

On the occupation time of 1-d exclusion processes

Patrícia Gonçalves
CMAT-Portugal

E-mail: patg@math.uminho.pt

Resumo

In this talk I will consider exclusion processes denoted by $\{\eta_t: t \geq 0\}$, evolving on the one-dimensional lattice Z and starting from the invariant state, namely, the Bernoulli product measure on $\{0,1\}^Z$ with constant parameter $\rho \in [0,1]$. The goal of the talk consists in establishing scaling limits of the functional $\Gamma_t(f) = \int_0^t f(\eta_s) ds$ for proper local functions f . When $f(\eta) := \eta(0)$, the functional $\Gamma_t(f)$ is called the occupation time of the origin. I will present a method that was recently introduced in Gonçalves and Jara (10) "Universality of the KPZ equation", from which we derive a local Boltzmann-Gibbs Principle for a general class of exclusion processes. For the occupation time of the origin, this principle says that the functional $\Gamma_t(f)$ is very well approximated to the density of particles. As a consequence, the scaling limits of $\Gamma_t(f)$ follow from the scaling limits of the density of particles. As examples I will present the symmetric simple exclusion, the mean-zero exclusion and the weakly asymmetric simple exclusion. For the latter, when the asymmetry is strong enough such that the fluctuations of the density of particles are given by the KPZ equation, we establish the limit of $\Gamma_t(f)$ in terms of this solution. The case of asymmetric simple exclusion will also be discussed. This is a joint work with Milton Jara (IMPA), with title "Scaling limits of additive functionals of interacting particle systems" and is accepted in 2012 for publication in Communications on Pure and Applied Mathematics.

O Perfil dos Campeões de Jogos Matemáticos

Paulo Infante ¹, Anabela Afonso ¹, Jorge Nuno Silva ²

¹ Escola Centro de Investigação em Matemática e Aplicações e Departamento de Matemática, ECT - Universidade de Évora

² Centro de História das Ciências da Universidade de Lisboa e Secção Autónoma de História e Filosofia das Ciências (SAHFC), FCUL

E-mail: pinfante@uevora.pt

Resumo

Nos últimos anos o interesse dos alunos pelos Jogos Matemáticos tem crescido exponencialmente. As escolas registam participações massivas dos seus alunos em atividades envolvendo os Jogos. Os próprios programas dos Cursos Profissionais do Ensino Secundário incluem um módulo de Jogos (módulo B5: Jogos e Matemática). O Campeonato Nacional de Jogos Matemáticos (CNJM) regista uma adesão de alunos que ultrapassou todas as expectativas, estimando-se que nas escolas o número de alunos jogadores que disputam as eliminatórias de apuramento para a final esteja próximo dos 100000, número muito superior ao de praticantes de qualquer outra atividade lúdica e desportiva em Portugal. Este Campeonato tem sido promovido, desde a sua primeira edição em 2004, pela Associação LUDUS, pela Sociedade Portuguesa de Matemática (SPM), pela Associação de Professores de Matemática (APM) e pela Ciência Viva. Em Março de 2012 disputou-se em Coimbra a 8ª edição, onde participaram quase 2500 alunos dos 3 ciclos do ensino básico e ensino secundário, pertencentes a quase 500 escolas de todo o país. Um número tão elevado de participantes é ainda mais significativo porque cada escola apenas pode inscrever um aluno por jogo e por nível de ensino.

Nesta comunicação apresentamos os primeiros resultados de um inquérito por questionário aplicado no último CNJM. Começamos por caracterizar os vencedores dos torneios que ficaram apurados para a finalíssima dos jogos que fizeram parte desta edição: Semáforo (1º ciclo), Cães e Gatos (1º e 2º ciclos), Ouri (1º, 2º e 3º ciclos), Hex (2º e 3º ciclos e secundário), Rastros (3º ciclo e secundário) e Avanço (secundário). Procuramos também avaliar a significância estatística de algumas associações, como por exemplo, analisar até que ponto o desempenho a disciplinas base como Matemática e Português está associado ao desempenho no campeonato. Finalmente, traçamos o perfil dos finalistas em função das preferências dos alunos, da frequência de prática de jogos matemáticos e outros jogos, do desempenho escolar a algumas disciplinas, das condições de realização da final, do grau de nervosismo e de algumas variáveis de natureza sociodemográfica.

