ENGG 107: Bayesian Statistical Modeling and Computation

Wrapping up Bayes/ Analytical Solutions (meeting #6)

Objectives for today:

1. Learn key concepts of analytical solutions to Bayesian inference
2. Learn implementation approaches for these analytical solutions
3. Identify strengths and weaknesses of these analytical approaches.

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	Designing a project	Setting up Computation		1	none
3	January 16, 2023	No class (MLK day)	Analytical Solutions / Intuition	Bayesian Workflow		2	1
4	January 23, 2023	Precalibration /(Bayes) Monte Carlo	MCMC Part 1	MCMC Part 2	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	Getting 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 of Reading Assignments

- Core:
  - D'Agostini, G. (2003). *Bayesian reasoning in data analysis: A critical introduction*. Singapore: World Scientific Publishing. (Chapter 6, an easy and elegant introduction).

- Supplementary (if you are interested in background and/or applications)
  - Kalman Filter. (n.d.). Retrieved November 29, 2022, from https://en.wikipedia.org/wiki/Kalman_filter
  - Chen, Z., & Others. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, *182*(1), 1–69.
  - Pei, Y., Biswas, S., Fussell, D. S., & Pingali, K. (2019). An elementary introduction to Kalman filtering. *Communications of the ACM*, *62*(11), 122–133. https://doi.org/10.1145/3363294
  - Enting, I. G., Trudinger, C. M., & Francey, R. J. (1995). A synthesis inversion of the concentration and delta ^13C of atmospheric CO_2. *Tellus Ser. B.*, *47*, 35–52.

Outline for today

1. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

Outline for today

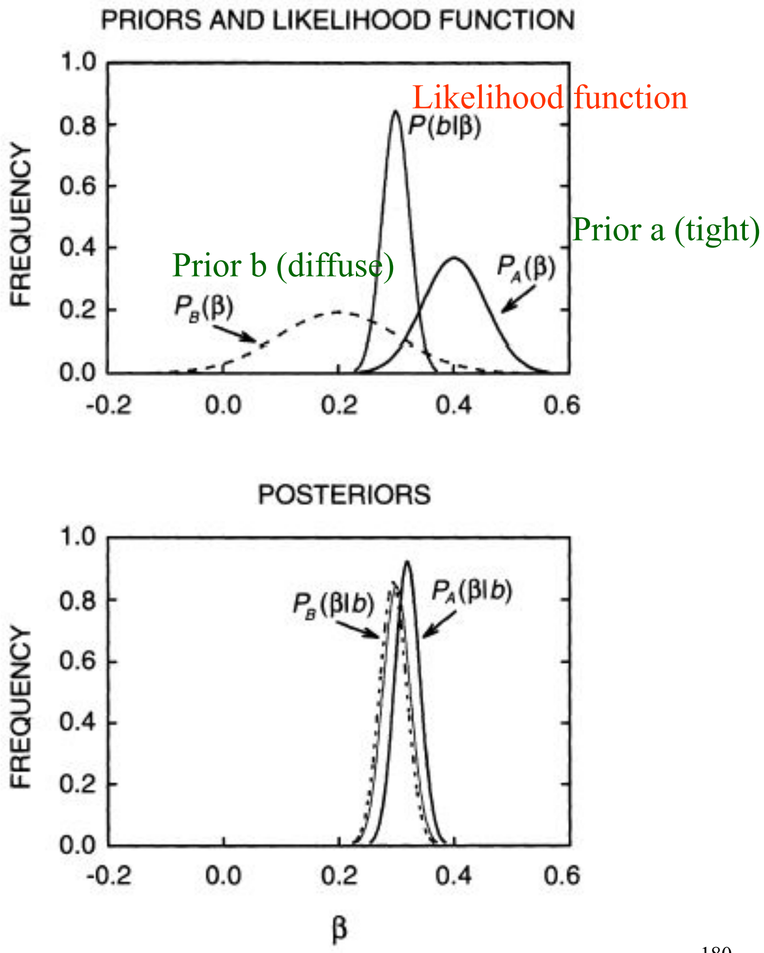
1. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

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).

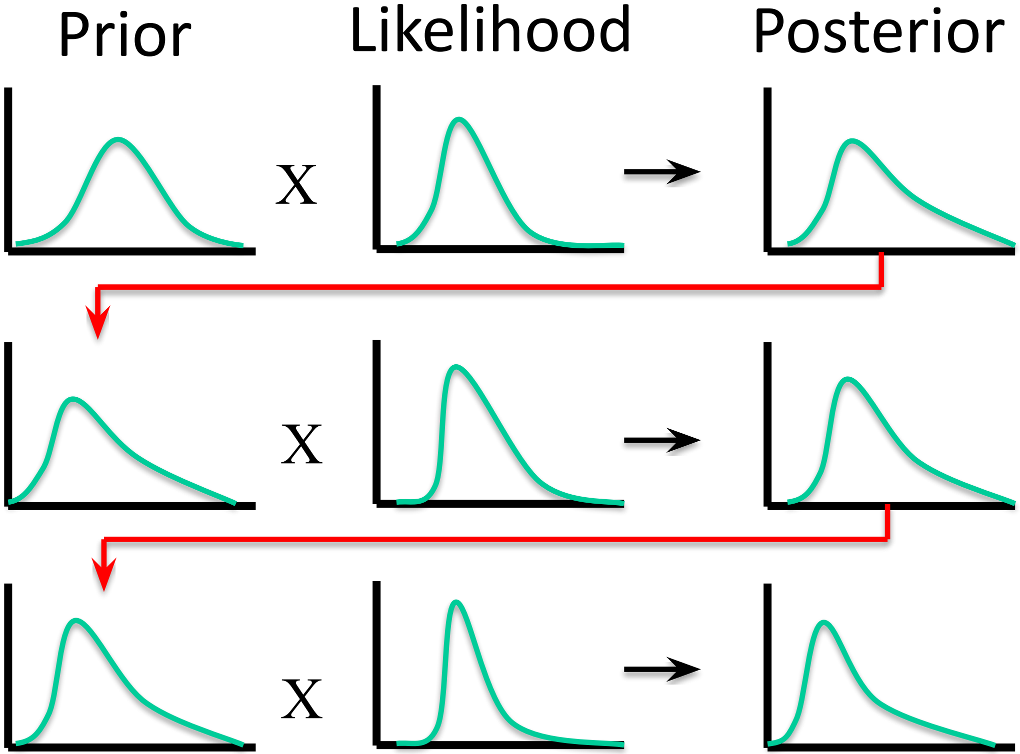
Ellison (1996)

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. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

A more general form of Bayes' rule

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

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

$\theta \Rightarrow$ parameter

x $\Rightarrow$ observation(s)

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

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

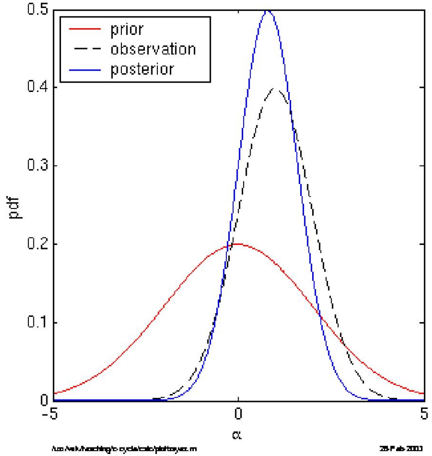
$P(\theta) \Rightarrow$ Prior probability of θ

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

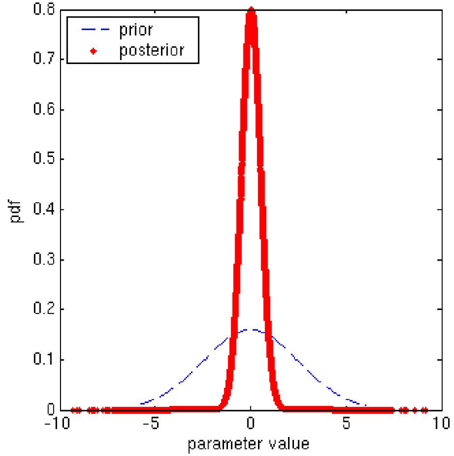
cf. D'Agostini, 2003, p. 36-37

Bayesian updating examples

Using analytical solution, works fine for normal distributions.

