

## **Engineering Tripos Part IIB, 4M24: Computational Statistics and Machine Learning, 2023-24**

### **Leader**

[Prof M Girolami](#) [1]

### **Timing and Structure**

Michaelmas term. 75% exam / 25% coursework. Lectures will be recorded.

### **Prerequisites**

3F3, 3F8, 3M1

### **Aims**

The aims of the course are to:

- Introduce students to foundational theoretical concepts and methodological tools essential for the successful development, analysis, and application of advanced Machine Learning and Computational Statistical methods.

### **Objectives**

As specific objectives, by the end of the course students should be able to:

- Introduce the students to the required statistical and mathematical concepts that underpin all rigorously designed Machine Learning and Computational Statistical methods that can be used practically across all the contemporary engineering sciences
- Introduce the students to advanced computational statistical inference methods required to design Machine Learning solutions to a range of challenging large scale engineering problems where data and models are synthesised

### **Content**

By successful completion of this course the student will have an appreciation and basic understanding of the mathematical, probabilistic, and statistical foundations of modern Computational Statistical Methods and recent developments in Machine Learning algorithms.

#### **Computational Statistics and Machine Learning**

**Lecture.1. Monte Carlo Methods - A** : Numerically computing integrals, the law of large numbers for Monte Carlo estimators, The Central Limit Theorem for Monte Carlo estimators.

**Lecture.2. Monte Carlo Methods - B** : Improving MC estimators, Importance Sampling, Control Variates to reduce variance of estimates.

**Lecture.3. Lebesgue Integral and Measure - A** : Difference between Riemann and Lebesgue Integral, why Lebesgue integral is required for machine learning and engineering, definition of Lebesgue integral.

**Lecture.4. Lebesgue Integral and Measure - B** : Definition of Lebesgue Measure, Radon-Nikodym derivative and change of measure, Measure theoretic basis of Probability (Kolmogorov), Random Variables.

**Lecture.5. Markov Chain Monte Carlo - A** : Definition of Markov chain and invariant distributions, presentation of the Metropolis and Hastings method.

**Lecture.6. Markov Chain Monte Carlo - B** : Metropolis Hastings in multiple dimensions, the Gibbs Sampler.

**Lecture.7. Vector, Metric, and Banach Spaces** : generalisation of Euclidean space in  $\mathbb{R}^3$  to infinite dimensional spaces, Completion of space and definition of Banach space of functions.

**Lecture.8. Hilbert Spaces** : Inner product space, definition of Hilbert space, Cauchy sequences and function approximation, Reproducing kernel Hilbert Space and function approximation.

**Lecture.9. Sobolev Spaces** : Definition of weak derivatives, understanding rates of convergence of function approximations based on properties of Sobolev space (smoothness)

**Lecture.10. Gaussian Measure in Hilbert Space** : Illustrating non-existence of Lebesgue Measure in function space, construction of finite Gaussian measure in Hilbert space, definition of Bayes rule (via Radon-Nikodym derivative) in Hilbert space employing Gaussian measure as reference - GP's.

**Lecture.11. MCMC in Hilbert space** : defining dimension invariant Markov transition kernel in Hilbert space and how overcomes degeneracy in high dimensions.

**Lecture.12. Langevin Dynamics Simulation I** - use of Langevin dynamics to simulate from a desired probability measure.

**Lecture.13. Langevin Dynamics Simulation II** - use of approximate Langevin dynamics to simulate from a desired probability measure.

**Lecture.14. Parallel Tempering** - introduction to simulation from multi-modal target probability measures.

**Further notes**

Machine Learning methods are having a major impact in every area of the engineering sciences. Machine Learning models and methods rely predominantly on Computational Statistics methods for model calibration, estimation, prediction and updating. Together Computational Statistics and Machine Learning are providing a revolution in the way mankind lives, works, communicates, and transacts.

Machine Learning methodology is not a magic wand that once waved will mysteriously solve long standing technical problems. There are underlying mathematical and statistical theories and principles which define these Machine Learning methods and it is important for the Machine Learning practitioner to have some understanding of them. This course is complementary to current Machine Learning modules in the Engineering Tripos.

This course will provide an overview and very basic introduction to a subset of the major theoretical and methodological ideas that underpin much of Machine Learning. It will provide the student with an appreciation of the possibilities and limitations of Machine Learning and Computational Statistics. This should be a launch pad for students wishing to gain a greater in-depth understanding of Machine Learning as both practitioner and researcher.

### Coursework

| Coursework                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Format                                             | Due date<br>& marks              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|----------------------------------|
| <p><b>Simulation Based Inference on Engineering Problem</b></p> <p>The synthesis of both data and formal mathematical models in defining a digital twin of an engineering problem will be presented. The design of the machine learning and computational statistical methods to characterise uncertainty in predictions and forecasts from the digital twin will be the main focus of this exercise.</p> <p><u>Learning objective:</u></p> <ul style="list-style-type: none"> <li>• To take an engineering problem and define appropriate mathematical, machine learning and data modelling strategies in studying the characteristics of the engineering system or artefact.</li> <li>• To successfully implement and deploy computational statistical methods in delivering an uncertainty quantification strategy in the specific engineering problem.</li> </ul> | <p>Individual Report</p> <p>anonymously marked</p> | <p>Wed week 9</p> <p>[15/60]</p> |

### Booklists

Shima, H. Functional Analysis for Physics and Engineering: An Introduction, CRC Press.

Biegler, L., Biros, G., Ghattas, O., Heinkenschloss, M., Keyes, D., Mallickj, B., Tenorio, L., van Bloemen Waanders,

- B., Willcox, K., and Marouk, Y. (2010). Large-Scale Inverse Problems and Quantification of Uncertainty. Wiley.
- Brooks, S., Gelman, A., Jones, G. L., and Meng, X. (2011). Handbook of Markov Chain Monte Carlo. CRC.
- Cotter, C. and Reich, S. (2015). Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge University Press.
- Law, K., Stuart, A., and Zygalakis, K. (2015). Data Assimilation: A Mathematical Introduction. Springer.
- Rogers, S. and Girolami, M. (2016). A First Course in Machine Learning, 2nd Edition. CRC.
- Sullivan, T. J. (2015). Introduction to Uncertainty Quantification. Springer.

## Examination Guidelines

Please refer to [Form & conduct of the examinations](#) [2].

Last modified: 30/05/2023 15:35

**Source URL (modified on 30-05-23):** <https://teaching23-24.eng.cam.ac.uk/content/engineering-tripos-part-iib-4m24-computational-statistics-and-machine-learning-2023-24>

## Links

[1] <mailto:mag92@cam.ac.uk>

[2] <https://teaching23-24.eng.cam.ac.uk/content/form-conduct-examinations>