統数研/CIGS　経済物理学とその周辺　2016/8/29-30

アローヘッド市場における株価変動の統計分布

～Lévy分布再考～

田中美栄子

mieko.tanakay@gmail.com

統計数理研究所/

（元）鳥取大学工学部知能情報工学科

概要

- 東証において，アローヘッドシステムという1ミリ秒間隔で株式売買を行える仕組みが稼働を始めたのが2010年1月であり，2015年9月にはその第2ステップとして0.5ミリ秒間隔にまで短縮された
- ．この高速取引によって価格変動の性質がどのように変わったのかを知るため，5秒足にサンプリングされた2013年のデータから分布関数を求めたところ，1990年頃の米国株価変動の解析結果により求められたα=1.4のLévy分布と同じであるという結果を得た.
- 更に1分足にサンプリングされた2015年のデータからも同様の結果を得た.

講演歴

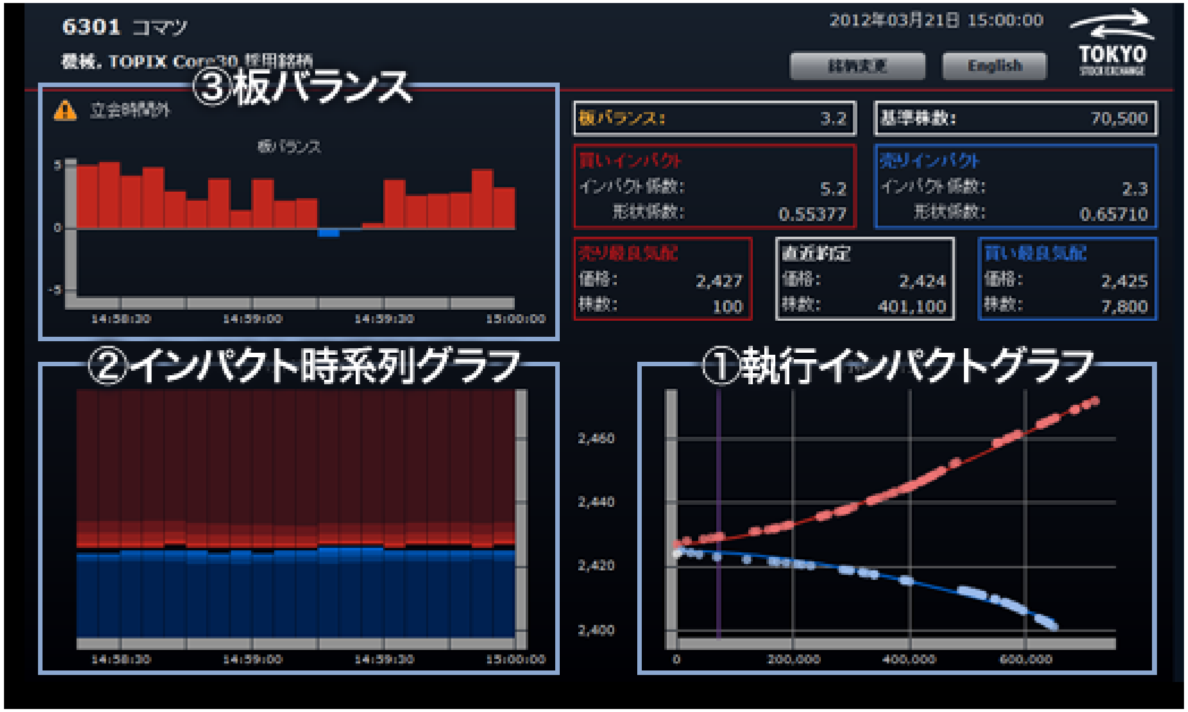
- 統数研研究会・立川　2016・1・9
- 日本物理学会・東北学院大　2016・3・20
- 電子情報通信学会・複雑コミュニケーション研究会・機械振興会館　2016・6・13
- APEC－SSS2016　東大本郷　2016・8・24-26
- ISM_CIGS　CIGS　2016・8・29-30

Arrowhead取引
について

wiki

- 2010年1月4日以降、東京証券取引所はarrowheadと呼ぶ千分の1秒単位での取引を可能するシステムに移行した。更に2015年9月24日以降，2千分の1秒単位での取引を可能にするarrowhead 2となって、超高速化が進行し，海外から多くのファンドが参加してシステムトレードを行うようになった結果，従来の取引環境とは異なる投資状況が生じている．このようなシステム下で行われている市場の価格変動の性質もおそらく以前とは大きく異なるものと考えられる.

東証 Market Impact View



1.執行インパクトグラフ
板の注文情報をもとに統計的手法を用いて処理し、執行インパクトを可視化

2.インパクト時系列グラフ
インパクトグラフをもとに、一定株数ごとに色の濃淡を付与し、時系列で可視化したグラフ

3.板バランス
注文状況から板のバランス（売買のバイアス）を表現

2012年3月30日 … この度、東証が開発した『東証 Market Impact View』は、目の前の注文状況から相場を分析するまさに"生放送" 型のテクニカル分析だという。目で追えないフル板状況を直感 的に把握することができる.「このテクニカル分析は、約定前の注文

検証データ

東証アローヘッド市場　5秒足

2013/4/1/9:00:05　～　2013/12/25/15:00:00

1銘柄:640800　データ

100銘柄

（4年生が毎日ダウンロード→卒論用）

機械学習による短期予測を試す

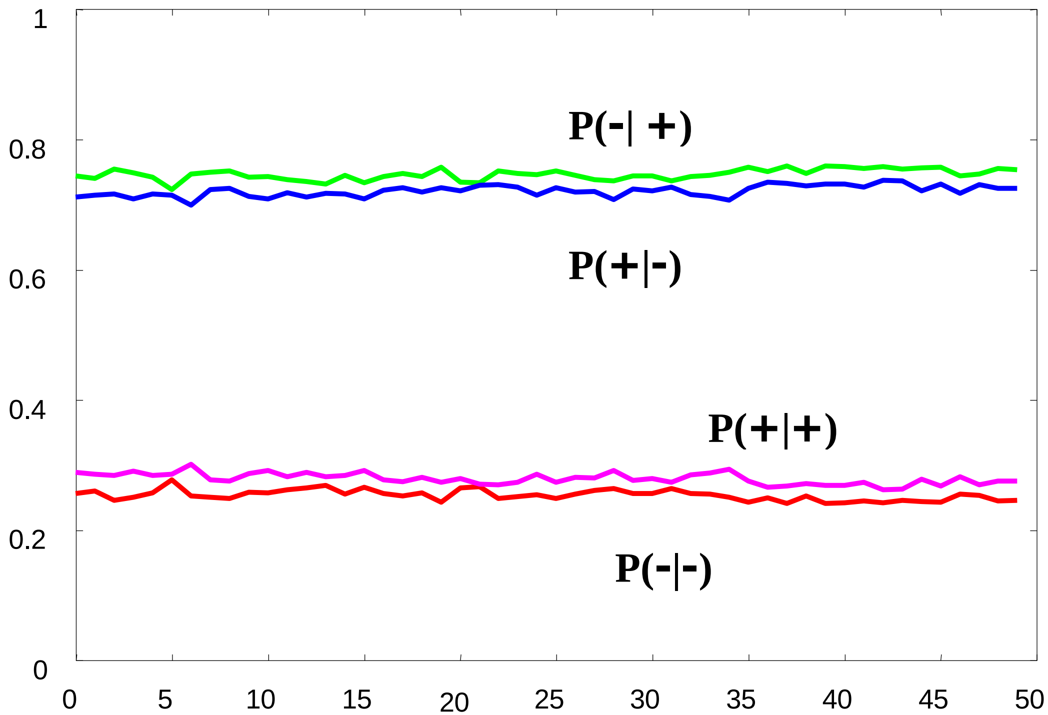
- 以前の研究結果を応用して安直に卒論を、、

- 過去に成功率の高かったパターンを学習し
- 状況に応じて適用する

- 条件付き確率を学習し、適用する
- 安定な確率を求められるのか？

- arrowhead以前のtick価格変動予測

事実：Tick価格の細かい動きの確率構造が長期にわたって驚くほど安定（但し2000tickの平均）



1tick先は確率7割強で反発方向に行く

P(－|＋)=0.7

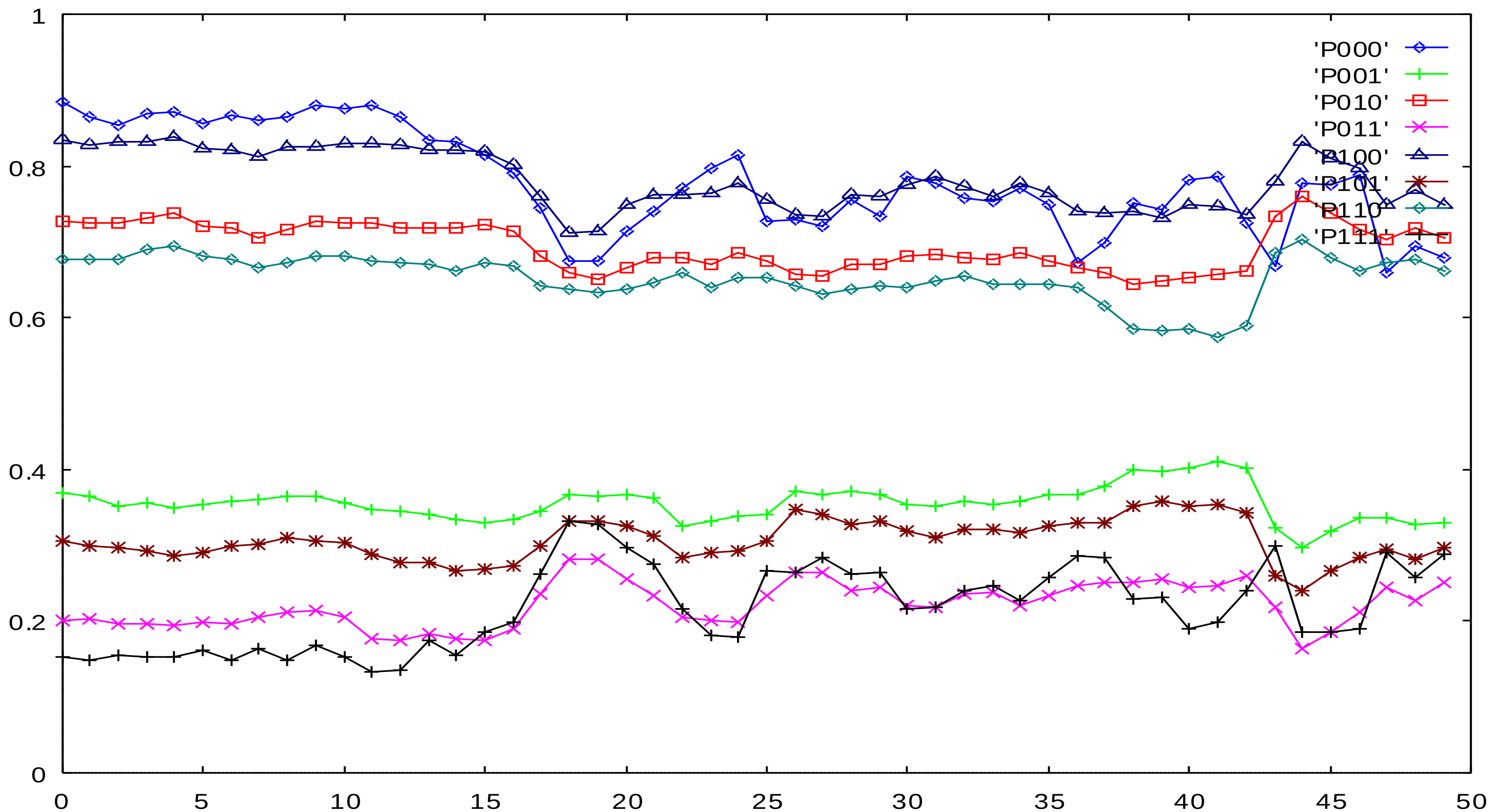
$$\bar{P} = \begin{pmatrix} 0.25 & \underline{0.73} \\ \underline{0.75} & 0.27 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.008 & 0.008 \\ 0.008 & 0.008 \end{pmatrix}$$

為替データにおける遷移確率の推移

(期間：95.01/02～95.09/05)

Depth=3
P(1|000),P(1|001),P(1|010), P(1|011), P(1|100),P(1|101),P(1|110), P(1|111)



予測戦略の学習

P(＋|ー)=P(ー|＋)=0.7

P(＋|＋)=P(ー|ー)=0.3

- 1tick前のパターンに基づいて1tick先を予測

P(ー|＋＋)=?

P(ー＋＋＋)=?

- 数ticks前からのパターンに基づいて1tick先を予測する戦略をGAで自動学習させる

（**前提条件を長くする**）

Tick価格には癖がある→　予測

- パターンの抽出　→　予測（ T−tick先）
- 様々なパターンの存在：　記憶長に依存
- とりあえずは短期予測を行う｛　ー，0，＋｝
- 戦略はオンラインで選ぶ（状況ごとに変化）
- 遺伝アルゴリズムによる学習
- 戦略の切り替えが重要（人工知能）

- パターン学習＋戦略の切替 = 価格変動規則

↓

予測の（物理的？）科学

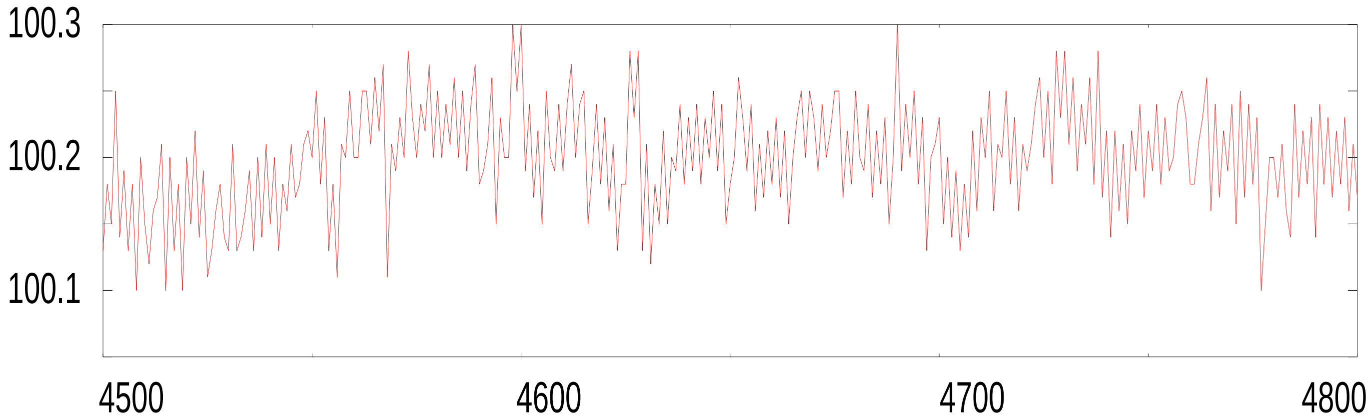
Data: Tick-wise Stock Price

NYSE, 8 symbols, 1993/1/1～1993/12/31

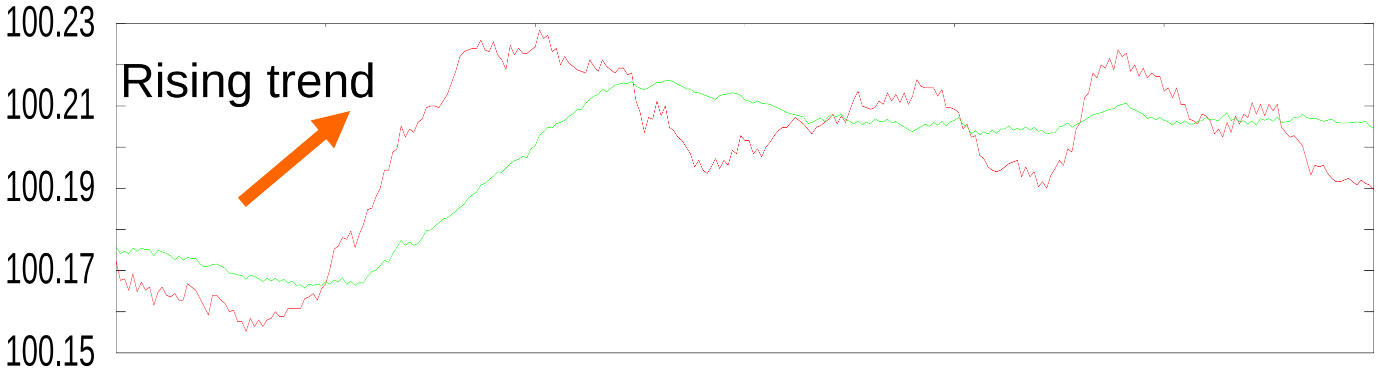
Stock symbol	Business type	Data size (ticks)	Tick interval (s/ticks)	10-ticks (minute)
BBY	retail	54821	109	18
SMRT	retail	12525	473	78
APC	oil	23685	253	42
BP	oil	73562	83	14
CA	computer	65051	92	15
IBM	computer	455233	14	2
F	automobile	194561	32	5
GM	automobile	277241	23	4

Example of trend indicator : MA

Tick price



MA
(Trend Indicators)



