

# 生物統計学

— データに基づく「よりよい推論」のために —

三中信宏

MINAKA Nobuhiro

独立行政法人 農業環境技術研究所 生態系計測研究領域 上席研究員  
東京大学大学院 農学生命科学研究科 生物・環境工学専攻 教授 [生態系計測学]  
東京農業大学大学院 農学研究科 客員教授 [応用昆虫学]

[minaka@affrc.go.jp](mailto:minaka@affrc.go.jp) (メール)

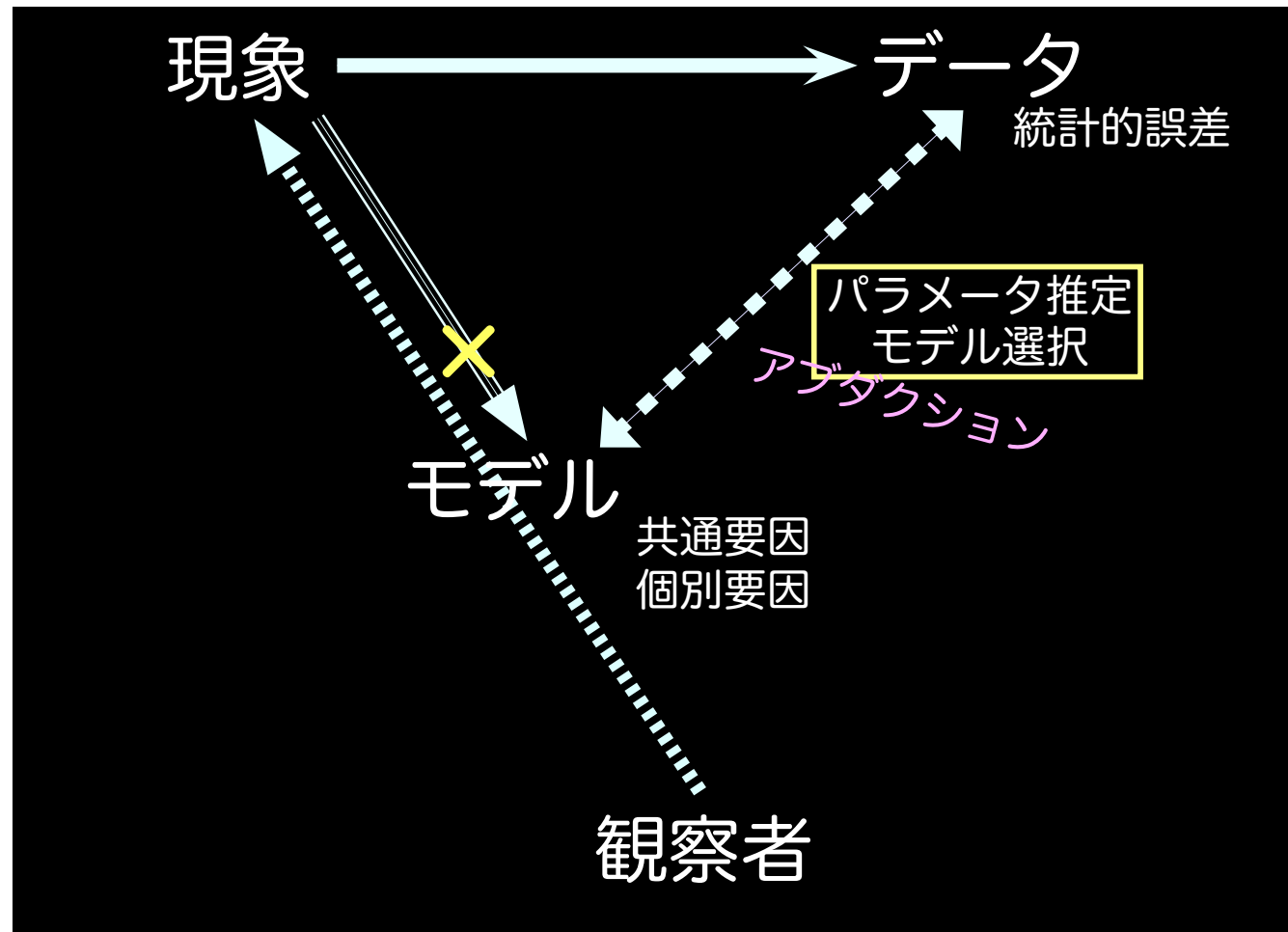
<http://twitter.com/leeswijzer> (ツイッター)

<http://cse.niaes.affrc.go.jp/minaka/> (ウェブサイト)

<http://d.hatena.ne.jp/leeswijzer/> (書評ブログ)

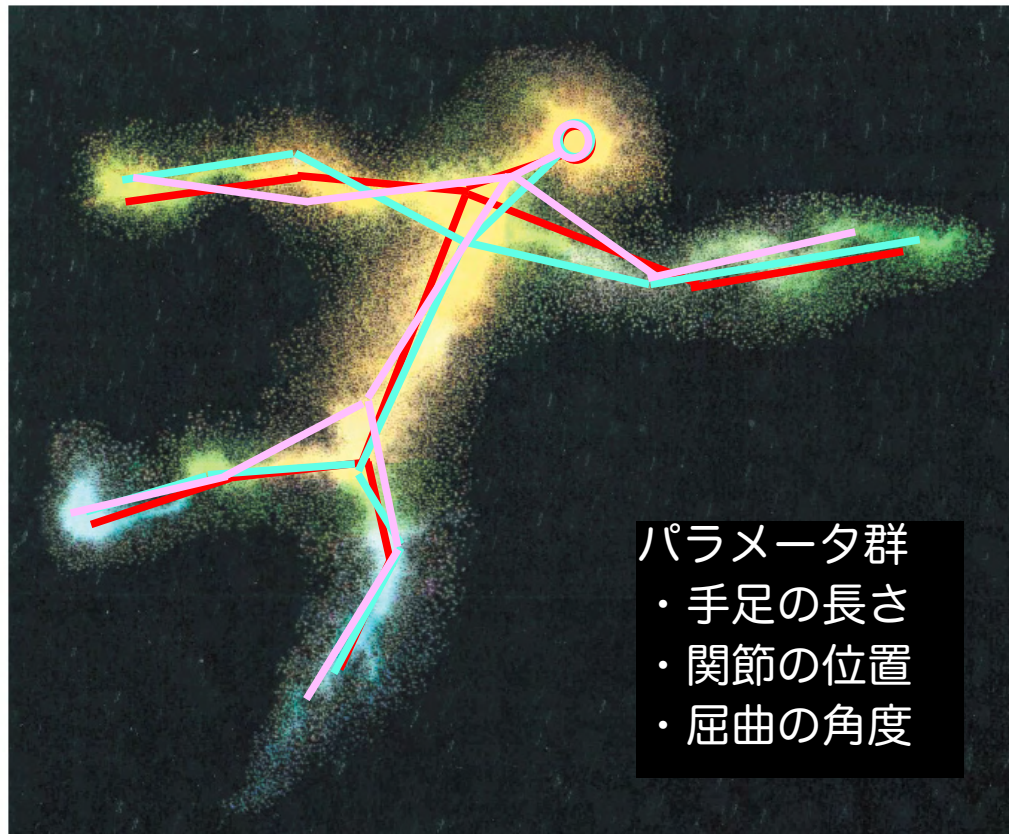
# 線形統計モデル (LM) への入り口

モデルとデータの接点に触る



# 線形統計モデル (LM) への入り口

## モデルの構造とパラメータ



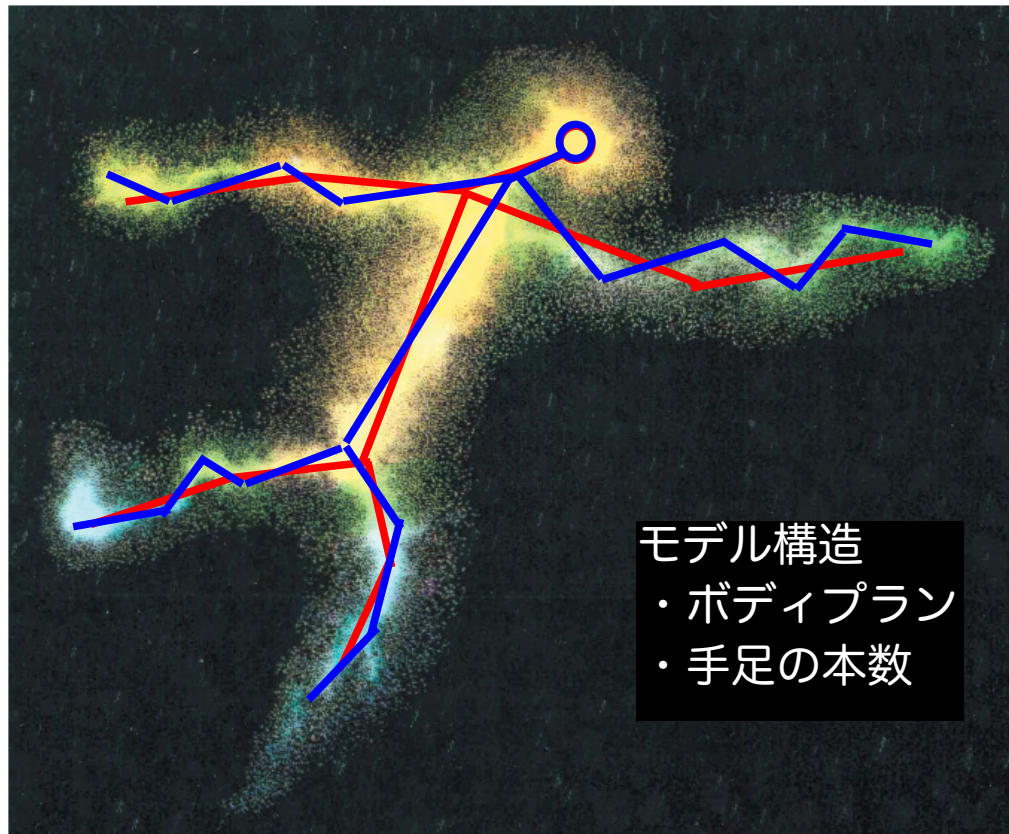
### パラメータ推定問題

モデルの「構造」を固定したとしても、パラメータ群の値は可変である。どのような基準で最適なパラメータ値を決定するのか。

計算すればなんとかなる

# 線形統計モデル (LM) への入り口

## モデルの構造とパラメータ



### モデル選択問題

モデルの基本となる「構造」そのものをどのような基準で決定すればいいのか。

やや難しいかも・・・

# 線形モデルを鍛えあげる

## 線形モデル (LM)

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (\epsilon_{ij} \sim N(0, \sigma^2))$$

データ 総平均  
処理効果 誤差項 誤差は正規分布をする

通常の「線形モデル」の仮定

- 処理効果はある実数（「パラメータ値」）である。
  - 誤差項は独立正規分布にしたがう。  
（「正規性」と「等分散性」）
- これらの仮定を緩めよう（一般化）**

