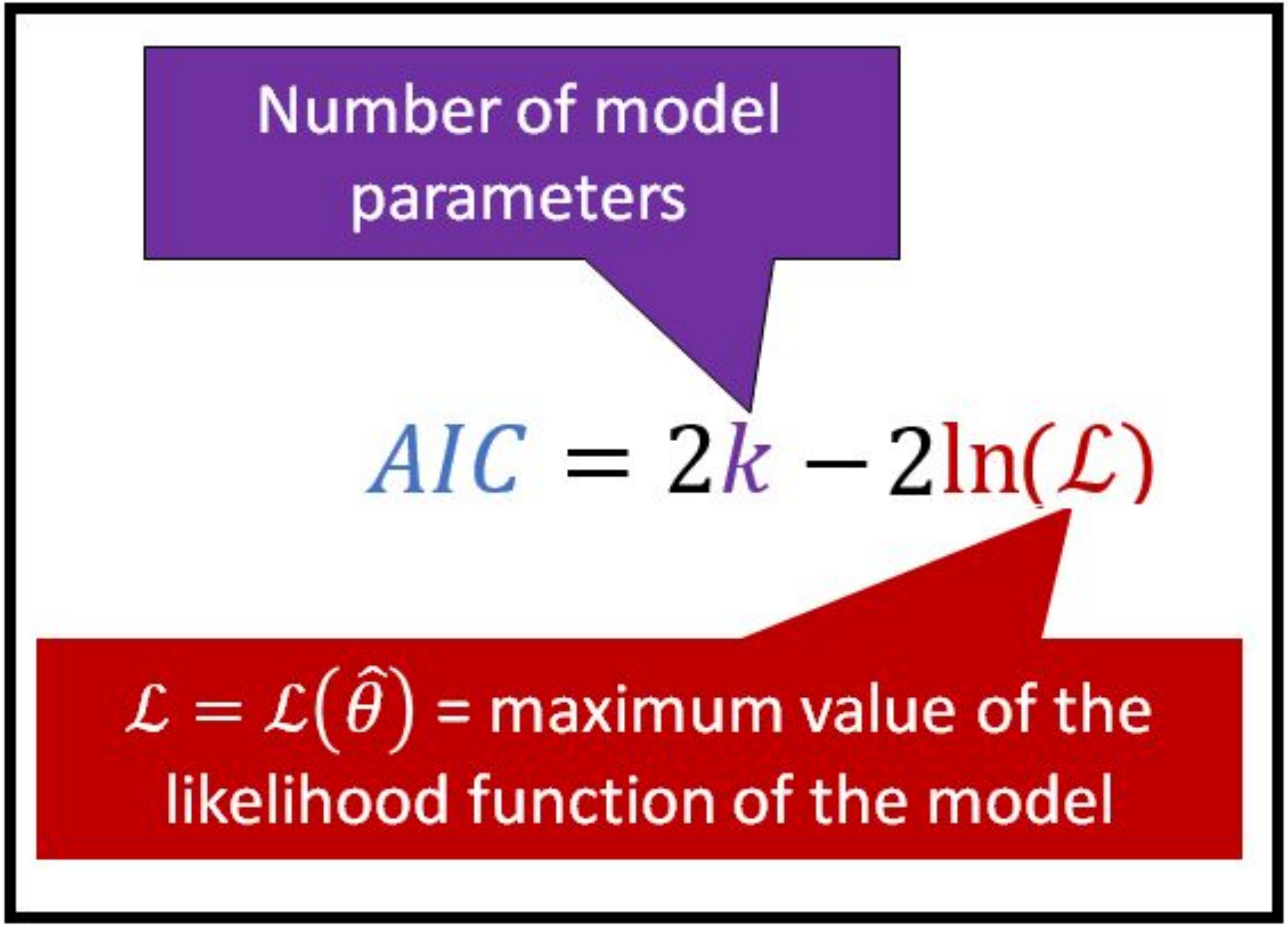
Is this the **best** model for my data?

“Best” may mean: good match + least complex

Model selection (e.g. AIC, BIC, Bayesian evidence / Bayes factors)

Alternative : use all the models!
(Weighted model combination, Bayesian model averaging)

The Akaike Information Criterion



Rewriting the AIC

- As model becomes more complex, the numerator grows.
- As max likelihood increases, the denominator grows

Number of model parameters

$$AIC = 2\ln\left(\frac{e^{k}}{\hat{\mathcal{L}}}\right)$$

Maximized Likelihood

To compare two models, use the difference of the AIC from each.

Reframed as an exponent, this gives the relative likelihood of model 1 to model 2. (can also be phrased as ‘model weights’ or ‘odds ratios’)

$$\boxed{Relative\ likelihood = \mathrm{e}^{(AIC_1 - AIC_2)/2}}$$

The Bayesian Information Criterion

Number of model parameters

Number of data points

$$BIC = k \ln(n) - 2\ln(\mathcal{L})$$

$\mathcal{L} = \mathcal{L}(\hat{\theta})$ = maximum value of the likelihood function of the model

Model Selection with Bayesian Evidence

Bayesian evidence (a.k.a. Bayes factors comparison)

The AIC and BIC use only the maximum likelihood values - not the full likelihood (or posterior) distribution.

These are therefore useful only when the likelihood is sharply single-peaked.

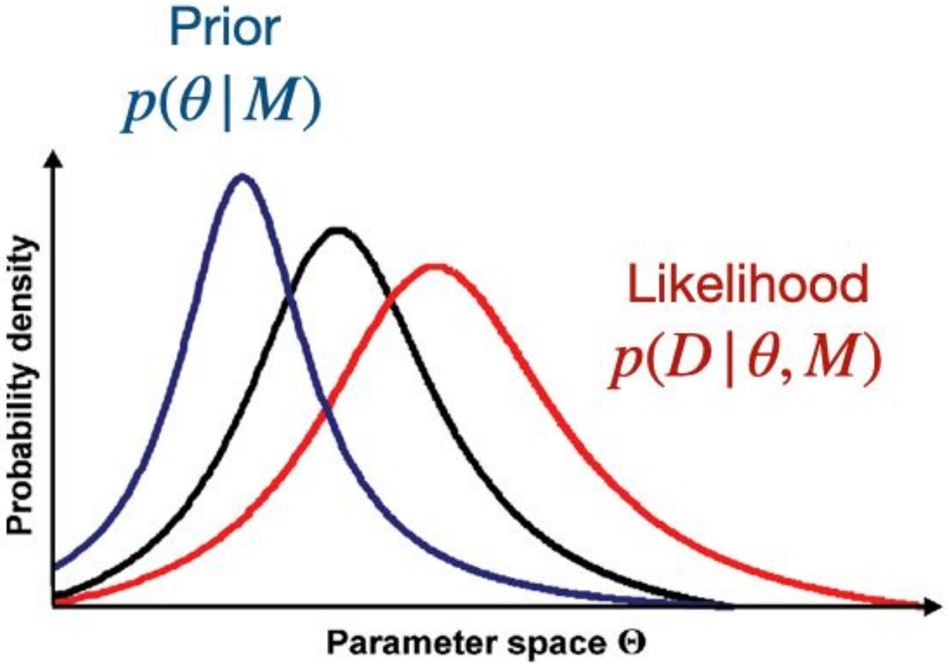
When the likelihood is more complex, or when you have strong informative priors, you should use a model comparison tool that incorporates that information, such as the Bayesian evidence (Bayes factors).

Bayes theorem :

$$\overbrace{P(A|B)}^{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}} * \overbrace{P(A)}^{\text{Prior}}}{\underbrace{P(B)}_{\text{Evidence}}}$$

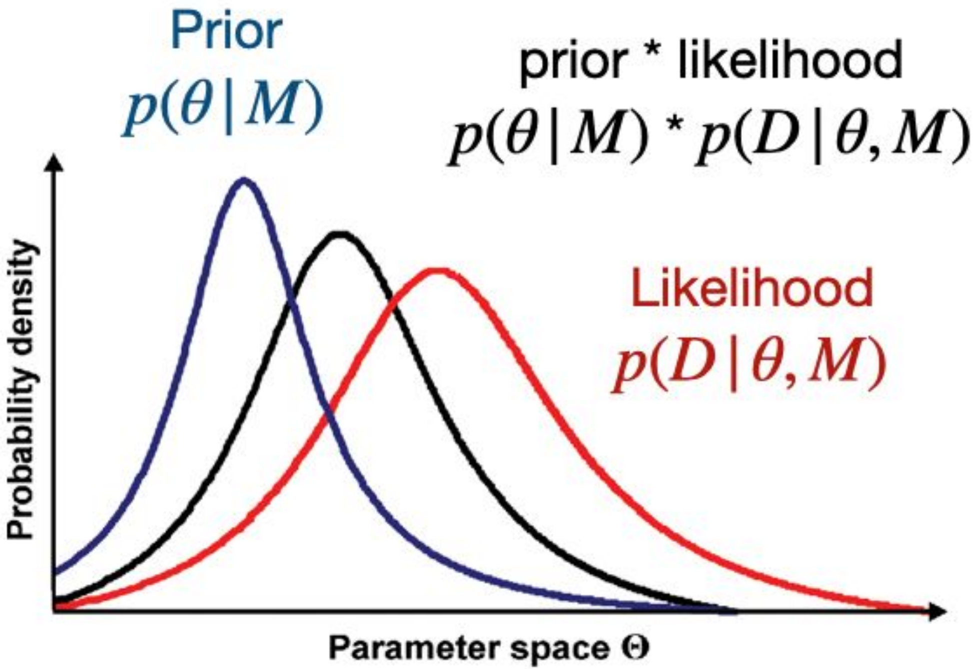
The denominator is the "Bayesian Evidence"

