TALENT COURSE 11
LEARNING FROM DATA: BAYESIAN METHODS AND MACHINE LEARNING

Lecture 1: Introduction

Daniel Phillips
Ohio University
TU Darmstadt
ExtreMe Matter Institute

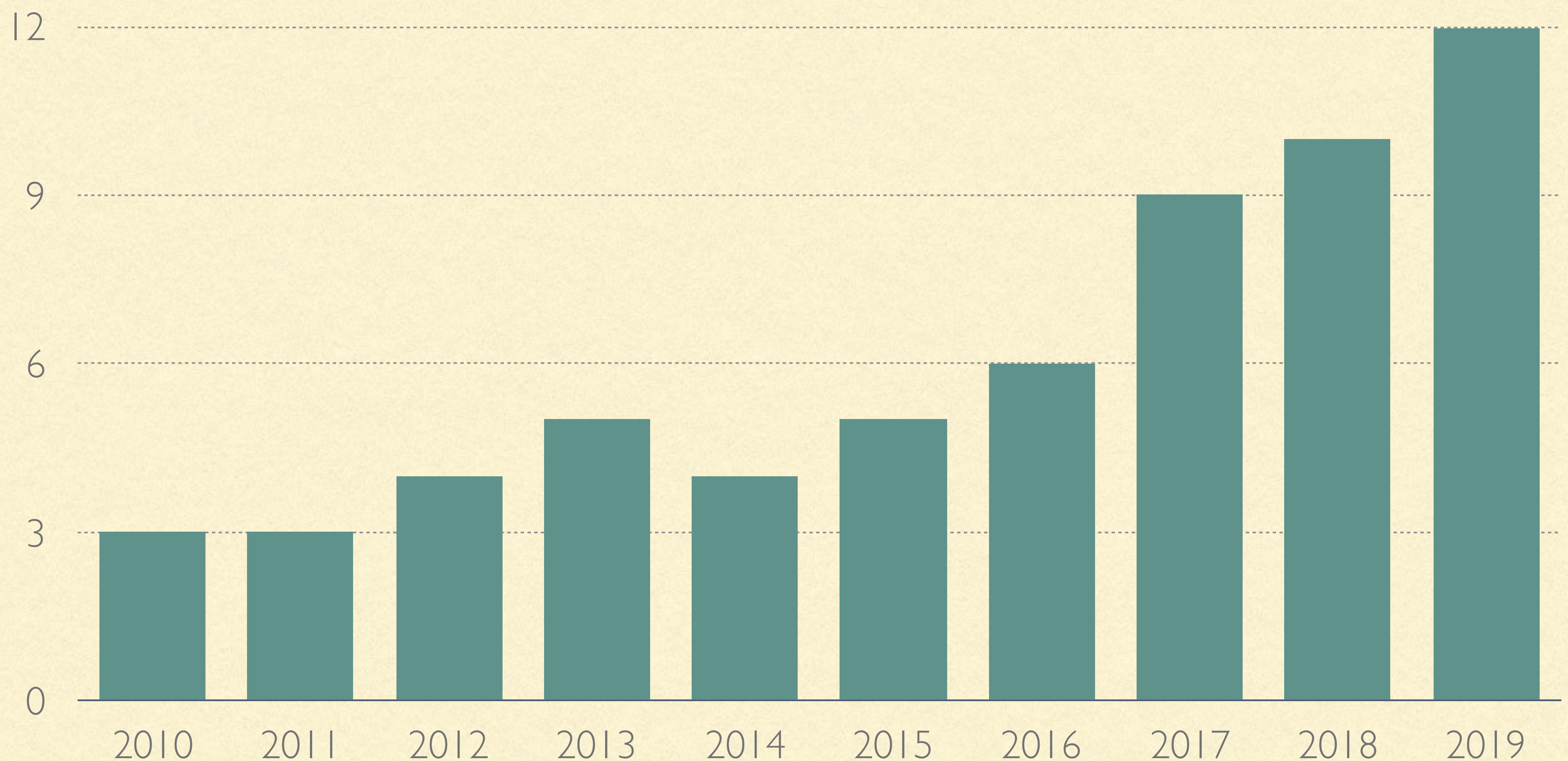
HOW DO YOU FEEL ABOUT STATISTICS?

- Disraeli (attr.): “There are three kinds of lies: lies, damned lies, and statistics.”
- Rutherford: “If your result needs a statistician then you should design a better experiment.”

Bayesian Methods: rules of statistical inference are an application of the laws of probability

- Laplace: “La théorie des probabilités n’est que le bon sens réduit au calcul”

BAYES ON NUCL-TH



EXAMPLES

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **485**, 5363–5376 (2019) doi:10.1093/mnras/stz654
Advance Access publication 2019 March 11

Equation of state sensitivities when inferring neutron star and dense matter properties

S. K. Greif,[1,2★†] G. Raaijmakers,[3★†] K. Hebeler,[1,2] A. Schwenk[1,2,4] and A. L. Watts [3]

[1] *Institut für Kernphysik, Technische Universität Darmstadt, D-64289 Darmstadt, Germany*
[2] *ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung GmbH, D-64291 Darmstadt, Germany*
[3] *Anton Pannekoek Institute for Astronomy, University of Amsterdam, PO Box 94249, NL-1090GE Amsterdam, the Netherlands*
[4] *Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, D-69117 Heidelberg, Germany*

