

Thematic Sessions

1 - Stochastic processes and their applications

29 Nov
15:30

Chairs: Alex Ramos and Pablo Rodriguez - Universidade Federal de Pernambuco
Auditorium Prof. Ruy Luis Gomes - DE-UFPE - Room 347

1.1 - Two models for bid-ask spread dynamics: asymptotic analysis of highly competitive and liquid fluctuations regimes

Anatoly Yambartsev - Universidade de São Paulo

We introduce two stochastic models to describe the dynamics of a limit order book with various types of liquidity. These models can be categorized as population models with uniform and geometrical catastrophes. We establish the validity of fundamental statistical principles, such as the law of large numbers, central limit theorem, and large deviations, for our catastrophe-based models. Additionally, we utilize our models to demonstrate how significant deviations in the spread and prices can occur. The work is done in collaboration with J. Hernandez, A. Logachov, and H. Rojas. The talk is based on [1], [2] and [3]. References: [1] H. Rojas, A. Logachov & A. Yambartsev, Order book dynamics with liquidity fluctuations: limit theorems and large deviations. arXiv preprint arXiv:2004.10632, 2020. [2] J. C. Hernandez, A. Logachov & A. Yambartsev, Bid-ask spread dynamics: large upward jump with geometric catastrophes, In preparation. [3] A. Logachov, O. Logachova & A. Yambartsev, The local principle of large deviations for compound Poisson process with catastrophes, Brazilian Journal of Probability and Statistics, v. 35, p. 205-223, 2021.

1.2 - Quasi-stationary distributions, extinction and resilience in ecological models

Jean-Rene Chazottes - CNRS & Ecole Polytechnique

We consider birth-and-death processes going eventually to extinction with probability one, accounting for the finiteness of resources. A classical result says that if we rescale such a process by a parameter K and fix a finite horizon time, then the resulting sample paths are very close to the solutions of a differentiable equation (deterministic dynamical system) with high probability, when K goes to infinity. The corresponding vector field is given by the difference between the birth and death rates. In particular, the limit when times goes to infinity does not commute with the limit when K goes to infinity. In this talk, I will try to explain what can be said for finite times and for finite K . A central object is the so-called quasi-stationary

distribution which plays the role of a stationary distribution when conditioning upon non-extinction. Another one is the mean-time to extinction. Finally, I will touch upon the following question. Observing a realization of the process, can we determine the so-called engineering resilience? This quantity can be defined as the reciprocal of the characteristic return time to, say, a fixed point after a (small) perturbation. To answer this question, we establish two relations which intermingle the resilience, which is a macroscopic quantity defined for the dynamical system, and the fluctuations of the process, which are microscopic quantities.

1.3 - The frog model on complete graphs

Gustavo Oshiro de Carvalho - Universidade de São Paulo

We consider a system of interacting random walks known as the frog model on complete graphs. Initially, an active particle is placed at one vertex of the n -complete graph \mathcal{K}_n , as well as one inactive particle at every other vertex. At each instant of time, each active particle may die with probability $1 - p$. Every active particle performs a simple symmetric random walk on \mathcal{K}_n until the moment it dies, activating all inactive particles it hits along its path. In this talk, we study some asymptotic properties, as n approaches infinity, of the number of visited vertices until all active particles have died.

1.4 - Evaluating dispersion strategies in growth models subject to geometric catastrophes

Fábio Prates Machado - Universidade de São Paulo

We consider stochastic growth models to represent population dynamics subject to geometric catastrophes. We analyze different dispersion schemes after catastrophes, to study how these schemes impact the population viability and comparing them with the scheme where there is no dispersion. In the schemes with dispersion, we consider that each colony, after the catastrophe event, has d new positions to place its survivors. We find out that when $d = 2$ no type of dispersion considered improves the chance of survival, at best it matches the scheme where there is no dispersion. When $d = 3$, based on the survival probability, we conclude that dispersion may be an advantage or not, depending on its type, the rate of colony growth and the probability that an individual will survive when exposed to a catastrophe.

