

Probability, Population Genetics and Evolution

CIRM, June 11 to 15, 2012

TITLES AND ABSTRACTS

Vincent Bansaye

Small positive values for supercritical branching processes in random environment

ABSTRACT: Branching process in random environment (BPRE) are a generalization of Galton-Watson processes where the reproduction law is picked randomly in an i.i.d manner in each generation. In the supercritical case, the population Z_n either explodes (with a positive probability) or becomes extinct. We are interested in the event when the process Z_n takes positive but bounded values for large times n . We are characterizing the asymptotic behavior of $P(0 < Z_n < k)$. We are also interested by the form of the tree of the population conditionally on $Z_n = k$. We will describe the position of the MRCA (most recent common ancestor) in the linear fractional case and two regimes appear. Finally, we will derive the rate function for lower large deviations of BPRE.

Nick Barton

Mathematical problems in population genetics

ABSTRACT: Population genetics deals with a broad range of interesting, but difficult, mathematical problems. These may be deterministic, involving multiple interacting genes, or stochastic, with an interaction between random drift and selection. I give a rough survey of the classical theory and its recent development, and discuss the relation between computational and analytical approaches.

Julien Berestycki

Branching processes and selection

ABSTRACT: Let us consider a model in which particles move in space and branch independently of one another but on which a selection mechanism is acting to keep the population size essentially constant. Several such models have recently been investigated by physicists (Brunet, Derrida et al.). Non-rigorous arguments led them to formulate several remarkable predictions for those systems. In particular they argued that the asymptotic genealogy of the particles should be a

universal object: the Bolthausen-Sznitman coalescent. I will give an overview of some of those conjectures as well as some recent related rigorous results.

Based on joint work with N. Berestycki and J. Schweinsberg.

Jean Bertoin

The cut-tree of large Galton-Watson trees and the Brownian CRT

ABSTRACT: This is a joint work with Grégory Miermont (Orsay). Consider the edge-deletion process in which the edges of some finite tree are removed one after the other in the uniform random order. Roughly speaking, the cut-tree then describes the genealogy of connected components appearing in this edge-deletion process. Our main result shows that after a proper rescaling, the cut-tree of a critical Galton-Watson tree with finite variance and conditioned to have size n , converges as $n \rightarrow \infty$ to a Brownian CRT in the weak sense induced by the Gromov-Prokhorov topology. This yields a multi-dimensional extension of a limit theorem due to Janson for the number of random cuts needed to isolate the root in Galton-Watson trees conditioned by their sizes.

Matthias Birkner

Ancestry in the face of competition

ABSTRACT: The genealogies in spatial models whose population sizes fluctuate stochastically because of local competition form relatively complicated random walks in a space-time dependent random environment. We consider the supercritical discrete-time contact process on \mathbb{Z}^d as the simplest non-trivial example of such a locally regulated population model (thinking of particles that compete for the resource "empty space") and study the dynamics of ancestral lineages sampled at stationarity, viz. directed random walk on a supercritical directed percolation cluster. We prove a law of large numbers and central limit theorem (both averaged and "quenched") via suitable regeneration constructions. Furthermore, we discuss approaches to extend these results to more general models that allow multiple occupancy of sites, and implications for the spatial distribution of (neutral) types.

Based on joint work with Jiří Černý, Andrej Depperschmidt and Nina Gantert.

Jochen Blath

An ancestral recombination graph for diploid populations with skewed offspring distribution

ABSTRACT: We consider a diploid biparental multilocus population model of Moran type, in which randomly chosen pairs of diploid individuals contribute offspring to the population. The number of offspring can be large, in particular relative to the total population size. Such 'heavily skewed' reproduction mechanisms have been considered by various authors recently, cf. e.g. Eldon and Wakeley (2008), and reviewed by Hedgecock and Pudovkin (2011). The chromosomes of each diploid offspring are derived from two distinct individuals, resulting in a separation of timescales phenomenon: ancestral lineages can only coalesce when in distinct individuals. We extend a result of Möhle (1998) to obtain convergence of the ancestral process to an ancestral

recombination graph. Due to diploidy and large offspring numbers, novel effects appear. For example, the marginal genealogy at each locus is given by a Ξ -coalescent necessarily involving simultaneous quadrifold multiple mergers, and different loci remain substantially correlated even as the recombination rate grows large. We compute correlations of coalescence times for two loci and discuss our findings for simulated data.

