

Allan 分散プロットの DNA 塩基配列データへの応用

Allan Varicance-like Plot for DNA Sequence Data

天野 晃

AMANO, Kou

理化学研究所・農業生物資源研究所

RIKEN / National Institute of Agrobiological Sciences

あらまし: Allan Variance (アラン分散)[1, 2] とは、発振器やセンサーの周波数特性に関する指標であり、これをプロットしたものは、Allan Variance Plot (アラン分散プロット、あるいは単にアランプロット) と呼ばれ、これらによる分析をアラン分散分析という。アラン分散分析では、時間領域において評価対象となるモデルもしくはセンサーのノイズ成分の定量評価ができる。これは、周波数領域におけるフーリエ解析と相補的役割をなす。このような時系列シーケンス解析の技術は、DNA 塩基配列データによく応用されるが、アラン分散分析が応用された報告はない。筆者は、アラン分散分析を DNA 塩基配列データ解析に応用する手法を検討中である。本報告では、この応用の基本的な考え方と、簡単な例を紹介する。

キーワード: Allan Variance, DNA Sequence

1 本研究の目的

本研究の目的は、ひとまずは、アラン分散分析の DNA 塩基配列データへの応用可能性を探ることである。現段階では、手法を検討中であり、本報告では検討中の手法を報告する。

2 Allan Variance Plot

アラン分散分析の対象は、時系列データである (オシレーターの発振データなど)。このデータを、 $\mathbf{a} = \{a_1, a_2, a_3, \dots\}$ とする。

このデータに対するアラン分散プロットは、以下の手順で行う。

1. 時間幅 (window の大きさ) w を決定する
2. ベクトル \mathbf{a} より、幅 w からなるすべてのサブリストを生成する
3. すべての (時間領域的に) 隣接するサブリスト (window) 間の平均差をとる
4. その平均差の値を 2 乗したものの平均に、 $1/2$ を掛けた値がアラン分散となる
5. 上記 1-4 のステップを、可能な w につき行う
6. w を横軸座標に、対応するアラン分散のルートを縦軸座標に、両対数プロットしたものがアラン分散プロットである

3 Allan Variance Plot の DNA 塩基配列データへの応用

アラン分散プロットは観測値シーケンスの局所平均差の平均をプロットしたものだといえるが、このデータを DNA 塩基配列に置き換えて考えるなら、全塩基配列における部分塩基配列の相異 (距離) の平均を得ることができれば、アラン分散様のプロットが得られる。

筆者は以下のような手順によりこのプロットが可能であると考ええる。

まず、DNA 塩基配列データを、 $\mathbf{a} = \{a_1, a_2, a_3, \dots\}$ とする (a は、塩基)。このデータに対するアラン分散様プロットは、以下の手順で行う。

1. 部分塩基配列 (window) の大きさ w 決定する
2. 全塩基配列データ \mathbf{a} より、幅 w からなるすべての部分塩基配列を生成する
3. すべての (配列領域的に) 隣接する部分塩基配列 (window) 間の距離をとる
 - 距離はオリゴヌクレオチド頻度ベクトルにおけるユークリッド距離とする
 - オリゴヌクレオチド頻度は部分配列において取りうるすべてのオリゴヌクレオチド長 n を考える
4. その距離を 2 乗したものの平均に、 $1/2$ を掛けた値がアラン分散様値となる
5. 上記 1-4 のステップを、可能な w 、 n につき行う
6. 各 n につき、 w を横軸座標に、対応するアラン分散様値のルートを縦軸座標に、(両対数) プロットしたものがアラン分散様プロットとなる

4 応用

当日の発表では、実際の DNA 塩基配列データに応用した例を紹介する。

謝辞

シュルンベルジェ株式会社 佐藤滋氏には、Allan Variance Plot の詳細を教授いただきました。感謝致します。

注・文献

- [1] Allan D. W. Statistics of Atomic Frequency Standards. *Proc. of the IEEE*, Vol. 54, No. 2, pp. 221–230. 1966.
- [2] IEEE Standard Specification Format Guid and Test Procedure for Signal-Axis Interferometric Fiber Optic Gyros, IEEE Std 952-1997. 1998.