1.5 - Sparse Markov Models for High-dimensional Inference

Daniel Yasumasa Takahashi - Universidade Federal do Rio Grande do Norte

Finite-order Markov models are well-studied models for dependent finite alphabet data. Despite their generality, application in empirical work is rare when the order d is large relative to the sample size n (e.g., $d = O(n)$). Practitioners rarely use higher-order Markov models because (1) the number of parameters grows exponentially with the order, (2) the sample size n required to estimate each parameter grows exponentially with the order, and (3) the interpretation is often difficult. Here, we consider a subclass of Markov models called Mixture of Transition Distribution (MTD) models, proving that when the set of relevant lags is sparse (i.e.,

$O(\log(n))$), we can consistently and efficiently recover the lags and estimate the transition probabilities of high-dimensional ($d = O(n)$) MTD models. Moreover, the estimated model allows straightforward interpretation. The key innovation is a recursive procedure for a priori selection of the relevant lags of the model. We prove a new structural result for the MTD and an improved martingale concentration inequality to prove our results. Using simulations, we show that our method performs well compared to other relevant methods. We also illustrate the usefulness of our method on weather data where the proposed method correctly recovers the long-range dependence.

1.6 - Community detection with maximum and integrated conditional likelihood for the Stochastic Block Model

Florencia Leonardi - Universidade de São Paulo

In this talk I will present the fundamental results in the literature on community detection for the Stochastic Block Model (SBM) with k communities, not necessarily symmetrical (all communities of the same size), focusing mainly on the methods of maximum and integrated conditional likelihood. This approach was previously considered by Chen and Bickel (2009). However, the proof of the consistency of the maximum likelihood estimator presents some points that are still open and not fully justified, as pointed out by Van der Pas and van der Vaart (2018). In this work, we show, using different concentration inequalities, that the maximum likelihood estimator is consistent above and at the phase transition threshold for networks with a logarithmic degree regime, completing the proof of Chen and Bickel (2009) and generalizing their results. We also extend these results to the integrated conditional likelihood estimator. This is joint work with Andressa Cerqueira (UFSCAR)

2 - Mathematical Modelling of Socioeconomic Events

Chair: Thyago C. C. Nepomuceno - Universidade Federal de Pernambuco
Auditorium Professor Ricardo de Carvalho Ferreira of CCEN-UFPE

29 Nov
15:30

2.1 - Patents, investments in R&D, and market value: an analysis by company size

Fernando H. Taques – Universidad Autónoma de Madrid

This research aims to understand whether organizational innovation can explain the market value of companies based on indicators of patents published, patents cited, and R&D expenditure. This empirical study performed a sample of 1,342 companies from 1985 to 2016 with fixed-effects and random-effects panel data models and quantile regression for panel data. There was partial support for the evidence that innovation positively affects companies' market value. This result could differ from the indicator applied in the estimates, methodology used and the company's size. Regardless of the empirical method, the total variables assets and gross revenue were able to explain the variations in the market value of organizations in most models. However, the coefficients found are different between the sample quantiles, suggesting that the company's size may be relevant.

2.2 - Perspectives for an AI-Based Decision Support Framework in Public Administration

Victor D. Heuer de Carvalho – Universidade Federal de Pernambuco

The integration of artificial intelligence is on the rise within public administration, calling for sophisticated tools to conduct a variety of analyses. These analyses span the utilization of textual content extracted from internal and external documents, such as web-scraped data, information from structured internal sources like relational databases, and multimedia-originated data, including videos, images, and sound recordings. This panel aims to introduce a framework that revolves around the comprehensive analysis of both structured and unstructured (textual) data, with a concurrent focus on incorporating geographic information.

2.3- Applying spatial decision support for maternal mortality analysis in a Brazilian state

Thiago Poletto – Universidade Federal do Pará

The interface between Decision Support Systems (DSS) and Geographical Information Systems (GIS) generates Spatial Decision Support Systems (SDSS) which aid in the decision-making process, particularly in situations where spatial attributes play a pivotal role in achieving accurate conclusions. By integrating conventional decision criteria with spatial criteria and visualization through the adoption of SDSS and geographic information systems technologies, a comprehensive analysis is ensured. The primary aim of this research is to exemplify an application of SDSS to support analyses and decisions on a specific issue in Public Health. The focus lies on assessing cities within the state of Pernambuco (Brazil) that exhibit the highest rates of maternal mortality. This approach promises to offer valuable insights and aid in making informed decisions to address critical maternal health issues in the region.

