

Department of Systems Engineering and Engineering Management

Seminar Series

An Afternoon with Lancelot James

Prof. Lancelot F. James

Professor of Information Systems and Operations Management
The Hong Kong University of Science and Technology

Date:	30 th October 2014 (Thursday)
Time:	2:30pm (Tea/Coffee service at 3:45pm)
Venue:	LT10, 4/F, AC1

Abstract

I. Stick-breaking, Dirichlet processes and all that. A retrospective. (2:30-3:45pm)

The Dirichlet Process has been the centerpiece of Bayesian Nonparametric Statistical Modelling since the work of Ferguson (1973). Nonparametric in this sense can be viewed as a natural extension of a Bayesian Dirichlet-Multinomial model to Dirichlet like priors over the infinite simplex. The area has matured substantially since that time and one now routinely sees exciting applications and innovations in related fields such as statistical Machine Learning and Natural language Processing. Broadly these techniques have played a role in what is now being called Big data problems. An important component of this explosion has been what is referred to as the stick-breaking representation of the Dirichlet process. I will talk about this, its origins and connections to population genetics and also related

processes. Furthermore we will discuss some developments since my work with Hemant Ishwaran in 2001, which helped to popularize many of these ideas in BNP and Statistical Machine Learning.

II. Stickbreaking Pitman-Yor and randomized generalized gamma priors. (4:00-5:00pm)

The stick-breaking representation of the Dirichlet process (DP) has been a fruitful tool for statistical inference and modeling as well as providing key insights into its underlying structure. There is a natural generalization of the DP which is referred to as Pitman-Yor process, wherein it also has a highly tractable stick-breaking representation. However this model also possesses features which distinguish itself from the DP, namely its power law behavior in relation to the size of clusters, which is highly relevant to certain language models and other entities. Formally stick-breaking of this form is derived from the notion of size-biased sampling from a discrete random probability measure. This framework is rather restrictive as the only models that possess a stick-breaking structure based on INDEPENDENT variables are in effect the DP and its extension the Pitman-Yor process. As such there are very few explicit representations for other models that might be of interest. Here I discuss results for a popular class of models, and an extended version of it, that actually contains the aforementioned processes.

About the Speaker

Prof. Lancelot F. James is currently a Professor in the Business School at the Hong Kong University of Science and Technology. He has been at HKUST since 2001. Prior to that, he was an Assistant Professor in the School of Engineering at the Johns Hopkins University in Baltimore.

He is well known for his work in Bayesian nonparametric statistics.

Since the early 2000s, he has advocated and developed ideas related to the usage of Chinese restaurant processes, stick-breaking priors, and Pitman-Yor processes. Since that time, these colourfully named

processes (the latter two named by James and his co-author Hemant Ishwaran in 2001) have played a major role in the development of intricate applications in Bayesian Statistical Machine Learning and Bayesian nonparametric methods in general, all of which are pertinent to the analysis of Big Data. Prof. James is an elected member of the International Statistical Institute.

Since 2008 he is an elected Fellow of the Institute of Mathematical Statistics where he is cited for contributions to Bayesian nonparametric statistics, the development of Poisson partition calculus for Levy processes, and for dedicated service to IMS.

Professor James was born on the island of Jamaica in the Caribbean Sea and grew up in Westchester County, New York.

Enquiry: 3442 8408

All are Welcome!

SEEM Seminar 2014-2015/015