XI Brazilian School of Probability

Maresias, SP, Brazil - August 5–11, 2007

Talks

Substitution operators

Andrei Toom
UFPE

Abstract: We study properties of substitution operators. A substitution operator acts on translation invariant measures on $A^{\mathbb{Z}}$, where A is any finite set called alphabet. Elements of A are called letters. Any finite sequence of letters is called a word. Given two words U and V (with a certain restriction on U, namely U being self-avoiding) and a number r in $[0,1]$, the substitution operator substitutes every entrance of U in the configuration by V with a rate r. A rigorous definition is based on approximation of measures by words. We prove that all substitution operators are continuous provided $r < 1$. We also prove that all substitution operators are convex. This allows to prove existence of some invariant measures.

A simple maximization model inspired by algorithms for the organization of genetic candidates in bacterial DNA

Andrew Hart
Universidad de Chile

Abstract: We propose a simple model for interaction between gene candidates in the two strands of bacterial DNA (deoxyribonucleic acid). Our model assumes that 'final' genes appear in one of the two strands, that they do not overlap (in bacteria there is only a small percentage of overlap), and that the final genes maximize the occupancy rate, which is defined to be the proportion of the genome occupied by coding zones. We are more concerned with describing the organization and distribution of genes in bacterial

DNA than with the very hard problem of identifying genes. To this end, an algorithm for selecting the final genes according to the previously outlined maximization criterion is proposed. We study the graphical and probabilistic properties of the model resulting from applying the maximization procedure to a Markovian representation of the genic and intergenic zones within the DNA strands, develop theoretical bounds on the occupancy rate (which, in our view, is a rather intractable quantity), and use the model to compute quantities of relevance to the Escherichia coli genome and compare these to annotation data. Although this work focuses on genomic modelling, we point out that the proposed model is not restricted to applications in this setting. It also serves to model other resource allocation problems.

Stochastic chains with variable length memory

Antonio Galves
IME-USP

Abstract: Stochastic chains with variable length memory define an interesting family of stochastic chains of infinite order on a finite alphabet. The idea is that for each past, only a finite suffix of the past, called *context*, is enough to predict the next symbol. The set of contexts can be represented by a rooted tree with finite labeled branches. The law of the chain is characterized by its tree of contexts and by an associated family of transition probabilities indexed by the tree.

These models were first introduced in the information theory literature by Rissanen (1983) under the name of *finite memory source* or *tree machines*. More recently they became quite popular in the statistics literature under the name of *Variable Length Markov Chains* coined by Bühlmann and Wyner 1999.

Rissanen (1983) introduced these models as an universal tool to perform data compression. Recently, they have been used to model up scientific data in areas as different as biology, linguistics and music.

My talk aims to introduce some of the basic ideas and results in the field, and to convince the audience that the class of stochastic chains with unbounded variable length memory constitutes an interesting object to be considered by probabilists.

The truncation's question and percolation of words

Bernardo Nunes Borges de Lima
UFMG

Abstract: In a site percolation model on $\mathbb{Z}^2$ equiped with all horizontal and vertical conections; given any binary sequence (word), it is possible to show that there exists some large integer N, such that is still possible to see this word, almost surely, if all conections whose length is bigger than N are supressed.

Growing Directed Networks: Limit in-degree distribution for arbitrary out-degree one

Daniel Fraiman
Universidad de San Andrés

Abstract: We compute the limiting in-degree distribution for a growing network process with directed edges and arbitrary out-degree distribution. In particular, under preferential linking, we find that if a process is such that each new node has a light tail (finite variance) out-degree distribution, then the corresponding in-degree one behaves as k^{-3}. Moreover, for an out-degree distribution with a scale invariant tail, $P_{out}(k) \sim k^{-\alpha}$, the corresponding in-degree distribution has exactly the same asymptotic behavior only if $\alpha < 3$ (infinite variance). Similar results are obtained when attractiveness is included. Finally, we present an application to the scientific publications network from ISI dataset. The results presented here explain the in/out-degree distribution observed in many real networks.

Density-profile processes

Eduardo Neves
IME-USP

Abstract: We introduce jump processes in $\mathbb{R}^k$, called *density-profile process*, to model biological signaling networks. They describe the macroscopic evolution of finite-size spin-flip models with k types of spins interacting through a non-reversible Glauber dynamics. We focus on the the k-dimensional empirical-magnetization vector in the thermodynamic limit, and

prove that, within arbitrary finite time-intervals, its path converges almost surely to a deterministic trajectory determined by a first-order (non-linear) differential equation. As parameters of the spin-flip dynamics change, the associated dynamical system may go through bifurcations, associated to *phase transitions* in the statistical mechanical setting.

