# モデリングの数理
# 第7回　拡散現象のモデリング3

拡散現象の応用
金融市場の価格変動
時系列の基本モデル
ARモデル
ARCH・GARCHモデル

第3Q
月・木、7・8限
(15:05-16:35)
西8号館　W834

2019年10月24日

担当：
高安 秀樹
特任教授
高安美佐子
教授

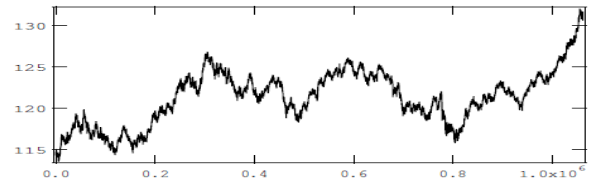
Research unit for Super-Smart-Society
ADAM

東京工業大学　科学技術創成研究院
ビッグデータ数理科学研究ユニット
情報理工学院 数理・計算科学系　高安美佐子研究室

## 拡散現象のモデルの応用

外国為替市場：通貨の交換をする市場、ドル円レート
ex. 1ドル＝115.105円
為替市場は、コンピュータネットワークとサーバー
月曜の朝から金曜の夜まで連続して取引できる
取引の単位は100万ドル（約1億円）
売買の取引が成立すると市場価格が決まる
一日約10000回程度価格の変化がある
世界経済の背骨のような重要な役割

（実データ，2001 年）
為替レート時系列

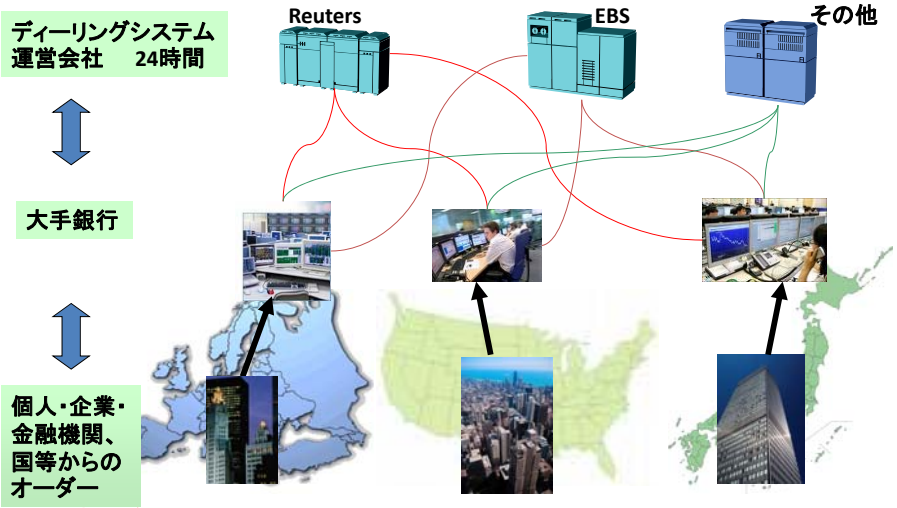



ランダムウォークとよく似た時系列

## 外国為替市場の仕組み

コンピュータネットワークで世界中とつながった巨大取引システム

連続的な世界最大の自由市場・眠らない市場・取引の規制がない




自動売買が解禁された現在、ネットワークの渋滞による100分の1秒の遅延も大問題！！

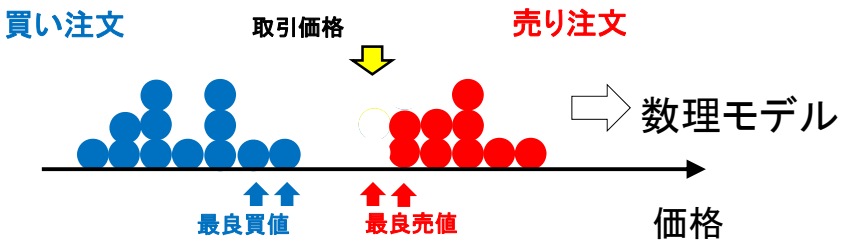
## 外国為替市場の注文の入れ方

買い注文を入れる：買いたい価格と本数を入れる

最良売値よりも安い価格なら、板に積まれる
最良売値以上の価格なら、最良売値で取引成立
キャンセルは、すでにある板注文を消去する

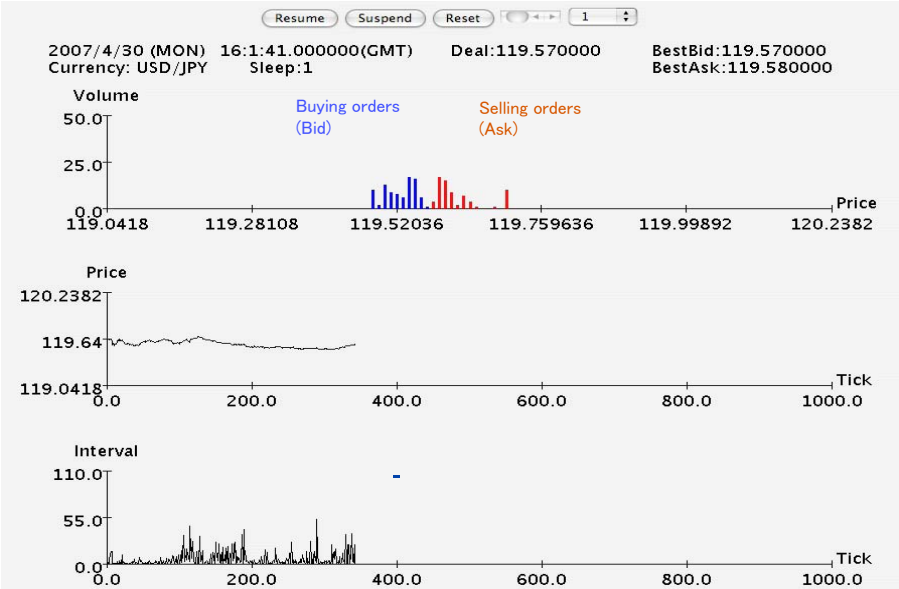
売り注文も同様

売り注文、買い注文、キャンセル注文

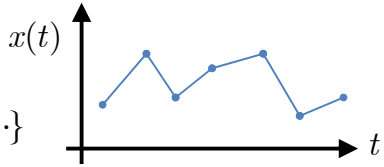


## Order book (EBS USD/JPY)

Unit bar=$1,000,000




## 時系列データのモデリング:




$$\{x(t_0), x(t_1), x(t_2), x(t_3), \cdots, x(t_N), \cdots\}$$

ここでは、ランダムウォークに近い特性を持つ、一般性の高い基本的な時系列データのモデルを紹介する

簡単のため、等間隔( Δt)の時系列データを考える

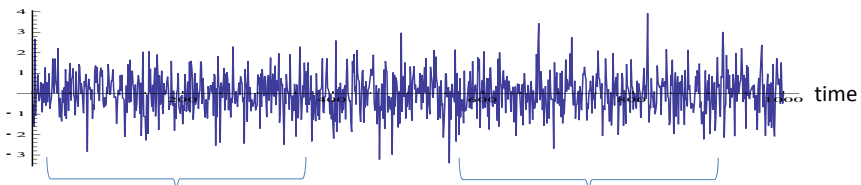
$$\{x(0), x(\Delta t), x(2\Delta t), \cdots, x(N\Delta t), \cdots\}$$

理論解析では、連続的な関数、$\{x(t)\}$ 、を想定する

## 第4回の講義で紹介した定常性を改めて議論する:

定常:時間の原点を移動しても統計的特性が変わらないこと

→ 分布、平均値、分散などが、観測区間に依存しない




観測区間 1

$\{x(0), x(\Delta t), x(2\Delta t), \cdots, x(N\Delta t), \cdots\}$

観測区間 2

$\{x(t'), x(t'+\Delta t), x(t'+2\Delta t), \cdots, x(t'+N\Delta t), \cdots\}$

統計的な性質が同じ

## 定常性の定義

強定常

全ての統計量が時間の原点に依存しない

時系列 $\{x(t)\}$に対して、時刻tで、値 $x$ をとる確率を $P(x,t)$とする
定常ならば、1体の分布が時刻に依存しない

$$P(x,t_1) = P(x,t_2) = P(x)$$

時刻 $t_1$ で値 $x_1$ 時刻 $t_2$ で値 $x_2$ を取る、という2体の分布 $P(x_1,t_1;x_2,t_2)$ も時間の原点に依存しない

$$P(x_1,t_1;x_2,t_2) = P(x_1,t_1+t;x_2,t_2+t)$$

理論的には便利な仮定

高次の分布に関しても同様

現実データでは限度がある

白色雑音は、強定常

白色雑音: $\varepsilon_t$ 各時刻でランダムに $p(\varepsilon)$ から無相関に値が選ばれる

実データで使いやすいのは：

**弱定常**　平均値と標準偏差が時間の原点に依存しない

$$< x(t) >\equiv \mu \ \ (\text{constant})$$

$$< x(t)^2 >_c \equiv \left\langle (x(t)-\mu)^2 \right\rangle = \sigma^2 \ \ (\text{constant})$$

ここで, <....> は、アンサンブル平均
（その意味は、次のスライドで）

観測が何度も繰り返せる、と想定する

Data 1: $\{x^{(1)}(0), x^{(1)}(\Delta t), x^{(1)}(2\Delta t), \cdots, x^{(1)}(N\Delta t), \cdots\}$

Data 2: $\{x^{(2)}(0), x^{(2)}(\Delta t), x^{(2)}(2\Delta t), \cdots, x^{(2)}(N\Delta t), \cdots\}$

...

Data K: $\{x^{(K)}(0), x^{(K)}(\Delta t), x^{(K)}(2\Delta t), \cdots, x^{(K)}(N\Delta t), \cdots\}$

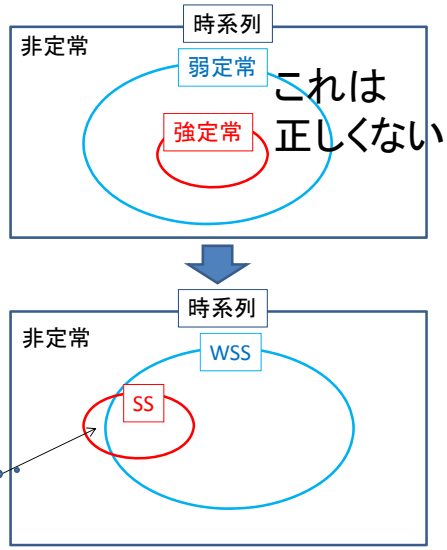
アンサンブル平均は、同じ条件での観測結果のサンプル平均

$$< x(N\Delta t) >\equiv \frac{1}{K}\sum_{k=1}^{K} x^{(k)}(N\Delta t) \qquad < x(N\Delta t)^2 >\equiv \frac{1}{K}\sum_{k=1}^{K} x^{(k)}(N\Delta t)^2$$

弱定常を確認するための統計テストはいろいろとある
弱定常性が確認されたデータに対し、強定常のモデルを適用することが多い

定常でない場合が、非定常
しかし、非定常かどうかの判断には注意が必要

普通は、強定常であれば必ず弱定常であると思われがち。しかし、強定常であるが、弱定常ではない場合がある。




問題: 小テストレベル
強定常であるが、弱定常ではないモデルの例を挙げよ

答え: 平均や分散が発散する分布に基づく白色雑音

例えば、次のような例を簡単に作ることができる

確率変数 $u(t)$を (0,1]の値をとる一様乱数とする
新たな確率変数 $x(t)$ を、この変数の逆数とする

$$x(t) \equiv 1 \,/\, u(t)$$

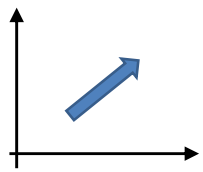
この確率変数の平均値は発散する

$$< x(t) >= \int_0^1 \frac{1}{u}\, du = [\log u]_0^1 = +\infty$$

平均値が発散しているので、弱定常とは言えない
しかし、同じ分布から独立に繰り返しサンプリングして時系列が作られるので、強定常である

理論的に平均値が無限大の場合にはどうなるか？
観測値はいつも有限
サンプル数を増やすと平均値が増加する傾向となり
非定常のように見えがち

サンプル平均



$$< x(t) >_k \equiv \frac{1}{k}\sum_i^k x(i\Delta t)$$

k: サンプル数

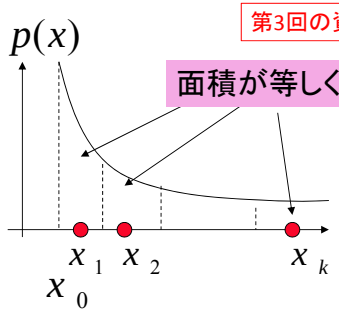
観測期間を長くすると、平均値が増大する傾向
（観測の原点の位置の問題ではない）

ベキ分布の場合にも、強定常になるモデルが作れる:
ベキ分布に従う白色雑音

**サンプル数 $k$ が有限のとき、$n$ 次のモーメントはどれくらいの大きな値になるだろうか？**

第3回の資料

$$< x^n >_k = \frac{1}{k}\sum_{j=1}^{k}(x_j)^n$$




サンプルを小さい順に並べる。

$$\{x_1, x_2, \cdots, x_k\}$$

一様なサンプリングならば、それぞれのサンプル点の周囲の確率は、$k$ 分の1程度になると期待される。

$$P(> x_k) = \left(\frac{x_k}{x_0}\right)^{-\alpha} \approx \frac{1}{k}$$

$$x_k \approx x_0 k^{1/\alpha}$$

$\alpha < n$ のとき

最大値は、右のような形で$k$とともに発散する。

$$< x^n >_k \simeq (x_0)^n k^{\frac{n}{\alpha}-1}$$



平均値が発散する場合の評価の方法

最大の値を与えるサンプルのおおよその値を見積もる

一様乱数の逆数の場合
区間 (0,1]の一様乱数からk個のサンプルを選んだとき
最も小さいサンプルの値は、おおよそいくつになるか？

考え方:
区間 (0,1]の一様乱数からk個のサンプルを選ぶ
区間をk個の区間に等分割、その中の真ん中を
代表サンプルと想定してみる

一番小さい値の区間は、(0,1/k]、その真ん中（中央値）の
サンプルの値は、1/(2k)
一様乱数の逆数の場合、最大サンプルの値は、
およそ2kと見積もれる。サンプル数kに比例して発散。