# 線形モデルを鍛えあげる

## 一般化線形モデル (GLM)

データ 総平均

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

処理効果 誤差項

~~$(\epsilon_{ij} \sim N(0, \sigma^2))$~~

誤差は正規分布をする

何が一般化されたのか？

- ・ 誤差項が正規分布でなくてもよい。
- ・ データのある関数変換（「リンク関数」）が線形モデルで記述できればよい



# 線形モデルを鍛えあげる

## 一般化加法モデル (GAM)

データ

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

総平均

処理効果

誤差項

~~$(\epsilon_{ij} \sim N(0, \sigma^2))$~~

誤差は正規分布をする

さらなる一般化とは？

- ・ 予測関数は単一の式で表示される関数である必要すらなく，スプライン関数によって区分的に構築された任意に関数（ただし加法性はある）でもかまわない。

# 線形モデルを鍛えあげる

## 混合効果モデル (mixed-effect model)

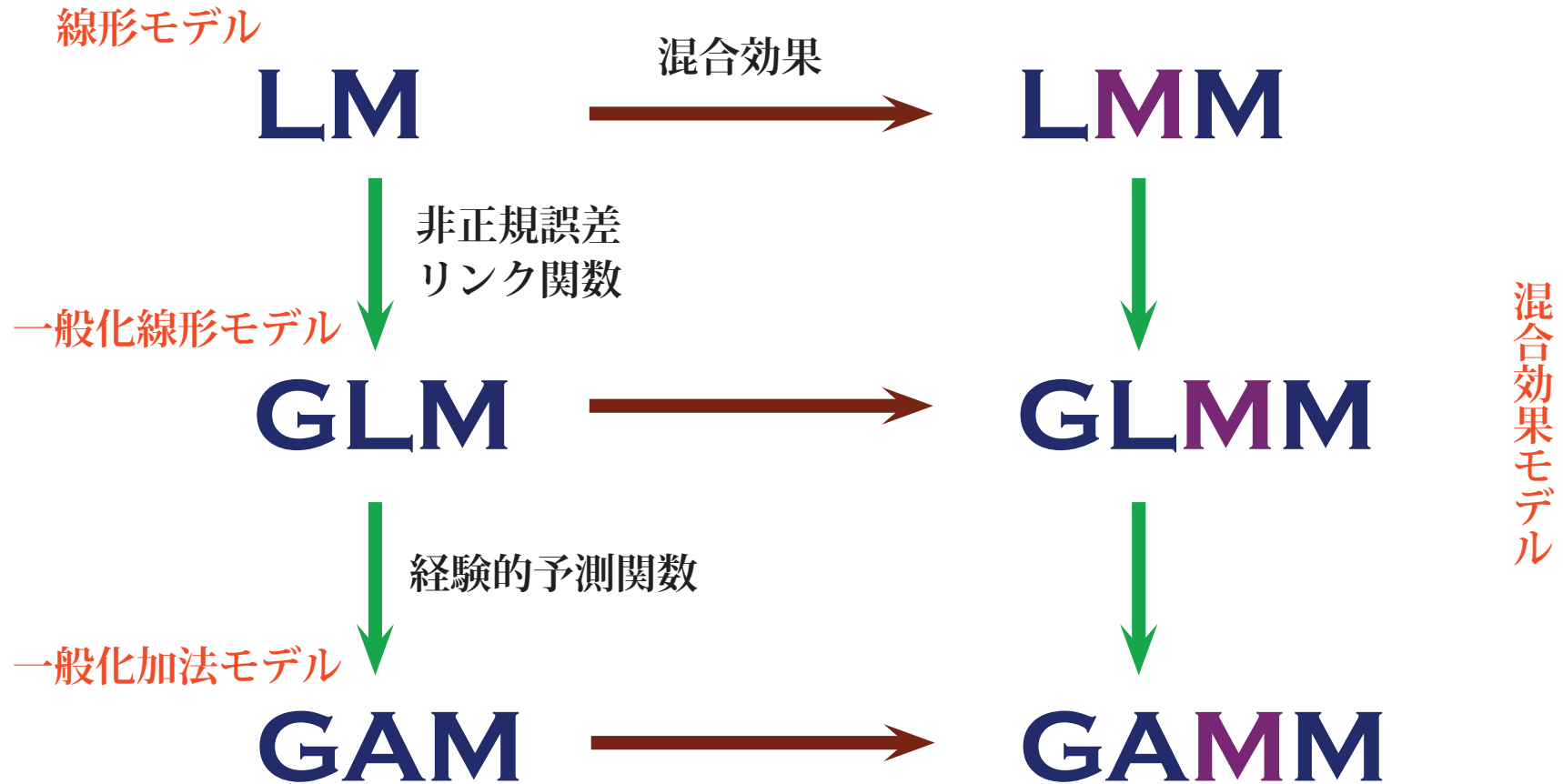
$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (\epsilon_{ij} \sim N(0, \sigma^2))$$

処理効果

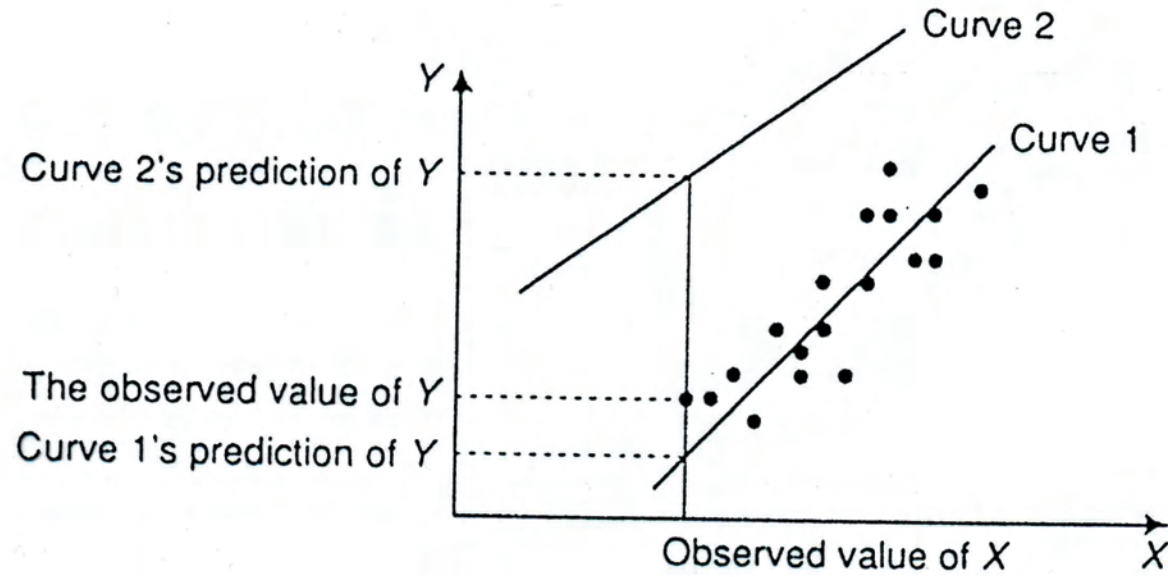
要因の「処理効果」は、ある実数の「パラメータ値」という定数として「固定効果」(fixed effect)をもつと仮定してきた。しかし、その処理効果が別の確率分布にしたがう“変量”とみなして、「ランダム効果」をもつと仮定すれば、より細かい変動や個体差をもモデリングできるだろう。



# 線形モデルを鍛えあげる

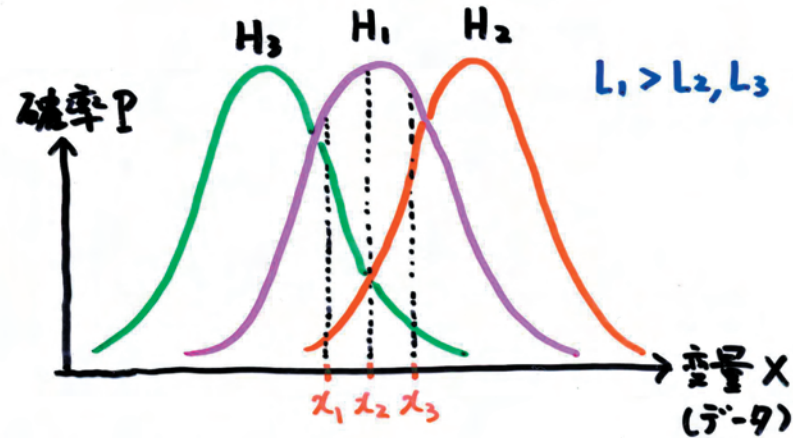


# Curve-fitting problem (2)



# 尤度 (Likelihood)

データが仮説に与える支持の強さ.