基本手法：予測遺伝子(P)の設計

全ての履歴に対応した
予測値を持つ遺伝子を設計

➡ 予測遺伝子が戦略となる

$P = p_1 p_2 \cdots p_{3^N}$

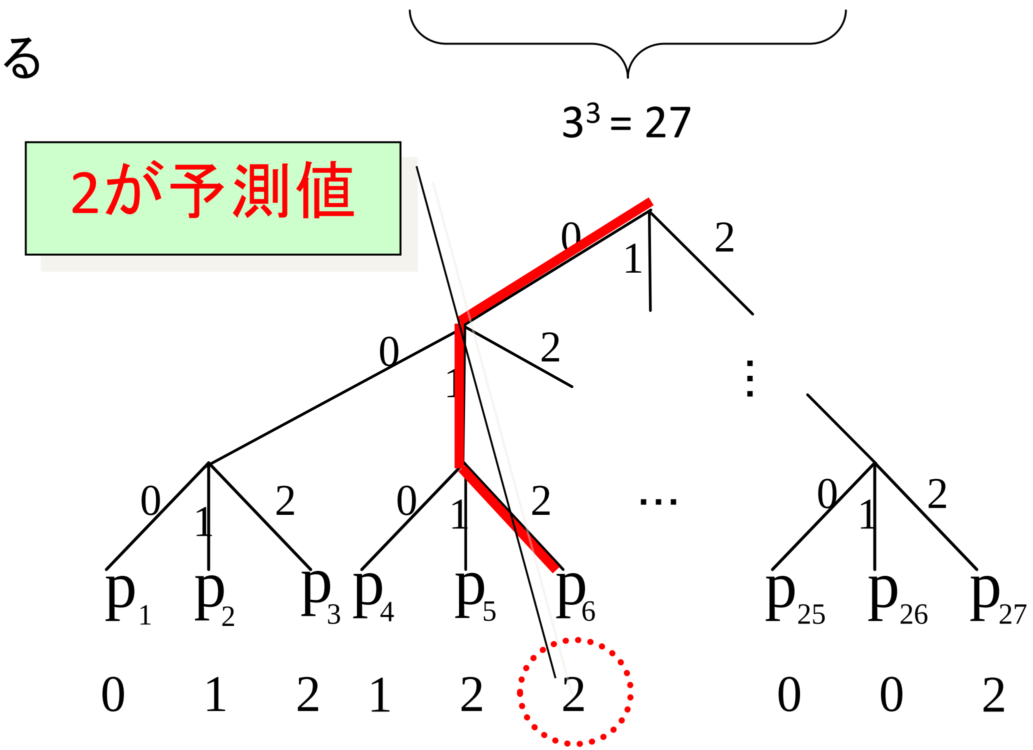
$p_j \in \{0,1,2\}$

遺伝子の長さ $= 3^N$

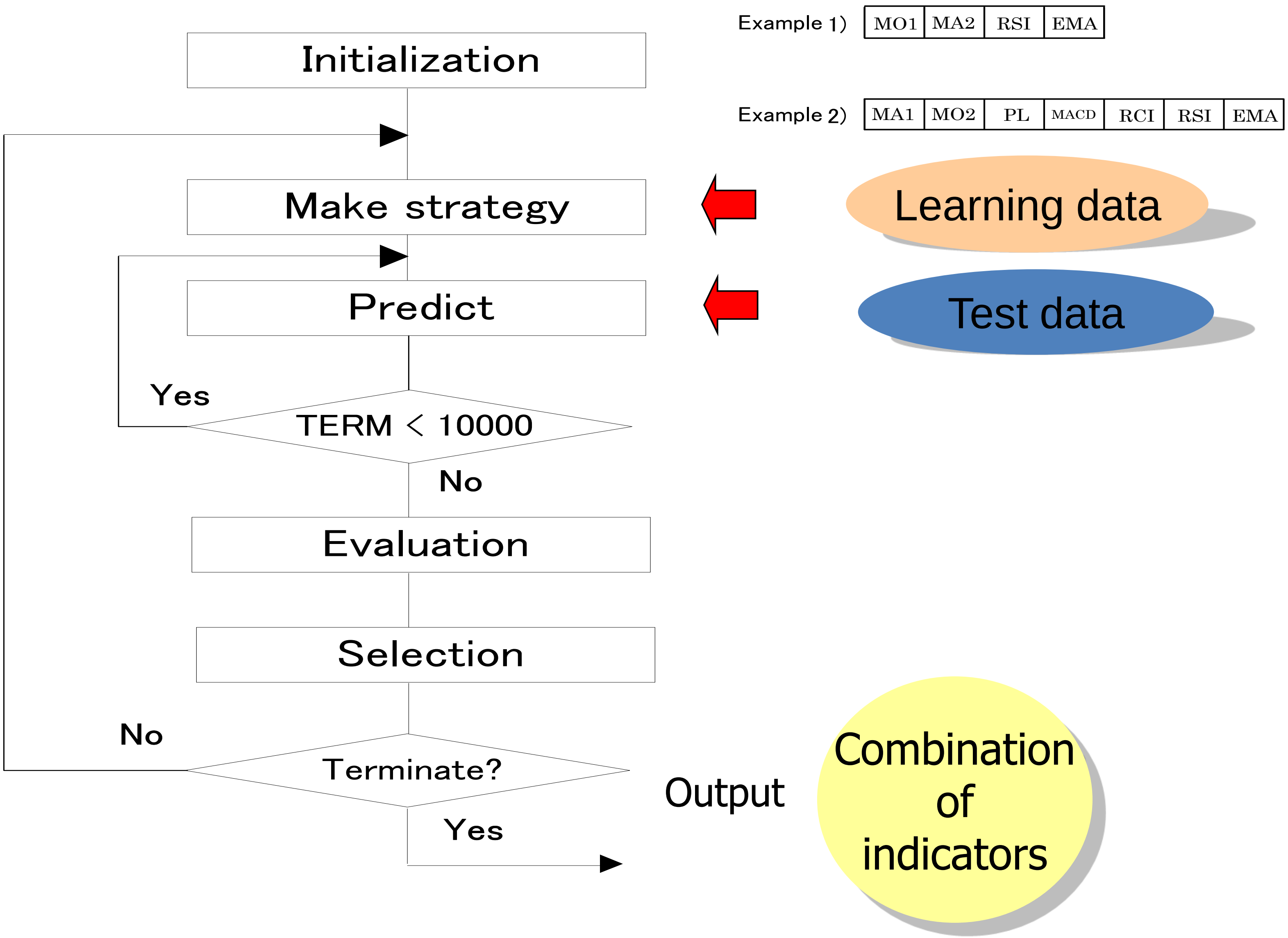
遺伝子の種類 $= 3^{3^N}$

Ex). N = 3,H=012の場合

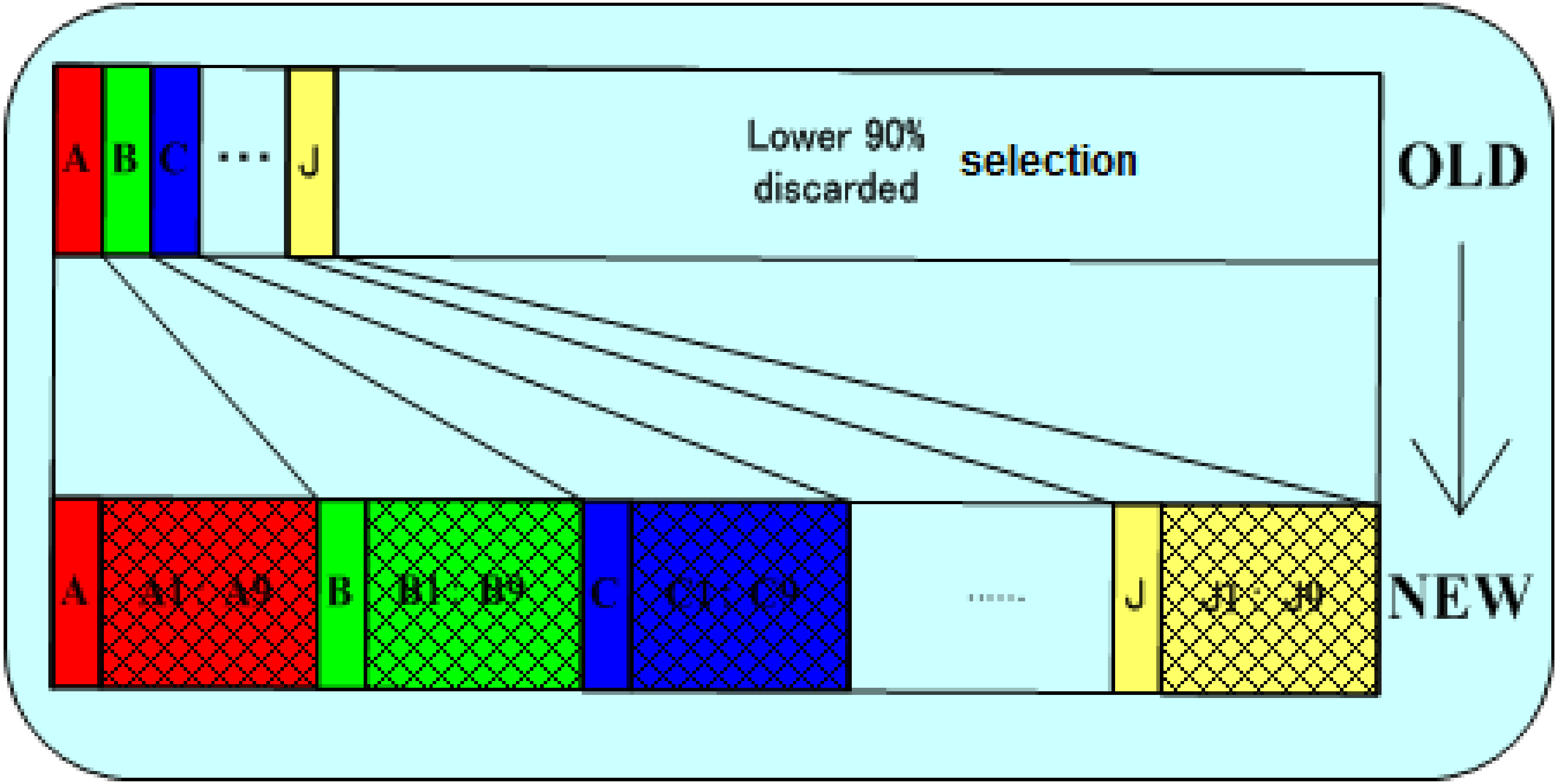
P = 012122・・・002



Flow of selecting indicators by GA

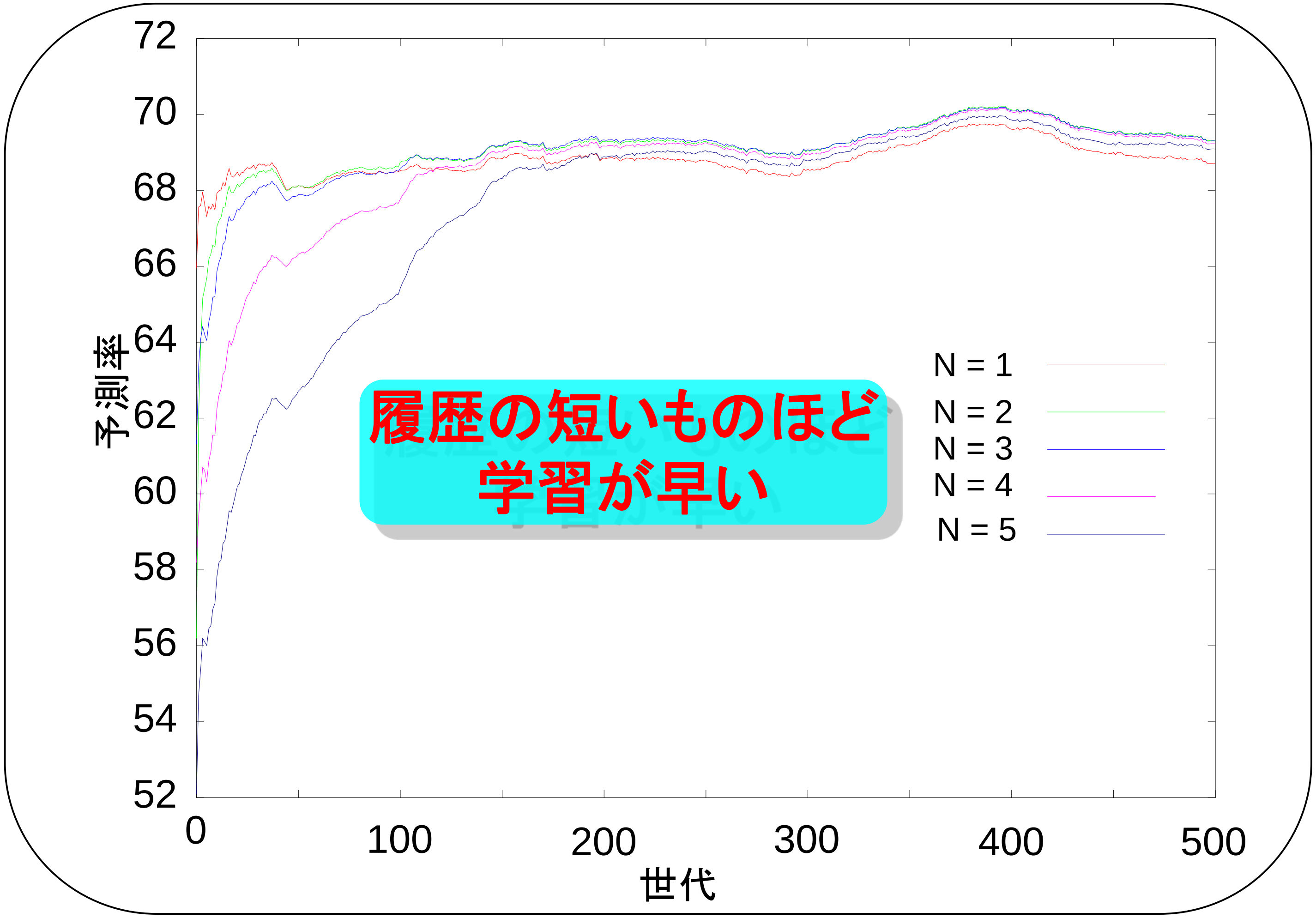


Generation change



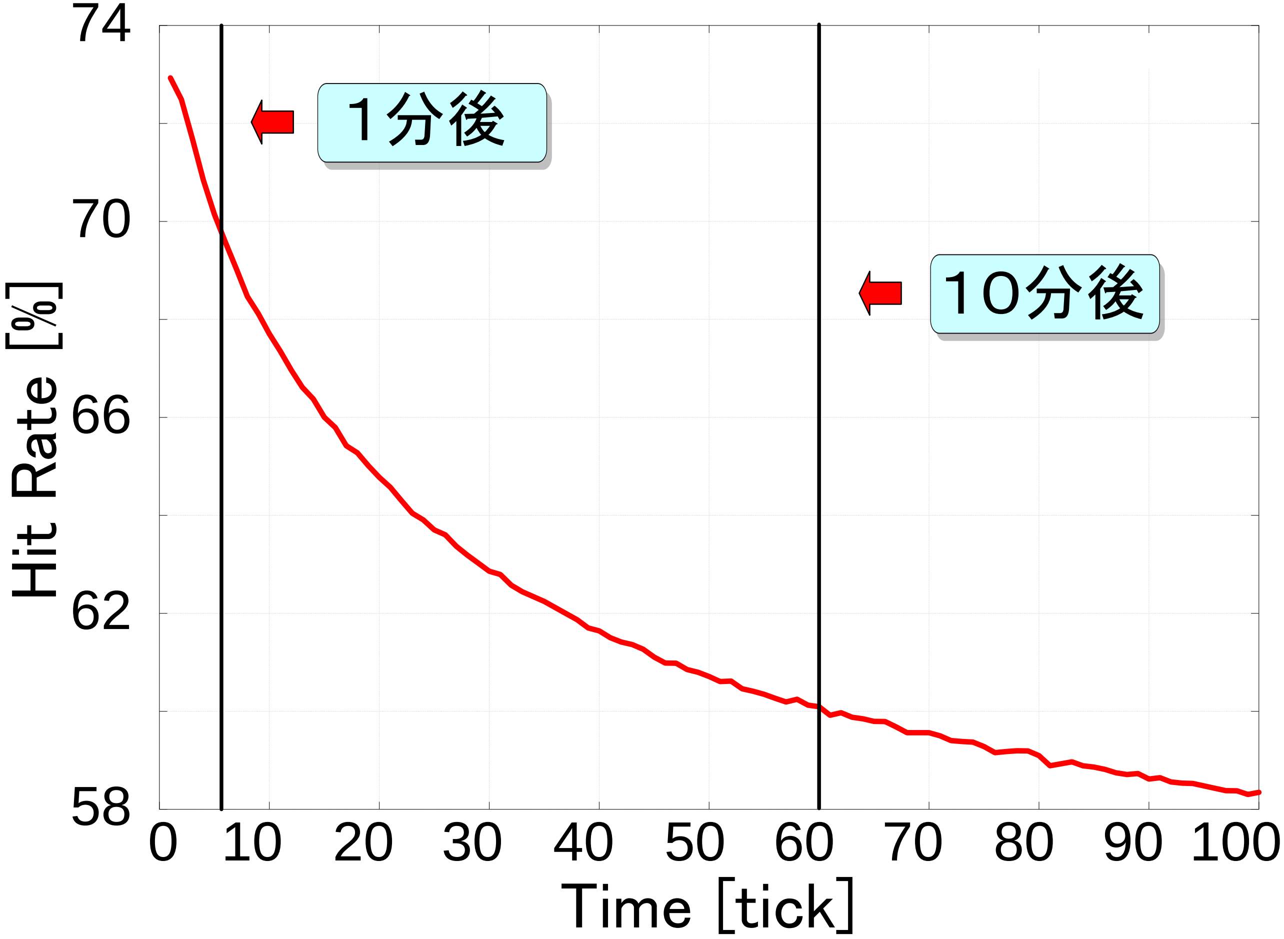
遺伝アルゴリズムによる学習に基づく予測結果

はじめの500世代までの予測率の推移



Hit Rate vs. Prediction Term

$/¥ rate(2000)

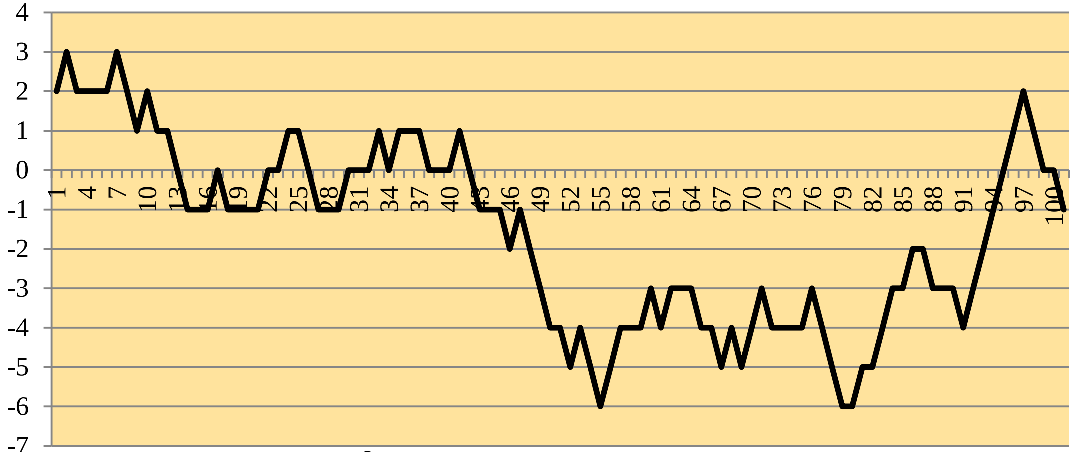


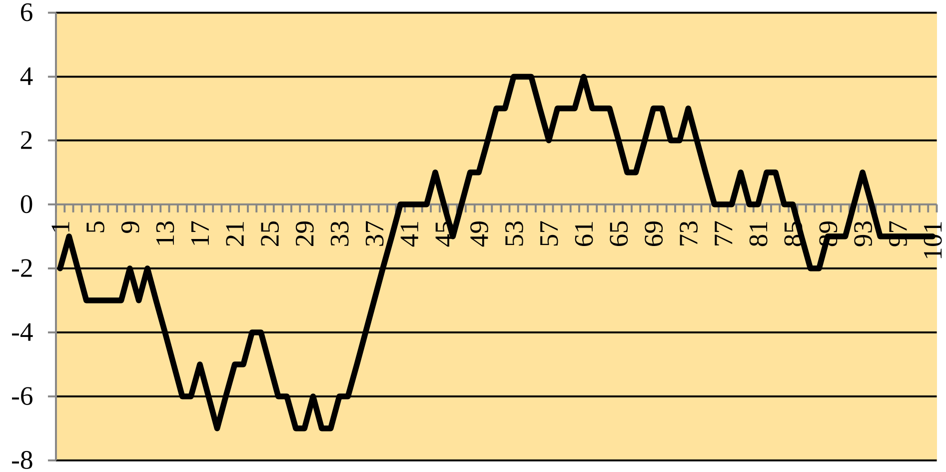
予測結果（上昇or下降）

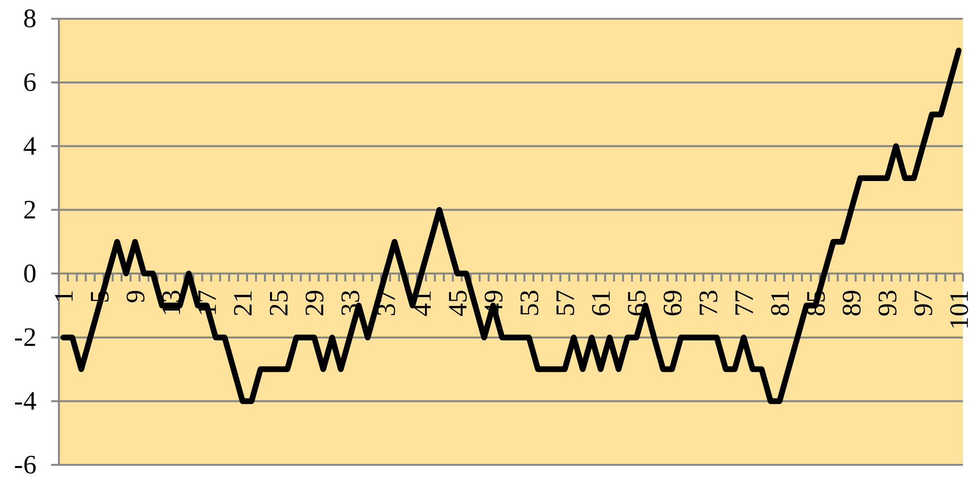
- 数十秒～数分程度のtick価格変動は予測可
- Arrowheadデータ（5秒毎のサンプルデータ）に適用しても良い結果が得られなかった、、
（卒論失敗）
- 何かが違う！

価格変動の原理を知る必要

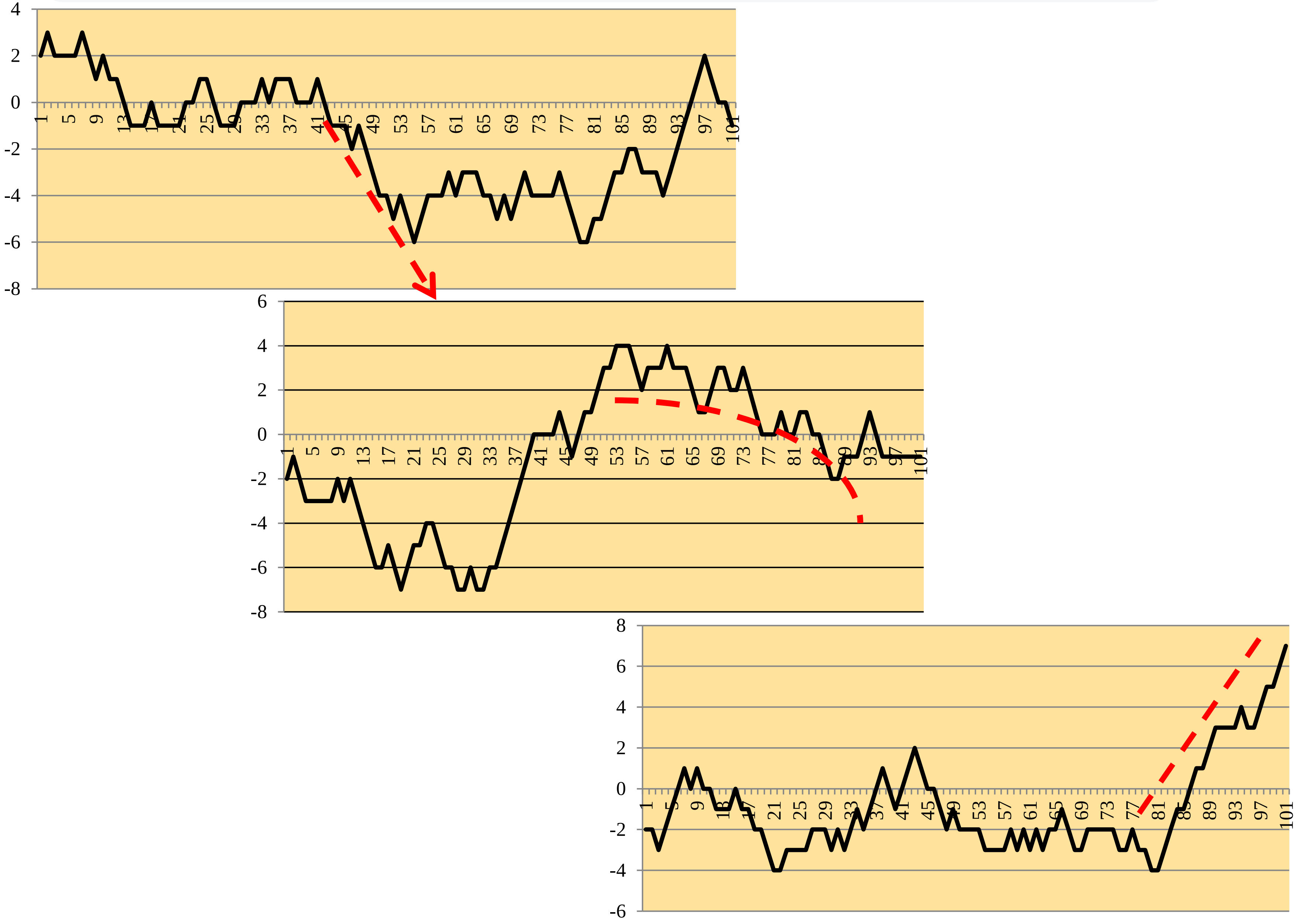
如何なる動きなのか？


4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7


6
4
2
0
-2
-4
-6
-8

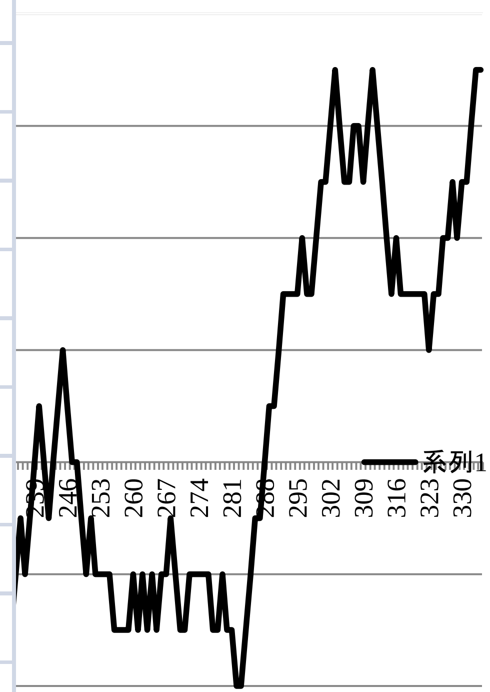

8
6
4
2
0
-2
-4
-6

グラフの傾向を見る線 ：どのようにも見える



乱数発生器による乱数

rand()	rand() %3	rand()%3-1	accum sum
21293	2	1	1
29441	2	1	2
19484	2	1	3
12759	0	-1	2
15400	1	0	2
6211	1	0	2
4144	1	0	2
15335	2	1	3
22704	0	-1	2
32520	0	-1	1
23789	2	1	2
32121	0	-1	1
21913	1	0	1
23571	0	-1	0
12369	0	-1	-1
2770	1	0	-1
1594	1	0	-1



価格変動の科学

フランスの数学者Bachelier(バシュリエ)が博士論文をEcole Normaleに提出（アンリ・ポアンカレは却下）

1900

価格変動はrandom walk

Random walk の効用

Black－Sholes formula
１９９７ Nobel Prize

デリバティブ価格を
σ（Volatility）から
決定できる

Myron S. Scholes(1941-)

ブラック･ショールズのオプション価格式

$$C(S,t) = \underline{N(d_1)S} - \underline{N(d_2)K}e^{-r(T-t)}$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right]$$

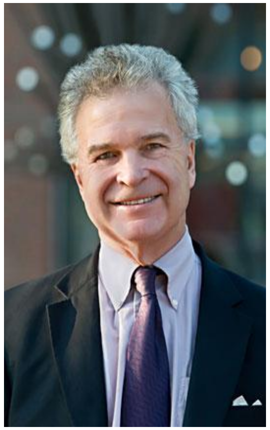
$$d_2 = d_1 - \sigma\sqrt{T-t}$$

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz$$

正規分布

C：コールオプション価格　q：原資産利回り　t：期間
S：原資産価格
N（d）：標準正規分布の累積確率密度関数　r：安全利子率（非危険利子率）
k：行使価格
d：累積密度関数N（）の変数　σ：ボラティリティ（予想変動率）

Harry Eugene Stanley　1941-

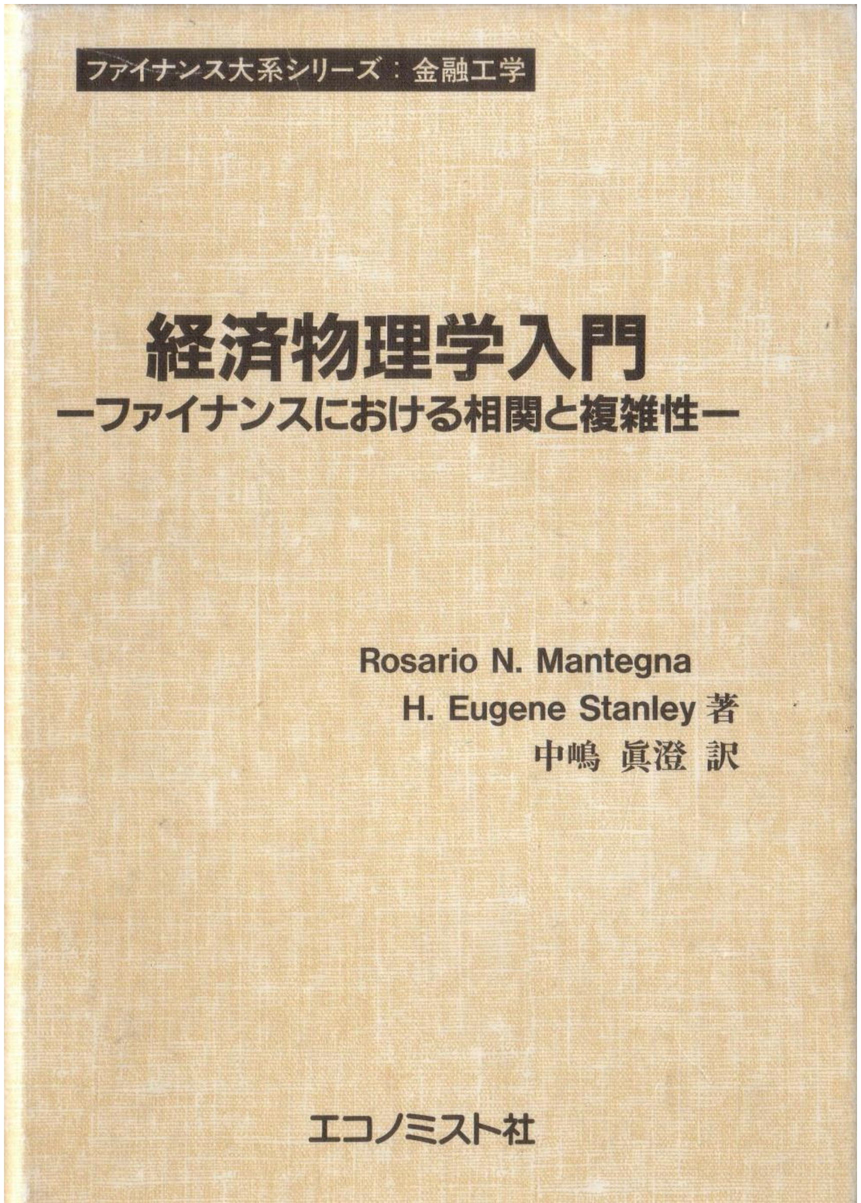


ジーン・スタンレー
経済物理学の父

Known for
Econophysics,
Statistical physics,
Alzheimer's

2004 Boltzmann Medal
2008 Julius Edgar Lilienfeld Prize
Teresiana Medal
Distinguished Teaching Scholar
Director's Award
Nicholson Medal
Memory Ride Award for Alzheimer Research

もうちょっとでノーベル賞


ファイナンス大系シリーズ：金融工学
経済物理学入門
ーファイナンスにおける相関と複雑性ー
Rosario N. Mantegna
H. Eugene Stanley 著
中嶋 眞澄 訳
エコノミスト社


ファイナンス・ライブラリー 6
金融リスクの理論
——経済物理からのアプローチ——
J.-P. ブショー／M. ポッター 著
森平爽一郎 監修　森谷博之／熊谷善彰 訳
朝倉書店

価格の実データは、
正規分布ではない、
という結果！
Mantegna-Stanley著
1995　Nature

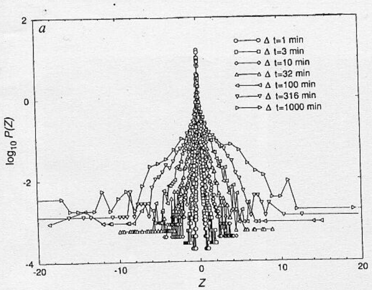
Scaling behaviour in the dynamics of an economic index

Rosario N. Mantegna & H. Eugene Stanley

Center for Polymer Studies and Department of Physics,
Boston University, Boston, Massachusetts 02215, USA

THE large-scale dynamical properties of some physical systems depend on the dynamical evolution of a large number of nonlinearly coupled subsystems. Examples include systems that exhibit self-organized criticality[1] and turbulence[2,3]. Such systems tend to exhibit spatial and temporal scaling behaviour—power-law behaviour of a particular observable. Scaling is found in a wide range of systems, from geophysical[4] to biological[5]. Here we explore the possibility that scaling phenomena occur in economic systems—especially when the economic system is one subject to precise rules,

NATURE · VOL 376 · 6 JULY 1995



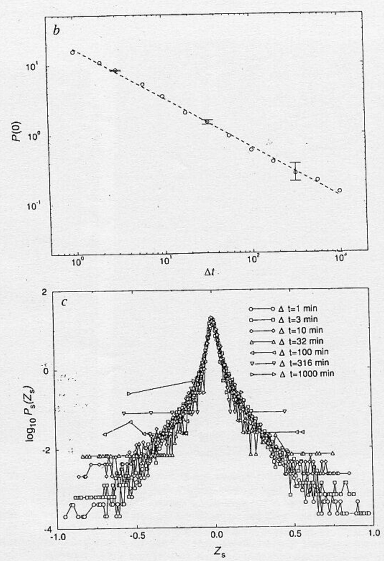


FIG. 1 *a*, Probability distributions $P(Z)$ of the S&P 500 index variations $Z(t)$ observed at time intervals Δt, which range from 1 to 1,000 min. By increasing Δt, a spreading of the probability distribution characteristic of a random walk is observed. *b*, Probability of return $P(0)$ of $Z(t)$ as a function of the time sampling intervals Δt. A scaling power-law behaviour is observed for time intervals spanning three orders of magnitude. The slope of the best-fit straight line is -0.712 ± 0.025. By using equation (2), from this value we obtain the index $\alpha = 1.40 \pm 0.05$ characterizing the Lévy distribution. Typical estimated errors are shown. *c*, Scaled plot of the probability distributions shown in *a*. All the data collapse on the $\Delta t = 1$ min distribution by using the scaling transformations of equations (3a) and (3b), with $\alpha = 1.40$. The points outside the average behaviour define the noise level of that specific distribution.

as is the case in financial markets[6-8]. Specifically, we show that the scaling of the probability distribution of a particular economic index—the Standard & Poor's 500—can be described by a non-gaussian process with dynamics that, for the central part of the distribution, correspond to that predicted for a Lévy stable process[9-11]. Scaling behaviour is observed for time intervals span-

processes have quite similar statistical properties in the high-frequency regime.