非定常なのか、平均値の発散なのかは、判断可能
例えば、サンプル数依存性があるかどうか
サンプル数を増やしたとき分布全体が収束しているか
（定常分布で平均値が発散していないか？）

もし、平均値の発散があるなら、変数変換で発散のない量に
変換しておく、という対処も考える
一様分布の逆数の場合は、その逆数なら扱いに問題はない
ベキ分布の場合は対数を取れば、指数分布になり扱いやすい

数理モデルは強定常である方が望ましい
（解きやすいし、理解しやすい）
データから明らかに非定常性がある場合には
その非定常の特性を別途モデル化することを考える
うまく非定常成分を取り除くことができるかもしれない

## 定常過程に変換できるような非定常過程

### 1. 線形トレンド:

$$x(t) \equiv \varepsilon(t) + at$$

$\varepsilon(t)$ : 白色雑音

$a$ : 定数係数

$at$ : トレンド（これ自体は非定常）

トレンドを除去すれば定常モデルとなる　$x'(t) \equiv x(t) - at$

### 2. 周期変動:

$$x(t) \equiv \varepsilon(t) + a\sin(2\pi t / T)$$

ここで、正弦波の代わりに、どんな周期関数が来てもよい
例えば、24時間の活動量の周期変動、曜日に依存した変動 ...

想定される周期変動を探索して見つけ、そのような周期変動を取り除くと、残渣の変動は定常

### 3. 差分時系列:

微分や差分をとると定常になるケース

$$x(t) \equiv \int_0^t \varepsilon(s)ds \quad \text{or} \quad x(N\Delta t) \equiv \sum_{i=1}^{N} \varepsilon(i\Delta t)$$

$\varepsilon(t)$ : 白色雑音

例えば、ランダムウォーク

$$\frac{d}{dt}x(t) = \varepsilon(t) \quad \text{or} \quad x(N\Delta t) - x((N-1)\Delta t) = \varepsilon((N-1)\Delta t)$$

時系列モデリングの基本的な考え方:
非定常の一般論はまだないので、定常過程に基づいてモデルを作る
1 : 取り除ける非定常効果は全て取り除き
2 : 残った変動をできるだけ単純で解きやすいモデルで近似する

---

定常の場合、時間依存性のない分布を想定できる

$(x,t)$　　$p(x,t)$

定常分布を想定する　$p(x)$

平均値、$< x >$、分散、$< x^2 >_c$ 、は時間に依存しない

2体の分布関数も、時間差だけの関数と想定できる

$p(x_1, t; x_2, t+T)$ : 時刻、$t$ 、に値、$x_1$ 、を観測し、
さらに、時刻、$t+T$、に値、$x_2$ を観測する確率

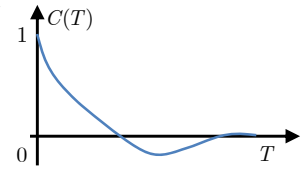
定常であれば、2体の分布は時間差だけの関数となる

$p(x_1, x_2; T)$

---

弱定常なら、下記の自己相関関数が導入できる
（平均と分散が有限となる）

$$C(T) \equiv \frac{< \{x(t) - \mu\}\{x(t+T) - \mu\} >}{\sigma^2}$$




この関数は [-1,+1]の値をとり、
$C(0) = 1$

$C(T) = 1$　完全相関 ( $x(t) = x(t+T)$ )

$C(T) = -1$　反相関 ( $x(t) = 2\mu - x(t+T)$ )

$C(T) = 0$　無相関 ($< \{x(t) - \mu\}\{x(t+T) - \mu\} >= 0$)

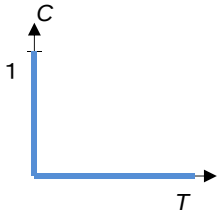
自己相関は弱定常の場合にしか適用できないことに注意

## 相関関数の基本的なふるまい

小テストレベル

$$C(T) \equiv \frac{<\{x(t)-\mu\}\{x(t+T)-\mu\}>}{\sigma^2}$$

**無相関**



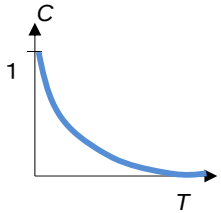

$$C(T) = \begin{cases} 1 & T = 0 \\ 0 & T \neq 0 \end{cases}$$

**無相関**

**すぐに0になる**

**白色雑音**

**短相関**



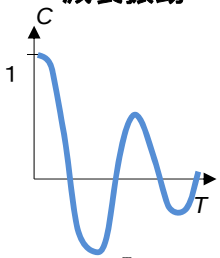

$$C(T) = e^{-\frac{T}{\tau}}$$

$\tau$ **:特性時間**

**指数減衰**

**片対数で直線**

**減衰振動**



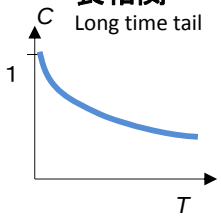

$$C(T) = e^{-\frac{T}{\tau}} \cos(w_0 T)$$

**$\omega_0$ :特性振動数**

**付きの指数減衰**

**長相関**

Long time tail




$$C(T) \propto T^{-\alpha}$$

**長時間相関**

**ロングテール**

**両対数プロットで直線**



## 自己相関関数と対になるパワースペクトルの紹介

パワースペクトルは理論でもデータ解析でも使える便利なツール

$$S(f) \equiv <|\tilde{x}(f)|^2>$$

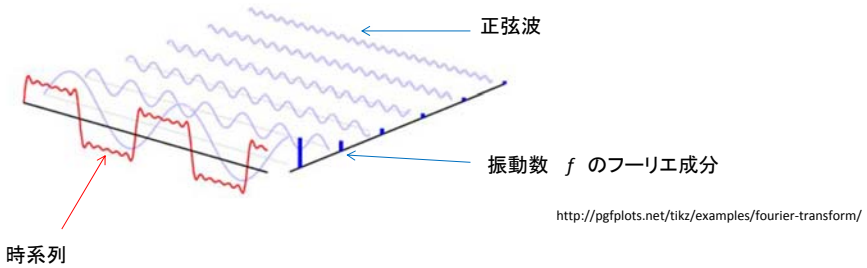
ここで, $\tilde{x}(f)$ は時系列 $\{x(t)\}$ のフーリエ変換

$$\tilde{x}(f) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)\,\mathrm{e}^{-ift}\,dt$$

