

# 短時間ノイズのパワースペクトル及び Detrended Fluctuation Analysis への影響

只木進一

佐賀大学総合情報基盤センター

## 概要

時系列データの時間解像度の改善は、短時間のノイズを時系列データに持ち込むことがある。こうした短時間ノイズが、パワースペクトル及び非定常時系列の長時間相関の解析手法の一つである Detrended Fluctuation Analysis (DFA) の結果に影響を与える。人工的な時系列を用いることで、この影響を議論する。短時間の粗視化をすることで、長時間の相関を復元することができる。

## Effects of short-range noises on the power spectrum and the Detrended Fluctuation Analysis

Shin-ichi Tadaki

Computer and Network Center, Saga University

## Abstract

Improving time resolution of data sometimes introduces short-range random noises into temporal data sequences. These noises affect the result of the power spectrum and the Detrended Fluctuation Analysis, which is one of useful methods for analyzing long-range correlations in non-stationary sequences. The effects are discussed based on artificial temporal sequences. The long range correlation is recovered by coarsening those noises.

## 1 はじめに

近年、時系列データなどに含まれる長時間相関の存在が、様々な分野で関心を集めている。様々なシステムの動的性質に、べき則揺らぎの存在が指摘されている。このような多様なシステムにべき則揺らぎが現れる起源も大きな関心の対象である。特に、近年、大量のデータを用いることで現実の系の性質を理解する試みが行われている。

このような時系列データは、通常は非定常的である。非定常時系列に含まれる長時間相関の解析手法の一つに Detrended Fluctuation Analysis (DFA) がある。元々、DFA は、デオキシリボ核酸 (DNA) 系列の解析として開発された手法である [1, 2]。この手

法は、これまで、様々な時系列の解析に使われ、べき則揺らぎの存在を明らかにしてきた。

時系列に DFA を適用する場合には、最初に、元の時系列の平均からの偏差和 (profile) を作成する。次に、時系列をある長さ  $l$  に、重複のないように分割し、各区間ごとに、profile の平均的挙動 (trend) を最小二乗法などで作成する。この trend を直線として求める方法を 1 次の DFA と呼ぶ。本稿では 1 次の DFA を用いた議論を行う。

各区間における profile の trend からのずれの標準偏差を  $F(l)$  とする。元の時系列がべき則揺らぎを有する場合には、 $F(l)$  は区間長  $l$  に対して

$$F(l) \sim l^\alpha. \quad (1)$$

のように振る舞う。これはパワースペクトル  $P(k)$  のべき指数と以下のように対応付けられる。

$$P(k) \sim k^{-\gamma}, \quad \gamma = 2\alpha - 1. \quad (2)$$

実測データの量的増加は、しばしば解像度の向上を伴う。このような解像度の改善は、長時間相関を解析する上でどのような影響を与えるであろうか。例として、インターネット流量の解析の場合を考えよう [3]。インターネット流量は、MRTG (Multi Router Traffic Grapher) と呼ばれるソフトウェアで収集される [4]。MRTG はスイッチやルータといった通信機器から SNMP (Simple Network Management Protocol) という方式でデータを収集する。デフォルトの設定では、流量が5分ごとの積算として得られる。この間隔を短くすると、様々な種類の乱れが入ってくる。たとえば、統計的揺らぎ、コンピュータや通信装置の非同期的挙動、利用者の非同期的挙動の影響である。

本稿では、DFA の解析結果に、このような短時間領域の乱れがどのように影響するかを調べることにする。そのために、長時間相関を有する人工的時系列を生成し、短時間領域の乱れの影響を検討する。また、短時間領域を積算し、観測時間を粗くすることで、元の相関が復元できることを示す。

## 2 Fourier Filter Method

べき則揺らぎを持つ時系列を人工的に作る手法の一つに Fourier Filter Method (FFM) がある [5, 6]。この手法に対して、相関の長さを拡張する方法も提案されている [7]。ここでは簡単のため、オリジナルの FFM を用いる。