Joint work with Matthias Birkner (Mainz), Bjarki Eldon (Oxford).

Nicolas Champagnat

Adaptive dynamics in an individual-based, multi-resources chemostat model

ABSTRACT: We consider an evolutionary model of population with competition for resources through a chemostat-type model, where individuals consume several common resources that are constantly supplied. This model describes for example the adaptation of bacteria interacting with their environment composed of resources. Bacteria are characterized by a continuous traits describing their consumption ability for each resource. The population of bacteria is assumed to follow a discrete stochastic dynamics, and the dynamics of resources concentrations is governed by deterministic ODEs. We consider fast resource and birth and death dynamics and a slow mutation rate. We prove that the population behaves on the mutation time scale as a jump process describing successive fast mutant invasions between evolutionary equilibria. In the small mutation steps limit, this process converges to an ODE known as the canonical equation, and we are able to characterize the trait values where evolutionary branching — a form of diversification or speciation — may occur.

This is joint work with Pierre-Emmanuel Jabin and Sylvie Méléard.

Graham Coop

Moving towards a general model of the coalescent with linked selection

ABSTRACT: Two major sources of stochasticity in the dynamics of neutral alleles result from the sampling due to finite population size (genetic drift) and the random genetic background of selected alleles on which neutral alleles are found (linked selection). There is now good evidence that linked selection plays an important role in shaping polymorphism levels in a number of species. One of the best investigated models of linked selection is the recurrent full sweep model, in which newly arisen selected alleles fix rapidly. However, the bulk of selected alleles that sweep into the population may not be destined for rapid fixation in the species. Here we develop a coalescent model that generalizes the recurrent full sweep model to the case where selected alleles do not sweep to fixation. We show that in a large population, only the initial rapid increase of a selected allele affects the genealogy at partially linked sites, such that the subsequent fate of the selected allele often does not matter. I will discuss the possible extensions of these results and highlight promising directions in developing a more flexible framework to describe the effects of linked selection on genomic patterns of diversity.

Jean-François Delmas

Record on Lévy trees and paradox on advantageous mutations

ABSTRACT: We will present a model of continuous state branching process (CSBP) with advantageous mutation based on a record process on Lévy tree (this model is closely related to a model of CSBP with neutral mutations). The advantage of the mutation for a sub-population will be characterized by a parameter of the branching mechanism. For well chosen parameters, the whole population is still a CSBP but with different branching mechanism. Of course: to the more advantageous mutations corresponds stochastically larger sub-populations. However, conditionally on the mean of the parameters of the branching mechanisms over the sub-populations (that is the average of the advantageous mutation), we get the sub-populations have the same distribution.

Steven N. Evans

Go forth and multiply?

ABSTRACT: Organisms reproduce in environments that vary in both time and space. Even if an individual currently resides in a region that is typically quite favorable, it may be optimal for it to “not put all its eggs in the one basket” and disperse some of its offspring to locations that are usually less favorable because the effect of unexpectedly poor conditions in one location may be offset by fortuitously good ones in another.

I will describe joint work with Peter Ralph and Sebastian Schreiber (both at University of California, Davis) and Arnab Sen (Cambridge) that combines stochastic differential equations, random dynamical systems, and even a little elementary group representation theory to explore the effects of different dispersal strategies.

Warren Ewens

On the deterministic theory of population genetics and its possible stochastic extensions.

ABSTRACT: Deterministic population genetics theory is now largely motivated by the need to analyze data particularly from the human species, and has thus moved to a whole-genome, non-random-mating analysis. This talk discusses aspects of this theory and possible extensions of it to the corresponding stochastic theory.

Robert Griffiths

The Λ -Fleming-Viot process and a connection with Wright-Fisher diffusion

ABSTRACT: The d -dimensional Λ -Fleming-Viot generator acting on functions $g(x)$, with x being a vector of d allele frequencies, can be written as a Wright-Fisher generator acting on functions g with a modified random linear argument of x induced by partitioning occurring in the Λ -Fleming-Viot process. The eigenvalues and right eigenvectors are easy to see from the representation.

