Random Measures and Measure-Valued Processes
Titles and abstracts

- **Jochen Blath** (TU Berlin)

  On the scaling limit of the interface of the symbiotic branching model

  We investigate the limiting process of the interface of the continuous-space symbiotic branching model in the negative correlation regime under diffusive rescaling. In particular, we provide tightness of the rescaled systems in suitable topologies and uniqueness of the limiting martingale problem. We then discuss some properties of the shape of the limit.

  This is joint work (in progress) with Matthias Hammer (TU Berlin) and Marcel Ortgiese (TU Berlin)

- **Donald Dawson** (Carleton University)

  A dual approach to models with multilevel selection

  The question of "levels of selection" has received a great deal of attention and controversy in the biological literature but relatively little attention in the mathematical literature. In this lecture we brie y review some background and formulate some models of multitype populations undergoing mutation and selection including multilevel selection. The main tool is a related dual processes which is used to analyse their behaviour.

- **Jean-François Delmas** (Cermics Paris)

  Conditioned Galton-Watson trees

  We describe the limits of ALL critical or subcritical Galton-Watson trees conditioned to have a large number of nodes or a large number of leaves or more generally a large number of nodes having a degree in a given set. The limit is an infinite tree with an infinite spine in the generic case (which includes critical Galton-Watson trees) or an infinite tree with a node of infinite degree in the non-generic case. Most of those results are well known. We present an elementary proof for those results. It relies on a particularly handy family of sets of trees which is convergence determining, the Dwass formula and the strong ratio limit property for random walks.

- **Alison Etheridge** (Oxford University)

  The spatial Lambda-Fleming-Viot process and friends (provisional title)

  The spatial Lambda-Fleming-Viot process refers to a whole class of models that have been introduced to model frequencies of different genetic types in a population evolving in a spatial continuum. In this talk we examine some instances of that process and their relationship to more familiar stochastic processes.

- **Shizan Fang** (Université de Bourgogne)

  Fokker-Planck equations on Lie groups

  We will use De Giorgie approximation to construct Fokker-Planck equations on Lie groups and related topics.

- **Stefano Favaro** (University of Torino and Collegio Carlo Alberto)

  Bayesian nonparametric inference for discovery probabilities

  Species sampling problems have a long history in ecological and biological studies and a number of statistical issues, including the evaluation of species richness, the design of sampling experiments and the estimation of rare species variety, are to be addressed. Such inferential problems have recently emerged also in genomic applications, however exhibiting some peculiar features that make them more challenging: specifically, one has to deal with very large genomic libraries containing a huge number of distinct genes and only a small portion of the library has been sequenced. These aspects motivate the Bayesian nonparametric approach we undertake, since it allows to achieve the degree of flexibility typically needed in this framework. Given an initial observed sample of size n, focus will be on prediction of a key aspect of the outcome from an additional sample of size m, namely the so-called discovery probability. In particular, conditionally on the observed initial sample, we derive a novel estimator of: i) the probability of detecting at the $(n + m + 1)$-th observation a new species and ii) the probability of detecting at the $(n+m+1)$-th observation a species that have been observed with any given frequency in the enlarged sample of size $(n+m)$. We also characterize the large m asymptotic behavior of the discovery probabilities in i) and ii) by means of almost sure fluctuation limits and large deviation principles. The proposed analysis lead to quantify both the rate at which rare species are detected and the achieved sample coverage of abundant species, as m increases. Natural applications are represented by the estimation of the probability of discovering rare genes within genomic libraries and the results are illustrated by means of two Expressed Sequence Tags datasets.

- **Clément Foucart** (Université Paris Nord)

The impact of selection in the Lambda-Wright-Fisher model

The purpose of the talk is to study some asymptotic properties of the Λ-Wright-Fisher process with selection. The resampling mechanism is governed by a finite measure Lambda on [0,1] and the selection by a parameter alpha. According to the measure Lambda, some particular behaviours in the frequency of the allele can occur. In some precise cases, the selection coefficient alpha may be large enough to compensate the random genetic drift. In other words, for certain selection pressure, the disadvantaged allele will vanish asymptotically. This phenomenon has been first discovered by Der Esptein and Plotkin when Lambda is a Dirac measure. We study the dual process of the Λ-Wright-Fisher process with selection and prove this result through martingale arguments.