まず、相関のないランダムな長さ  $T$  の実時系列  $\{u_t\}$  ( $t = 0, 1, \dots, T-1$ ) を作成する。その相関関数は

$$C_\tau = \frac{1}{T} \sum_{t=0}^{T-1} u_t u_{t+\tau}. \quad (3)$$

である。相関関数の Fourier 成分は以下のように、時系列の Fourier 成分  $\hat{u}_k$  で表される。

$$\hat{C}_k = \frac{1}{\sqrt{T}} \sum_{\tau=0}^{T-1} e^{-2\pi i k \tau / T} C_\tau = \frac{1}{\sqrt{T}} \hat{u}_k \hat{u}_{-k}, \quad (4)$$

Fourier 成分  $\hat{u}_k$  の振幅を変化させることでべき則揺らぎを導入できることが解る。そのためのフィルタ  $S(k)$  を定義する。

$$S(k) = k^{-\gamma}. \quad (5)$$

このフィルタを使って新しい Fourier 成分を定義する。

$$\hat{\eta}_k = S^{1/2}(k) \hat{u}_k \quad (6)$$

これに対応する時系列  $\{\eta_t\}$  はべき則揺らぎを有する。

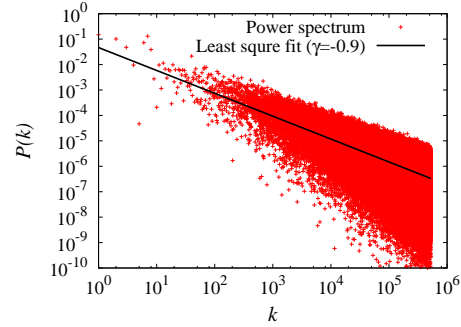


図 1:  $\gamma = 0.9$  の FFM によって生成した時系列に対するパワースペクトル。直線は最小二乗法によるフィットである。

このようにして、生成した時系列 ( $\gamma = 0.9$ ) に対するパワースペクトルを図 1 に示す。全データ点に対して最小二乗法で傾きを求めると、 $\gamma = 0.9$  を再現することができる。このデータに対して DFA を実行すると、対応する指数  $\alpha = 0.95$  を得る (図 2)。

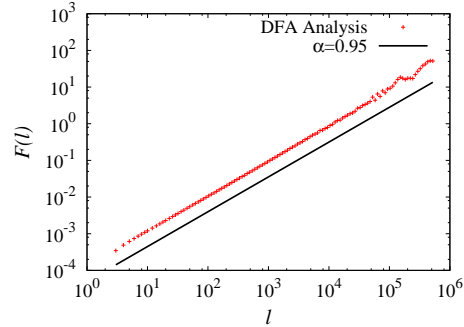


図 2:  $\gamma = 0.9$  の FFM によって生成した時系列に対する DFA 解析の結果。対応する指数  $\alpha = 0.95$  を得ることが出来る。

## 3 短時間の雑音

短時間の領域に雑音が入った場合を考えるために、フィルタを以下のように変形する。

$$S(k) = k^{-\gamma} + k_c^{-\gamma}, \quad (7)$$

ここで、 $k_c$  は定数であり、べき則揺らぎのカットオフに対応する。 $k_c$  より波数が大きい、つまり短時間領域では、フィルタ  $S(k)$  はほぼ一定となる。 $k_c$  より波数が小さい、つまり長時間領域では、べき則に従うことになる (図 3)。