The moment dual process in the Fleming-Viot process is the usual Λ -coalescent back in time. In the Wright-Fisher representation using a different set of polynomials as test functions produces

a dual death process which has a similarity to the Kingman coalescent and decreases by units of one.

An application in the infinitely-many-alleles Λ -Fleming-Viot process is finding an interesting identity for the frequency spectrum of alleles that is based on size-biasing.

Joachim Hermisson

Evolutionary rescue in structured populations

ABSTRACT: We consider a population in a rapidly deteriorating structured environment. The population is doomed for extinction unless it is rescued through the origin and establishment of a beneficial “rescue mutation”. We use branching process theory to calculate the probability for evolutionary rescue and discuss the dependence of the rescue probability on various genetic and environmental factors. Several counterintuitive predictions result.

This is joint work with Hildegard Uecker.

Paul Joyce

Characterizing the Distribution of Lysis Time and Burst Size in Lytic Phage

ABSTRACT: A lytic phage (virus) invades its host, replicates within the host and lyses the cell. Understanding this process is fundamental to phage biology. We present a statistical model for analyzing data that tracks this process. Either ‘Missing data’ or ‘censored data’ challenge any statistical modeling effort. The data used for this model has both. We address the challenges in 3 steps: (1) estimating the number of phage in each well, (2) estimating the lysis time, and (3) estimate the burst size. A rigorous validation process using simulations offers credibility to the lysis time & size estimates for a particular virus and indicates the methods may be extended to other lytic phage.

Joint work with Craig Miller, Univ. of Idaho, and Daniel Weinreich, Brown Univ.

Götz Kersting

The total external branch length of evolving Kingman coalescent trees

ABSTRACT: n -coalescent trees are models for the genealogy of n individuals. For the Kingman coalescent they result from the underlying Moran dynamics, with time ranging from $-\infty$ to ∞ . In this set-up different coalescent trees arise at different moments, such that one ends up with a stationary tree-valued process, the evolving Kingman n -coalescent. Different aspects of this process have been investigated for $n \rightarrow \infty$. We study the evolution of the total external branch length, which is the sum of all branches ending in an individual. Properly scaled it converges to a stationary Gaussian process with a.s. continuous paths and the covariance function $1/(1+|t-s|)^2$.

Joint work with Iulia Stanciu.

Fima Klebaner

The Long Run Age Structure of Population-Dependent General Branching Processes in Environments with A High Carrying Capacity

ABSTRACT: The age structure of populations supercritical below and subcritical above a *carrying capacity* is investigated, the result being two laws of large numbers, one as time passes and the capacity increases, provided the starting population is not little, the other concerning populations starting small but viewed at an *evolutionary time scale* where the unit equals the carrying capacity, tending to infinity. The resulting asymptotic age-distributions are similar and deterministic but their for initial conditions, which in the second case are intrinsically random. Joint work with K. Hamza (Monash) and P. Jagers (Chalmers).

Steve Krone

Antibiotic resistance plasmids and spatial structure

ABSTRACT: Bacterial plasmids are circular extra-chromosomal genetic elements that code for simultaneous resistance to multiple antibiotics and are thought to be one of the most important factors in the alarmingly rapid loss of our arsenal of antimicrobial drugs. Plasmids propagate horizontally by infectious transfer, as well as vertically during cell division. Horizontal transfer requires contact between donor and recipient cells, and so spatial structure can play a key role in mediating the spread of antibiotic resistance genes. We will discuss ODE and stochastic spatial models of plasmid population dynamics, as well as empirical results. As an example of the effects of spatial structure, we will use the spatial model to evaluate the effectiveness of a commonly used estimate of plasmid transfer efficiency when applied to surface-associated populations. This example points out the importance of modeling spatial and temporal heterogeneities in growth and infection rates.

Tom Kurtz

Filtering and models in population biology

ABSTRACT: The simplest derivations of lockdown constructions for population models are based on filtering arguments. Some of the background of these methods will be discussed along with extensions to models in random environments and sampling at multiple time points.