- **Robert Griffiths** (Oxford University)

  Generalized Wright-Fisher Diffusion processes

  A generalization of the d-dimensional Wright-Fisher diffusion process is to a process $\{P(t),\ t \geq 0\}$ with a generator

  $$\mathcal{L} = \frac{1}{2} \sum_{i,j \in [d]} \sigma_{ij}(p) \frac{\partial^2}{\partial p_i \partial p_j} + \sum_{i \in [d]} \mu_i(p) \frac{\partial}{\partial p_i},$$

  where

  $$\sigma_{ij}(p) = \sum_{r \in [d]} p_r \mathrm{Cov}\big(\zeta_{ri} - p_i \zeta_r, \zeta_{rj} - p_j \zeta_r\big).$$

  $\left\{ \{\zeta_{rk}\}_{k \in [d]};\ r \in [d] \right\}$ are collections of non-negative random variables, independent between different $r \in [d]$ with $\zeta_r = \sum_{k \in [d]} \zeta_{rk}$. Typically $\mu_i(p)$ has a form which reflects selection in the model. The process can be derived from the limit in a conditioned discrete multitype branching process in a constant-sized population of individuals of d-types. This talk will describe the discrete model limit; give examples of such processes; and briefly discuss a connection between multitype diffusion branching processes with non-identical rates. This is joint research with Alison Etheridge.

- **Kenji Handa** (Department of Mathematics Saga University)

  Ergodic properties for a class of generalized Fleming-Viot processes

  We discuss a Markov jump process which is regarded as a variant of the Fleming-Viot process with parent-independent mutation. After identifying its unique stationary distribution, an ergodic property is shown for the generalized Fleming-Viot process. The proof is based on both the distinguished relationship to a measure-valued branching process with immigration whose jump mechanisms are governed by certain stable laws and the spectral gap estimate for the branching process.

- **Lancelot F. James** (Hong Kong University of Science and Technology)

Another look at Pitman and Yor's Proposition 21

In this talk we will focus on another result of Pitman and Yor's 1997 paper on the remarkable two parameter Poisson Dirichlet family, with law denoted as $PD(\alpha, \theta)$ for $\theta > -\alpha$, which is Proposition 21. As read, it describes a representation of $PD(\alpha, \theta)$ mass partitions, for the less general range of $\theta \geq 0$, in terms of a generalized gamma subordinator whose jumps are restricted to a (random) gamma distributed time. However a closer look at the proof reveals a much larger class where one can replace the gamma time by an arbitrary non-negative random variable ζ. I call this class $PG(\alpha, \zeta)$. Noting that while this class is extremely general, it does not contain the $PD(\alpha, \theta)$ for the case where $-\alpha < \theta < 0$, I construct a new class called extended $PG(\alpha, \zeta)$ or $EPG(\alpha, \zeta)$, which contains the most popular discrete random probability measures and many more. Although less friendly analytically than the $PD(\alpha, \theta)$ class, various remarkable properties of this class are revealed. From a practical point of view an explicit stick-breaking representation is derived for the entire range of parameters. This is not achieved by working directly with formulae for the various joint densities which as it turns out is fairly impossible. Furthermore Markov chains are derived which establish a direct link between Markov chains derived from stick-breaking, i.e. successive size biased sampling excursions of excursion lengths and excision(deletion)/insertion, as described in Perman, Pitman and Yor (1992) and Markov Chains derived from successive usage of coagulation fragmentation operators described in Bertoin and Goldschmidt and Dong, Goldschmidt and Martin. This has potential implications to certain types of fragmentation trees and coalescents. There are other interesting results as well. Naturally these results are suggestive of many possible new models for applications in Bayesian Nonparametrics/Machine Learning, where $PD(\alpha, \theta)$ bridges are often referred to as Pitman-Yor processes, although it is presented for a wider audience. The name Pitman-Yor process was coined by Hemant Ishwaran and myself in a 2001 JASA paper.