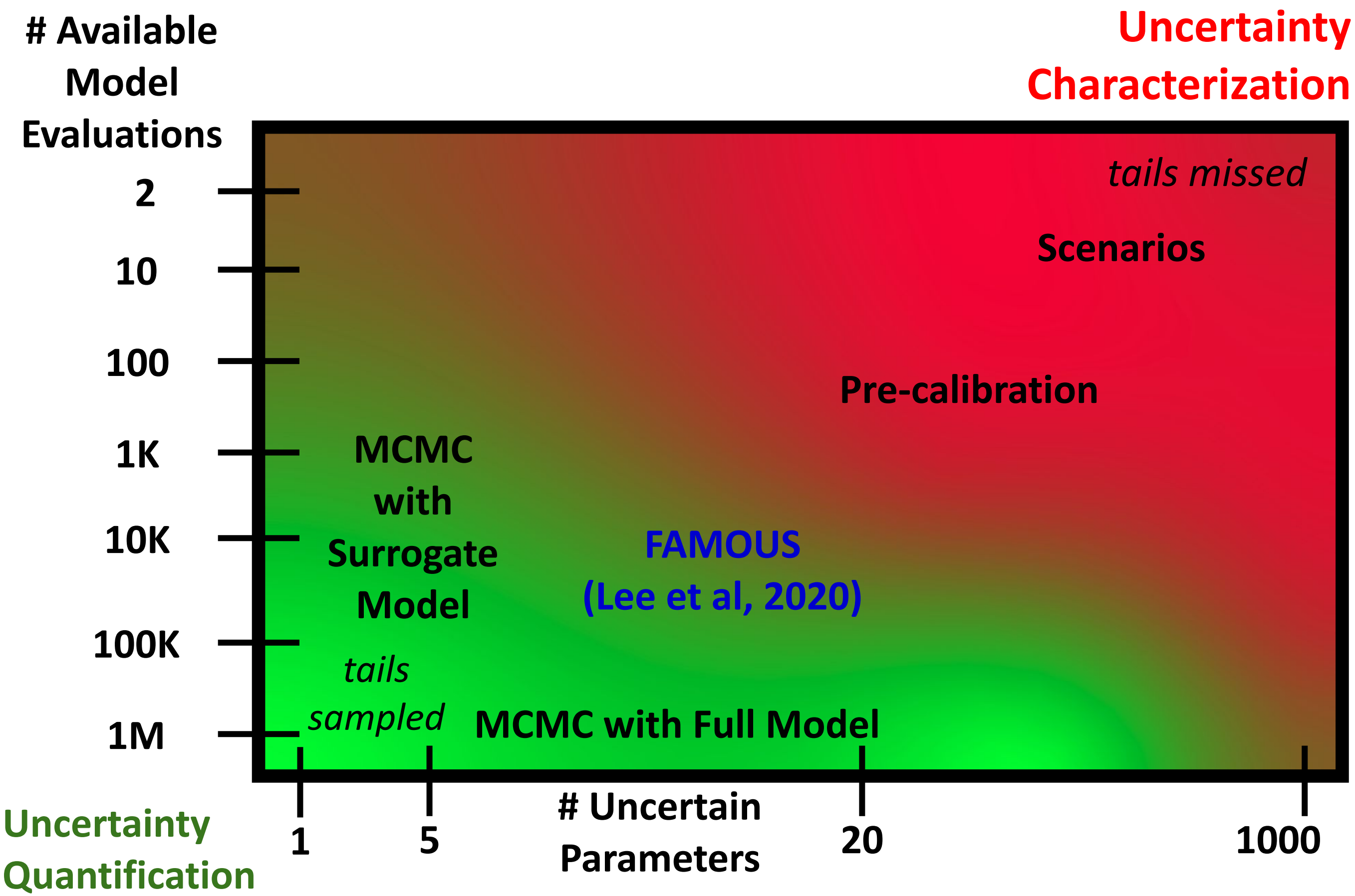


Using numerical solution, works fine for all distributions, as long as the dimension of the problem is low and/or the model is fast.



