ENGG 107: Bayesian Statistical Modeling and Computation

The Bayesian Approach (II)

Meeting #3

Objectives for today:

1. Recap of the introduction to the Bayesian approach
2. Refining our understanding of the Bayesian approach

Draft Schedule

Week	Week starting	Class 1	Class 2	Class 3	Notes	Assignment out	Assignment due
1	January 2, 2023	No Class	Introduction to the class	Bayesian Approach I			none
2	January 9, 2023	Bayesian Approach II	Picking a project	Setting up Computation		1	none
3	January 16, 2023	No class (MLK day)	Analytical Solutions / Kalman	Precalibration /(Bayes) Monte Carlo		2	1
4	January 23, 2023	MCMC Part 1	MCMC Part 2	Bayesian Workflow	X hour class	3	2
5	January 30, 2023	Students pitch project ideas	Writing a method section	Catching up / review		4	3
6	February 6, 2023	Convergence diagnostics	Checking Assumptions	Deep Uncertainty		5	4
7	February 13, 2023	Geeting Solid Priors	Links to Decision-Making	Links to Decision-Making		6	5
8	February 20, 2023	Model Choice	Emulation	Sensitivity analysis		7	6
9	February 27, 2023	Communication of results	Student Presentations	Student Presentations		8	7
10	March 6, 2023	Class Debriefing / Resarch links	No Class	No Class		none	8

Review

1. What is probability?
2. What is a Bayesian approach?
3. What is the Bayes equation and how can you derive it?
4. What are key differences between a Bayesian and a Frequentist approach?
5. How to define and use (in a simple example)
    a. prior
    b. likelihood
    c. probability

Outline for today

1. Review of reading assignment
2. Why may we need a Bayesian approach?
3. How does a Bayesian approach work?
   a. prior
   b. likelihood
   c. posterior
   d. Bayesian updating
4. The pesky aspects of computation

Outline for today

1. Review of reading assignment
2. Why may we need a Bayesian approach?
3. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
4. The pesky aspects of computation

Review of Reading Assignments

- Core
  - van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. Nature Reviews Methods Primers, 1(1), 1–26. https://doi.org/10.1038/s43586-020-00001-2
- Background (if you are interested)
  - Freedman, D. (1997). Some Issues in the Foundation of Statistics. In B. C. van Fraassen (Ed.), Topics in the Foundation of Statistics (pp. 19–39). Springer Netherlands. https://doi.org/10.1007/978-94-015-8816-4_4
  - Jefferys, W. H., & Berger, J. O. (1992). Ockham’s razor and Bayesian analysis. American Scientist, 80(1), 64–72. https://www.jstor.org/stable/pdf/29774559.pdf

Reading Review Questions

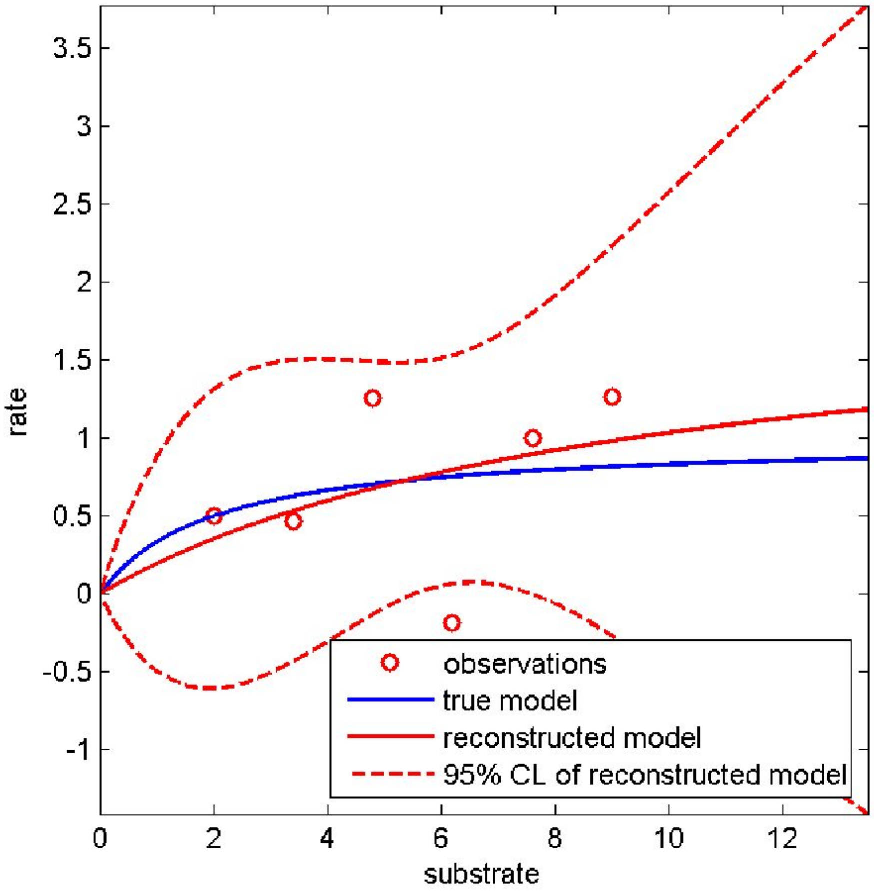
1. What are the main points?
2. What remains unclear?
3. What questions does this raise?
4. What does this mean for your research / project design?

