Research into randomness

Professor Dirk Kroese explains the goals and approach of his research into Monte Carlo methods, and describes some of the real-world applications of his work and the challenges his group has had to overcome

To begin, could you outline the aim of your current research project?

It is not uncommon for certain Monte Carlo algorithms to run for weeks or even months. Using sophisticated mathematical ideas and techniques it is possible to devise Monte Carlo algorithms that offer orders of magnitude improvement in efficiency. In addition, such improved Monte Carlo techniques can be applied to a whole range of new applications in engineering, physics, telecommunications, finance and computational biology.

How is Monte Carlo simulation is used in the real world?

Monte Carlo techniques can be used in three different ways. First, we can simulate the random system of interest to understand how it works. For example, if we have a mathematical model for star cluster formations, we may wish to generate various random star clusters and see how they compare with an actual system. Secondly, we can use Monte Carlo to estimate numerical quantities that can be expressed in terms of a random system. A third way in which Monte Carlo is used is to optimise complicated functions. The idea is to search a large space of objects in a randomised way in order to find one or more objects with optimal properties. A typical example is the optimal design of an electricity network or power plant, where many variables (eg. position, capacities) have to be considered so as to minimise operating costs.

What is the importance of the development of the Cross-Entropy (CE) method?

The CE method is a very general and easy method to speed up Monte Carlo simulation, in particular for rare-event simulation. It was long known that one could increase the efficiency of a Monte Carlo estimation method by using a technique called importance sampling, where the system is simulated under different parameters, while compensating for the introduced bias. However, it was not well-known how to choose good alternative parameters in an importance sampling scenario. The CE method gives a simple automated procedure to find the optimal parameters.

Could you highlight the most challenging aspects of your work?

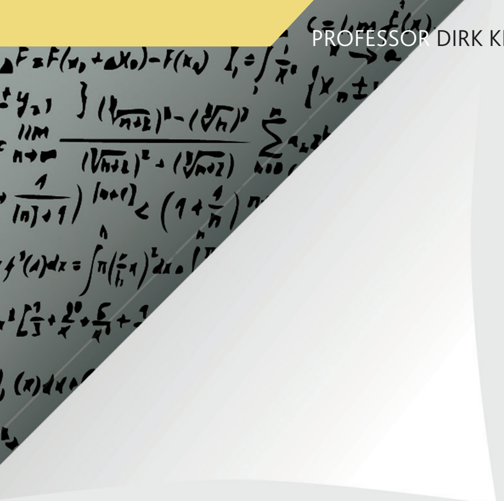
The most challenging aspect is finding the right mathematical framework for the problem at hand, be it an estimation, optimisation or counting problem. Once the mathematical framework for the simulation is established it becomes possible to do probabilistic and statistical analyses of how to best find the quantity of interest. From there, it is a relatively small step to construct a simple algorithm and then to implement and test it. A good example is the simulated annealing algorithm, which was conceived from mathematical models in statistical physics, and is now being used as a simple algorithm to optimise complicated functions. Similarly, the CE method has its foundations in information theory, where the cross-entropy measures the distance between two sampling distributions.

How do you see this work evolving in the future?

The project is continuing to branch into new areas of research. For example, our group at The University of Queensland has started collaborative work with Professor Volker Schmidt (Ulm University) into Monte Carlo methods for spatial processes. Another area of development which I believe will become very important in the future is the parallelisation of Monte Carlo methods; many Monte Carlo algorithms appear ideally suited to massive parallel computing.

Finally, what recommendations would you give to those seeking an entry to the field?

I believe that the best way to learn about randomness (probability and statistics, if you wish) is to conduct simple Monte Carlo experiments. Although computers are increasingly being used in the teaching of statistics, it is usually in the form of black box statistical software. My new book, *Statistical Modelling and Computation*, which will be published by Springer in late 2013/early 2014, takes a more inclusive approach to statistics, where the mathematical and Monte Carlo aspects are integrated, without relying on statistical packages.