We have undertaken a statistical study of timescales as short as 1 min, a value close to the minimum time needed to perform a transaction in the market. Specifically, we investigate the dynamics of a price index of one of the largest financial markets

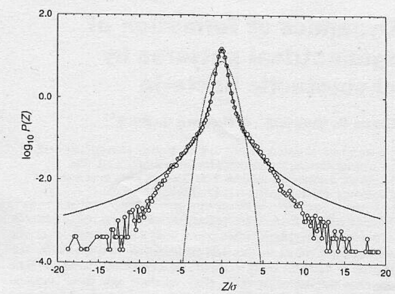


FIG. 2 Comparison of the $\Delta t = 1$ min probability distribution with the symmetrical Lévy stable distribution of index $\alpha = 1.40$ and scale factor $\gamma = 0.00375$ (solid line). The scale factor γ is obtained from equation (2) by using the experimental values of α and $P(0)$. The dotted line shows the gaussian distribution with standard deviation σ equal to the experimental value 0.0508. The variations of price are normalized to this value. Approximately exponential deviations from the Lévy stable profile are observed for $|Z|/\sigma \gtrsim 6$.

1,000 min. The number of data in each set decreases from the maximum value of 493,545 ($\Delta t = 1$ min) to the minimum value of 562 ($\Delta t = 1{,}000$ min).

Figure 1*a* is a semilogarithmic plot of $P(Z)$ obtained for seven different values of Δt. As expected for a random process, the distributions are roughly symmetrical and are spreading when Δt increases. We also note that the distributions are leptokurtic, that is, they have wings larger than expected for a normal process. A determination of the parameters characterizing the distributions is difficult if one uses methods that mainly investigate the wings of distributions, especially because larger values of Δt imply a reduced number of data.

Therefore we use a different approach: we study the 'probability of return' $P(Z=0)$ as a function of Δt. With this choice we investigate the point of each probability distribution that is least affected by the noise introduced by the finiteness of the experimental data set. In Fig. 1*b*, we show $P(0)$ versus Δt in a log–log plot. The data are fitted well by a straight line of slope -0.712 ± 0.025. We observe a non-normal scaling behaviour (slope $\neq -0.5$) in an interval of trading time ranging from 1 to 1,000 min. This experimental finding agrees with the theoretical model of a Lévy walk or Lévy flight[9-11]. In fact, if the central region of the distribution is well described by a Lévy stable symmetrical distribution,

$$L_\alpha(Z, \Delta t) \equiv \frac{1}{\pi} \int_0^\infty \exp(-\gamma \Delta t q^\alpha) \cos(qZ) dq \qquad (1)$$

of index α and scale factor γ at $\Delta t = 1$, where $\exp(-\gamma \Delta t |q|^\alpha)$ is

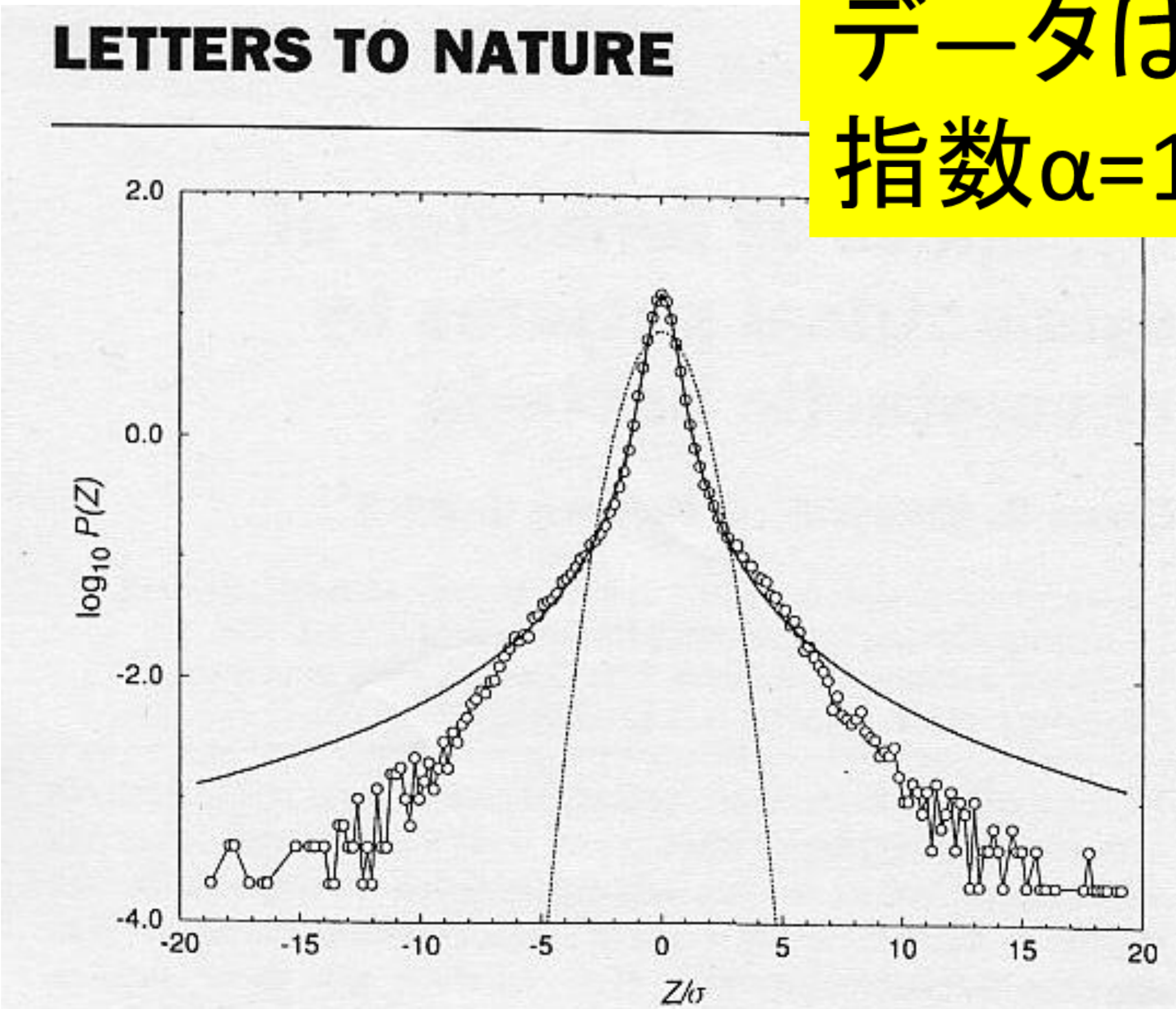
dynamics of the probability distribution $P(Z)$ of the random process over time intervals spanning three orders of magnitude.

In Fig. 2, we compare the probability distribution observed for $\Delta t = 1$ min with the Lévy stable distribution of index $\alpha = 1.40$. Note that the solid line is not simply a 'fit' to the data; rather, the appropriate scale factor $\gamma = 0.00375$ is obtained by using the experimental value of $P(0)$ and equation (2). A good agreement with the Lévy (non-gaussian) profile is observed for almost three orders of magnitude when $|Z|/\sigma \lesssim 6$ and an approximately exponential fall-off from the stable distribution is observed for $|Z|/\sigma \gtrsim 6$; here $\sigma = 0.0508$ is the standard deviation. Our results show a clear deviation of the tails of the distribution from the Lévy profile.

The Lévy distribution has an infinite second moment (if $\alpha < 2$). But our experimental finding of an exponential (or stretched exponential) fall-off implies that the second moment is finite, thereby resolving the question of how one could get around the problem of an infinite variance if the Lévy distribution is used to describe the price difference distribution[24]. This conclusion might at first sight seem to contradict our observation of Lévy scaling of the central part of the price difference distribution over fully three orders of magnitude. However, there is no contradiction, because (for example), a recent study[27] finds that Lévy scaling may hold over a long period of time for the dynamics of 'quasi-stable' stochastic processes having a finite variance.

By using the Berry–Esseen theorem[28,29], we can estimate that the maximal time needed to observe convergence for the price differences to a gaussian process is of the order of 1 month. This

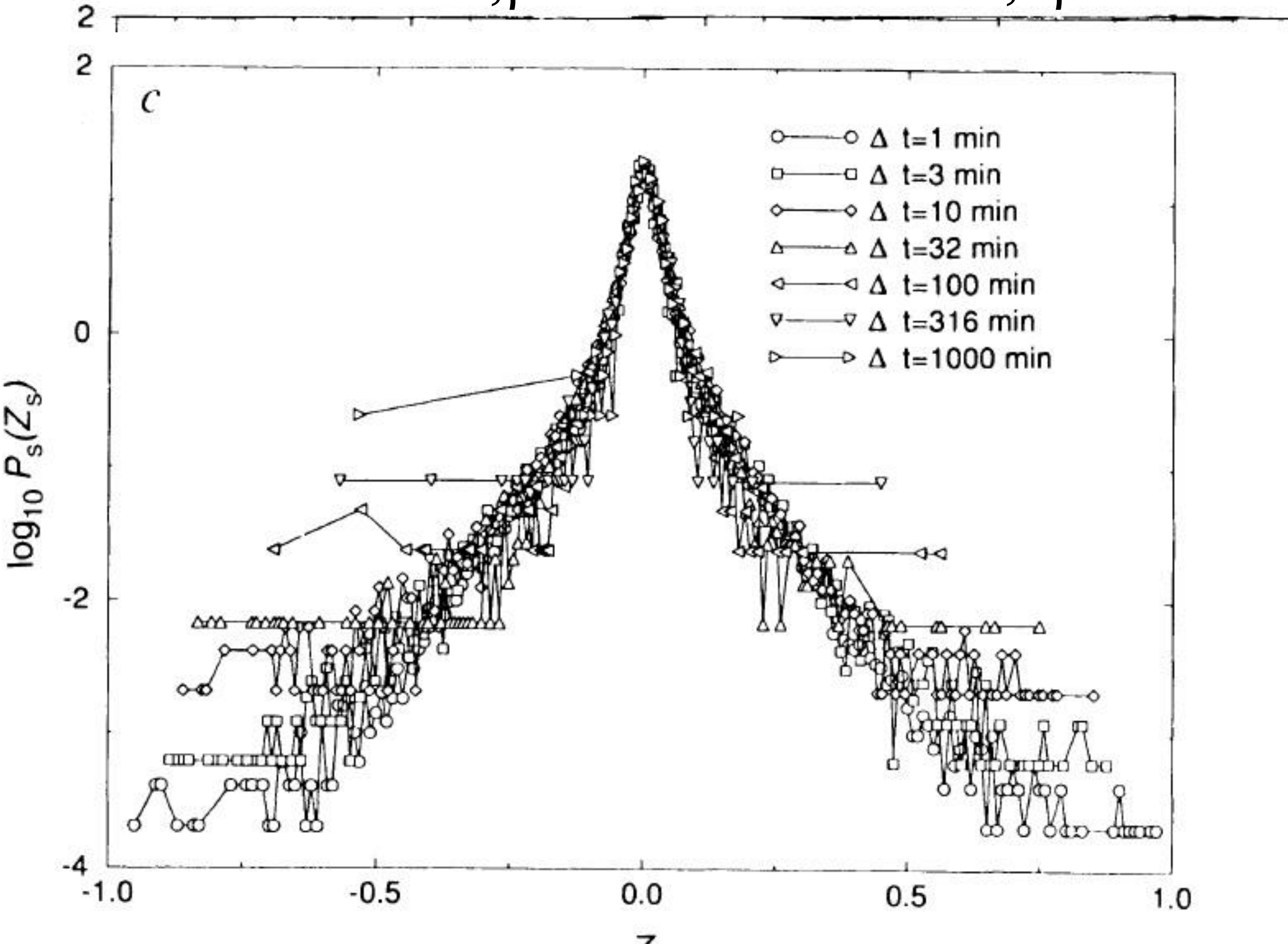
実データは正規分布から大きく乖離



データは別の分布で
指数α=1.4のLevy分布

正規分布なら片対数で $y = -x^2$ （放物線） となる

自己相似性 $P_{\alpha,\beta}(x) = \lambda^{1/\alpha} P_{\alpha,\lambda\beta}(\lambda^{1/\alpha} x)$



原点回帰率　P(0) のΔt　依存性

両対数で直線になり、
勾配　が　−1/α

$$\log\left(P_{\Delta t}(0)\right) = -\frac{1}{\alpha}\log(\Delta t) + \log C$$

Scaling behaviour in the dynamics of an economic index

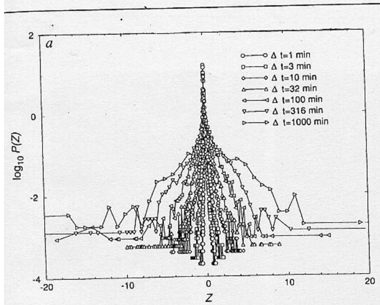
Rosario N. Mantegna & H. Eugene Stanley

Center for Polymer Studies and Department of Physics,
Boston University, Boston, Massachusetts 02215, USA

THE large-scale dynamical properties of some physical systems depend on the dynamical evolution of a large number of nonlinearly coupled subsystems. Examples include systems that exhibit self-organized criticality[1] and turbulence[2,3]. Such systems tend to exhibit spatial and temporal scaling behaviour—power-law behaviour of a particular observable. Scaling is found in a wide range of systems, from geophysical[4] to biological[5]. Here we explore the possibility that scaling phenomena occur in economic systems—especially when the economic system is one subject to precise rules,

NATURE · VOL 376 · 6 JULY 1995

価格の実データは、
正規分布ではない、
という結果！
Mantegna-Stanley著
1995　Nature



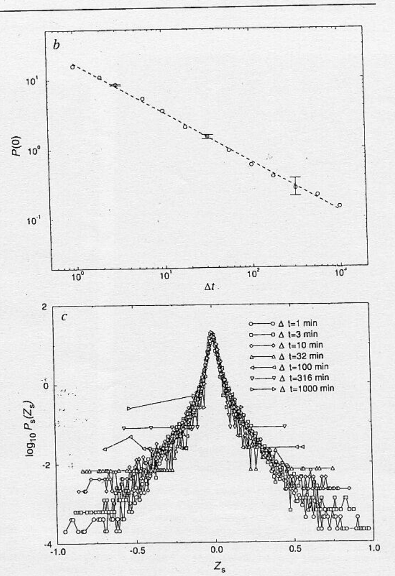


FIG. 1 *a*, Probability distributions $P(Z)$ of the S&P 500 index variations $Z(t)$ observed at time intervals Δt, which range from 1 to 1,000 min. By increasing Δt, a spreading of the probability distribution characteristic of a random walk is observed. *b*, Probability of return $P(0)$ of $Z(t)$ as a function of the time sampling intervals Δt. A scaling power-law behaviour is observed for time intervals spanning three orders of magnitude. The slope of the best-fit straight line is -0.712 ± 0.025. By using equation (2), from this value we obtain the index $\alpha=1.40\pm0.05$ characterizing the Lévy distribution. Typical estimated errors are shown. *c*, Scaled plot of the probability distributions shown in *a*. All the data collapse on the $\Delta t=1$ min distribution by using the scaling transformations of equations (3a) and (3b), with $\alpha=1.40$. The points outside the average behaviour define the noise level of that specific distribution.

as is the case in financial markets[6-8]. Specifically, we show that the scaling of the probability distribution of a particular economic index—the Standard & Poor's 500—can be described by a non-gaussian process with dynamics that, for the central part of the distribution, correspond to that predicted for a Lévy stable process[9-11]. Scaling behaviour is observed for time intervals span-

processes have quite similar statistical properties in the high-frequency regime.