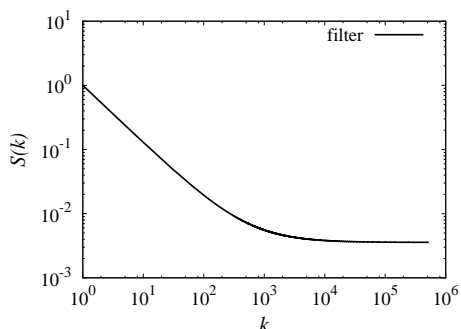


図 3: カットオフのあるフィルタ。  $k_c = 10^3$  としている。

このようにして生成した時系列に対するパワースペクトルを図 4 に示す。データ点全体を使って傾きを求めると  $\gamma = 0.02$  とほとんどフラットになってしまう。つまり、短時間領域の雑音が、全体のべき則揺らぎを覆い隠してしまう。 $k < 10^3$  の領域内のデータだけを使って傾きを求めると  $\gamma = 0.66$  と小さな値となる。つまり、パワースペクトルを見ても、べき則揺らぎの存在を明かにすることができない。

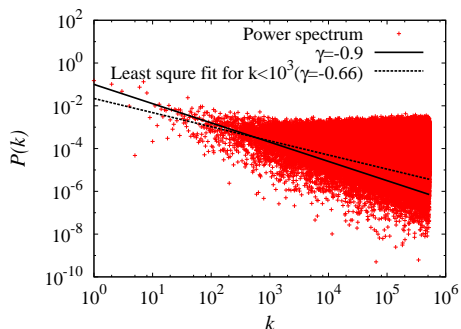


図 4: カットオフを入れたフィルタを用いて作成した時系列のパワースペクトル。実線は  $\gamma = 0.9$  を表している。 $k < 10^3$  の領域だけを使って傾きを求めたものを波線で表す。このときの指数は  $\gamma = 0.66$  である。

このような場合に、DFA はべき則揺らぎの存在を見いだす有効な方法である。DFA の解析結果を図 5 に示す。二つの領域のクロスオーバーが見える。短い区間長に対応する領域の傾きは 0.5 であり、ラン

ダムな時系列に対応する。長い区間長に対応する領域の傾きは 0.95 であり、フィルタで導入したべき則揺らぎに対応している。二つの領域は  $k_c$  に対応する区間長で接合している。

つまり、べき則揺らぎのある時系列に短時間領域のノイズがある場合、DFA を用いることで、べき則揺らぎの部分を見いだすことが可能である。

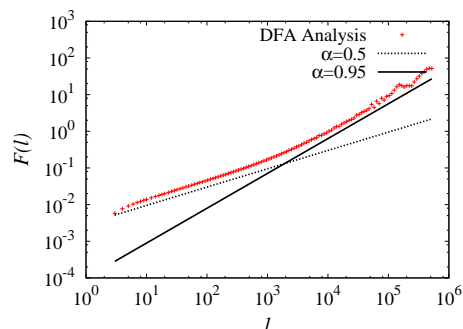


図 5: カットオフを入れたフィルタを用いて作成した時系列の DFA 解析の結果。

## 4 粗視化による解析

短時間の雑音に加わった時系列が得られたときに、その影響を除いて長時間相関を求める方法のもっとも簡単な方法は、短時間領域を粗視化することです。本節では、そのような粗視化によって、元の長時間相関が復元することを確認する。

前節で扱った、べき則揺らぎのある時系列に短時間の雑音に加わった時系列を粗視化した場合を考える。長さ  $T/k_c$  の区間ごとの積算量を新たな時系列とする。これに対するパワースペクトルを図 6 に示す。元のべき則揺らぎ  $\gamma = -0.9$  をほぼ再現することができた。

この時系列に対する DFA の結果を図 7 に示す。こちらも、対応する傾きを再現することが出来る。

## 5 まとめと議論

様々な時系列に含まれる長時間相関やべき則揺らぎに関心が寄せられている。計測データの増大は、しばしば、計測精度、特に時間解像度の改善をもたらすことがある。一方、時間解像度の改善は、様々な非同期雑音成分の混入を招き、長時間の挙動の解析を困難にする可能性がある。