Análise Classificatória de Dados de Natureza Complexa: Uma Aplicação do Coeficiente de Afinidade

Áurea Sousa¹, Helena Bacelar-Nicolau², Fernando C. Nicolau³, Osvaldo Silva⁴

¹ Universidade dos Açores, Departamento de Matemática, CEEApIA, CMATI

² Universidade de Lisboa, Faculdade de Psicologia, Laboratório de Estatística e Análise de Dados

³ Universidade Nova de Lisboa, FCT, Departamento de Matemática

⁴ Universidade dos Açores, Departamento de Matemática, CMATI

E-mails: aurea@uac.pt, hbacelar@fp.ul.pt, geral@datascience.org, osilva@uac.pt

Resumo

A Análise de Dados Simbólicos (ADS) é um domínio recente na área da descoberta do conhecimento e da gestão de dados. O seu primeiro passo consiste na extração de informações úteis a partir de bases de dados de elevada dimensão, como em *Data Mining*, de modo a sumariar esses dados em termos dos seus conceitos mais relevantes. O segundo passo em ADS é aplicar novas ferramentas às informações extraídas, visando estender o *Data Mining* ao *Knowledge Mining*.

O coeficiente de afinidade generalizado ponderado é uma medida de semelhança apropriada à Análise Classificatória (*Cluster Analysis*) de dados de natureza complexa (dados simbólicos), mesmo no caso de dados heterogéneos e de elevada dimensão. Neste trabalho, é dada uma atenção especial às medidas de proximidade entre objectos simbólicos, entre as quais o coeficiente de afinidade generalizado ponderado. É ilustrada a aplicação da Análise Classificatória Hierárquica Ascendente (*ACHA*) a dados de natureza complexa, retirados da literatura, com base no coeficiente de afinidade generalizado ponderado e em critérios de agregação clássicos e probabilísticos, estes últimos no âmbito da metodologia VL. Os resultados obtidos são comparados com os de outros autores.

Bibliografia

Bacelar-Nicolau, H., Nicolau, F., Sousa, A., Bacelar-Nicolau, L. (2009). Measuring similarity of complex and heterogeneous data in clustering of large data sets. *Biocybernetics and Biomedical Engineering*, v. 29, n. 2, 9-18, 2009.

Bock, H. -H., Diday, E. (Eds.) (2000). *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data*, Series: Studies in Classification, Data Analysis, and Knowledge Organization, Berlin: Springer-Verlag.

Lerman, I.C. (1981). *Classification et Analyse Ordinale des Données*, Dunod, Paris, 1981.

Nicolau, F., Bacelar-Nicolau, H. (1999). Clustering Symbolic Objects Associated to Frequency or Probability Laws by the Weighted Affinity Coefficient. In: *Applied Stochastic Models and Data Analysis. Quantitative Methods in Business and Industry Society*, H. Bacelar-Nicolau, F. C. Nicolau and Jacques Janssen (Eds.), INE, Lisbon, Portugal, 155-158.

Sousa, A. (2005). *Contribuições à Metodologia VL e Índices de Validação para Dados de Natureza Complexa*. Tese de Doutoramento, Universidade dos Açores, Ponta Delgada.