Hydrodynamics for non-conservative interacting particle systems

Glauco Valle
UFRJ

Abstract: We will discuss some non-conservative interacting particle systems whose hydrodynamic behavior is known. These models are hard to analyse since usual methods to prove the hydrodynamic limit cannot be applied directly and each case should be considered separatedly with its own particularities. The problem is also connected to other fenomena of interest as the study of the microscopic evolution of the boundary in physical systems with more than one state.

Two-component symmetric exclusion process with open boundaries

Gunter Schuetz
Forschungszentrum Jülich

Abstract: We consider a simple symmetric exclusion process with two species A, B of particles with different hopping rates D_A, D_B. We study open boundaries with different reservoir densities which drive the system into a non-equilibrium steady state. For investigating the dynamics we derive the hydrodynamic limit. This yields a system of coupled non-linear diffusion equations for the coarse-grained particle densities. We obtain exact results for the length-dependent self-diffusion coefficients and for the stationary solution. In the steady state we find a boundary-induced first-order phase transition as the total exterior density gradient between the system boundaries is varied. At one boundary a boundary layer develops inside which the current flows against the local density gradient. Generically the width of

the boundary layer and the bulk density profiles do not depend on the two hopping rates. At the phase transition line, however, the individual density profiles depend strongly on the ratio D_A/D_B.

Local Thermodynamic Equilibrium for some stochastic models of Hamiltonian origin

Krishnamurthi Ravishankar
SUNY-New Paltz

Abstract: We consider a class of 1-D stochastic models that are realizations of Hamiltonian models of heat conduction and prove that in the infinite volume limit local thermodynamic equilibrium is attained with linear energy profile.

Walking on Delaunay triangulations

Leandro Pimentel
IME-USP

Abstract: In this talk we will study some geometrical and combinatorial aspects of random Delaunay triangulations of the plane. This is a joint work with Raphael Rossignol (Université de Neuchâtel CH).

The Continuous Broken Lines Process

Leonardo Rolla
IMPA

Abstract: We introduce a spatial process of continuous broken lines and derive some of its properties. The broken lines generalize the Young diagram and are useful for computing last passage percolation values and finding maximal oriented paths. For a class of passage time distributions we can deduce a law of large numbers from the reversibility of the Markov chain associated to the process. One application is a simple proof of the explicit law of large numbers found by K. Johansson for the case of geometric and

exponential distributions. Joint work with V. Sidoravicius, D. Surgailis and M. E. Vares.

Minima in Branching Random Walks

Louigi Addario-Berry
University of Oxford

Abstract: We determine the expected value of the minimum M_n of a supercritical branching random walk after n steps, conditioned on survival, to within $O(1)$, assuming a natural condition on the exponential moment generating function of the step size. Under the same assumption, we derive bounds on the probability that M_n is very far from its expected value, showing in particular that all higher centralized moments of M_n are finite. This is joint work with Bruce Reed.

Billiards in a general domain with random reflections

Marina Vachkovskaia
UNICAMP

Abstract: We study stochastic billiards on general tables: a particle moves according to its constant velocity inside some domain $D \subset I\!R^d$ until it hits the boundary and bounces randomly inside according to some reflection law. We assume that the boundary of the domain is locally Lipschitz and almost everywhere continuously differentiable. The angle of the outgoing velocity with the inner normal vector has a specified, absolutely continuous density. We construct the discrete time and the continuous time processes recording the sequence of hitting points on the boundary and the pair location/velocity. We mainly focus on the case of bounded domains. Then, we prove exponential ergodicity of these two Markov processes, we study their invariant distribution and their normal (Gaussian) fluctuations. Of particular interest is the case of the cosine reflection law: the stationary distributions for the two processes are uniform in this case, the discrete time chain is reversible though the continuous time process is quasi-reversible. Also in this case, we give a natural construction of a chord "picked at random" in D, and we study

the angle of intersection of the process with a $(d-1)$-dimensional manifold contained in D.