References

[1] BERTOIN, J. and GOLDSCHMIDT, C. (2004). Dual random fragmentation and coagulation and an application to the genealogy of Yule processes. In *Mathematics and computer science III: Algorithms, Trees, Combinatorics and Probabilities*, M. Drmota, P. Flajolet, D. Gardy, B. Gittenberger (editors), pp. 295–308. Trends Math., Birkhäuser, Basel.

[2] DONG, R., MARTIN, J. and GOLDSCHMIDT, C. (2005). Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees. *Ann. Appl. Probab.* **16** 1733-1750.

[3] ISHWARAN, H. and JAMES. L.F.. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.

[4] PERMAN, M., PITMAN, J. and YOR, M. (1992). *Size-biased sampling of Poisson point processes and excursions. Probab. Theory Related Fields.* **92**, 21-39.

[5] PITMAN, J. and YOR, M.. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.

- **Antonio Lijoi** (University of Pavia)

  Vectors of random probability measures for Bayesian inference

  Random probability measures are key ingredients in Bayesian inference since their law acts as a prior distribution, the most well-known example being the Dirichlet process. A considerable body of the recent Bayesian nonparametric literature is on the proposal and the study of dependent random probability measures that are suited for the analysis of non-exchangeable data. In this talk, classes of priors based on transformations of vectors of dependent completely random measures will be presented. Characterizations of the posterior distribution will be displayed and used for estimating quantities of statistical interest in density estimation, clustering and survival analysis. If time allows, applications of the proposed methodology will be illustrated on real and simulated datasets.

- **Sylvie Méléard** (École Polytechnique, Palaiseau)

  Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks

  This talk presents a work in progress with Sylvain Billard, Régis Ferrière and Chi Viet Tran. Much of the genetic diversity measured by population geneticists involves molecular sequences assumed to be selectively neutral. A longstanding question in evolutionary theory is understanding how variation in such a marker evolves.This raises the motivation for modeling a co-evolutionary process of adaptive trait and neutral marker dynamics. The model will relax the key assumptions of constant population size and selection coefficients to study the dynamics and the fixation of the neutral marker. We start with an individual-based model where individuals are characterized by an adaptive trait that influences ecological interactions among individuals, and by a genetic marker that is selectively neutral. The ecological and evolutionary dynamics of the population result from clonal reproduction, mutation and competition between individuals. The main assumption is that the marker mutation process is much faster than the trait mutation process but much slower than the ecological time scale of birth and death events. Therefore, there are three time scales in the model. We study the joint process of trait and marker

dynamics on the trait mutation time scale. The population size stabilizes in a neighborhood of the ecological equilibrium and jumps to another equilibrium when a successful trait mutant goes to fixation in the population. This is the well known Trait Substitution Sequence dynamics of the adaptive trait. The TSS is a jump process on the trait space, whose jumps correspond to invading mutant traits. The study of the marker distribution during the invasion period requires careful consideration of the individual process and of the different scales involved. Until the next jump of the TSS, the marker evolves as a stochastic distribution-valued process. In the case where the marker mutation effects are continuous and small, this process is a Fleming-Viot process whose drift and covariance depend on the resident adaptive trait. Therefore, the collated dynamics defines a measure-valued diffusive process with jumps that we call Substitution Fleming-Viot Process (SFVP). From a biological standpoint, we recover the hitch-hiking phenomenon and genetical bottleneck and explain the restoration of diversity.

- **Martin Möhle** (Mathematical Institute Eberhard Karls, University of Tübingen)

  Conditions for exchangeable coalescents to come down from infinity

  An improved condition to come down from infinity for exchangeable coalescent processes with simultaneous multiple collisions of ancestral lineages is provided. For so-called non-critical coalescents this leads to an improved necessary and sufficient condition for the coalescent to come down from infinity. An analog conjecture for the full class of exchangeable coalescents is presented. New examples of critical coalescents are studied.