We have undertaken a statistical study of timescales as short as 1 min, a value close to the minimum time needed to perform a transaction in the market. Specifically, we investigate the dynamics of a price index of one of the largest financial markets

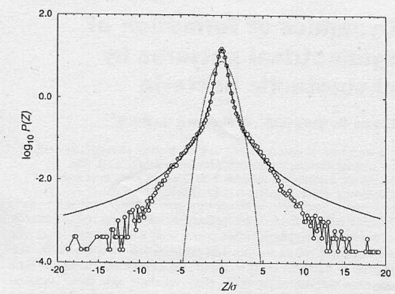


FIG. 2 Comparison of the $\Delta t=1$ min probability distribution with the symmetrical Lévy stable distribution of index $\alpha=1.40$ and scale factor $\gamma=0.00375$ (solid line). The scale factor γ is obtained from equation (2) by using the experimental values of α and $P(0)$. The dotted line shows the gaussian distribution with standard deviation σ equal to the experimental value 0.0508. The variations of price are normalized to this value. Approximately exponential deviations from the Lévy stable profile are observed for $|Z|/\sigma \gtrsim 6$.

1,000 min. The number of data in each set decreases from the maximum value of 493,545 ($\Delta t=1$ min) to the minimum value of 562 ($\Delta t=1{,}000$ min).

Figure 1*a* is a semilogarithmic plot of $P(Z)$ obtained for seven different values of Δt. As expected for a random process, the distributions are roughly symmetrical and are spreading when Δt increases. We also note that the distributions are leptokurtic, that is, they have wings larger than expected for a normal process. A determination of the parameters characterizing the distributions is difficult if one uses methods that mainly investigate the wings of distributions, especially because larger values of Δt imply a reduced number of data.

Therefore we use a different approach: we study the 'probability of return' $P(Z=0)$ as a function of Δt. With this choice we investigate the point of each probability distribution that is least affected by the noise introduced by the finiteness of the experimental data set. In Fig. 1*b*, we show $P(0)$ versus Δt in a log–log plot. The data are fitted well by a straight line of slope -0.712 ± 0.025. We observe a non-normal scaling behaviour (slope$\neq-0.5$) in an interval of trading time ranging from 1 to 1,000 min. This experimental finding agrees with the theoretical model of a Lévy walk or Lévy flight[9-11]. In fact, if the central region of the distribution is well described by a Lévy stable symmetrical distribution,

$$L_\alpha(Z,\Delta t)\equiv\frac{1}{\pi}\int_0^\infty \exp\left(-\gamma\Delta t q^\alpha\right)\cos(qZ)\,dq \qquad (1)$$

of index α and scale factor γ at $\Delta t=1$, where $\exp(-\gamma\Delta t|q|^\alpha)$ is

dynamics of the probability distribution $P(Z)$ of the random process over time intervals spanning three orders of magnitude.

In Fig. 2, we compare the probability distribution observed for $\Delta t=1$ min with the Lévy stable distribution of index $\alpha=1.40$. Note that the solid line is not simply a 'fit' to the data; rather, the appropriate scale factor $\gamma=0.00375$ is obtained by using the experimental value of $P(0)$ and equation (2). A good agreement with the Lévy (non-gaussian) profile is observed for almost three orders of magnitude when $|Z|/\sigma\lesssim6$ and an approximately exponential fall-off from the stable distribution is observed for $|Z|/\sigma\gtrsim6$; here $\sigma=0.0508$ is the standard deviation. Our results show a clear deviation of the tails of the distribution from the Lévy profile.

The Lévy distribution has an infinite second moment (if $\alpha<2$). But our experimental finding of an exponential (or stretched exponential) fall-off implies that the second moment is finite, thereby resolving the question of how one could get around the problem of an infinite variance if the Lévy distribution is used to describe the price difference distribution[24]. This conclusion might at first sight seem to contradict our observation of Lévy scaling of the central part of the price difference distribution over fully three orders of magnitude. However, there is no contradiction, because (for example), a recent study[27] finds that Lévy scaling may hold over a long period of time for the dynamics of 'quasi-stable' stochastic processes having a finite variance.

By using the Berry–Esseen theorem[28,29], we can estimate that the maximal time needed to observe convergence for the price differences to a gaussian process is of the order of 1 month. This

日本株（平均株価）でも同じ結果

- 日大理工・島田一平研究室。柳川一貴修論
  1996年の東京株価指数（1分足）
  α = 1.41のLevy分布

柳川一貴,「短い時間での価格変動を起こす要因」

- 日本大学大学院理工学研究科平成9年度修士論文

日本物理学会秋の分科会（神戸大学,1997年9月16日）講演予稿集52(2-3),8a-YD-9;

- 総合研究大学院大学共同研究"新分野の開拓"1997-2002, 小グループ「経済学」第１回研究会(1998年7月14日東京駅ホテル)における講演.

総研大「新分野の開拓」1998-2002 →統数研「経済物理とその周辺」

- 1998　総研大研究会「新分野の開拓」
  東京ステーションホテル
- 1999　総研大研究会「新分野の開拓」
  丸ノ内ホテル、熱海、福岡大、
  総研大葉山キャンパス
- 2002~　統数研研究会「経済物理とその周辺」

Elliot W. Montroll

http://en.wikipedia.org/wiki/Elliott_Waters_Montroll

(5/4, 1916-12/3, 1983)
Known for
Traffic flow analysis
Notable awards
Lanchester Prize('59)

Kenneth.Wilson,M.Fisher, L.Kadanoff 等に影響.
マンハッタン島トンネル設計

価格変動のスケーリング
所得分布のスケーリング
繰り込み群のアイデア

株取引の高速化

2010年東証アローヘッド
→株価の注文処理がミリ秒単位で可能

株価変動の確率分布に
変化はないのか？

レヴィ分布

$$P_{\alpha,\beta}(Z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}} dk$$

Z : 株価の変動

$$Z(t) = \log X(t+\Delta t) - \log X(t)$$

$X(t)$:時刻tにおける株価

レヴィ分布

安定分布

$$P_{\alpha,\ \beta}(Z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\,dk$$

P : 出現確率

Z : 確率変数(価格変動)

α : 裾の広がりを表す

β : 尺度を表す

レヴィ分布

$$P_{\alpha,\ \beta}(Z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\,dk$$

$0 < \alpha \leq 2$ の範囲

ローレンツ分布

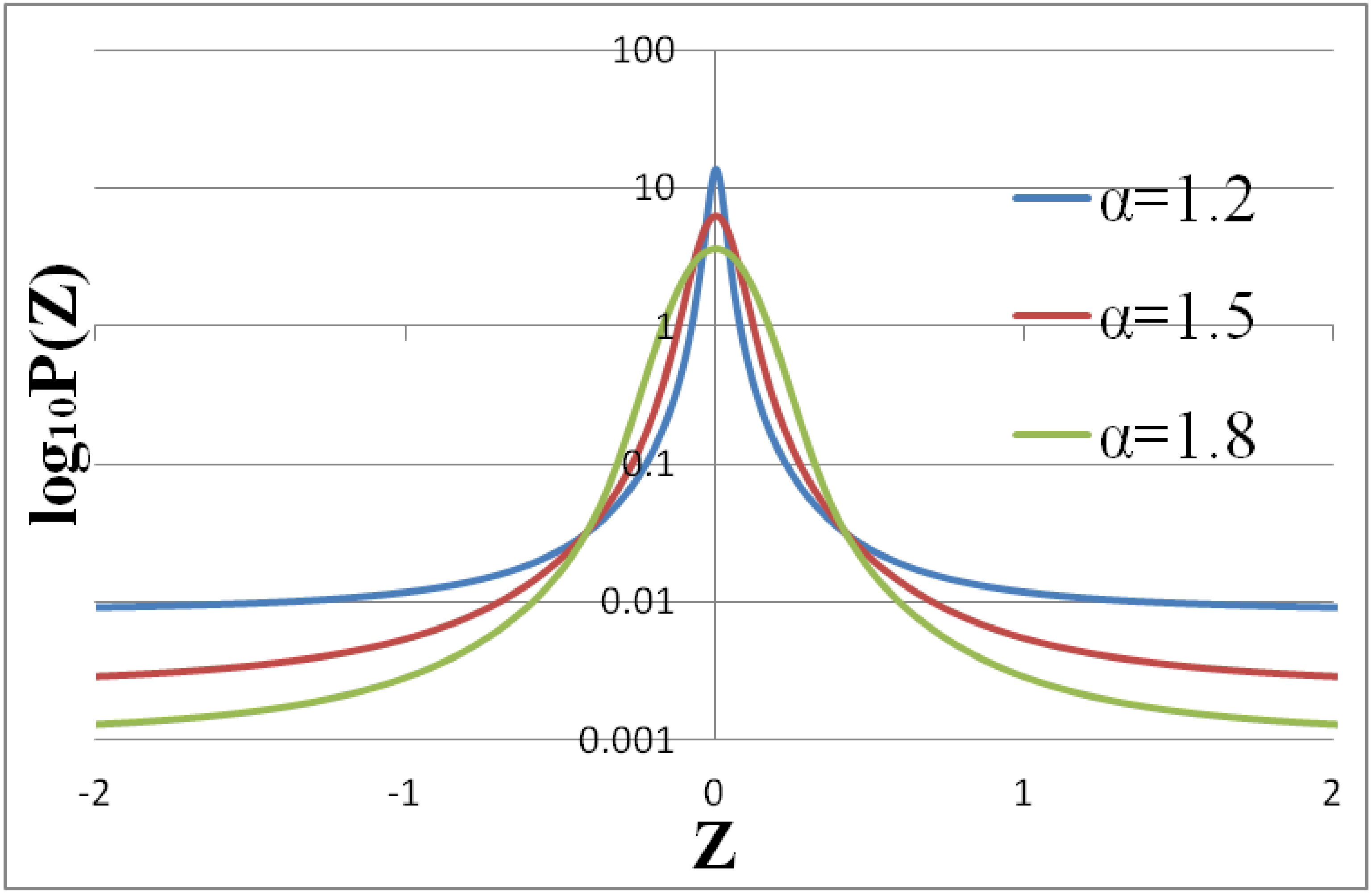
α=1.0 $$P_{\alpha,\ \beta}(Z) = \frac{\beta}{\pi}\frac{1}{\beta^2 + Z^2}$$

正規分布

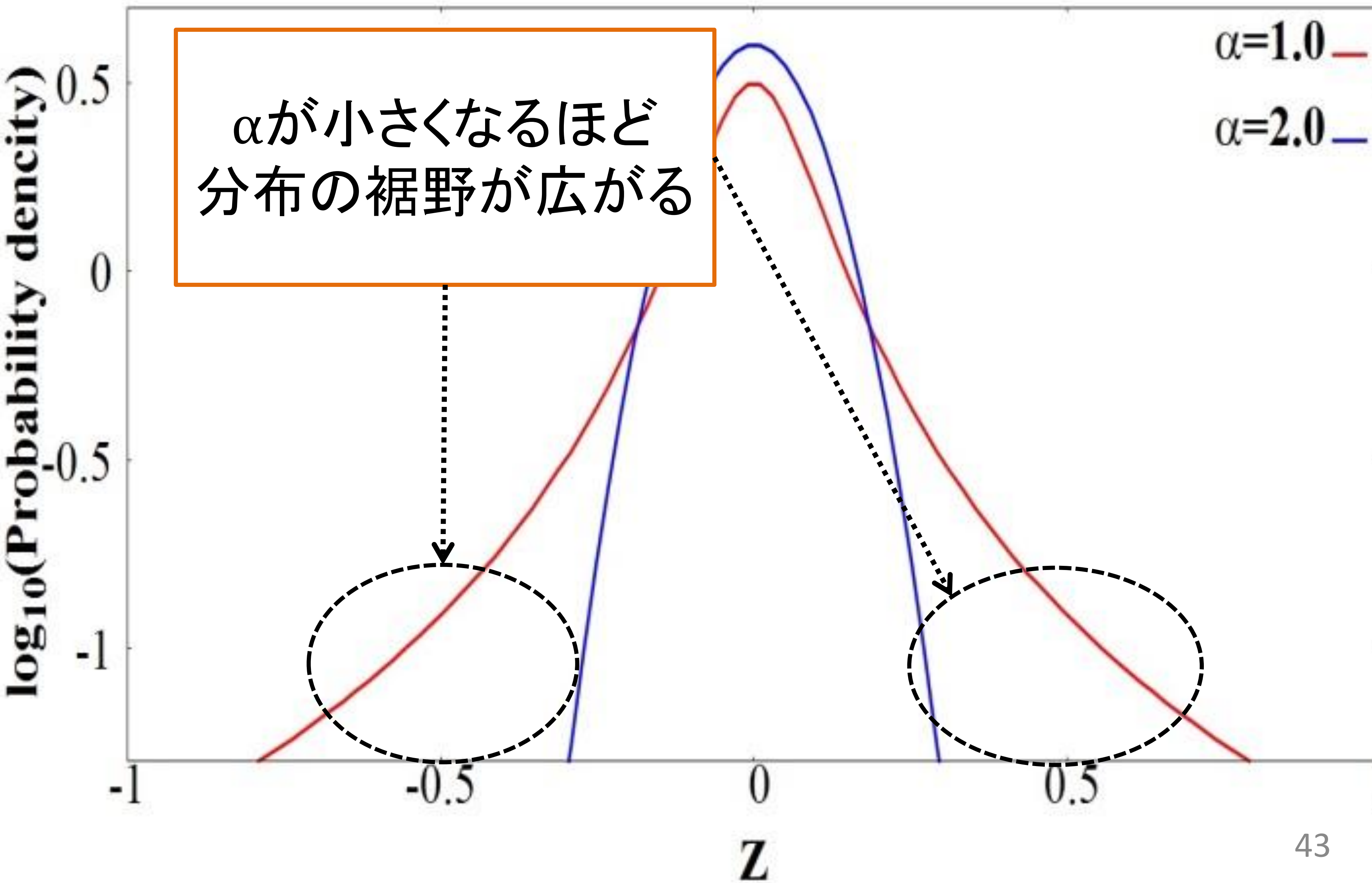
α=2.0 $$P_{\alpha,\ \beta}(Z) = \frac{1}{2\sqrt{\pi\beta}}\exp\left(-\frac{Z^2}{4\beta}\right)$$

レヴィ分布

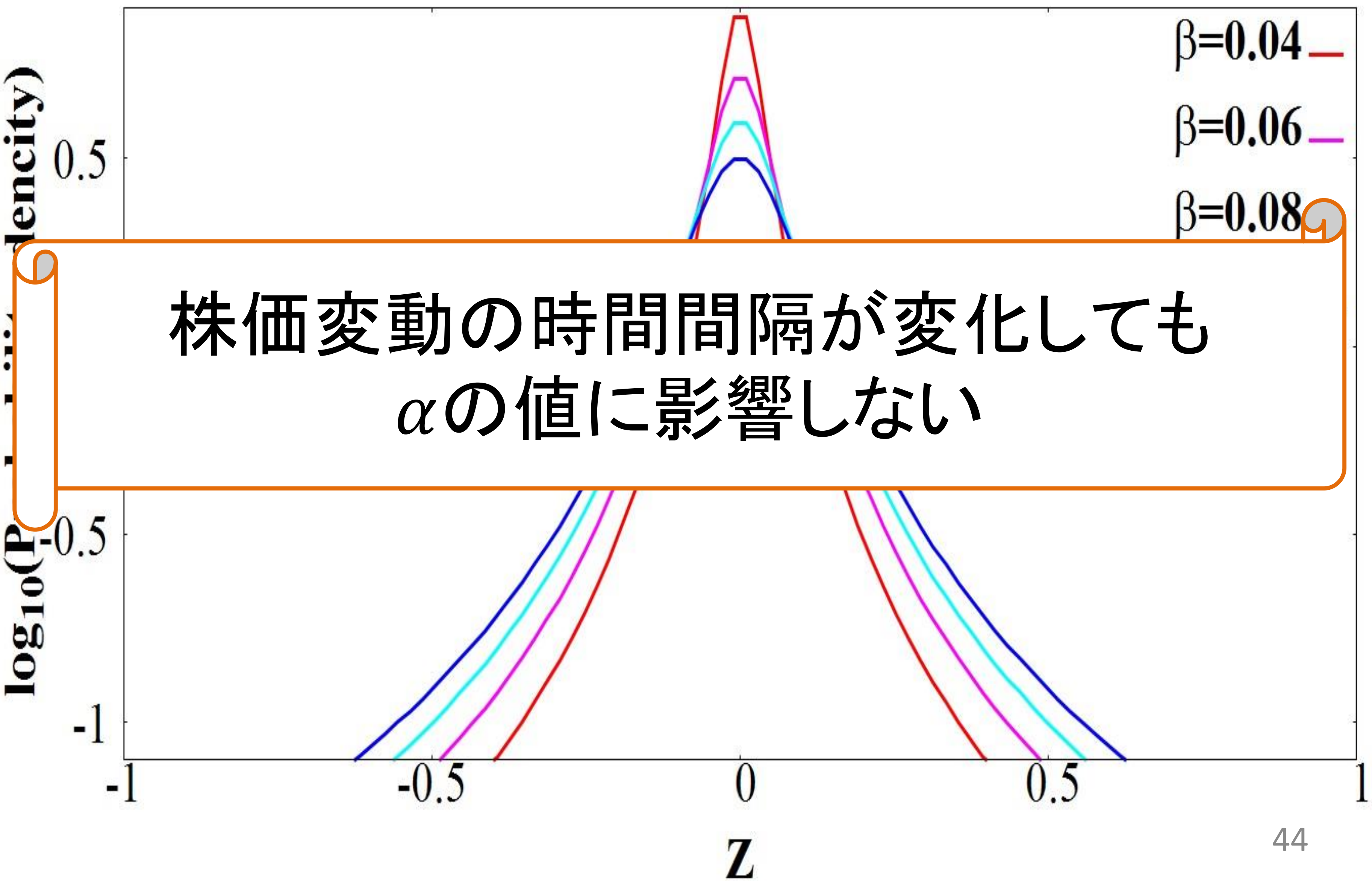
$$P_{\alpha,\beta}(Z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\,dk$$