Amaury Lambert

Coalescent point processes and phylogenies

ABSTRACT: A coalescent point process is a planar, ultrametric tree where the coalescence times between two consecutive tips are independent, identically distributed random variables. We consider any branching phylogenetic tree model where: 1) life lengths all have the same distribution, which is not necessarily exponential and may possibly depend on birth time; 2) the birth rate possibly also depends on the time variable; 3) species sampling can be incomplete. We show that under any such model, the reconstructed tree (or tree spanned by sampled tips, i.e., extant species) is always a coalescent point process. We characterize the common distribution of

coalescence times in the following three special cases: 1) the homogeneous case: lifetime distributions and birth rate do not depend on time; 2) the Markovian case: time-inhomogeneity of lifetime distributions is due to inhomogeneous, instantaneous death rates; 3) bottlenecks: the time inhomogeneity of lifetime distributions is due to independent thinning of lineages at fixed times. Finally, we show some interesting extensions/applications to : a) epidemiology inference from viral sequence data, b) the protracted speciation model, and c) phylogenetic trees with population dynamics.

Hans Metz

Conflict between alleles and modifiers in the evolution of genetic polymorphisms

ABSTRACT: The Canonical Equation of adaptive dynamics is a differential equation for the adaptive change of observable traits over evolutionary time, derivable through subsequent limits from individual-based trait-differentiated population models including rare mutations with small effect. In this talk I describe an example of how this CE can be used to arrive at a somewhat unexpected biological conclusion. In diploid Mendelian populations the average offspring number in the branching process by which a mutant allele invades into some resident population can be decomposed into a micro-gametic and a macro-gametic contribution (in humans through sperm or eggs). The standard population genetics assumption is that these two contributions are equal. However, realistically speaking this is generically never the case. I therefore do not assume equality. Focus on some protected polymorphism, i.e., a polymorphism such that each of the two alleles has average offspring number larger than 1 in the ecological and genetic environment produced by a population of homozygotes of the other allele.

We can then write down two different CEs for the evolutionary change of the polymorphism, one for the case where only the focal alleles evolve, the other for the case where only the rest of the genome evolves (usually referred to as modifier evolution). Not only are the CE's different, they also have different equilibria. Hence there is an intra-genomic conflict between the two sorts of players. Of course, for the description of reality both CE's have to be combined into one, with the contributions of the two mechanisms weighted with the respective relative mutation frequencies. As the rest of the genome will in general produce more relevant mutations than the focal alleles, we may expect that overall the modifiers will win. However, when the population has settled at a modifier dominated ESS, still once in a while a mutant in the focal alleles will produce a change in phenotypes that will quickly be undone again by the modifiers. The effect will be a continual turnover of the genome with no visible effect.

Although from an ecological perspective the inferred arms race should be generic, it does not occur in the standard simplified models from the literature: if one argues backwards to see under which conditions it will not occur, the main commonly used simplifications pop out. For the connection with the real world the most interesting case is when there are two separate sexes with their own genotype to phenotype maps (males can have different values of the trait than females). This leads to the prediction that hermaphroditic species (e.g. most seed plants, but also quite some animals) should have faster genome turnover than species with separate sexes.

Martin Moehle

On Compound Poisson Population Models

ABSTRACT: Compound Poisson population models are particular conditional branching process models. All these models belong to the richer class of exchangeable Cannings models. In the first part of the talk particular compound Poisson population models are analyzed, namely skewed Wright-Fisher models and skewed Dirichlet models. The models in the domain of attraction of the Kingman coalescent are characterized and it is shown that these models are never in the domain of attraction of any other continuous-time coalescent process. Results are obtained characterizing which of these models are in the domain of attraction of a discrete-time coalescent with simultaneous multiple mergers of ancestral lineages. The second part of the talk discusses more general compound Poisson models and focuses in particular on symmetric models. A criterion is provided ensuring convergence to the Kingman coalescent. A key analytic tool in the proof is the saddle point method. The talk finishes with a discussion of more complicated compound Poisson models not having a saddle point in the open interval $(0, r)$, where r denotes the radius of convergence of the power series defining the compound Poisson model.

Main parts of this talk are joint work with Thierry Huillet, Université de Cergy-Pontoise.

Pleuni Pennings

Quantifying the evolution of drug resistance in HIV and learning from deleterious mutations

ABSTRACT: I am interested in understanding the evolution of drug resistance in HIV. Resistance can evolve as a response to treatment in the virus every patient. One major question is whether resistance mutations tend to stem from standing genetic variation (mutations which are present in the viral population of the patient prior to treatment start) or new mutations. I have made some progress on this question by using epidemiological data. I find that drug resistance is more likely to evolve in the first year of treatment versus later years, and the increased risk of evolution in the first year can be directly linked to standing genetic variation.