found by integrating the likelihood * prior over all parameter space.



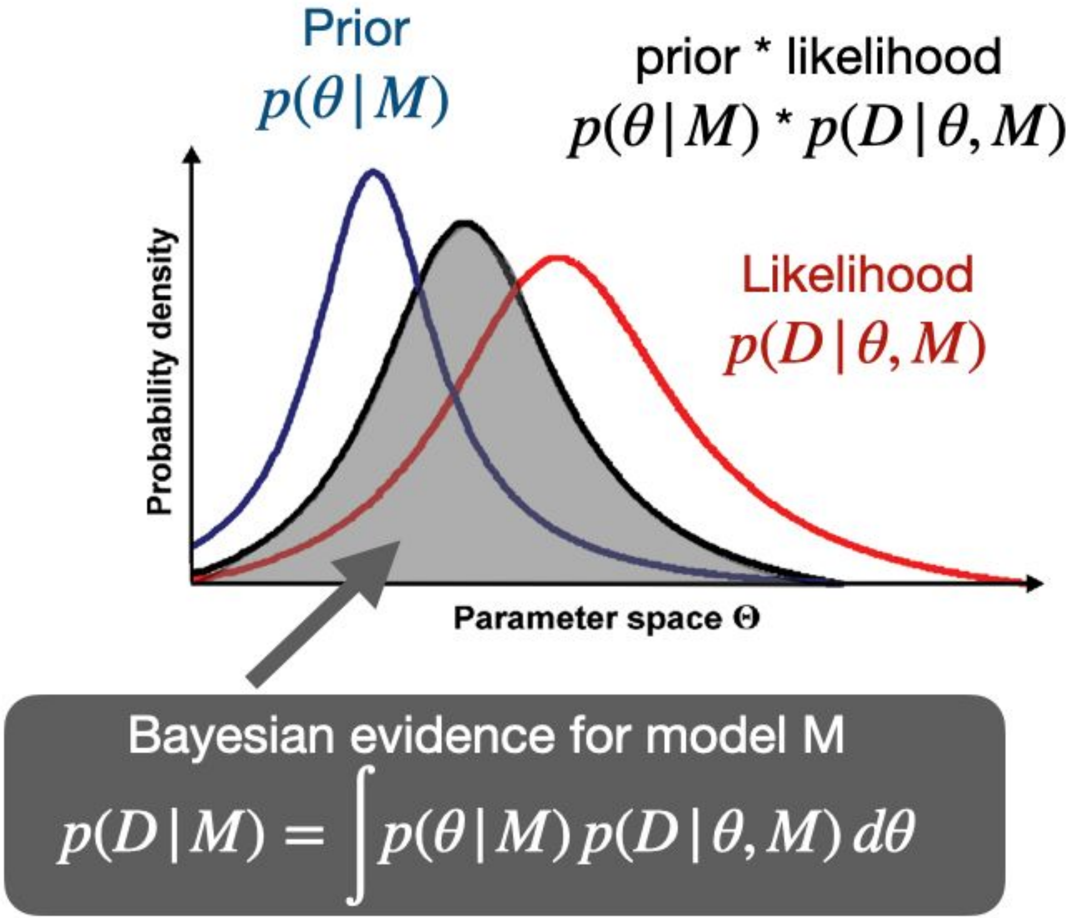
The denominator is the “Bayesian Evidence”

found by integrating the likelihood * prior over all parameter space.



The denominator is the "Bayesian Evidence"

found by integrating the likelihood * prior over all parameter space.



The Bayes Factor, K, is simply a ratio of Bayesian evidence from two models, compared to the same data

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)}$$

The Bayes Factor, K, is simply a ratio of Bayesian evidence from two models, compared to the same data

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\int \Pr(\theta_1|M_1)\Pr(D|\theta_1, M_1)\,d\theta_1}{\int \Pr(\theta_2|M_2)\Pr(D|\theta_2, M_2)\,d\theta_2} = \frac{\frac{\Pr(M_1|D)\Pr(D)}{\Pr(M_1)}}{\frac{\Pr(M_2|D)\Pr(D)}{\Pr(M_2)}} = \frac{\Pr(M_1|D)}{\Pr(M_2|D)}\frac{\Pr(M_2)}{\Pr(M_1)}$$

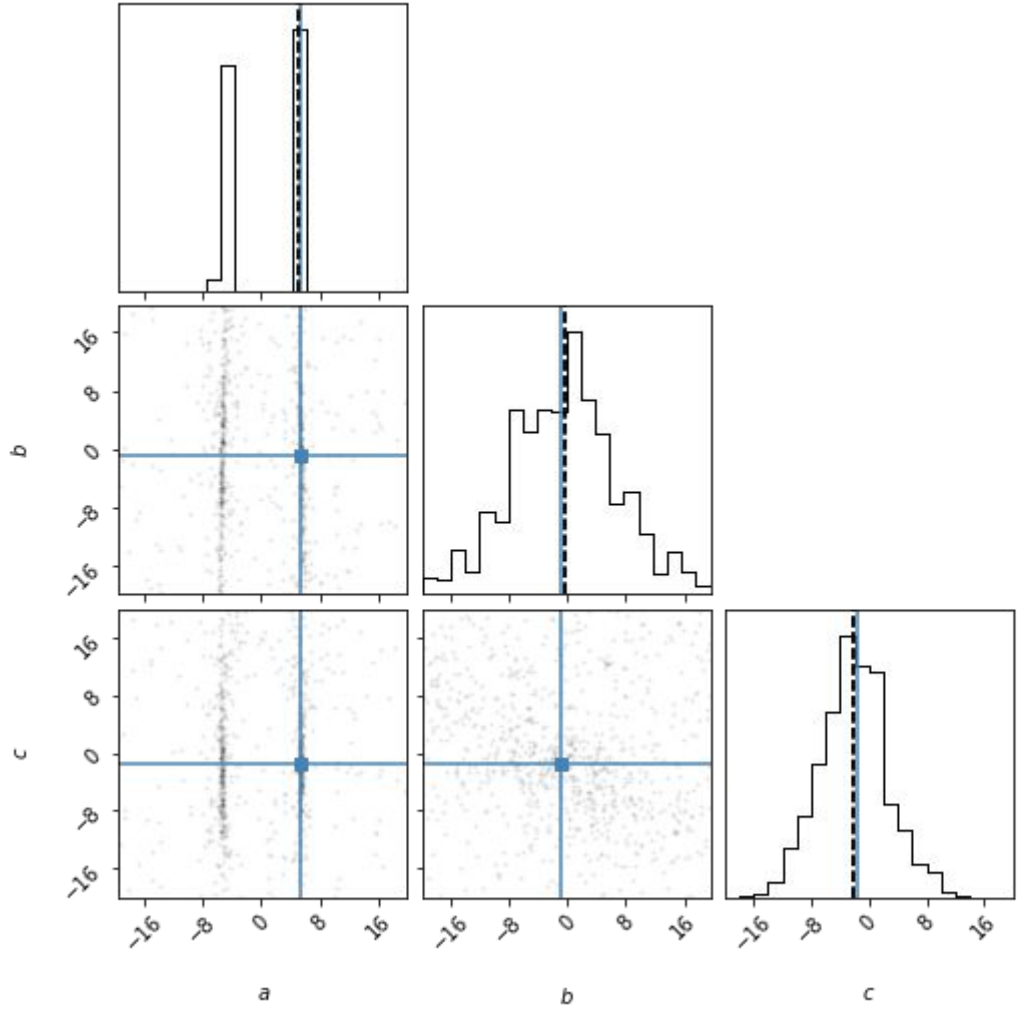
How to compute the Bayesian evidence

- Define your priors carefully
- Use a tool like MCMC or Nested Sampling to measure the likelihood over all of parameter space
- Approximate the integral from the set of likelihood measurements recorded in your MCMC sampling chain

Some beautiful and instructive MCMC algorithm visualizations:

http://chi-feng.github.io/mcmc-demo

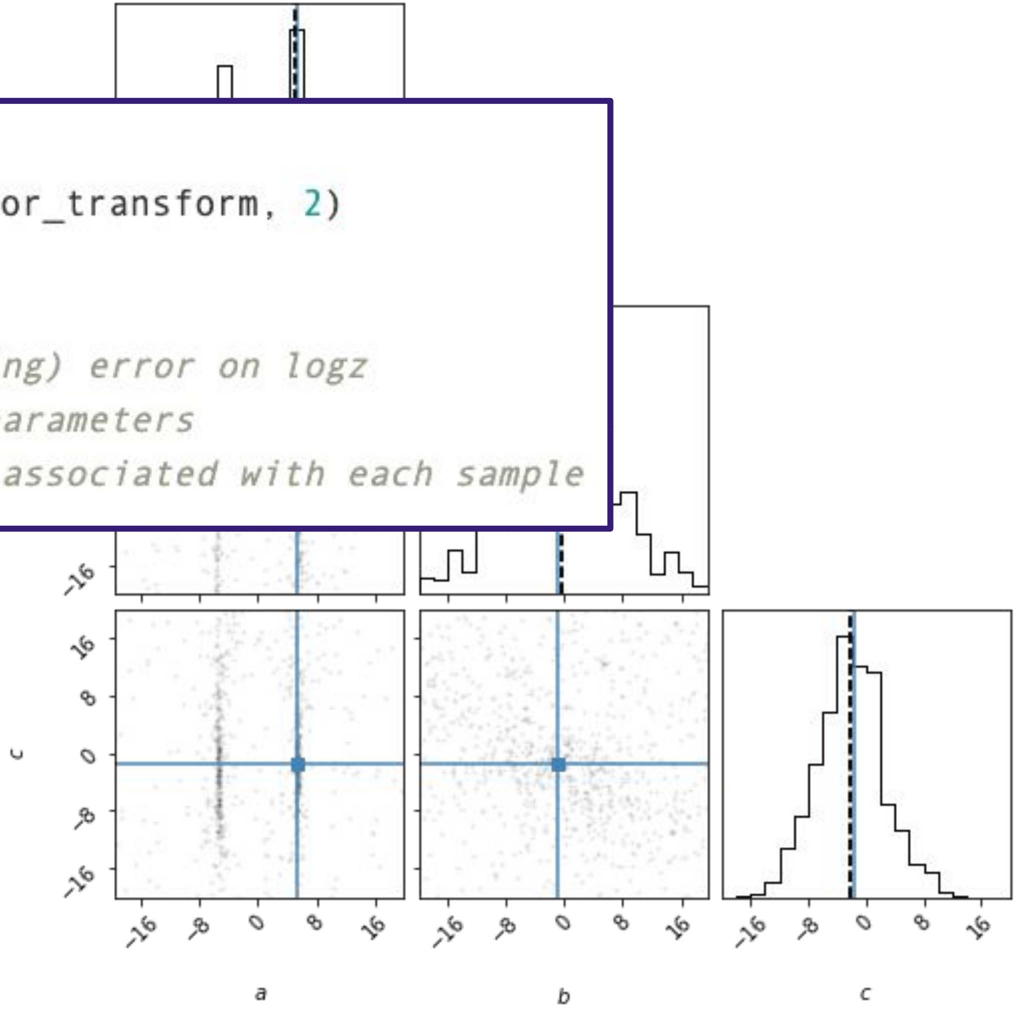
The ‘nestle’ package used in our nested sampling notebook returns both the array of log(likelihood) values and the integral, the log(evidence).



Nestle results:

```
# Run nested sampling.
result = nestle.sample(loglike, prior_transform, 2)

result.logz      # log evidence
result.logzerr   # numerical (sampling) error on logz
result.samples   # array of sample parameters
result.weights   # array of weights associated with each sample
```



Resources

AIC Article by Sachin Date on the Medium blog “TowardsDataScience”

The emcee and Nestle packages for MCMC and Nested Sampling to measure Bayesian evidence

Book : Sivia, D. and Skilling, J. "Data Analysis: A Bayesian Tutorial"

Some “tutorial” papers:

- On AIC / BIC :
  - Wagenmakers and Farrell 2004
  - Symonds and Moussalli 2010
- Bayesian model averaging
  - Hoeting, Madigan, Raftery & Volinsky 2010