

Professor Arnaud Doucet  
Department of Statistics  
Oxford University, UK  
doucet@stats.ox.ac.uk

October 28, 2018

Reference for Mr. Rob Cornish

I am writing this letter in very strong support of the application of Rob Cornish for an internship at DeepMind. I am currently a Statutory Professor (i.e. Professor with Chair) in the Department of Statistics of Oxford University and I am supervising the PhD of Rob. I am a specialist of computational statistics, statistical machine learning and Monte Carlo methods.

Rob has obtained two BSc degrees with top marks: one in sciences from Melbourne University specializing in Electrical Engineering and the other one in Mathematics and Statistics from Monash University. He has received numerous prizes during his undergraduate studies in both Electrical Engineering and Mathematics. He has done additionally internships in a programming language group, in robotics and computer vision. One of this internships has led to an international publication. He became interested at the time in Machine Learning and applied for a position in the Center for Doctoral Training in Autonomous Intelligent Machines and Systems run by the Department of Engineering in Oxford University. This is a very selective program which receives several hundred applications per year but he was logically admitted given his achievements.

The first year of the CDT is mostly devoted to courses and, towards the end of this year, Rob started his PhD under the supervision of Frank Wood in the department of Engineering. During his time under the supervision of Frank Wood, he worked on various topics related to probabilistic programming, gradient descent techniques and Monte Carlo methods. For example, the *ICML* 2018 paper “On Nested Monte Carlo Estimators” investigates the properties of estimators appearing in applications as diverse as variational auto-encoders, probabilistic programming and Bayesian experimental design. It provides a careful analysis of these estimators, shows when they are consistent and provide guidelines to avoid a number of common pitfalls. This is an excellent and thorough piece of work. The *ICLR* 2018 paper is another illustration of the work of Rob. This paper entitled “Online learning rate adaptation with hypergradient descent”. This paper presents a simple methodology to improve the convergence rate of gradient-based optimizers including SGD, ADAM and SGD with acceleration by learning the learning rate using another gradient descent technique. The excellent empirical performance of this method have been demonstrated on deep networks.

Frank Wood left Oxford for UBC in March this year. He had a large research group and many of his students had to find new supervisors in Oxford. When Frank asked me whether I wanted to supervise one of them, I admit that I was not enthusiastic as I already had a fair number of students. However, he told me that I should really meet his brightest student, Rob, before deciding. He was right. I was really impressed after my first meeting with Rob and agreed to supervise him. He has ever since exceeded all my expectations.

As soon as Rob joined my group, he also joined the reading group we had on probability in high dimensions. Not only did Rob catch up extremely quickly with the highly technical material but he was also one of the only two students who could solve the most challenging exercises.

On the research side, he has proved very impressive too despite having changed PhD subject. We have been working together on the development and analysis of Markov chain Monte Carlo (MCMC) schemes to carry out Bayesian inference for “big data” problems. In these scenarios, it is well-known that standard methods such as the Metropolis–Hastings scheme are far too costly. Techniques based on sub-sampling/mini-batches have thus also been developed in this Bayesian context, this includes Stochastic Gradient Langevin Dynamics (SGLD) and Firefly. However, these methods suffer from various drawbacks; e.g. SGLD does not preserve the posterior as invariant distribution (if one wants to obtain consistent estimate then one loses the standard Monte Carlo rate of convergence) while Firefly can easily get stuck. Recently, alternative methods based on non-reversible MCMC schemes have been proposed to tackle these problems but these continuous-time algorithms are difficult to analyze and their behaviour in big data scenarios is not yet understood. Rob has developed a novel class of discrete-time MCMC algorithms based on subsampling ideas. These new algorithms are easy to put in practice. Moreover, Rob has been able to analyze them theoretically, establishing geometric convergence of these schemes under reasonable and verifiable assumptions. This is a truly remarkable achievement. We are currently writing up this material for *ICML*.

I can only recommend Rob Cornish in the strongest possible terms for an internship at DeepMind. He is a very bright and creative young researcher. He has additionally excellent programming skills (including C++, TensorFlow and PyTorch). Moreover, from a personal point of view, he has excellent interpersonal skills.

Yours sincerely,

A handwritten signature in black ink, appearing to be 'AD' or 'Arnaud Doucet'.

Arnaud Doucet