

Proposta para Dissertação de Mestrado

Título: Melhoria da TestH, uma Biblioteca em C para Geração de Processos Auto-Semelhantes e Estimação do Parâmetro de Hurst

Orientador e Contacto

Doutor Pedro R. M. Inácio (inacio@di.ubi.pt)

Sumário / Enquadramento

As séries auto-semelhantes são conjuntos de valores com ordem temporal que apresentam uma estrutura de auto-correlação muito particular e que define que cada valor da série depende de valores anteriores com mais ou menos intensidade. A aplicação destas séries em várias áreas do conhecimento e a sua implicação nos resultados de análises estatísticas têm vindo a ser alvo de estudo e interesse mais ou menos constante ao longo dos últimos 70 anos. Séries de valores auto-semelhantes com distribuição gaussiana podem ser vislumbradas em concretizações de vários processos naturais e artificiais, nomeadamente no caudal dos rios ou nos preços de ações na bolsa de valores. Há 25 anos, a descoberta da natureza auto-semelhante do tráfego de rede aproximou ainda mais esse estudo da disciplina da Informática, e nomeadamente da área da monitorização e análise de tráfego de rede.

Motivados pela importância dessas séries, pelo que podem significar em termos de análise estatística, e pela necessidade de as ter em conta durante o estudo de outros problemas, foram sendo desenvolvidos, ao longo dos anos, diversos métodos para (i) medir a intensidade da auto-semelhança numa determinada série de valores e (ii) simular séries auto-semelhantes em computador. Cada um dos métodos propostos é normalmente apresentado com as suas vantagens e desvantagens (e.g., alguns podem operar valor-a-valor, de modo incremental, enquanto que outros só produzem resultados após todos os valores terem sido processados), sendo possível encontrar implementações destes métodos disponíveis na Internet, inclusive no âmbito de algumas ferramentas que os permite utilizar de um modo integrado, recorrendo ainda a uma interface gráfica. Contudo, não existem bibliotecas em C devidamente comentadas e testadas que implementem todos (ou quase todos) os métodos referidos.

Objetivos

Este projeto visa a implementação de vários métodos descritos na literatura para estimação do parâmetro de Hurst, que é o parâmetro que mede a intensidade da auto-semelhança numa série de valores, e para simulação computacional destas séries. A implementação deve ser feita em ANSI C, indo de encontro à implementação já feita e apresentada em [1]. Os algoritmos a implementar são os que ainda não estão presentes na versão que pode ser obtida do *Uniform Resource Locator* (URL) <https://github.com/diogo-fernan/testh>.

Parte do esforço de investigação estará no entendimento dos métodos, nor-

malmente baseados em estatísticas calculadas para a série de valores e para transformações dessa mesma série, bem como na sua otimização e adequação à forma de operar da ferramenta. O segundo objectivo deste projeto refere-se à documentação e ao formalismo que acompanha o trabalho, que deve, por exemplo, estar ao nível da implementação do conjunto de testes disponível em <http://simul.iro.umontreal.ca/testu01/tu01.html>. Esta referência sobre testes estatísticos desenvolvida em linguagem C pode, de resto, ser tomada como uma linha de orientação acerca do que se pretende deste trabalho. O cldoc pode ser considerado como um possível candidato para produção da documentação.

Todos os métodos implementados têm de ser validados, devendo esta fase ser relatada na forma de um artigo científico, cuja submissão numa revista científica internacional será avaliada atempadamente. Espera-se que o(a) aluno(a) fortaleça as suas competências na área da análise estatística aplicada e fortaleça as suas aptências de programação, bem como melhorar o domínio da língua Inglesa (a documentação e a dissertação devem ser escritas em inglês). Quem envereder por este projecto terá a oportunidade de se inserir numa equipa de investigação científica dinâmica e disposta a partilhar ideias e conhecimento.

Tarefas

São propostas as seguintes tarefas para a execução do trabalho de investigação e de desenvolvimento, conducente à elaboração da dissertação de mestrado:

Tarefa 1 Familiarização com o tema e com a implementação TestH; estudo do estado da arte (1 mês).

Tarefa 2 Estudo pormenorizado de vários métodos para simulação de séries auto-semelhantes e implementação dos mesmos (2 meses).

Tarefa 3 Estudo pormenorizado de vários métodos para estimação do parâmetro de Hurst e sua implementação (2 meses).

Tarefa 4 Teste, validação e comparação de todos os métodos ou algoritmos presentes na ferramenta (2 meses).

Tarefa 5 Melhoria da documentação e aprimoramento das bibliotecas (1 mês).

Tarefa 6 Escrita da dissertação de mestrado e de um artigo científico (2 meses).

Cronograma

A tabela seguinte representa a calendarização prevista para a execução das tarefas, em que a execução de uma dada tarefa num determinado mês é assinalada com uma cruz (x).

Tarefa \ Mês	Set.	Out.	Nov.	Dez.	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.
1	x										
2		x	x								
3				x	x	x					
4							x	x			
5									x		
6										x	x

Resultados Esperados

- Um pacote com o código fonte da implementação dos diversos estimadores e geradores, que também deve incluir a sua completa documentação;
- Um artigo num encontro ou revista internacional com arbitragem científica, apresentando o estudo efetuado no âmbito deste mestrado;
- Uma dissertação de mestrado.

Escrita da Dissertação e de Artigos Científicos em Língua Inglesa

A dissertação de mestrado e o artigo científico resultantes da realização do plano de trabalho proposto deverão ser objecto de escrita em Língua Inglesa, tendo em vista a divulgação internacional do trabalho científico desenvolvido. O título da dissertação em Língua Inglesa deverá ser o seguinte: *“Improvement of TestH, a C Library for Generating Self-Similar Series and Estimating the Hurst Parameter”*.

Referências

- [1] Diogo A. B. Fernandes, Miguel Neto, Liliana F. B. Soares, Mário M. Freire and Pedro R. M. Inácio, A Tool for Estimating the Hurst Parameter and for Generating Self-Similar Sequences, in Proceedings of the 46th Summer Computer Simulation Conference 2014 (SCSC 2014), Monterey, CA, USA, July 6-10, 2014.