仮説  $H_1$   
 $H_2$   
 $H_3$  } 確率分布の平均値が異なる。  
パラメータ

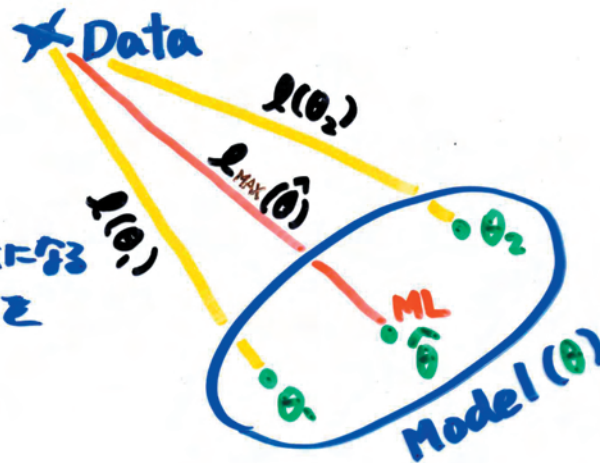
$$\text{尤度 } L = P(x_1) \cdot P(x_2) \cdot P(x_3)$$

(確率積)

↓  
最尤推定: 尤度  $L$  が最大となる仮説を  
受容する.

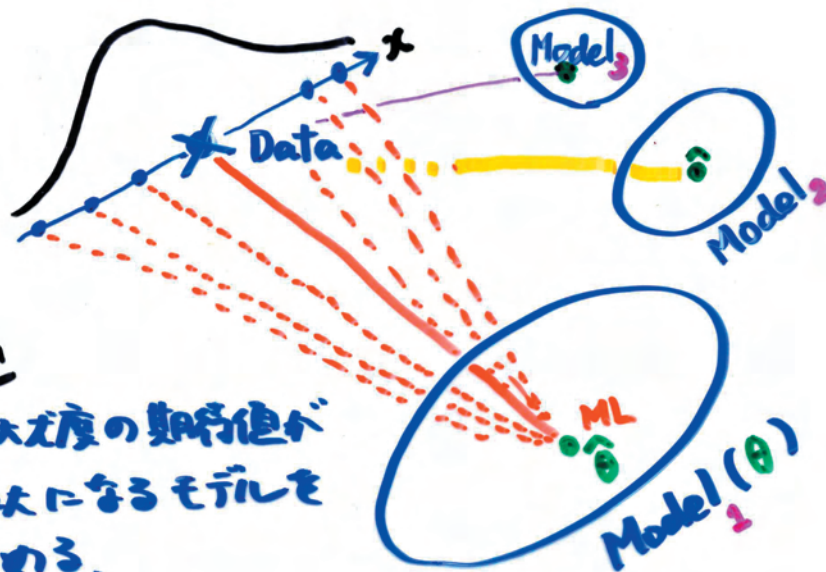
### 最尤法

尤度が最大になる  
パラメータを  
求める。

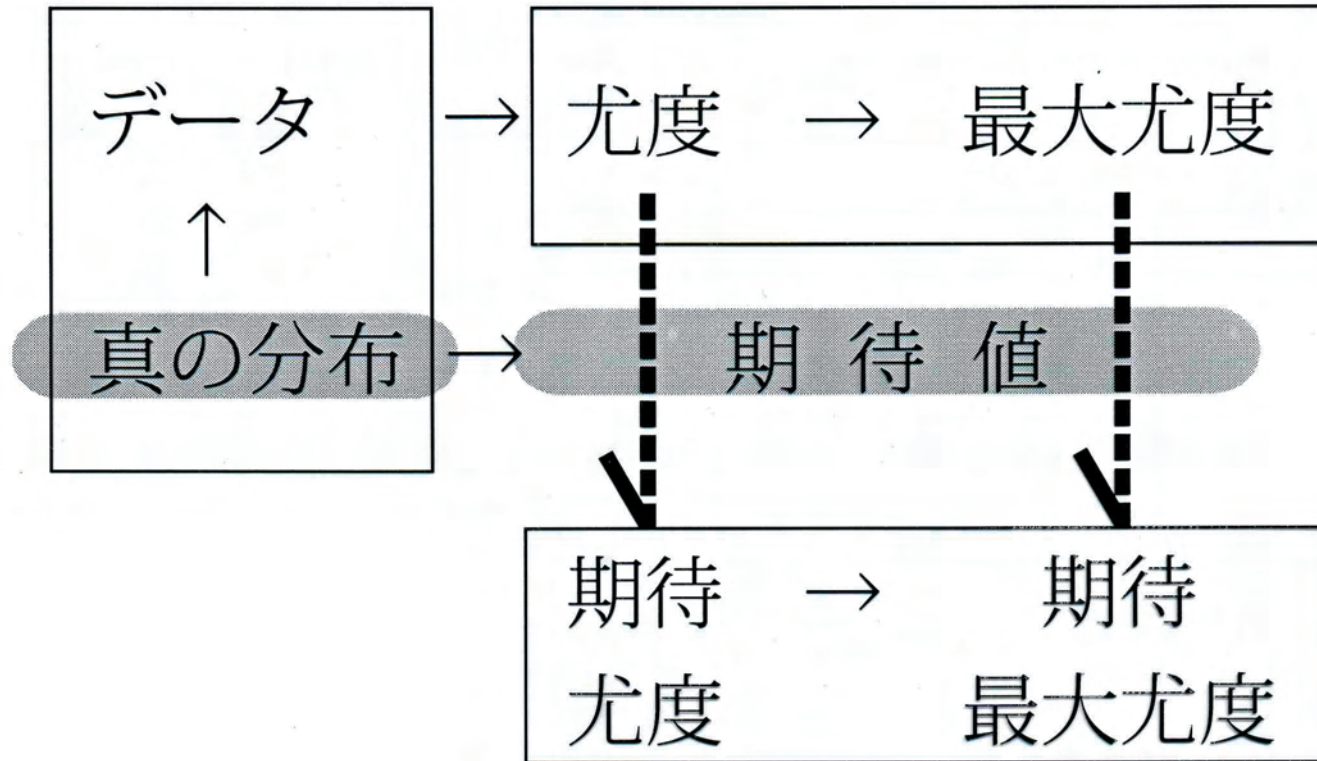


### AIC

最大尤度の期待値が  
最大になるモデルを  
求める。

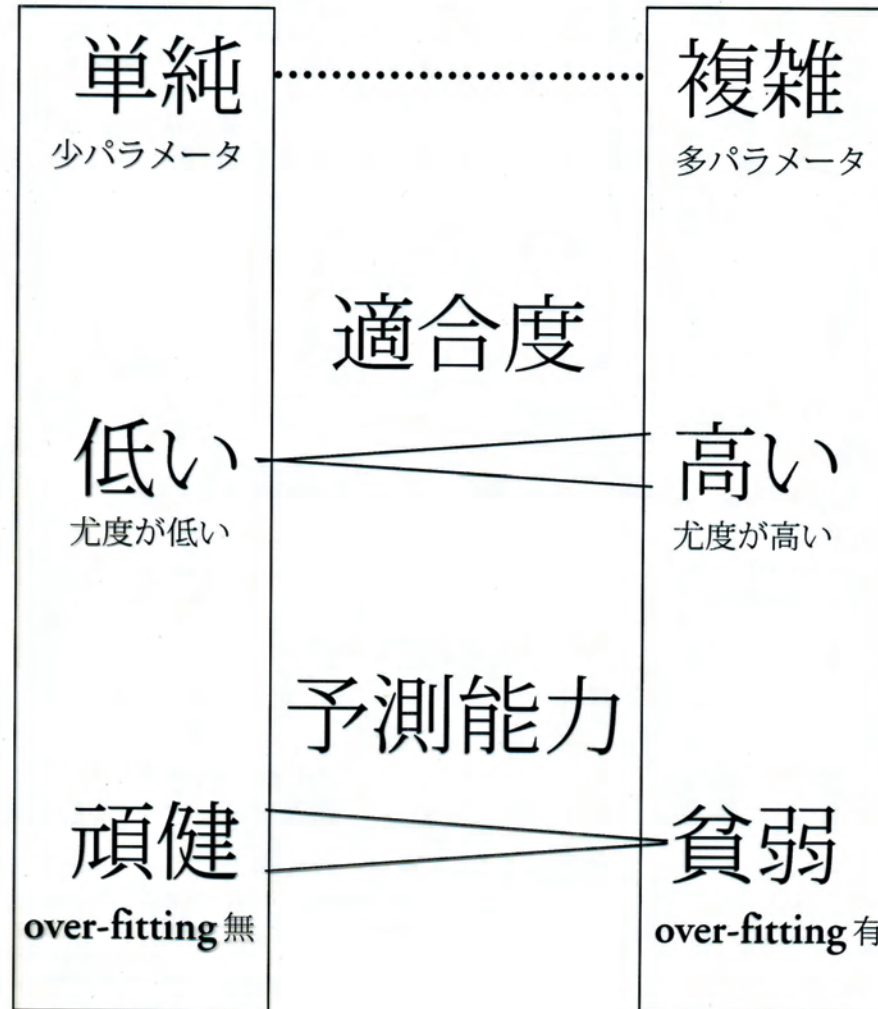


# 最尤法



# AIC

# トレード・オフ



モデルの予測確度

=

期待尤度 - パラメータ数

(適合度)

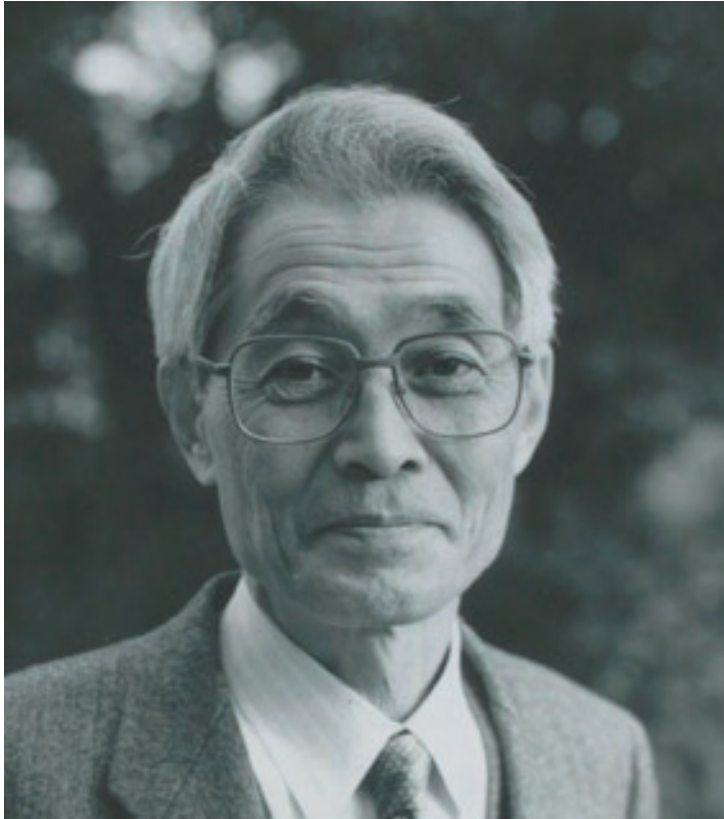
(複雑度)



**AIC**

(赤池情報量基準)





良いモデルは  
「予測確度」  
(predictive accuracy)  
が大きい

[赤池弘次]