

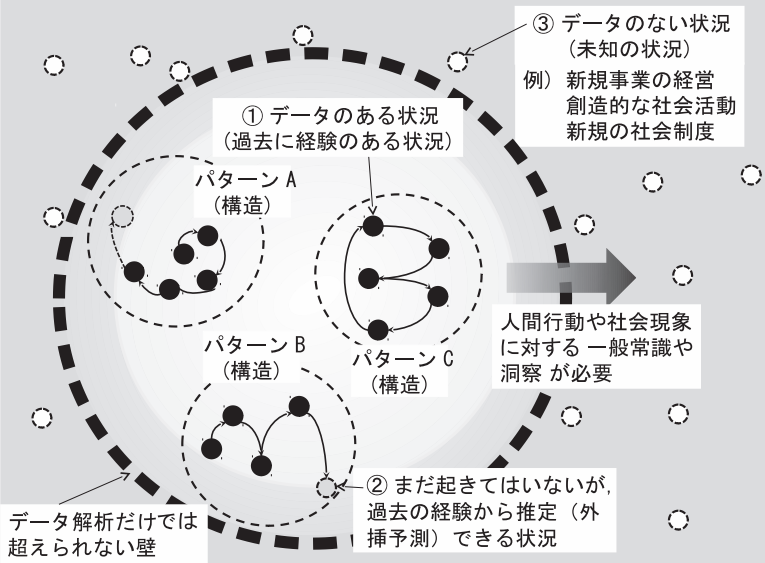
大規模データで社会を予測できるか？

ツイッターなどのソーシャルデータによる社会予測の成功例

- インフルエンザの流行 (Ginsberg ほか, 2009)¹⁰⁾
- 雇用状況 (Ettredge ほか, 2005)⁶⁾
- 映画の興行収入 (Asur ほか, 2010)¹⁾
- 株式市場 (Bollen ほか, 2011)²⁾
- 選挙結果 (Tumasjan ほか, 2010)³⁹⁾ など

しかし、最近ではソーシャルデータによる予測能力に疑問も…

- 音楽やビデオゲーム、映画のヒットの予測：
従来からの単純な回帰分析に勝てない (Goel ほか, 2010)¹¹⁾
- 選挙予測：
世論調査会社の予測に勝てない (Gayo-Avello, 2011)⁷⁾



社会現象において既存データからの 外挿予測が困難な理由

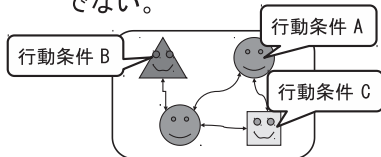
- ① 構成要素である個人行動
が複雑



- ・さまざまな行動条件
- ・ルールが変化

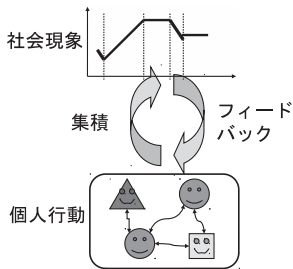
行動の基本方程式が存在しない。

- ② 要素間で行動ルールが均質
でない。



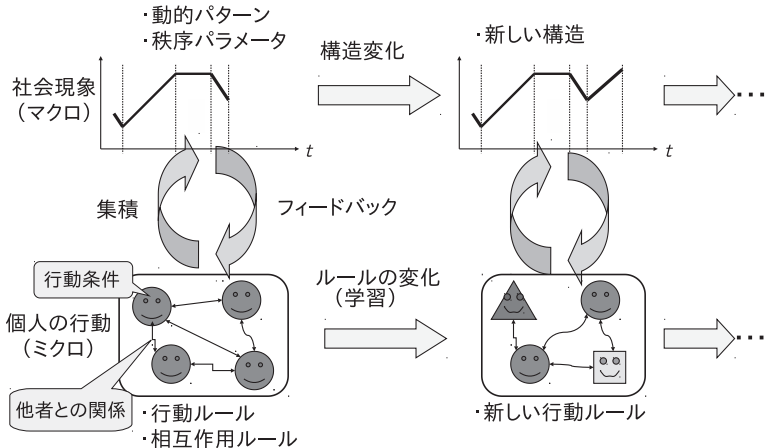
現象全体でみるべき条件の
組合せ数 $\{A \times B \times C\}$ が増大

- ③ ミクロ・マクロループが
存在

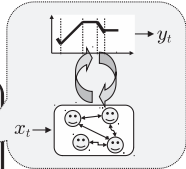
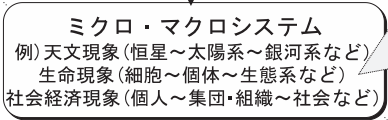
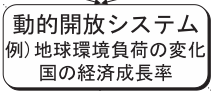
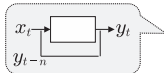
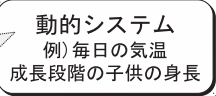
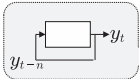


ミクロな個人行動とマクロな
社会現象が相互に影響

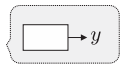
マイクロ・マクロ ループ



システムの分類



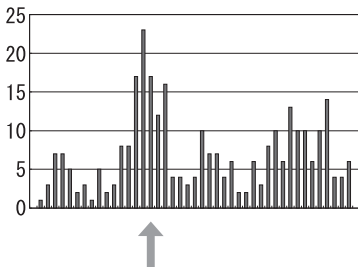
閉鎖システム



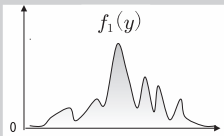
観測データ y

10	5回
20	3回
30	6回
40	13回
50	23回
60	16回
70	3回
80	9回

統計分布・基本統計量

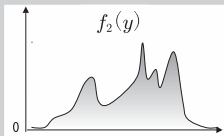


システムのモデル化: 確率密度関数 $f(y)$ の推定



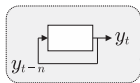
+

混合



正規分布? べき分布?...

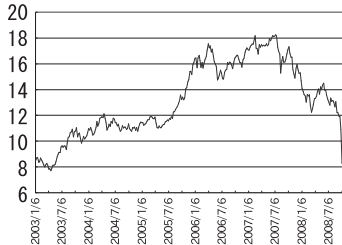
動的システム



時系列データ y_t

1月 6日	8 470円
1月13日	8 690円
1月20日	8 731円
1月27日	8 339円
2月 3日	8 448円
2月10日	8 701円
2月17日	8 513円
2月24日	8 363円

× 10⁵ 例) 週次の株価データ



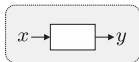
システムモデル化: 時系列解析

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-r})$$

例) 自己回帰(AR)モデル, 移動平均(MA)モデル,
自己回帰和分移動平均(ARIMA)モデルなど

入力データと出力データ
例) 経済指標 x から株価 y を説明する。

開放システム



指標 x_1	指標 x_2	指標 x_3
9.42 %	-6.23 %	2.60
22.8 %	-5.13 %	2.10
10.8 %	-9.16 %	1.87
11.7 %	-10.4 %	0.59
8.50 %	-9.89 %	2.00
12.6 %	-12.7 %	1.34

株価 y
8 470円
8 690円
8 731円
8 339円
8 448円
8 701円

システムのモデル化: 入出力関係

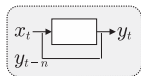
$$y = f(x_n)$$

例1) 線形回帰分析

例2) ニューラルネットワーク

入力時系列データと出力時系列データ
例) 経済指標 x_t から株価 y_t を説明

動的開放システム



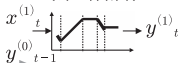
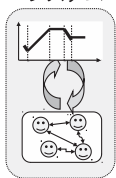
	指標 $x_{1,t}$	指標 $x_{2,t}$	指標 $x_{3,t}$	株価 y_t
1月 6日	9.42 %	-6.23 %	2.60	8 470円
1月13日	22.8 %	-5.13 %	2.10	8 690円
1月20日	10.8 %	-9.16 %	1.87	8 731円
1月27日	11.7 %	-10.4 %	0.59	8 339円
2月 3日	8.50 %	-9.89 %	2.00	8 448円
2月10日	12.6 %	-12.7 %	1.34	8 701円

システムのモデル化: 入出力関係

$$y_t = f(x_t, x_{t-1}, \dots, x_{t-\tau_1}, y_{t-1}, y_{t-2}, \dots, y_{t-\tau_2})$$

例) 外生入力を持つ時系列モデル
回帰的ニューラルネットワーク

マイクロ・マクロ
システム



マクロ
(第1階層)

ミクロ
(第0階層)



	$y^{(0)}_t$	$x^{(1)}_{1,t}$
t_1	3.11	-6.23
t_2	3.27	-5.13
t_3	2.80	-9.16

	$x^{(0)}_{1,t}$	$x^{(0)}_{2,t}$
t_1	9.42	-6.23
t_2	22.8	-5.13
t_3	10.8	-9.16


$y^{(1)}_t$
3.27
2.80
3.01

$y^{(0)}_t$
3.27
2.80
3.01

システムのモデル化：マイクロ・マクロループ

例) エージェントシミュレーション

分析対象のシステム

マクロシステム  $y^{(1)}$

または

ミクロシステム  $y^{(0)}$

出力/観測
データ

変換/
計算

システムの状況を
反映した特徴量
 (z_1, z_2, \dots, z_n)

特徴空間

グループ A

グループ C

グループ B

- システムと外部環境がある程度独立
- システムの挙動が安定的

閉鎖システム
とみなす。

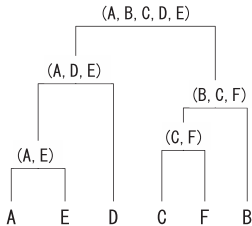
データ分類手法(クラスタ分析など)
を用いて似た状況をグループ分け

クラスタ分析の種類

階層クラスタ分析

類似度の高い組合せから順に
クラスタを作る方法

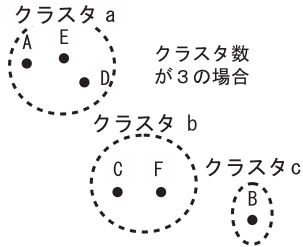
類似度
が低い
組合せ



類似度
が高い
組合せ

非階層クラスタ分析

与えられたクラスタ数に合わせて
クラスタに分ける方法



A



B



C



D



E



F

分析対象のシステム

特徴量
(面積, 明るさ)

A  (0.2, 1.0)

B  (1.2, 0.2)

C  (0.8, 0.3)

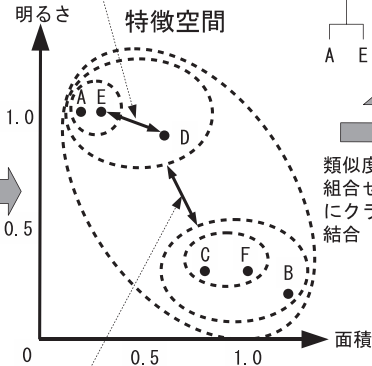
D  (0.6, 0.9)