- **Peter Orbanz** (Columbia University)

  Random measures generating binary matrices and their application in Bayesian statistics

  A class of problems that has recently attracted considerable attention in statistics involves random binary matrices. These matrices can be regarded as generalizations of exchangeable partitions, to the case where blocks of the partition are not disjoint; each element of the underlying set can be contained in multiple, possibly overlapping blocks. The Bayesian approach to these problems gives rise to random discrete measures with weights in the unit interval, but no global normalization constraint. The prototypical model—roughly the analogue of the Poisson-Dirichlet—is a Poisson random measure whose intensity resembles a beta distribution, and which has become known as the 'beta process' in statistics. I will survey this class of models and also tie in some recent results.

- **Juan Carlos Pardo** (CIMAT Guanajuato)

  Total internal and external lengths of the Bolthausen-Sznitman coalescent

  In this talk, we study a weak law of large numbers for the total internal length of the Bolthausen-Szmitman coalescent. As a consequence, we obtain the weak limit law of the

centered and rescaled total external length. The latter extends results obtained recently by Dhersin & Möhle.

This is a joint work with Arno Siri-Jegousse (CIMAT) and Goetz Kersting (University of Frankfurt)

- **Peter Pfaffelhuber** (University of Freiburg)

  Some large deviation results in Kingman's coalescent

  Kingman's coalescent is well-known as the dual to the Fleming-Viot measure-valued process. In this talk we consider some large deviation results in this random tree, which are based on the following almost sure limit laws: 1) Let N_t be the number of blocks at small time t. Then, $tN_t \to 2$ as $t \to 0$. 2) Let $U_1, ..., U_{N_t}$ be the family sizes of the N_t families at time t. Then, $n \sum_{i=1}^{N_t} U_i^2 \to 2$ as $t \to 0$. Connected to 1), a principle of large deviations is shown, while for 2), we provide some large deviation bounds. This is joint work with Andrej Depperschmidt and Annika Scheuringer.

- **Lea Popovic** (Concordia University)

  Ancestral Features of Multitype Branching Trees

  We develop a multi-type coalescent point process description of the standing population of a general branching tree with finitely many different types. This process can be used to determine various statistical properties of the ancestral tree of the standing population, e.g. the time to the most recent common ancestor (MRCA) of individuals of different types, and times of MRCAs of individuals of the same type. We investigate the dependence of these properties on the offspring distribution and show how asymmetrical type distributions vastly affect the properties of the ancestral tree. Explicit formulae for two-type branching process are given in the special case of multitype linear-fractional offspring distribution.

  This is joint work with Mariolys Rivas

- **Igor Pruenster** (University of Torino & Collegio Carlo Alberto)

  Frequentist asymptotic behaviour of Gibbs-type priors

  Gibbs-type priors represent a natural generalization of the Dirichlet process and are increasingly used in the context of Bayesian nonparametric infernce. In this talk their asymptotic behaviour will be studied from a frequentist point of view. In particular, we shall focus on posterior consistency and display conditions under which the posterior distribution accumulates in suitable neighborhoods of what is assumed to be the 'true' data generating distribution. Examples featuring inconsistency will also be discussed in

view of the valuable insight they provide on general modeling issues that arise in Bayesian nonparametric inference.

- **Matteo Ruggiero** (University of Torino & Collegio Carlo Alberto)

  Optimal filtering and the dual process

  We link optimal filtering for hidden Markov models to the notion of duality for Markov processes. We show that when the signal is dual to a process that has two components, one deterministic and one a pure death process, and with respect to functions that define changes of measure conjugate to the emission density, the filtering distributions evolve in the family of finite mixtures of such measures and the filter can be computed at a cost that is polynomial in the number of observations. Examples include models where the signal is a Cox-Ingersoll-Ross diffusion or a K-dimensional Wright-Fisher diffusion. We will also discuss extensions to signals driven by Fleming-Viot processes with parent independent mutation.

