

複雑理工学専攻とデータ駆動科学

岡田真人

東京大学 大学院新領域創成科学研究科
複雑理工学専攻

自己紹介(理論物理)

- 大阪市立大学理学部物理学科 (1981 - 1985)
- 大阪大学大学院理学研究科物理専攻 (1985 - 1987)
 - 希土類元素の光励起スペクトルの理論
- 三菱電機 (1987 - 1989)
 - 化合物半導体(半導体レーザー)の結晶成長
- 大阪大学大学院基礎工学研究科生物工学 (1989 - 1996)
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
- 理化学研究所 脳科学総合研究センター (2001 - 04/06)
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻 (2004/07 -)

内容

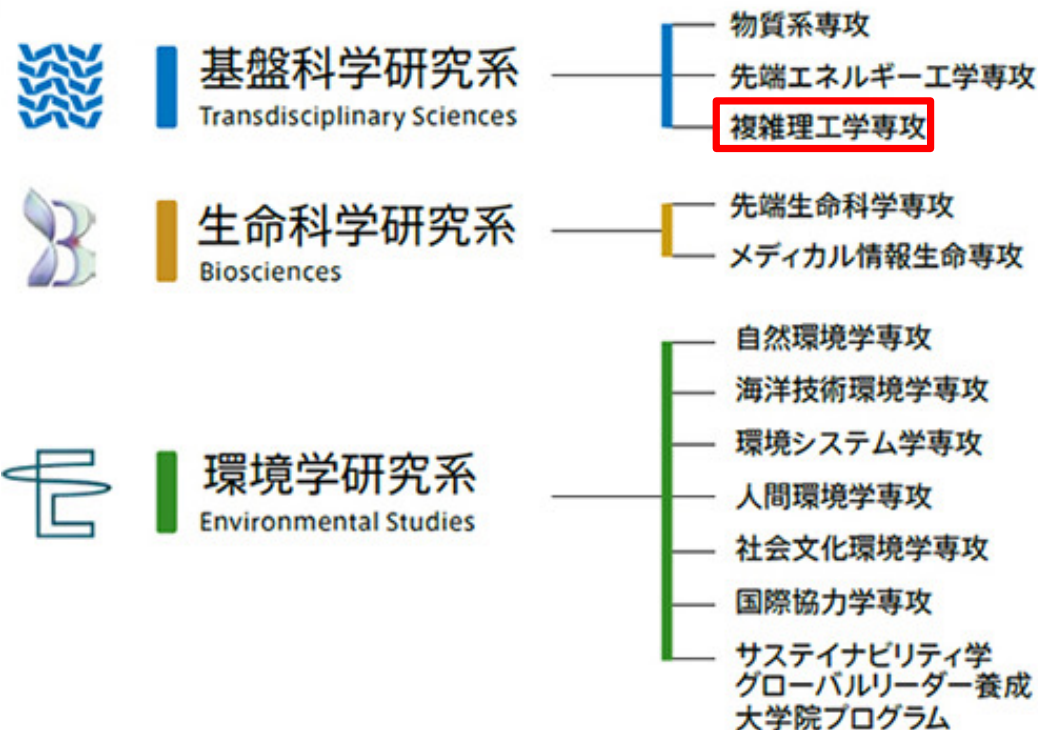
- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

新領域創成科学研究科の組織

複雑理工学専攻
Complexity

新しい学問領域の創成を目指す**新領域創成科学研究科**は「基盤科学」「生命科学」「環境学」の3つの研究系で構成されており、既存の個別学問分野から派生する**未開拓の領域を研究・教育の対象**とし、人類が解決を迫られている課題に取り組んでいます。

大学院新領域創成科学研究科



研究科附属施設

生涯スポーツ健康科学研究センター／オーミクス情報センター／バイオイメーjingセンター
ファンクショナルプロテオミクスセンター／革新複合材学術研究センター



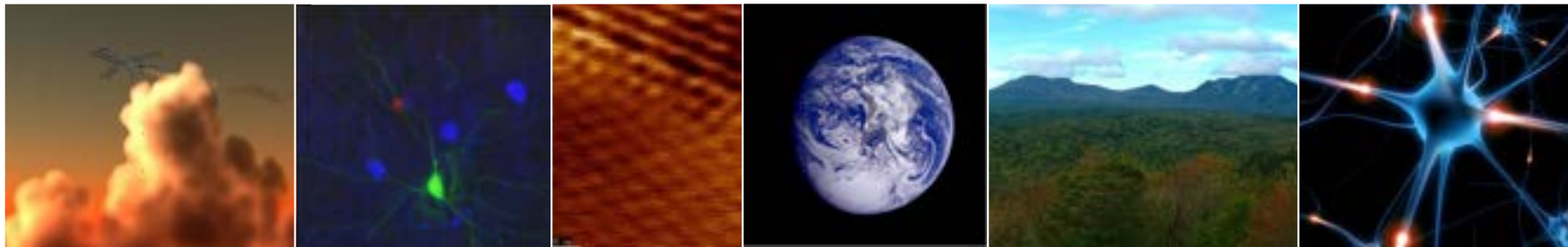
複雑理工学とは？

複雑理工学専攻
Complexity



理学・工学にまたがる
様々な学術分野で普遍的
に存在する**複雑性の解明**

実世界には、多数の非線形要素が強く相互作用する**複雑系**が普遍的に存在し、それらは単純な現象の線形的な重ね合わせでは表現・理解することはできません。



このような**複雑系**を包括的に理解し応用することこそが、新たな科学・技術を創成する原動力となります。

複雑理工学専攻の教育・研究の目的

- ◇理学と工学を融合した新しいアプローチによる「**複雑性**」の解明
- ◇ナノから宇宙にわたる**マルチスケール複雑系**の学融合の推進
- ◇**複雑系の理論・要素技術**の構築
- ◇**複雑系を取り扱う新しいパラダイム**を創成できる研究者・技術者の養成

複雑理工学専攻の構成

複雑理工学専攻
Complexity

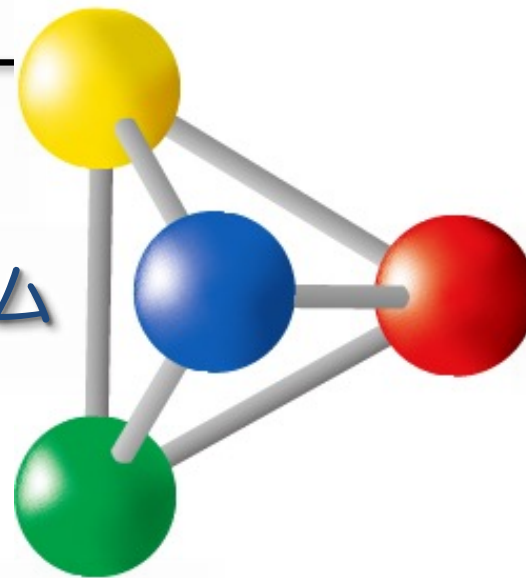
本専攻では「**脳・バイオ**」「**アストロバイオロジー**」「**極限物質**」の3つのモジュールを中心に、ナノから宇宙にわたるマルチスケール複雑系の学融合を推進し、データ駆動科学を用いて、新しい複雑系科学・技術の創成を目指しています。また、これらの分野に共通する数理・情報・データ駆動科学などの理論と技術に基づく「**複雑系プラットフォーム**」を構築することにより研究と人材育成を促進し、新たな展開を図っています。

脳・バイオモジュール

複雑系プラットフォーム

極限物質モジュール

アストロバイオロジーモジュール



- 機械学習などの**人工知能**を使い，各学問分野の問題を解いていくというアプローチ
- 実験/計測/計算データの**背後にある潜在的構造**の抽出に関して，データが対象とする学問に**依存しない普遍的な学問体系**
- 同じアルゴリズムがスケールや対象を超えて，有用であることが多いという経験的事実を背景として，その理由を問い，背後にある普遍性から，**データ解析自体を学問的对象**とする枠組み

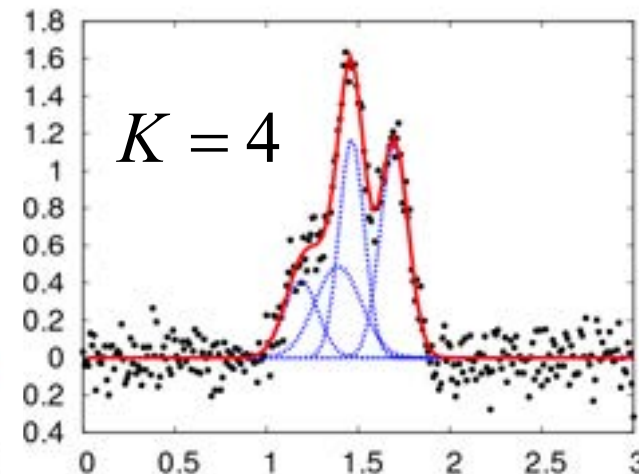
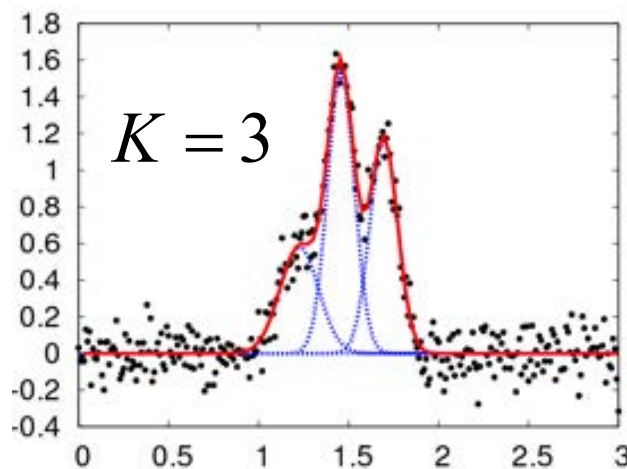
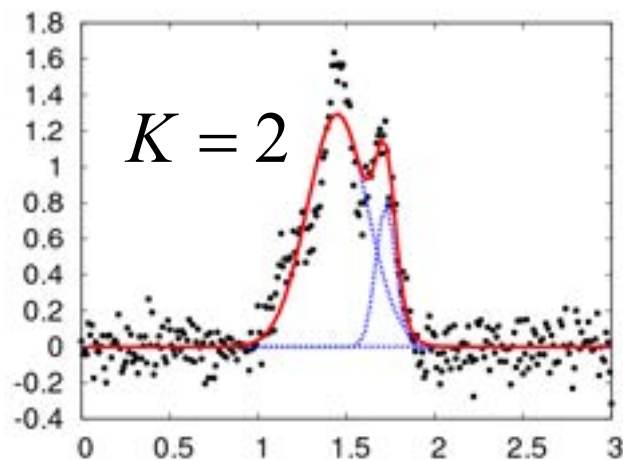
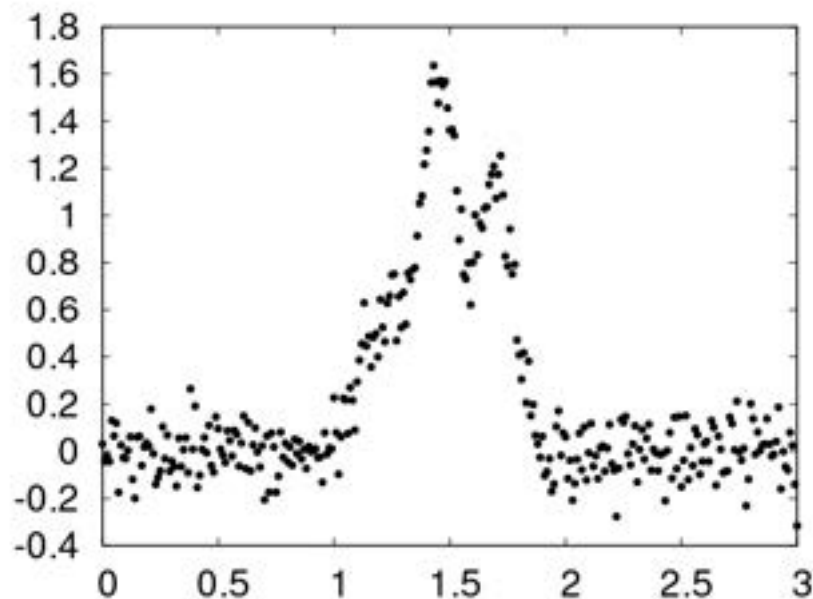
複雑理工学専攻での データ駆動科学の水平展開

