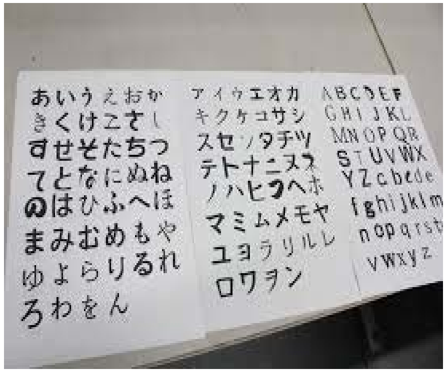
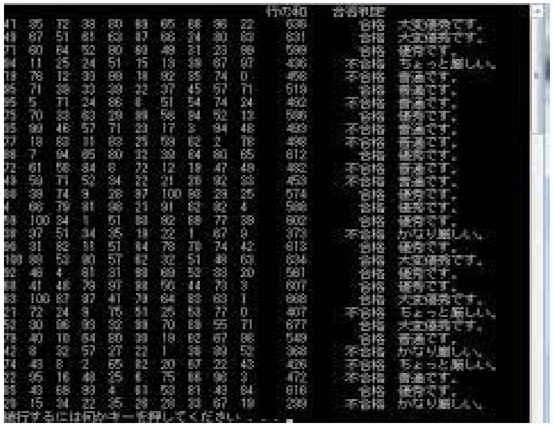
スケール不変性の検証方法とDNA配列の解析

SP 10144　吉田俊介　非線形物理学研究室

はじめに

我々の身の周りには**ランダムな配列**が多く存在している。





そのランダムな配列にも**何かしらのパターン**があるかどうか調べるためにその配列を**0,1の数列**にして解析するという研究が行われてきた。

目的

```
ORGANISM  Clostridium botulinum A2 str. Kyoto
1100111111110111111010100110100110100010101000101111111111001001
1011000110010100010010101010100000110111010111110111011010101000
1011110001111110110000011000000011000110100010111011111010000011
1000000000010010001011010000000100011101000000100011110010101100
0110000001011001111011111111100000010011101100000110011010110001
0011110100000000001111100000010001010000001111010011001011110000
1101110111100000000101001010001111011010011011000000000000010000
0000011010010010000000111001010011010001001000001011101101010110
0011110001010111111001110111010010000000100100100011101110101010
0101111110010010010000011011101100000101010011111110111010101001
0000110011110010000011101110010101011011001101101010110101001
0011110001111000100111010101111111100000001110100010001100110 01
1110100000000100100000010100000111000000100010100000010100010111 10
1100110010010111001110110110000000000000000011100000000001111111 01
0111111001010110101000001100100010001001101011000100110111111 10
000000001011101111100000001000110000011100011100010000000111010011
0010111000101110011110000011000001001101111000000000000001100101 0
0000000110001011111100100000100100101010100110100000011000110001
0111101000110000000001111001000000100100111111101101110110010000
```

図:実際の0,1数列にしたDNA配列

その従来の研究から0,1の数列のパターンは、
ある長さLにおける1の個数の分布の分散V_LがLのベキ乗に
従うことが知られてきた。
また、ベキ性を持つことからスケール不変性であることも同時にいえる。
本研究では、従来の方法ではなく新たな手法として
相関関数のスケーリング解析を提案し0,1数列に変換された
DNA配列がスケール不変性をもつかどうかを検証する。