  Joint work with Omiros Papaspiliopoulos

- **Dario Spanò** (Department of Statistics, University of Warwick)

  Polynomial Spectrum of time-dependent Gamma and Dirichlet random measures

  I will give an overview of some results and open problems related to the question: when do symmetric transition kernels for (continuous time)-dependent random measures, with gamma or Dirichlet stationary law, map polynomials to polynomials of the same degree? All kernels with such a property are interesting as they allow for a computable series representation and reveal a tractable combinatorial and probabilistic interpretation in connection with hypergroup theory. They are therefore potentially useful for applications e.g. as time-dependent parameters in nonparametric Bayesian inference. I will show that Markov polynomial kernels with gamma invariant law can be interpreted as the transition function of some subordinated Dawson-Watanabe measure-valued branching diffusion with immigration. Similarly, subordinated Fleming-Viot processes with neutral parent independent mutation yield Markov polynomial kernels in the Dirichlet family. The connection between the two families is still to be fully understood but some insight is provided with the help of some bits of special function theory and some bits of Population Genetics.

- **Anja Sturm** (University of Göttingen)

  Long term behavior of subcritical contact processes

  In this talk we consider the long-time behavior of the law of a contact process started with a single infected site, distributed according to counting measure on the lattice. This

distribution is related to the configuration as seen from a typical infected site and gives rise to the definition of so-called eigenmeasures, which are possibly infinite measures on the set of non empty configurations that are preserved under the dynamics up to a multiplicative constant. We show that contact processes on general countable groups have in the subcritical regime a unique spatially homogeneous eigenmeasure. We also discuss applications of this result, in particular regarding the behavior of the exponential growth rate of the process as a function of its death rate.

This is joint work with Jan Swart (UTIA Prague).

- **Wei Sun** (Concordia University)

Markov Processes and Semi-Dirichlet Forms

Dirichlet forms (more generally, semi-Dirichlet forms) are used to study Markov processes. Two of their advantages are in handling the cases of differential operators with singular coefficients and infinite dimensional state spaces. Roughly speaking, if a Markov process is time-reversible then it is associated with a symmetric Dirichlet form and hence can be studied by the Dirichlet forms theory. However, many measured-valued Markov processes are not time-reversible. For example, a Fleming-Viot process is time-reversible if and only if the mutation, selection, and recombination have a special form. In this talk, we will consider the relationship between Markov property and symmetry. We will introduce our recent results on representation of Markov processes, Hunts hypothesis (H) for Markov processes and Getoors conjecture for Levy processes, Fukushimas decomposition for semi-Dirichlet forms.

- **Yee Whye Teh** (Oxford)

Modelling Genetic Variations using Fragmentation-Coagulation Processes

Abstract:

Hudson's coalescent with recombination is a well-known model of genetic variation in populations with recombination. With growing amounts of population genetics data, demand for probabilistic models to analyse such data is strong, and the coalescent with recombination is a very natural candidate. Unfortunately posterior inference in the model is intractable, and a number of approximations and alternatives have been proposed.

A popular class of alternatives are based on hidden Markov models (HMMs), which can be understood as approximating the tree-structured genealogies at each point of the chromosome with a partition of the observed haplotypes. However HMMs suffer from two problems. Firstly, they are parametric and requires either a user specified number of states or expensive model selection procedures. Secondly, due to the way HMMs parametrize

partitions using latent states, they suffer from significant label switching issues affecting the quality of posterior inferences.

We propose a novel Bayesian nonparametric model for genetic variations based on Markov processes over partitions called fragmentation-coagulation processes. Our model is based on a particularly simple class of Berestycki's exchangeable fragmentation-coalescence processes with nice properties: they are reversible and stationary Markov process which evolves via binary fragmentations and coagulations. Statistically, our model can infer the number of states easily and automatically, and does not suffer from the label-switching issues of HMMs. Inference is achieved using an efficient Gibbs sampling algorithm, and we report encouraging results on genotype imputation.