Next, I planned to use population genetic data to estimate the probability that drug resistance mutations are present prior to treatment initiation. Results from such population genetic approach should be compatible with results from the epidemiological approach. It turned out that the population genetic approach is at the same time more & less feasible than expected. I will explain some of the practical and theoretical issues in this project. It got us (Sergey Kryazhimskiy, John Wakeley and me) interested in how deleterious mutations are affected by bottlenecks and selective sweeps. Deleterious mutations (such as non-synonymous mutations) may be more useful than expected for learning about sweeps and bottlenecks.

Peter Pfaffelhuber

Selective sweeps in structured populations

ABSTRACT: If a highly beneficial mutant arises in a natural population, it may happen that eventually all individuals are mutants. This scenario – termed a selective sweep – is well-studied for panmictic populations of constant size. Here, we will examine this model in a population which is distributed on two (or more) islands. Our main tool is the ancestral selection graph,

which is the extension of Kingman’s coalescent to models with selection. We state a convergence result for the duration of the selective sweep, depending on the migration rate between islands.

This is joint work with Cornelia Pokalyuk, Andreas Greven and Anton Wakolbinger.

Lea Popovic

Stochastically induced bistability in Density Dependent Population Processes on Multiple Scales

ABSTRACT: We study a stochastic two-species interacting population system where there are two sets of mechanisms involved. One set consists of interactions which governs the overall dynamics of species amounts in the system. The other set consists of resampling, branching or splitting mechanisms which yield unbiased perturbative changes to species amounts. Our results show that in a system with a large but bounded capacity only a combination of these two sets of interactions can lead to stochastically induced bistability. In fact, depending on the relative rates between the two sets of interactions, there are two ways in which bistability can occur with distinct signatures.

Joint work with John McSweeney.

Peter L. Ralph

Exploring recent relatedness – IBD and biparental ancestry

ABSTRACT: Long segments of genome shared (IBD) between individuals, which can be common in large enough samples, have the potential to tell us about population history extremely recently, an exciting new frontier for population genetics. I will discuss some theory relating distributions of IBD lengths to various coalescent time distributions while thinking explicitly about biparental ancestry, and point out difficulties and directions in which the way forward is not clear. I will illustrate with empirical patterns of relatedness coming from the past tens of generations in humans from across Europe.

Serik Sagitov

Interspecies correlation for Brownian traits

ABSTRACT: A simple way to model phenotypic evolution is to assume that after splitting, the trait values of the sister species diverge as independent Brownian motions. Relying only on a prior distribution for the underlying species tree (conditioned on the number n of extant species) we study the random vector (X_1, \dots, X_n) of the observed trait values. We derive compact formulae for the variance of the sample mean and the mean of the sample variance for the vector (X_1, \dots, X_n) . The key ingredient of these formulae is the correlation coefficient between two trait values randomly chosen from (X_1, \dots, X_n) . This interspecies correlation coefficient takes into account not only variation due to the random sampling of two species out of n available and the stochastic nature of Brownian motion but also the uncertainty in the phylogenetic tree. The latter is modeled by a (supercritical or critical) conditioned branching process. In the critical case we modify the Aldous-Popovic model by assuming a proper prior for the time of origin.

This is a joint work with Krzysztof Bartoszek.

Jay Taylor

Gene Flow and the Population Genetics of Human Infectivity in East African Sleeping Sickness

ABSTRACT: The kinetoplastid parasites that cause African sleeping sickness have been traditionally classified into three sub-species: *Trypanosoma brucei gambiense*, which causes a chronic form of the disease in western Africa; *T. b. rhodesiense*, which causes an acute form of the disease in eastern and southern Africa; and *T. b. brucei*, which infects cattle and wildlife, but which is usually not human infective. However, recent studies indicate that *T. b. rhodesiense* is paraphyletic relative to *T. b. brucei*, and suggest that human infectivity may be frequently conferred on the progeny of *T. b. brucei* x *T. b. rhodesiense* by the gain of a single sub-telomerically located gene known as SRA. In this talk, I will discuss some questions concerning the population genetics of the SRA gene and the effect that gene flow between human and non-human infective populations might have on the efforts to treat and eradicate the acute form of human sleeping sickness.