E  (0.3, 1.0)

F  (1.0, 0.2)

1. 特徴量の決め方

2. 特徴量間の類似度の決め方



3. クラスタ間の距離の決め方

階層クラスタ分析



類似度の高い
組合せから順
にクラスタに
結合

1. 特徴量の決め方

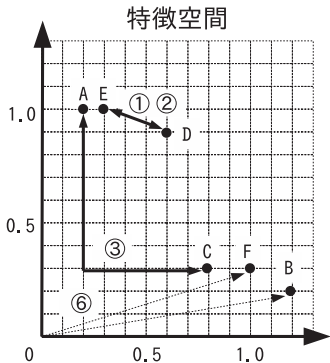
(a) 特徴選択

- 一般的に無関係な種類のデータを除外する。
 - 例) 投票行動での星座 など
- データの事前分析で特徴選択を行う。
 - 目的変数との関連性が一定以上の特徴量を選択 (フィルタ法)
 - 目的変数をうまく推定できる特徴量の組合せを選択 (ラッパ法)

(b) 特徴抽出 (特徴量の合成)

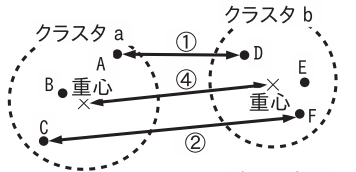
- 社会現象に関係する一般的な指標をデータから計算する。
 - 例) エンゲル係数
- 統計的手法もしくは機械学習を用いて、特徴量を抽出する。
 - 例1) 主成分分析
より少ない種類の合成変数 (主成分) に縮約
 - 例2) オートエンコーダ
ニューラルネットワークを用いて与えられた次元数の特徴量を抽出

2. 特徴量間の類似度(距離)の決め方



- ① ユークリッド距離：日常でよく用いられる2点間の距離
- ② 平方ユークリッド距離：ユークリッド距離の値を2乗したもの
- ③ マンハッタン距離：各次元の座標の差の絶対値の和
- ④ ミンコフスキー距離：ユークリッド距離の一般化
- ⑤ マハラノビス距離：変数間の相関や分散を考慮して一般化した距離
- ⑥ コサイン類似度：ベクトルどうしのなす角度の近さ

3. クラスタ間の類似度(距離)の決め方



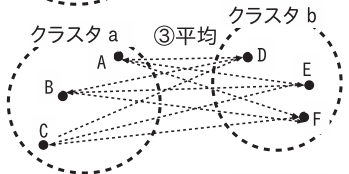
① 最短距離法：二つのクラスタ a, b の要素間の距離の中で最小値

② 最長距離法：二つのクラスタ a, b の要素間の距離の中で最大値

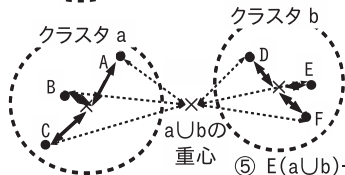
③ 群平均法：二つのクラスタ a, b の要素間の距離の平均値

④ 重心法：二つのクラスタ a, b の重心間の距離

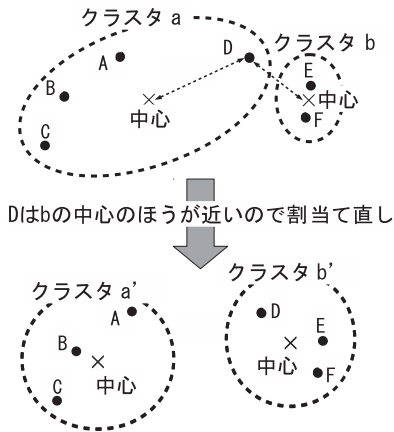
⑤ ウォード法：結合後のクラスタのばらつき(重心との距離の2乗の和)の、結合前のクラスタのばらつきの和に対する増加分



⑤ $E(a \cup b) - (E(a) + E(b))$



非階層クラスタ分析 K-means法の手続き



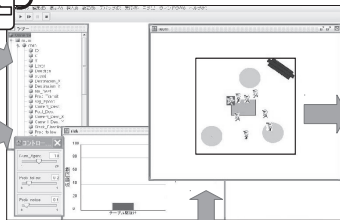
- ① すべての特徴量をランダムに K 個のクラスタに割り振る。
- ② 各クラスタの重心を計算して、それを中心と呼ぶ。
- ③ 自分から一番近いクラスタ中心のクラスタへ割り当て直す。
- ④ 割当ての変化がなくなるまで、上記②と③を繰り返す。

データ解析とシミュレーションの統合

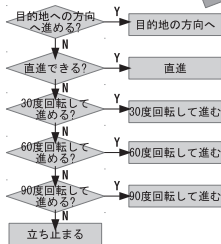
エージェント移動モデル



エージェントシミュレーション

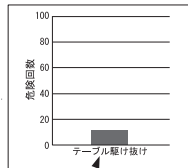


データを分類して
モデル作成

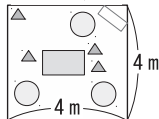


エージェント移動ルール

リスク評価



レイアウト



評価したいレイアウトを
デザイン

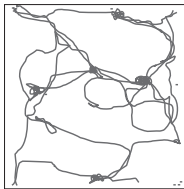


リスク状況

多解像度クラスタ分析法

- ① 移動データを複数の解像度(区画サイズ)で分割

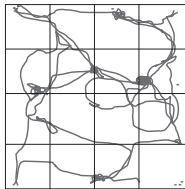
解像度0



解像度1



解像度2

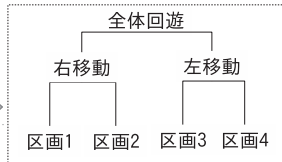


- ② 各区画で特徴量を計算 {速度ベクトルの向き, 平均速度, データ数}

- ③ 全区画の特徴量を用いて階層クラスタ分析



- ④ 解像度の違いで分割情報利得の増減を見る。

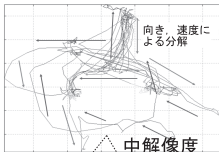


ケース1

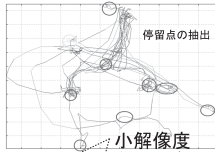
① 分析 解像度 粗い 1.5m x 1.5m



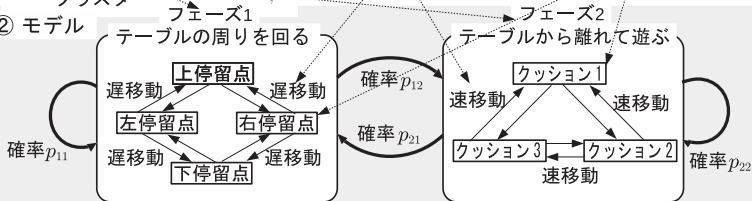
解像度 0.5m x 0.5m



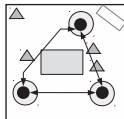
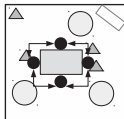
解像度 0.3m x 0.3m



② モデル

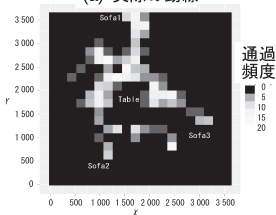


③ シミュレーション

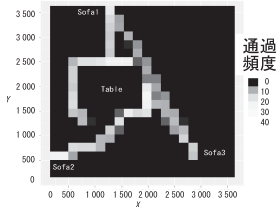


ケース1のシミュレーション結果

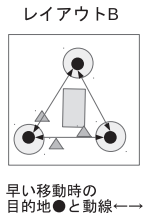
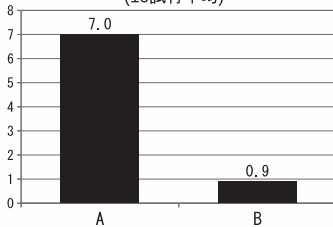
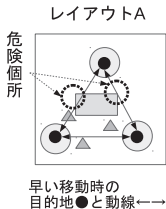
(a) 実際の動線



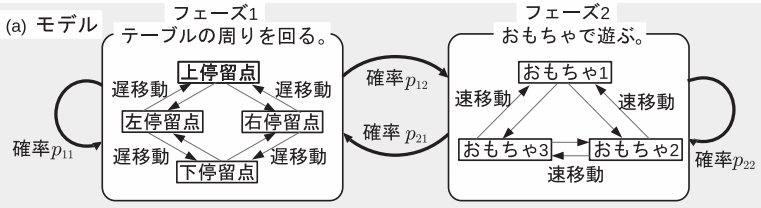
(b) シミュレーション結果



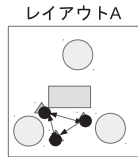
(c) 各レイアウトでの危険場面の回数
(10試行平均)



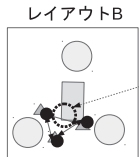
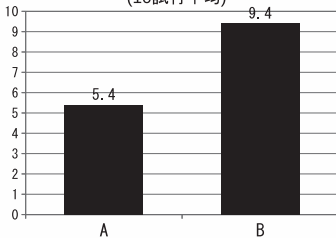
ケース2の移動モデルとシミュレーション結果



(b) 各レイアウトでの危険場面の回数
(10試行平均)



早い移動時の
目的地●と動線←→

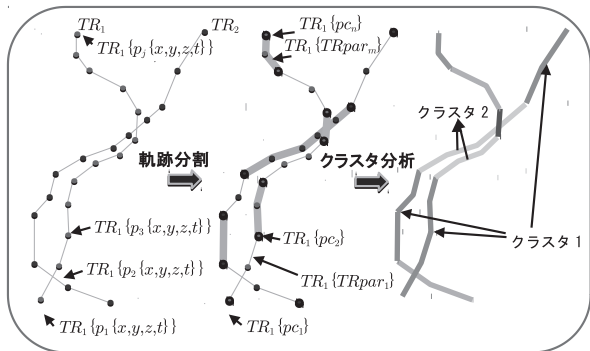


危険箇所

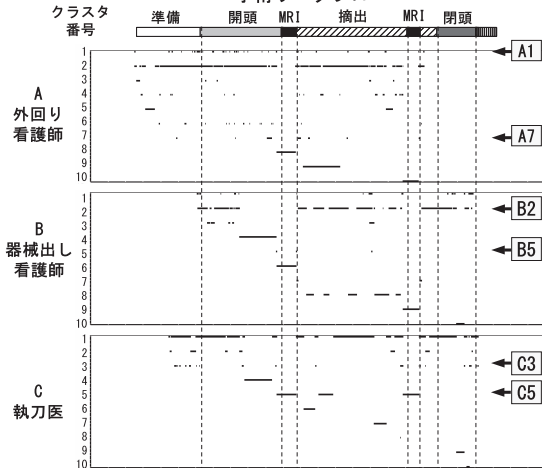
早い移動時の
目的地●と動線←→

軌跡クラスタ分析

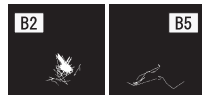
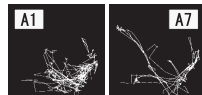
1. 軌跡分割法による軌跡のセグメンテーション
2. 分割された軌跡の特徴量（移動距離，平均速度など）をもとに K-means法（非階層クラスタ分析）による分類



手術ワークフロー



各クラスターの
部屋の上から
見た動きの例



各クラスターが発生した時刻

Speed Setting

Speed [100]

1 100

Repeat Setting

1

Random Seed Setting

 AUTO

1303118796894

Event Tag Setting

 Person x Person Person x Equipment Equipment x Equipm..