- データ駆動科学の核融合グループとアストロバイオロジーグループへの水平展開を行なっている.
- まずその皮切りとして, 分光学を取り上げた.
- 分光学はスケールや対象を超えて広く使われている。
- マルチスケールを取り扱う複雑理工でのデータ駆動科学の水平展開にうってつけの題材である.

内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

分光学とは: スペクトル分解



Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

共同研究者

- | | |
|--------|-------------------------------|
| 永田 賢二 | 国立研究開発法人物質・材料研究機構 |
| 村岡 怜 | 東京大学 大学院新領域創成科学研究科 |
| 本武 陽一 | 一橋大学大学院 ソーシャル・データサイエ
ンス研究科 |
| 杉田 精司 | 東京大学 大学院理学系研究科 |
| 佐々木 岳彦 | 東京大学 大学院新領域創成科学研究科 |

分光学の諸問題: アンケート

- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。
- そのような方は、一度ベイズ計測をお試しください。

[情報計測] 計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用

[← トップに戻る](#)

▣ 領域活動についてはこちらをご覧ください。

戦略目標

「材料研究をはじめとする最先端研究における計測技術と高度情報処理の融合」

研究総括



雨宮 慶幸(公財) 高輝度光科学研究センター 理事長)

副研究総括



北川 源四郎(東京大学 数理・情報教育研究センター 特任教授)

[情報計測] 平成29年度採択課題

[← トップに戻る](#)

岡田 真人

ベイズ推論とスパースモデリングによる計測と情報の融合

研究代表者

岡田 真人



東京大学
大学院新領域創成科学研究科
教授

主たる共同研究者

桑谷 立	海洋研究開発機構 海域地震火山部門 グループリーダー代理
長尾 大道	東京大学 地震研究所 准教授
成田 憲保	東京大学 大学院総合文化研究科 教授
日野 英逸	情報・システム研究機構 統計数理研究所 教授

内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

スペクトル分解の定式化

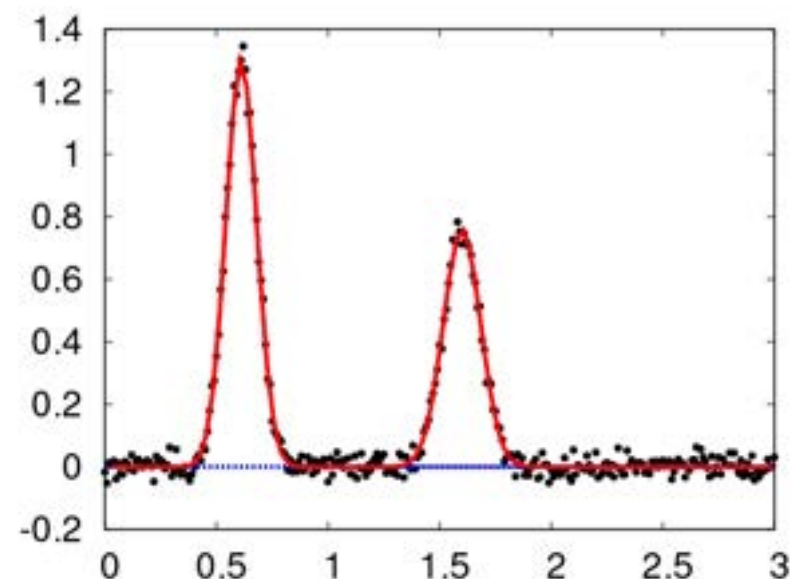
ガウス関数(基底関数)の足し合わせにより, スペクトルデータを近似

観測データ: $D = \{x_i, y_i\}_{i=1}^n$

x_i : 入力 y_i : 出力

$$f(x; \theta) = \sum_{k=1}^K a_k \exp\left(-\frac{b_k (x - \mu_k)^2}{2}\right)$$

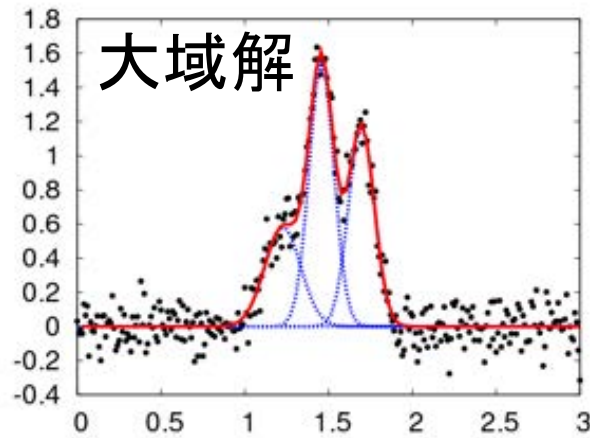
$$\theta = \{a_k, b_k, \mu_k\} \quad k = 1, \dots, K$$



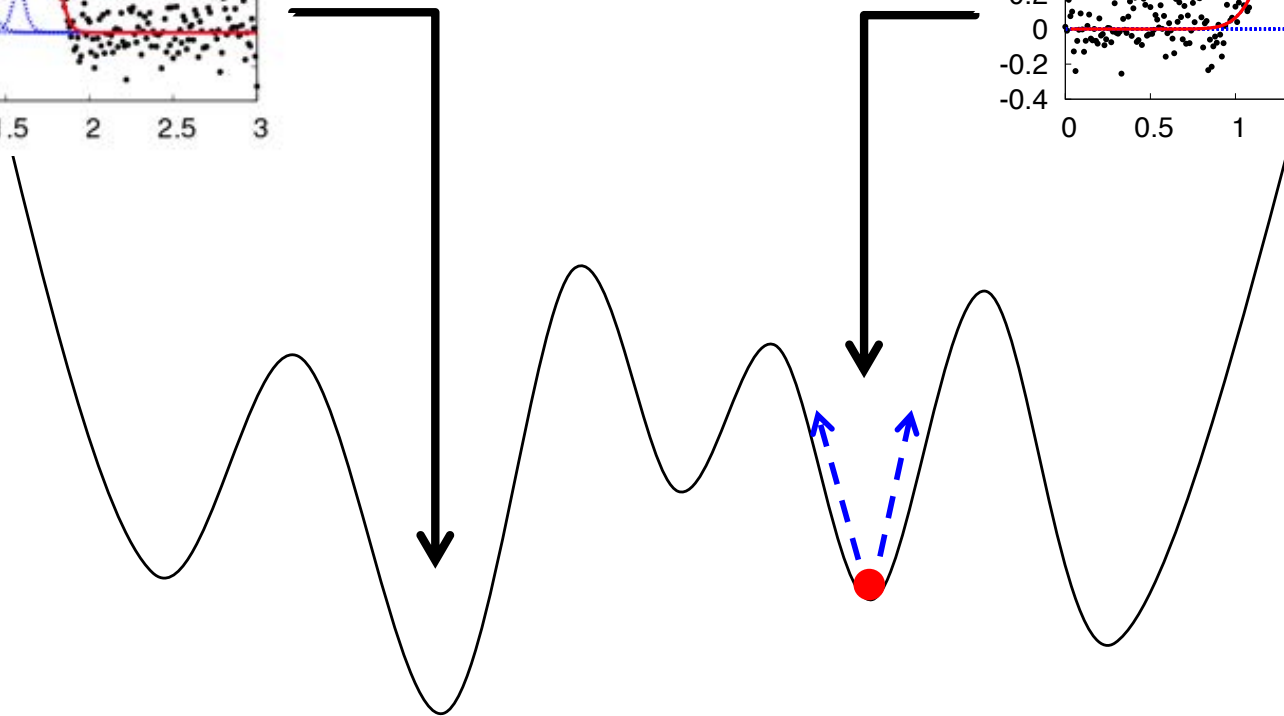
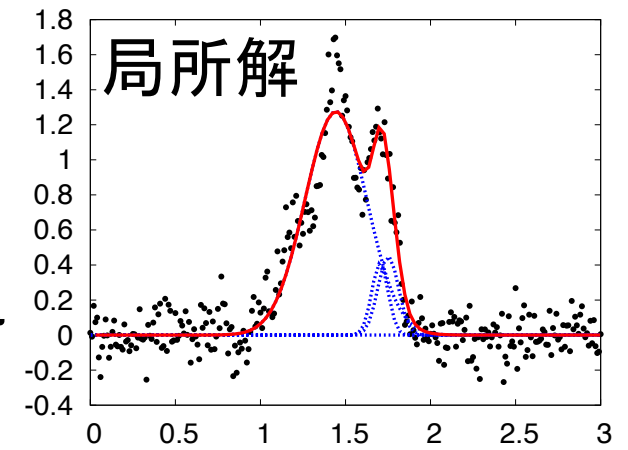
二乗誤差を最小にするようにパラメータをフィット(最小二乗法)