Accepted 2019 March 4. Received 2019 March 4; in original form 2018 December 13

ABSTRACT

Understanding the dense matter equation of state at extreme conditions is an important open problem. Astrophysical observations of neutron stars promise to solve this, with *Neutron Star Interior Composition Explorer* poised to make precision measurements of mass and radius for several stars using the waveform modelling technique. What has been less clear, however, is how these mass–radius measurements might translate into equation of state constraints and what are the associated equation of state sensitivities. We use Bayesian inference to explore and contrast the constraints that would result from different choices for the equation of state parametrization; comparing the well-established piecewise polytropic parametrization to one based on physically motivated assumptions for the speed of sound in dense matter. We also compare the constraints resulting from Bayesian inference to those from simple compatibility cuts. We find that the choice of equation of state parametrization and particularly its prior assumptions can have a significant effect on the inferred global mass–radius relation and

LEARNING OUTCOMES

Upon completion of this course students should be able to:

- Apply the rules of probability to derive posterior probability distributions for simple problems involving prior information on parameters and various standard likelihood functions.
- Perform Bayesian parameter estimation, including in cases where marginalization over nuisance parameters is required.
- Use Monte Carlo sampling to generate posterior probability distributions and identify problems were standard sampling is likely to fail.
- Compute an evidence ratio and explain what it means.
- Explain machine learning from a Bayesian perspective and employ a testing and training data set to develop and validate a Gaussian-process model.
- Employ these methods in the context of specific nuclear-physics problems.
- Be able to understand, appreciate, and criticize the growing literature on Bayesian statistics and machine learning for low-energy nuclear physics applications.

My personal goal for you: think about inferences and learning from a Bayesian perspective

WEEKLY SCHEDULE

- Today: Introduction, probabilities, Bayes' theorem (DP), Parameter Estimation 1 (RJF)
- Tuesday: More on Bayesian ideas (DP), Parameter Estimation 2 (RJF)
- Wednesday: Sampling 1 (RJF), Why Bayes is Better 1 (CF)
- Thursday: Parameter Estimation 3 (RJF), Sampling 2 (CF)
- Friday: Why Bayes is Better 2 (CF), EFT & Bayesian methods (RJF)
- Week 2: Model Selection, Maximum Entropy, More Sampling, Gaussian Processes
- Week 3: Model checking, Bayesian optimization, Bootstrap, TBD
- Mini-projects each week. First one “handed out” tomorrow

ICE BREAKER TIME….

DAILY SCHEDULE

- https://nucleartalent.github.io/Bayes2019/
- Material delivered through github
- git tutorial tonight at 6 pm
- Social evening: ????

INFERENCE

- Deductive inference. Cause→Effect.
- Inference to best explanation. Effect→Cause.
- Scientists need a way to:
  a) Quantify the strength of inductive inferences;
  b) Update that quantification as they acquire new data.

PROBABILITY DISTRIBUTION FUNCTIONS (PDFS)

- pr(A|B) reads "probability of A given B"
- Simplest examples are discrete, but physicists often interested in continuous case, e.g., parameter estimation
- When integrated, continuous pdfs become probabilities$\Rightarrow$pdfs are NOT dimensionless, even though probabilities are
- 68%, 95%, etc. intervals can then be computed by integration
- Certainty about a parameter corresponds to $pr(x)=\delta(x-x_0)$
- Let's play

PRODUCT AND SUM RULE

- Product rule:

$$\mathrm{pr}(A\,|\,B\,.\,I)\mathrm{pr}(B\,|\,I) = \mathrm{pr}(A\,|\,I)$$

- Sum rule:

$$\mathrm{pr}(A\,|\,I) + \mathrm{pr}(\bar{A}\,|\,I) = 1$$

Extends to:

$$\sum_j \mathrm{pr}(A_j\,|\,I) = 1$$

for a complete, discrete, mutually exclusive set of possibilities $\{A_j\}$

Time to <u>play</u> with this a bit…

BAYES' THEOREM

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A|B) = \frac{\mathrm{pr}(B|A)\mathrm{pr}(A)}{\mathrm{pr}(B)}$$

Likelihood

Prior

$$\mathrm{pr}(\mathrm{hypothesis}|\mathrm{data}) = \frac{\mathrm{pr}(\mathrm{data}|\mathrm{hypothesis})\mathrm{pr}(\mathrm{hypothesis})}{\mathrm{pr}(\mathrm{data})}$$

Posterior

Normalization

Probability as degree of belief cf. frequentist view

MARGINALIZATION

$$\sum_j \mathrm{pr}(A_j \,|\, I) = 1$$

$$\Rightarrow \mathrm{pr}(B \,|\, I) = \sum_j \mathrm{pr}(B, A_j \,|\, I)$$

Continuous versions:

$$\int dx \, \mathrm{pr}(x \,|\, I) = 1$$

$$\Rightarrow \mathrm{pr}(y \,|\, I) = \int dx \, \mathrm{pr}(y, x \,|\, I)$$

COIN TOSSING EXAMPLE

- Is this a fair coin?
- Deductive argument:

$$\text{Fair} \Rightarrow \text{pr(Heads)} = \text{pr(Tails)} = 0.5$$

$$\Rightarrow \text{pr}(R \text{ heads out of } N \text{ tosses} \,|\, \text{fair coin}) = \binom{N}{R}(0.5)^R(0.5)^{N-R}$$

- Is sum rule obeyed here?
- More generally

$$\text{pr}(R \text{ heads out of } N \text{ tosses} \,|\, p_H) = \binom{N}{R} p_H^R (1 - p_H)^{N-R}$$

COIN TOSSING TAKE AWAYS

- $pr(p_H|data,I)$ is the product of the binomial distribution and the prior
- The frequentist result corresponds to a particular choice of prior
- Can do analysis sequentially or all at once
- MUST NOT bootstrap
- Let's play