A transfer matrix related to exactly integrable stochastic process of particles of arbitrary size

Matheus Jatkoske Lazo
UFSM/Unipampa

Abstract: We introduce a special family of exact integrable interacting vertex models that generalizes the well known six-vertex model. In addition to the usual nearest-neighbor interactions among the vertices, there exist extra hard-core interactions among pair of vertices at larger distances. The associated row-to-row transfer matrices are diagonalized by using the recently introduced matrix product ansatz. Similarly as the relation of the six-vertex model with the XXZ quantum chain, the row-to-row transfer matrices of these knew models are also the generating functions of an infinite set of commuting conserved charges. Among these charges we identify the integrable generalization of the XXZ chain that contains hard-core exclusion interactions among the spins. These quantum chains already appeared in the literature and are related to exactly integrable stochastic process of particles of arbitrary size. In the present work we explains their exact integrability.

Insensitive queuing networks

Matthieu Jonckheere
CWI

Abstract: A large variety of communication systems, including telephone and data networks, can be represented by so-called Whittle networks. The stationary distribution of these networks is insensitive, depending on the service requirements at each node through their mean only. These models are of considerable practical interest as derived engineering rules are robust to the evolution of traffic characteristics. We identify the class of load balancing and bandwidth allocations policies which preserve insensitivity and characterize optimal strategies in some specific cases.

Connected allocation to Poisson process in $\mathbb{R}^2$

Maxim Krikun
Institut Elie Cartan, Univ. Nancy 1

Abstract: We construct a scheme allowing to determinstically allocate in a translation-invariant way cells of unit area to the points of a Poisson point process in $\mathbb{R}^2$, thus answering a question by Hoffman-Holroyd-Peres in [math.PR/0505668]. The proposed solution makes use of the Riemann map from the plane minus the minimal spanning forest of the Poisson point process to the halfplane.

Fluctuations of the equilibrium profile of condensating zero range processes

Michalis Loulakis
University of Crete

Abstract: For certain choices of the jump rate function the equilibrium measures of a zero range process exhibit a phase-transition in density. There is a critical particle density below which the equilibrium particle density has a uniform macroscopic profile, but above which particles form a bulk at the critical density and the remaining ones condensate on a single randomly chosen site. We study the thermodynamic limit of the fluctuations around this profile. Joint work with Ines Armendariz.

Random walk in random environment with asymptotically zero perturbation

Mikhail Menshikov
University of Durham, UK

Abstract: We give criteria for ergodicity, transience and null recurrence for the random walk in random environment on $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$, with reflection at the origin, where the random environment is subject to a vanishing perturbation. Our results complement existing criteria for random walks in random environments and for Markov chains with asymptotically zero drift,

and are significantly different to these previously studied cases. Our method is based on a martingale technique — the method of Lyapunov functions.

New developments for estimating probabilistic context trees.

Nancy Garcia
UNICAMP

Abstract: Probabilistic context trees define an interesting family of stochastic chains of finite or infinite order on a finite alphabet. The idea is that for each past, only a finite suffix of the past, called context, is enough to predict the next symbol. These suffixes can be represented by a countable tree of finite contexts. To completely define the model we associate to each context a transition probability.

The CLT for Shannon-McMillan-Breiman

Nicolai Haydn
University of Southern California

Abstract: The theorem of Shannon-McMillan-Breiman states that for every generating finite partition on an ergodic system, the exponential decay rate of the measure of cylinder sets equals the metric entropy almost everywhere. We prove that the measure of cylinder sets are lognormally distributed for strongly mixing systems and show that the rate of convergence is polynomially. This then in particular implies the LIL and the weak invariance principle for the information function.

Slow to start traffic model and $M/M/1$ queue

Pablo Augusto Ferrari
IME-USP

Abstract: We consider a system of ordered cars moving in $\mathbb{R}$ from right to left. Each car is represented by a point in $\mathbb{R}$; two or more cars can occuppy

the same point and cannot overpass. Cars have two possible velocities: either 0 or -1. A car needs an exponential time of mean 1 to pass from speed 0 to speed -1 (emphslow-to-start). Car n, say, travels at speed -1 until it (possibly) hits the stopped car $n-1$ to its left. After the departure of car $n-1$, car n waits an exponentially distributed random time to change its speed to -1, travels at this speed until it hits again stopped car $n-1$ and so on. We consider that initially cars are distributed in $\mathbb{R}$ according to a Poisson process of parameter $\lambda < 1$ and show that every car will be stopped only a finite number of times and that the final relative positions of the cars is again a Poisson process with parameter λ. To do that, we map the trajectories of the cars with a $M/M/1$ stationary queue as follows. Space in the traffic model is time for the queue. The initial positions of the cars correspond to the arrival process of the queue and the final relative positions of the cars correspond to the departure process of the queue. Joint work with E. Pechersky and F. Castellares

Diffusive-ballistic transition for a class of $2D$ self-repelling random walks

Remy Sanchis
UFMG

Abstract: We exibit a model of self-repelling $2D$ random walks that have diffusive behaviour for high temperatures and ballistic behaviour for low temperatures.

A Topological Correction for a Graph Based Spatial Scan Cluster Detection Algorithm

Sabino Ferreira
UFMG

Abstract: Many spatial cluster finder algorithms do not have adequate procedures for controlling the shapes of the clusters found. The cluster candidates may sometimes spread through large portions of the map, making it difficult for the practitioner to assess the geographical meaning of the solution. We propose a novel scan statistic algorithm for finding irregularly

shaped spatial clusters in a map divided in a finite number of regions, whose adjacency is defined by a graph structure, by means of a new penalty function, the cohesion penalty correction. Based on the graph topology, the cohesion correction was developed to avoid the excessive irregularity of the clusters. The cohesion correction is compared with the geometric concept of compactness correction, which was used previously as a penalty function. We show that the cohesion correction has advantages over the compactness correction, boosting the power to detect elongated clusters, and being less computer-intensive. A multi-objective genetic algorithm is used to compute the solutions, consisting of the Pareto-set of clusters candidates. The goal of the cluster finder algorithm is to maximize two objectives: the scan statistic and the cohesion of the graph structure. The statistical significances of the clusters in the Pareto-set are estimated through Monte Carlo simulations, using Gumbel's approximation for the scan statistic distribution, which are used as a criterion for choosing the best solution.

A criterion for transience of multidimensional branching random walks in random environment

Sebastian Müller
University Münster

Abstract: We develop a criterion for transience for a general model of branching Markov chains. In the case of multi-dimensional branching random walk in random environment this criterion becomes explicit.

Gene Expression Network Analysis

Serban Nacu
Ecole Normale Superieure

Abstract: There is a strong genetic component in cancer and other diseases. Often the disease involves dysfunction in the cellular machinery: genes in normal and disease cells behave in different ways. Those differences can be measured using microarray technology, which in the recent past has become a fundamental technique in biology and medicine.

A typical microarray measures expression levels for 20,000 genes at the same time: this massive parallel power also raises important statistical and computational problems. Standard microarray analysis treats each gene separately, and ranks them according to some measure of differential expression. But in reality genes interact, they act in concert rather than alone; an analysis that accounts for this has the potential to be more statistically accurate and biologically meaningful.

We introduce a method called GXNA (Gene eXpression Network Analysis). GXNA uses a gene interaction graph to search for clusters of related genes that are differentially expressed. It has several desirable features, such as fast runtimes and the computation of objective, permutation-based significance levels, and it shows promising results when applied to data sets involving cancer and the human immune system.

This is joint work with Rebecca Critchley-Thorne, Peter Lee, and Susan Holmes.

On a queueing network with a single server and varying regimes

Serguei Popov
IME-USP

Abstract: We consider a queueing system with one moving server and several stations having queues of infinite capacity. The server serves all customers at a station (including those that might have come after the server began its work there), and then jumps instantly to the next station. The served customers may leave the system, or be transferred to another station, with given probabilities (similarly to what happens, e.g., in Jackson networks). When the server comes to a station, it chooses the service rate and the transition parameters at random; those remain valid during the whole stay of the server in that station. For this model, we give criteria for recurrence, transience, and the existence of sth moment of the return time.

Condensation in zero-range processes

Stefan Grosskinsky
University of Cambridge

Abstract: The zero-range process is an interacting particle system with unbounded local state space and conserved number of particles. It is simple enough to have stationary product measures, but under certain conditions on the jump rates it exhibits a condensation transition. We show that in the context of the equivalence of ensembles this condensation can be characterized as a large deviation with respect to the maximal grand-canonical measure. Since this has a sub-exponential tail, the entropy of the condensate vanishes in the hydrodynamic limit. In the generalized case where the jump rates may also depend on the system size, the condensate can contribute to the entropy and the transition can be discontinuous.

Condensation for a fixed number of independent random variables

Valentin Sisko
UFF

Abstract: A family of m independent identically distributed random variables indexed by a chemical potential $\varphi \in [0, \gamma]$ represents piles of particles. As φ increases to γ, the mean number of particles per site converges to a maximal density ρ_c.