発表の流れ

- スケール不変性とは？
- 従来の方法とは？
- 0、1の数列のスケール不変性とは？
- 相関関数のスケーリング解析とは？
- 検証結果

スケール不変性とは？

図1:カリフラワー

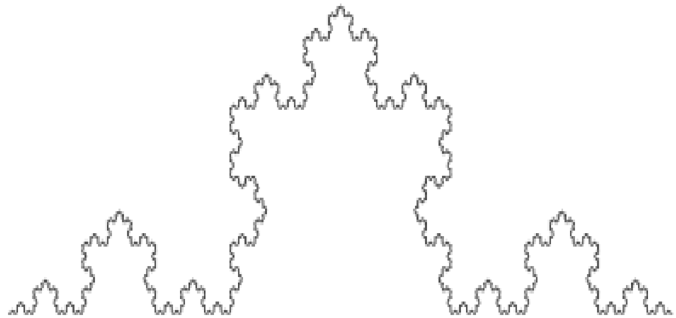


図2:コッホ曲線

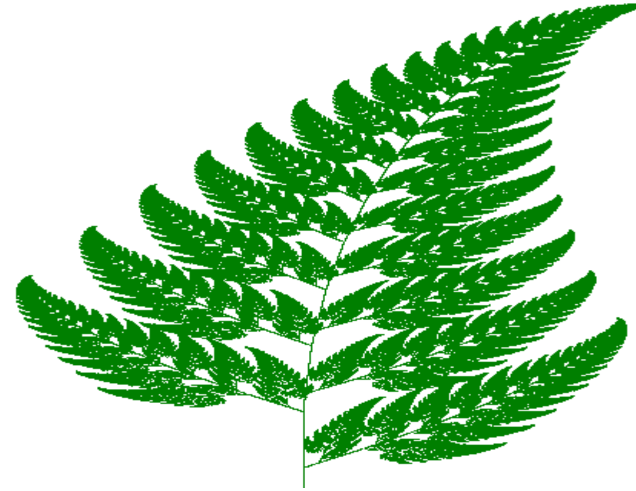


図3:シダの葉

一部を拡大しても全体に見え、縮小してもその
一部と区別がつかない。

スケール不変性

$V_L(N_1) \propto L\hat{}\,\alpha$

従来の方法から知られてきたこと

0010100001010100100001001010001010010100101010101100 10100
1001001010101000010111010110100101010101010······

従来の方法だと、、

·········0100010101010010010········

L

長さLにおける1の個数(N_1)の分散$V_L(N_1)$がべき性に従う

$1 < \alpha < 2$

$V_L(N_1) \propto L$

ノーマル拡散相

αによって配列のパターンは決まってしまう

$V_L(N_1) \propto L^{\alpha}$

スーパー拡散相

0,1の数列のスケール不変性とは
べき性をもつこと

00101000010101001000010010100010100101001010101011001010010010010101010000101110101101001010101010101010……