$$E(\theta) = \frac{1}{n} \sum_{i=1}^n \left(y_i - f(x_i; \theta) \right)^2$$

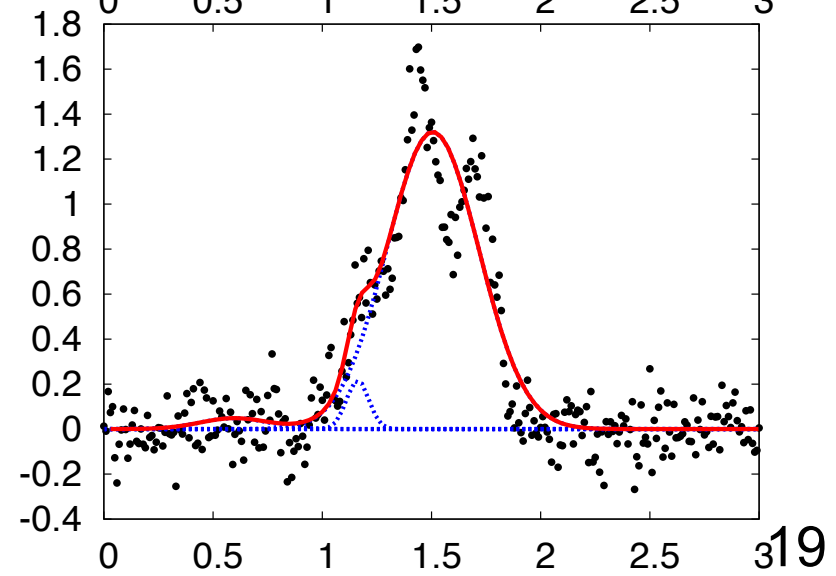
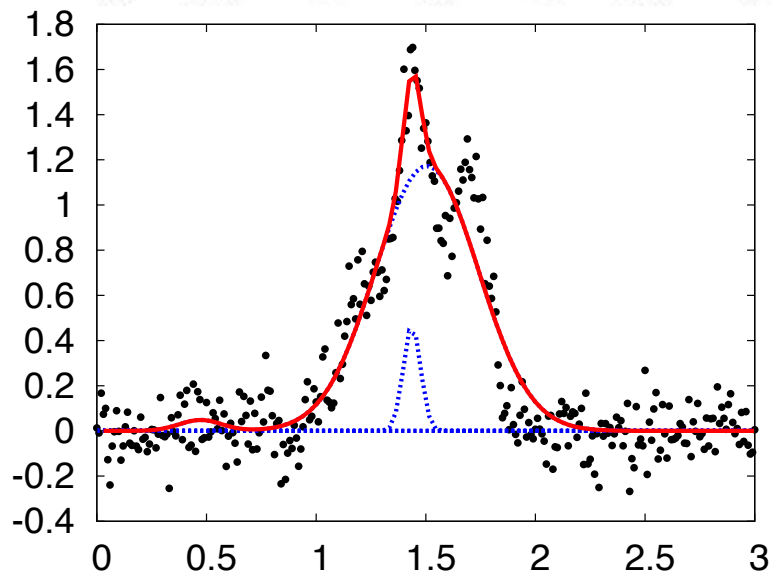
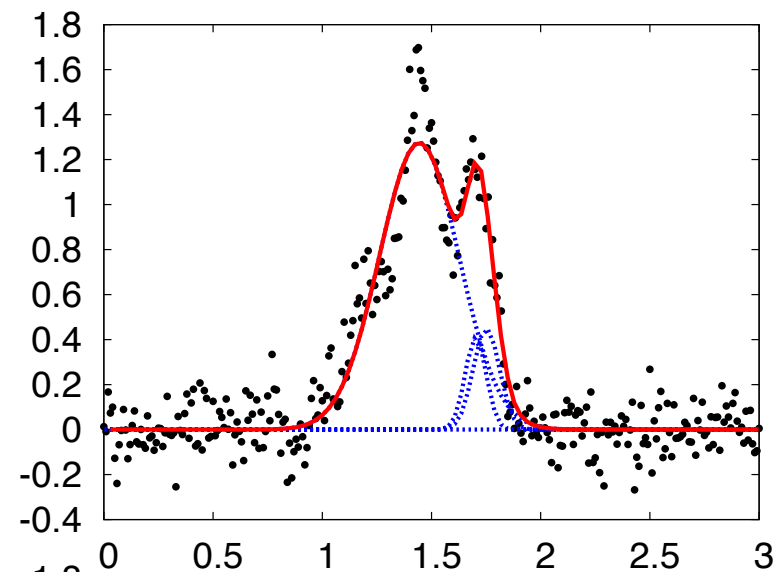
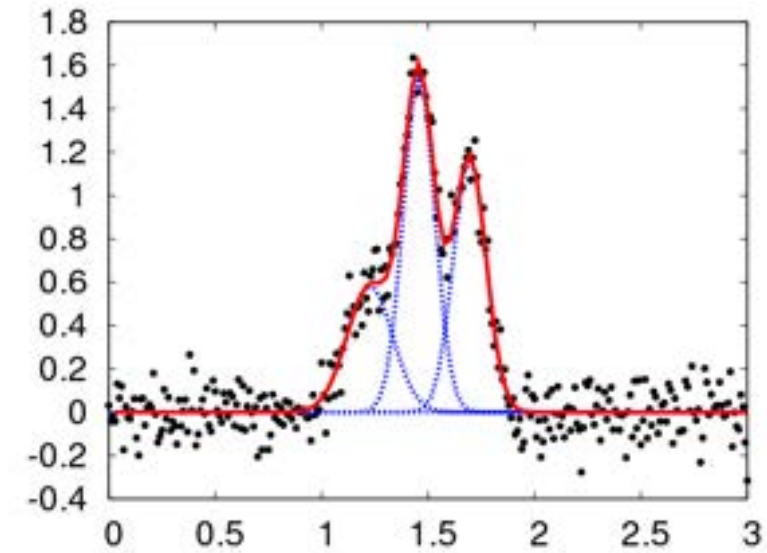
誤差関数は局所解を持つ



＜通常の最適化法＞
e.g., 最急降下法



ローカルミニマム



分光学の諸問題:アンケート

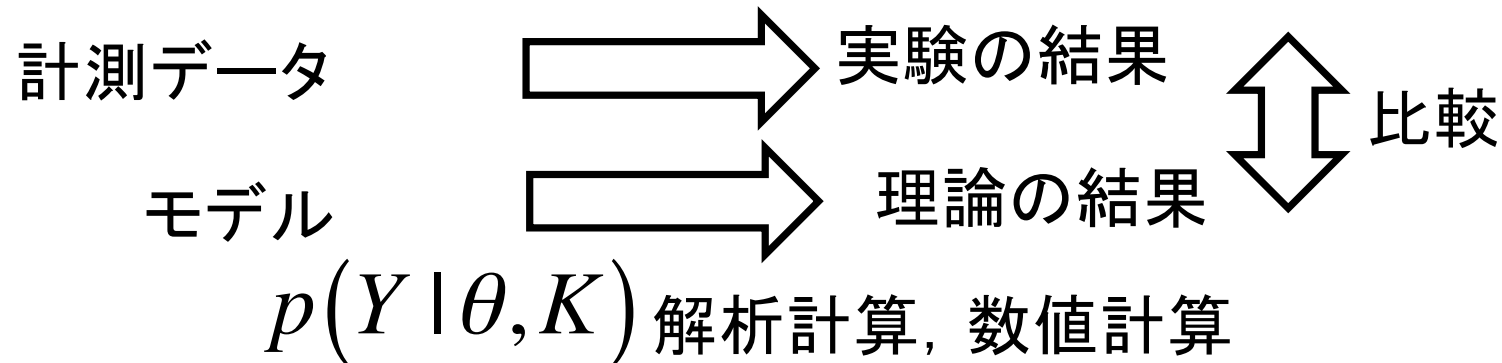
- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。
- そのような方は、一度ベイズ計測をお試しください。

内容

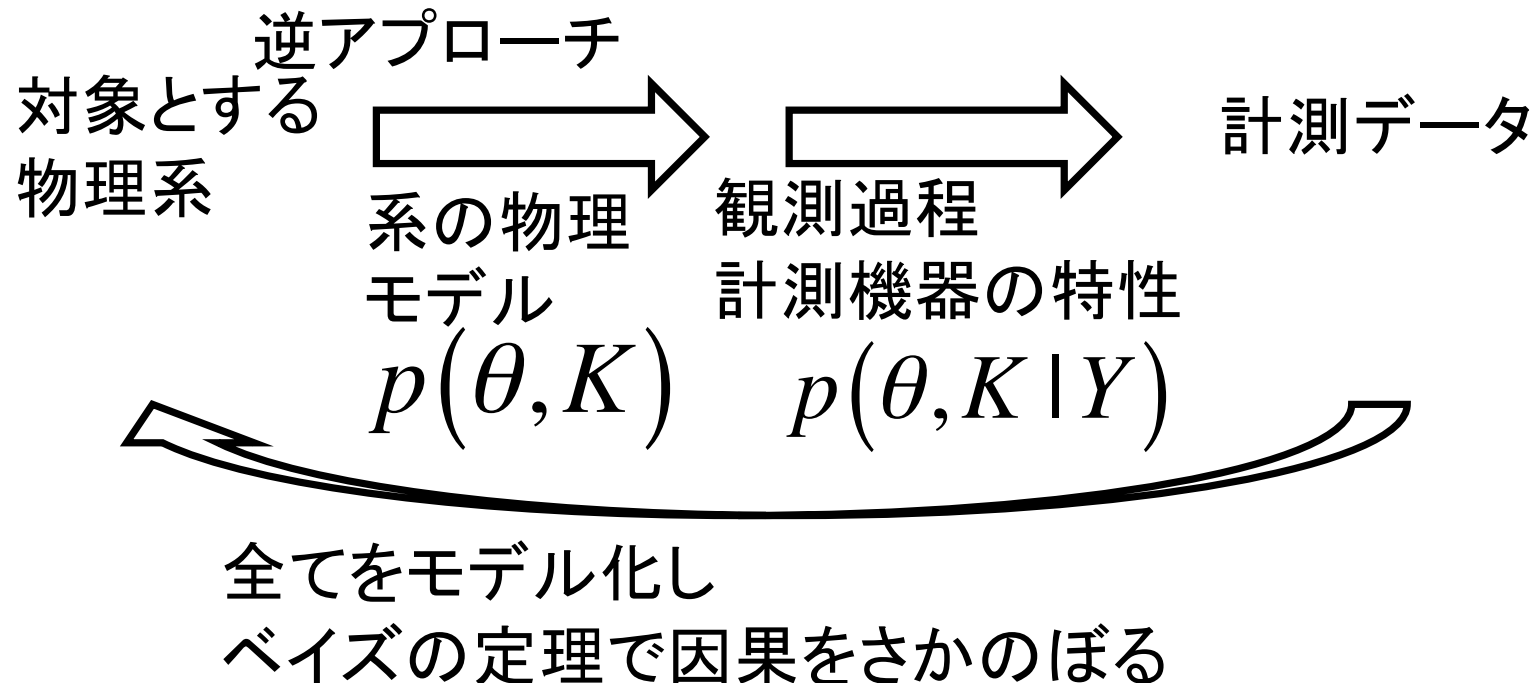
- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - **ベイズ計測**
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

ベイズ計測

順アプローチ



逆アプローチ



確率的定式化

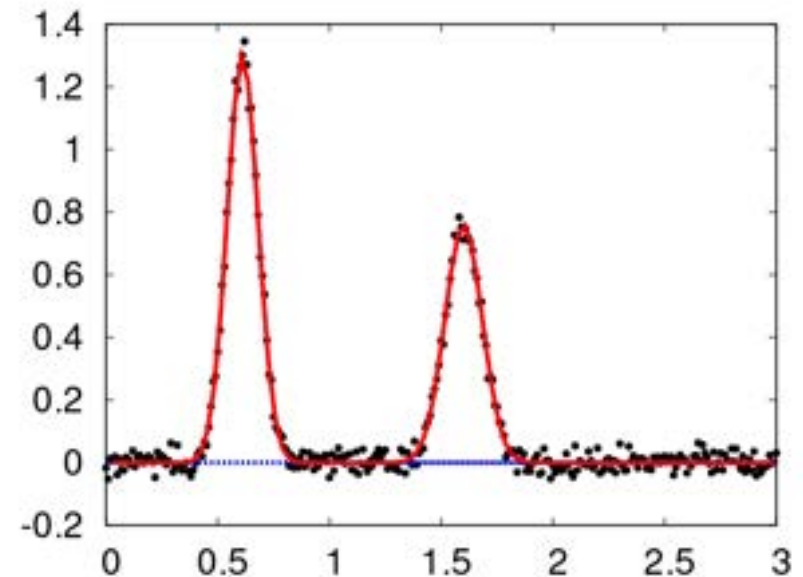
出力は, 入力からの応答とノイズの足し合わせにより生成

⇒出力は, 確率変数である.

$$y_i = f(x_i; \theta) + \varepsilon$$

ノイズが正規分布であるとするとき,

$$p(y_i | \theta) \propto \exp\left(-\frac{1}{2}\left(y_i - f(x_i; \theta)\right)^2\right)$$



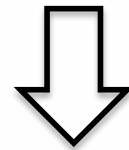
それぞれの出力 y_i が, 独立であるとするとき,

$$p(Y | \theta) = \prod_{i=1}^n p(y_i | \theta) \propto \exp(-nE(\theta)) \quad Y = \{y_1, \dots, y_n\}$$
$$E(\theta) = \frac{1}{2n} \sum_{i=1}^n \left(y_i - f(x_i; \theta)\right)^2$$

ボルツマン分布

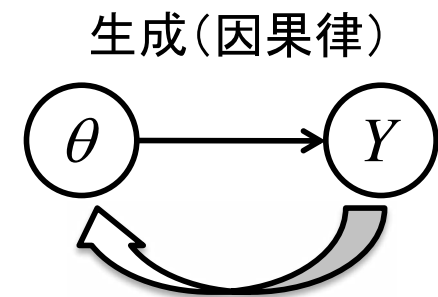
ベイズ推論：因果律を組み込んでデータ解析

$$p(Y, \theta) = p(Y | \theta)p(\theta) = \underline{p(\theta | Y)p(Y)}$$



<ベイズの定理>

$$p(\theta | Y) = \frac{p(Y | \theta)p(\theta)}{p(Y)} \propto \exp(-nE(\theta))p(\theta)$$



$p(\theta | Y)$: 事後確率。データが与えられたもとでの, パラメータの確率.

$p(\theta)$: 事前確率。あらかじめ設定しておく必要がある。
これまで蓄積されてきた科学的知見

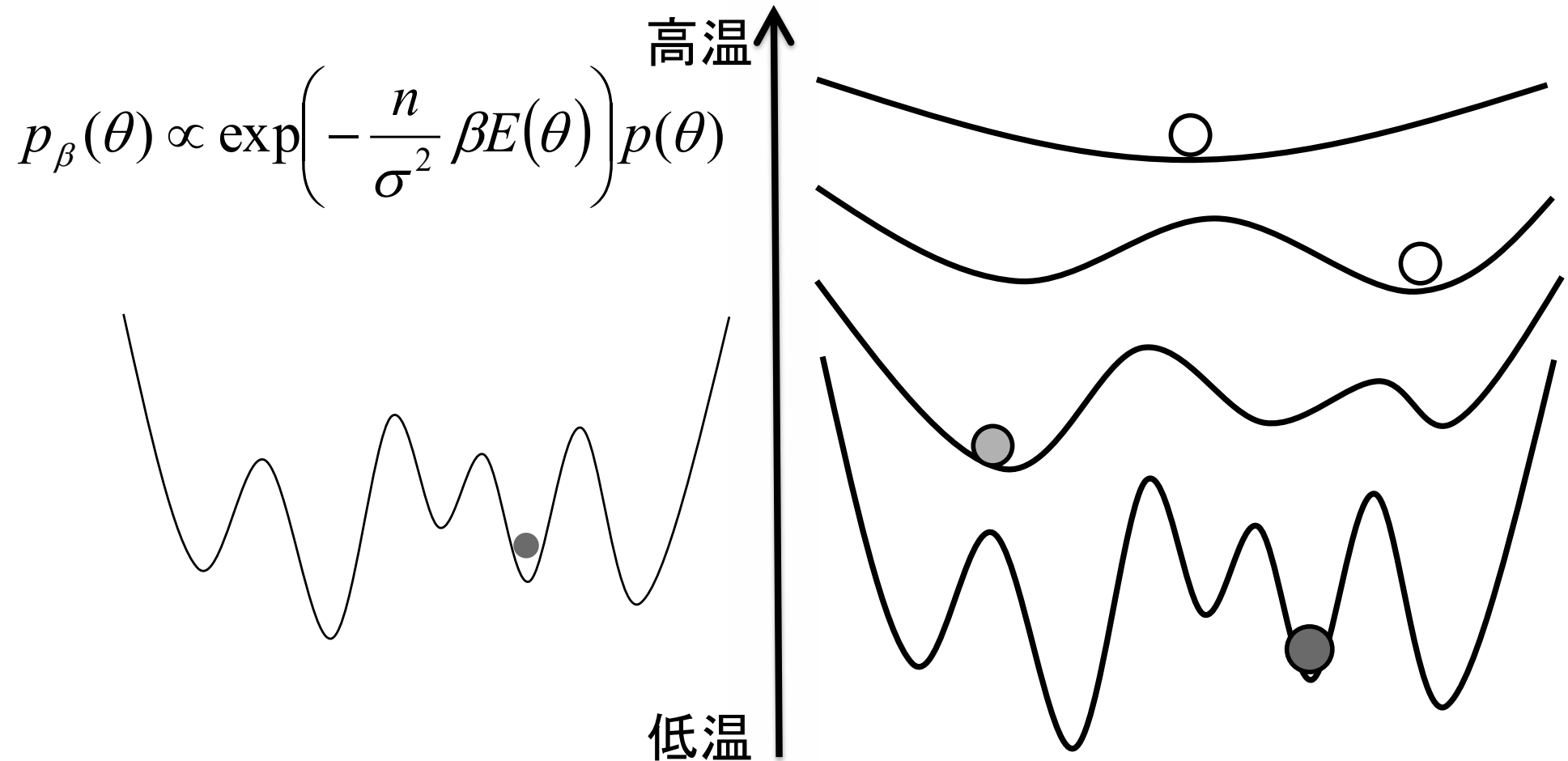
内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