Footmark Display Step

100

Module Display Setting

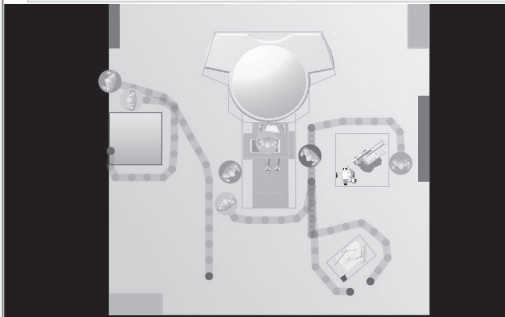
 Agent Layout HeatMap

Image Display Setting

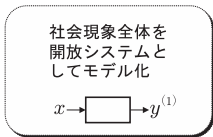
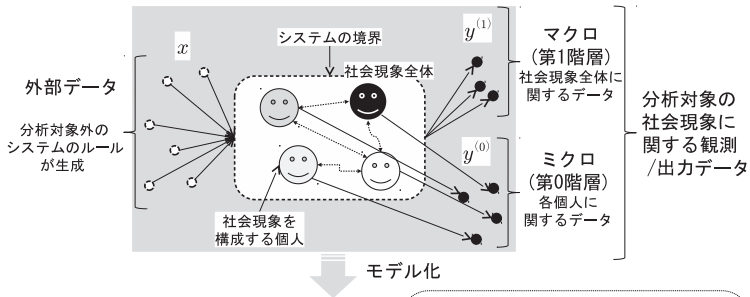
 Agent Equipment Area

Room

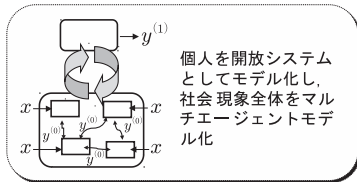
TIME 00:56:30 Repeat: 1 Finished



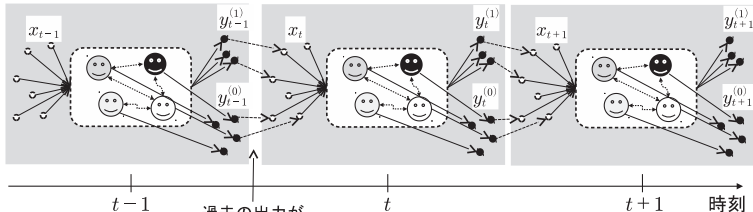
関連する外部データが存在するシステムのモデル化



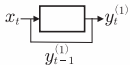
または



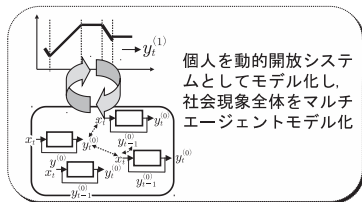
関連する外部データが 時間順序を持つシステムのモデル化



社会現象全体を動的開放システムとしてモデル化



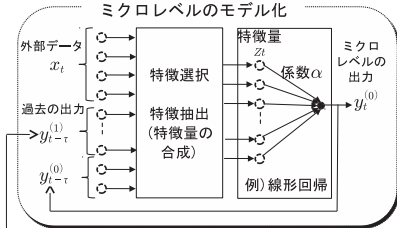
または



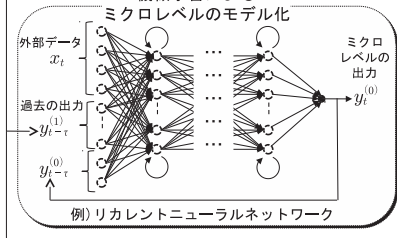
個人を動的開放システムとしてモデル化し、社会現象全体をマルチエージェントモデル化

マイクロレベルとマクロレベルのモデル化

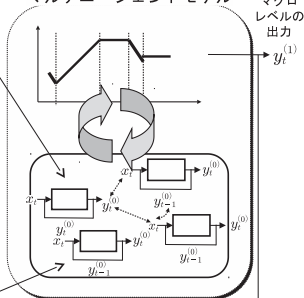
特徴選択と多変量解析による
マイクロレベルのモデル化



機械学習による
マイクロレベルのモデル化

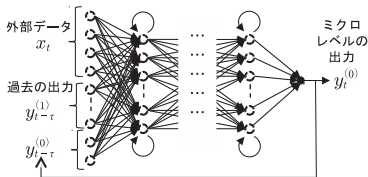
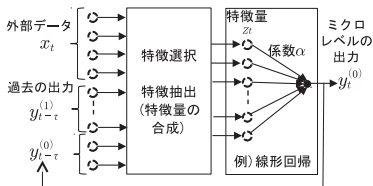


マイクロとマクロの関係性の
マルチエージェントモデル

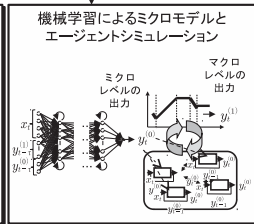
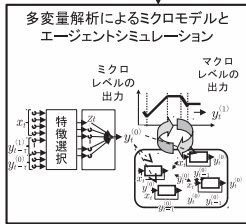
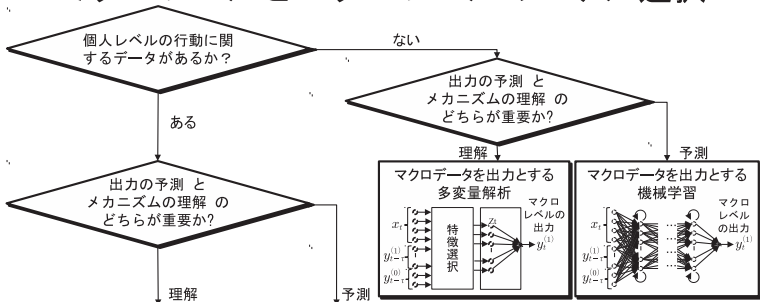


多変量解析と機械学習の違い

	多変量解析(統計学的手法)	機械学習
おもな目的	<ul style="list-style-type: none"> 入力データと出力データの関係性を説明すること データ全体の傾向を調べる 	<ul style="list-style-type: none"> 入力データから出力データを予測すること (部分的でも)データの特徴を抽出すること
特徴選択	<ul style="list-style-type: none"> 適切な変数を事前に選択する必要がある。 	<ul style="list-style-type: none"> 変数を事前に選択する必要がない。
モデル選択	<ul style="list-style-type: none"> 統計指標によるモデル選択の指針があることが多い。 	<ul style="list-style-type: none"> 適切なモデルを事前に選択する必要がある。
困難点	<ul style="list-style-type: none"> 変数選択が難しい。 	<ul style="list-style-type: none"> モデル選択が難しい。

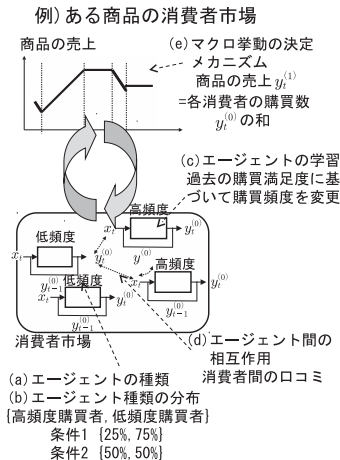


マイクロレベルとマクロレベルのモデル選択



エージェントモデルの要素

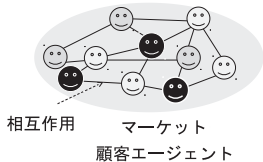
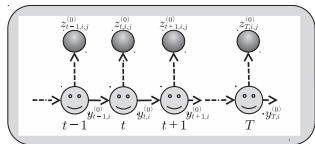
- (a) エージェントの種類
- ・ ミクロデータの分析結果に基づき、入出力変数や入出力関係を選択して、エージェントタイプを構築する。
- (b) エージェント種類の分布
- ・ 社会現象中のエージェントタイプの割合
 - ・ 決め方：固定，実データ，エージェントの学習
- (c) エージェントの学習
- ・ どんな基準で自分の行動ルールを変更するか。
- (d) エージェント間の相互作用
- ・ 相互作用ネットワークと相互作用ルール
- (e) マクロ挙動の決定メカニズム
- ・ 個人の行動や相互作用から社会現象全体の挙動が決まる構造



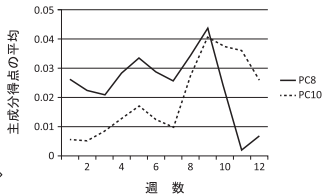
マーケティングシミュレーション

• エージェントシミュレーション

購買行動モデル



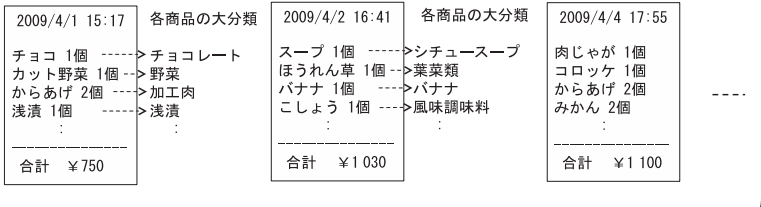
特定商品の
各ブランドのシェア



ID-POSデータの解析

「いつ、どこで、だれが、なにを、いくつ」購買したのかの履歴が蓄積されたデータ

顧客ID n のレシートデータ



一定期間での集約

顧客ID n の一定期間での大分類ごとの購買回数

チョコレート	3回
野菜	3回
加工肉	1回
浅漬	1回
:	

ID-POSデータ解析

データ期間7か月で購買回数が50回以上の購買客のレシートデータを分析

	チョコレート	野菜	加工肉	...	風味調味料
期間1, 顧客1	3	3	2	...	2
期間1, 顧客2	1	3	3	...	0
:	:	:	:	...	:
期間1, 顧客N	0	7	6	...	0
:	:	:	:	...	:
期間t, 顧客i	0	5	5	...	1
:	:	:	:	...	:
期間T, 顧客N	1	5	0	...	2