時系列が離散的な場合は、次のように離散的に定義される

$$x[n] = \Delta t x(n\Delta t) \qquad \tilde{x}_{\Delta t}(f) \equiv \sum_{n=-\infty}^{\infty} x[n]\,\mathrm{e}^{-ifn}$$

## フーリエ変換によって、与えられた時系列を正弦波成分に分解する




http://pgfplots.net/tikz/examples/fourier-transform/

$|x(f)|$ 振動数 $f$ 成分の波の振幅

$|x(f)|^2$ その波のエネルギーに比例する量

パワースペクトル、とよぶ

## 典型的なパワースペクトル

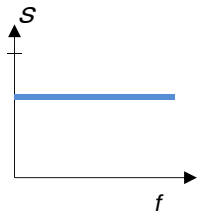
$$S(f) \equiv <|\tilde{x}(f)|^2> \qquad \tilde{x}(f) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{-ift}dt$$

パワースペクトルは

偶関数

$S(f) = S(-f)$

無相関

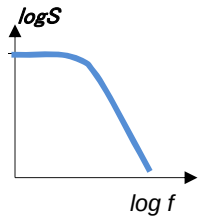



$S(f) = const.$

**白色雑音**

**（あらゆる振動数の成分を等しく持つ）**

短相関

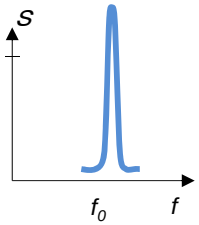



$$S(f) \propto \frac{1}{1+(\tau f)^2}$$

$\tau$ **:特性時間**

**fの大きいところはベキ減衰　1/f²**

振動




$S(f) \approx \delta(f - f_0)$

**$f_0$ :特性振動数**

長相関

$S(f) \propto f^{-\beta}$

**ベキ減衰**

**両対数プロットで直線的**

## ウィーナー・ヒンチンの定理

弱定常過程の場合、パワースペクトルは自己相関数のフーリエ変換で得られ、逆フーリエ変換で、パワースペクトルから自己相関数を求められる（両者は同じもの）

$$\int_{-\infty}^{\infty} C(T)e^{-ifT}dT = \frac{2\pi}{\sigma^2}S(f)\ ,\ \int_{-\infty}^{\infty} S(f)e^{ifT}df = \sigma^2 C(T)$$

$$\int_{0}^{\infty} C(T)\cos(fT)dT = \frac{\pi}{\sigma^2}S(f)$$

証明は次のページ



## ウィーナー・ヒンチン定理の証明

$$< \tilde{x}(f)\tilde{x}(f') >$$

$$=< \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t)x(t')e^{-ift}e^{-if't'}dtdt' >$$

$$= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} < x(t)x(t') > e^{-i(ft+f't')}dtdt'$$

$$= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{< (x(t)-\mu)(x(t')-\mu) > +\mu < x(t)+x(t') > -\mu^2\}e^{-i(ft+f't')}dtdt'$$

$$= \frac{\sigma^2}{4\pi^2}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(t-t')e^{-if(t-t')-i(f+f')t'}dtdt' + 4\pi^2\mu^2\delta(f)\delta(f')\}$$

$$= \frac{\sigma^2}{4\pi^2}\int_{-\infty}^{\infty} C(T)e^{-ifT}dT\int_{-\infty}^{\infty} e^{-i(f+f')t'}dt' + \sigma^2\mu^2\delta(f)\delta(f')$$

$$= \frac{\sigma^2}{4\pi^2}\int_{-\infty}^{\infty} C(T)e^{-ifT}dT\cdot 2\pi\delta(f+f') + \sigma^2\mu^2\delta(f)\delta(f')$$

デルタ関数の基本性質

$$< \int_{-\infty}^{\infty} e^{-ift}dt >= 2\pi\delta(f)$$

$$\mu \equiv < x(t) >$$

$T = t - t'$ 弱定常

平均値は振動数0の成分のみに寄与

一方、

$$< \tilde{x}(f)\tilde{x}(f') >=< \tilde{x}(f)\tilde{x}(-f) > \delta(f+f') =<| \tilde{x}(f) |^2> \delta(f+f') = S(f)\delta(f+f')$$

振動数 f≠0 に対し、次の関係が成り立つ

$$\tilde{x}(-f) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{ift}dt = \tilde{x}(f)^*$$

複素共役

$$\int_{-\infty}^{\infty} C(T)e^{-ifT}dT = 2\int_{0}^{\infty} C(T)\cos(fT)dT = \frac{2\pi}{\sigma^2}S(f)$$



## Autoregressive models 自己回帰モデル

次のような線形な時系列モデルを考える

AR(n) モデル , n: 自然数

$$X_t = c + \sum_{i=1}^{n}\omega_i X_{t-i} + \varepsilon_t$$

ここで

$\omega_1, \cdots, \omega_n$ は重みを表すパラメータ,

$c$ 定数,

$\varepsilon_t$ 白色雑音

白色雑音（無相関な正規分布を仮定することが多い）.

$$< \varepsilon_t \varepsilon_{t'} >= \sigma_\varepsilon{}^2 \delta_{t,t'}$$

第6回の講義で出てきたランジュバン方程式も差分では、ARモデル になる

小テストレベル

$$m\frac{d}{dt}v(t) = -\nu v(t) + F(t)$$

## 最も簡単な場合, AR(1) 、c=0.

$$X_t = \omega_1 X_{t-1} + \varepsilon_t$$

$\varepsilon_t$ 白色雑音

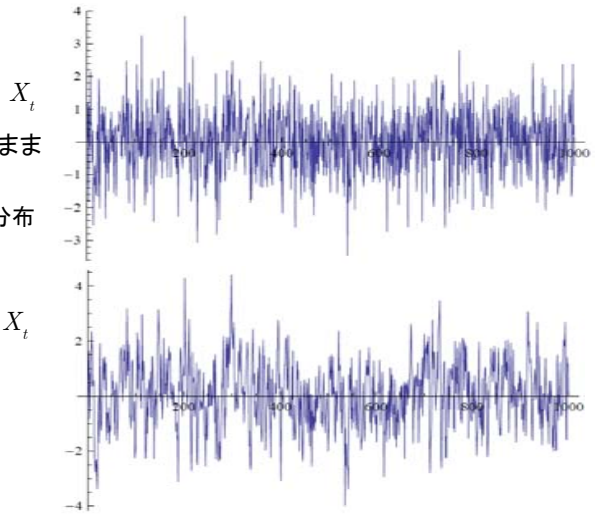
例えば、:

$\omega_1 = 0$ 白色雑音そのまま

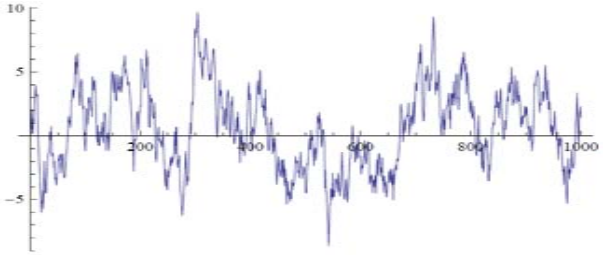
N(1,0): 平均0分散1の正規分布

$< \varepsilon >= 0, < \varepsilon^2 >= 1$

$\omega_1 = 0.6$

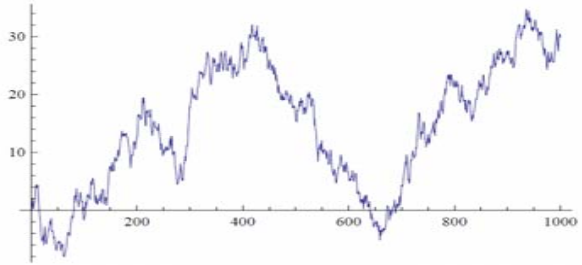


$\omega_1 = 0.95$



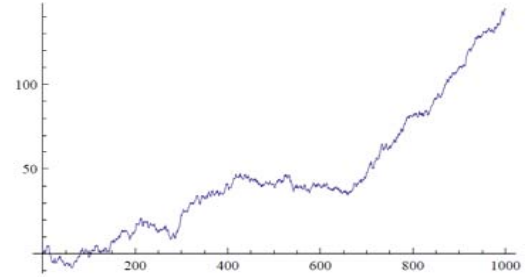
$\omega_1 = 1.0$

ランダムウォーク



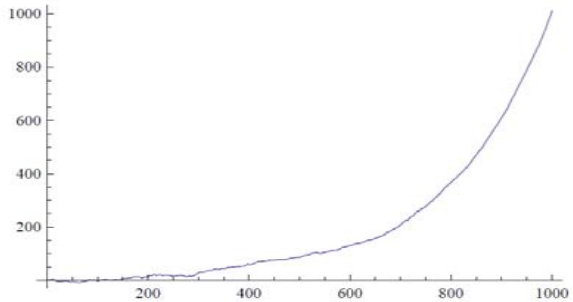
$\omega_1 = 1.0025$

発散傾向



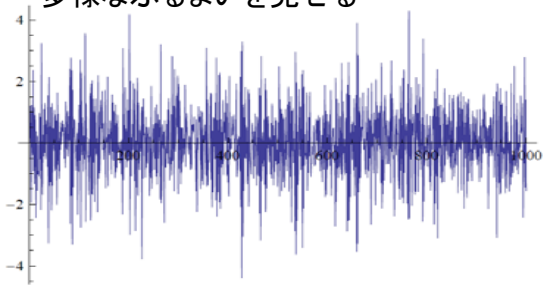
$\omega_1 = 1.005$

指数的な発散

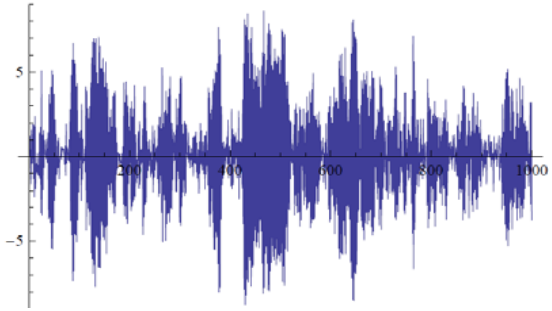


負の場合　$\omega_1$：　非常に単純なモデルだが、
多様なふるまいを見せる

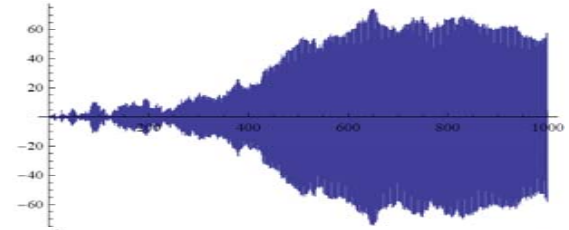
$\omega_1 = -0.6$



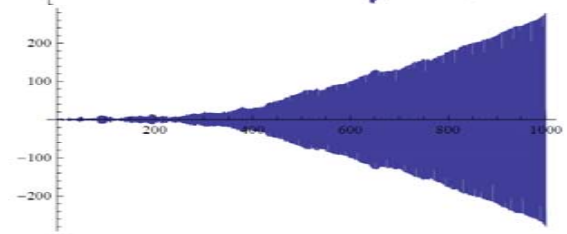
$\omega_1 = -0.95$



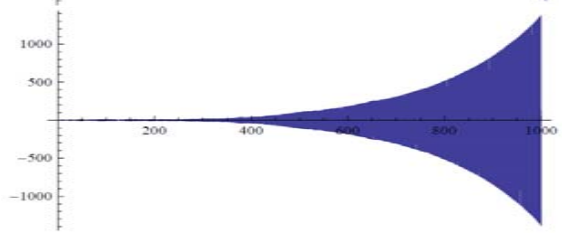
$\omega_1 = -1.0$



$\omega_1 = -1.0025$



$\omega_1 = -1.005$



AR(1)モデルで作り出される時系列が（弱）定常になる条件

$-1 < \omega_1 < 1$

証明:

小テストレベル

$$X_t = \omega_1 X_{t-1} + \varepsilon_t$$
$$= \omega_1(\omega_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$
$$= \omega_1^2(\omega_1 X_{t-3} + \varepsilon_{t-1}) + \omega_1\varepsilon_{t-1} + \varepsilon_t$$
$$\vdots$$

代入の繰り返し

$$= \omega_1^t\varepsilon_0 + \omega_1^{t-1}\varepsilon_1 + \cdots + \omega_1\varepsilon_{t-1} + \varepsilon_t$$
$$= \sum_{s=0}^{t}\omega_1^{t-s}\varepsilon_s$$

$$< X_t >= \sum_{s=0}^{\infty}\omega_1^{t-s} < \varepsilon_0 >= 0$$

$$X_{t+T} = \sum_{s=0}^{t+T}\omega_1^{t+T-s}\varepsilon_s$$

$$< X_t X_{t+T} >= \omega_1^{2t+T} < \varepsilon_0^2 > + \omega_1^{2(t-1)+T} < \varepsilon_1^2 > + \cdots + \omega_1^T < \varepsilon_t^2 >$$
$$= \sigma_\varepsilon^2\omega_1^T\frac{\omega_1^{2t}-1}{\omega_1^2-1}$$

等比級数の和

ここで、$-1 < \omega_1 < 1$ 、ならば、右辺は有限の値に収束

$\omega_1^{2t} \to 0$

$$< X_t X_{t+T} >= \frac{\sigma_\varepsilon^2}{1-\omega_1^2}\omega_1^T$$

時間差 T のみの関数で、時刻 t を含まない

ギリギリのパラメータ $\omega_1 = 1$ ではランダムウォーク

$$X_t = \varepsilon_0 + \varepsilon_1 + \cdots + \varepsilon_{t-1} + \varepsilon_t$$
$$< X_t >= 0$$

$$X_{t+T} = \varepsilon_0 + \varepsilon_1 + \cdots + \varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1} + \cdots + \varepsilon_{t+T}$$
$$< X_t X_{t+T} >=< \varepsilon_0^2 > + < \varepsilon_1^2 > + \cdots + < \varepsilon_t^2 >$$
$$= (t+1)\sigma_\varepsilon^2$$
$$< X_t^2 >= (t+1)\sigma_\varepsilon^2$$

分散は、時間に比例して増加する
ランダムウォークは非定常、その差分は定常

$\omega_1 > 1$ のとき、時系列は非定常で指数発散

$$X_t = \omega_1^t\varepsilon_0 + \omega_1^{t-1}\varepsilon_1 + \cdots + \omega_1\varepsilon_{t-1} + \varepsilon_t$$
$$< X_t >= 0$$
$$< X_t^2 >= \sigma_\varepsilon^2\frac{\omega_1^{2t}-1}{\omega_1^2-1} \approx \frac{\sigma_\varepsilon^2}{\omega_1^2-1}\omega_1^{2t}$$

右辺の見積もりは、時間差でなく、原点からの経過時間とともに
指数的に発散する
時間の原点の位置に依存するので、典型的な非定常

もうひとつのギリギリの場合 $\omega_1 = -1$

$$X_t = (-1)^t\varepsilon_0 + (-1)^{t-1}\varepsilon_1 + \cdots + (-1)\varepsilon_{t-1} + \varepsilon_t$$
$$< X_t >= 0$$
$$< X_t^2 >= t\sigma_\varepsilon^2$$

ランダムウォークの場合と同じく、時間差の関数ではなく、
原点からの時間の関数：非定常

$\omega_1 < -1$ の場合、振動しながら発散する
分散は、指数関数的に発散

$$< X_t^2 >= \sigma_\varepsilon^2\frac{\omega_1^{2t}-1}{\omega_1^2-1} \approx \frac{\sigma_\varepsilon^2}{\omega_1^2-1}\omega_1^{2t}$$

項の数を増やした一般の AR(n) モデルの定常条件は
次のように考えられる

$$X_t = \sum_{i=1}^{n} \omega_i X_{t-i} + \varepsilon_t$$

確率変数を除いた決定論的な斉次多項式を考える

$$X_t = \sum_{i=1}^{n} \omega_i X_{t-i}$$

線形方程式なので、次のように一般解を求められる

$$X_t = \sum_{i=1}^{n} c_i \lambda_i^t$$

ここで、$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$、は、次の特性方程式の根

$$\lambda^n = \sum_{i=1}^{n} \omega_i \lambda^{n-i}$$

$\{c_1, c_2, \cdots, c_n\}$ は、初期値を満たすように決められる

$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ は、特性根、固有値とよばれる

特性根の絶対値の最大値, $\Lambda \equiv \max\{|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|\}$ ,
がAR（n）の挙動を決める

$$X_t = \sum_{i=1}^{n} \omega_i X_{t-i} + \varepsilon_t$$

小テストレベル

$\Lambda < 1$ ：AR(n) は定常

$\Lambda = 1$ ：AR(n) は非定常
ランダムウォークの場合
（単位根とよばれる）

$\Lambda > 1$ ：AR(n) は非定常
絶対値をとった量, $|X_t|$, は指数発散

AR(n) モデルは、白色雑音、ランダムウォーク、振動、
指数発散など、さまざまな状態を記述できる汎用性の
高い時系列モデル

平均値が0であるような定常的な時系列データに対して、
最適な自己回帰モデルを導入する方法

与えられた時系列 $\{x_0, x_1, x_2, x_3, \cdots, x_t, \cdots\}$

AR（n）モデル $X_t = \sum_{i=1}^{n} \omega_i X_{t-i} + \varepsilon_t$

誤差を小さく $< (X_t - x_t)^2 >$

nは幾つにすればいいか、重みはどう決めればいいか？
自己相関関数（自己共分散）から決められる

$$C_T^* \equiv \frac{1}{N} \sum_{t=0}^{N} x_t x_{t+T}$$

自己相関関数はデータから
求める

$$X_t = \sum_{i=1}^{n} \omega_i X_{t-i} + \varepsilon_t$$

$X_0$ を AR(n) にかける

$$X_T X_0 = \sum_{i=1}^{n} \omega_i X_{T-i} X_0 + \varepsilon_T X_0$$

for T=1,2,…,n-1

白色雑音に関する平均をとる $\{\varepsilon_t\}$ .

$$< X_T X_0 > = \sum_{i=1}^{n} \omega_i < X_{T-i} X_0 > + < \varepsilon_T X_0 >$$

0

$$C_T = \sum_{i=1}^{n} \omega_i C_{T-i}$$

モデルから計算される自己相関
関数の満たす関係式

データとモデルが整合する条件、$C_T = C_T^*$ 、T=1,2,…,n-1

n個のパラメータに対し、n-1 個の方程式 $\{\omega_1, \omega_2, \cdots, \omega_n\}$

もうひとつの方程式は、$C_0$ 、に対して求める

$$C_0 =< X_t^{\,2} >$$

$$=< X_t(\sum_{i=1}^{n}\omega_i X_{t-i} + \varepsilon_t) >$$

$$= \sum_{i=1}^{n}\omega_i < X_t X_{t-i} > + < (\sum_{i=1}^{n}\omega_i X_{t-i} + \varepsilon_t)\varepsilon_t >$$

$$= \sum_{i=1}^{n}\omega_i C_i + \sigma_\varepsilon^{\,2} \qquad < \varepsilon_t^{\,2} >= \sigma_\varepsilon^{\,2}$$

これで、未知数の数と方程式の数が一致（ユール・ウォーカー法）

$$C_0 = \sum_{i=1}^{n}\omega_i C_i + \sigma_\varepsilon^{\,2}$$

$$C_T = \sum_{i=1}^{n} a_i C_{T-i}$$

$$\longrightarrow \begin{pmatrix} C_0 & C_1 & \cdots & C_{n-1} \\ C_1 & C_0 & \cdots & C_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n-1} & C_{n-2} & \cdots & C_0 \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}$$

逆行列を解けば、重み $\{\omega_1, \omega_2, \ldots, \omega_n\}$ が求められる

## レビンソンのアルゴリズムという便利な解法が知られている

次のような代入の繰り返しによって定義される。（詳細は省略）

$$C_0 =< v(t)^2 >$$

n=1 から順に大きくして求めていく

$$\omega_n^{(n)} = \frac{C_n - \sum_{k=1}^{n-1}\omega_k^{(n-1)} C_{n-k}}{(\sigma^{(n-1)})^2} \qquad \text{(7A)}$$

$$(\sigma^{(n)})^2 = (\sigma^{(n-1)})^2\{1 - (\omega_n^{(n)})^2\} \qquad \text{(7B)}$$

$$\omega_k^{(n)} = \omega_k^{(n-1)} - \omega_{n-k}^{(n-1)}\omega_n^{(n)} \qquad \text{(7C)}$$

$$C_0 = (\sigma^{(0)})^2$$

↓ (7A) n=1

$$a_1^{(1)}$$

↓ (7B) n=1

$$(\sigma^{(1)})^2$$

↓ (7A) n=2

$$a_2^{(2)}$$

↓ (7B) n=2

$$(\sigma^{(2)})^2$$

↓ (7C) n=2

$$a_1^{(2)} \Rightarrow \text{n=3,4,}$$

これらが、次のユール・ウォーカー方程式を満たすことは代入によって確認できる（ここでは、省略）

$$(\sigma^{(n)})^2 = C_0 - \sum_{k=1}^{n}\omega_k^{(n)} C_k \qquad C_j = \sum_{k=1}^{n}\omega_k^{(n)} C_{j-k}$$