レプリカ交換モンテカルロ法 ランダムスピン系の知見から

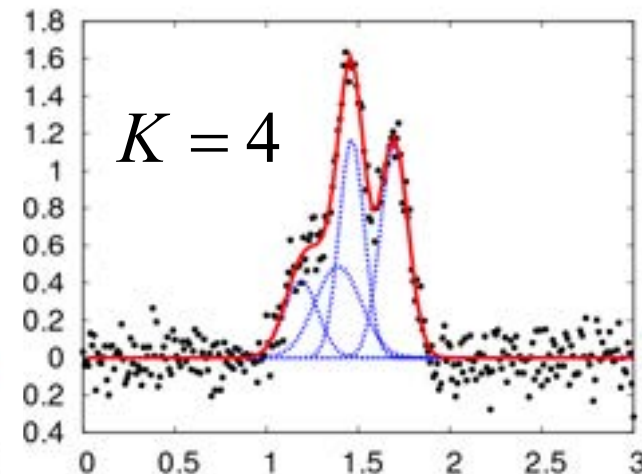
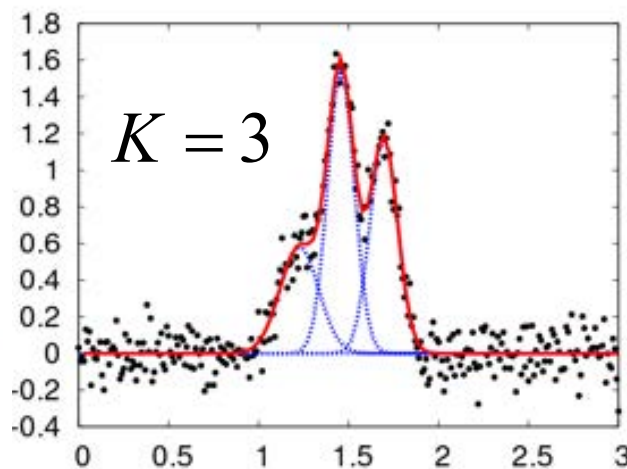
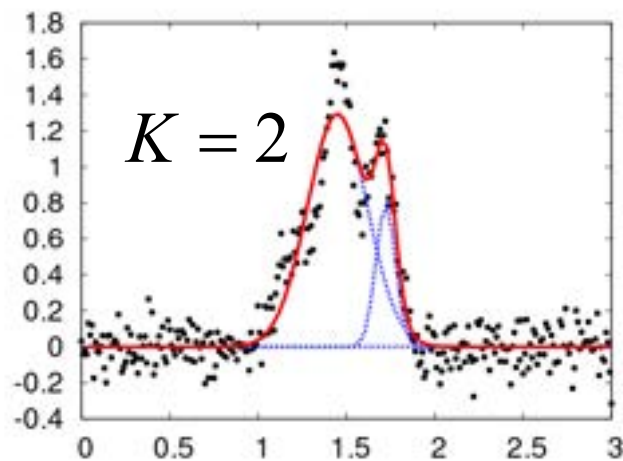
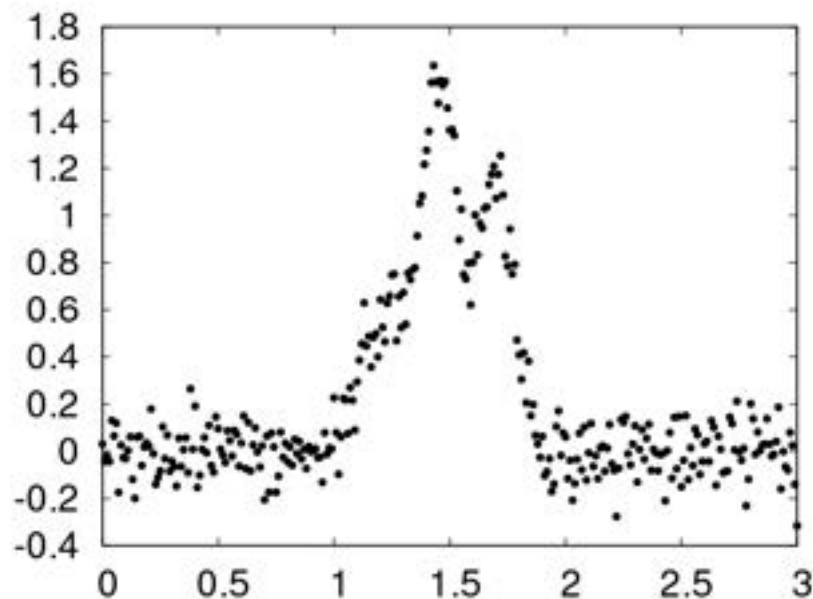
メトロポリス法

レプリカ交換モンテカルロ法



K. Hukushima, K. Nemoto, *J. Phys. Soc. Jpn.* **65** (1996).

例題1:スペクトル分解



Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

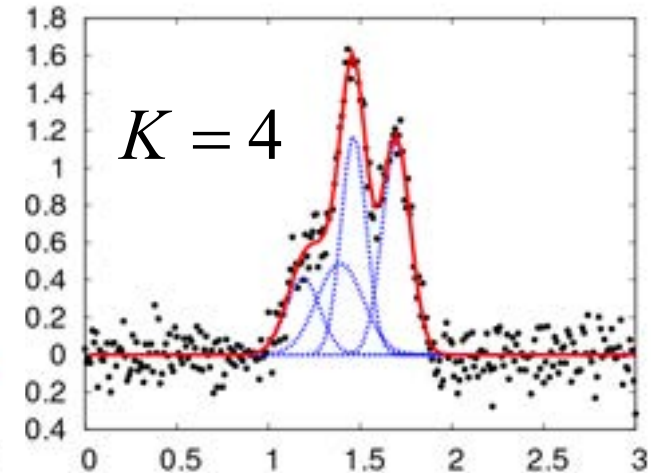
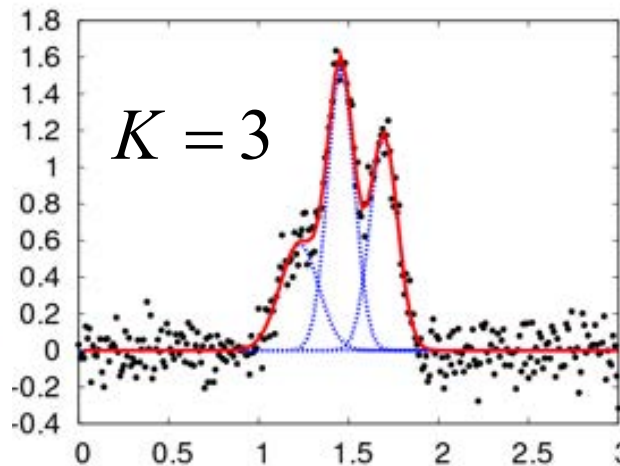
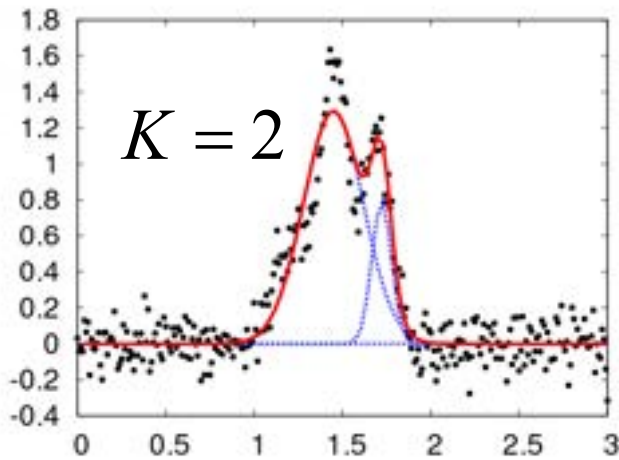
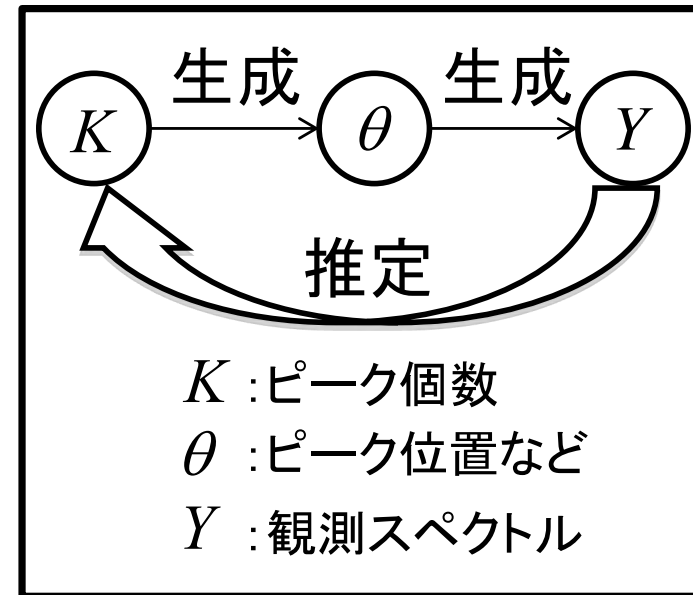
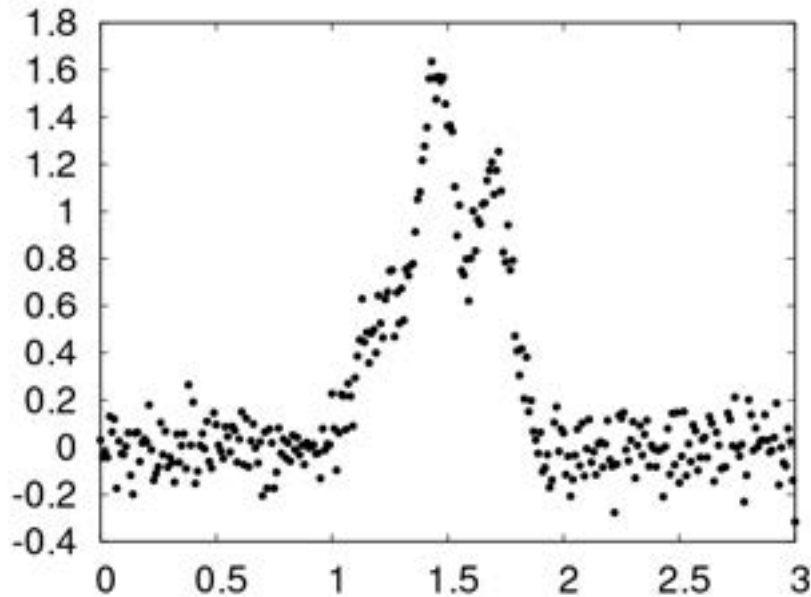
分光学の諸問題: アンケート

- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。
- そのような方は、一度ベイズ計測をお試しください。

内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

より深い構造をさぐる: モデル選択



Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

モデル選択

1. 欲しいのは $p(K|Y)$

2. θ がないぞ

3. $p(K, \theta, Y)$ の存在を仮定

$$p(K, \theta, Y) = p(Y | \theta, K) p(K)$$

$$p(Y | \theta, K) = \prod_{i=1}^n p(y_i | \theta) \propto \exp(-nE(\theta))$$

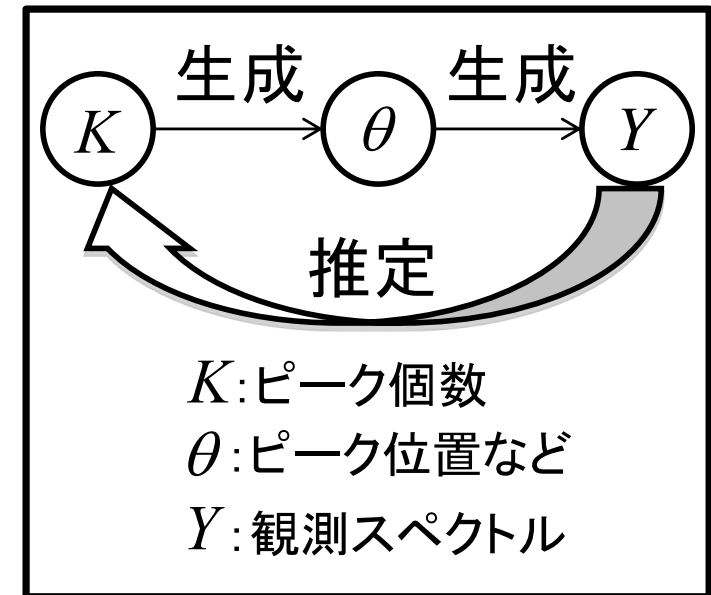
4. 無駄な自由度の系統的消去: 周辺化, 分配関数

$$p(K, Y) = \int p(K, \theta, Y) d\theta$$

$$p(K | Y) = \frac{p(Y | K) p(K)}{p(Y)} \propto p(K) \int \exp(-nE(\theta)) p(\theta) d\theta$$

$$F(K) = -\log \int \exp(-nE(\theta)) p(\theta) d\theta$$

自由エネルギーを最小にする個数 K を求める.