60種類の商品カテゴリ



特徴選択

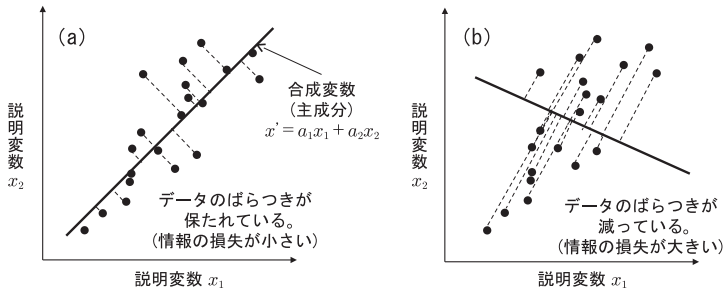
共購買(併買)パターンによる購買特徴

例) 野菜類と肉類, 酒類と菓子, 野菜と調味料などの共購買

主成分分析

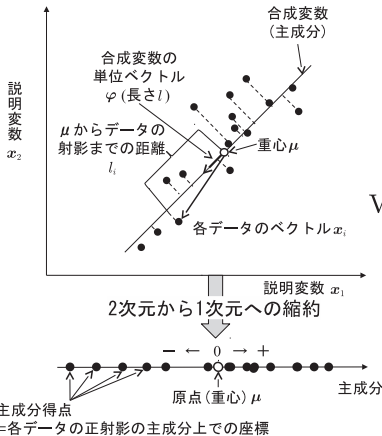
- 多種類の説明変数を、できるだけ情報を損わずに少ない種類の合成変数(主成分)に縮約する。
 - 情報とはデータのばらつき(分散)
 - 合成変数は元の変数に重みをつけて統合したもの
 - 合成変数間には相関がない(直交している)。

例) 2次元のデータを1次元の合成変数に縮約する場合



主成分分析

- 全データの合成変数への射影のばらつきを最大化する。



各データ x_i の射影までの距離

$$l_i = (x_i - \mu) \cdot \varphi$$

距離 l_i の分散 Var を最大化する。
合成変数の単位ベクトル φ を求める。

$$\begin{aligned} \text{Var}(l_i) &= \frac{1}{n} \sum_{i=1}^n l_i^2 = \frac{1}{n} \sum_{i=1}^n [(x_i - \mu) \cdot \varphi]^2 \\ &= \varphi^T \underbrace{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T}_{\text{分散共分散行列}} \varphi \end{aligned}$$

Var を最大化する φ は
分散共分散行列の固有ベクトル

主成分分析による購買行動の特徴選択

元の説明変数
60種類の商品カテゴリ

7か月間1.2万人分の購買データを主成分分析
15個の主成分(固有値が1.0以上)に縮約

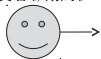
チョコレート
野菜
加工肉
浅漬
シチュー・スープ
葉菜類
バナナ
玉子
果菜類
おかず
キムチ
:
風味調味料



	ラベル	関連する商品カテゴリ
1	昼食夕食	(+) おにぎり,和洋菓子,サラダ,菓子パン,総菜,寿司
2	女性: 酒-ソフトドリンク	(+) チューハイ,ビール,アイス,ポテトチップス,女性雑誌 (-) コーヒー,紅茶,お茶,サンドイッチ
3	デザート-パスタ	(+) クッキー,ビスケット,アイス (-) 焼そば,冷し麺,調理麺パスタ
4	男性: 酒-菓子	(+) チューハイ,ビール,たばこ,炭酸水,ポテトチップス,カップ麺 (-) ソフトドリンク,チョコレート,キャンディ,パスタ
:	:	:
15	ペットボトル-チルド飲料	(+) ソフトドリンク(900ml,1.5L),総菜パン,食卓パン,女性雑誌 (-) チルド飲料(牛乳,紅茶,野菜),焼そば,寿司,新聞

各購買者の各期間の購買行動は15次元の主成分得点に縮約される

購買者 i , 期間 t



購買行動

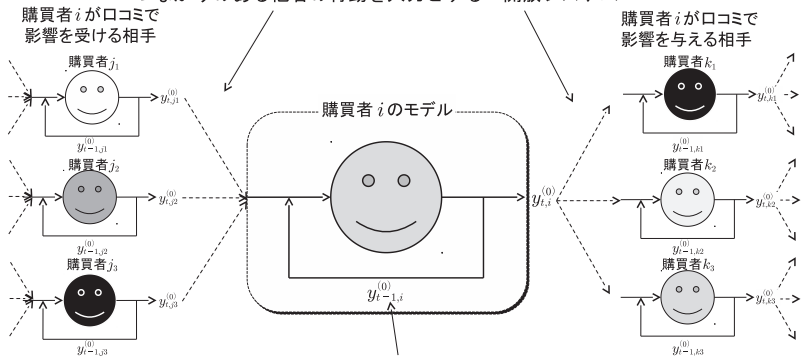
チョコレート	3回
野菜	3回
加工肉	1回
浅漬	1回
:	:

$$\rightarrow y_{t,i}^{(0)} = \begin{pmatrix} y_{1,t} = 0.026 \\ y_{2,t} = 0.601 \\ \vdots \\ y_{15,t} = 0.218 \end{pmatrix}$$

購買行動のマイクロモデル(動的開放システム)

- 購買行動 $y_{t,i}^{(0)}$ = 購買者 i の期間 t に購買した品目と数量
- 購買行動決定の際に、消費者のロコネットワークの影響が大きい。

つながりのある他者の行動を入力とする→開放システム



- 過去の自分の購買行動に執着する場合がある(粘着的な購買)
 - 過去の自分の購買行動 $y_{t-1,i}^{(0)}$ を入力とする→動的システム

態度変容モデル(全体)

エージェント*i*の*t*週目における状態変数の第*m*主成分の値

$$y_{m,t,i}^{(0)} = a_{m0} + \sum_{l=1}^{15} a_{ml} y_{l,t-1,i}^{(0)} + \epsilon_{m,t,i}$$

- (a) 過去の自分からの影響・・・1週前からのみ影響（自己回帰モデル）
係数 a を実データからの分析により決定（係数 a は全消費者で同一とした）

$$+ \beta \sum_{j=1}^3 (y_{m,t-1,\text{neighbor}(i,j)}^{(0)} - y_{m,t-1,i}^{(0)})$$

- (b) 他人からの影響

$\text{neighbor}(i,j)$ = 消費者 i とつながっている j 人目の消費者
スモールワールドネットワークを仮定

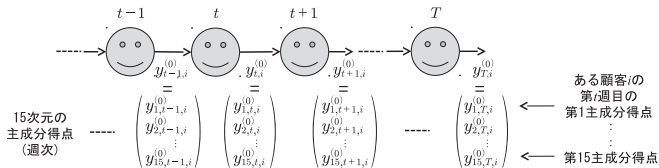
態度変容モデル(粘着性・習慣性)

- 過去の自分の購買行動に執着する場合がある(粘着的な購買)
 - 過去の自分の購買行動 $y_{l,t-1,i}^{(0)}$ を入力とする→動的システム

$$a_{m0} + \sum_{l=1}^{15} a_{ml} y_{l,t-1,i}^{(0)} + \epsilon_{m,t,i}$$

(a) 過去の自分からの影響・・・1週前からのみ影響(自己回帰モデル)
 係数 a を実データからの分析により決定(係数 a は全消費者で同一とした)

※すべての週で1回以上購買した顧客5937人のPOSデータの分析



線形回帰式に対する最小二乗法

各データ点での誤差の二乗和 S を最小にする係数 a を求める。

線形回帰式 ($i=1\dots n$ のデータ点)

$$y_i = a_0 + \sum_{l=1}^p a_l x_{li} + \epsilon_i = \tilde{y}_i + \epsilon_i$$

すべてのデータ点についてまとめる。

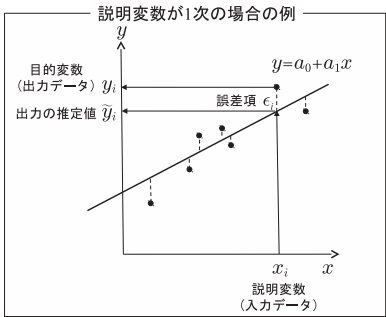
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

\downarrow \quad \downarrow \quad \downarrow \quad \downarrow

$$y = X a + \epsilon$$

誤差の二乗和

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 = (\mathbf{y} - \mathbf{X}\mathbf{a})^T(\mathbf{y} - \mathbf{X}\mathbf{a})$$



線形回帰式に対する最小二乗法

誤差の二乗和 S を係数ベクトル a で偏微分する。

$$\begin{aligned}\frac{\partial S}{\partial a} &= \frac{\partial}{\partial a} (\mathbf{y} - \mathbf{X}\mathbf{a})^T (\mathbf{y} - \mathbf{X}\mathbf{a}) \\ &= -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{a} = 0\end{aligned}$$

係数 a は次式の正規方程式の解となる。

$$\mathbf{X}^T \mathbf{X} \mathbf{a} = \mathbf{X}^T \mathbf{y}$$

$$\begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \cdots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \cdots & \sum_{i=1}^n x_{i1}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{i1}x_{ip} & \cdots & \sum_{i=1}^n x_{ip}^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ip}y_i \end{pmatrix}$$

態度変容モデル(他者からの影響)

- 購買行動決定の際に、消費者のロコミネットワークの影響が大きい。
 - つながりのある他者の行動を入力とする→開放システム

• スモールワールドネットワーク

- 全エージェントが自分の隣の k 個のエージェントとリンクがある状態から、割合 p でリンクを選択してランダムに接続先を変更する。
- 現実の人間関係の特徴とよく似たネットワーク構造になる。

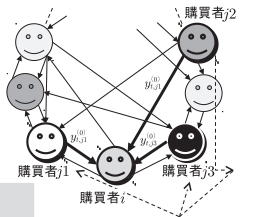
相手 $y_{m,t-1,neighbor(i,j)}^{(0)}$ と自分 $y_{m,t-1,i}^{(0)}$ の購買行動の差を見て

- 差を縮めるように購買行動 $y_{m,t,i}^{(0)}$ を変化 ($\beta > 0$)
- 差を拡大するように購買行動 $y_{m,t,i}^{(0)}$ を変化 ($\beta < 0$)

$$+ \beta \sum_{j=1}^3 (y_{m,t-1,neighbor(i,j)}^{(0)} - y_{m,t-1,i}^{(0)})$$

(b) 他人からの影響

$neighbor(i,j)$ = 消費者 i とつながっている j 人目の消費者
スモールワールドネットワークを仮定



購買者 i がロコミで影響を受ける相手

$neighbor(i,1) = j1$

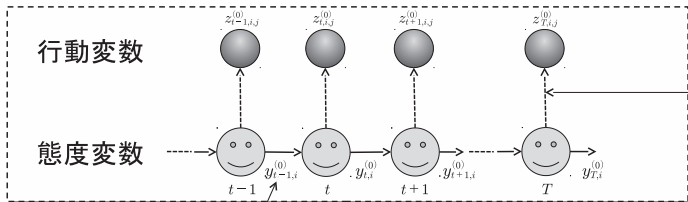
$neighbor(i,2) = j2$

$neighbor(i,3) = j3$

エージェントの購買行動モデル

購買行動モデル(ポワソン過程+k近傍法)

