



# “Longa Dependência em Sequências de DNA Utilizando a Metodologia de Mudança de Regimes”

Pietro Tiaraju Giavarina dos Santos <sup>1</sup>, Raquel Romes Linhares <sup>2</sup>

1. Acadêmico de Estatística da UFRGS. Bolsista BIC/UFRGS  
2. Doutora em Matemática. Professora Adjunta da Universidade Federal de Uberlândia (UFU)



## INTRODUÇÃO

Uma sequência de DNA é composta de bases A (adenina), C (citosina), T (timina) e G (guanina). A fim de aplicar métodos numéricos a uma sequência de nucleotídeo é necessário transformá-la em uma sequência numérica. Se considerarmos a abordagem de transformação numérica arbitrária para A, C, T, G e depois utilizarmos a análise espectral, o resultado irá depender de cada transformação em particular, portanto não sabemos se a existência de longa dependência em sequência de DNA é induzida pela transformação. Nossa proposta é utilizar a metodologia de mudança de regimes para analisar a longa dependência em sequência de DNA. Nesta metodologia, se a duração dos regimes de uma série temporal tem uma distribuição de cauda pesada com parâmetro  $\alpha$  em  $(0,2)$ , então a série temporal apresenta característica de longa dependência. Ao se aplicar qualquer transformação linear que preserva a propriedade de variância finita na série temporal, ela igualmente preserva a propriedade de longa dependência. Portanto, nosso interesse é aplicar esta metodologia em sequências de DNA, para saber se a longa dependência em sequência de DNA é propriedade da sequência, ou se ela é induzida pela transformação.

## OBJETIVO

Nosso objetivo é analisar a longa dependência em sequências de DNA utilizando a metodologia de mudanças de regimes proposta por Liu (2000), para saber se a longa dependência em sequência de DNA é propriedade da sequência, ou se ela é induzida pela transformação.

## METODOLOGIA

Primeiro apresentamos a definição de distribuições estáveis, que é uma classe rica de distribuições de probabilidade que permitem caudas pesadas e têm fascinantes propriedades matemáticas. Após, é apresentado o processo de mudança de regime e suas propriedades. Através dele é mostrado que a longa dependência é preservada, quando é aplicada qualquer transformação linear que preserva a propriedade de variância finita no processo pré-definido com mudança de regimes.

## RESULTADOS

Geramos séries temporais  $\alpha$ -estáveis simétricas de média 0, tamanho amostral 1000 e 1000 replicações. Calculamos o valor médio, variância, desvio quadrático médio e vício. Através disso, conseguimos visualizar que os melhores resultados foram obtidos para  $\alpha$ s próximos aos extremos, especialmente para  $\alpha$ s grandes, próximos de 2. Os piores casos foram observados para  $\alpha$ s próximos de 1. Para análise da taxa de decaimento da cauda consideramos o estimador utilizado para gerar as séries temporais  $\alpha$ -estáveis. Testamos as hipóteses seguintes: o estimador resultar em 2 ou num valor menor que 2. Com isso, foi passado para análise uma série de sequências de DNA pré-estabelecidas e verificamos que, a nível de significância de 5%, para todas as sequências de DNA estudadas que a longa dependência nas vinte e quatro sequências é realmente propriedade da sequência e não da transformação utilizada.



MODALIDADE  
DE BOLSA

**BIC/UFRGS**