



<b>Evento</b>	Salão UFRGS 2013: SIC - XXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2013
<b>Local</b>	Porto Alegre - RS
<b>Título</b>	Longa Dependência em Sequências de DNA Utilizando a Metodologia de Mudança de Regimes
<b>Autor</b>	PIETRO TIARAJU GIAVARINA DOS SANTOS
<b>Orientador</b>	RAQUEL ROMES LINHARES

Uma seqüência de DNA é composta de bases A (adenina), C (citosina), T (timina) e G (guanina). A fim de aplicar métodos numéricos a uma seqüência de nucleotídeo é necessário transformá-la em uma seqüência numérica (Crato et al., 2010 e Crato et al., 2011). Se consideramos a abordagem de transformação numérica arbitrária para A,C,T,G e depois utilizamos a análise espectral, o resultado irá depender de cada transformação em particular, portanto não sabemos se a existência de longa dependência em seqüência de DNA é induzida pela transformação. Nosso objetivo é analisar a longa dependência em seqüência de DNA utilizando metodologia de mudança de regimes proposta por Liu (2000). 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 a característica de longa dependência. Ao se aplicar qualquer transformação linear que preserve a propriedade de variância finita na série temporal, ela igualmente preserva a propriedade de longa dependência. Portanto, nosso interesse é aplicar esta metodologia em seqüências de DNA, para saber se a longa dependência em seqüência de DNA é propriedade da seqüência, ou se ela é induzida pela transformação.

Descrevemos de forma resumida, as metodologias utilizadas na pesquisa: apresentamos as distribuições estáveis e um estimador para o parâmetro de estabilidade  $\alpha$ , denotado por  $\hat{\alpha}_{fce}$ . Definimos um processo estocástico com mudança de regime e apresentamos algumas de suas propriedades.

Nosso objetivo principal foi analisar a longa dependência em seqüência de DNA utilizando metodologia de mudança de regimes. Consideramos que a mudança de regime ocorre em uma seqüência de DNA, quando há uma mudança na base da seqüência. Seja  $\{w_k\}_{k=1}^m$  os  $m$  diferentes regimes da seqüência de DNA. Seja  $\{T_k\}_{k=1}^m$  as durações de regimes para a seqüência de DNA. Investigamos a propriedade de longa dependência, focalizando nossa atenção no parâmetro estável  $\alpha$  da série temporal  $\{T_k\}_{k=1}^m$ , para vinte e quatro seqüências de DNA. Para a análise da taxa de decaimento da cauda, consideramos o estimador  $\hat{\alpha}_{fce}$ . Observamos que  $\hat{\alpha}_{fce} \in (1, 2)$  a nível de significância de 5%, para todas as seqüências de DNA aqui estudadas. Portanto, pelo mecanismo de mudança de regime proposto por Liu (2000), as seqüências de DNA aqui analisadas, apresentam a característica de longa dependência e que qualquer transformação linear  $f(\cdot)$  aplicada na seqüência de DNA, que preserve variância finita, igualmente preserva a propriedade de longa dependência da seqüência de DNA. Portanto, a longa dependência nas vinte e quatro seqüências de DNA consideradas neste trabalho é realmente propriedade da seqüência e não da transformação utilizada.

## Referências

- Crato, N., R.R. Linhares e S.R.C. Lopes (2010). “Statistical Properties of Detrended Fluctuation Analysis”. *Journal of Statistical Computation and Simulation*, Vol. **80**(6), 625-641.
- Crato, N., R.R. Linhares e S.R.C. Lopes (2011). “ $\alpha$ -Stable Laws for Noncoding Regions in DNA Sequences”. *Journal of Applied Statistics*, Vol. **38**(2), 261-271.
- Liu, M. (2000). “Modeling Long Memory in Stock Market Volatility”. *Journal of Econometrics*, Vol. **99**, 139-171.