αの値について



βの値について



株価変動の時間間隔が変化しても
αの値に影響しない

レヴィ分布

レヴィ分布を解析的に求めるのは困難

➡ 数値積分にて求める

$$P_{\alpha,\ \beta}(Z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\,dk$$

変形

$$P_{\alpha,\ \beta}(Z) = \frac{1}{\pi}\int_{0}^{\infty} e^{-\beta k^{\alpha}} \times \cos(kZ)\,dk$$

$e^{-\beta k^{\alpha}} = 10^{-8}$ となるように積分範囲を定め

50000分割して足し合わせる

使用データ

平均値のデータ
と
業種の異なる3銘柄

銘柄名	銘柄番号	業種
関西電力	9503	電力･ガス
日産自動車	7201	自動車･輸送機
東芝	6502	電機･精密

銘柄ごとに分布の違いがあるか？

ヒストグラム

価格変動の対数収益を
確率変数Zとする

確率変数Z

$$Z(t) = \log\big(x(t)\big) - \log\big(x(t-1)\big)$$

x : 価格
t : 時刻

ヒストグラム

確率変数Z

$$Z(t) = \log(x(t)) - \log(x(t-1))$$

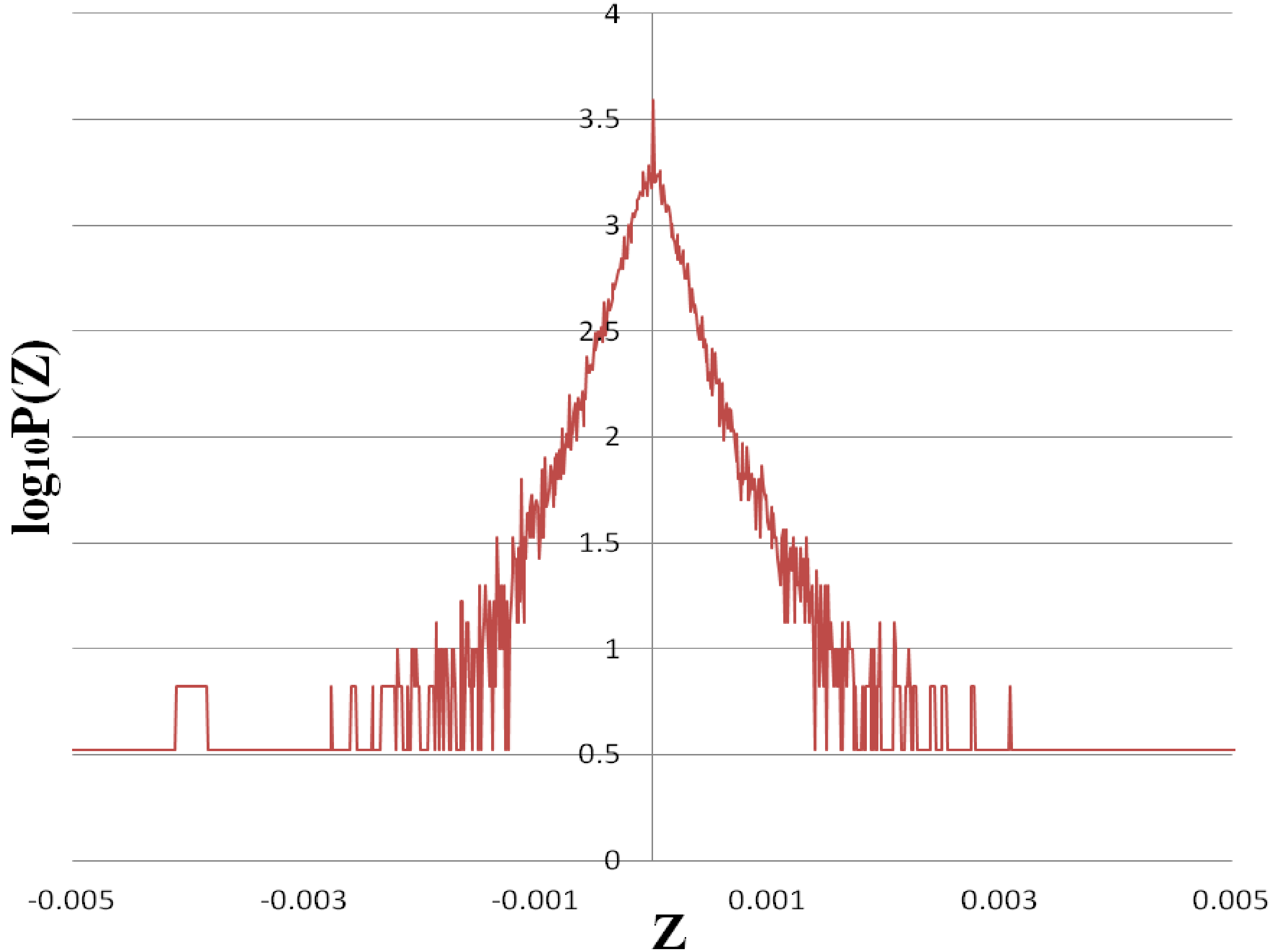
Zの範囲 [-0.01, 0.01]

平均値データ 分割数 2000

各銘柄 分割数 100

Z=0付近に出現頻度0
が出ないように分割数を決定

ヒストグラム



平均値(分割数2000)のヒストグラム

検証データ

東証アローヘッド市場　5秒足

2013/4/1/9:00:05　～　2013/12/25/15:00:00

1銘柄:640800　データ

100銘柄

時間解像度　Δt

Δt	時間間隔(秒)	データ数
1	5	640800
3	15	213600
6	30	106800
12	60	53400
24	120	26700

対数収益

銘柄iの株価：$X_i(t)$ $(i = 1,2,\cdots 100)$

5秒間隔：$\Delta t = 1$

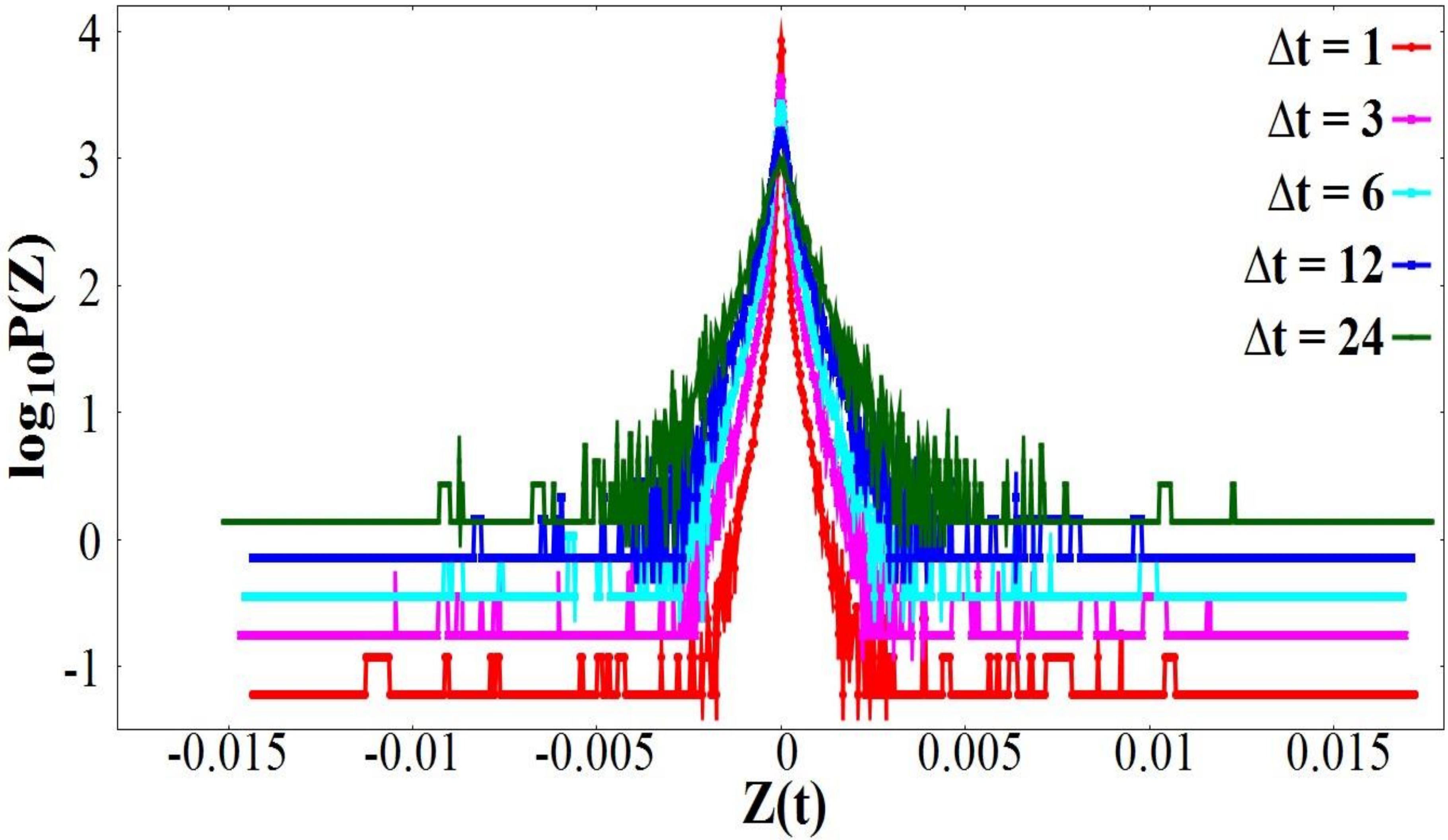
対数収益： $Y_{i,\Delta t}(t) = \log X_i(t + \Delta t) - \log X_i(t)$

$$\Delta t = 1 \ \rightarrow \ Y_{i,1}(t) = \log \frac{X_i(t+5)}{X_i(t)}$$

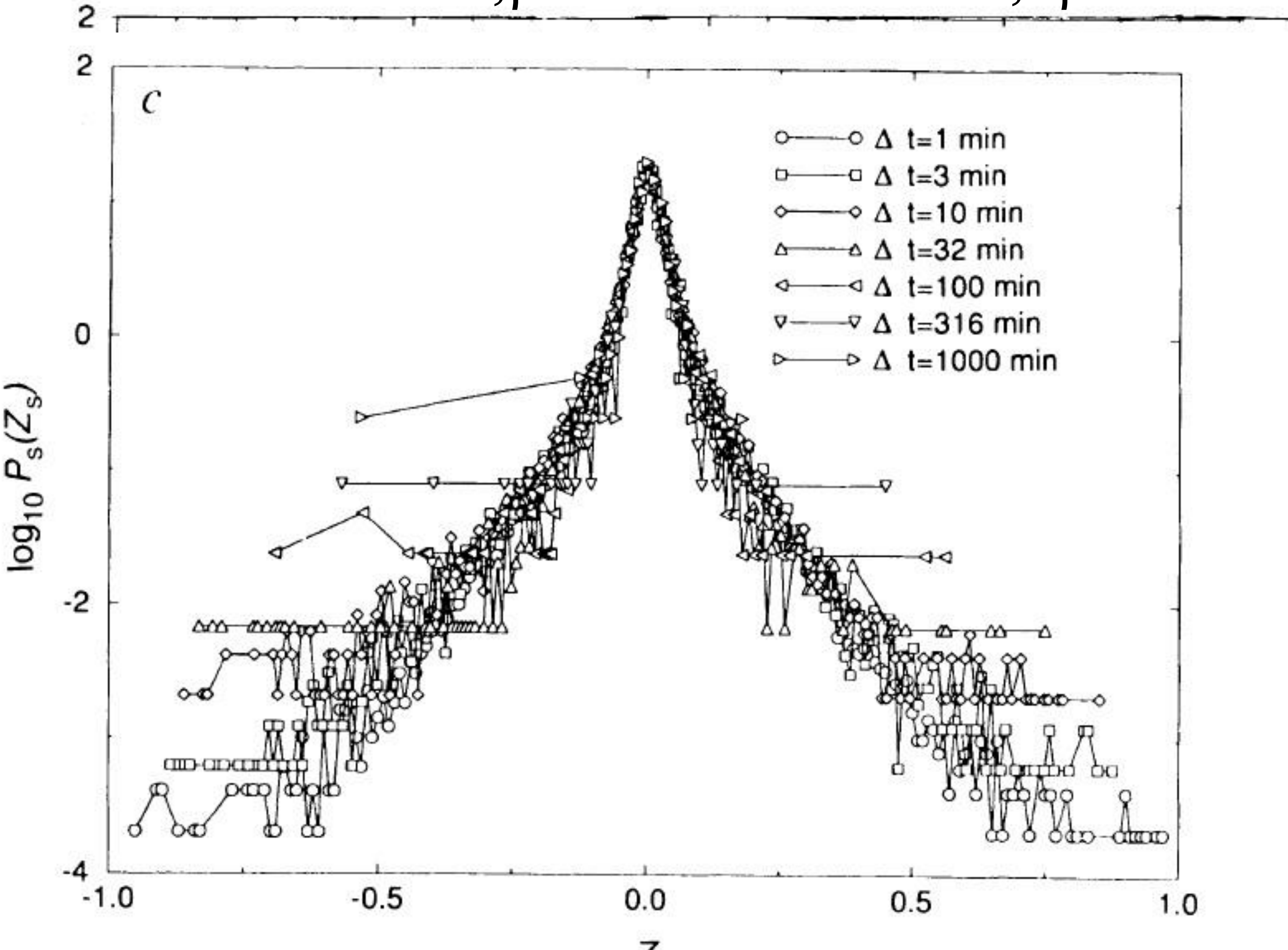
$$\Delta t = 3 \ \rightarrow \ Y_{i,3}(t) = \log \frac{X_i(t+15)}{X_i(t)}$$

平均：$Z_{\Delta t}(t) = \frac{1}{100} \sum_{i=1}^{100} Y_{i,\Delta t}(t)$

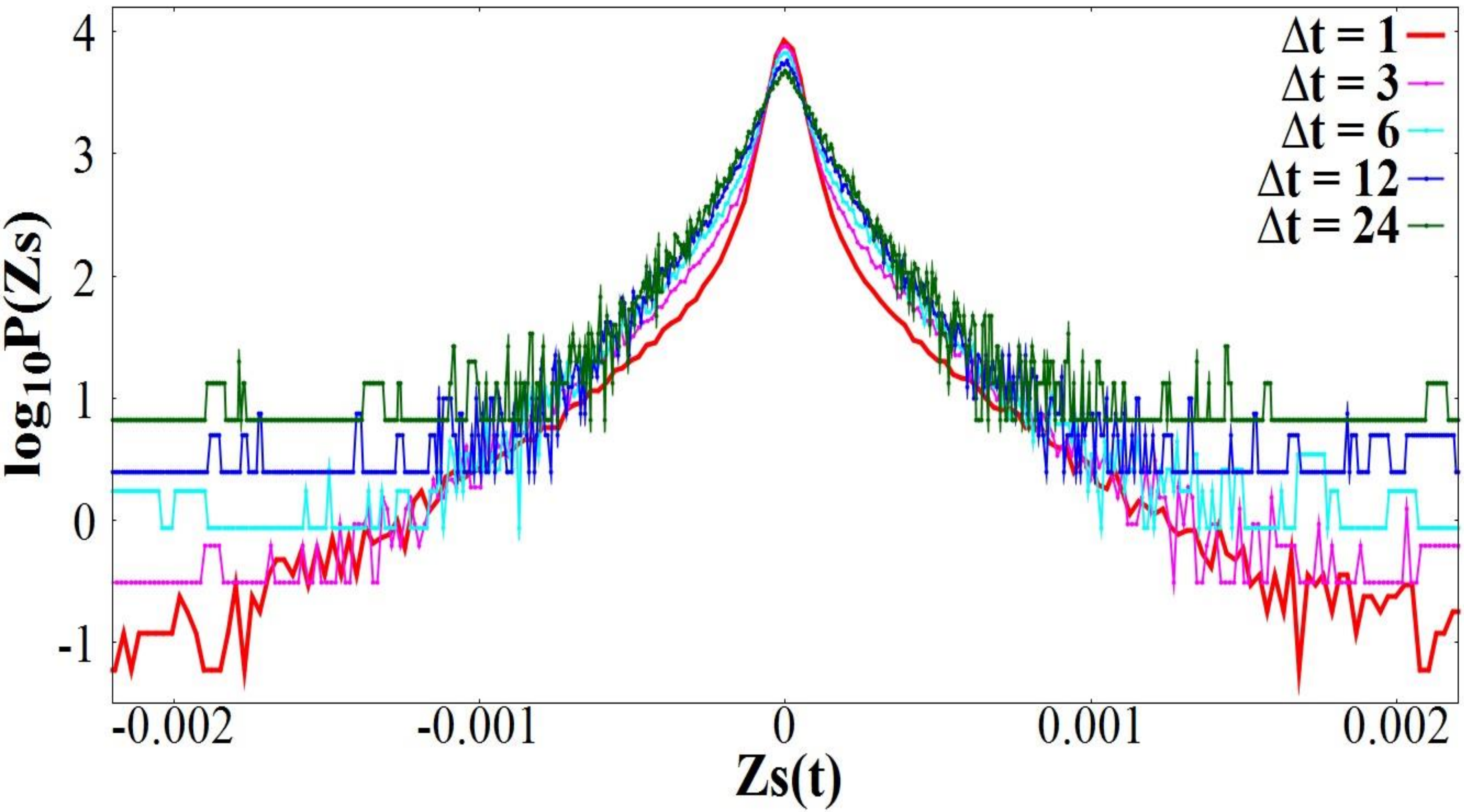
ヒストグラム　　藤井猛・卒論・2016年2月



自己相似性 $P_{\alpha,\beta}(x) = \lambda^{1/\alpha} P_{\alpha,\lambda\beta}(\lambda^{1/\alpha} x)$

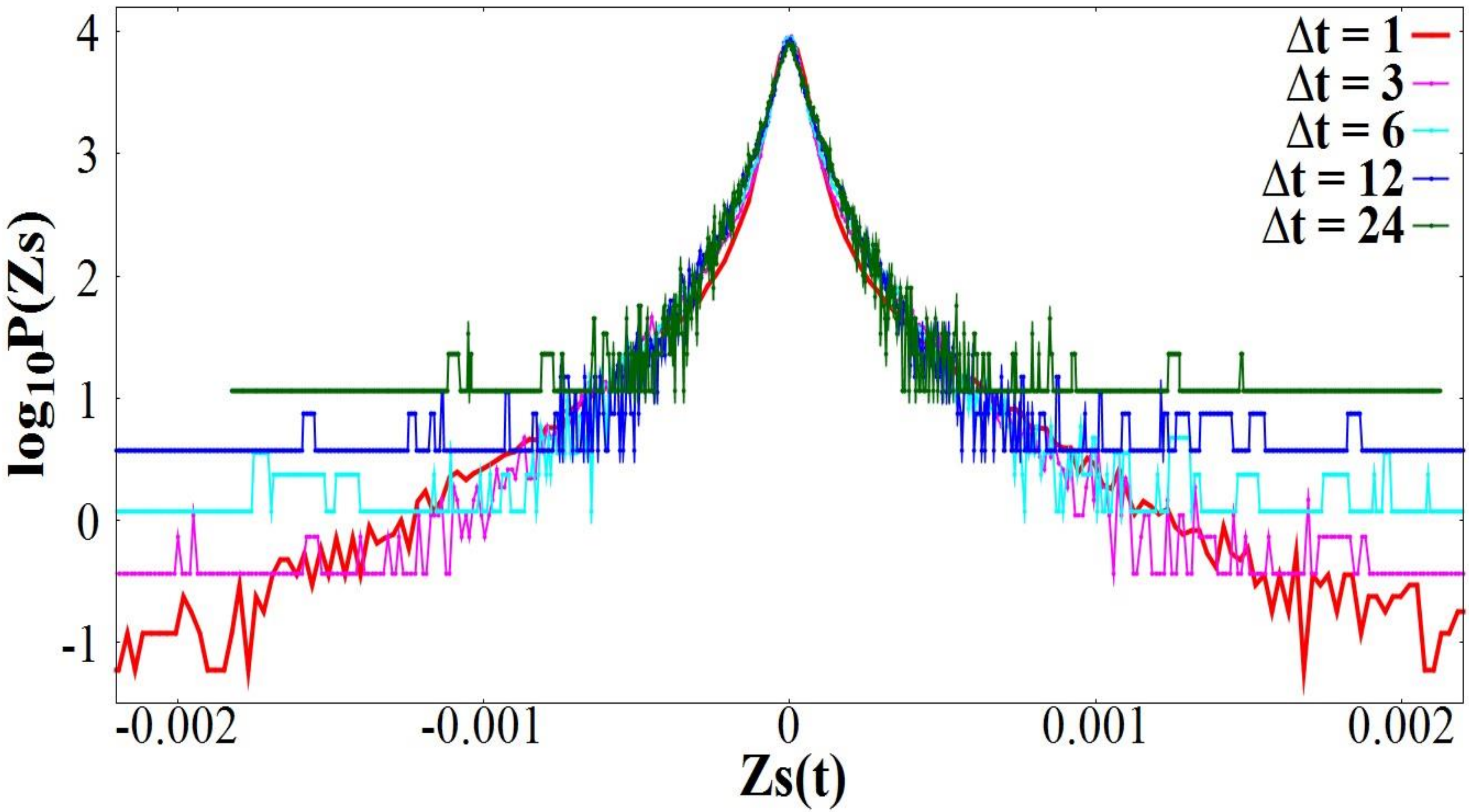


$\alpha = 2.0$



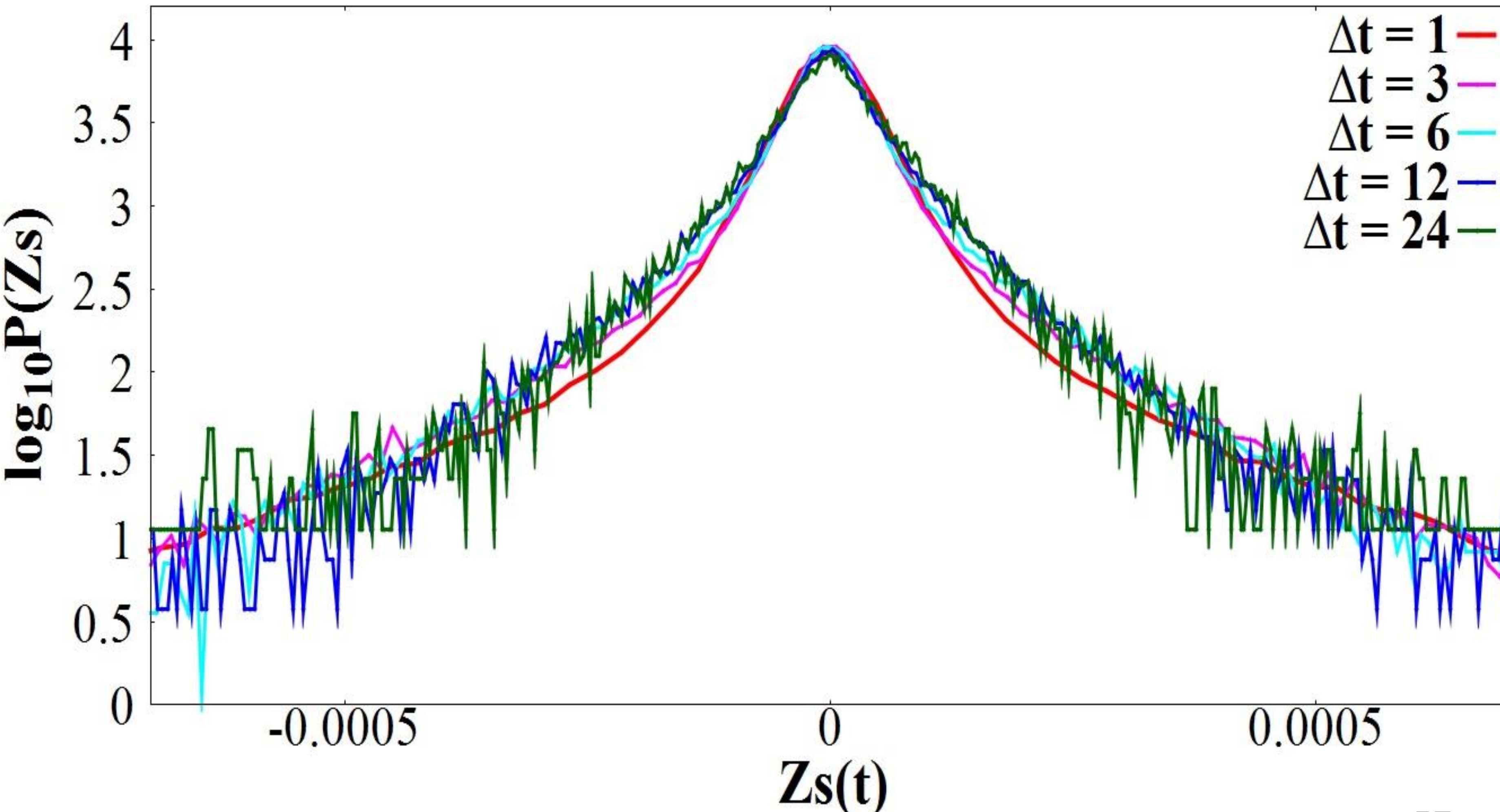
4
3
2
1
0
-1
log10P(Zs)
-0.002
-0.001
0
0.001
0.002
Zs(t)
Δt = 1
Δt = 3
Δt = 6
Δt = 12
Δt = 24

$\alpha = 1.5$

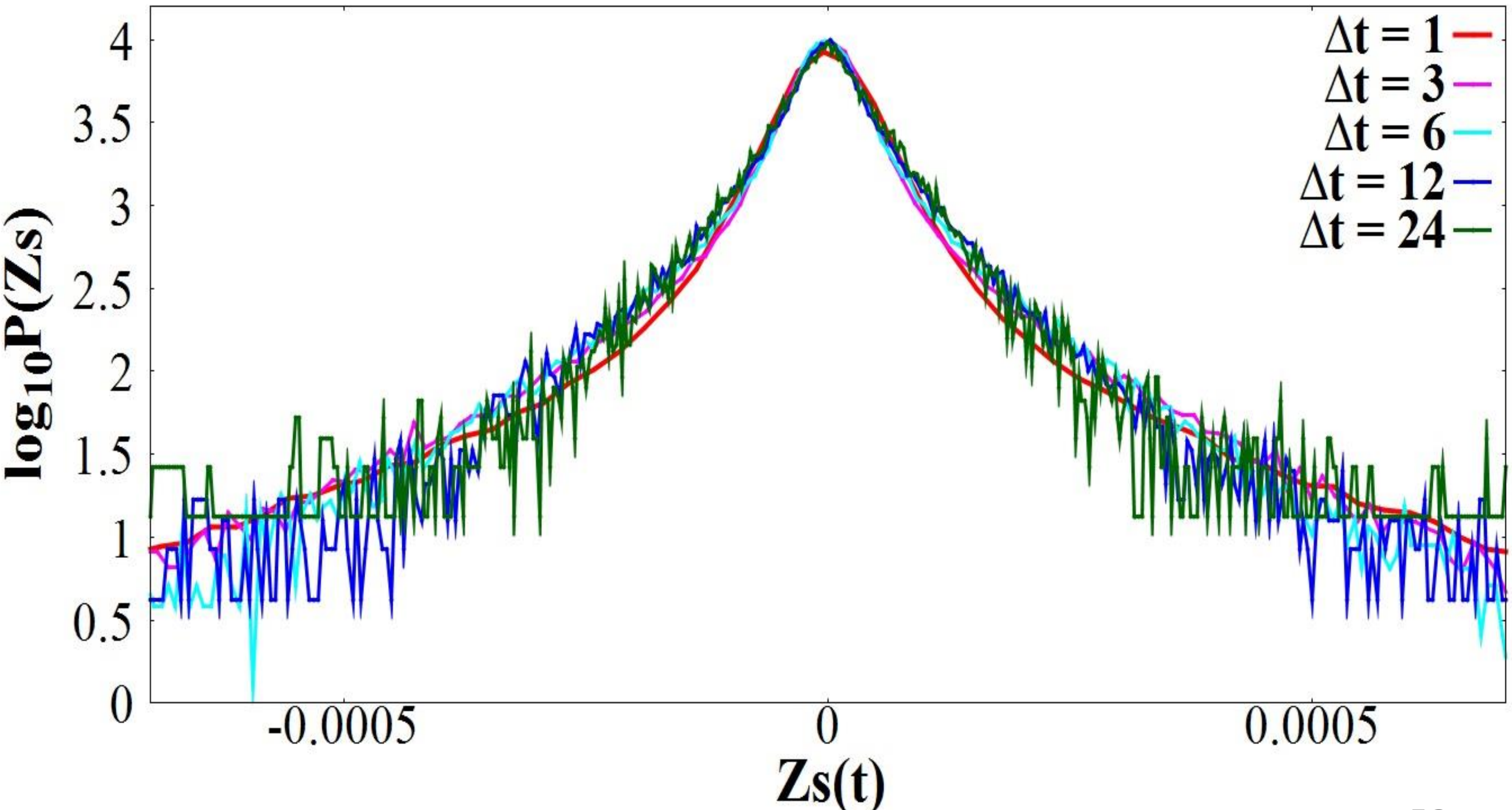


Δt = 1
Δt = 3
Δt = 6
Δt = 12
Δt = 24
log10P(Zs)
Zs(t)
4
3
2
1
0
-1
-0.002
-0.001
0
0.001
0.002

$\alpha = 1.5$ （拡大）

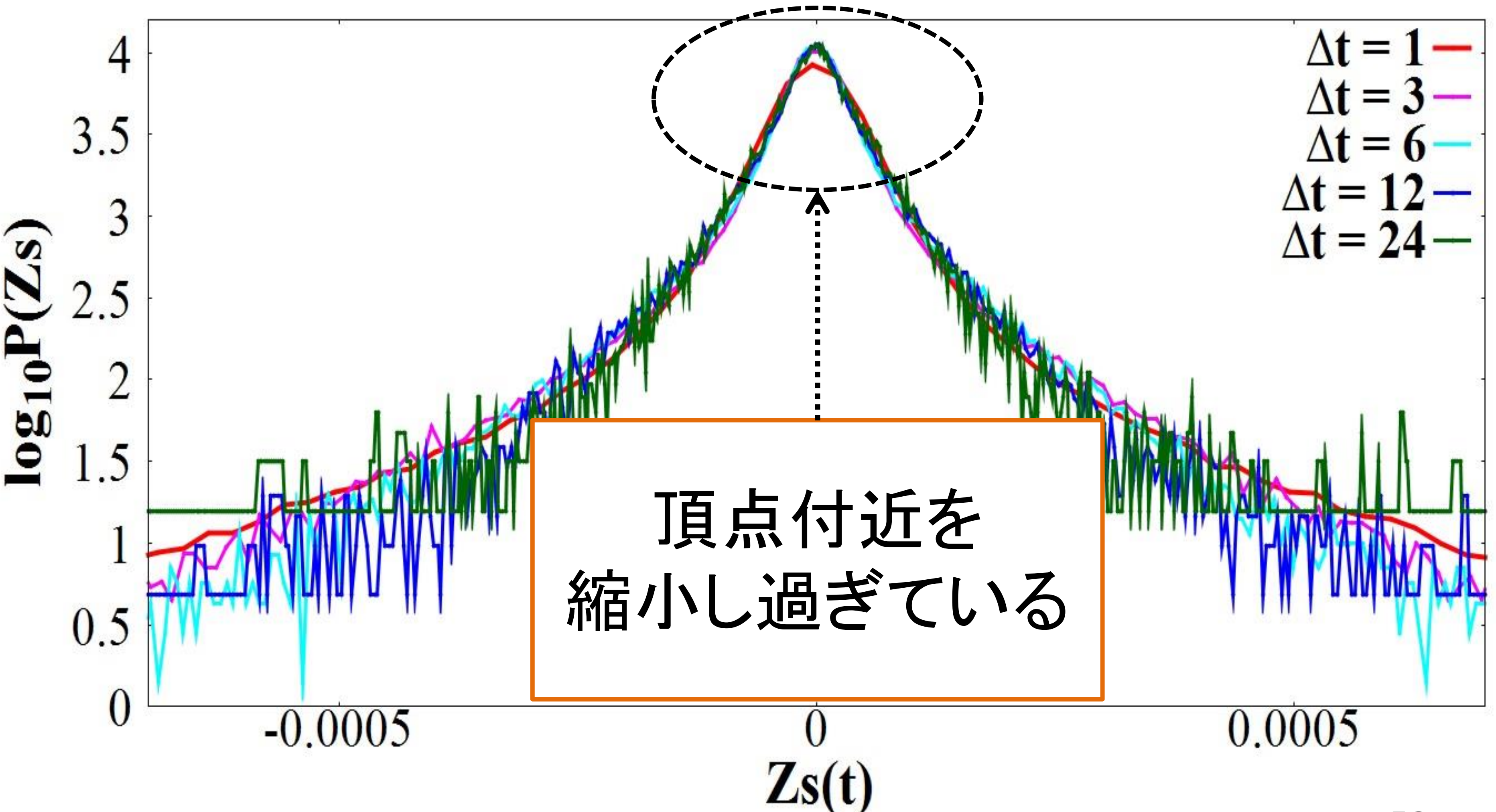

Δt = 1
Δt = 3
Δt = 6
Δt = 12
Δt = 24
log10P(Zs)
Zs(t)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.0005
0
0.0005

$\alpha = 1.4$ （拡大）



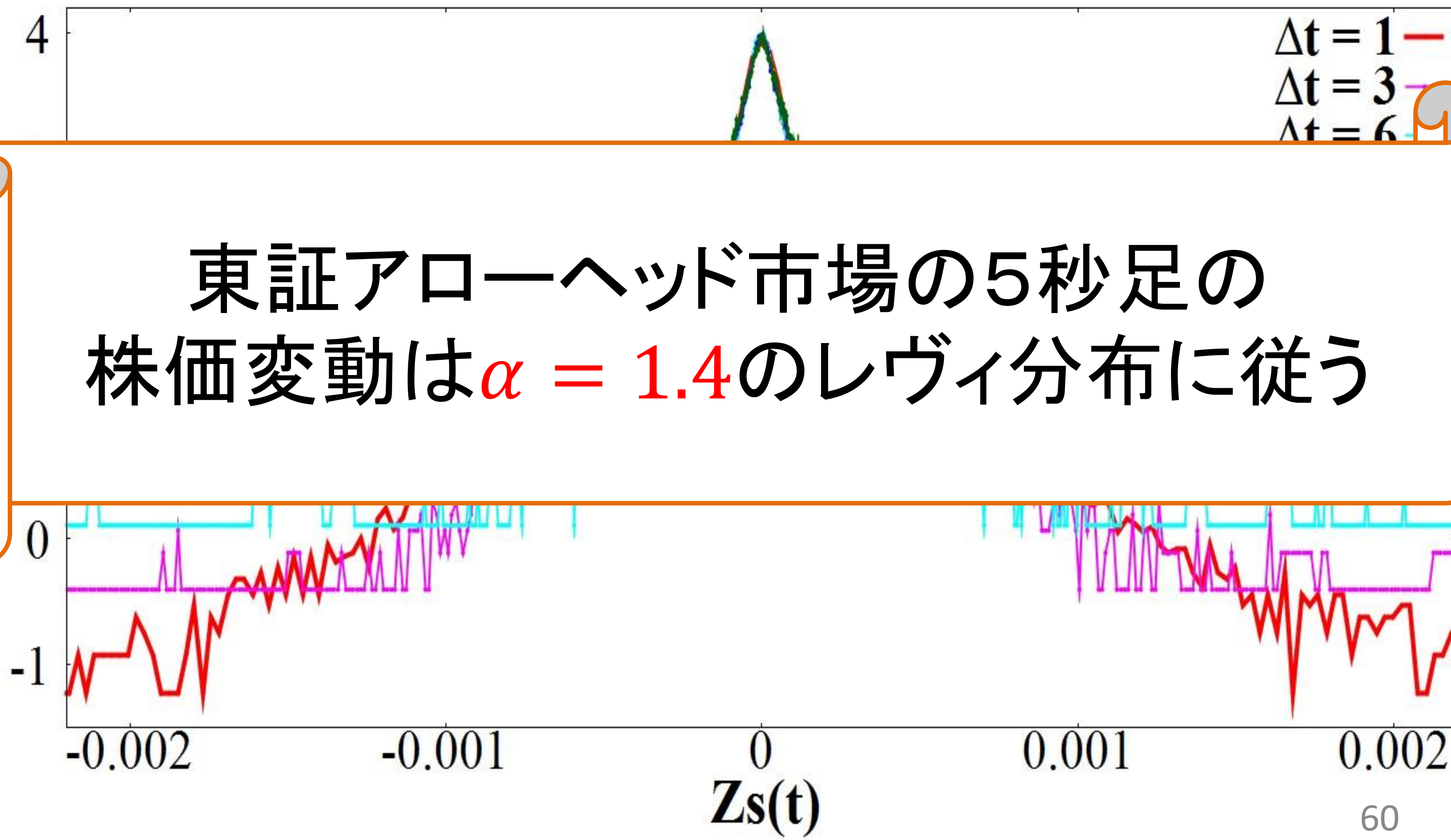
Δt = 1
Δt = 3
Δt = 6
Δt = 12
Δt = 24
log10P(Zs)
Zs(t)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.0005
0
0.0005

$\alpha = 1.3$ （拡大）



Δt = 1
Δt = 3
Δt = 6
Δt = 12
Δt = 24
log10P(Zs)
Zs(t)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.0005
0
0.0005
頂点付近を
縮小し過ぎている

結果



東証アローヘッド市場の5秒足の
株価変動は$\alpha = 1.4$のレヴィ分布に従う

原点回帰率　P(0)　の　Δt　依存

- 自己相似性

$$P_{\alpha,\beta}(x) = \lambda^{1/\alpha} P_{\alpha,\lambda\beta}(\lambda^{1/\alpha} x)$$

Return rate to the origin: P(0) depends on Δt

- Self-similarity

$$P_{\alpha,\beta}(x) = \lambda^{1/\alpha} P_{\alpha,\lambda\beta}(\lambda^{1/\alpha} x)$$

$$P_{\alpha,\Delta t}(0) = \lambda^{1/\alpha} P_{\alpha,\lambda\Delta t}(0)$$

$$\log\left(P_{\Delta t}(0)\right) = -\frac{1}{\alpha}\log(\Delta t) + \log C$$

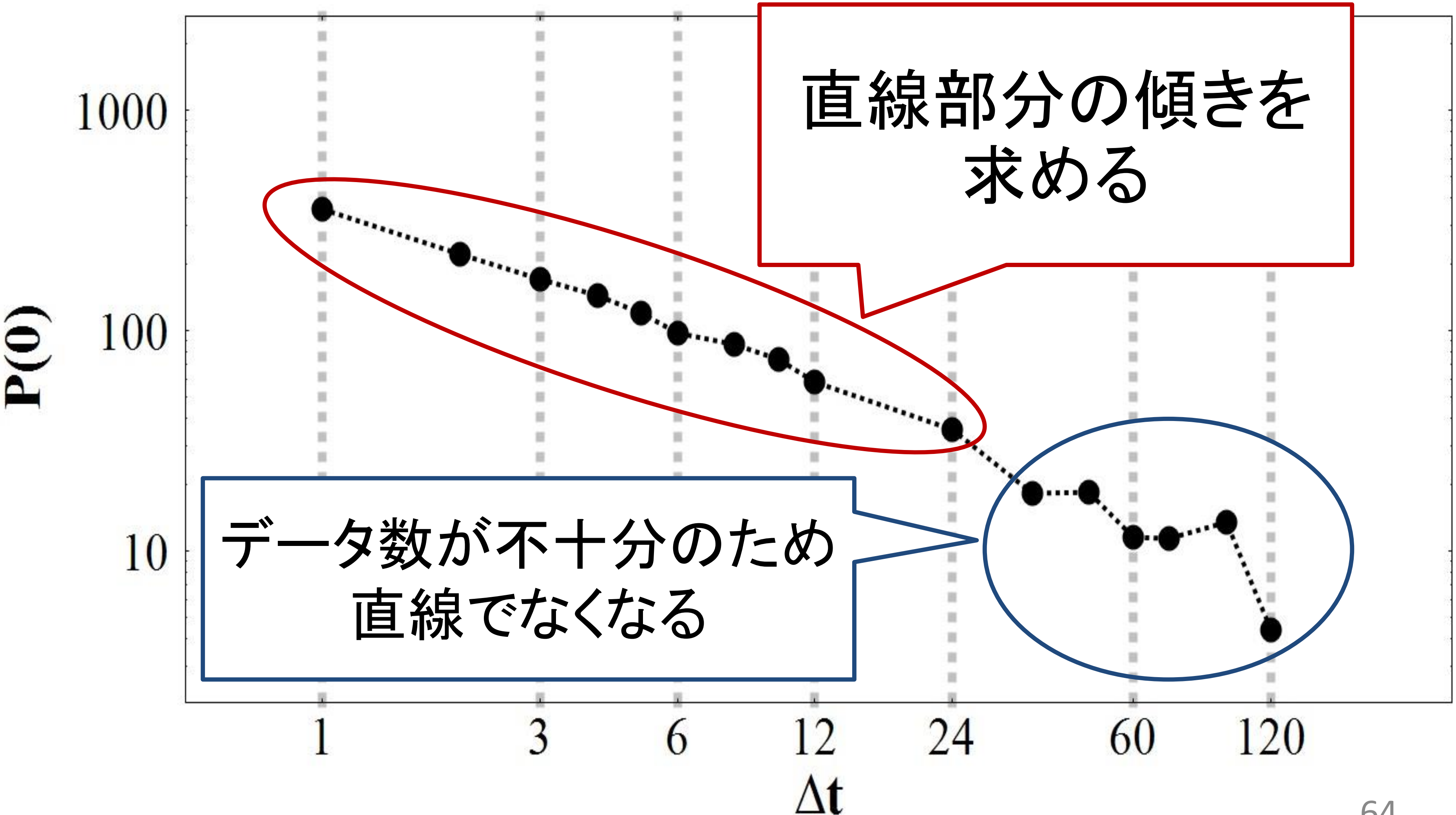
検証: αの値

$$\log\left(P_{\Delta t}(0)\right) = -\frac{1}{\alpha}\log(\Delta t) + \log C$$

原点回帰確率: $\mathrm{P}(Z = 0)$

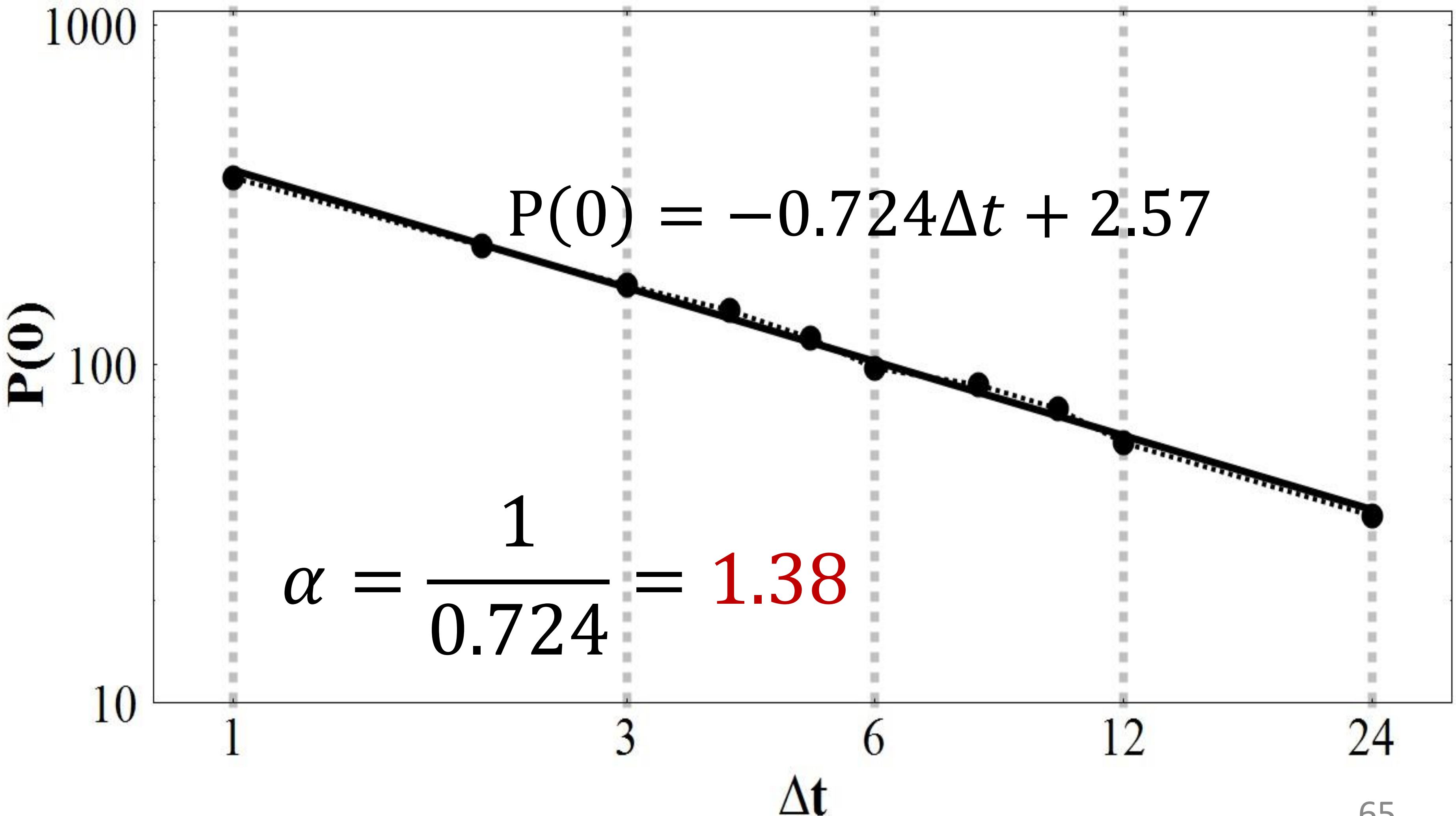
Δtを様々な値に変えて両対数グラフ
→近似直線の傾きからαの値を算出

原点回帰確率



直線部分の傾きを
求める
データ数が不十分のため
直線でなくなる
P(0)
1000
100
10
1
3
6
12
24
60
120
Δt

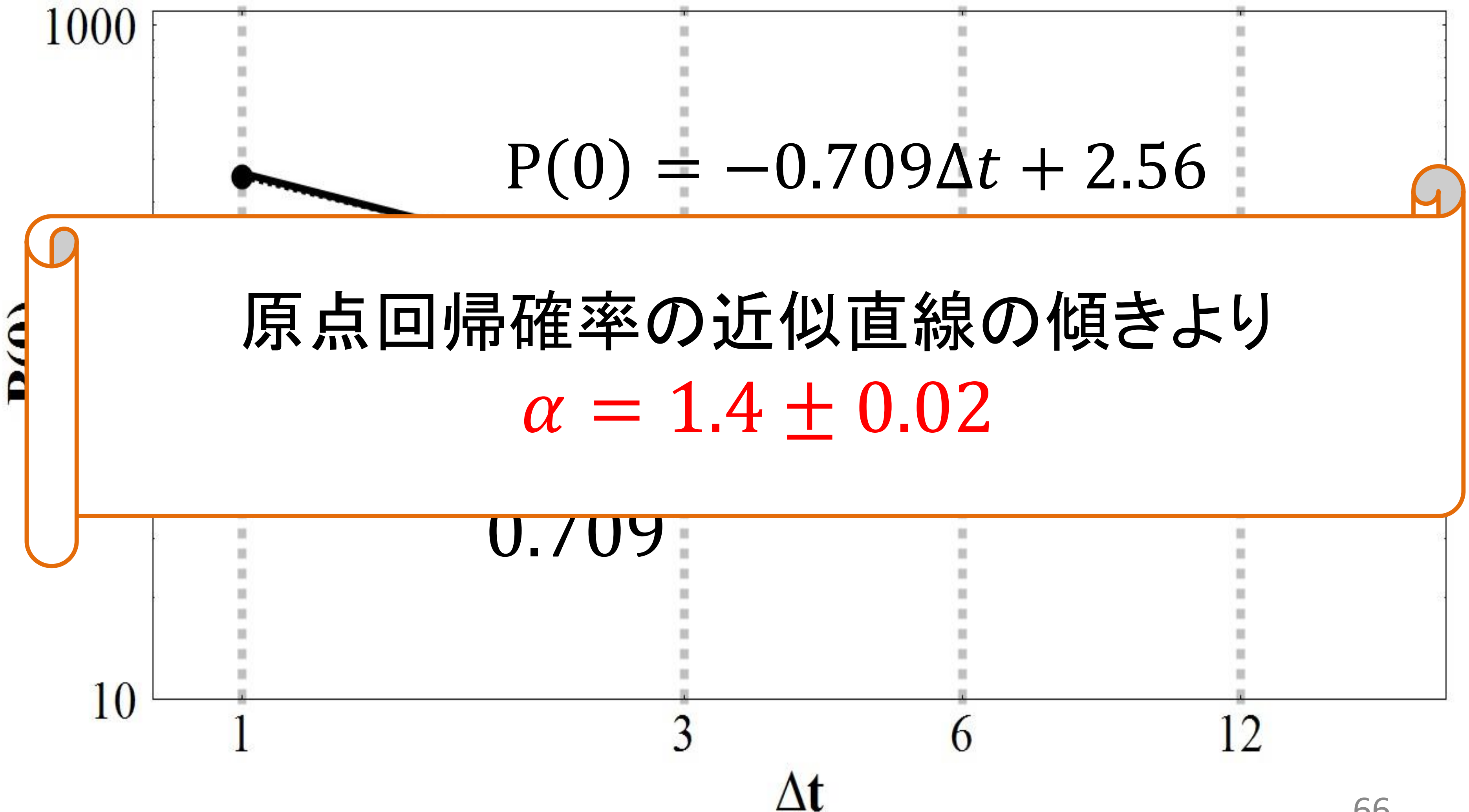
$\Delta t = 24$までの傾き



$P(0) = -0.724\Delta t + 2.57$

$\alpha = \frac{1}{0.724} = 1.38$

$\Delta t = 12$までの傾き



原点回帰確率の近似直線の傾きより

$\alpha = 1.4 \pm 0.02$

KL情報量を用いた定量化

- Kullback–Leibler divergence （KL情報量）

を用いて

真の分布（データによるもの）と
モデル関数

の差を測る

Kullback Leibler divergence

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

p(x) : 真の分布 (histogram)

q(x) : 比較対象の分布(Levy distribution)

KL情報量　D(p||q)を最小にする関数 q(x) を選ぶ

１９９５年の１分データによる検証

使用データ

東証1部の1分足データ

銘柄数	440銘柄
期間	2015年6月16日から 2015年11月4日まで
データ長	29386

使用データ

市場全体について調べるため
東証1部全440銘柄の平均をとる

KL情報量最小化

銘柄	α	β	KL情報量
平均値	1.40	5.4×10^{-6}	0.039
関西電力	1.55	10.0×10^{-6}	0.286
日産自動車	1.65	3.9×10^{-6}	0.423
東芝	1.55	8.8×10^{-6}	0.156

平均値と日産自動車の
結果について詳しく見ていく

吉井勝俊修論
2016.2.12

KL情報量を最小化

平均値のヒストグラム
分割数とαとβの推移

分割数	α	β	KL情報量
100	1.75	0.5×10^{-6}	0.095
500	1.55	1.9×10^{-6}	0.072
1000	1.40	5.4×10^{-6}	0.027
2000	1.40	5.4×10^{-6}	0.039

Summary

Arrowhead Market (Tokyo SE)

5 second sample data, 2013

Lévy distribution of $\alpha = 1.4$

S&P500 index (USA 1980's)

1 minute data

Statistical distribution unchanged under the new arrowhead trading system

まとめ

東証アローヘッド市場

株価変動の統計的性質は
取引が高速化されても変化しない

1分足の株価変動

?