Comparação de Pares de Partições em Análise Classificatória

Osvaldo Silva¹, Helena Bacelar-Nicolau², Fernando C. Nicolau³, Áurea Sousa⁴

¹Universidade dos Açores, Departamento de Matemática, CMATI

² Universidade de Lisboa, Faculdade de Psicologia, Laboratório de Estatística e Análise de Dados

³ Universidade Nova de Lisboa, FCT, Departamento de Matemática

⁴ Universidade dos Açores, Departamento de Matemática, CEEApIA, CMATI

E-mails: osilva@uac.pt, hbacelar@fp.ul.pt, geral@datascience.org, aurea@uac.pt

Resumo

A comparação de pares de partições ajuda a compreender as relações entre métodos de classificação, parâmetros e características dos dados. A qualidade dos resultados de uma Análise Classificatória pode ser avaliada usando, no caso da comparação de duas partições, diversos coeficientes clássicos no contexto de três abordagens, com base, respectivamente, na contagem de pares, no emparelhamento das classes e na variação da informação. Os coeficientes permitem assim avaliar o grau de concordância entre as duas partições em avaliação, tendo em atenção diferentes informações. Cada um desses índices assume um determinado valor, de acordo com os aspectos presentes na sua expressão analítica, e alguns apresentam intervalos de variação diferentes e não variam no intervalo previsto mas somente numa parte desse intervalo. É apresentada, nesta comunicação, uma abordagem para a comparação de partições com base na semelhança VL (*Validade da Ligação*), a qual tem, entre outras, a vantagem de uniformizar a escala de medida.

Serão apresentados exemplos de aplicação da metodologia, no âmbito da qual foram testadas as abordagens clássicas e a abordagem VL e tecidas algumas considerações sobre os resultados obtidos.

Bibliografia

Bacelar-Nicolau, H. (1980). Contribuições ao Estudo dos Coeficientes de Comparação em Análise Classificatória, Tese de Doutoramento, FCL, Universidade de Lisboa.

Lerman, I. C. (1981). Classification et Analyse Ordinale des Données. Paris, Dunod.

Nicolau, F. (1983). Cluster Analysis and Distribution Function. Methods of Operations Research, 45, 431-433.

Nicolau, F.; Bacelar-Nicolau, H. (1998). Some Trends in the Classification of Variables. In: Data Science, Classification, and Related Methods, C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. H. Bock, Y. Baba (Eds.), Springer-Verlag, 89-98.

Silva, O. (2012). Contributos para a Avaliação e Comparação de Partições em Análise Classificatória. Tese de Doutoramento, Universidade dos Açores, Ponta Delgada.

Smoothing windows for the synthesis of Gaussian stationary random fields using circulant matrix embedding

Vladas Pipiras
University of North Carolina at Chapel Hill, USA

Email: pipiras@email.unc.edu

Resumo

When generating Gaussian stationary random fields, a standard method based on circulant matrix embedding usually fails because some of the associated eigenvalues are negative. The eigenvalues can be shown to be nonnegative in the limit of increasing sample size. Computationally feasible large sample sizes, however, rarely lead to nonnegative eigenvalues. In this talk, I will discuss how the standard circulant embedding method can be modified to overcome this problem. The basic idea is to smooth the embedded circulant matrix at the boundary by using a cutoff window or overlapping windows over a transition region. This straightforward approach works for many examples of interest, with the overlapping windows performing consistently better. (This is joint work with H. Helgason and P. Abry.)

Seleção de variáveis baseada na informação mútua: O que ainda não foi dito

Cláudia Pascoal¹, M. Rosário de Oliveira¹, Rui Valadas², António Pacheco¹
¹CEMAT e Departamento de Matemática, IST, Universidade Técnica de Lisboa
²IT e IST, Universidade Técnica de Lisboa

Emails: claudiapascoal@ist.utl.pt, rsilva@math.ist.utl.pt, rui.valadas@ist.utl.pt,
apacheco@math.ist.utl.pt

Resumo

Problemas reais envolvendo grandes conjuntos de dados são cada vez mais frequentes. Uma análise cabal dos mesmos implica a determinação das variáveis mais importantes na explicação do fenómeno em causa. Esta escolha tem a vantagem de diminuir a complexidade do problema, melhorar o desempenho do método estatístico em uso, permitir uma melhor visualização dos resultados e ainda uma mais fácil interpretação dos mesmos. As capacidades de armazenamento e registo de informação conduzem a conjuntos de dados cada vez maiores, contendo muitas variáveis irrelevantes e outras redundantes para o problema em estudo. A identificação deste subconjunto óptimo é conhecida por seleção de variáveis. Em problemas envolvendo grandes conjuntos de dados é frequente determinar este subconjunto antes da análise dos dados propriamente dita, sendo estes denominados métodos de filtragem.