$z_{t,i,j}^{(0)}$ エージェント*i*の*t*週目におけるブランド*j*の購買確率



$y_{m,t,i}^{(0)}$ エージェント*i*の*t*週目における商品カテゴリ間の共購買パターン

態度変容モデル(自己回帰+スモールワールド)

$$y_{m,t,i}^{(0)} = a_{m0} + \sum_{l=1}^{15} a_{ml} y_{l,t-1,i}^{(0)} + \epsilon_{m,t,i} + \beta \sum_{j=1}^3 (y_{m,t-1,\text{neighbor}(i,j)}^{(0)} - y_{m,t-1,i}^{(0)})$$

自己回帰 スモールワールド

購買行動モデル

1. 各エージェント*i*が*t*期に特定商品群を購入する総数 $\omega_{t,i}^{(0)}$ を決定

- 総数 $\omega_{t,i}^{(0)}$ が *n* 個である確率は期待値 $\lambda_{t,i}^{(0)}$ のポアソン分布に従う。

$$\text{Prob}(\omega_{t,i}^{(0)} = n) = \frac{[\lambda_{t,i}^{(0)}]^n e^{-\lambda_{t,i}^{(0)}}}{n!}$$

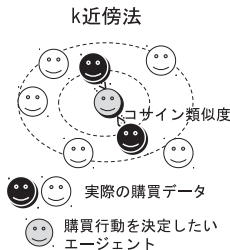
- エージェント*i*の*t*期の総購買数の期待値 $\lambda_{t,i}^{(0)}$ は、態度変数 $y_{m,t,i}^{(0)}$ の線形回帰で推定

$$\lambda_{t,i}^{(0)} = \alpha_0 + \sum_{m=1}^{15} \alpha_m y_{m,t,i}^{(0)}$$

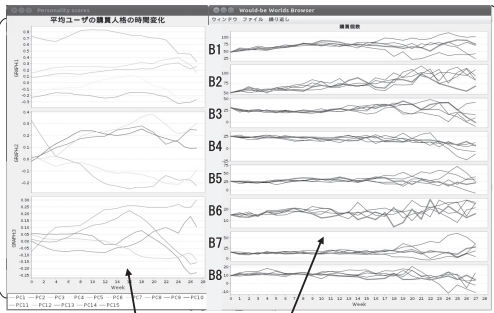
2. 各ブランドの購入比率を購買態度から推定

k近傍法により決定：実際の購買データから態度変数が類似している*k*人の買い方を参照する。

係数 α_m やパラメータ *k* の値は特定消費群を購入した
29 780人の購買データから推定



出力画面



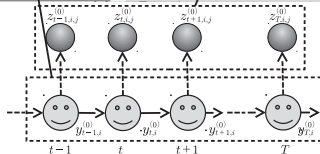
5 937個の
顧客エージェントの
15次元の態度変数の
平均推移

※右の線グラフの
購買総数が結果
となるとき
の態度変数の推移
を表す

5 937個の
顧客エージェントの
各ブランド
購入総数

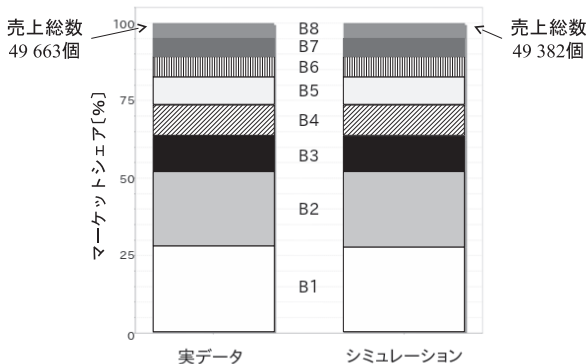
行動変数

態度変数



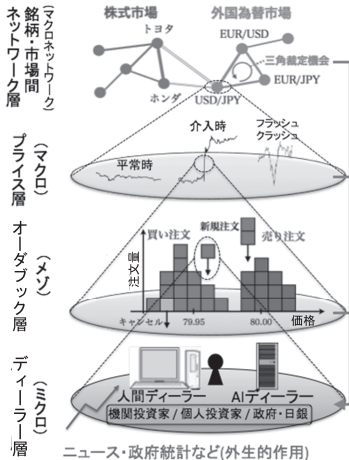
ブランドごとのシェア比較

7か月間全体での各ブランドの売上合計の比率を、実データとシミュレーションで比較した。



金融市場のミクロ・マクロ構造の俯瞰

金融市場の階層構造



解析データ

各銘柄
価格データ
(外為, 日経)

価格データ
(外為, 日経)

オーダーブック
データ

各ディーラーの
売買行動履歴

モデル

動的
ネットワーク
モデル

時系列モデル

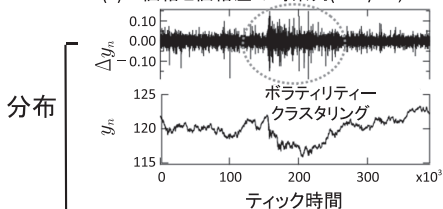
オーダーブック
モデル

エージェント
ベース
モデル

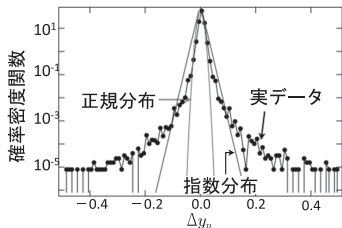
マクロ・ミクロ間の横断的理解

金融市場の代表的経験則(価格変動の分布)

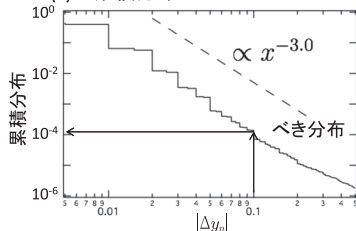
(a) 価格と価格差の時系列(USD/JPY, 2000~2005)



(b) 確率密度関数

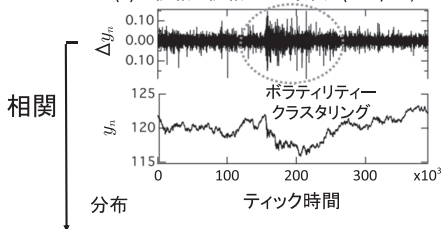


(c) 累積分布

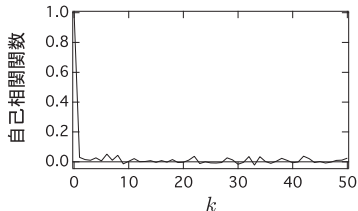


金融市場の代表的経験則(価格変動の相関)

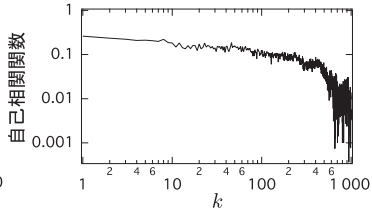
(a) 価格と価格差の時系列(USD/JPY, 2000~2005)



(b) 自己相関関数(価格差)



(c) 自己相関関数(ボラティリティ)



問: 下の数列はある規則に従って並んでいる
□に入る数字を予測せよ。

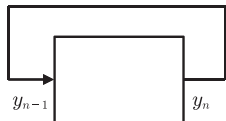
(1) 1, 3, 5, 7, 9, □

(2) 1, 2, 4, 8, 16, □

(3) 1, 2, 5, 10, 17, □

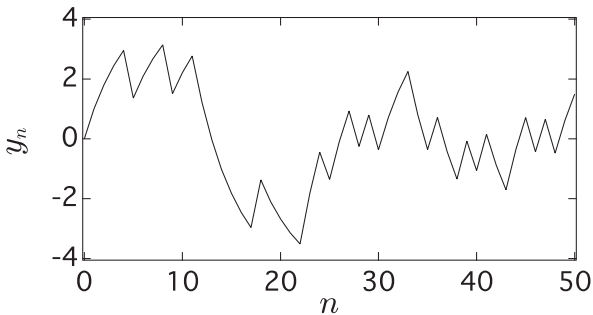
(4) 1, 0, -1, 0, -1, -2, -1, □

(5) 1, 1.1, -0.89, -1.089, 0.0089, □



この問題は左図に示すフィードバックシステムの時間発展を予想していることに対応する。

自己回歸 (AR) 過程 (確率論的)



$$y_n = \underbrace{0.9y_{n-1}}_{\text{決定論的}} + \underbrace{f_n}_{\text{確率論的}}$$

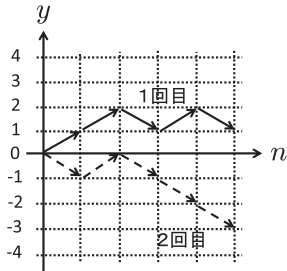
ランダムウォーク(酔歩)

偏りのないコインを投げて表が出たら+1(確率1/2), 裏が出たら-1(確率1/2)進む確率過程を考える。

例:

n	1	2	3	4	5
1回目 (実線)	表	表	裏	表	裏
2回目 (点線)	裏	表	裏	裏	裏

図示



この確率過程は、ランダムウォークと呼ばれ次式の漸化式に従う。

時間発展方程式

$$y_n = y_{n-1} + f_n$$

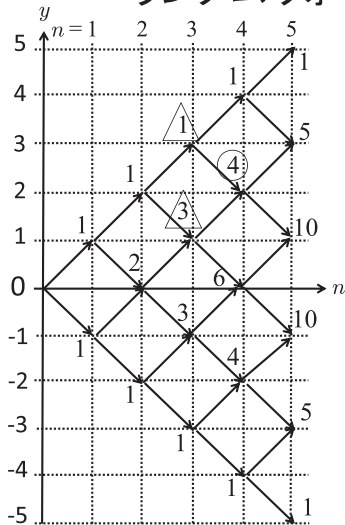
確率項

$$f_n = \begin{cases} +1 & (\text{確率 } 1/2) \\ -1 & (\text{確率 } 1/2) \end{cases}$$

