The University of Tokyo

EMC ベイズ推定を全てのユーザーへ

片上 舜
東京大学　大学院新領域創成科学研究科

自己紹介

- 東京大学・大学院理学系研究科 岡田研 (2016 ~ 2022)
  - 学位論文

    「ベイズ推論による物理モデルに対するパラメータ分布推定」
- 東京大学・大学院新領域創成科学研究科 助教（2022/04 ~ ）
  - 物理計測データに対してのベイズ解析
  - ベイズ計測オープンソースソフトウェア

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。

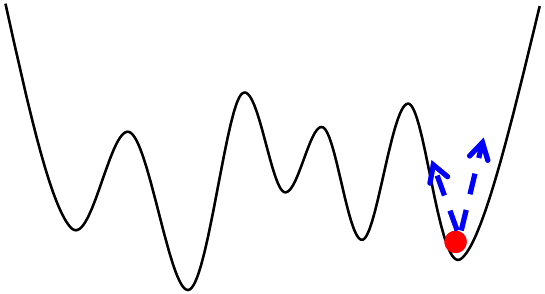
- そのような方は、一度ベイズ計測をお試しください。

ベイズ推定って、どうやって実装したらいいのか。

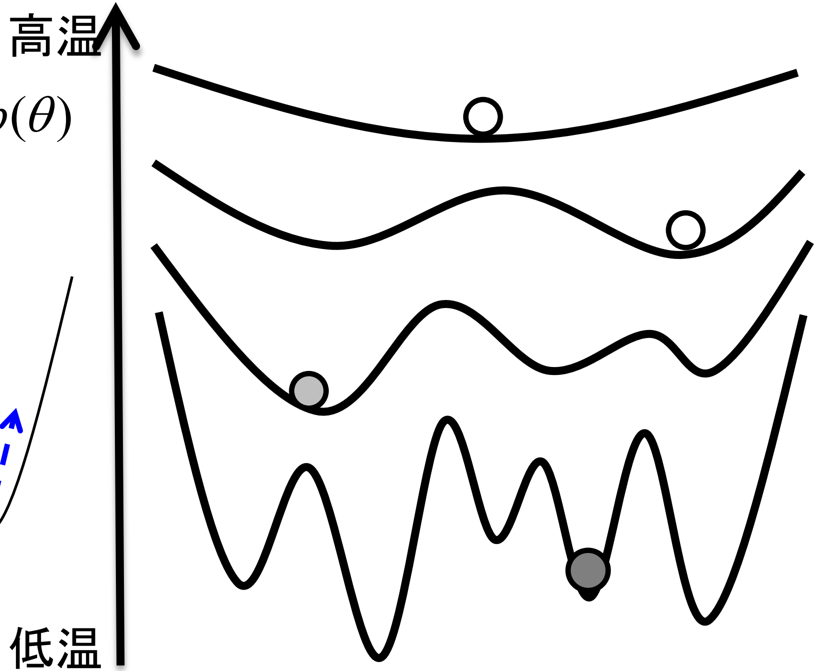
レプリカ交換モンテカルロ法
ランダムスピン系の知見から

メトロポリス法

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



レプリカ交換モンテカルロ法



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

情報統計力学(Statistical Mechanical Informatics)へ

01 計測科学における理想的なベイズ推論ライブラリ

ベイズ推論の高速な実行

- 高速かつ効率的なEMCの計算実装
- 迅速に実装可能かつ柔軟なモデル構築がフレームワーク
- 実行解析結果の可視化

ベイズ推論

$$P(\theta|D) \propto P(D|\theta) \times P(\theta)$$

事後分布　尤度(モデル)　事前分布

θ : 物理量 (モデルパラメータ)
D : 計測データ

モデルの実装 → データの取り込み → 解析結果の確認

ベイズ推論ワークフロー

ベイズ計測オープンソースライブラリの構築

ベイズ計測の勉強の流れ

1. 線形回帰モデル$y=ax+b$のベイズ計測の解析計算
2. 線形回帰モデル$y=ax+b$をレプリカ交換モンテカルロ法で数値解析し、1の解析結果と同じ結果が出ることを確認する