従来の方法だと、、

………0100010101010100100010………

L

長さLにおける1の個数(N_1)の分散$V_L(N_1)$

$1<\alpha<2$

$V_L(N_1) \propto L$

ノーマル拡散相

αによって配列のパターンは決まってしまう

$V_L(N_1) \propto L^{\alpha}$

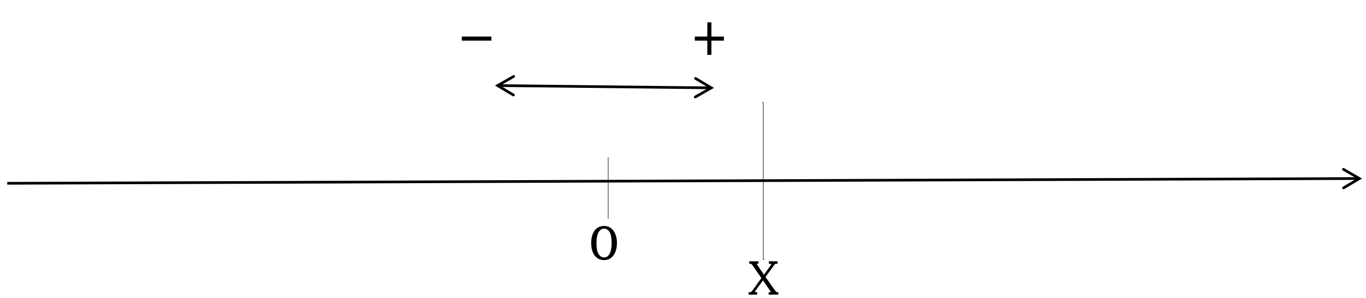
スーパー拡散相

従来の方法の例

例　ランダムウォーク

1の個数: N_1　右に1進む
−1の個数: N_0　左に1進む

1秒ごとに
確率1/2で1か−1か振る舞う



t秒後にいる位置 $X = N_1 - N_0 = 2N_1 - t$

分散 $V(X) = 4 \cdot V(N1) \propto t$

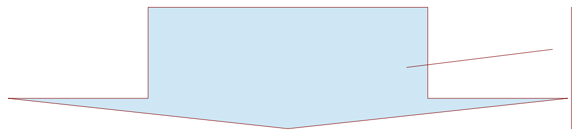
スケール不変性

ノーマル拡散相

DNA配列のスケール不変性
（従来の方法）

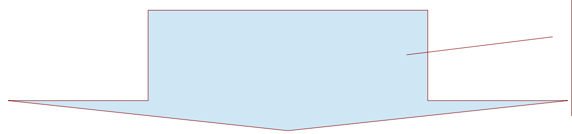
長さ100万以上

ACGTGTACGTACGTACGTACCGAGTCGATGATCGTAGTACGATCGATGA GAGTG
CGTACGTACCGAGTCGATGATCGTAGTACGATCGATG・・・・・・



{A,G},{T,C}→0,1

001010000101010010000100101000101001010010101010110010100100
10010101010000101110101101001010101010101010・・・・・・