Made in Monte Carlo

Researchers at **The University of Queensland** are focusing on increasing the efficiency of a group of computational and mathematical techniques which have broad applications in science, business and technology

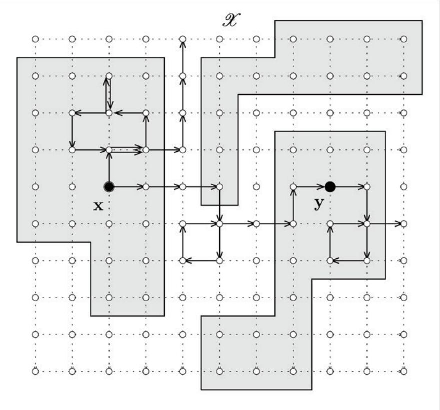
THE WORLD AROUND us is full of systems and processes which we would consider to be random. Randomness is a central component of our natural environment; in science it exists from biology, to physics and astronomy. Understanding patterns in randomness is a mathematical and statistical challenge – a challenge all statisticians and probabilists are attempting to meet.

Monte Carlo algorithms are a collection of mathematical tools that are used to understand and predict complex random systems. While these methods have far-reaching applications across much of science and business, they are far from perfect, necessitating their improvement. This has been recognised by Professor Dirk Kroese and colleagues from the School of Mathematics and Physics at The University of Queensland.

Kroese has always been attracted to mathematics and particularly the understanding of randomness: "I find it fascinating that the precise rules of probability and statistics can be used to understand the seemingly unpredictable world of randomness," he explains. This passion naturally led Kroese to the examination of the Monte Carlo methods – which, while hugely powerful in predicting the outcomes of random systems, are in need of refinement: "With the growing complexity of today's quantitative problems, traditional Monte Carlo methods increasingly face difficulties in terms of speed and accuracy, despite the continual advances in computing power," he observes. Although Monte Carlo methods can be extremely fast, in certain applications they can take up many hours and even days. Due to this and other limitations, Kroese has set his sights on modifying current Monte Carlo methods to create the next generation of algorithms which promise to be more efficient.

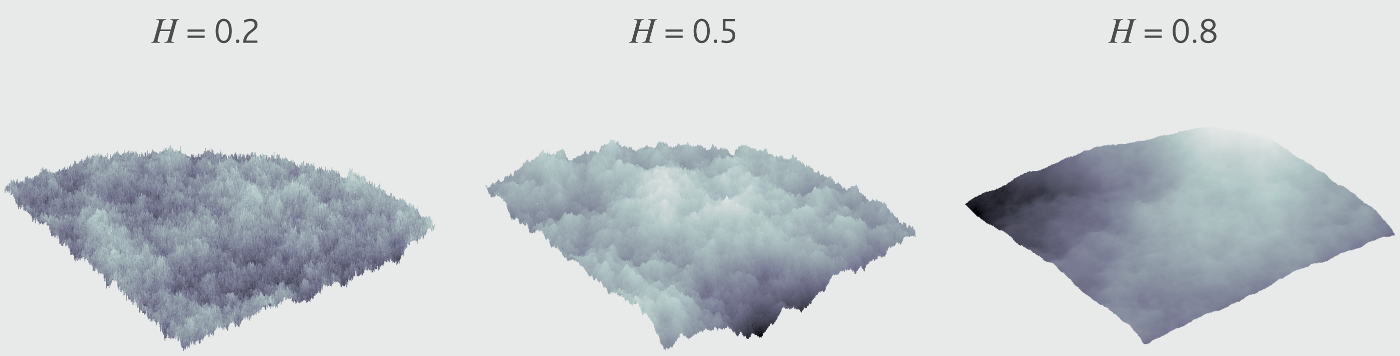
EXPANDING EFFICIENCY AND EMPLOYMENT

Not only do improved Monte Carlo methods enjoy the benefits of reduced computational time and capacity, they also have a broader potential application: "Such improved Monte Carlo techniques can be applied to a whole range of new applications in engineering, physics, telecommunications, finance and computational biology," Kroese points out. His work to improve the functionality of Monte Carlo algorithms has taken several approaches – two of the most promising of these are the cross-entropy (CE) method and the splitting method. The first of these is an improvement in the ability to predict the likelihood of an event with a tiny probability: "The significance of the CE method is that it provides an easy and principled way to estimate the importance parameters by using the same simulation algorithm," he notes. This method is important for understanding how to gain the optimal control over otherwise unpredictable scenarios or processes such as disease outbreaks and business market predictions.

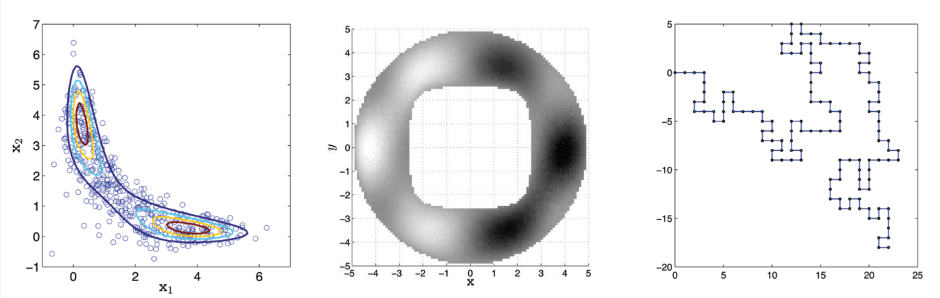


PATH OF A HIT-AND-RUN SAMPLER

A Monte Carlo simulation is the running of a random algorithm – that is, one which is not deterministic, but makes random choices at various stages. While a Monte Carlo simulation is essentially one long and continuous calculation, the splitting method breaks this calculation into multiple copies at various stages. This technique is used to make highly improbable events more likely. Crucially, this reduces the time taken to complete simulations and as such makes Monte Carlo estimations more efficient and cost-effective. This method has been enhanced by a modern approach: "The splitting technique was known



SIMULATION OF FRACTIONAL BROWNIAN LANDSCAPES UNDER DIFFERENT SMOOTHNESS PARAMETERS



EXAMPLES OF RANDOM SAMPLING IN STATISTICS, ENGINEERING, AND MATHEMATICS

as early as the 1950s, but recent theoretical insights have made it much more attractive for both rare-event simulation and randomised optimisation," Kroese highlights.

OUTPUT AND COLLABORATION

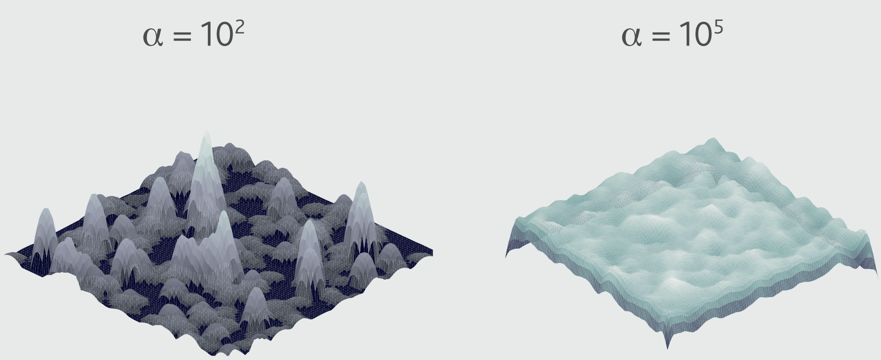
The project being conducted by Kroese and his colleagues is nearing completion in its current format. The team has enjoyed a large amount of success: much like the focus of the work, the research itself is highly efficient, with current research interests leading to 23 publications. This scale of academic output across a range of mathematical and statistical disciplines has required a joint effort, as Kroese fully appreciates: "This project has had valuable input from many people, especially from my former PhD students Zdravko Botev, Tim Brereton, Thomas Taimre and Josh Chan". In addition to the efforts of his students, Kroese has enjoyed the benefits of several academic collaborations with global experts in stochastic simulation. These collaborators include the late Reuven Rubinstein from the Technion, Israel; Søren Asmussen at Aarhus University, Denmark; Peter Glynn of Stanford University in the US; and Pierre L'Ecuyer, based at the University of Montreal, Canada. This breadth and depth of expertise has allowed Kroese and his research group to make valuable contributions to the efficiency of Monte Carlo methods – promising benefits across a huge range of industry and science.

In spite of these achievements, the group from The University of Queensland are far from done. Kroese and his colleagues have produced a recent proposal to improve Monte Carlo methods concerned with the optimisation of two- and three-dimensional spatial systems. Such systems are hugely diverse and range from electricity networks to functional materials used in solar cells. Monte Carlo methods in this context can be extremely challenging. Any meaningful improvements would notably enhance our ability to understand complex spatial systems.

Kroese and his colleagues' work has branched into many aspects of statistical and mathematical optimisation, but the central aim of all of these sub-projects remains the same: "The ultimate impact, I hope, is that people see randomness not as an unavoidable nuisance, but as a powerful tool for computation and understanding," he elaborates. Kroese hopes that his work will empower others to apply a new generation of Monte Carlo algorithms to systems which are otherwise extremely challenging to understand. Only then can we get a firm grip on elucidating the randomness existing in our lives today, but more importantly, employ this randomness to understand complex processes.

ETYMOLOGY

Monte Carlo simulation is the act of carrying out random experiments on a computer. It operates by performing multiple trial runs using random variables. The simulation's name derives eponymously from the district in the principality of Monaco. Monte Carlo is famous for its casinos. Gambling pursuits such as roulette, dice and slot machines epitomise the application of randomness.



SIMULATION OF LEVY RANDOM FIELDS

INTELLIGENCE

IMPROVED MONTE CARLO METHODS, FOR ESTIMATION, OPTIMISATION, AND COUNTING

OBJECTIVES

The current prevalence of Monte Carlo methods in science is a tribute to their usefulness for solving real-world problems. However, as the complexity of such problems continues to grow, classical methods increasingly run into problems in terms of speed and accuracy. The last few years have seen exciting new theoretical and conceptual developments in Monte Carlo techniques, providing dramatic improvements in performance. The aim of the project is to develop new theory and applications for the next generation of Monte Carlo methods, and to provide much-needed generic algorithms and software for solving complex estimation, optimisation, and counting problems over a wide range of applications.

KEY COLLABORATORS

Zdravko Botev, Thomas Taimre, Joshua Chan, Tim Brereton, Reuven Rubinstein, Søren Asmussen, Peter Glynn, Pierre L'Ecuyer and **Volker Schmidt**

FUNDING

Australian Research Council (Discovery Project DP0985177)

CONTACT

Professor Dirk P Kroese

School of Mathematics and Physics
The University of Queensland
Brisbane 4072
Australia

www.maths.uq.edu.au/~kroese

DIRK P KROESE is a professor of Mathematics and Statistics at the School of Mathematics and Physics of The University of Queensland, Australia. He is currently the recipient of a Professorial Fellowship (2009-13) of the Australian Research Council. His present interest lies in the area of Monte Carlo methods, which involves the simulation of random systems in order to solve numerical problems in science, engineering and finance. Kroese has made significant international contributions to Monte Carlo simulation and applied probability, and is a pioneer of the cross-entropy method, which is being used by many researchers and practitioners across the world. His monographs, *Simulation and the Monte Carlo Method, 2nd Ed.* and *The Cross-Entropy Method* (both with Rubinstein), have highly influenced the field of simulation and computational statistics in recent years, as evidenced by over 4,000 citations. His most recent monograph, *Handbook of Monte Carlo Methods*, with Taimre and Botev (see www.montecarlohandbook.org), provides a comprehensive state-of-the-art account of the major topics in Monte Carlo.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA