

Stochastic models on Biology

Organizer: Arno Siri-Jegousse (IIMAS-UNAM, MX)

For decades now, random models applied to biology have taken an important place in the worldwide probability community. Random models can apply to a wide range of life science thematics such as evolution, population genetics, ecology... On the other side, a large amount of probabilistic tools have been used to study such phenomena and biological applications have motivated the advanced study of such thematics as measure-valued processes, branching and coagulation processes, random partitions, stochastic differential equations... In Mexico, a contributed session with this thematic was organized by A. Wakolbinger during the SPA 2011. Since this time the presence of biological applications to probability has been permanent. The aim of this proposal is to keep on informing the Latin American community about the area, by inviting young and confirmed researchers from various regions of the globe, providing some diverse visions of the thematic. C

Inferring demographic histories using coalescent hidden Markov Models

Matthias Steinrücken (University of Chicago, US)

Inference of historical demographic events from contemporary genomic sequence data has received a lot of attention in recent years. A particular focus has been on the recent exponential growth of population size in humans. This recent growth strongly impacts the distribution of rare alleles, which are of importance when studying disease related genetic variation. The popular method PSMC (Li and Durbin, 2011) is used to infer population sizes from a sample of two chromosomes. However, the small sample size severely limits the power of this method in the recent past.

To improve inference in the recent past, we extend the Coalescent Hidden Markov model approach of PSMC to larger sample sizes. Since using the full genealogical trees relating the sample at each locus is computationally prohibitive, we introduce a flexible mathematical framework to employ different representations of these local trees. In particular, we present the implementation of this framework using the height of the local trees (TMRCA), corresponding to PSMC for sample size 2, and using the total branch length of the local trees.

We evaluate the different representations in simulation studies and applications to genomic variation data from diverse human populations. We discuss potential extension of the framework to infer divergence times and migration rates in structured populations, and employing the posterior distribution of the local trees to detect regions under selection.

The symmetric coalescent

Verónica Miró Piña (UNAM, MX)

In this work, we study a family of population models with demographic bottlenecks, i.e. in which, at exponentially distributed times, the population size is substantially reduced. The genealogy of this process is described by a new class of coalescents with simultaneous multiple collisions, that we call the symmetric coalescents.

Symmetric coalescents are characterized by a possibly infinite measure over the non negative integers. We call them symmetric because they are the unique family of exchangeable coalescents satisfying a

symmetry property on their coagulation rates: they are invariant under any transformation that consists in moving one element from one block to another without changing the total number of blocks.

We illustrate the diversity of behaviors of this family of processes by introducing and studying a one parameter subclass, the (β, S) -coalescents. We compute estimates for their total coalescence rate and their length.

Finally, we compare our predictions to genomic data from different species, to test whether demographic bottlenecks have occurred in the past.

This is joint work with Adrián González Casanova and Arno Siri-Jégousse

Genealogies in pushed waves

Sarah Pennington (University of Bath, MX)

Consider a population in which each individual carries two copies of each gene, and suppose that a particular gene occurs in two different types, a and A . Suppose that individuals carrying AA have a higher evolutionary fitness than aa individuals, and that aA individuals have a lower evolutionary fitness. We can model this situation using a stepping stone model on the integers, and show that (under certain conditions) as the number of individuals at each site goes to infinity, the genealogy of a sample of type A genes from the population (under a suitable time scaling) converges to a Kingman coalescent.

Joint work with Alison Etheridge.

The Wright-Fisher model with efficiency

Adrián González-Casanova (UNAM, MX)

In populations competing for resources, it is natural to ask whether consuming more or less resources provides any selective advantage. To answer this question, we propose a Wright-Fisher model with two types of individuals: the inefficient individuals, those who need more resources to reproduce and can have a higher growth rate, and the efficient individuals. In this model, the total amount of resources N , is fixed, and the population size varies randomly depending on the number of efficient individuals. We show that, as N increases, the frequency process of efficient individuals converges to a diffusion which is a generalisation of the Wright-Fisher diffusion with selection. The genealogy of this model is given by a branching-coalescing process that we call the Ancestral Selection/Efficiency Graph, and that is an extension of the Ancestral Selection Graph ([Krone and Neuhauser (1997a), Krone and Neuhauser (1997b)]). It turns out that, in evolving populations, inefficiency can arise as a promoter of selective advantage and not necessarily as a trade-off.

This talk is based in joint work with Veronica Miró Pina and Juan Carlos Pardo