初期条件

$$y_0 = 0$$

ランダムウォークの経路数



各地点の数字はランダムウォークの経路数を表す。

例えば $n=4$ で $y=2$ に着く経路は $\{+++-, ++-+, +-++, -++++\}$ の4通り

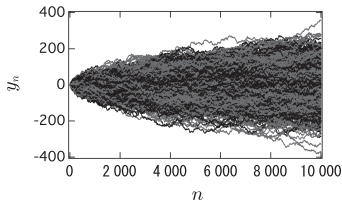
ランダムウォークの経路数はパスカルの三角形に従う。

$n=4$ で $y=2$ に着く確率は

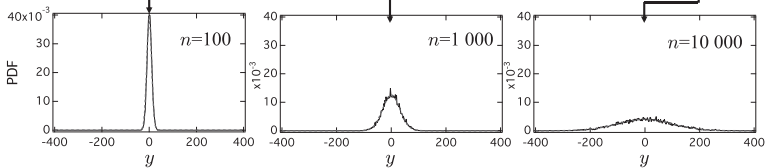
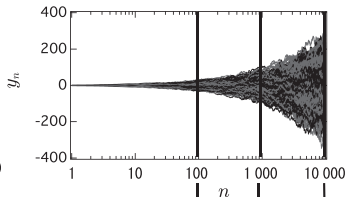
$$\frac{4}{1 + 4 + 6 + 4 + 1} = \frac{1}{4}$$

ランダムウォークの分布

ランダムウォーク 1 000 試行



ランダムウォーク 1 000 試行(横軸対数)



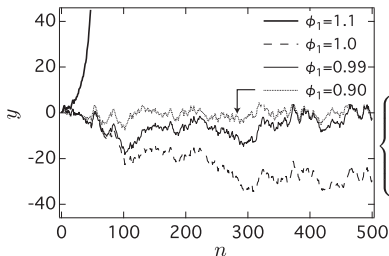
破線は $g(y_n, n) = \frac{1}{\sqrt{2\pi n}} e^{-\frac{y_n^2}{2n}}$ (平均 0, 分散 n) の正規分布

自己回帰(AR)モデル

$$y_n = \phi_0 + \sum_{i=1}^k \phi_i y_{n-i} + f_n$$

f_n は正規分布に従う
ホワイトノイズ

$k=1$ (AR(1))において ϕ_1 を変化させたときのシミュレーション結果



$$(\phi_0 = 0, \langle f_n \rangle = 0, \langle f_n^2 \rangle = 1)$$

$\phi_1 = 1$ を境にダイナミクスがまったく異なる。

- $1 < \phi_1$ $\dots +\infty$ または $-\infty$ へ発散
- $\phi_1 = 1$ \dots ランダムウォーク
- $-1 < \phi_1 < 1$ \dots ガウス過程 (定常分布を持つ)
- $\phi_1 < -1$ \dots 振動発散

$|\phi_1| < 1$ のとき AR(1) の平均値と分散は以下ようになる。

平均値: $\frac{c}{1-\phi}$, 分散: $\frac{\sigma_f^2}{1-\phi^2}$

分散は時間 n に依存しない!
(ランダムウォークのときは n に比例)

Yule-Walker方程式

ラグ k までの自己共分散 C と重み ϕ , ノイズの分散 σ_f^2 の関係式

$$\begin{pmatrix} C_0 & C_1 & C_2 & \cdots & C_k \\ C_1 & C_0 & C_1 & \cdots & C_{k-1} \\ C_2 & C_1 & C_0 & \cdots & C_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ C_k & C_{k-1} & C_{k-2} & \cdots & C_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\phi_1 \\ -\phi_2 \\ \cdots \\ -\phi_k \end{pmatrix} = \begin{pmatrix} \sigma_f^2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Yule-Walker方程式を用いた時系列データからの
AR(k)過程の推定手順

0. 時系列が定常であることを確認する(例えば単位根検定)。
1. ラグ k までの自己共分散を計算する。
2. 1の自己共分散をYule-Walker方程式へ代入し連立1次方程式を解く。
3. $\phi_1, \phi_2 \cdots \phi_k$ と σ_f^2 が得られAR(k)過程が決まる。
4. AICなどを用いて最適な次数 k を決定する。

Levinson-Durbinの漸化式

時系列データからAR過程を推定する際に次数 k が高い場合やさまざまな次数 k を計算し比較したい場合に有用

k 次の AR モデルの j 番目の AR 係数を $\phi_{j,k}$, 分散を σ_k^2 と表す。

$$\left[\begin{array}{l}
 1. \tilde{C}_0, \dots, \tilde{C}_M \text{ を時系列データから計算する。} \\
 2. \tilde{\sigma}_0^2 = \tilde{C}_0 \\
 3. \tilde{\phi}_{k:k} = \left(\tilde{C}_k - \sum_{j=1}^{k-1} \tilde{\phi}_{j:k-1} \cdot \tilde{C}_{k-j} \right) (\tilde{\sigma}_{k-1}^2)^{-1} \quad \leftarrow k \leftarrow k+1 \\
 4. \tilde{\phi}_{i:k} = \tilde{\phi}_{i:k-1} - \tilde{\phi}_{k:k} \tilde{\phi}_{k-i:k-1} \quad (i < k) \\
 5. \tilde{\sigma}_k^2 = \tilde{\sigma}_{k-1}^2 (1 - \tilde{\phi}_{k:k}^2) \quad \text{-----}
 \end{array} \right.$$

パラメータ推定の順序：上の手順を繰り返すと○の中の数字の順番でパラメータが推定される。

	分散	1番目の係数	2番目の係数	3番目の係数	...
1次のARモデル	② σ_1^2	① $\tilde{\phi}_{1:1}$			
2次のARモデル	⑤ σ_2^2	④ $\tilde{\phi}_{1:2}$	③ $\tilde{\phi}_{2:2}$		
3次のARモデル	⑨ σ_3^2	⑧ $\tilde{\phi}_{1:3}$	⑦ $\tilde{\phi}_{2:3}$	⑥ $\tilde{\phi}_{3:3}$	

⋮

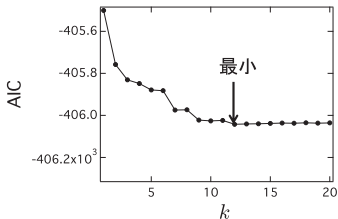
ARモデルのパラメータ推定結果

- (a) 価格差(ΔP)に対してARモデルのパラメータを Levinson-Durbin のアルゴリズムを用いて推定を行った結果

	$\tilde{\sigma}_f$	$\tilde{\phi}_1$	$\tilde{\phi}_2$	$\tilde{\phi}_3$	$\tilde{\phi}_4$...
1	1.47×10^{-4}	0.999 96				
2	1.47×10^{-4}	1.030 21	-0.030 25			
3	1.47×10^{-4}	1.029 72	-0.013 64	-0.016 11		
4	1.47×10^{-4}	1.029 59	-0.013 75	-0.007 78	-0.008 09	

⋮

- (b) 次数(k)を変化させたときのAIC



- (c) パラメータの推定値

$\tilde{\sigma}_f$	0.000 147		
$\tilde{\phi}_1$	1.029	$\tilde{\phi}_7$	-0.020
$\tilde{\phi}_2$	-0.014	$\tilde{\phi}_8$	0.012
$\tilde{\phi}_3$	-0.007	$\tilde{\phi}_9$	-0.018
$\tilde{\phi}_4$	0.003	$\tilde{\phi}_{10}$	0.005
$\tilde{\phi}_5$	-0.004	$\tilde{\phi}_{11}$	0.008
$\tilde{\phi}_6$	0.015	$\tilde{\phi}_{12}$	-0.009

ARCHモデル

ARCH(q)モデル

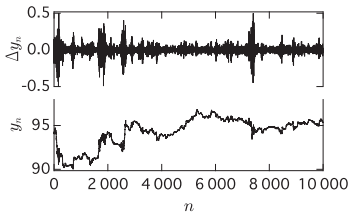
$$y_{n+1} = y_n + \sigma_n f_n$$

$$\sigma_n^2 = \sigma_0^2 + \sum_{m=1}^q \alpha_m (\sigma_{n-m} f_{n-m})^2$$

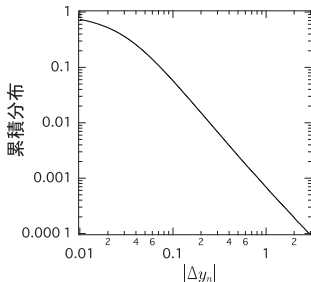
ARCHモデルのシミュレーション結果

$$\sigma_0 = 0.01, \alpha_1 = \dots = \alpha_{10} = 0.1$$

(a) 価格変動(下)と価格差(上)



(b) ボラティリティーの累積分布



感染症対策とシミュレーション

- 新型インフルエンザウイルスによるパンデミックの危険性
 - 突然変異・遺伝子再集合による新種の出現。免疫を持たないために、容易に世界的感染拡大(パンデミック)が起こりうる。
 - A (H1N1): スペイン風邪, 季節性, 2009年新型ブタ
 - A (H3N2): 香港かぜ, 季節性
 - A (H5N1): おもに野鳥間で伝染。90年代後期に、ヒトへの感染力を獲得した高病原株が登場
 - A (H7N9): 2013年に中国・香港で局地的流行が発生。鳥インフル
 - シミュレーションの感染症対策への貢献
 - 介入政策目標の設定 (e.g. 特定の学校群での学級閉鎖)
 - シナリオ分析
 - 流行予測
- 現時点での
パンデミック
の潜在的脅威

SIRモデル

$$\frac{dS}{dt} = -\lambda IS$$

$$\frac{dI}{dt} = \lambda IS - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

+非一様性



$$\frac{dS_i}{dt} = -\lambda_i (I_i + \sum_{j \in A(i)} \epsilon_{ij} I_j) S_i$$

$$\frac{dI_i}{dt} = \lambda_i (I_i + \sum_{j \in A(i)} \epsilon_{ij} I_j) S_i - \gamma I_i$$

$$\frac{dR_i}{dt} = \gamma I_i$$

+確率的変動



$$S_n = S_{n-1} - \#[S \rightarrow I]_n$$

$$I_n = I_{n-1} + \#[S \rightarrow I]_n - \#[I \rightarrow R]_n$$

$$R_n = R_{n-1} + \#[I \rightarrow R]_n$$

$$\#[S \rightarrow I]_n \sim \text{Poisson}[\text{平均} = \lambda I_{n-1} S_{n-1} \Delta t]$$

$$\#[I \rightarrow R]_n \sim \text{Poisson}[\text{平均} = \gamma I_{n-1} \Delta t]$$

+確率的変動



$$S_{i,n} = S_{i,n-1} - \#[S \rightarrow I]_{i,n}$$

$$I_{i,n} = I_{i,n-1} + \#[S \rightarrow I]_{i,n} - \#[I \rightarrow R]_{i,n}$$

$$R_{i,n} = R_{i,n-1} + \#[I \rightarrow R]_{i,n}$$

$$\#[S \rightarrow I]_{i,n} \sim \text{Poisson}[\text{平均} = \lambda_i (I_i + \sum_{j \in A(i)} \epsilon_{ij} I_j) S_i \Delta t]$$

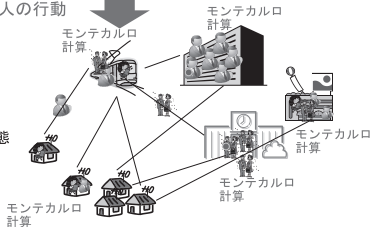
$$\#[I \rightarrow R]_{i,n} \sim \text{Poisson}[\text{平均} = \gamma I_{i,n-1} \Delta t]$$

+個人の行動



個人ベースシミュレーション

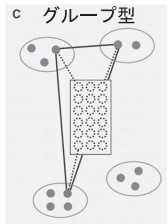
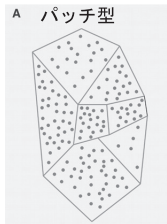
- すべての感受性者に対してコイントスでつぎのステップでの感染/非感染、状態遷移を決める。



個人ベースシミュレーションの類型

Riley (2007), Science 316, 1298

- パッチ型：パッチ内のメンバは等しい感染力を受ける。
- 距離型：感染者からの距離に反比例する感染力を受ける。
- グループ型：グループ内での感染伝播 + なんらかのグループ間作用
- ネットワーク型：ネットワークにより感染路を直接表現する。



東南アジアにおけるH5N1封込め可能性(1)

Ferguson+(2005), Science 437, 209-214

• 背景

- 鳥インフルエンザA(H5N1)に、強毒性の変種が東南アジアで見ついている。現時点では、安定的なヒト-ヒト間の感染伝達は確認されていないが、獲得時の被害・可能な対応策を事前に試算しておくことは有益である。

• 課題

- 感染者確認の早期に効率的よく抗ウイルス剤を配布することで、国全体(引いては地球全体)への感染拡大を防ぐことができるかを、シミュレーションによって調べる。

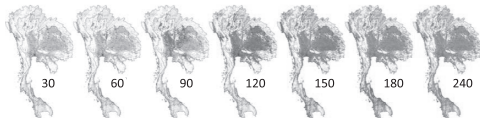
• 結論

- ウイルスの感染力が $R_0 \leq 1.9$ であれば、90%の確率で封込めに成功する。

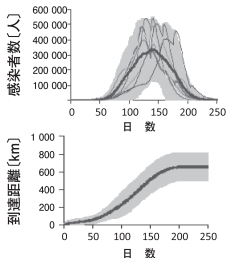
東南アジアにおけるH5N1封込め可能性(2)

- シミュレーション設定：非介入時の動向
 - シミュレーション対象：タイを中心とする東南アジア一帯
 - 人口(エージェント)：85万 (タイの人口に匹敵)
 - ウィルスの感染力：基礎再生産定数 R_0 を単位として、 $R_0 = 1.1 \sim 1.9$
 - 最初の30日は、初期感染者を導入した地区に留まるが、60~90日目に国全体に拡大する。

■ 感染者 ■ 回復者



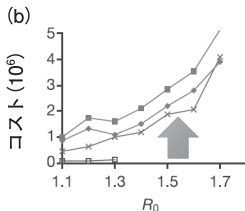
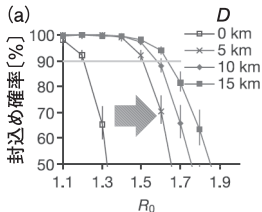
$R_0 = 1.5$ の場合の感染拡大の様子



東南アジアにおけるH5N1封込め可能性(3)

抗ウイルス剤の予防的投与対象者の 選び方と封込め可能性の条件

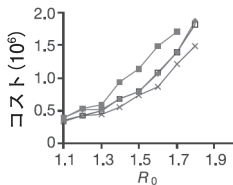
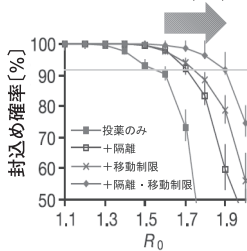
- 包括的投与
 - $R_0 < 3.6$ であれば封込め可能
(抗ウイルス剤の理論性能の表示)
- 社会的選定：患者が属する家族，学校，会社
 - $R_0 \leq 1.25$ が封込め確率90%達成に必要
 - (右図で $D = 0\text{km}$ に対応)
- 地理的選定：患者の半径 D 圏内
 - 封込め確率90%達成条件を緩和できるが，
必要な接種コストは増大
 - $R_0 \leq 1.5$, $D = 5\text{km}$ で，平均 2×10^6 コースが必要
最悪，平均の10倍が必要になる。



抗ウイルス剤の節約を検討する必要がある。

東南アジアにおけるH5N1封込め可能性(4)

- 複合戦略と封込め可能性
 - 節約：地理的選定 + 患者に最近接の m [人] に制限
 - 消費量が半分程度に抑えられる。
 - 隔離：節約($m=50\ 000$ 人)しつつ患者が発見された5km圏内で21日間、5割の会社、9割の学校を閉鎖
 - $R_0 \leq 1.7$ で封込め確率90%が実現
 - 移動制限：節約($m=50\ 000$ 人)しつつ患者が発見された5km圏内で21日間、流入出量を80%に制限
 - $R_0 \leq 1.7$ で封込め確率90%が実現
 - 隔離 + 検疫
 - $R_0 \leq 1.9$ で封込め確率90%が実現



モデリング/シミュレーションの構成(1)

シミュレーションの流れ

- 各個人(エージェント) $i = 1, \dots, N$ について:
 - 感受性 (S) のとき: 確率 $P_{S \rightarrow E, i} = 1 - \exp(-\lambda_i \Delta t)$ で曝露状態 (E) へ遷移
 - λ_i を感染力 (force of Infection) という。
 - 曝露状態 (E) のとき:
 - 最初にこの状態に遷移したとき: ワイブル分布 $h(\cdot | t_0, \gamma, \eta)$ から滞在時間 T_E をサンプリングする (詳細は後述)。
 - この状態への滞在時間が T_E 経過してから有感染状態 (I) へ遷移
 - 有感染力状態 (I) のとき:
 - この状態に遷移してからの経過時間を t とすると, 他者に及ぼす感染力は $\kappa(t)$ に比例して減少する。 $\kappa(t)$ は10日で打ち切られた対数正規関数(後述)。
 - 10日後に回復状態 (R) へ遷移
- $t \leftarrow t + \Delta t$ として最初に戻る (この研究では $\Delta t = 0.25$ 日)

モデリング/シミュレーションの構成(2)

- 感染伝播のモデリングに必要な要素
 - 潜伏期間 (時間を確率変数とする分布) $h(t)$
 - 発症期間 (分布/時間) $\kappa(t)$
 - 感染力 (上の二つから合成) λ



- 未感染者が時間 τ だけ経過後に感染者になる確率

$$P_{S \rightarrow I} = 1 - \exp(-\lambda_i \tau) \doteq \lambda_i \tau$$

「 $n \rightarrow \infty$ 回の被感染イベントをすべて回避した」の余事象の確率

$$1 - P_{S \rightarrow I} = \left(1 - \lambda \frac{\tau}{n}\right) \times \left(1 - \lambda \frac{\tau}{n}\right) \times \cdots \times \left(1 - \lambda \frac{\tau}{n}\right) \rightarrow \exp(-\lambda \tau)$$

※ シミュレーションは $\tau = 0.25$ 日をタイムステップとして実行

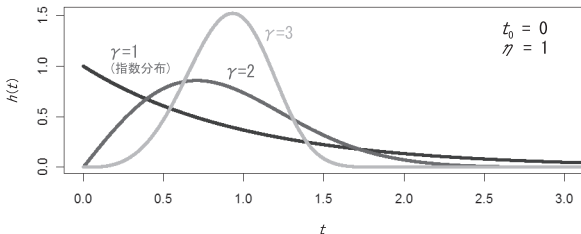
潜伏期間のモデリング(1)

- ワイブル分布

$$h(t|t_0, \gamma, \eta) = \frac{\gamma}{\eta} \left(\frac{t-t_0}{\eta} \right)^{\gamma-1} \exp \left(- \left(\frac{t-t_0}{\eta} \right)^\gamma \right)$$

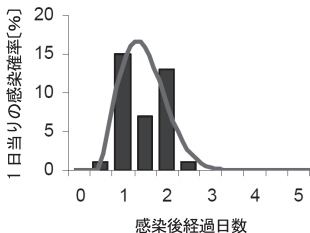
γ : 形状パラメータ
 η : スケールパラメータ
(t_0 : 感染成立時刻)

- 機械の故障時間のモデリングに使われる: 製造(感染)からの経過時刻が $t \sim t + \Delta t$ の間に確率 $h(t) dt$ で故障(発症)する。



潜伏期間のモデリング(2)

- 推定手段: 航空機内で1人の感染源から37人に伝染した事例
 - Moser+(1979), American Journal of Epidemiology 110, 1-10
 - 曝露期間がフライト時間の高々12時間と短いので, 2次感染者の出現時刻のヒストグラムをそのまま潜伏期間(as 確率変数)の実現値と見ることができる。



ベストフィット

$$h(t|t_0 = 0.5 \text{ 日}, \gamma = 2.21 \text{ 日}, \eta = 1.1 \text{ 日})$$



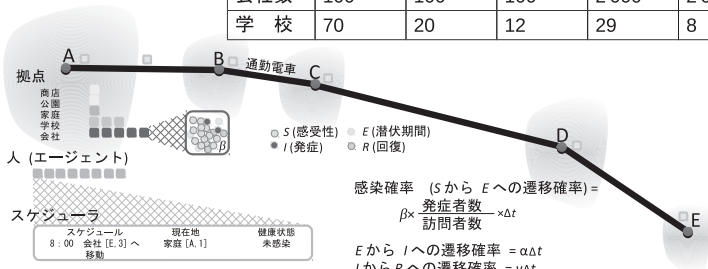
シミュレーションで使う感染力獲得待ち時間

$$T_{\text{latent}} = 1.48 \pm 0.47 \text{ 日}$$

シミュレーションの構成要素

約122万人のシミュレーション

	A	B	C	D	E
人口	571 641	176 866	138 684	314 861	44 680
会社数	100	100	100	2 000	2 000
学校	70	20	12	29	8



感染確率 (SからEへの遷移確率) = $\beta \times \frac{\text{発症者数}}{\text{訪問者数}} \times \Delta t$

EからIへの遷移確率 = $\alpha \Delta t$

IからRへの遷移確率 = $\gamma \Delta t$

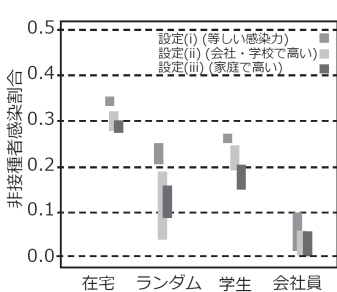
$1/\alpha$ = 平均潜伏期間 (~ 3日)

$1/\gamma$ = 平均発症期間 (~ 3日)