自由エネルギーの数値的計算法

レプリカ交換法の性質を巧妙に使う

自由エネルギー:

$$F = -\log \int \exp\left(-\frac{n}{\sigma^2} E(\theta)\right) p(\theta) d\theta$$

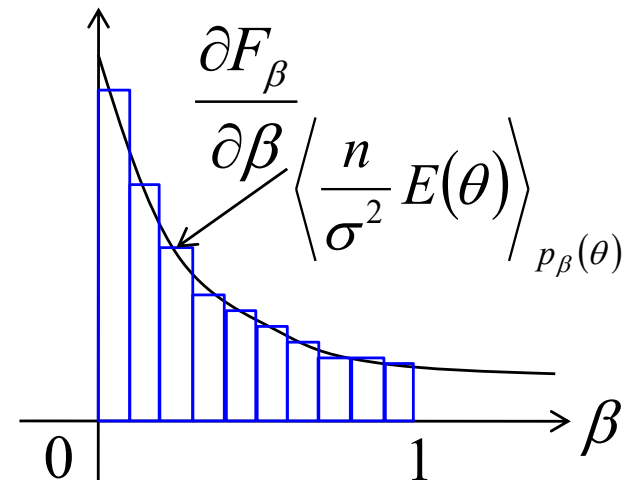
以下のように、補助変数 β を導入する。 β : 逆温度

$$F_\beta = -\log \int \exp\left(-\frac{n}{\sigma^2} \beta E(\theta)\right) p(\theta) d\theta \quad (F_{\beta=0} = 0)$$

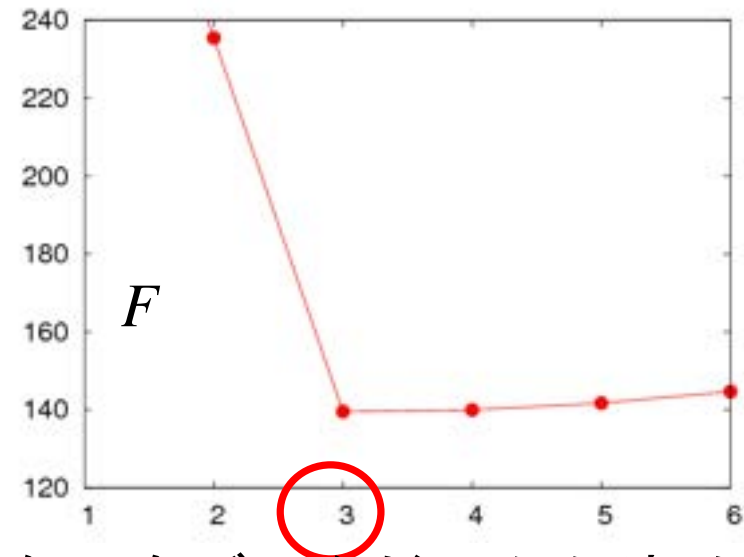
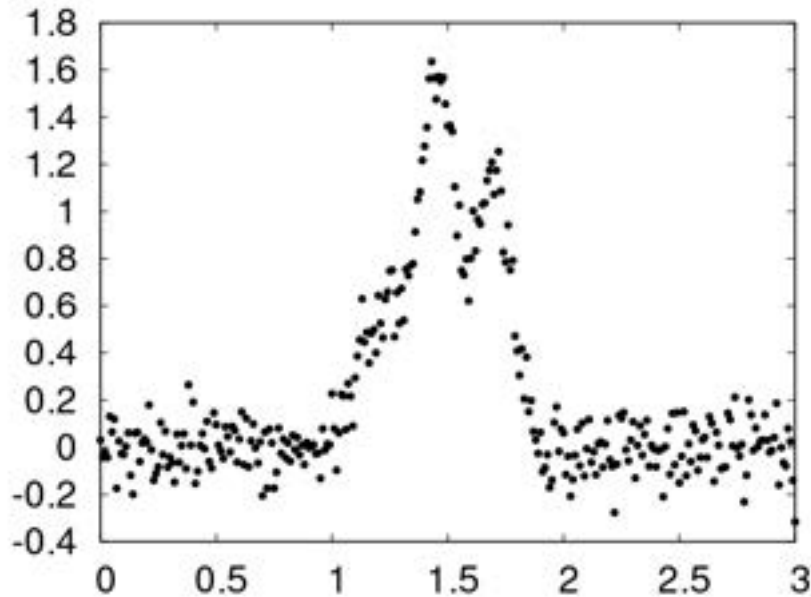
$$F = F_{\beta=1} = \int_0^1 d\beta \frac{\partial F_\beta}{\partial \beta} \quad \begin{array}{l} \text{たくさんの温度でのシミュレーションが必要} \\ \rightarrow \text{各温度でのエネルギー平均(すでにやってる)} \end{array}$$

$$\frac{\partial F_\beta}{\partial \beta} \dots \begin{array}{l} \text{確率分布 } p(\theta; \beta) \text{ に従う} \\ \text{二乗誤差 } \frac{n}{\sigma^2} E(\theta) \text{ の期待値} \end{array}$$

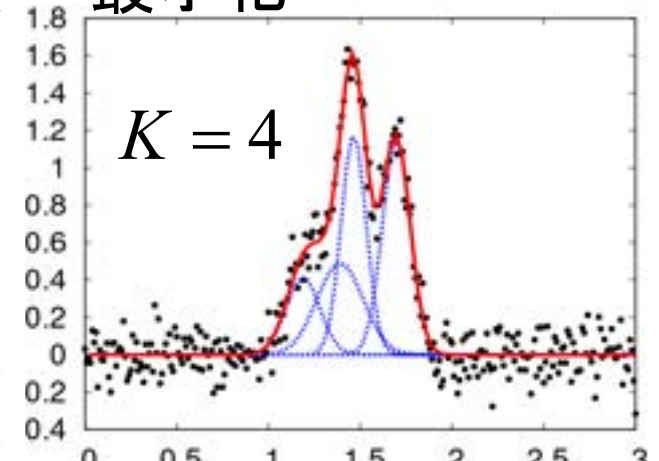
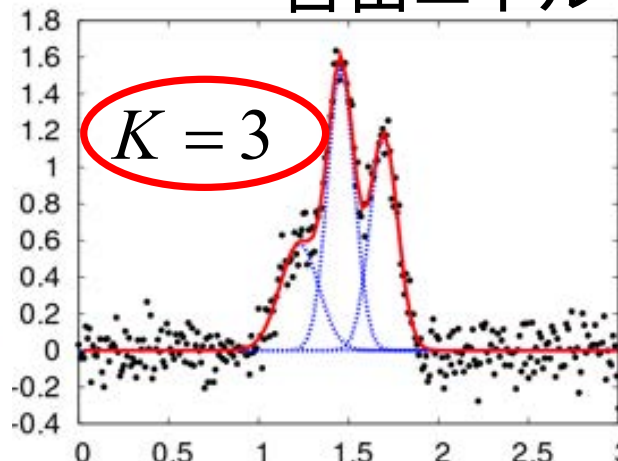
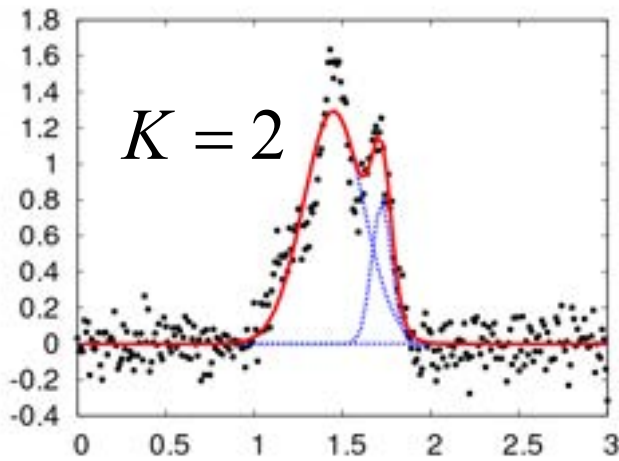
$$p_\beta(\theta) \propto \exp\left(-\frac{n}{\sigma^2} \beta E(\theta)\right) p(\theta)$$



スペクトル分解



最適な K をデータだけから決める
自由エネルギー最小化



Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

分光学の諸問題: アンケート

- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。
- そのような方は、一度ベイズ計測をお試しください。

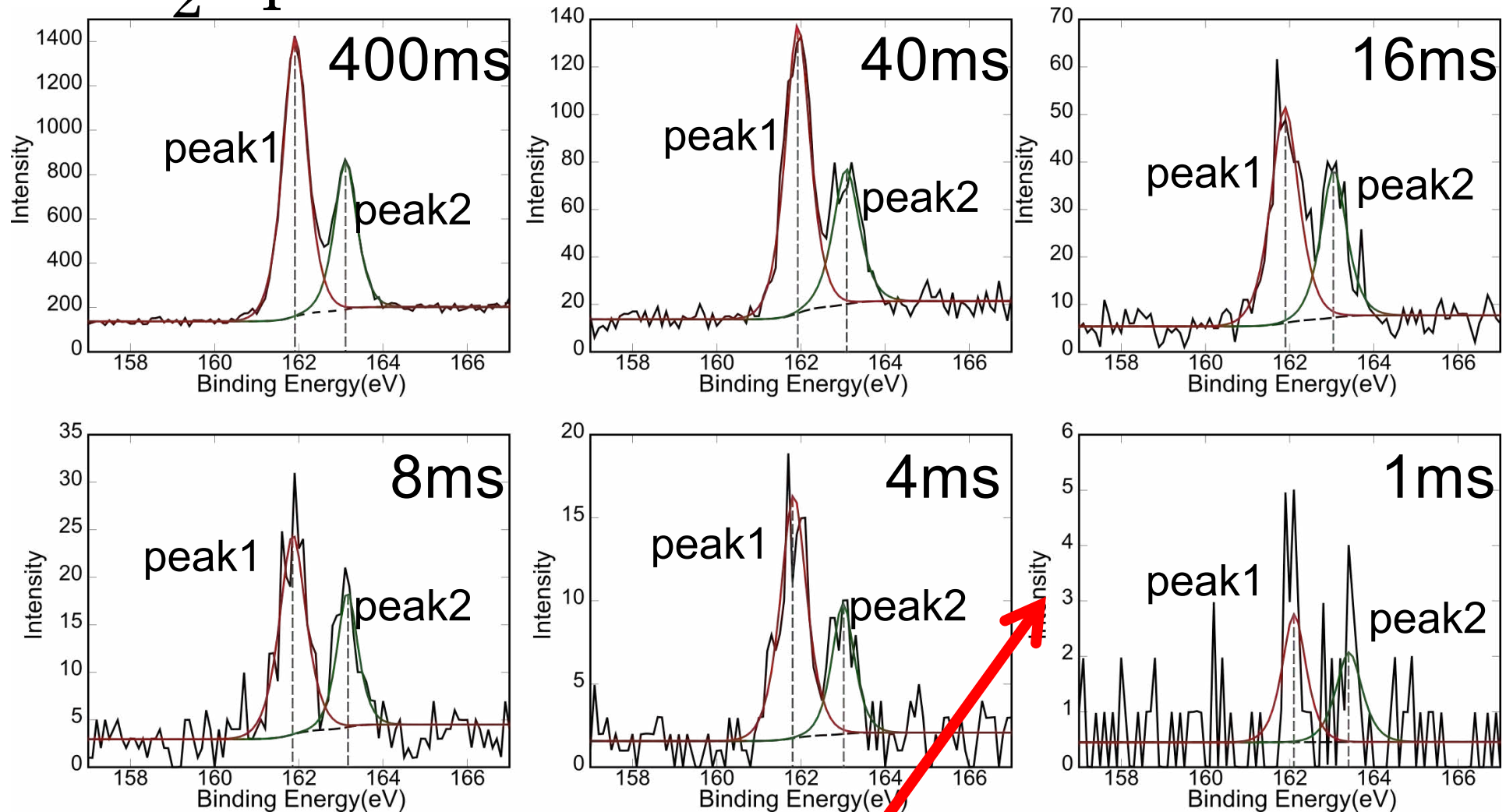
内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

例題2: XPS

どこまで時間窓を小さくできるか

MoS₂ 2p



ここまで

ベイズ推論の拡張性

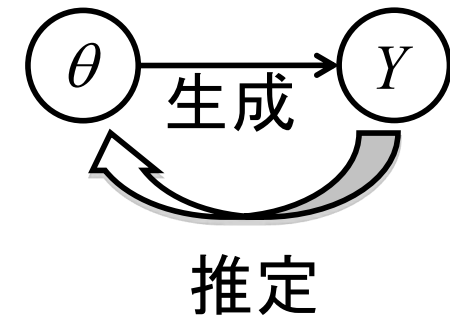
光電子の量子性を考慮する(ポアソン分布)

■ 事後確率: $p(\theta | Y) = \frac{p(Y | \theta)p(\theta)}{p(Y)}$ $Y = \{y_i\}_{i=1}^n$

$$p(\theta | Y) = \frac{1}{p(Y)} \exp\left(-\frac{n}{\sigma^2} E(\theta)\right) p(\theta)$$

これまでの

$$E(\theta) = \frac{1}{2} \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$



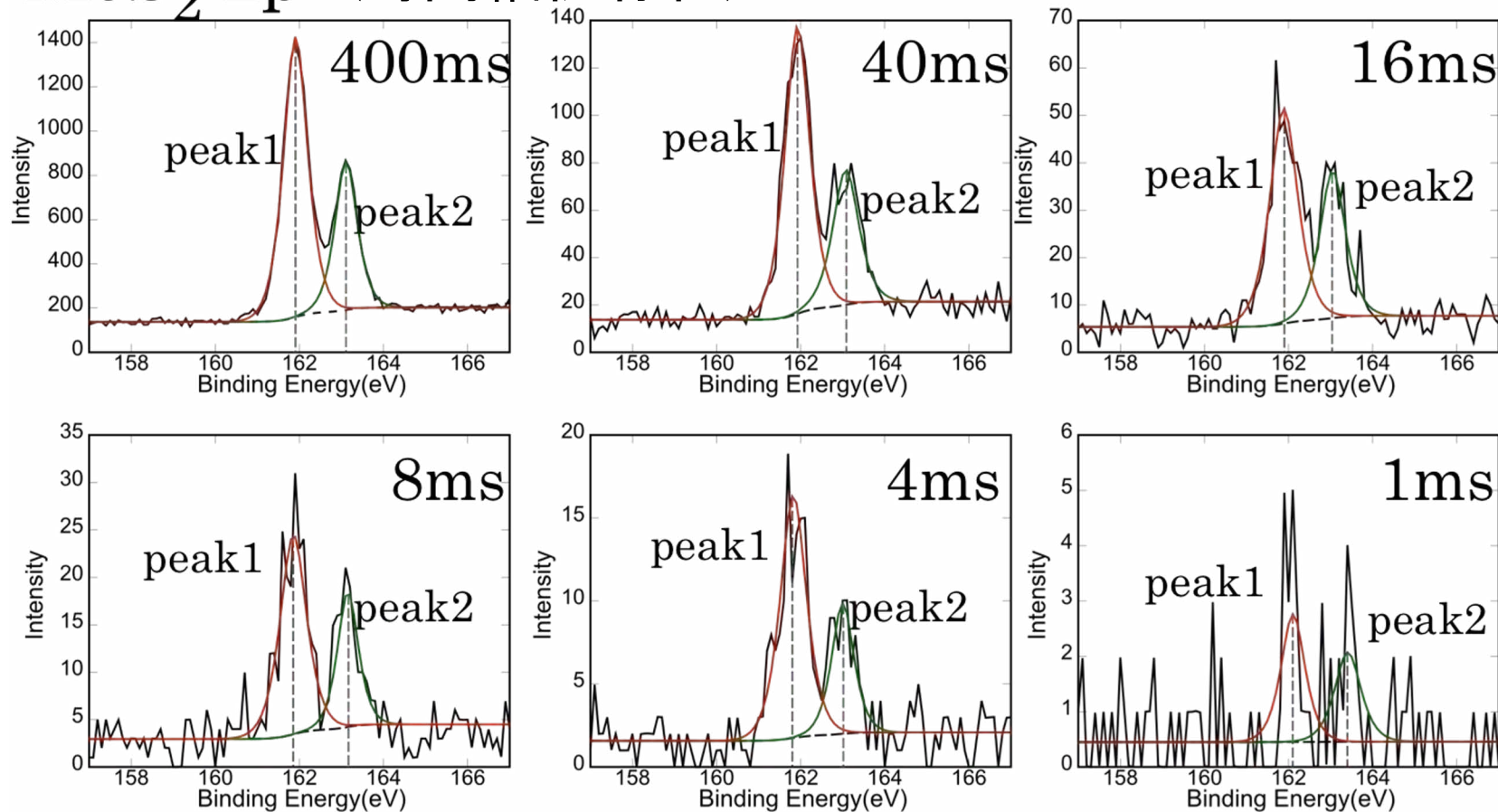
θ : ピーク位置など
 Y : 観測スペクトル

$$E(\theta) = \sum_{i=1}^n \left(y_i \log f(x_i; \theta) + f(x_i; \theta) + \sum_{j=1}^{y_i} \log(j) \right)$$

に変更するだけ

ベイズ計測 スペクトル分解のベイズ理論(Nagata *et al.* 2019)

MoS₂ 2p (時間幅依存性)



ベイズ計測

スペクトル分解のベイズ理論(Nagata *et al.* 2019)

ベイズ計測: ベイズ推論によって, ピーク位置のベイズ事後確率を計算

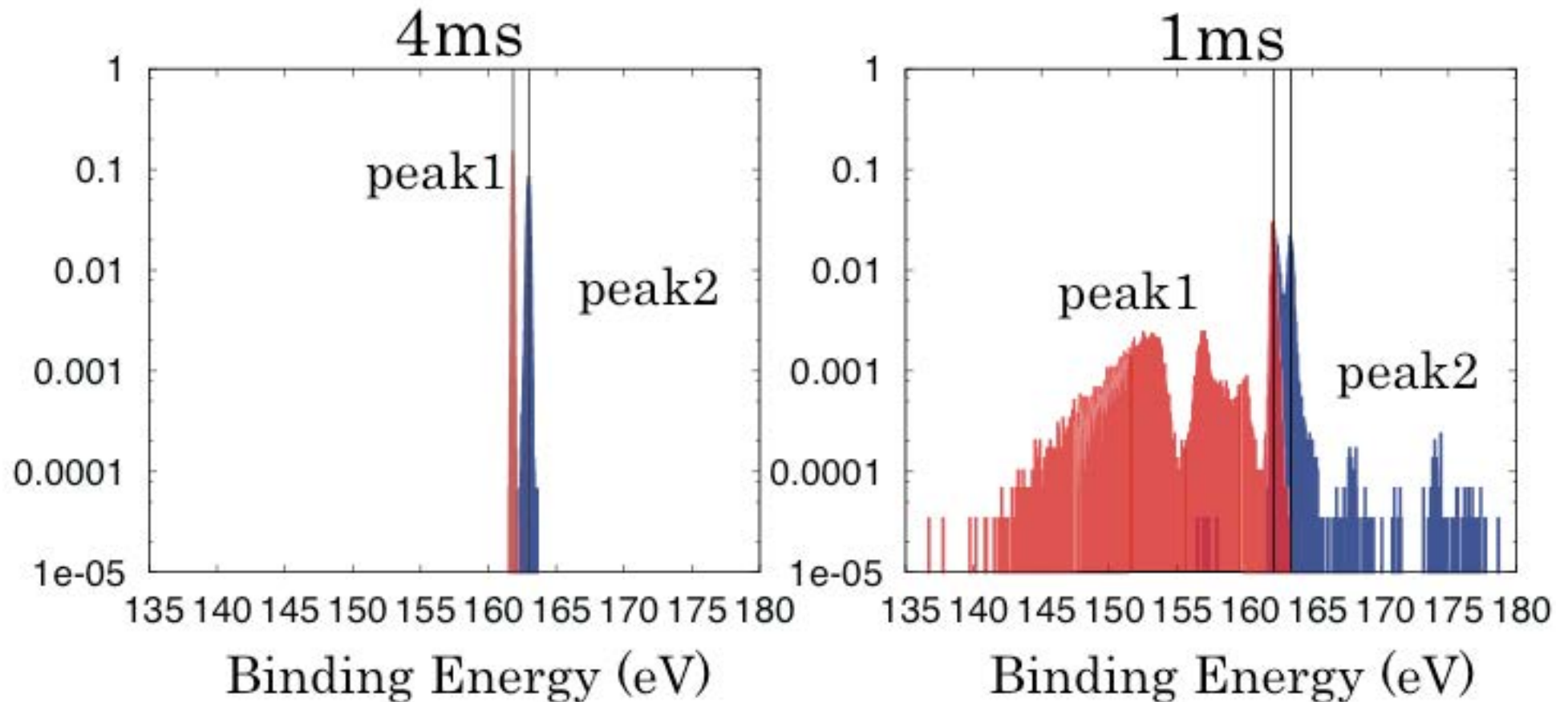
(時間幅依存性)