Os problemas de classificação, têm grande relevância prática. Dado um conjunto de características (variáveis) e a classe a que cada objecto pertence têm por objectivo construir regras de classificação que permitem atribuir um novo objecto a uma das classes em estudo. Um subconjunto popular dos métodos de seleção de variáveis consiste em escolher as variáveis que optimizam uma dada função objectivo. Esta função pondera dois critérios: a relevância e a redundância de cada variável. A relevância é quantificada por uma medida de associação entre a variável candidata e a classe a que cada objecto pertence. A redundância é uma medida de quanto a informação relevante da variável candidata é tida em conta pelas variáveis já seleccionadas. Existe uma família de métodos de seleção de variáveis que escolhem como medida de relevância a informação mútua e divergem na quantificação da medida de redundância entre as variáveis. Os diversos autores baseiam-se em estudos de simulação para advogar as vantagens dos seus métodos em comparação com os já existentes na literatura. No entanto, e apesar dos cenários de simulação se basearem em variáveis aleatórias contínuas, o uso de estimadores de entropia e informação mútua construídos para variáveis aleatórias discretas introduzem enviesamentos que confundem os resultados obtidos.

Neste trabalho comparam-se vários métodos de seleção de variáveis a partir dos resultados teóricos, associados a dois cenários distintos. Esta comparação elimina possíveis enviesamentos devidos aos estimadores das diversas funções objectivo em estudo, clarificando-se resultados contraditórios existentes na literatura.

Políticas Óptimas de Investimento em Novas Tecnologias em Ambiente Aleatório

Cláudia Nunes

CEMAT e Departamento de Matemática, Instituto Superior Técnico, Universidade
Técnica de Lisboa

Email: cnunes@math.ist.utl.pt

Resumo

Na literatura de opções reais respeitante a políticas óptimas de investimento em novas tecnologias é usual assumir-se que novas tecnologias ficam disponíveis de acordo com um processo de Poisson homogéneo; em particular tal significa que após o aparecimento de uma nova tecnologia é tão provável que nova fique disponível em qualquer instante. Este princípio não é de todo razoável no panorama atual das inovações tecnológicas. Um dos paradigmas é a política de lançamento de novos produtos da empresa Apple; uma vez lançado uma nova versão de um iPad, por exemplo, os consumidores sabem que provavelmente uma nova versão só estará disponível passado cerca de 1 ano.

Nesta apresentação generaliza-se o modelo de forma a contemplar cenários em que a probabilidade de ocorrer uma inovação tecnológica altera-se com o tempo. De cada vez que há uma inovação tecnológica é possível investir, pagando para tal um custo irreversível I mas, em contrapartida, aumentando a eficiência associada ao processo. Desta forma o problema de investimento numa nova tecnologia é um problema de decisão, no âmbito dos problemas de controlo estocástico, sendo possível a sua resolução por via puramente matemática. Ao decisor apresentam-se, em cada instante, duas acções possíveis: i) investir em nova tecnologia ou ii) adiar a decisão de investimento; a decisão de investimento resulta de um balanço entre o custo de investimento e os ganhos esperados futuros.

Analisa-se diferentes cenários, dependendo do decisor esperar ou não uma descida da intensidade com que novas tecnologias ficam disponíveis. Analisa-se os efeitos da variância entre duas chegadas consecutivas na decisão de investimento, e mostra-se que uma maior variância acelera a opção de investimento no caso da intensidade aumentar.

No geral, mostra-se que um aumento da incerteza acelera a decisão de investimento, facto que é expectável na Teoria das Opções Reais.