従来の方法だと、、

・・・・・・・・010001010101010010010・・・・・・・・

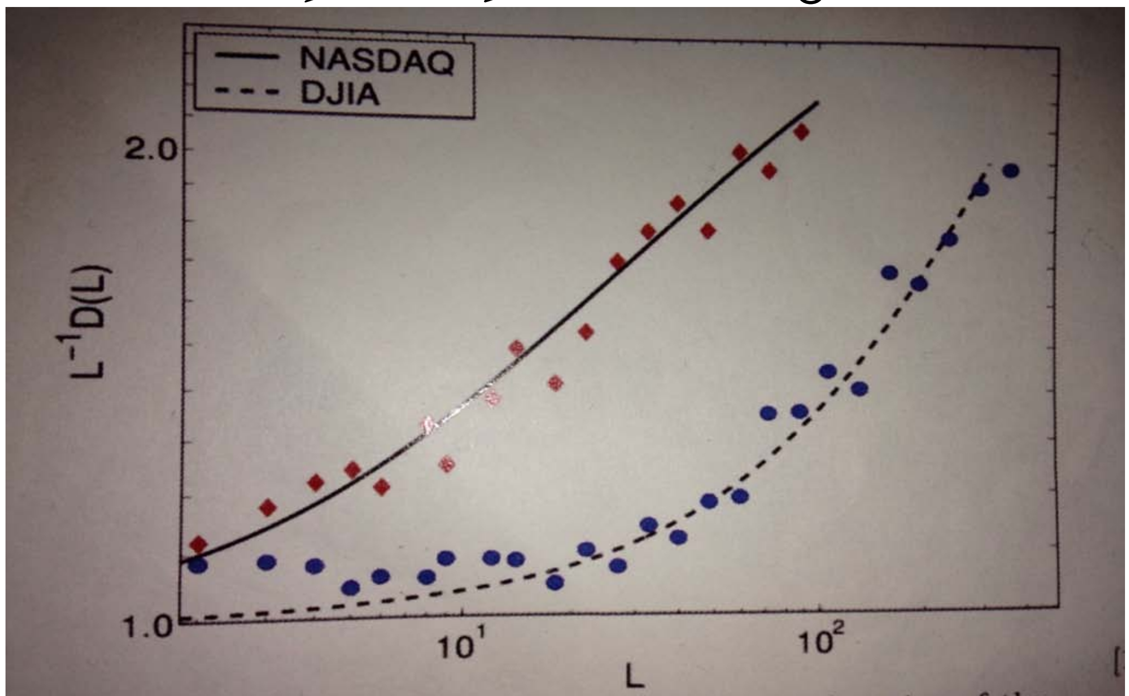
$\longleftrightarrow$

L

1の個数（N_1）の分散V_L（N_1）→ベキ性を持てばいい

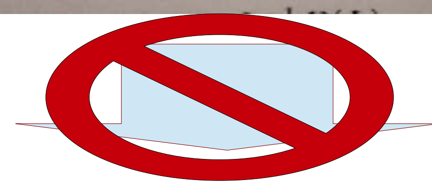
しかし、従来の方法だとバラつきが多いように見える。

縦軸Lにおける分散



横軸　log（L）

図:株のべき性のグラフ



怪しい

$$V_L(N_1) \propto L^{\alpha}$$

相関関数のスケーリング解析

従来の方法と相関関数の関係

$$X(s) \in \{0,1\}, s \in \{1,2,3,\cdots,T\}$$

$$V_L(N_1) \propto L^{\alpha} \quad \xleftrightarrow[\alpha = p+1]{} \quad \mathrm{Cor}(t) \propto t^{p-1}$$

共分散

$$\mathrm{Cor}(t) \equiv \frac{1}{T-t} \sum_{s=1}^{T-t} X(s)X(s+t) - \langle X(s) \rangle \langle X(s+t) \rangle$$

$$\langle X(s) \rangle \equiv \frac{1}{T-t} \sum_{s=1}^{T-t} X(s) \qquad \langle X(s+t) \rangle \equiv \frac{1}{T-t} \sum_{s=1}^{T-t} X(s+t)$$

相関関数のスケーリング解析

DNA配列の相関関数$\mathrm{Cor}(t) \propto t^{p-1}$　として

積分相関時間（緩和時間）

$$\tau(t) \equiv \frac{M_0(t)}{\mathrm{Cor}(0)}$$

2次モーメント相関時間（相関時間）

$$\xi(t) \equiv \sqrt{\frac{M_2(t)}{M_0(t)}}$$

n次モーメント

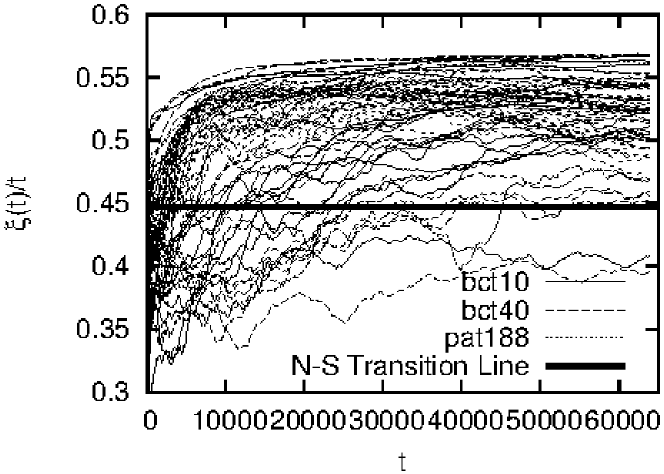
$$M_n(t) \equiv \sum_{s=0}^{t-1} \mathrm{Cor}(s) s^n$$

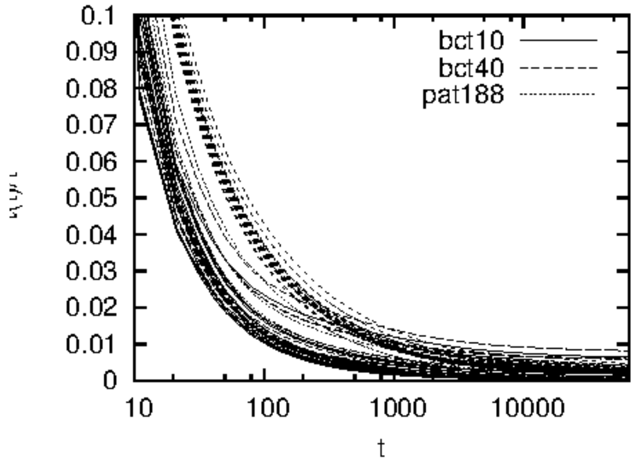
もし、スケール不変性なら

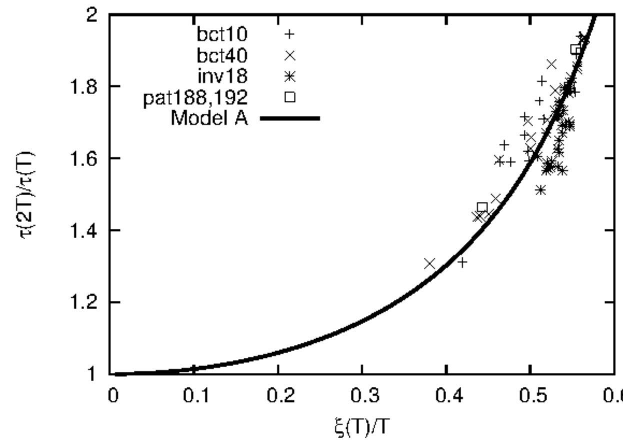
$$\lim_{t\to\infty} \xi(t)/t = \sqrt{\frac{p}{p+2}} \qquad \lim_{t\to\infty} \tau(t)/t = 0 \qquad \frac{\tau(2t)}{\tau(t)} = 2^p$$

となることが分かっている。

解析結果







結論

（1）相関関数のスケーリング解析により、DNAはスケール不変性をもつものが存在することが検証でき、本手法の有効性が確かめられた。

（2）相関時間ξ/tの極限値から、相関が非常につよく、スケール不変なものはほとんどスーパー拡散相にあることが分かった。

（3）解析したものには明らかにスケール不変でないものも多い。スケール不変性の生物学的な意味は不明である。

- 終わり

0, 1配列でのスケール不変性の検証方法

$$X(t) \in \{0,1\}, t \in \{1,\cdots,T\}$$

$V_L(N_1)$

$V_L(N_1) \propto L$

$$V_L(N_1) \propto L^{\alpha} \qquad \longleftrightarrow \qquad \mathrm{Cor}(t) \propto t^{p-1}$$

$$\alpha = p+1$$

$V_L(N_1) \propto L$

$V_L(N_1) \propto L^{\alpha}$

$$\lim_{t\to\infty} \xi(t)/t = \sqrt{\frac{p}{p+2}} \qquad \lim_{t\to\infty} \tau(t)/t = 0 \qquad \frac{\tau(2t)}{\tau(t)} = 2^p$$

$$\mathrm{Cor}(t) \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s)X(s+t) - <X(s)><X(s+t)>$$

$$<X(s)> \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s) \qquad <X(s+t)> \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s+t)$$

n次モーメント

$$M_n(t) \equiv \sum_{s=0}^{t-1} \mathrm{Cor}(s)s^n$$

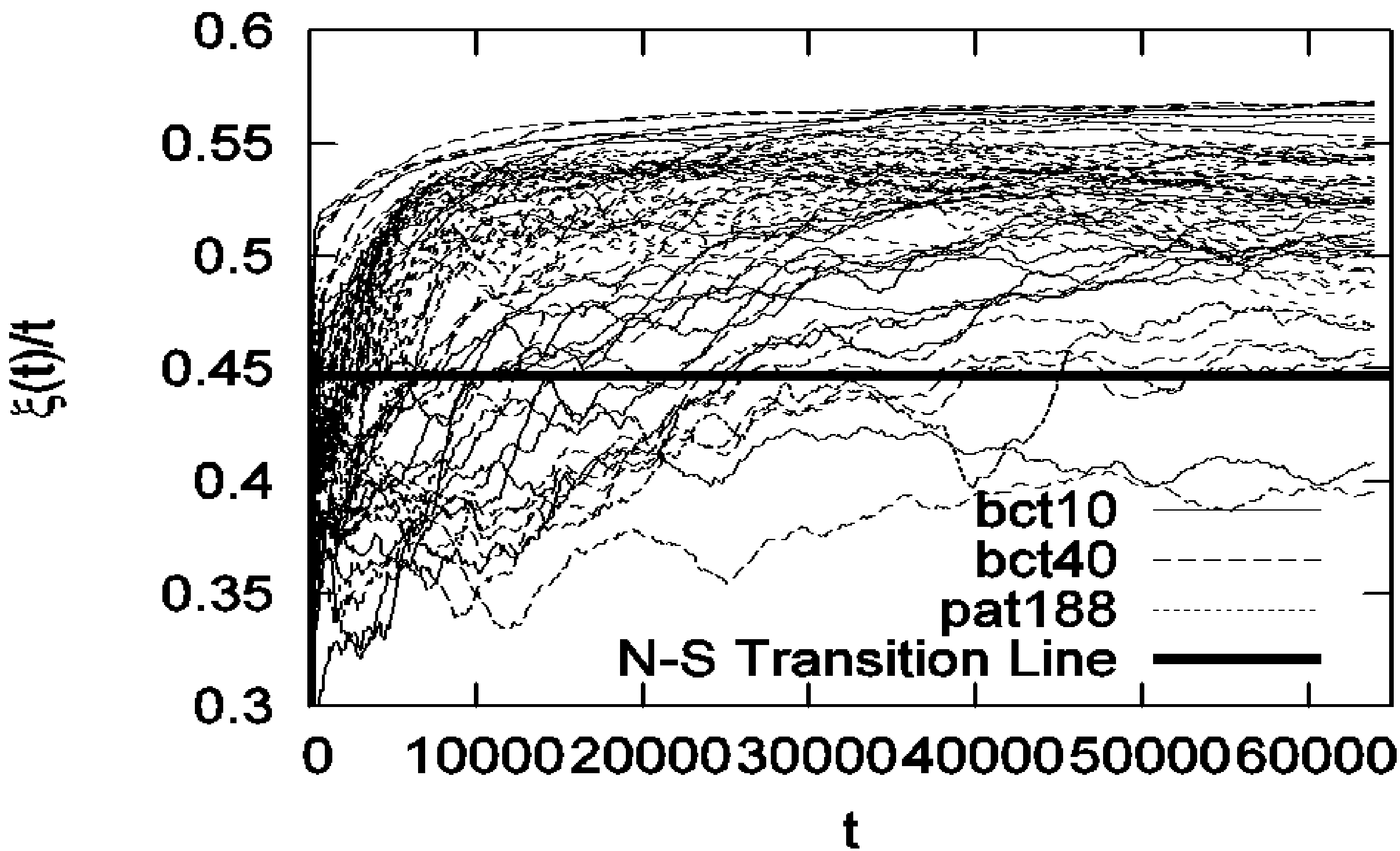
積分相関時間（緩和時間）

$$\tau(t) \equiv \frac{M_0(t)}{\mathrm{Cor}(0)}$$

2次モーメント相関時間（相関時間）

$$\xi(t) \equiv \sqrt{\frac{M_2(t)}{M_0(t)}}$$

結果1
(相関時間ξ(t))/t



0.6
0.55
0.5
0.45
0.4
0.35
0.3
ξ(t)/t
0
10000
20000
30000
40000
50000
60000
t
bct10
bct40
pat188
N-S Transition Line

0, 1配列でのスケール不変性の検証方法

$$X(t) \in \{0, 1\}, t \in \{1, \cdots, T\}$$

$V_L(N_1)$

$V_L(N_1) \propto L$

$$V_L(N_1) \propto L^{\alpha} \qquad \mathrm{Cor}(t) \propto t^{p-1}$$

$$\alpha = p + 1$$

$V_L(N_1) \propto L$

$V_L(N_1) \propto L^{\alpha}$

$$\lim_{t\to\infty} \xi(t)/t = \sqrt{\frac{p}{p+2}} \qquad \lim_{t\to\infty} \tau(t)/t = 0 \qquad \frac{\tau(2t)}{\tau(t)} = 2^p$$

$$\mathrm{Cor}(t) \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s)X(s+t) - < X(s) >< X(s+t) >$$

$$< X(s) > \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s) \qquad < X(s+t) > \equiv \frac{1}{T-t}\sum_{s=1}^{T-t} X(s+t)$$

n次モーメント

$$M_n(t) \equiv \sum_{s=0}^{t-1} \mathrm{Cor}(s) s^n$$

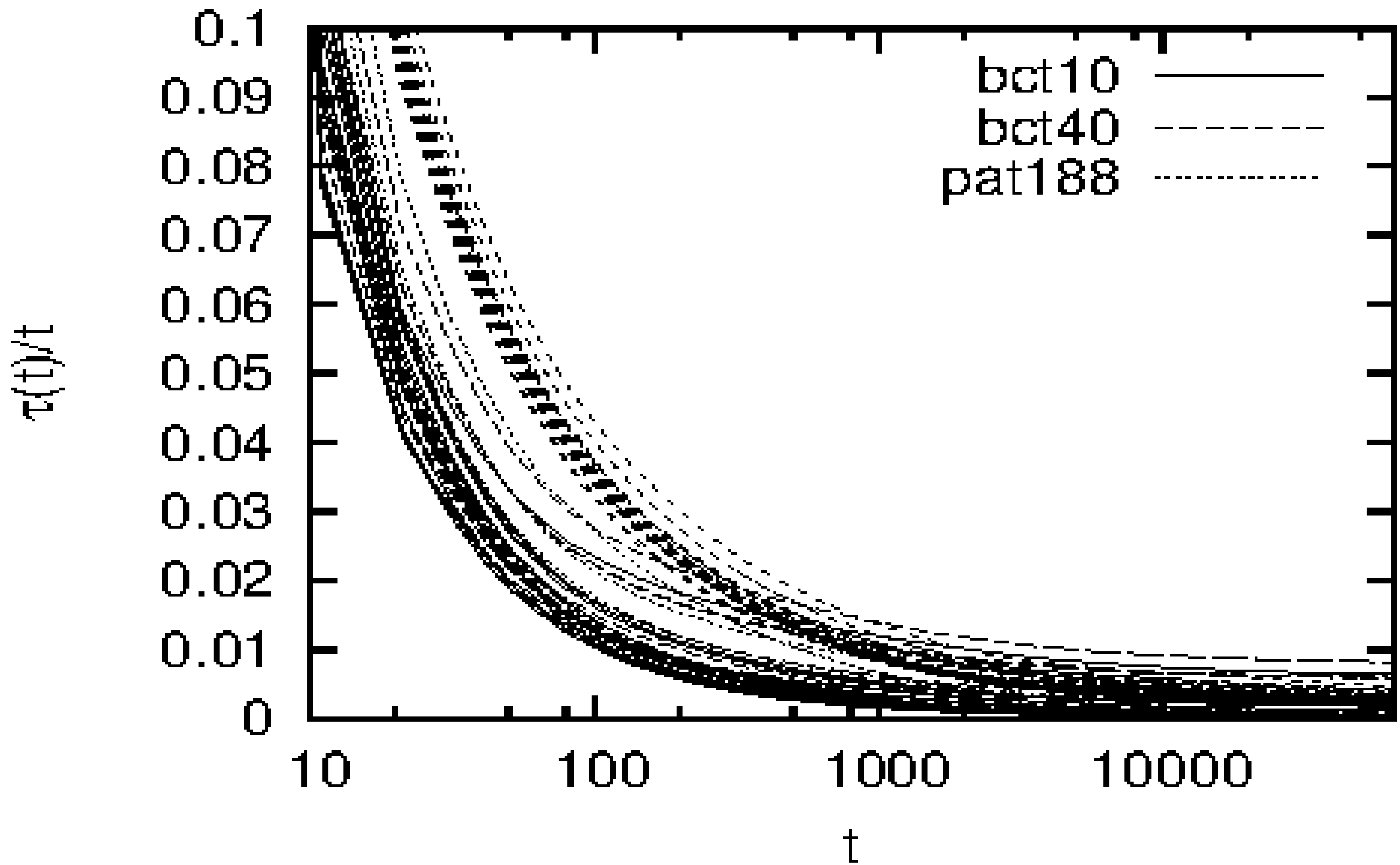
積分相関時間（緩和時間）

$$\tau(t) \equiv \frac{M_0(t)}{\mathrm{Cor}(0)}$$

2次モーメント相関時間（相関時間）

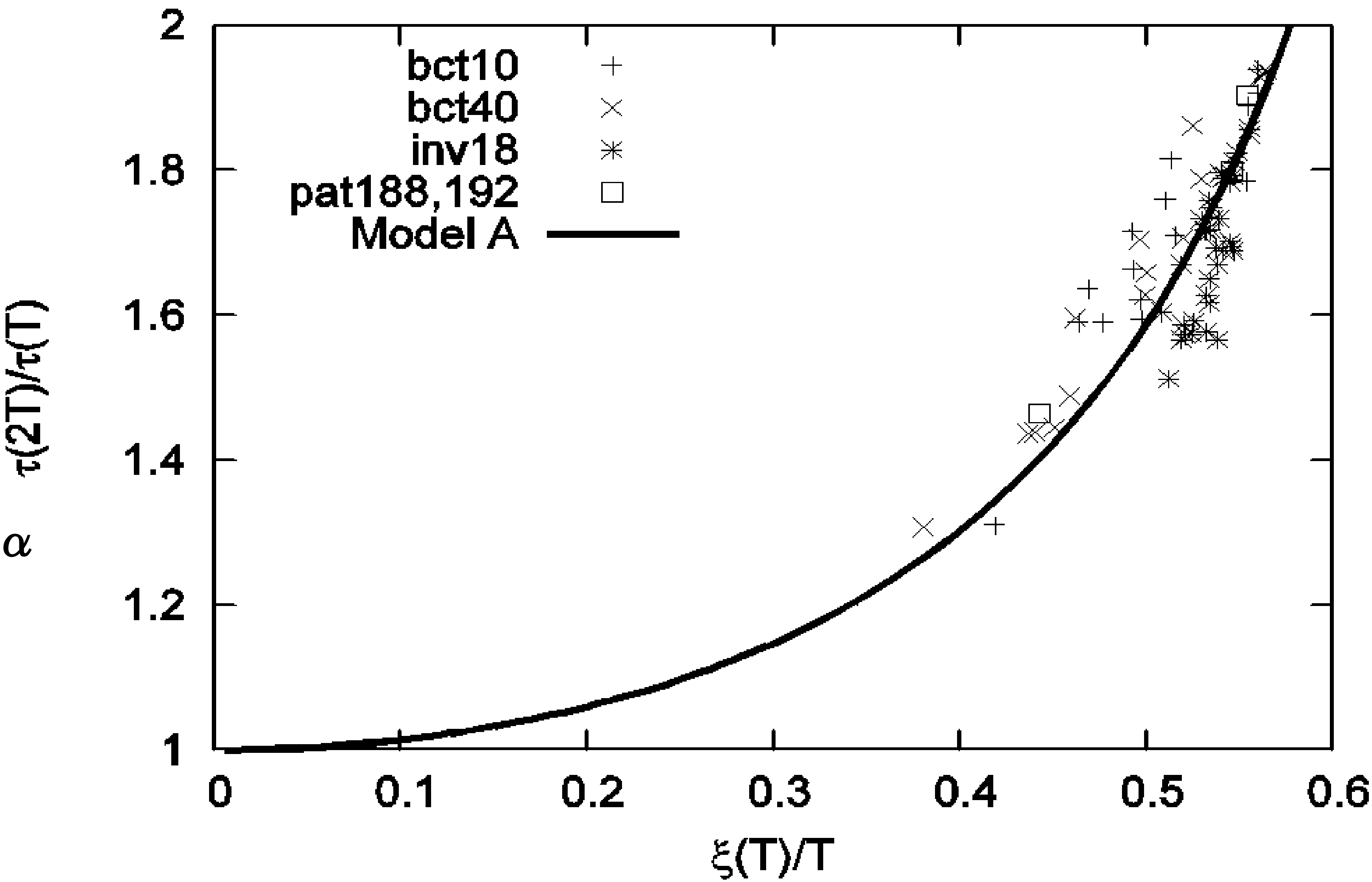
$$\xi(t) \equiv \sqrt{\frac{M_2(t)}{M_0(t)}}$$

結果2
(緩和時間 τ (t)/t)


bct10
bct40
pat188
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
10
100
1000
10000
τ(t)/t
t

結果3

2^α



bct10
bct40
inv18
pat188,192
Model A
τ(2T)/τ(T)
ξ(T)/T
2
1.8
1.6
1.4
1.2
1
0
0.1
0.2
0.3
0.4
0.5
0.6