Speeding up Monte-Carlo Integration

Francois-Xavier Briol

University of Warwick, Coventry, UK.

Monte Carlo methods are one of the main tools of modern statistics and applied mathematics. They are commonly used to approximate integrals, which allows statisticians to solve many tasks of interest such as making predictions or inferring parameter values of a given model. However, the recent surge in data available to scientists has led to an increase in the complexity of mathematical models, rendering them much more computationally expensive to evaluate. This has a particular bearing on Monte Carlo methods, which will tend to be much slower due to the high computational costs.

This talk will introduce a Monte Carlo integration scheme which makes use of properties of the integrand (e.g. smoothness or periodicity) in order to obtain fast convergence rates in the number of integrand evaluations. This will allow users to obtain much more precise estimates of integrals for a given number of model evaluations. Both theoretical properties of the methodology, including convergence rates, and practical issues, such as the tuning of parameters, will be discussed. Finally, the proposed algorithm will be illustrated on a Bayesian inverse problem for a PDE model of subsurface flow.

How can statistics change healthcare: the importance of clinical trials

Olympia Papachristofi

University of Leeds, Leeds, UK.

Clinical trials research is the backbone of primary research that informs healthcare practice in in the UK; if appropriately-designed, clinical trials generate the best scientific evidence of an intervention?s clinical and cost-effectiveness. Consequently, they are a key source of information for healthcare providers, NHS practitioners and patients, and the principal drivers in future decision making in healthcare.

In this talk, I will first briefly discuss the importance of clinical trials and cover the basic statistical principles in their design and analysis, providing examples from surgery where clinical trials led to changes in practice. I will then give an overview of my academic background in Biostatistics, and the incentives and tools it provided me with in order to pursue a career as Medical Statistician. I will draw on the work undertaken during my PhD to demonstrate how methodological developments in the design and analysis of trials in complex settings, such as surgery, can be achieved. I will then provide insights of my day-to-day responsibilities as a Senior Medical Statistician within an academic Clinical Trials Unit, and discuss the scope for methodological work within this environment. I will

conclude by giving examples of how some of the methodological work I developed during my PhD could be applied to directly tackle real problems, using examples from my current role.

A Modern Introduction to Randomized Coordinate Descent Methods

Peter Richtárik

University of Edinburgh, Edinburgh, UK.

Randomized coordinate descent (RCD) methods are algorithms of choice for many practical optimization problems arising in engineering, computer science, statistics, applied mathematics and machine learning. They are especially well suited for solving very large dimensional problems ? problems with millions or billions of variables (?coordinates"). A basic variant of RCD in each iteration updates a single coordinate of the decision vector, chosen uniformly at random. This is done either by performing a one-dimensional minimization or, preferably, by applying a simple closed-form formula. RCD methods can therefore be seen as iterative randomized decomposition techniques, which operate by reducing a large-dimensional problem into a sequence of randomly generated one-dimensional problems. A successful and popular strategy for speeding up RCD, both in theory and in practice, is to sample the coordinates with specific problem-dependent non-uniform probabilities. Another very successful strategy is to update a random subset of coordinates of a fixed size instead, possibly in parallel or in a distributed environment. The latter strategy is very closely related to a popular machine learning technique known as mini-batching. RCD methods can be accelerated in the sense of Nesterov, which leads to further theoretical and practical benefits.

In this talk I will give an accessible introduction into the field of randomized coordinate descent algorithms from a modern perspective for which we coined the phrase ?arbitrary sampling?. The methods I shall describe can at every iteration pick and update a random subset of the coordinates (by design, this can be done in parallel), chosen in an i.i.d. fashion according to an arbitrary random set-valued mapping (aka arbitrary sampling). That is, one may in theory assign a unique probability to each of the exponentially many subsets of the set of coordinates, and pick and update coordinates in this manner. With this approach, the line separating the world of randomized and deterministic methods is fully removed ? and what emerges is a unified theory including both randomized and deterministic first-order algorithms in special cases. I shall describe standard, accelerated and primal-dual variants developed in my group ? with the latter being a state-of-the-art method for training linear predictors.

Fast Multilevel Optization Algorithms for Large Scale Machine Learning

Vahan Hovhannisyan

Imperial College London, London, UK.

We present our recent work on multilevel optimization algorithms, focusing on large scale problems arising from machine learning and computer vision applications. First, we propose a multilevel algorithm for solving convex composite optimization problems, particularly for the face recognition problem. The second algorithm uses the idea of multilevel modelling within the (convex and non-convex) robust PCA framework. Both methods exploit the fact that many applications that give rise to large-scale problems can be modelled using varying degrees of fidelity. We show theoretical convergence guarantees as well as numerical experiments revealing the superiority of multilevel algorithms.

Using model selection and changepoints to answer questions about climate

Rebecca Killick

Lancaster University, Lancaster, UK.

All of us will have heard the warnings and numerous claims from scientists around the world about climate change. Some of the analyses are just the equivalent of statistical rubbish but many have sound statistical basis but lead to different conclusions about the data. One such area is around how to categorize the changes we see in our environment and even whether such changes exist. There is debate in the community about whether shifts are due to external forcing (e.g. climate shifts) or random reorganization of the system.

When analyzing large numbers of time series (as if common in climate applications) it is undesirable to have to inspect each series manually to decide on the appropriate model form. In this talk, I will propose an automatic methodology for formal detection of abrupt changes and for distinguishing between forced and unforced (i.e. random) reorganization of the system. For illustration, the methodology is applied to a well-documented marine regime shift in the North Pacific and global mean surface temperature.

What is Data Science?

Neal Ó Riain

Pivigo

The exponential increase in computing power through the 20th century, combined with the accumulation of recorded data at the dawn of the 21st, has given birth to the new field of data science. The growth of this field of has brought about new ways of studying the world, giving us new tools and methods to examine questions that until recently were not quantifiable. Data science offers the opportunity for those who are considering leaving academia, but who want to continue in a research capacity, to utilise their expertise and knowledge in solving interesting problems with real-world applications.

At Pivigo, we support academics in this transition to industry through our S2DS, the largest data science boot-camp in Europe. S2DS is a 5-week programme, hosted in London, during which the participants gain practical data science experience through working on industry projects with our partner companies. Working with these companies - including large businesses like British Gas, Royal Mail, and KPMG; charitable organizations like Just Giving and the NCVO; and exciting start-ups like Growth Intelligence and HAL24K - provides you with real world data science experience that will help accelerate your career transition.