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

Figure: 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/



Background:
Lee, B. S., Haran, M., Fuller, R. W., Pollard, D., & Keller, K. (2020). A fast particle-based approach for calibrating a 3-D model of the Antarctic ice sheet. *The Annals of Applied Statistics*, *14*(2), 605–634. https://doi.org/10.1214/19-AOAS1305
Sharma, S., Lee, B. S., Hosseini-Shakib, I., Haran, M., & Keller, K. (2023). Neglecting model parametric uncertainty can drastically underestimate flood risks. *Earth's Future*, *11*(1). https://doi.org/10.1029/2022ef003050

Outline for today

1. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

What are examples of analytical approaches?

- Closed form likelihood function with
  - improper prior
  - uniform prior
  - conjugate prior
- Kalman Filter
- “Synthesis inversion” (Enting 1995)

Enting, I. G., Trudinger, C. M., & Francey, R. J. (1995). A synthesis inversion of the concentration and delta ^13C of atmospheric CO_2. *Tellus Ser. B.*, *47*, 35–52.

Outline for today

1. How does a Bayesian approach work?
   a. prior
   b. likelihood
   c. posterior
   d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are key assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

What are conjugate priors?

6.4 Conjugate priors

In Sec. 6.3 we introduced with a practical example the concept of *conjugate priors*: a prior such that the product likelihood $\times$ prior, i.e. the posterior, belongs to the same family as the prior. This is a well known technique to simplify the calculations.

D'Agostini, G. (2003). *Bayesian reasoning in data analysis: A critical introduction*. Singapore: World Scientific Publishing.

Simple case of Gaussian Priors and Likelihood

6.3 Combination of several measurements – Role of priors

Let us imagine making a second set of measurements of the physical quantity, which we assume unchanged from the previous set of measurements. How will our knowledge of μ change after this new information? Let us call $x_2 = \overline{q}_{n_2}$ and $\sigma_2 = \sigma'/\sqrt{n_2}$ the new average and standard deviation of the average (σ' may be different from σ of the sample of n_1 measurements), respectively. Applying Bayes' theorem a second time we now have to use *as initial distribution the final probability of the previous inference*:

$$f(\mu\,|\,x_1,\sigma_1,x_2,\sigma_2,\mathcal{N}) = \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\,\sigma_2^2}\right] f(\mu\,|\,x_1,\mathcal{N}(\cdot,\sigma_1))}{\int \frac{1}{\sqrt{2\,\pi}\,\sigma_2}\exp\left[-\frac{(x_2-\mu)^2}{2\,\sigma_2^2}\right] f(\mu\,|\,x_1,\mathcal{N}(\cdot,\sigma_1))\,\mathrm{d}\mu}\,. \tag{6.7}$$

The integral is not as simple as the previous one, but still feasible analytically. The final result is

$$f(\mu\,|\,x_1,\sigma_1,x_2,\sigma_2,\mathcal{N}) = \frac{1}{\sqrt{2\,\pi}\,\sigma_A}\exp\left[-\frac{(\mu - x_A)^2}{2\,\sigma_A^2}\right], \tag{6.8}$$

where

$$x_A = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}\,, \tag{6.9}$$

$$\frac{1}{\sigma_A^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\,. \tag{6.10}$$

D'Agostini (Ed.). (2003). Gaussian Likelihood. In *Bayesian Reasoning in Data Analysis* (pp. 123–139). World Scientific.

Link to the Kalman Filter

6.3.1 *Update of estimates in terms of Kalman filter*

An interesting way of writing Eq. (6.9) is considering x_1 and x_A the estimates of μ at times t_1 and t_2, respectively before and after the observation x_2 happened at time t_2. The uncertainties about μ at t_1 and t_2 are $\sigma_\mu(t_1) = \sigma_1$ and $\sigma_\mu(t_2) = \sigma_A$, respectively. Indicating the *estimates* at different times by $\hat{\mu}(t)$, we can rewrite Eq. (6.9) as