2.4 - Hierarchical Game: An application of a three-person game in the pricing of product warranty

Henrique Pinto dos Santos Zaidan - Universidade Federal Rural de Pernambuco

This paper is a hierarchical game defined to a manufacturer, an agent, and a consumer applied to the product warranty. Through a mix of non-cooperative and cooperative solutions, equilibrium strategies are developed. The model brings a coalition between the manufacturer and the agent. The manufacturer defines the sale price of the product (including the base warranty costs), whereas the agent prices maintenance services. Finally, the Shapley value redistributes the total gain between the participants, assigning the equilibrium prices.

3 - Statistics applied to biologically and socially motivated problems

30 Nov
15:30

Chair: Viviane Oliveira - Uninversidade Federal Rural de Pernambuco
Auditorium Prof. Ruy Luis Gomes - DE-UFPE - Room 347

3.1 - Citation counts: the world of sleeping beauties and shooting stars

José Fernando Fontanari - Universidade de São Paulo

Citation-based measures are widely used as quantitative proxies for subjective factors such as the importance of a paper or even the worth of individual researchers. Here we analyze the citation histories of papers published in journals of the American Physical Society between 1960 and 1968 and argue that state-of-the-art models of citation dynamics and algorithms for forecasting nonstationary time series are very likely to fail to predict the long-term (50 years after publication) citation counts of highly-cited papers using citation data collected in a short period (say, 10 years) after publication. This is so because those papers do not exhibit distinctive short-term citation patterns, although their long-term citation patterns clearly set them apart from the other papers. We conclude that even if one accepts that citation counts are proxies for the quality of papers, they are not useful evaluative tools since the short-term counts are not informative about the long-term counts in the case of highly-cited papers.

3.2 - Phenotypic evolution as an Ornstein-Uhlenbeck process: the effect of environmental variation and phenotypic plasticity

Paulo Roberto de Araujo Campos - Universidade Federal de Pernambuco

Understanding how a population of living organisms evolves in a gradually changing environment is a critical question in evolution, especially important in the current context of climate change. At the phenotypic level, important insights have been inferred from two analytical and tractable models: the Brownian motion (BM) and Ornstein-Uhlenbeck (OU). The OU process is a modification of the random walk, in which the particle is subject to a harmonic-like potential, tending to drift toward the process's mean. The Brownian motion was the first proposed trait evolution model, essentially viewed as a neutral model. The OU process was introduced as an alternative to the BM to incorporate stabilizing selection, i.e., whereby a trait is attracted to a selection optimum.

We modified the standard (OU) model to address environmental variation by taking a moving phenotypic optimum and endowing organisms with phenotypic plasticity. These two processes lead to an effective fitness landscape that differs from the original.

We observe that the environmental variation and phenotypic plasticity result in skewed phenotypic distributions. The skewness of the resulting phenotypic distributions strongly depends on the rate of environmental variation and strength of selection. When generalized to more than one trait, the phenotypic distributions are affected by the magnitude of the rate of environmental variation and its direction. A remarkable feature of our predictions is the existence of an upper bound for the critical rate of environmental variation to allow population persistence, even if there is no cost associated with phenotypic plasticity.

3.3 - Multifractal polarization time series of fish groups

Nathan Lima Pessoa - Universidade Federal Rural de Pernambuco

The dynamical behavior of groups of zebrafish (*Danio rerio*) in a circular tank was captured. The polarization of the group was measured as a function of time. We employed the Multifractal Detrended Fluctuation Analysis (MF-DFA) to the polarization time series. We observed that they exhibit multifractal behavior and that it becomes weaker as the number of fish N increases, which leads us to the conclusion that the presence of multiple different long-range correlations for small and large fluctuations is a characteristic of groups of few fish and as the number of fish increases, the corresponding series tend to be monofractal. We confirm this observation by showing that the distributions of increments of the polarization time series are well-described by q -Gaussian probability density functions, with the parameter q approaching the value of 1 (Gaussian behavior) for many fish systems. These results bring a new perspective to the comprehension of the dynamics of fish groups.

3.4 - Impacts of collective behavior on socioeconomic dynamics

Mateus Francisco Batista Granha - Universidade Federal de Pernambuco

Fundamental mechanisms of human group behavior shape diverse aspects of socioeconomic relations, organization, and dynamics. Recent global events such as the financial crisis, the COVID-19 pandemic recession, and the post-Brexit uncertainty in European markets highlight the profound impact of these systems on modern culture. Using an agent-based model within a random network framework, we elucidate the dynamics of asset prices in stock markets. Our model assumes that two categories of investors largely influence financial markets — noise traders and fundamentalists — representing prominent trading strategies. Individuals adopt one of two opinion states related to buying or selling an asset, influenced by a socioeconomic anxiety parameter, denoted as q . Our model reproduces real-world market features for a finite level of social anxiety and for a small fraction of interacting fundamentalist agents.