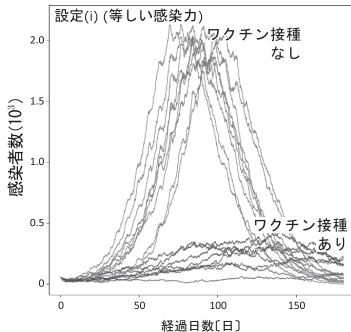
伝染モデル (場所ごとに確率的SEIR)



接種対象の選択と集団免疫



(a)

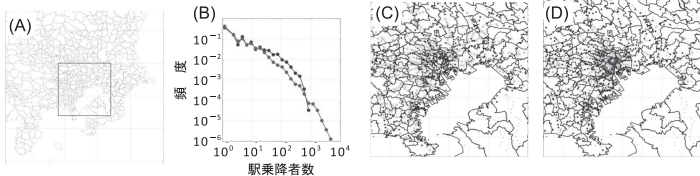


(b)

- (a): シミュレーション終了時(経過日数180日目)における非接種者に占める感染者の割合。色の違いは、感染力の設定方法の違い。
- (b): 設定(i)において会社員に集中投与した場合の感染者数の変化。感染流行を止めるとはできないが、感染の爆発が回避できていることがわかる。

首都圏鉄道網上で感染伝播(1)

- パーソントリップ調査にもとづく首都圏1 435駅をノード(地区)とする鉄道によるネットワーク上で139 841エージェントが行動するときの感染伝播シミュレーション
- エージェントは地区 i の家庭と地区 j の学校/会社とを往復する。
- ノード内の家庭または学校/会社に存在する感染者に比例して感受性者へ確率的に伝染する点は、事例2に類似



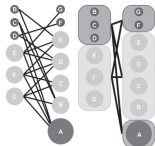
- Yashima, K. and Sasaki, A : Epidemic process over the commute network in a metropolitan area. PLoS ONE, **9**, 6, pp.1-13 (2014)

首都圏鉄道網上での感染伝播(2)

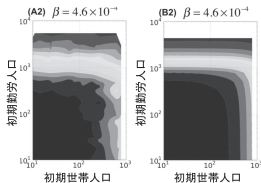
- 主結果: 大規模流行が発生する条件をエージェントシミュレーションで計算するとともに, 同一規模の地区を統合した, 粗視化モデルとの比較によって大規模流行発生確率は, 感染力と初期感染者が出現した地区の人口規模で(ほぼ)決定されることを示した。

エージェント 粗視化モデル

世帯人口 勤労人口 世帯人口 勤労人口

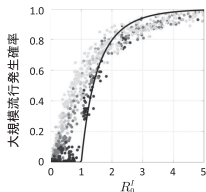


※大きさ/色は人口規模を表現



0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

ある感染力設定のもとでの
大規模流行発生確率



初期感染地区での再生産定数

10⁻⁴ 10⁻³ 10⁻² 10⁻¹

感染伝達係数 β

パンデミック・リアルタイム監視と予測

Ong et al., PLoS ONE 5(4) : e10036, 2010

- 2009年インフルエンザA(H1N1)パンデミックに際して、シンガポールにおける調査体制の形成と粒子フィルタによるベイズ予測の試み
- データ 535人の総合診療医(GPFD)/診療所にe-mailで呼びかけ、インフルエンザ様疾患(ILI)の週間発生件数を得た。
 - 23人の医師から協力の回答があった。
 - 急性呼吸器疾患(ARI)が伝統的に用いられている、パンデミック・インフルエンザがベースラインに埋もれてしまうため、独自に調査を行った。

症例報告シート

(性別, 出生年, 民族, 体温, 抗ウイルス剤投与の有無)

Doctor's Name: _____ Clinic ID: _____
Clinic Stamp: Date submitted: _____
Data date: _____
Total attendance seen: _____

Fill in data for all acute respiratory infections (ARI) seen on above date.
Fax form to 62581527 or email this file to jmmmy_ong@tsh.com.sg.
Turn on CAPS LOCKS if entering data.

S/No	Sex (M / F)	Year of birth	Ethnicity (C / M / I / O)	S'porean or PR vs others (S / O)	Temperature (degrees C)	Oseltamivir or Zanamivir prescribed (Y / N)

協力医療機関の分布



リアルタイム監視と予測（予測モデル）

- 予測モデル：確率的SEIRモデル



$$S_t = S_{t-1} - A_t, \quad A_t \sim \text{Bin}(S_{t-1}, 1 - \exp\{-[\epsilon + \beta I_{t-1}]/N\})$$

$$E_t = E_{t-1} + A_t - B_t, \quad B_t \sim \text{Bin}(E_{t-1}, 1 - \exp\{-\lambda\})$$

$$I_t = I_{t-1} + B_t - C_t, \quad C_t \sim \text{Bin}(I_{t-1}, 1 - \exp\{-\gamma\})$$

$$R_t = R_{t-1} + C_t$$

(総人口 $N = S_t + E_t + I_t + R_t$)

潜伏期間 λ^{-1} , 感染期間 γ^{-1} , 感染伝達率 β , ベースライン ILI 発生率,
1 週間に新たに出現した感染者 A_t , 発症者 B_t , 回復者 C_t

$\text{Bin}(n, p)$: 二項分布

リアルタイム監視と予測(観測モデル)

- 観測モデル (シミュレーションとデータとの対応付け)

$$D_t \sim \text{Pois} \left[N_t \delta_{t(\bmod 7)} \left(\phi + \frac{0.83 \times I_t}{1730} \right) \right]$$

報告された時刻 t における新規感染者数(データ) D_t

時刻 t に回答した診療所の数 N_t (≤ 23)

受診率=0.83, シンガポールの総合診療所数=1730

$\text{Pois}(\lambda)$: ポアソン分布

- 予測分布=(データ+予測・観測モデル) × 粒子フィルタ

$$\Pr(I_t | D_s, \dots, D_1) \approx \frac{1}{N} \sum_{t=1}^N \mathbb{1}(I_t = I_t^{(t)}), \quad t > s$$

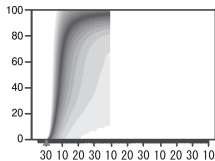
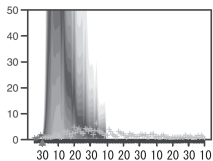
ここで, $\mathbb{1}(\text{真}) = 1$, $\mathbb{1}(\text{偽}) = 0$

予測時刻

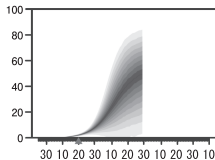
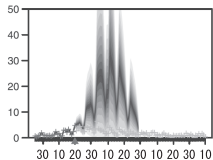
感染者数 I_t

累積感染者数

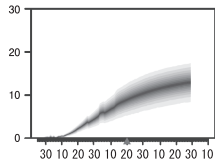
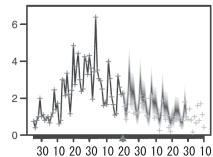
6月30日



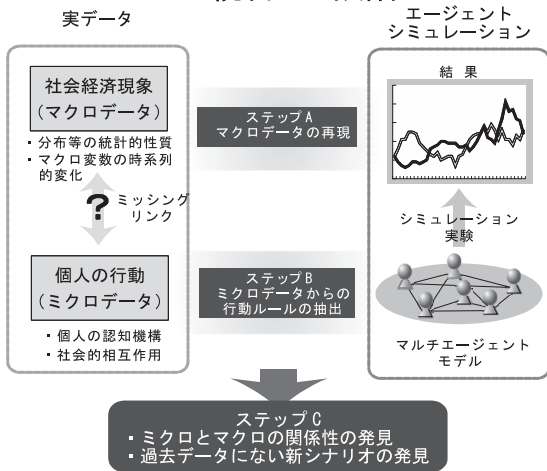
7月20日



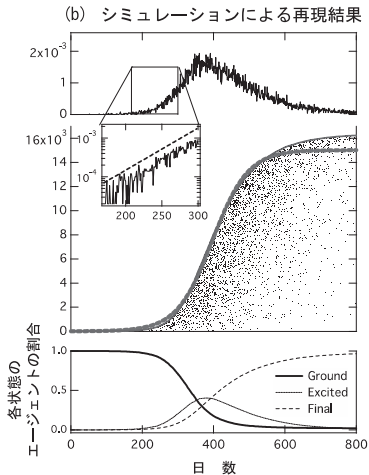
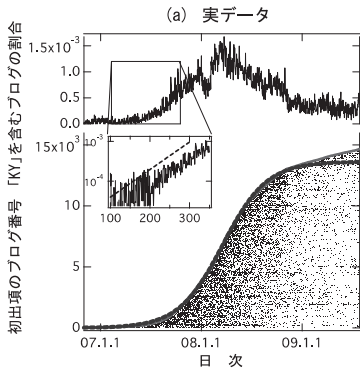
8月20日



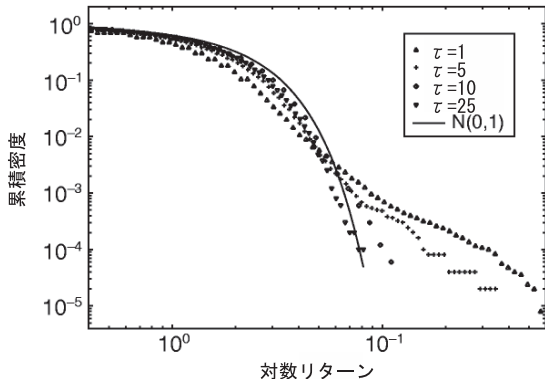
データ解析とエージェントシミュレーション の統合の3段階



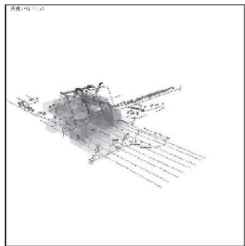
流行語の出現シミュレーション



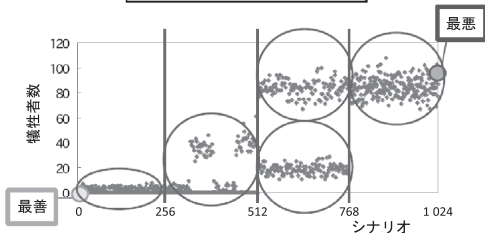
価格変動の累積密度分布 人工市場シミュレーション



避難シミュレーションと犠牲者数

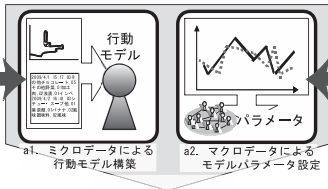


避難シミュレーション

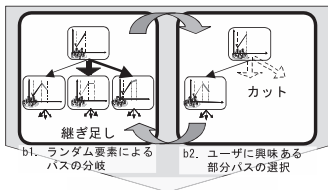


(a) 実データに基づく
モデル構築

ステップ B
マイクロデータからの
行動ルールの抽出



(b) 特定ケースの
作成



(c) エージェントの
内部状態分析



特定エピソードの作成