$$\begin{aligned}\hat{\mu}(t_2) &= \frac{\sigma_\mu^2(t_1)}{\sigma_x^2(t_2)+\sigma_\mu^2(t_1)}\,x(t_2) + \frac{\sigma_x^2(t_2)}{\sigma_x^2(t_2)+\sigma_\mu^2(t_1)}\,\hat{\mu}(t_1) \\ &= \hat{\mu}(t_1) + \frac{\sigma_\mu^2(t_1)}{\sigma_x^2(t_2)+\sigma_\mu^2(t_2)}\,[x(t_2)-\hat{\mu}(t_1)] \qquad (6.11) \\ &= \hat{\mu}(t_1) + K(t_2)\,[x(t_2)-\hat{\mu}(t_1)] \qquad (6.12)\end{aligned}$$

$$\sigma_\mu^2(t_2) = \sigma_\mu^2(t_1) - K(t_2)\,\sigma_\mu^2(t_1)\,, \qquad (6.13)$$

where

$$K(t_2) = \frac{\sigma_\mu^2(t_1)}{\sigma_x^2(t_2)+\sigma_\mu^2(t_1)}\,. \qquad (6.14)$$

Indeed, we have given Eq. (6.9) the structure of a *Kalman filter* [65]. The new observation 'corrects' the estimate by a quantity given by the *innovation* (or *residual*) $[x(t_2)-\hat{\mu}(t_1)]$ times the *blending factor* (or *gain*) $K(t_2)$. For an introduction about Kalman filter and its probabilistic origin, see Refs. [66] and [67].

D'Agostini (Ed.). (2003). Gaussian Likelihood. In *Bayesian Reasoning in Data Analysis* (pp. 123–139). World Scientific.

Outline for today

1. How does a Bayesian approach work?
    a. prior
    b. likelihood
    c. posterior
    d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

Advantages of analytical approaches

- Fast and precise
- Analytical tractable (insights, …)
- Not impacted by numerical artifacts
- Can provide a great positive control

Outline for today

1. How does a Bayesian approach work?
   a. prior
   b. likelihood
   c. posterior
   d. Bayesian updating
2. The pesky aspects of computation
3. What are examples of analytical approaches?
4. What are assumptions for these approaches?
5. What are advantages of analytical approaches?
6. When not to rely on analytical approaches?

When not to rely on analytical approaches?

- The required assumptions are too poor an approximation.
- You cannot find one ;-)

Getting intuition by doing approximate Bayesian analysis by hand

Can you safely make it an airport one flight hour away?

- One key input: how much gas do we have?
- Imperfect information with priors
- Average fuel use=10 gal/hr
- Required reserve: 45 minutes when you arrive.
- Consider left tank only
- Consider left and right tank.

Picture: Klaus Keller

Case studies
Please sketch out your priors, likelihood functions, and posteriors

- "Kalman"
  - Observation at 18 gallons with Gaussian standard deviation of 10 gallons
  - Gaussian prior at 15 gallons, standard deviation of 10 gallons.
- "Uniform"
  - Same observation
  - Uniform prior of 0 to 25 gallons (empty to full tank, capacity)
- "Deep Uncertainty"
  - Same observation
  - You and the passenger remember that the person who flew before you in the aircraft did add gas to the tank. Your passenger is certain that the prior pilot was in a hurry and only filled up the tank to ¼ of the capacity just before you take the observation. You, on the other hand, are certain that the previous pilot filled up the tank to the maximum of the capacity.
  - What priors do you choose to approximate this situation?
  - What are your posteriors?

Reading Assignments

- Core:
  - Please review carefully the **data analysis checklist handout.**

- Supplementary (if you are interested in background)
  - Doss-Gollin, J., & Keller, K. (2022). A subjective Bayesian framework for synthesizing deep uncertainties in climate risk management. Earth's Future. https://doi.org/10.1029/2022ef003044
  - Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., et al. (2020, November 3). Bayesian Workflow. arXiv [stat.ME]. Retrieved from http://arxiv.org/abs/2011.01808
  - Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. Psychological Methods, 22(2), 240–261. https://doi.org/10.1037/met0000065

Review - I

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?

Review II

Do you know

1. how to interpret and define a prior and a simple likelihood function?
2. examples and advantages of analytical solutions to the Bayes equation?
3. reasons to switch to numerical solutions to Bayes equation?
4. reasons why numerical solutions to Bayes equations may be hard to find?