30 Nov
15:30

4 - Recent advances in statistical modeling

Chair: Aldo M. Garay - Universidade Federal de Pernambuco
Auditorium Professor Ricardo de Carvalho Ferreira of CCEN-UFPE

4.1 - Partially linear models with p-order autoregressive skew-normal errors

Clécio da Silva Ferreira - Universidade Federal de Juiz de Fora

This paper proposes partially linear models with random errors following p -order autoregressive (AR) with skew-normal errors. The maximum likelihood estimators are derived from the Expectation-Maximization algorithm, which have analytic expressions for the M and E-steps. The estimation of the effective degrees of freedom

concerning the nonparametric component are obtained based on a linear smoother. The conditional quantile residuals are used for the construction of simulated confidence bands to assess departures from the error assumptions, as well as autocorrelation and partial autocorrelation graphs are considered to check adequacy of the AR error structure. A simulation study is carried out to evaluate the efficiency of the EM algorithm. Finally, the methodology is illustrated by a real data set on cardiovascular mortality.

4.2 - Approximate Bayesian Estimation of Stochastic Volatility in Mean Models Using Hidden Markov Models: Empirical Evidence from Emerging and Developed Markets

Carlos Antonio Abanto Valle - Universidade Federal do Rio de Janeiro

The stochastic volatility in mean (SVM) model proposed by Koopman and Uspensky (2002) is revisited. This paper has two goals. The first is to offer a methodology that requires less computational time in simulations and estimates compared with others proposed in the literature as in Abanto-Valle et al. (2021) and others. To achieve the first goal, we propose to approximate the likelihood function of the model applying Hidden Markov Models (HMM) machinery to make possible Bayesian inference in real-time. We sample from the posterior distribution of parameters with a multivariate Normal distribution with mean and variance given by the posterior mode and the inverse of the Hessian matrix evaluated at this posterior mode using importance sampling. Further, the frequentist properties of estimators are analyzed conducting a simulation study. The second goal is to provide empirical evidence estimating the SVM model using daily data for five Latin American stock markets, USA, England, Japan and China. The results indicate that volatility negatively impacts returns, suggesting that the volatility feedback effect is stronger than the effect related to the expected volatility. This result is similar to the findings of Koopman and Uspensky (2002), where the respective coefficient is negative but non statistically significant. However, in our case, all countries (except Peru and China) present negative and statistically significant effects. Our results are similar to those found using Hamiltonian Monte Carlo (HMC) and Riemannian HMC methods based on Abanto-Valle et al. (2021).

4.3 - A joint analysis proposal of nonlinear longitudinal and time-to-event right-, interval-censored data

Rolando de la Cruz Mesía - Universidad Adolfo Ibáñez

Pregnancy in-vitro fertilization (IVF) cases are associated with adverse first-trimester outcomes in comparison to spontaneously achieved pregnancies. Human chorionic gonadotrophin beta subunit (beta-HCG) is a well-known biomarker for the diagnosis and monitoring of pregnancy after IVF. Low levels of beta-HCG during this period are related to miscarriage, ectopic pregnancy, and IVF procedure failures. Longitudinal profiles of beta-HCG can be used to distinguish between normal and abnormal pregnancies and to assist and guide the clinician in better management and monitoring of post-IVF pregnancies. Therefore, assessing the association between longitudinally measured beta-HCG serum concentration and time to early miscarriage is of crucial interest to clinicians. A common joint modeling approach is to

use the longitudinal beta-HCG trajectory to determine the risk of miscarriage. This work was motivated by a follow-up study with normal and abnormal pregnancies where beta-HCG serum concentrations were measured in 173 young women during a gestational age of 9-86 days in Santiago, Chile. Some women experienced a miscarriage event, and their exact event times were unknown, so we have interval-censored data, with the event occurring between the last time of the observed measurement and ten days later. However, for those women belonging to the normal pregnancy group; that is, carrying a pregnancy to a full-term event, right censoring data are observed. Estimation procedures are based on the Stochastic Approximation of the EM (SAEM) algorithm.
