

2019年度夏季実習3日目

生化学反応系の数理モデル -パラメータフィッティング-

東京大学 理学系研究科 黒田研

大野 聡

今日使うファイル類は

<http://kurodalab.bs.s.u-tokyo.ac.jp/class/Summer/2019/Day3/>
に置いてあります（テキストエンコーディングはSJIS）

今日の流れ

今日使うファイル類は下記にあります

<http://kurodalab.bs.s.u-tokyo.ac.jp/class/Summer/2019/Day3/>

I. 講義

0. introduction
1. 細胞運命決定機構とRas, Rap1
2. パラメータ推定
3. 最適化

II. 課題

1. Ras, Rap1のODEモデルを作成
2. 1つ or 2つの未知パラメータを推定
3. 4つの未知パラメータをEPで推定

「課題」「発展課題」ができたならTAのチェックを受けてください。

14:30までに終了したい場合は「発展課題」まで終わらせてください。

14:30以降に終了する場合は「課題」まで終わらせてください。

終わったら、授業の途中で帰っても構いません。

I. 講義

0. introduction

1. 細胞運命決定機構とRas, Rap1
2. パラメータ推定
3. 最適化

前回の復習

生化学反応を常微分方程式（ODE）で記述



数値計算で解く（数値積分）



解（反応系の振る舞い）の解析

そもそも、なぜモデルを作るの？

現象を定量的に理解したい

- 観測データの再現には、どのパラメータが重要？

予測したい

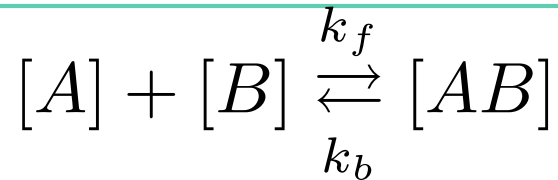
- 面白い現象を見るには、どんな実験条件が適切？

実験で確かめるのが困難なことを知りたい

- コスト・技術・倫理などの問題で、そもそも実験できないことを知りたい

モデルを使ったシミュレーション

反応式

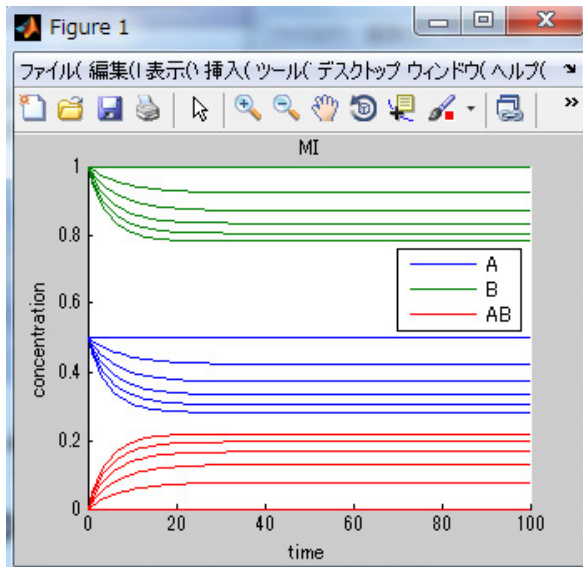


モデル(数式)

$$\frac{d[AB]}{dt} = k_f[A][B] - k_b[AB]$$

数値シミュレーション

モデル = 構造 + パラメータ
(定性的) (定量的)



モデルさえ決まってしまうば
様々な条件下での振る舞いが
シミュレーション可能！

⇒ 解析、実験デザインに有用

構造は既知とする。 パラメータはどうやって決める？

パラメータの決め方

- 文献・データベースを参照
→ 実験条件がばらばら。使って大丈夫？
- (速度定数などを)実験的に測る
→ 手間がかかる。そもそも測定できる？

そもそもモデルとは、現実の系を簡略化したもの。

真の速度定数がわかっていても、そのモデルで現象を再現できるとは限らない。

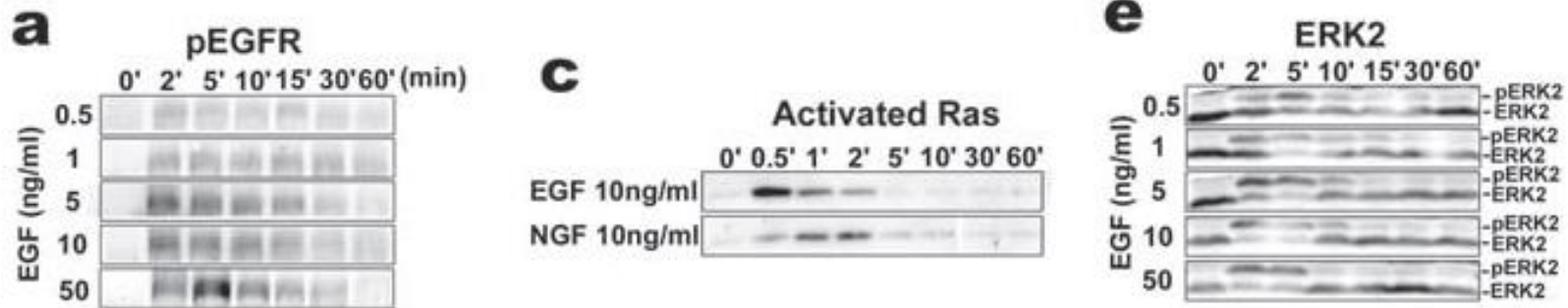
では、パラメータはどうやって決めるか？

パラメータは、**再現したい現象とモデル**に依存して決まる。

例：

細胞を刺激することで測定された
各シグナル分子の時間波形を再現するモデルが欲しい

実験データ：ウエスタンブロッティング（WB）や
イメージングなどにより取得

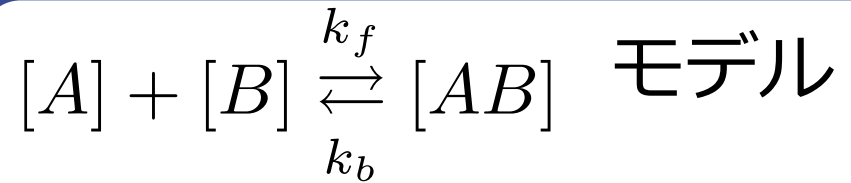
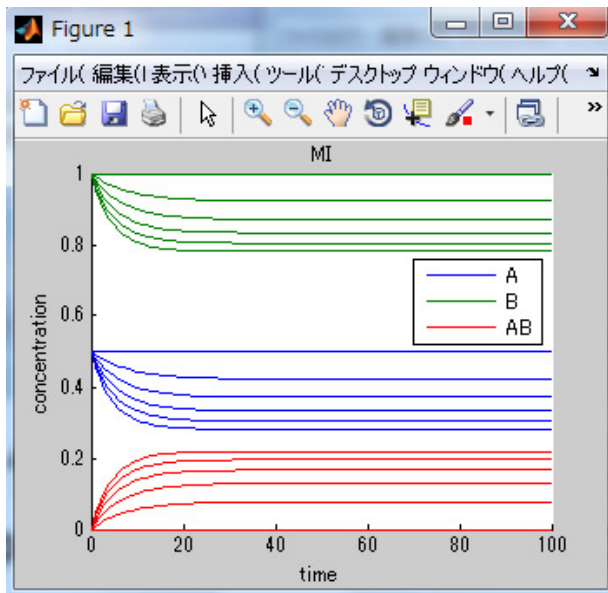


モデルの構造を決定

➡ このモデルで実験データを再現するパラメータを推定

今日のテーマ

シグナル分子活性の経時変化のデータから、
生化学反応の速度パラメータを推定する



$$\frac{d[AB]}{dt} = k_f [A][B] - k_b [AB]$$

I. 講義

0. introduction

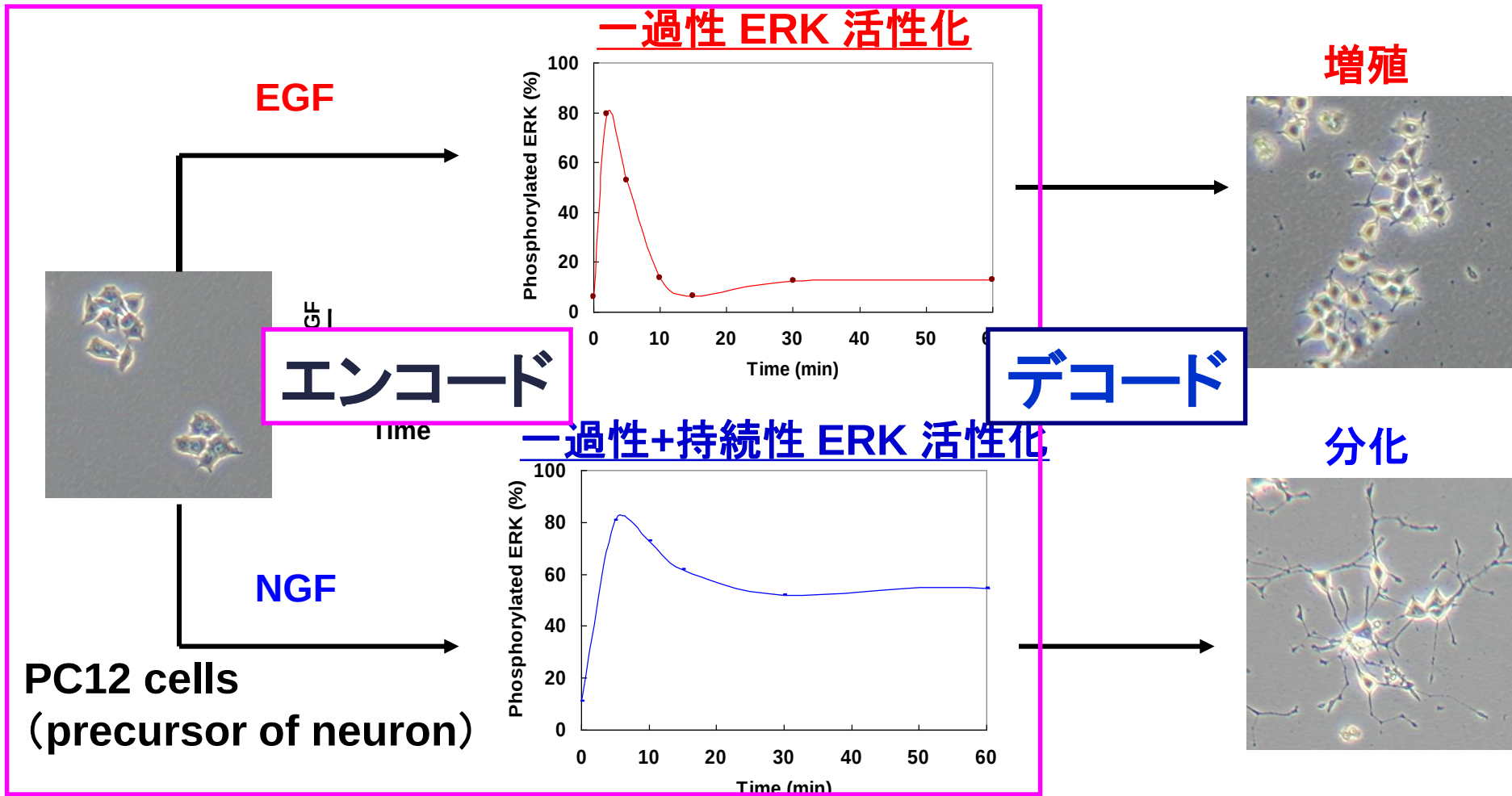
1. 細胞運命決定機構とRas, Rap1

2. パラメータ推定

3. 最適化

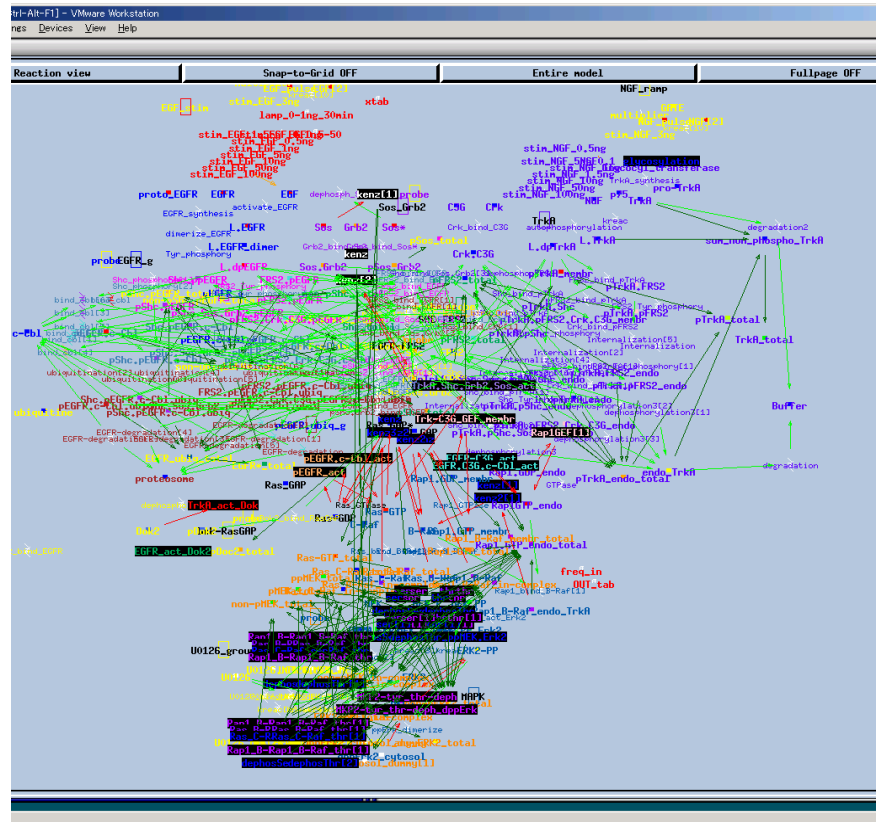
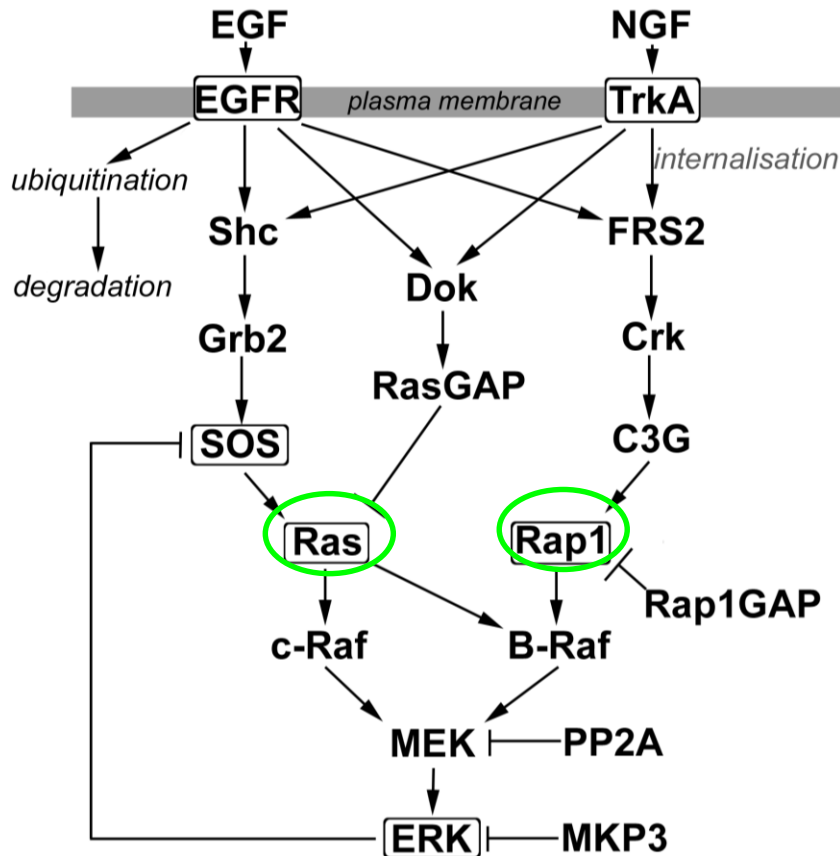
細胞運命決定機構：増殖と分化のスイッチ

ERKの時間波形が異なるだけで異なる現象を制御



Ras経路とRap1経路の生化学反応モデル

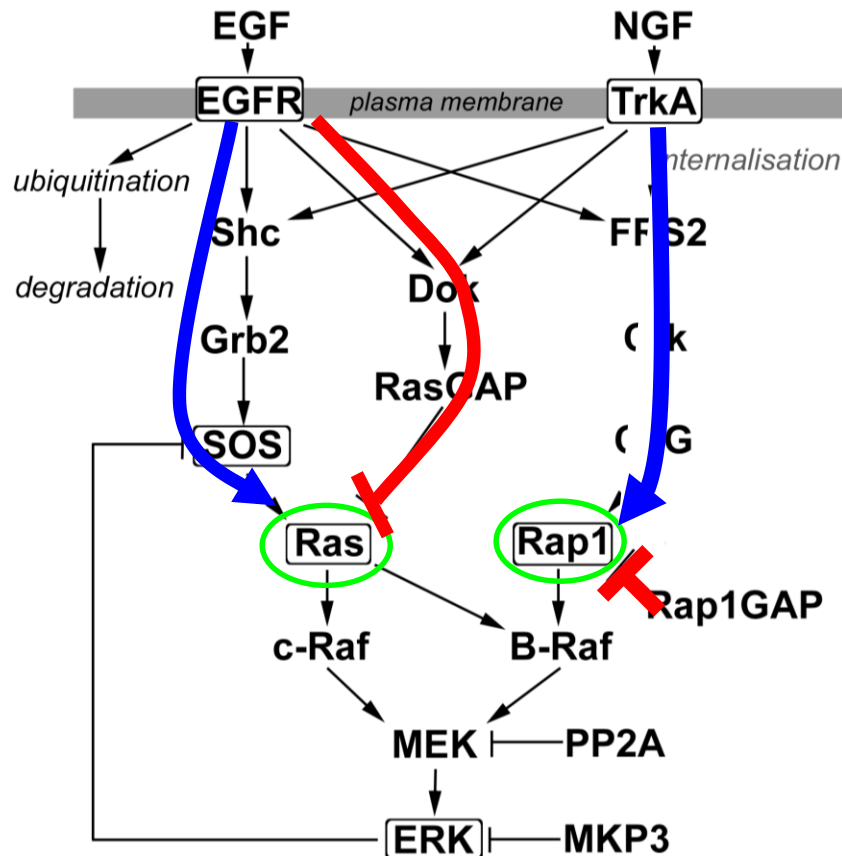
Erkのシグナル伝達経路



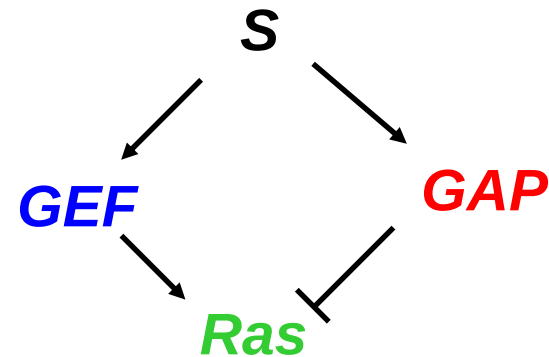
変数・パラメータが多すぎる！ どの作用が大事なの？

モデルの簡略化

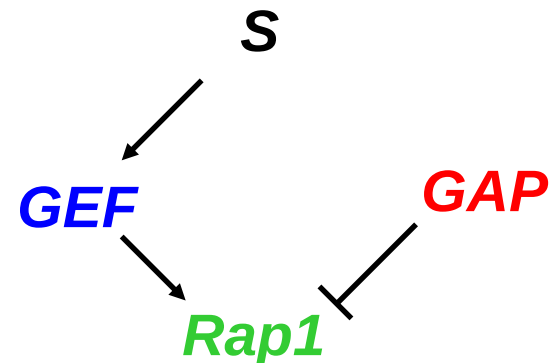
Erkのシグナル伝達経路



シンプルRasモデル

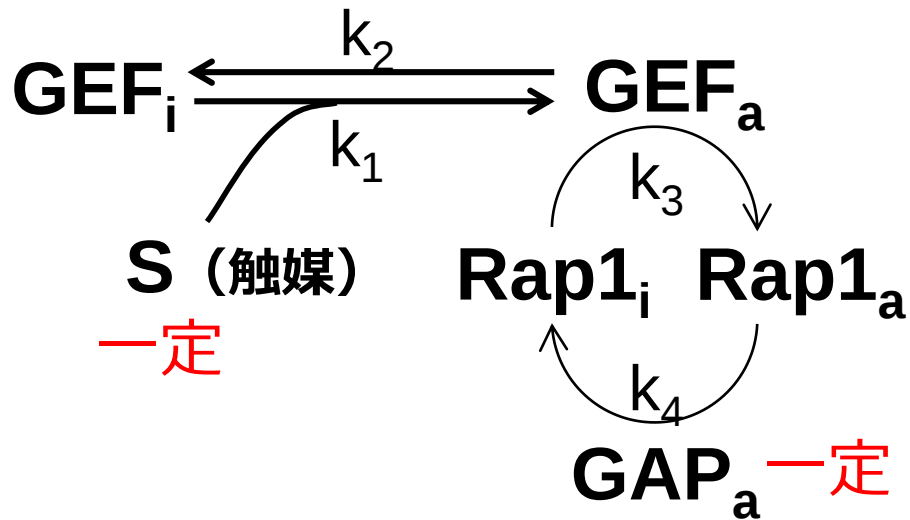


シンプルRap1モデル



シンプルRap1モデルとシンプルRasモデル

シンプルRap1モデル



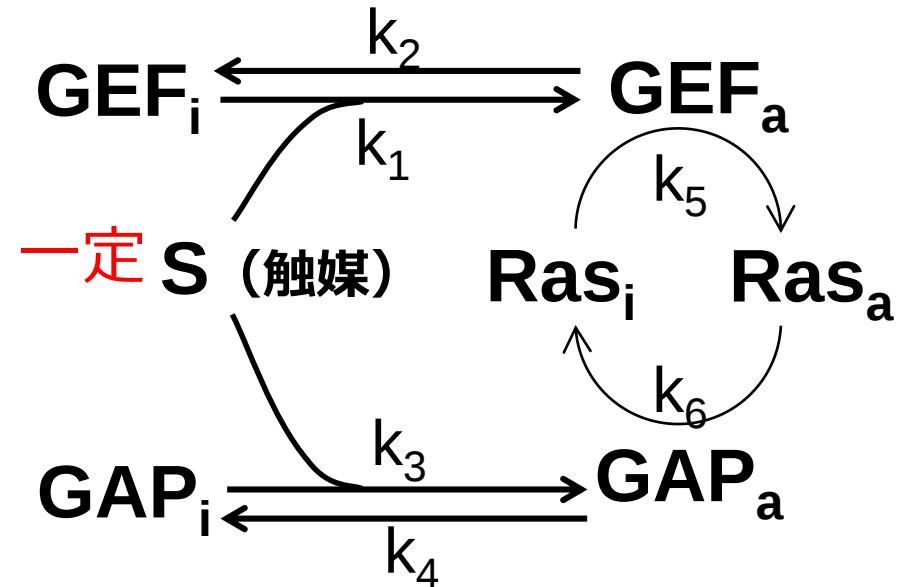
不活性型と活性型の総和は保存

$$\text{GEF}_i + \text{GEF}_a = 1$$

$$\text{GAP}_i + \text{GAP}_a = 1$$

$$\text{Rap1}_i + \text{Rap1}_a = 1$$

シンプルRasモデル



不活性型と活性型の総和は保存

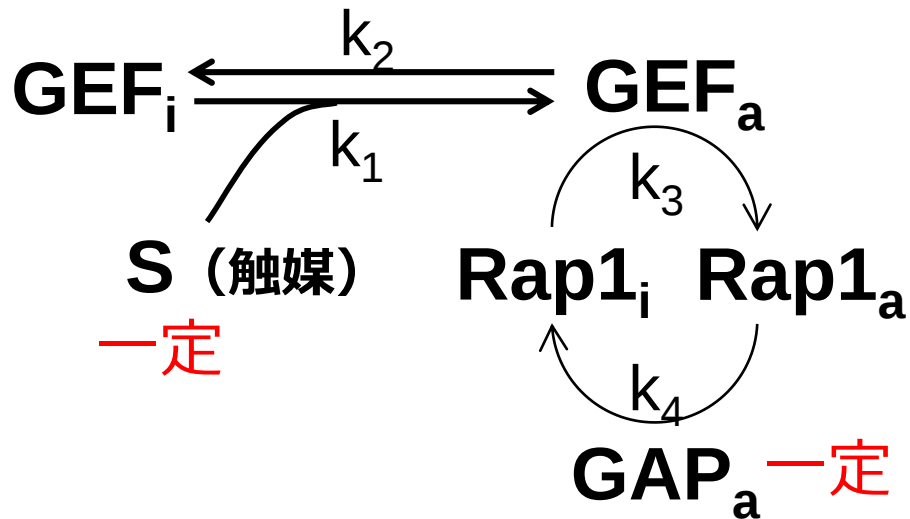
$$\text{GEF}_i + \text{GEF}_a = 1$$

$$\text{GAP}_i + \text{GAP}_a = 1$$

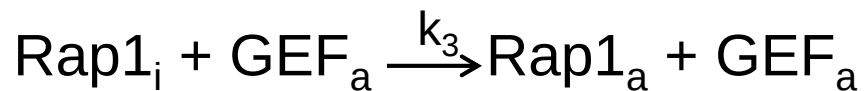
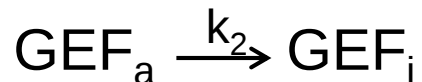
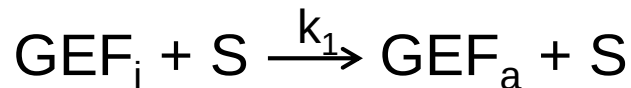
$$\text{Ras}_i + \text{Ras}_a = 1$$

シンプルRap1モデルとシンプルRasモデル

シンプルRap1モデル

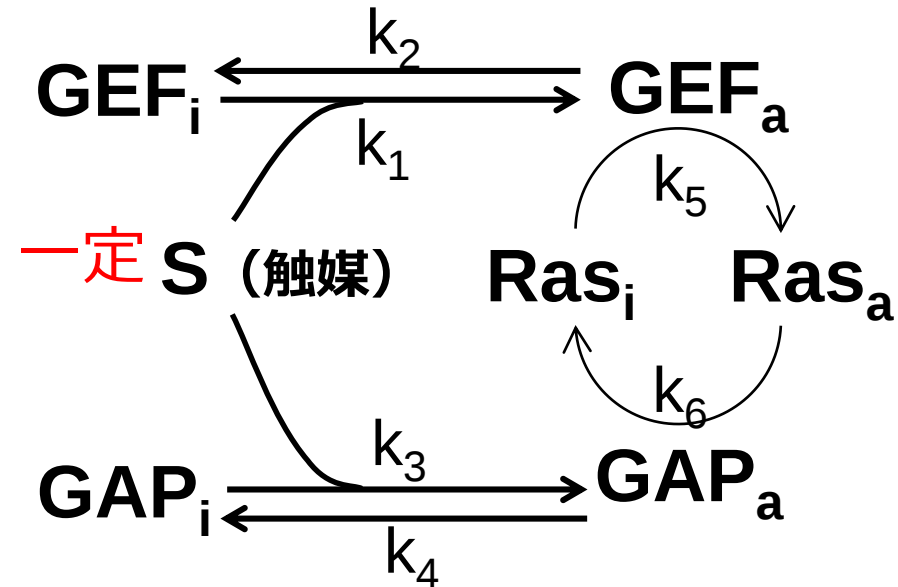


反応式は、



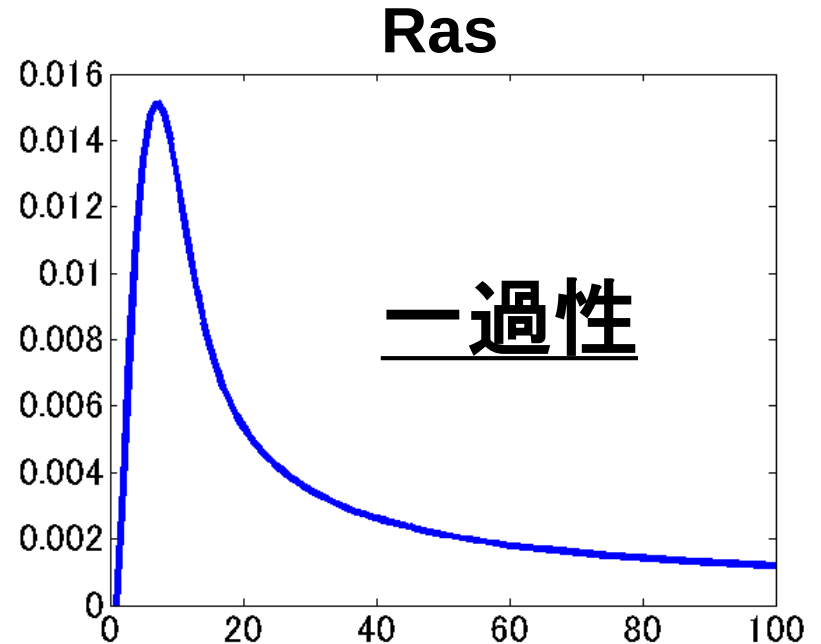
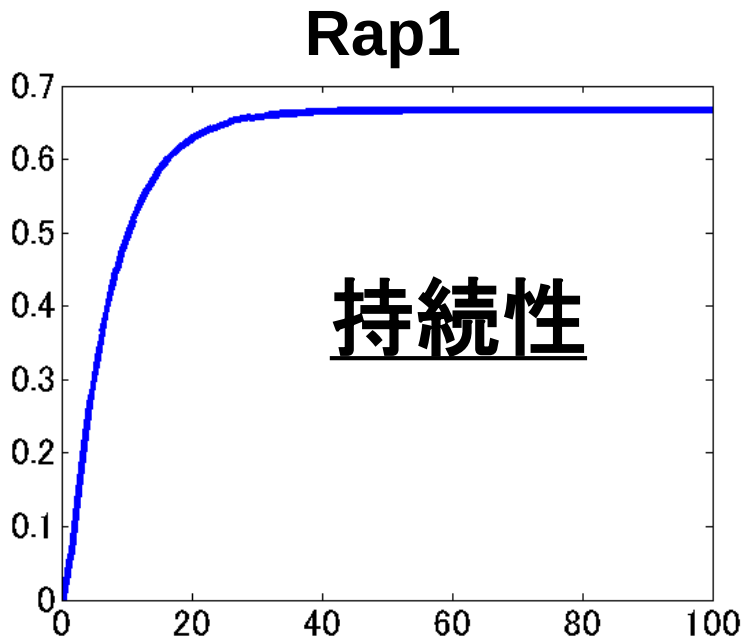
... のように書ける。

シンプルRasモデル



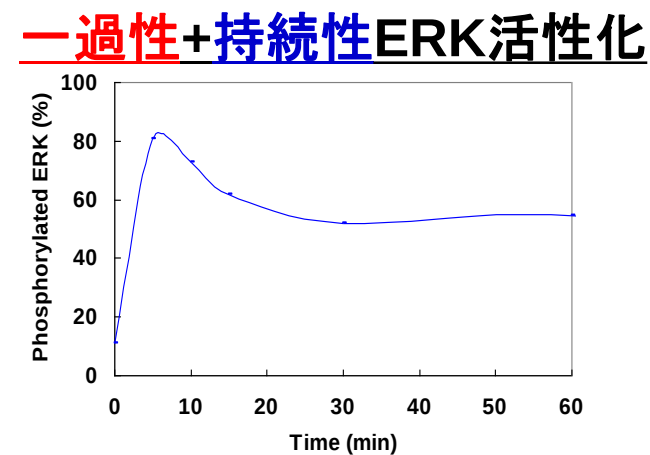
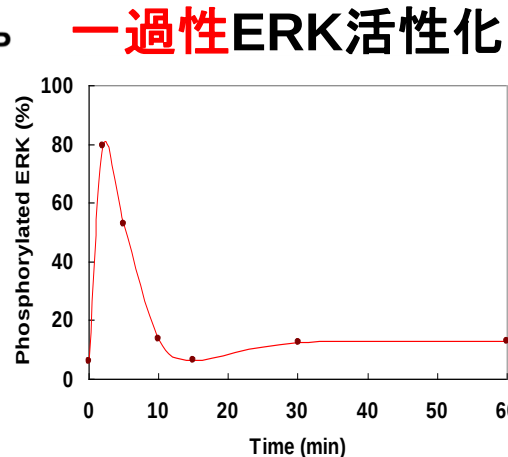
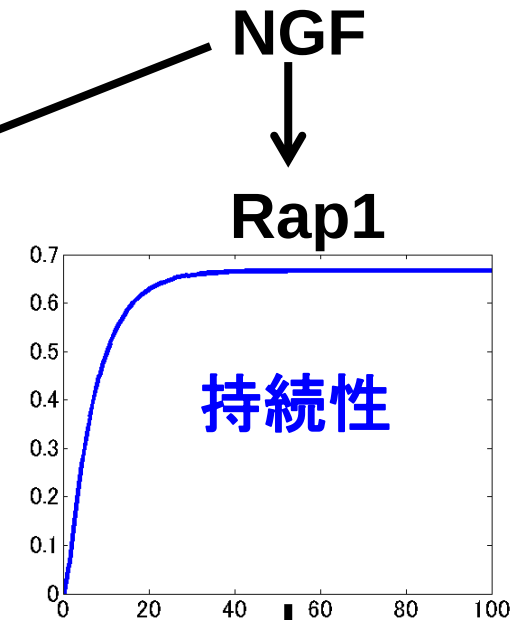
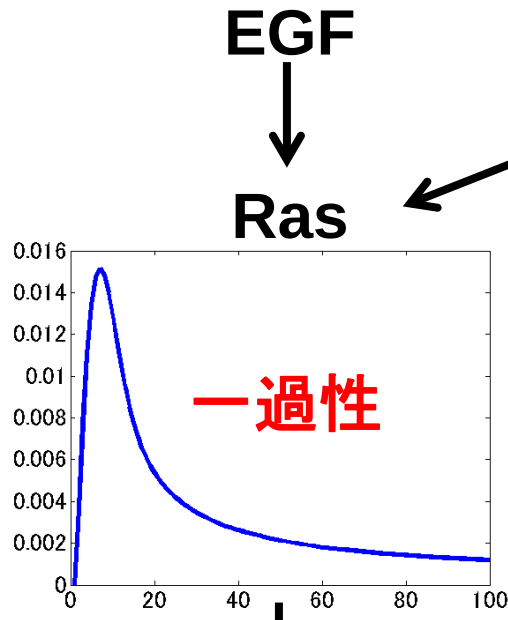
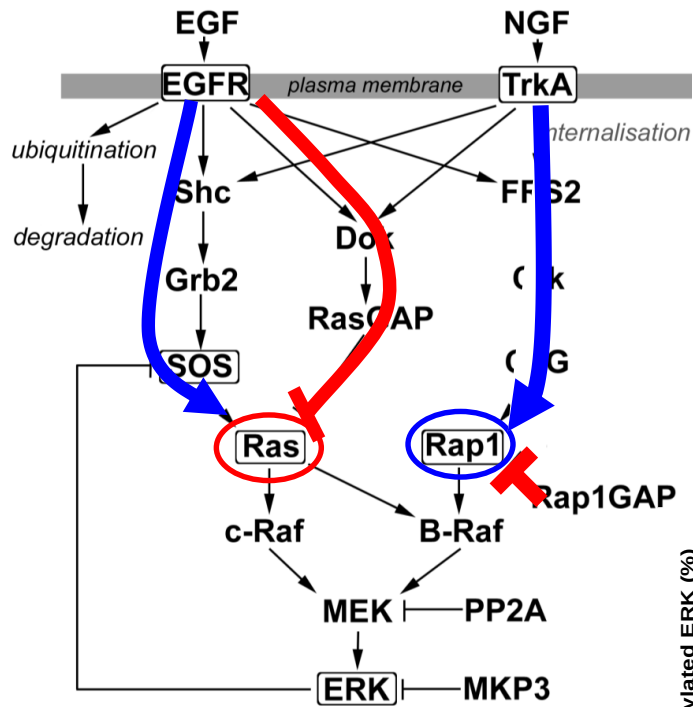
課題1では実際にモデルを作ってシミュレーションを実行してもらいます。

RapとRasの応答の違い



ERKの**一過性活性化**はRasに、ERKの**持続性活性化**はRap1に依存しているようだ。

Erk経路のブロック線図



I. 講義

0. introduction

1. 細胞運命決定機構とRas, Rap1

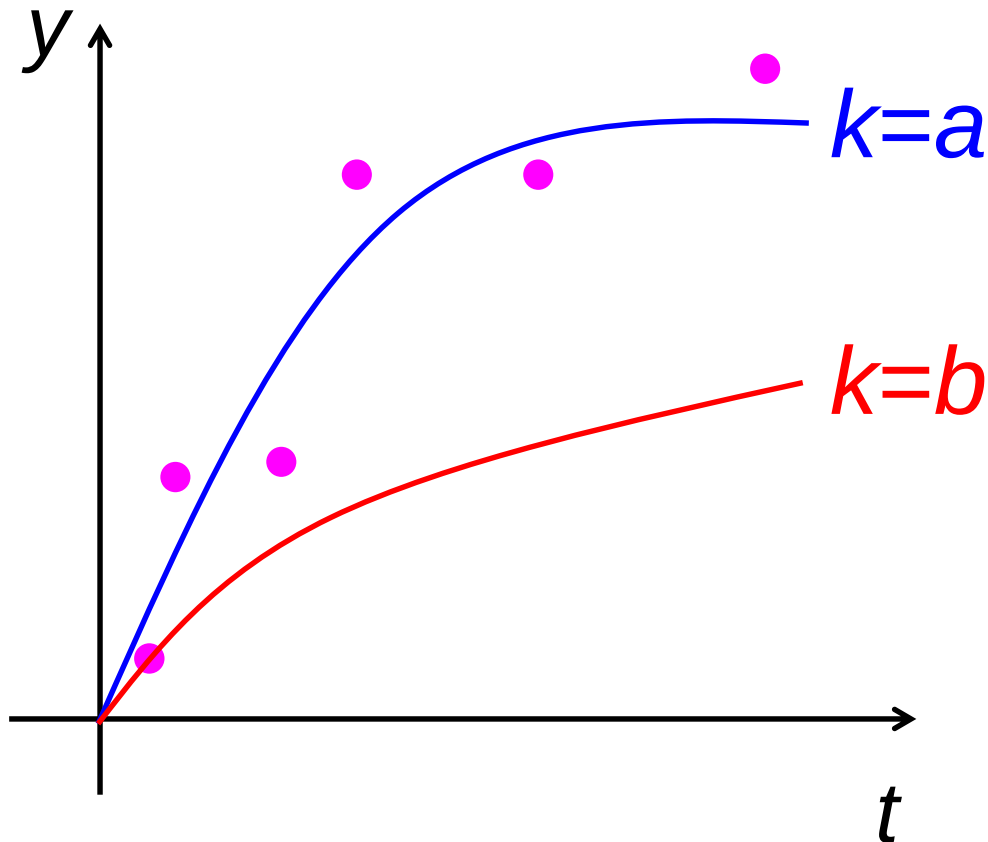
2. パラメータ推定

3. 最適化

パラメータの決め方

● データ点

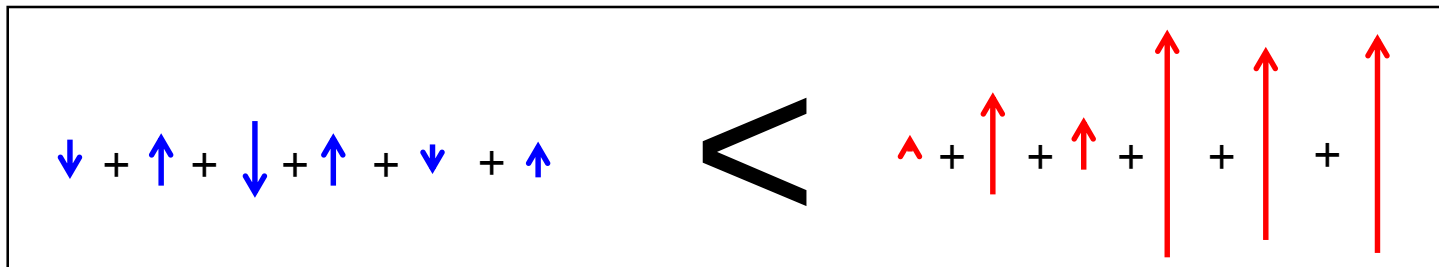
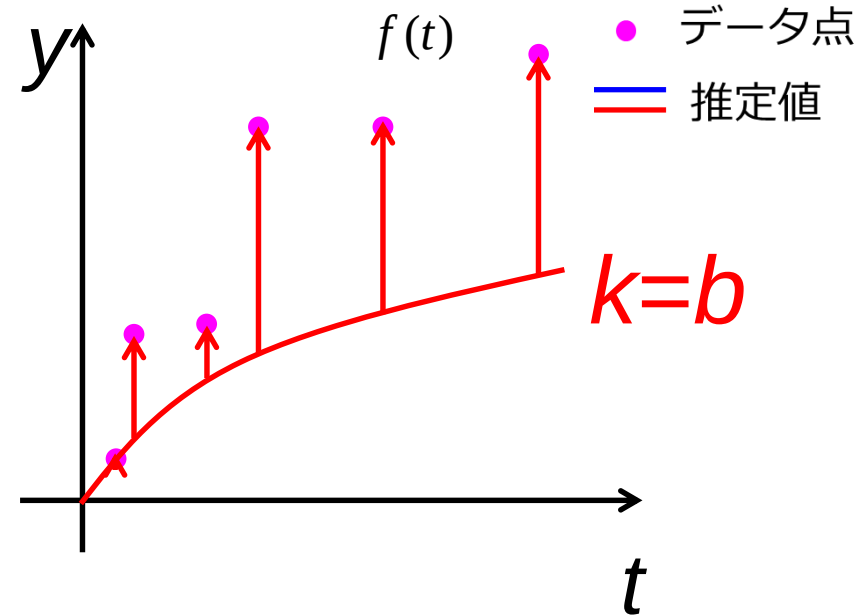
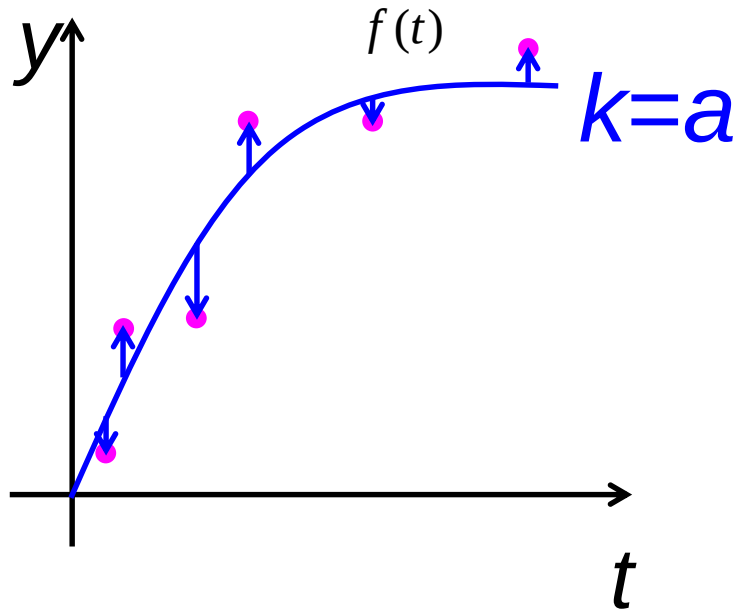
— モデルの出力



$k=a$ と $k=b$ のうち、
どちらがよい
パラメータか？

客観的かつ定量的に
議論するには
どうすればよいか？

パラメータの決め方



$k=a$ のほうがよさそう

客観的かつ定量的な指標→評価関数

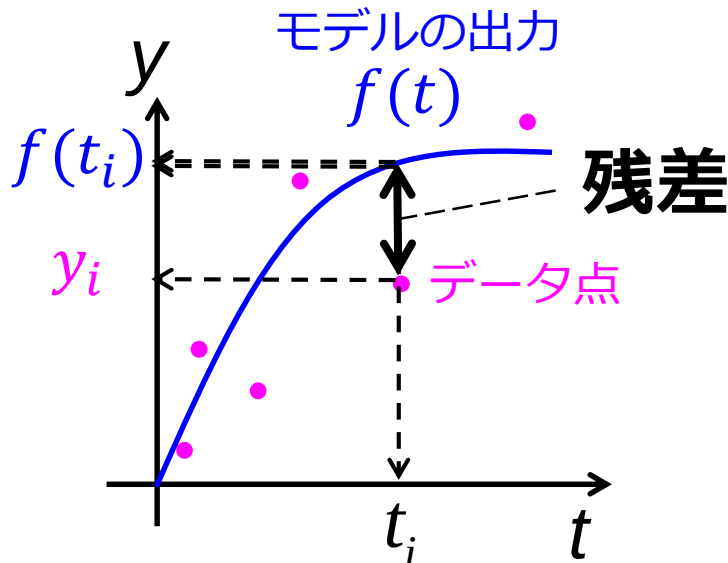
評価関数の例：残差二乗和

○残差

$y_i - f(k, t_i)$
データ点 (実験値) モデルの出力 (推定値)

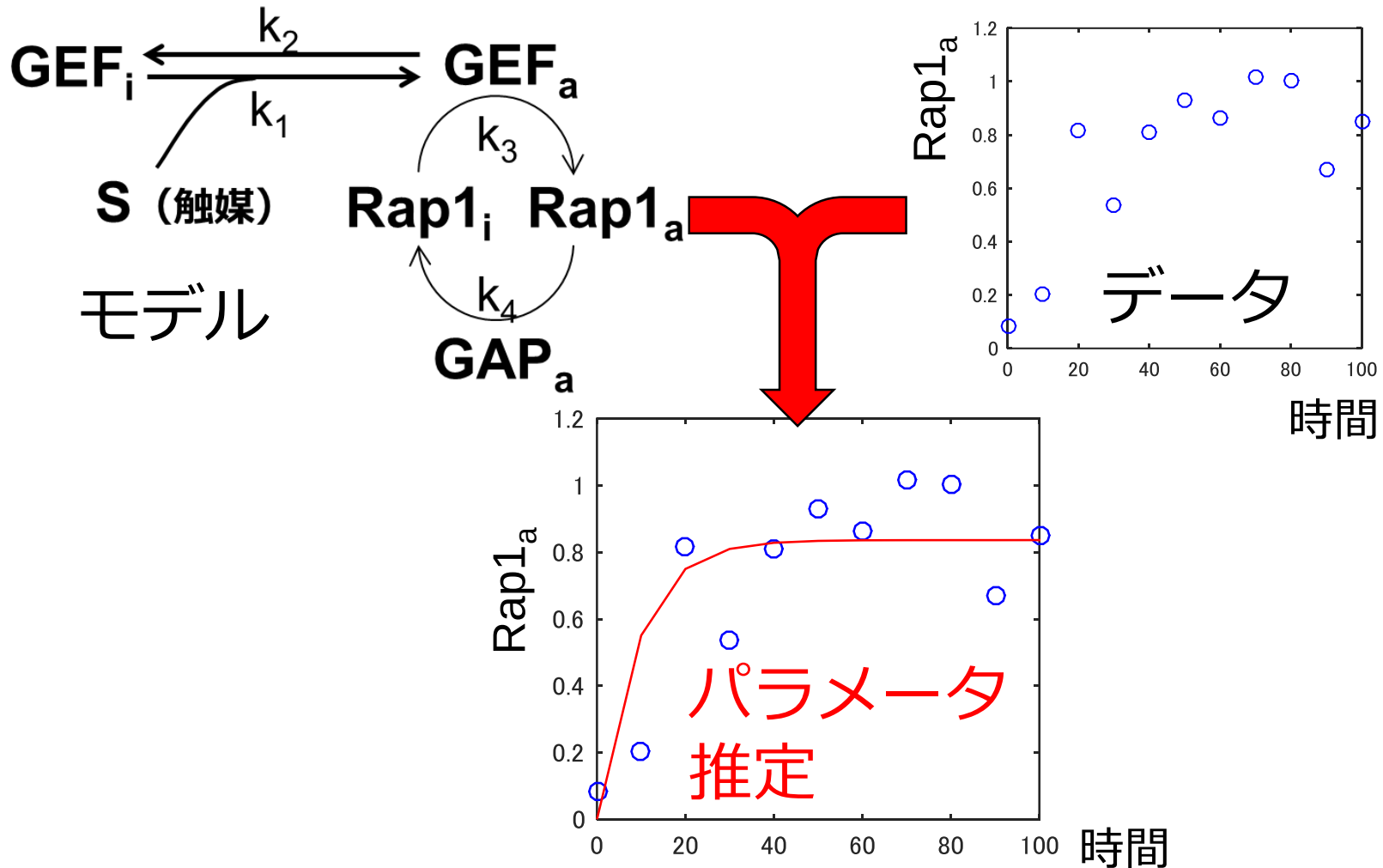
k : パラメータ、 t : 時間
モデルの出力（シミュレーションで得られる分子の時間波形）と、データ点の差

○残差二乗和 $\varepsilon^2 = \sum_{i=1}^N (y_i - f(k, t_i))^2$ データ点の数 ユークリッド距離に相当



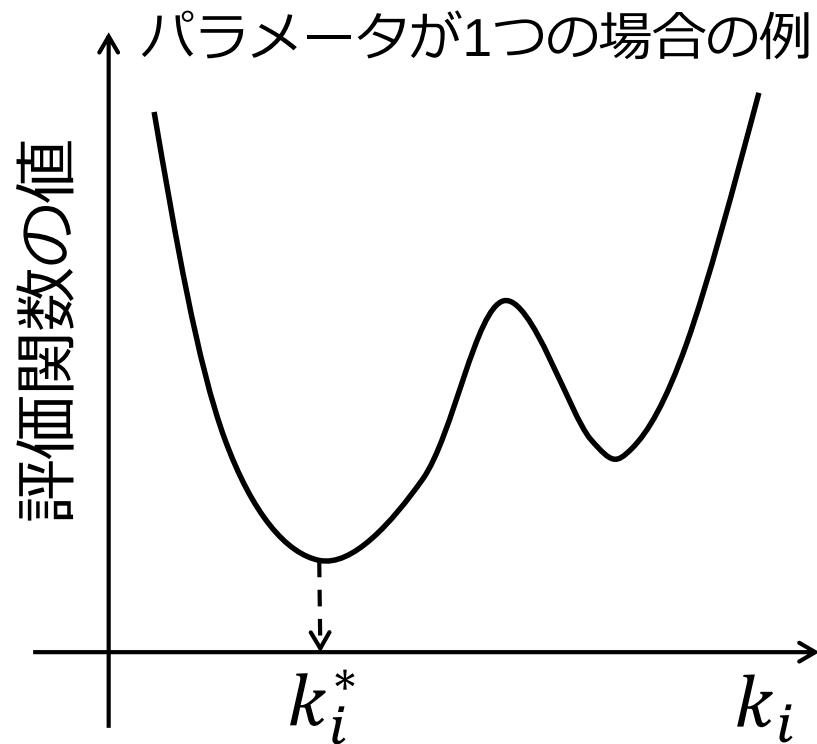
残差二乗和を評価関数として、
最小にするようにパラメータを
決める方法を最小二乗法という。

モデルとデータから、最小二乗法により パラメータを推定する

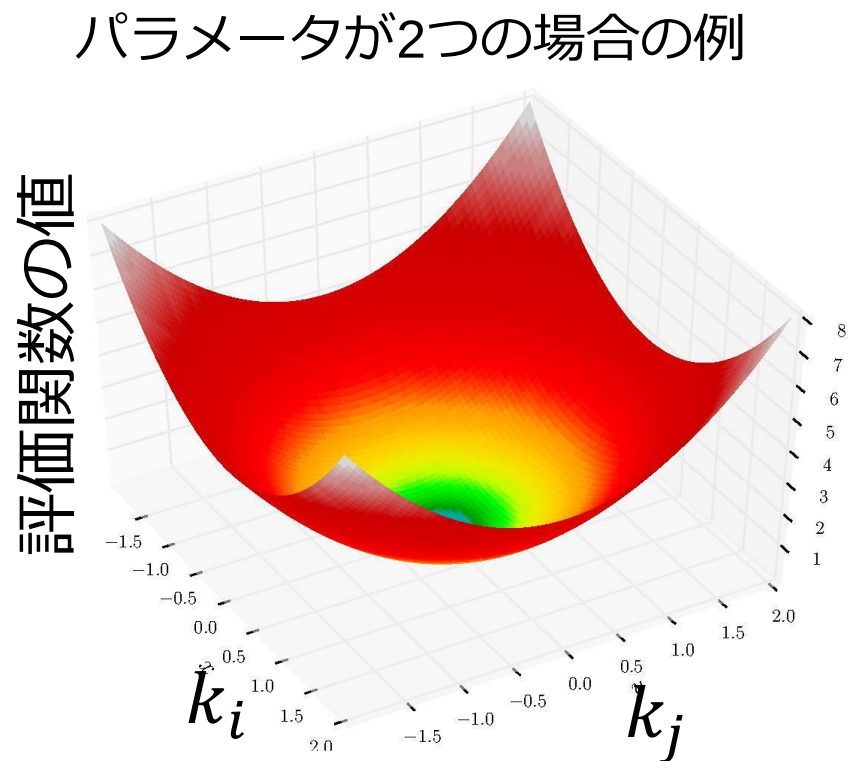


課題2で実際にパラメータ推定を行います。

評価関数の景観 (landscape)



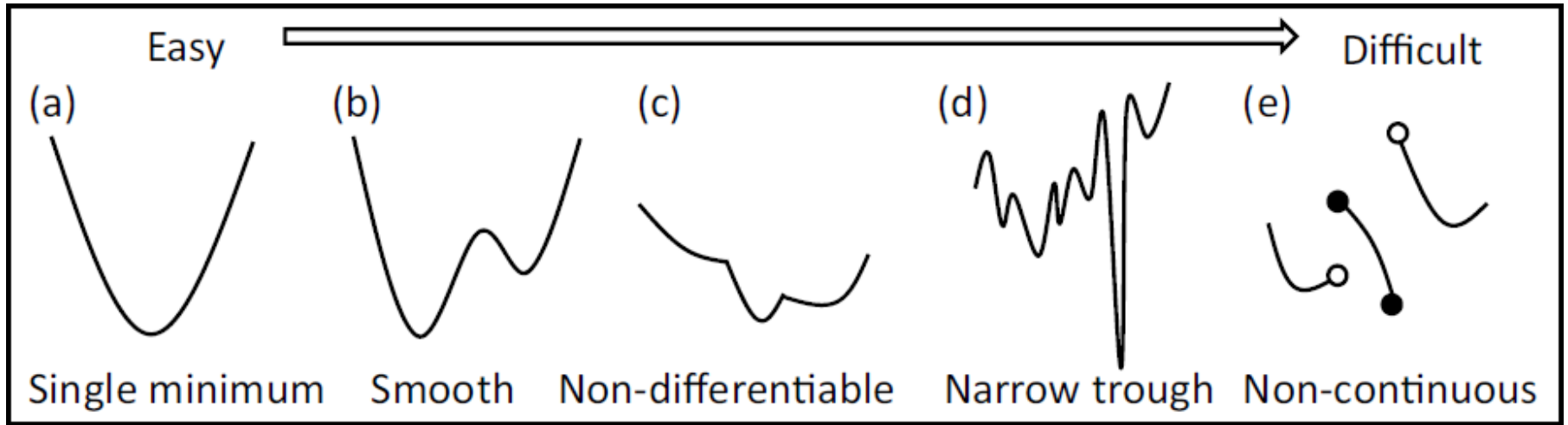
パラメータ k_i の最適解



評価関数の景観は、モデルの構造・データ・推定するパラメータに依存して決まる。

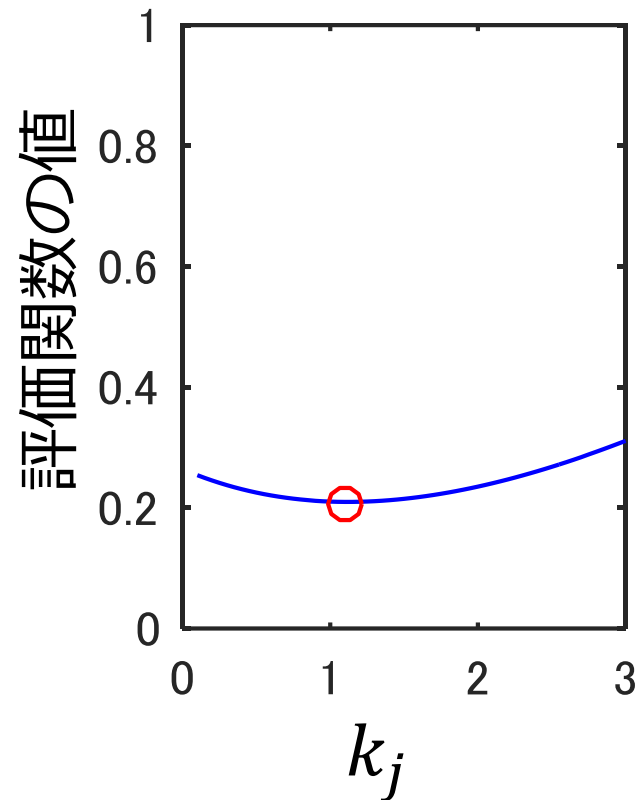
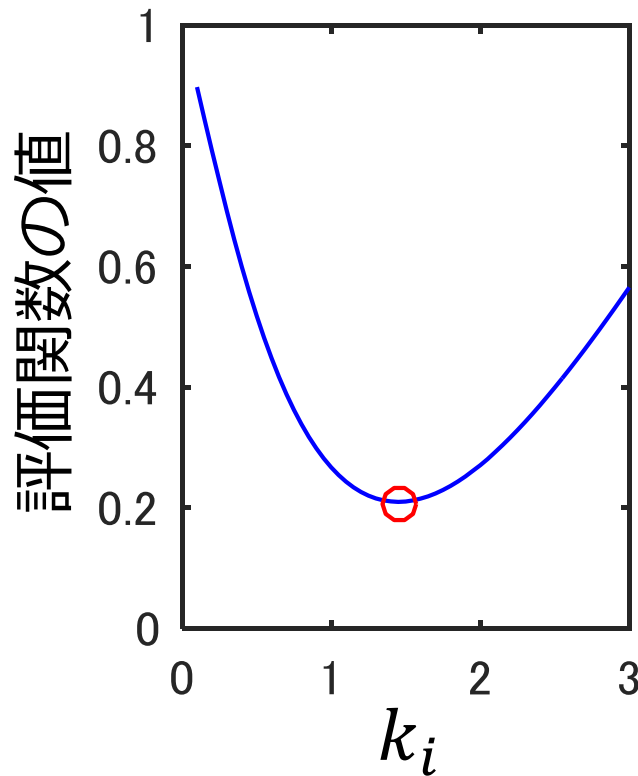
評価関数の景観と最適化の難しさ

右のものほど最適化が難しい



K. Hoki and T. Kaneko, "Large-Scale Optimization for Evaluation Functions with Minimax Search", Journal of Artificial Intelligence Research (JAIR), 2014, Volume 49, pages 527-568

パラメータによって評価関数の景観が異なる



k_i と比較して...

k_j の変化に対する評価関数の値の変化は小さい

⇒ k_j の推定値は、誤差の影響を受けやすく、信頼性が低い

I. 講義

0. introduction

1. 細胞運命決定機構とRas, Rap1

2. パラメータ推定

3. 最適化

パラメータの数が多いと推定が大変...

パラメータの数が少ない

→ for loopでしらみつぶし (力まかせ探索)

パラメータ数が多い

→ しらみつぶしに探すのは難しい

単純に、パラメータの数だけ for loopを回すと…？

→ (パラメータ候補の数) (パラメータ数) : 膨大な回数

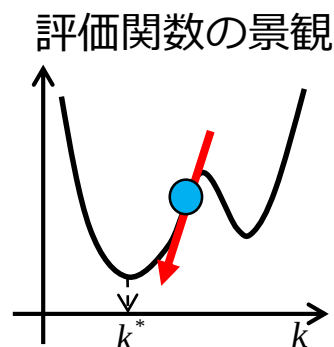
**どうやってデータをよく再現する
(評価関数の値が小さい)
パラメータを探すか？**

最適化手法を使って評価関数の値を最小化

変数が連続値をとる場合について

勾配 (微分)を用いた最適化

- 最急降下法
- 共役勾配法
- ニュートン法
- 信頼領域法
- など



局所的な最適化に向いている。
収束性や収束速度など理論も進んでいる。

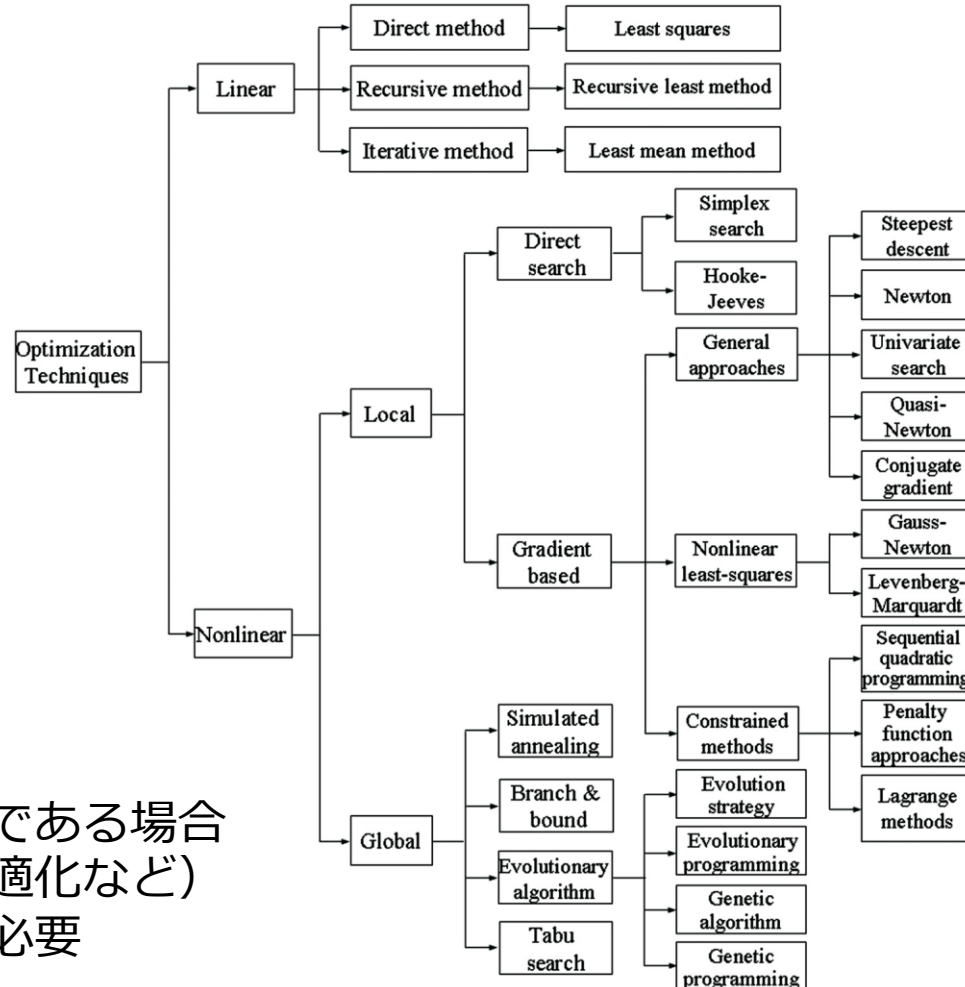
確率を用いた最適化

- Nelder-Mead法
- シミュレーテッドアニーリング (SA)
- タブー探索
- 遺伝的アルゴリズム (GA)
- 進化的プログラミング (EP)
- 粒子群最適化 (PSO)
- など

大域的な最適化に向いている。ただし、得られた解の最適性は保証されない。

参考：もう少し細かい最適化の分類

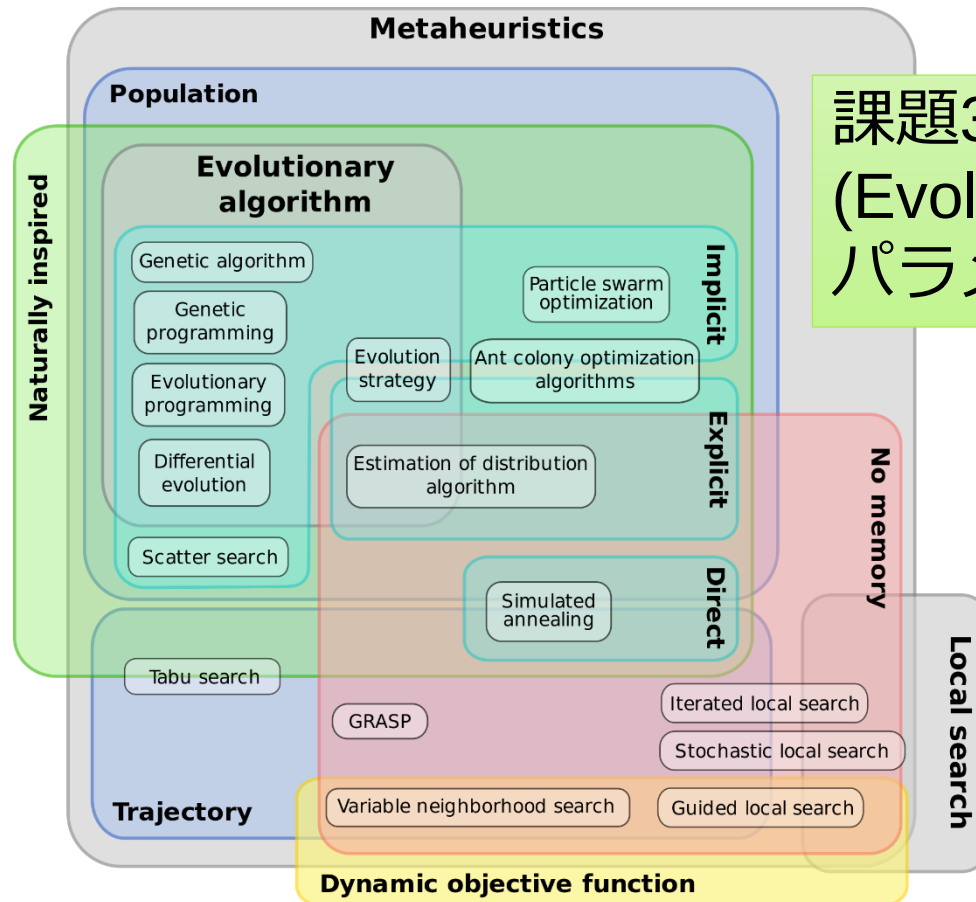
変数が連続値をとる場合について



※変数が離散値である場合
(組み合わせ最適化など)
は別の最適化が必要

メタヒューリスティクスの最適化

特定の問題に限定されず、どのような問題に対しても汎用的に対応できるように設計された最適化手法



課題3では進化的プログラミング (Evolutionary programming, EP)でパラメータを推定します。

Evolutionary programming(EP, 進化的計算)

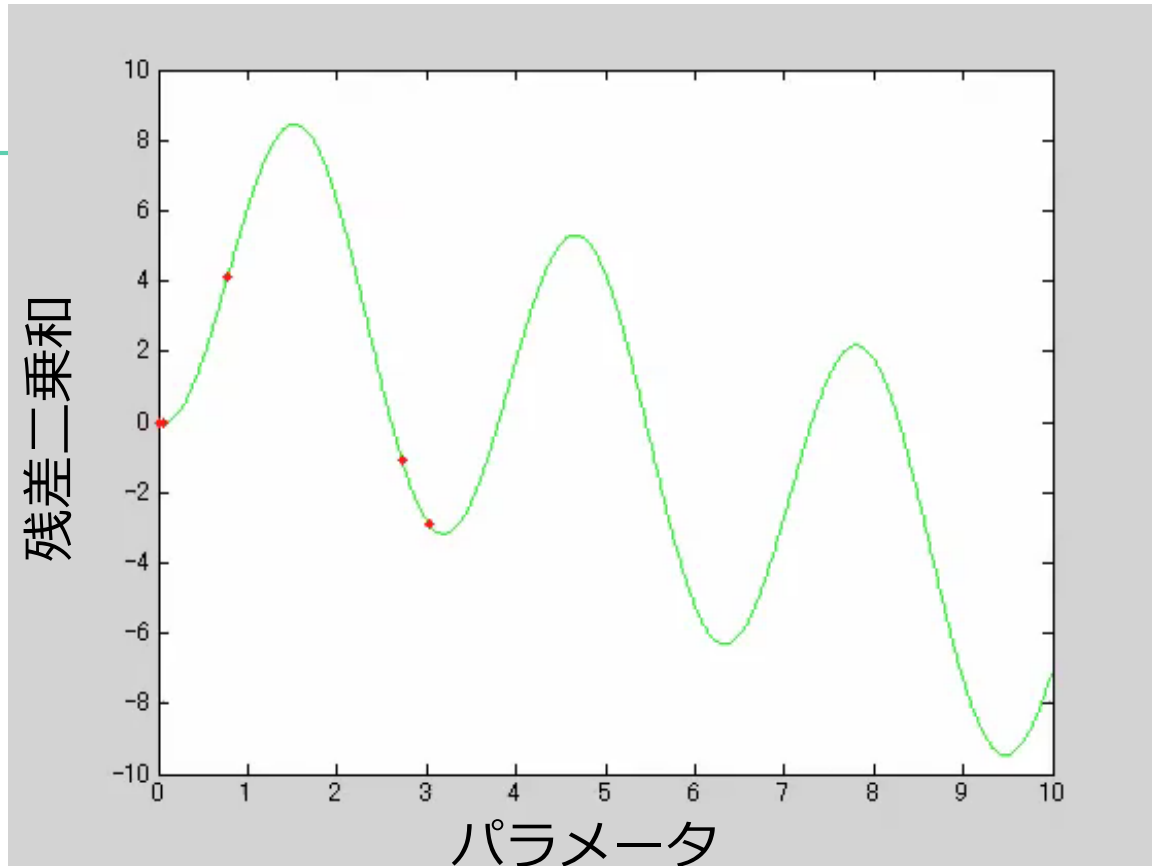
- 最適化アルゴリズムの一種
- 生物学における自然淘汰を模して作られている
- メタヒューリスティック(（ある程度）汎用性のある経験的な)

○大雑把な説明

- 1) パラメータを個体とみなして, 最初に N 個の個体がある
- 2) 各個体が子をつくる. 各子には確率的な変異が入る $\Rightarrow 2N$ 個体
- 3) 残差二乗和が小さいパラメータ $\rightarrow$ 優れている個体 $\rightarrow$ 生き残る
残差二乗和が大きいパラメータ $\rightarrow$ 劣っている個体 $\rightarrow$ 淘汰
 $\rightarrow$ 個体数を N 個にする.
- 4) 2)と3)を十分長く繰り返す
- 5) 最終的に最も優れた個体を, パラメータとして採用

EPのムービー

パラメータ数:1
個体数(N): 5



- 1) パラメータを個体とみなして, 最初に N 個の個体がある
- 2) 各個体が子をつくる. 各子には確率的な変異が入る $\Rightarrow 2N$ 個体
- 3) 残差二乗和を小さいパラメータ $\rightarrow$ 優れている個体 $\rightarrow$ 生き残る
残差二乗和が大きいパラメータ $\rightarrow$ 劣っている個体 $\rightarrow$ 淘汰
 $\rightarrow$ 個体数を N 個にする.
- 4) 2)と3)を十分長く繰り返す
- 5) 最終的に最も優れた個体を, パラメータとして採用

もう少し詳細なEPの説明

1. 乱数を初期値として, N 個の個体 (解の候補) を生成する.
2. 個体のそれぞれのコピーをつくる.
3. 2.で作った各コピーに正規乱数を加える. (変異)
4. 3.の操作で作られた新しい N 個の個体と, 元の個体を混ぜた $2N$ 個の各個体に対してスコアを求める. スコアは以下のよう
に決める.
 - a. スコアを決める対象の個体を除いて, $2N$ 個体の集団から無作為に q 個体を選ぶ.
 - b. 選んだ q 個体のうち, スコアを決める対象の個体より評価関数の値 (RSSなど) が悪い個体 (最小二乗法の場合は, 対象の個体よりRSSの値が大きい個体) の数をその個体のスコアとする.
5. スコアの下位 N 個の個体を削除する. (淘汰)
6. 終了条件を満たすまで2. ~7. (1世代に相当) の操作を繰り返す.
7. 残った中で最も評価関数の値が良い個体を“解”とする.

※EPには, バリエーションがいくつか存在する.

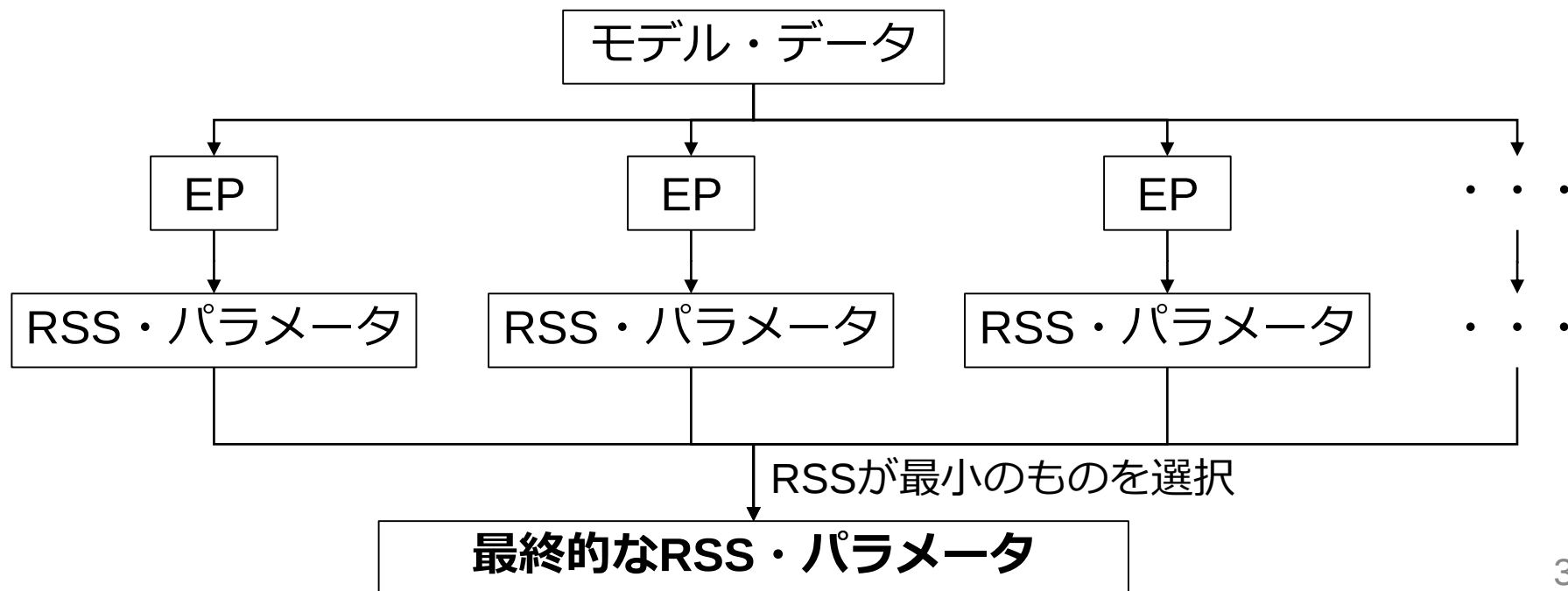
EPは複数回実行する

EPでは乱数を用いる

- ・ 初期値生成、個体生成時の変異

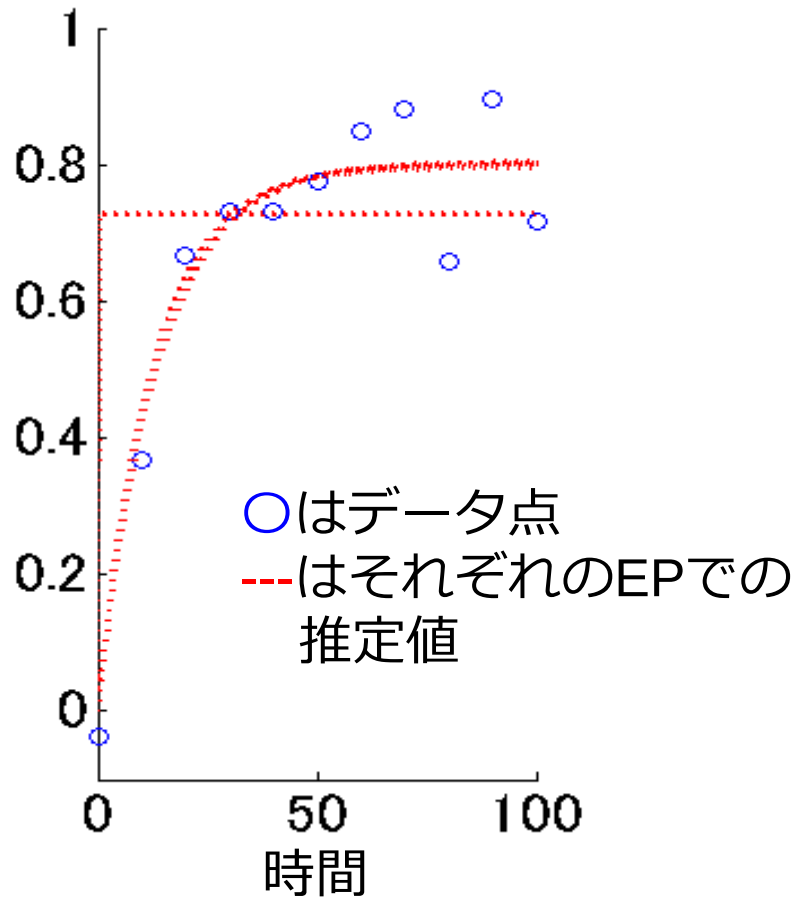
そのため、EPは実行するたびに結果が異なる。

そこで、複数回EPを実行して、その中で最も残差二乗和が小さいものを最終的な解とする。

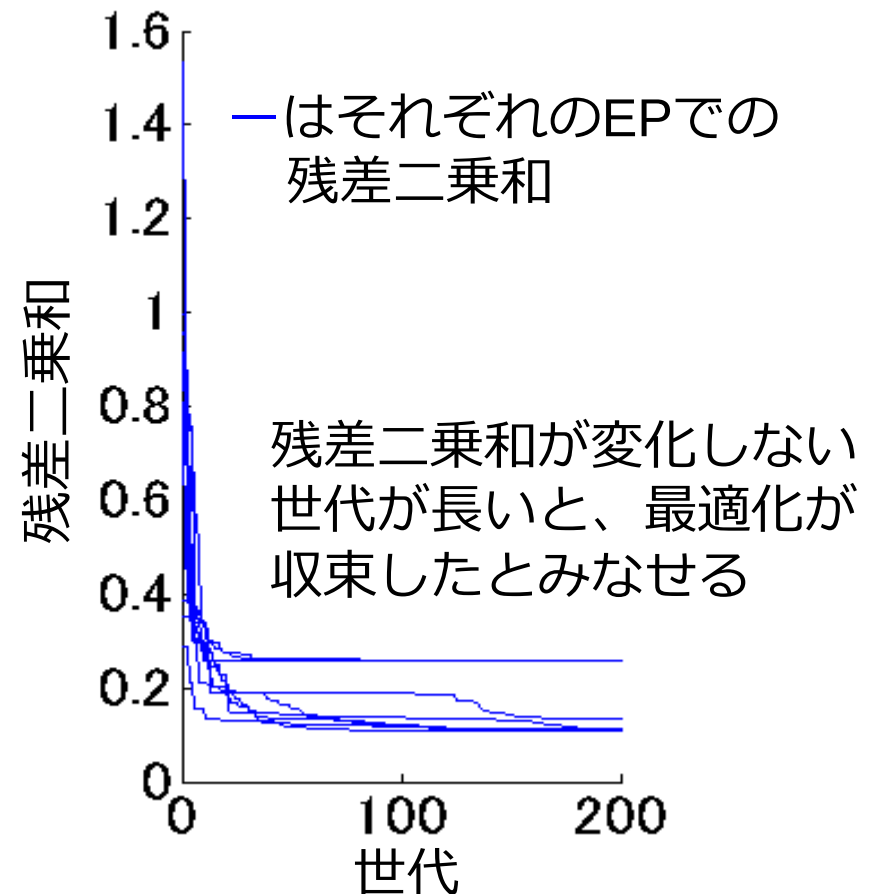


EPを実行したときに着目すること

フィッティングは妥当か？



最適化の世代は収束に十分か？



注意点

今回は、どのデータにどのモデルを使うかは予めわかっていた。

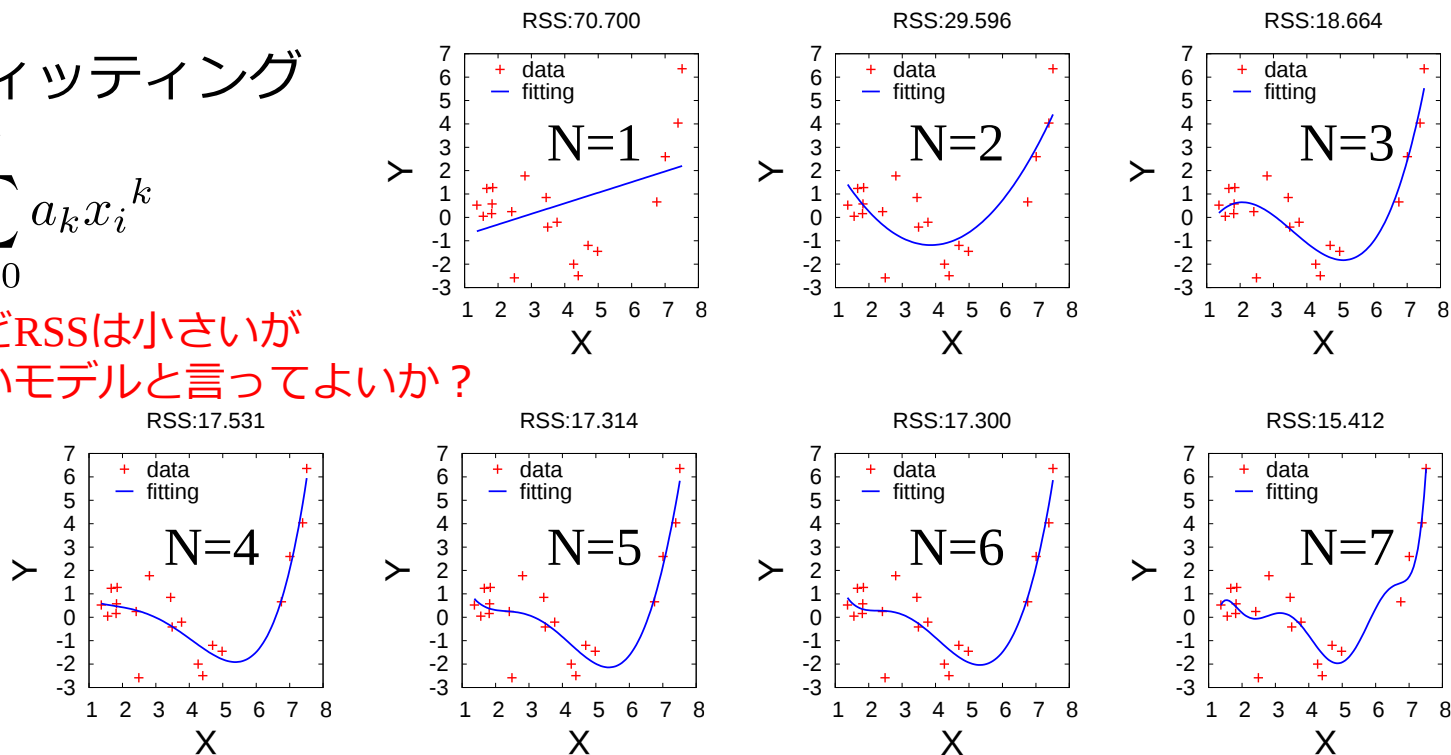
しかし、現実的にそのようなことは、文献などからその系について信頼性の高い情報がある場合に限られる。

データからモデルを決めるには、別の手法が必要
(交差検証、情報量規準に基づく選択など)。

例：多項式フィッティング

$$\hat{y}_i = \sum_{k=0}^N a_k x_i^k$$

次数Nが大きいほどRSSは小さいが
Nが大きいのが良いモデルと言ってよいか？



パラメータを推定して、（やっと）モデルができた。
その次は？

⇒次回の講義でモデルを使った
様々なシミュレーションをします。