ユール・ウォーカー法により推定された重みを用いた誤差の評価

$$< (X_t - x_t)^2 >$$

$$=< \{X_{t+\Delta t} - \sum_{k=1}^{n}\omega_k x_{t-(k-1)\Delta t}\}^2 >$$

$$= C_0 - 2\sum_{k=1}^{n} a_k C_k + \sum_{k=1}^{n}\sum_{j=1}^{n} a_k a_j C_{k-j}$$

この誤差の分散を重みのパラメータ、$\{a_j\}$ で微分すると

$$\frac{\partial < (X_t - x_t)^2 >}{\partial a_j} = -2\{C_j - \sum_{k=1}^{n} a_k C_{j-k}\} \qquad C_j = \sum_{k=1}^{n} a_k C_{j-k}$$

となるので、ユール・ウォーカー方程式を満たすときには、この微分が0になる。すなわち、**ユール・ウォーカー方程式で重みを決めれば、予測誤差の分散は最小になる。**

## 実データ解析への適用方法まとめ

1：与えられた時系列 $\{x_0, x_1, x_2, x_3, \cdots, x_t, \cdots\}$ から

2：相関関数 $\{C_T\}$ を求める（十分大きなTまで）

3：ユール・ウォーカー方程式をレビンソンのアルゴリズムで解く

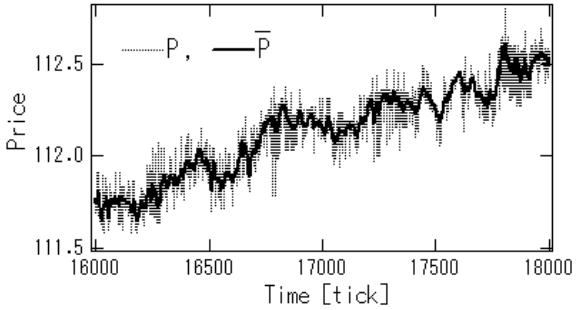
AR（n）モデルの重み、$\{\omega_k\}$、が小さいnから順次求められる

4：異なるnで、誤差の大きさを比較し、 $< (X_t - x_t)^2 >$

誤差が十分小さいなかで最小のnを採用する

この手法は、**定常的な時系列**に対して、システマチックに最適な**移動平均をとる手法**として、非常に強力

## 外国為替レート時系列の生データと ARモデルによる最適移動平均曲線




ノイズ除去実現！

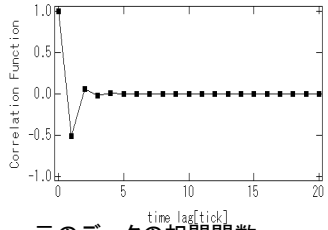
$$P(t) = \overline{P(t)} + f(t)$$

$$\overline{P(t)} = \sum_{k=0}^{K} \omega_k \cdot P(t-k)\ ,\ \ \langle f(t) f(t+T) \rangle \simeq 0$$

## 移動平均を差し引いた残差 $\delta P(t)$ の相関はほぼ完全に0になる

$$\Delta P(t) = P(t+1) - P(t)$$

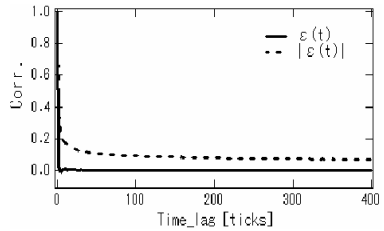
$$\langle \Delta P(t) \Delta P(t+T) \rangle \neq 0$$




元のデータの相関関数

$$\delta P(t) = P(t) - \overline{P(t)} = F(t)$$

$$\langle \delta P(t) \delta P(t+T) \rangle = 0$$




残差成分の相関関数

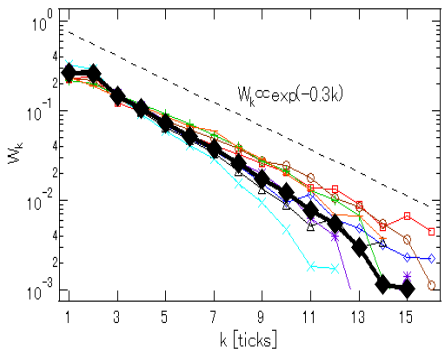
ノイズ除去実現！

$$P(t) = \overline{P(t)} + f(t)$$

$$\overline{P(t)} = \sum_{k=0}^{K} \omega_k \cdot P(t-k)\ ,\ \ \langle f(t) f(t+T) \rangle \simeq 0$$

## 円ドル市場での最適移動平均の重み関数 $\omega_k$

(年度ごとに最適移動平均の重み関数をプロット)




(◇:1995, □:1996, △:1997, ×:1998, ＊:1999,
○:2000, ＋:2001, ー:2002(Jan-Feb), ◆:1995-2002(Feb))

いつでも、ほぼ、平均値30秒程度の指数関数的な重みが付いていることがわかる

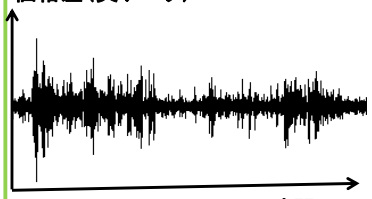
数分で重みはほとんど0になる

ディーラーは せいぜい過去 数分間しか見 ていない！？

## 金融工学の時系列モデル（ARCH・GARCH）

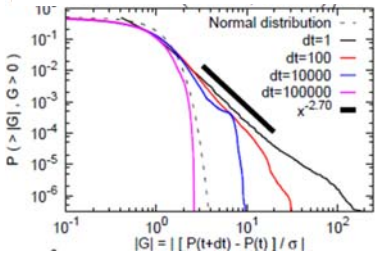
✓価格差の時系列は塊（Cluster）になりやすい。
✓価格差はべき分布
これらの特性を取り込むために、ARCH・GARCHモデルが導入されている。