Outline for today

1. Review of reading assignment
2. Why may we need a Bayesian approach?
3. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
4. The pesky aspects of computation

Challenge: We know that negative parameters and rates are nonsense.

How can we use this independent information?



Motivation for Bayesian Analyses

- You run an experiment and estimate decay rates. Some of the confidence limits for the decay rates derived from your frequentist software include negative values.
- You run an experiment to estimate diffusion rates. Again, some confidence limits for the decay rates derived from your frequentist software include negative value
- Your physical knowledge tells you that this is impossible.
- How can you represent this physical knowledge in a statistically sound way?
- Bayes to the rescue

"Thomas Bayes [..] c. 1701 – 7 April 1761 [..] was an English statistician, philosopher and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: Bayes' theorem."

Credit: Picture: https://en.wikipedia.org/wiki/Thomas_Bayes#/media/File:Thomas_Bayes.gif, public domain, accessed Dec 28 2022, quote: https://en.wikipedia.org/wiki/Thomas_Bayes, accessed Dec 28 2022

Frequentist vs. Bayesian Differences

TABLE 1. Some fundamental differences between frequentist and Bayesian statistical inference in their uses and interpretations of statistical concepts and terms.

Concept or term	Frequentist interpretation	Bayesian interpretation
Probability	Result of an infinite series of trials conducted under identical conditions	The observer's degree of belief, or the organized appraisal in light of the data
Data	Random (representative) sample	Fixed (all there is)
Parameters	Fixed	Random
$k\%$ confidence interval	This interval will include the true value of a given parameter in $k\%$ of all possible samples	$k\%$ of the possible parameter values will fall within the confidence (credibility) interval
Treatment of nuisance parameters	Conditions on sufficient statistics or maximum likelihood estimate	Integrates over all possible values
Conclusion	$P(x \mid H)$	$P(H \mid x)$

Ellison, A. M. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications: A Publication of the Ecological Society of America*, *6*(4), 1036–1046. https://doi.org/10.2307/2269588

Balancing Errors

"Howard Raiffa (1968, p. 264) noted that statistics students learn the importance of constantly balancing making an **error of the first kind (that is, rejecting the null hypothesis when it is true)** and an **error of the second kind, that is, accepting the null hypothesis when it is false** (Fig. 22.1). Raiffa thought it was John Tukey who suggested that practitioners all too often make **errors of a third kind: of solving the wrong problem**. Raiffa nominated a candidate for the **error of the fourth kind: solving the right problem too late**. John Tukey believed that it was better to find an approximate answer to the right question, than the exact answer to the wrong question, which can always be made precise."