Joint work with Lloyd Elliott and Charles Blundell.

- **Anton Wakolbinger** (University of Frankfurt)

  The spatial Lambda-Fleming-Viot process: an event-based construction and a lookdown representation

  We report on the construction a measure-valued equivalent to the spatial Lambda-Fleming-Viot process (SLFV) introduced by Etheridge (2008). By fixing the realization of the sequence of reproduction events we obtain a quenched evolution of the local genetic diversities. To this end, we use a particle representation which highlights the role of the genealogies in the attribution of types (or alleles) to the individuals of the population. This construction also enables us to clarify the state-space of the SLFV and to derive several path properties of the measure-valued process as well as of the labeled trees describing the genealogical relations between a sample of individuals. We complement it with a look-down construction which provides a particle system whose empirical distribution at time t, seen as a process in t, has the law of the quenched SLFV. This is joint work with Amandine Veber.

- **Anita Winter** (University Duisburg-Essen)

  Convergence of bi-measure R-trees and the subtree prune process

  In 1998 Aldous and Pitman constructed a tree-valued Markov chain by pruning off more and more subtrees above randomly chosen edges of a Galton-Watson tree. More recently Abraham, Delmas and He considered a similar process, where the cut-points are chosen in a degree-dependent way. In the same spirit prunings of continuum trees were studied by various authors. However, so far no precise link between the prunings of discrete and continuum trees has been given.

  In this talk we encode trees as metric measure spaces and equip them additionally with a pruning measure, and provide a topology on the space of bi-measure R-trees. We then

construct THE subtree prune process and show that convergence of initial states implies convergence of the whole bi-measure valued paths.

(joint work with Wolfgang Löhr and Guillaume Voisin)

- **Jie Xiong** (University of Macau and University of Tennessee)

  Some nonlinear SPDEs from measures valued processes

  In this talk, I will discuss three classes of nonlinear SPDEs arising from the study of continuous state branching processes in random environment. Some techniques developed for the study of these SPDEs will be introduced.

- **Feng Yu** (University of Bristol)

  Decoupling of linkage disequilibrium via recombination

  Natural selection is central to the understanding of how species evolve and adapt. The advantage of sexual reproduction (compared to asexual reproduction) can be understood via the decoupling of linkage disequilibrium between allele frequencies at linked selected loci, thus increasing the fitness variance of the population. We present two settings where this effect can be verified by explicit calculations. The first setting focuses the accumulation of deleterious and advantage mutations, where we employ Girsanov transformation to change the selected measure to the neutral one. The second setting turns to the case of stationary distribution of allele frequencies at two selected loci undergoing mutation, selection and recombination. The technique employed in both settings is an expansion in small Ns and Nr. If time allows, we will also present some results on expansion results with large Ns (strong selection).

- **Xiaowen Zhou** (Concordia University)

  The Modulus of Continuity for Λ-Fleming-Viot Processes with Brownian Spatial Motion

  A Fleming-Viot process is a probability-measure-valued stochastic process for population genetics. It describes the evolution of relative frequencies for different types of alleles in a large population that undergoes reproduction and mutation.

  In this talk we first briefly go over the Λ-coalescent of multiple collisions and the lookdown representation of Donnelly and Kurtz for Λ-Fleming-Viot process with Brownian spatial motion. An ancestry process can be recovered from the lookdown representation to keep track of genealogy of the individuals in the population. We then present a modulus of continuity result on the ancestry process, which can be applied to study the modulus of continuity and disconnectedness for the Λ-Fleming-Viot support.

This talk is based on Liu and Zhou [1,2] and work in progress.

References

[1] H. Liu and X. Zhou (2012). Compact support property of the Λ-Fleming-Viot process with underlying Brownian motion, *Electron. J. Probab.*, *17*, No. 73, 1-20.

[2] H. Liu and X. Zhou (2013). Some support properties for a class of Λ-Fleming-Viot processes. Submitted. Available at http://arxiv.org/abs/1307.3990.