価格差（実データ）

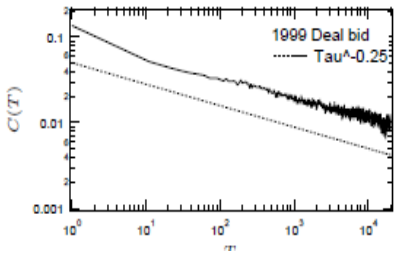


Tick時間

価格差のべき分布（実データ）




Volatilityの自己相関関数（実データ）



# ARCH, GARCHモデル

Generalized Auto-Regressive
Conditional Heteroscedasticity

**2003年　ARCHモデル（1982）を提案したRobert Engleは**
**ノーベル経済学賞受賞。**

$$P(t+\Delta t) = P(t) + \sigma(t+\Delta t)\cdot f(t+\Delta t)$$

$$\sigma(t+\Delta t)^2 = \sigma_0^{\ 2} + c_1\{P(t)-P(t-\Delta t)\}^2 + d_1\sigma(t)^2$$
$$= \sigma_0^{\ 2} + (c_1\cdot f(t)^2 + d_1)\sigma(t)^2$$

**価格と分散の連立自己回帰モデル**

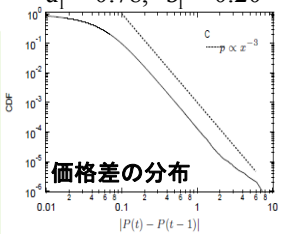
$\{f(t)\}$　：平均0、分散1の正規分布に従う独立な乱数

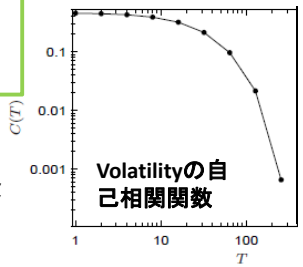
$d_1 = 0$　：ARCHモデル

ARCHモデルに大きな分散が持続しやすい効果を付加。

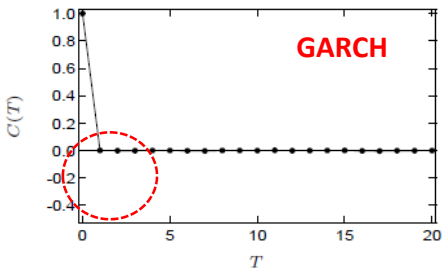
ただし、価格差の自己相関はARCHと同じく指数減衰となる。

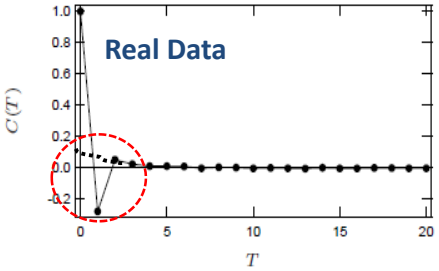
$a_1 = 0.78,\quad b_1 = 0.20$



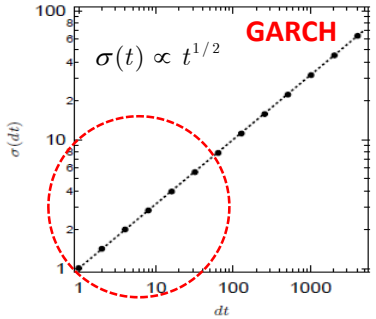



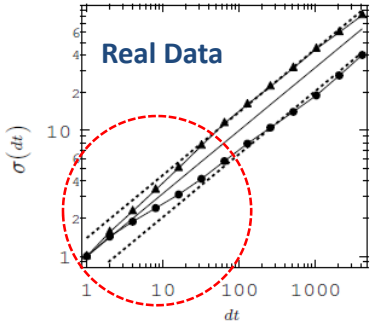

## 価格差の自己相関関数




## 価格の拡散



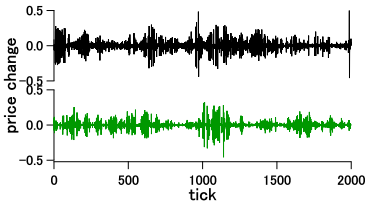

## しかし、現実の市場は単純なランダムウォークではない

**ボラティリティクラスタリング**

**価格差のゆらぎ幅が相関を持って塊になって変動する性質**

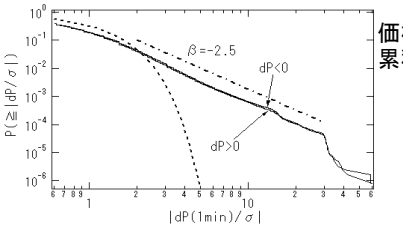
**⇒金融工学のARCHモデル**

**⇒経済物理学のモデル**



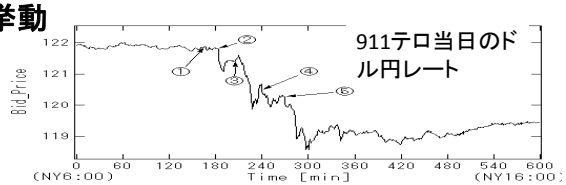

**価格差分布がファットテール**

**（価格差はベキ分布で近似できる）**




**暴騰暴落時には動力学挙動**

**（確率を超えた動きを見せる）**




## スケジュール　（小テスト＝最後の25分で基本確認）

| | | |
|---|---|---|
| 1 モデリングとは | 9月２６日木 | |
| 2 現象の観測と基本モデル１ | ９月３０日月 | |
| 3 現象の観測と基本モデル２ | １０月　３日木 | |
| 4 現象の観測と基本モデル３ | １０月　７日月 | 小テスト |
| 5 拡散現象のモデリング１ | １０月１７日木 | |
| 6 拡散現象のモデリング２ | １０月２１日月 | |
| 7 拡散現象のモデリング３ | １０月２４日木 | |
| 8 拡散現象のモデリング４ | １０月２８日月 | 小テスト |
| 9 分枝・凝集現象のモデリング１ | １０月３１日木 | |
| 10分枝・凝集現象のモデリング２ | １１月　４日月（振替休日） | 小テスト |
| 11相転移現象のモデリング１ | １１月　７日木 | |
| 12相転移現象のモデリング２ | １１月１１日月 | |
| 13相転移現象のモデリング３ | １１月１４日木 | 小テスト |
| 14複雑ネットワークのモデリング１ | １１月１８日月 | |
| 15複雑ネットワークのモデリング２ | １１月２１日木 | 小テスト |
| 予備日　１１月２５日月 | １１月２８日木 | |

2019

9 September

10 October

11 November

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 敬老の日 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 秋分の日 | 24 | 25 | 26 1 | 27 | 28 |
| 29 | 30 2 | 1 | 2 | 3 3 | 4 | 5 |
| 6 | 7 4 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 体育の日 | 15 | 16 | 17 5 | 18 | 19 |
| 20 | 21 6 | 22 即位礼正殿の儀 | 23 | 24 7 | 25 | 26 |
| 27 | 28 8 | 29 | 30 | 31 9 | 1 | 2 |
| 3 文化の日 | 4 10 | 5 | 6 | 7 11 | 8 立冬 | 9 |
| 10 | 11 12 | 12 | 13 | 14 13 | 15 七五三 | 16 |
| 17 | 18 14 | 19 | 20 | 21 15 | 22 | 23 勤労感謝の日 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |