





OK so far, you have a model and you can generate predictions given some parameters:

We can adopt a **qualitative** approach:

- Can my model reproduce some particular aspects of my dataset? (e.g. a positive relationship between two quantities, a hump-shaped trend for some variable...)

- For what range of parameters does this happen?



Livingston et al (2012) *Nat. Comm.* (spatial coexistence of two competitors) We can adopt a **qualitative** approach:

- Can my model reproduce some particular aspects of my dataset? (e.g. a positive relationship between two quantities, a hump-shaped trend for some variable...)

- For what range of parameters does this happen?

Generalized models/qualitative models can be useful here (e.g. Yeakel et al. 2011 Theor. Ecol.)



FIGS. 1, 2, and 3 show, respectively, the contours of motion for the conservative razor's-edge case, the case recognizing simple diminishng returns, and the case where returns are increasing at intermediate scales near the equilibrium.

We can adopt a **qualitative** approach:

- Can my model reproduce some particular aspects of my dataset? (e.g. a positive relationship between two quantities, a hump-shaped trend for some variable...)

- For what range of parameters does this happen?

Generalized models/qualitative models can be useful here (e.g. Yeakel et al. 2011 Theor. Ecol.)

We can also adopt a more quantitative approach (stats)

MODEL	DATASET



Note that in this context, the 'model' is intended as a blackbox, it can thus be any type of model:

- a static model (classical 'statistical models" are good examples: regressions, ANOVAs...)
- dynamical models, mathematical or simulatory
- artificial intelligence models such as neural networks
- ...

For simplicity, we can decompose the process into four steps of increasing ambition:



Major problems:

- we may obtain **biased** estimates
- the model may not be **identifiable**





#### Common approaches:

- minimize some distance between predictions and data (e.g. least squares)
- maximize likelihood
- maximize **posterior probability** (Bayesian approaches)

Least squares:

- + flexible, robust enough, very fast minimization techniques
- + equivalent to ML under certain assumptions\*
- not applicable to all models
- somewhat ad-hoc: other distances could be used (absolute differences...)

Maximum likelihood:

- + fully general, intuitive, solid theoretical grounding
- + consistent (asymptotically unbiased... if model is true)
- can be hard to compute and maximize



IX. On the Mathematical Foundations of Theoretical Statistics.

By R. A. FISHER, M.A., Fellow of Gonville and Caius College, Cambridge, Chief Statistician, Rothamsted Experimental Station, Harpenden.

Communicated by Dr. E. J. RUSSELL, F.R.S.

Received June 25,-Read November 17, 1921.

Least squares:

- + flexible, robust enough, very fast minimization techniques
- + equivalent to ML under certain assumptions\*
- not applicable to all models
- somewhat ad-hoc: other distances could be used (absolute differences...)

#### Maximum likelihood:

- + fully general, intuitive, solid theoretical grounding
- + consistent (asymptotically unbiased... if model is true)
- can be hard to compute and maximize

Posterior probability (Bayesian approaches):

- + supplements ML with **prior knowledge** on parameter values
- + efficient sampling algorithms (priors guide the exploration of parameter space)
- + prior distributions can alleviate non-identifiability issues
- supplements ML with **prior knowledge** on parameter values
- can be slow to converge

Bayes (1763) Phil. Trans. Royal Soc. LII. An Effay towards folving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. Dear Sir, Read Dec. 23, I Now fend you an effay which I have 1763. I found among the papers of our deceafed friend Mr. Bayes, and which, in my opinion,

# What do you do if you cannot compute the likelihood?

Imagine you can roll a dice but you do not know whether it is fair or not. What is the probability of getting a 2?



Roll it many times and count outcomes!

# What do you do if you cannot compute the likelihood?

#### ABC: approximate Bayesian Computation

Approximate likelihood by comparing simulation outputs to data based on some distance metrics

Approximate Bayesian Computation in Evolution and Ecology

Mark A. Beaumont

Department of Mathematics and School of Biological Sciences, University of Bristol, Bristol BS8 1TNW, United Kingdom; email: m.beaumont@bristol.ac.uk

Beaumont (2010) AREES

**STEP 2**: how to quantify the uncertainty in parameter estimates, and the quality of the fit (goodness of fit)?

#### Major problems:

- we want to construct good **intervals** around parameter estimates: there are many of them
- with not so much data, we may have very little **statistical power** to evaluate the goodness of the fit (and failing to reject is not accepting)

**STEP 2**: how to quantify the uncertainty in parameter estimates, and the quality of the fit (goodness of fit)?

#### Common approaches (uncertainty):

- resample dataset/refit model (bootstrap, jacknife...)
- use likelihood surface theory to get confidence intervals (Fisher information...)
- use Bayesian approaches to compute credible intervals

**STEP 2**: how to quantify the uncertainty in parameter estimates, and the quality of the fit (goodness of fit)?





STEP 3: how to compare different models together and select the 'best' (model selection)?

### Major problems:

- a more complex model will always fit the data better but...
- the **bias/variance trade-off**, or the **curse of dimensionality**: for a given amount of data, too simple a model will have little variance/high bias (**underfitting**), too complex a model will have low bias/huge variance (**overfitting**)
- in both cases we have poor estimation of parameters (total uncertainty = bias<sup>2</sup> + variance)
- in both cases, we'll have poor prediction power for future/other datasets
- we must find some intermediate level of complexity, i.e. make some compromise

STEP 3: how to compare different models together and select the 'best' (model selection)?



Model dimensionality (# parameters)



Model dimensionality (# parameters)

STEP 3: how to compare different models together and select the 'best' (model selection)?

### Common approaches:

- Split dataset (cross validation, training/test datasets...)
- is a particular parameter 'significant' or not? (model simplification)
- Likelihood ratio tests
- Information criteria for model comparison (AIC...)
- Regularization or penalization techniques (ridge regression, LASSO...)

#### Tredennick et al. (2021) Ecology

STEP 4: how to use several alternative models rather than just one (multimodel inference)?

#### Major problems:

- How to reduce model selection bias (see Freedman's paradox)?
- How to include model selection uncertainty?

- How to combine the estimates or predictions from different competing models, and combine them in an optimal way?

STEP 4: how to use several alternative models rather than just one (multimodel inference)?

#### Common approaches:

- Take all models and do some ad-hoc consensus (e.g. average or median prediction)
- Use **model-averaging** and ensemble techniques



Dormann et al. (2018) *Ecological Monographs* 

**SAR**: Guilhaumon et al. (2008) PNAS **TPC**: Padfield et al. (2021) *Methods in Ecology & Evolution* 

STEP 4: how to use several alternative models rather than just one (multimodel inference)?

#### Common approaches:

- Take all models and do some ad-hoc consensus (e.g. average or median prediction)

- Use **model-averaging** and ensemble techniques



Abboud et al. (2023) J. Math. Biol. (invasion of a plant pathogen)



PLOS COMPUTATIONAL BIOLOGY

#### EDUCATION

## Twelve quick tips for designing sound dynamical models for bioprocesses

Francis Mairet<sup>1</sup>, Olivier Bernard<sup>2,3,4</sup>\*

Mairet & Bernard (2019) PLoS Comput. Biol.



## **Tonight's guest lecture:**



Pr. Sarah Otto Department of Zoology University of British Columbia Vancouver VOL. 195, NO. 2 THE AMERICAN NATURALIST FEBRUARY 2020

SYMPOSIUM

## Theory in Service of Narratives in Evolution and Ecology\*

#### Sarah P. Otto<sup>†</sup> and Alirio Rosales

Department of Zoology and Biodiversity Research Centre, University of British Columbia

Otto & Day (2007)