Quote from: Singer, D. A. (2018). Solving the Wrong Resource Assessment Problems Precisely. In Handbook of Mathematical Geosciences (pp. 437–446). Springer, Cham. Retrieved from https://library.oapen.org/bitstream/handle/20.500.12657/22939/1007222.pdf?sequence=1#page=446

Testing a hypothesis: Type 1 and Type 2 errors

- Frequentist approach
  - Type 1 error:
    - Effect is noise but we assign significant connection.
    - Null-hypothesis is rejected, when it is actually true
    - “False positive”.
    - Steve Schneider claiming the ice ages are coming back in the 1970’s.
    - The probability of a Type 1 error is associated with the confidence level of correct statistical tests. For a 95% confidence, the probability of a Type 1 error is a = 5%.
  - Type 2 error:
    - Effect is real, but we do not assign a significant connection.
    - Null-hypothesis is accepted, when it is actually false.
    - “False negative”.
    - “We could not find evidence that perchlorate at this dose affects the unborn child”.
  - Scientists typically err on the side of type 2 errors and choose a low probability of a type 1 error (*e.g.*, “$p < 0.05$”).
- Optimal (or Bayesian) decision theoretical approaches
  - Design the strategy based on a utility function
  - Example: A hurricane is predicted to arrive in Miami with a probability of 20%.
  - Should you take action?
  - Maximize the utility of the decision consistent with your posterior.

You may apply Bayesian concepts without knowing it.

- Choosing (a set of possible) model structures can be interpreted as a Bayesian prior on an hyper-parameter.
- We rarely specify all possible alternative hypotheses.
- When you hear hoofs, think horses, not zebras.
  …or a knight with coconut shells

Outline for today

1. Review of reading assignment
2. Why may we need a Bayesian approach?
3. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
4. The pesky aspects of computation

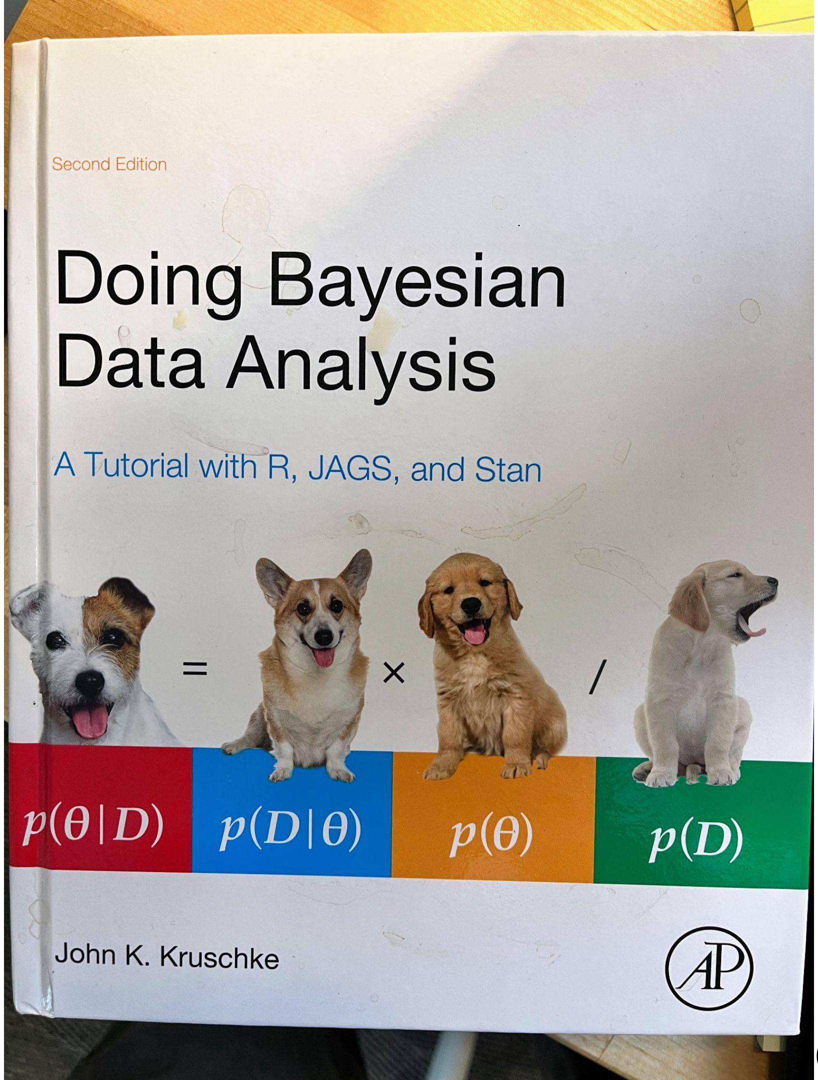
The German Yodeling Test

- There is genetic trait (i.e., exposure to German yodeling without headache) with a 0.5 % incidence rate in certain populations (let’s assume Germans). You, as a member of the population, test positive, but the test is only 98 % accurate (random error)

- What is the probability that you have the trait after one, two, and three test applications with a positive test result?

Please try to solve this warm-up problem
for the <u>first</u> yodeling test application.

Puppies to the rescue…



Recap: Solving the yodeling problem in a Bayesian way

$P(A|B) = P(B|A) * P(A) / P(B) = 98\% * 0.5\% / (98\% * 0.5\% + 2\% * 99.5\%) \approx 0.5\% / (0.5\% + 2\%) = 0.5 / (0.5 + 2) = 20\%$
A: Have the trait
B: Tested positive in one test

After one test, the probability of having the trait would be 20%. Similarly, after two tests,

$P(A|B) = P(B|A) * P(A) / P(B) = 98\% * 98\% * 0.5\% / (98\% * 98\% * 0.5\% + 2\% * 2\% * 99.5\%) \approx 92.346\%$
A: Have the trait
B: Tested positive in two tests

After two tests, the probability of having the trait would be 92.346%. After three tests,

$P(A|B) = P(B|A) * P(A) / P(B) = 98\% * 98\% *98\% * 0.5\% / (98\% * 98\% *98\% * 0.5\% + 2\% * 2\% * 2\% * 99.5\%) \approx 99.831\%$
A: Have the trait
B: Tested positive in three tests

After three tests, the probability of having the trait would be 99.831%

Credits: Email from your fellow student ang.xiao.th@dartmouth.edu on Jan 5 2023

What might a prior mean? (A partial list)

1. <u>A summary of previous, independent information</u>
2. <u>An objective or subjective statement of belief</u>
   (Objective vs Subjective Bayesians, *cf.* Efron 1986)
3. <u>Non-informative (ignorance) range (uniform, improper)</u>
   Careful, assessing the bounds of your ignorance can lead to interesting conclusions (cf. climate sensitivity results)

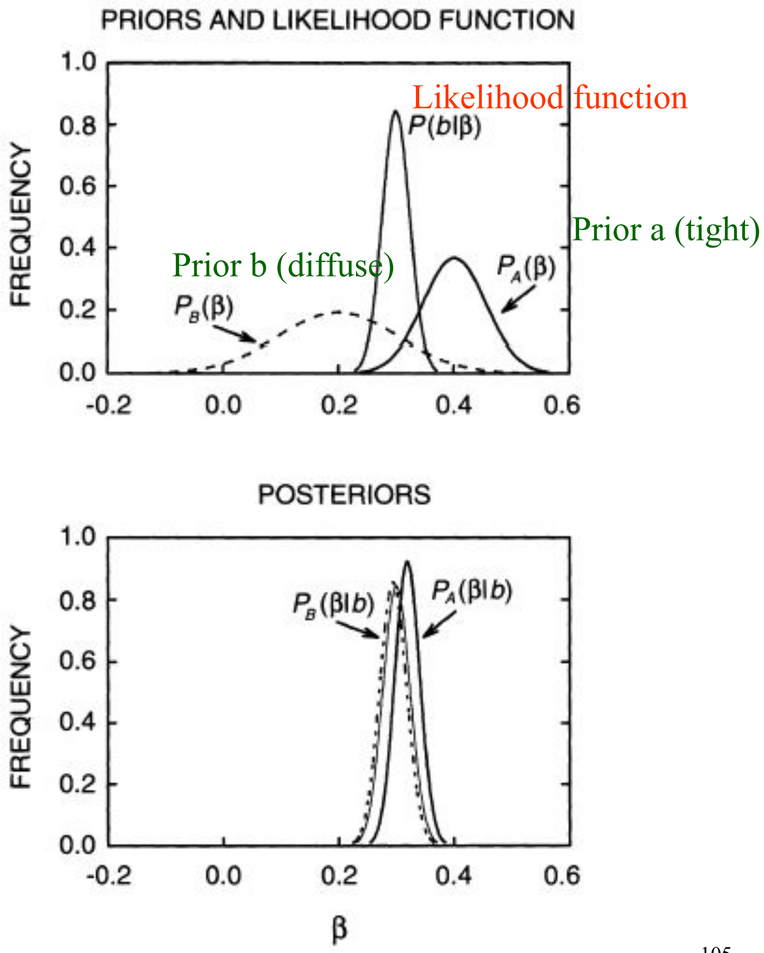
Efron, B. (1986). Why Isn't Everyone a Bayesian? *The American Statistician*, *40*(1), 1–5. https://doi.org/10.2307/2683105

Where do we get the likelihood from?

- A likelihood is a conditional probability.
  - P(rain) = 0.3 (probability)
  - P(rain | rain yesterday) = 0.6 (conditional probability a.k.a. likelihood)
  - careful, in standard english, likelihood is sometimes used synonymously with
- Remember the yodeling calculation from your fellow student?
  - “P(A|B) = P(B|A) * P(A) / P(B) = 98% * 0.5% / (98% * 0.5% + 2% * 99.5%) ≈ 0.5% / (0.5% + 2%) = 0.5 / (0.5 + 2) = 20%
  - A: Have the trait
  - B: Tested positive in one test”
- For a single observation and just observation errors, the likelihood is just the pdf of the observation error.
- For N independent observations, the likelihood is the product of the single pdf values.
- Why is the assumption of independence important?
- How can one test this assumption?

cf. Wilks, p. 19, 108, Ang Xiao

Bayesian updating



Side Note: A Bayesian analysis with infinitely diffuse prior recovers the Frequentist's maximum likelihood solution (*cf*. for example, Efron, 1986).

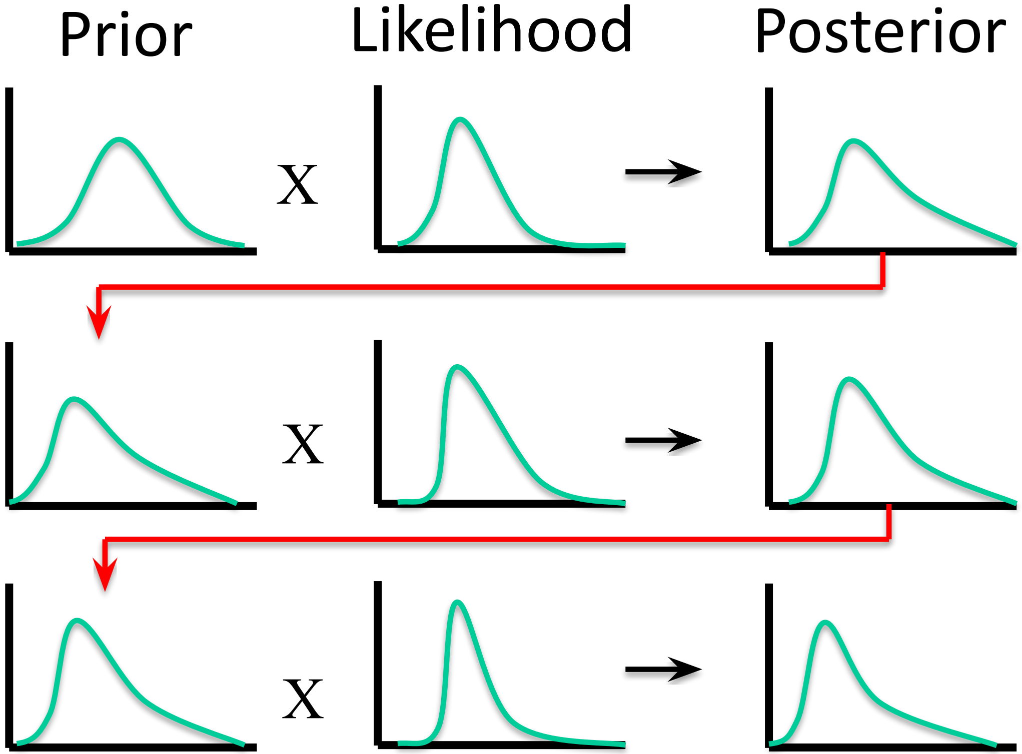
Ellison (1996)