ベイズ計測

スペクトル分解のベイズ理論(Nagata *et al.* 2019)

ベイズ計測: ベイズ推論によって, ピーク位置のベイズ事後確率を計算
(時間幅依存性)



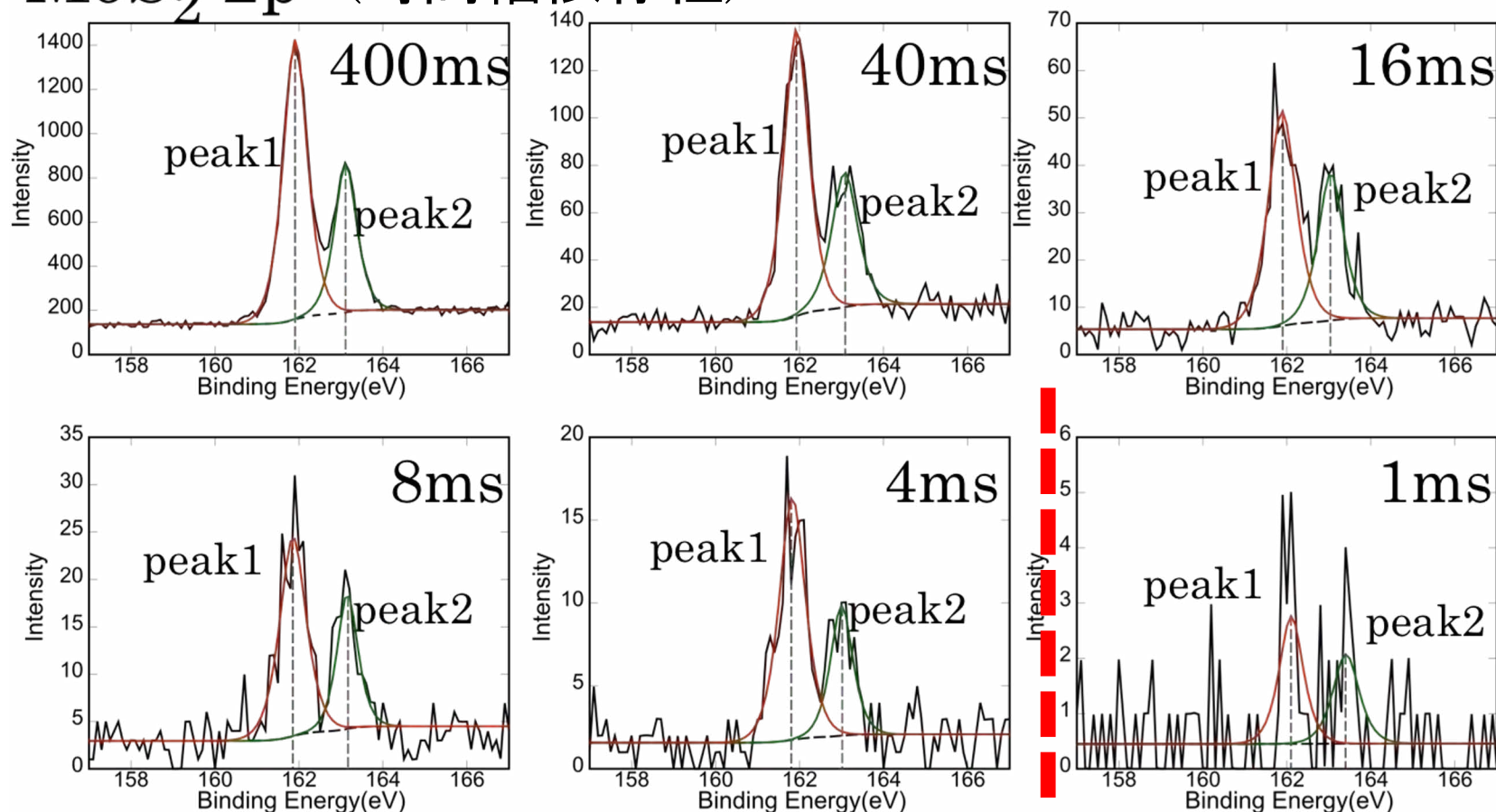
ベイズ計測

スペクトル分解のベイズ理論(Nagata *et al.* 2019)

ベイズ計測: ベイズ推論によって, ピーク位置のベイズ事後確率を計算

戦略目標: **計測限界を定量的に評価**できる枠組みの提案

MoS₂ 2p (時間幅依存性)



分光学の諸問題: アンケート

- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。
- そのような方は、一度ベイズ計測をお試しください。

内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

SPring-8全ビームラインベイズ化計画

敬称略

情報と放射光研究者のマッチング

メスバウアー

BL35XU

岡田研学生+筒井

小角散乱

BL08B2

岡田研学生+桑本

BL19B2

XAS測定

BL37XU

岡田研学生+水牧

BL39XU

放射光ユーザーへの展開

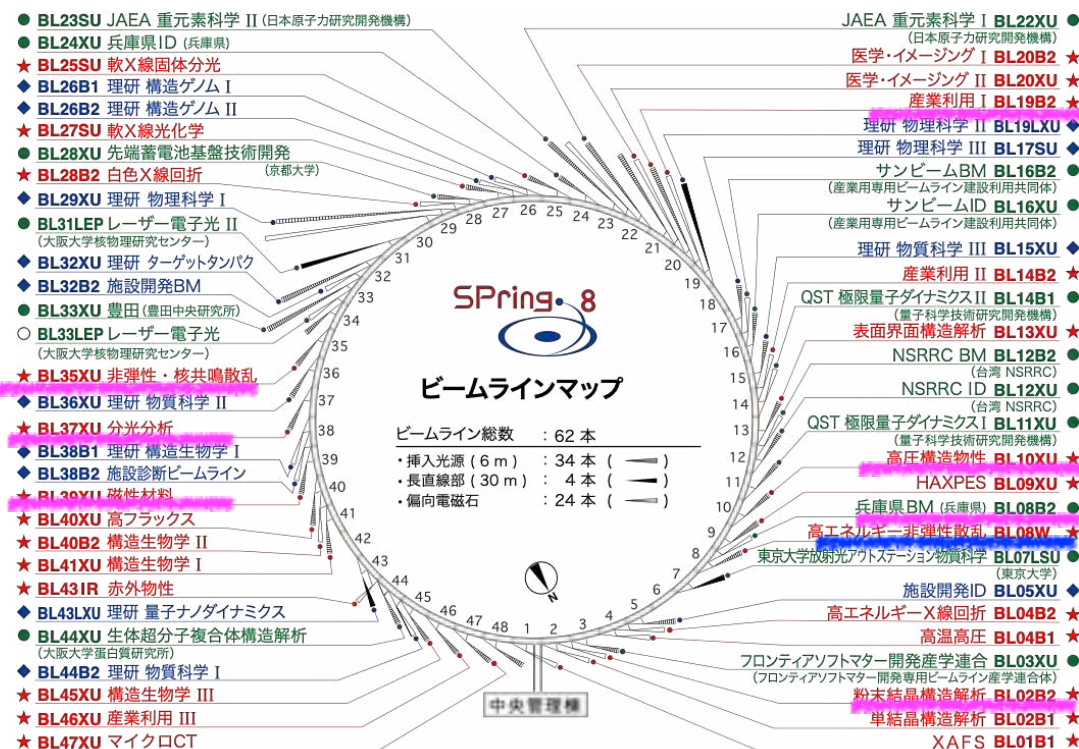
時分割XRD

BL02B2

横山優一+河口彰吾、沙織

BL10XU

ユーザー: 公立大、東工大



赤色BLが共用BL(JASRI担当): 計26本

全BL本数: 62本

来年度には過半数をこえる予定

年度	2021	2022	2023
導入	2	8	14
全BL	26	26	26

内容

- 自己紹介
- 複雑理工学専攻とデータ駆動科学
- スペクトル分解
 - スペクトル分解の従来法
 - ベイズ計測
 - レプリカ交換モンテカルロ法
 - モデル選択
 - 計測限界
- SPring-8全ビームラインベイズ化計画
- 高次元データ駆動科学教育プログラム

高次元データ駆動科学教育プログラム

- 「推論する」「測定する」「計算する」ことについて学
融合的なカリキュラムで視野の広い人材を育てま
す.
- 今日の内容はデータ駆動科学入門Iで講義します
- プログラム修了証書がもらえます
 - 短期集中講義と基盤系各専攻の通常講義
 - 短期集中講義4単位以上の履修
 - 通常講義を含め合計6単位
 - 新領域創成科学研究科長よりプログラム修了書
- ホームページ
<http://sasakilab.k.u-tokyo.ac.jp/HD3/>