本研究では、べき則揺らぎを持つ人工的時系列と、

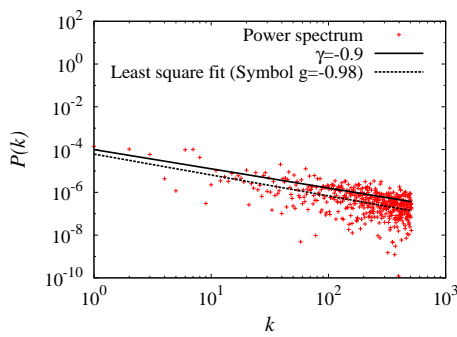


図 6: 長さ  $T/k_c$  の区間ごとの積算値に対するパワースペクトル。実線は  $\gamma = -0.9$ 、波線は全点に対するフィットにより得た傾き  $\gamma = -0.98$ 。

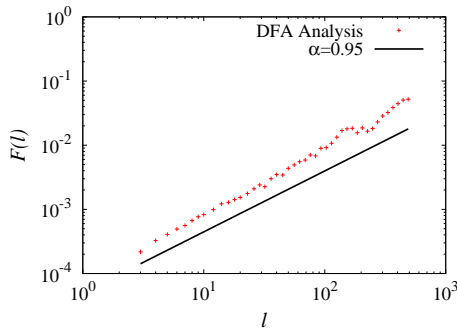


図 7: 長さ  $T/k_c$  の区間ごとの積算値に対する DFA の結果。

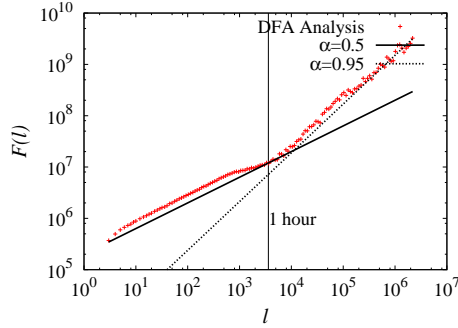


図 8: 秒単位で記録したメール送信量に対する DFA の結果。

折れ曲がりの見える点は約 1 時間である。それよりも長い時間領域では、 $\alpha = 0.95$  程度の傾きであり、長時間相関の存在が見える。

つまり、電子メール送信の流量計測の場合には、1 時間よりも短い時間で計測しても、接続機器や通信の非同期性から、揺らぎが大きくなり、長時間相関に関しては有用な知見が得られない。

今の場合、データ点の数は  $10^7$  個のオーダーである。一方、データの最小時間解像度は秒であり、 $10^3$  個程度を粗視化することで、長時間相関に関しては有用な知見を得ることになる。本稿で検討した、 $10^6$  個のデータのうち、 $10^3$  個の長さの部分に短時間ノイズがある場合に対応している。

さらに短時間の雑音がある場合について、パワースペクトル及び Detrended Fluctuation Analysis (DFA) の結果を検討した。

短時間雑音は、パワースペクトル解析を非常に困難とし、雑音成分が主成分として表れてしまう。一方、DFA においては、短時間のランダムな挙動と長時間のべき則に対応した挙動のクロスオーバーとして表れ、長時間相関があることを発見することができる。クロスオーバーする点は、雑音成分が入る特徴的長さに対応している。つまり、短時間雑音のある場合には、パワースペクトル解析より DFA 解析が有用である。

最後に、このような事例が発生する可能性について検討する。電子メールの送信用サーバには、メール送信の記録が秒単位で記録される。約 2 月にわたるメール送信量に対する DFA の結果を図 8 に示す。本稿で人工的時系列を解析した際と同様に、短時間側の傾きが  $\alpha = 0.5$  となる領域が見える。データの

## 参考文献

- [1] C.-K. Peng, et al., Phys. Rev. E **49**, 1685 (1994).
- [2] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, Chaos **5**, 82 (1995).
- [3] S. Tadaki, J. Phys. Soc. Japan **76**, 044001 (2007).
- [4] T. Oetiker, <http://oss.oetiker.ch/mrtg/>
- [5] C.-K. Peng, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. A **44**, R2239 (1991).
- [6] S. Prakash, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. A **46**, R1724 (1992).
- [7] H. A. Makse, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. E **53**, 5445 (1996).