Sequential Bayesian Updating

Probability it will rain given it rained yesterday…

…and that it is cloudy this morning

…and the wind is coming from the southwest



Outline for today

1. Review of reading assignment
2. Why may we need a Bayesian approach?
3. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
4. The pesky aspects of computation

A more general form of Bayes' rule

$$P(\theta \mid x) = \frac{P(x \mid \theta) \cdot P(\theta)}{\int P(x \mid \theta) P(\theta) \, d\theta}$$

P(θ|x) $\Rightarrow$ probability of parameter value θ, given observation x

$\theta \Rightarrow$ parameter

x $\Rightarrow$ observation(s)

P(x|θ) $\Rightarrow$ probability of observation x, given θ

a.k.a. the likelihood of x (discussed below)

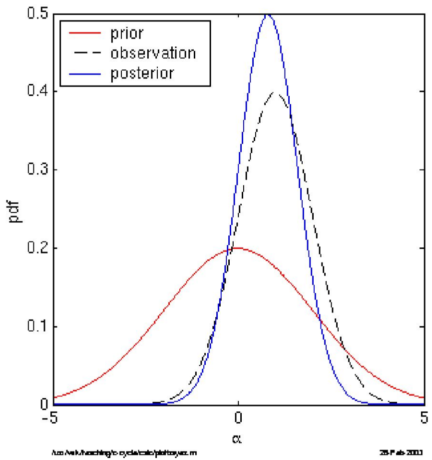
P(θ) $\Rightarrow$ Prior probability of θ

Nasty little detail: How to evaluate the pesky little integral…

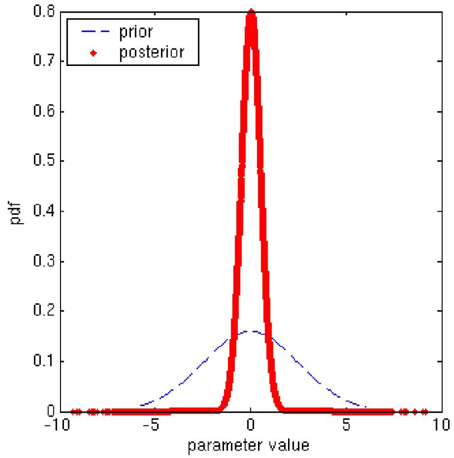
cf. D'Agostini, 2003, p. 36-37

Bayesian updating examples

Using analytical solution, works fine for normal distributions.



Using numerical solution, works fine for all distributions, as long as the dimension of the problem is low and/or the model is fast.



High-dimensional parameter spaces and computationally slow models can pose nontrivial computational challenges for uncertainty characterization.