Chi Viet Tran

Limit theorems of historical processes in population dynamics

ABSTRACT: We consider the evolving genealogy of a birth and death process with trait structure and ecological interactions. Traits are hereditarily transmitted from a parent to its offspring unless a mutation occurs. The dynamics may depend on the trait of the ancestors and on its past and allows interactions between individuals through their lineages. We define an interacting historical particle process describing the genealogies of the living individuals. For large populations with small individuals with allometric demographics, this individual-based process can be approximated by a nonlinear historical superprocess. Our convergence theorem is illustrated by examples of current interest in Biology. This is a joint work with Sylvain Billiard, Régis Ferrière and Sylvie Méléard.

Vladimir Vatutin

Critical branching process with two types of particles evolving in asynchronous random environments

ABSTRACT: A two-type pure decomposable branching process in random environment is considered. Each particle of this process may produce offspring of her own type only. Let $\exp\{X_k(i)\}$ be the mean number of children produced by a particle of type $i = 1, 2$ of generation k . Assuming that $X_k(2) = -X_k(1)$ with probability 1 and that the random walk $S_n(1) = X_1(1) + \dots + X_n(1)$ specified by the random environment is oscillating we study the joint conditional distribution of the number of particles in the population at moments $nt, 0 < t \leq 1$ given that both processes survive up to moment $n \rightarrow \infty$. Under the same conditioning we find the asymptotic representation for the joint conditional distribution of the population sizes of both types at the moments when the environment is very unfavorable for the particles of the first type. It is shown that the process has unusual properties which may be treated as bottlenecks and periods of growth in a model of predator-pray coexistence.

Amandine Véber

On the usefulness of genealogical trees

ABSTRACT: In this talk, we shall review some recent results on the spatial Λ -Fleming-Viot process. This measure-valued process models the evolution of the genetic diversity of a population living in a spatial continuum. We shall see that the genealogical trees associated to the forwards-in-time evolution are of great help in understanding the asymptotic behaviour of the population with and without the influence of a selection pressure, as well as for setting up robust methods to infer the relevant parameters of local evolution.

These results are joint work with N. Barton, A. Etheridge, J. Kelleher and F. Yu.

John Wakeley

Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent

ABSTRACT: A conceptual flaw in the backward-time approach to population genetics called coalescent theory as it is applied to diploid bi-parental organisms will be discussed. Specifically, the way random models of reproduction are used in coalescent theory is not justified. Instead, the population pedigree for diploid organisms—that is, the set of all family relationships among members of the population—although unknown, should be treated as a fixed parameter, not as a random quantity. Gene genealogical models should describe the outcome of the percolation of genetic lineages through the population pedigree according to Mendelian inheritance. Simulated pedigrees, some of which are based on family data from 19th century Sweden, show that in many cases the (conceptually wrong) standard coalescent model is difficult to reject statistically, and in this sense may provide a surprisingly accurate description of gene genealogies on a fixed pedigree. The differences between the fixed-pedigree coalescent and the standard coalescent are illustrated by heuristic analysis and further simulation results. Differences are apparent in recent past generations, roughly $< \log_2(N)$ generations, but then disappear as genetics lineages are traced into the more distant past.

Anita Winter

Tree valued spatial Lambda-Cannings dynamics

ABSTRACT: We study the evolution of genealogies for interacting spatially structured Lambda-Cannings models which are also known as generalized Fleming-Viot processes. These are the limit processes of individual-based population models where individuals carry a type, and are replaced by descendants of possibly very sizable offspring. The spatial interaction is due to migration through geographic space. We show that the dual to these tree-valued spatial Lambda-Cannings dynamics are tree-valued spatial Lambda-coalescents, and conclude from here the convergence of the fixed time genealogies to the genealogy of an infinitely old population as time tends to infinity. Depending on the strength of migration the latter consists either of a single or of multiple families. We then study the populations on large tori in \mathbb{Z}^d with $d \geq 2$. Depending on the rescaling we find global features which are universal for all Lambda-Cannings dynamics and local features which heavily depend on the measure Lambda.

Joint work with with Andreas Greven and Anton Klimovsky