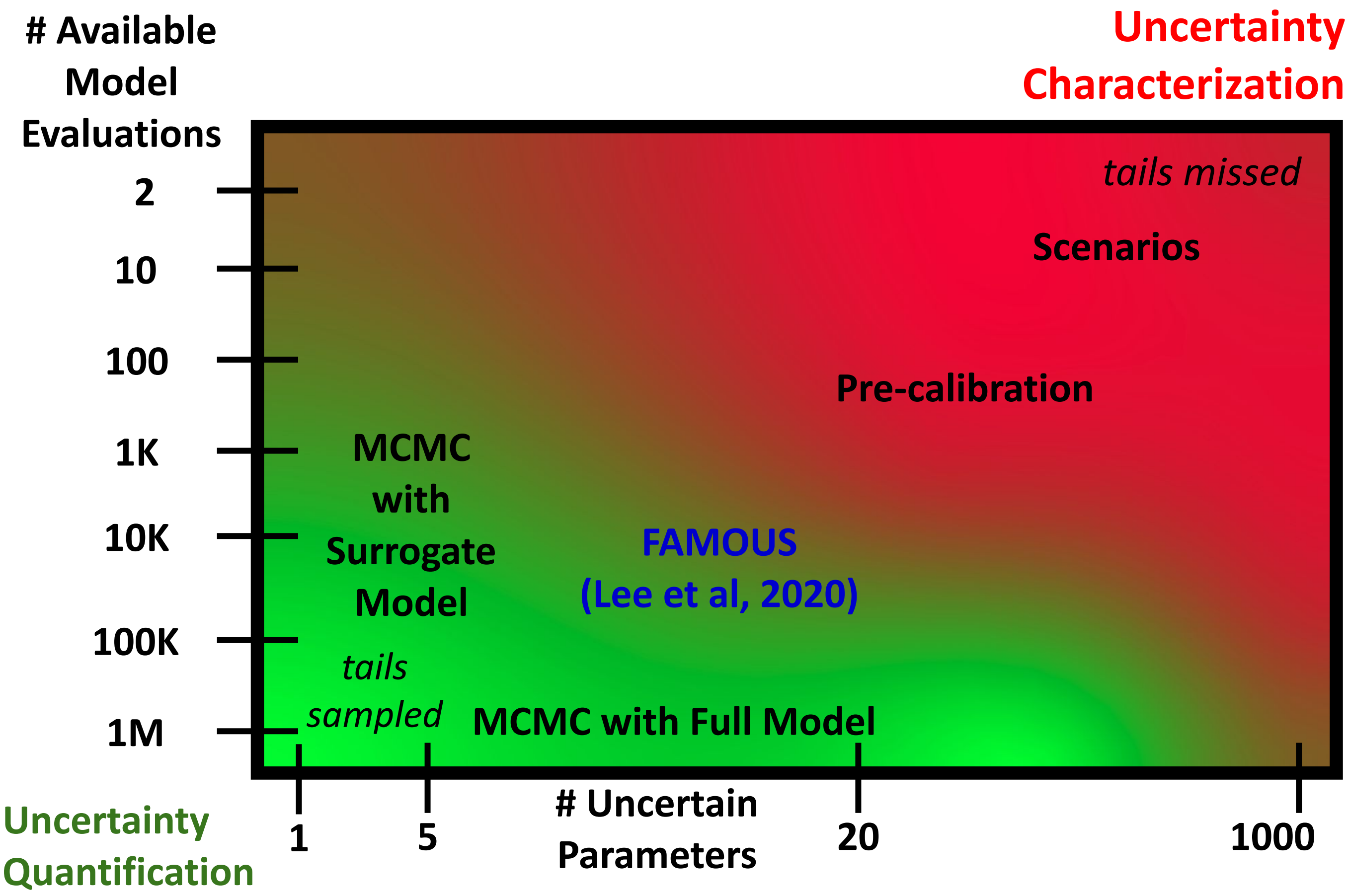


Figure modified from: Reed, P. M., Hadjimichael, A., Malek, K., Karimi, T., Vernon, C. R., Srikrishnan, V., et al. (2022). *Addressing Uncertainty in MultiSector Dynamics Research*. Retrieved from https://uc-ebook.org/

Reading Assignments

- Core:
  - NSF. (n.d.). What is GRFP? Retrieved November 29, 2022, from https://www.nsfgrfp.org/
  - COMPASS Science Communication, Inc. (2017). *The Message Box Workbook*. COMPASS. Retrieved from https://www.compassscicomm.org/

- Supplementary (if you are interested in background)
  - Stockes, D. E. (1997). *Pasteur's Quadrant*. Washington, DC: Brookings Institution Press.
  - Tijssen, R. J. W. (2018). Anatomy of use-inspired researchers: From Pasteur's Quadrant to Pasteur's Cube model. *Research Policy*, *47*(9), 1626–1638. https://doi.org/10.1016/j.respol.2018.05.010
  - NSF. (n.d.). *PROPOSAL AND AWARD POLICIES AND PROCEDURES GUIDE*. Retrieved from https://www.nsf.gov/pubs/policydocs/pappg22_1/index.jsp accessed Nov 29 2022

Review

1. What defines a Frequentist and a Bayesian approach?
2. What are common errors in statistical inference?
3. How may a Bayesian approach help you to avoid Type III errors?
4. What is a prior, likelihood function, and posterior?
5. Why may it be difficult to perform statistical inference for real-world models?