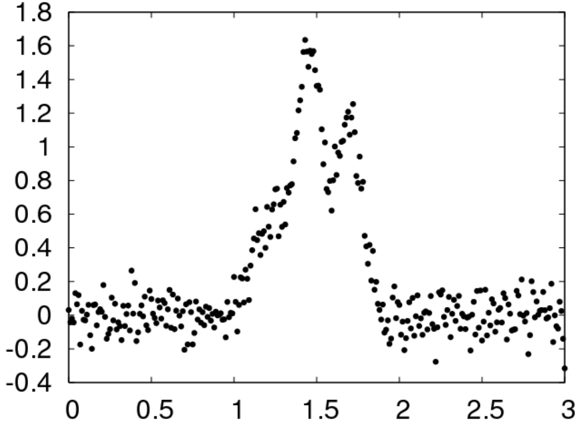
   汎用プログラム: 次の片上さんの講演

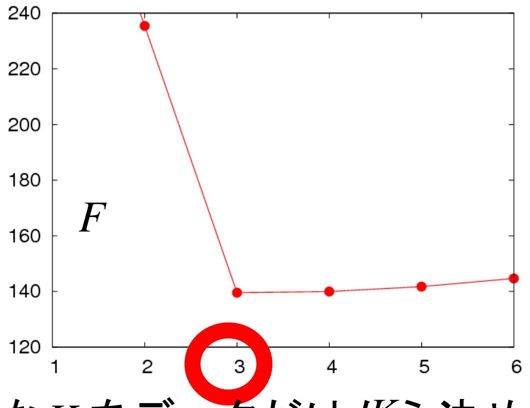
3. ベイズ的スペクトル分解をレプリカ交換モンテカルロ法で数値解析する

   汎用プログラム: 次の片上さんの講演

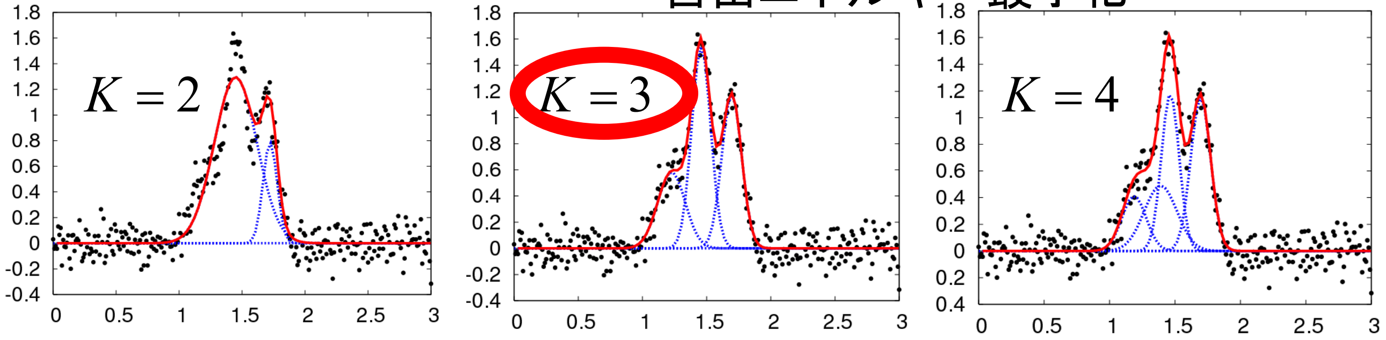
4. 各自のテーマに入る。

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

スペクトル分解の定式化

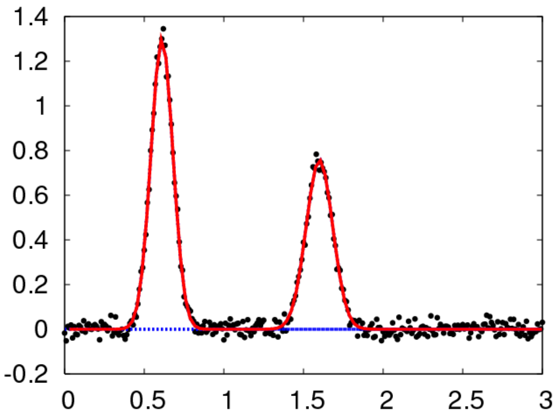
ガウス関数(基底関数)の足し合わせにより，スペクトルデータを近似

観測データ:$D=\{x_i, y_i\}_{i=1}^{n}$

x_i :入力　y_i :出力

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

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



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）

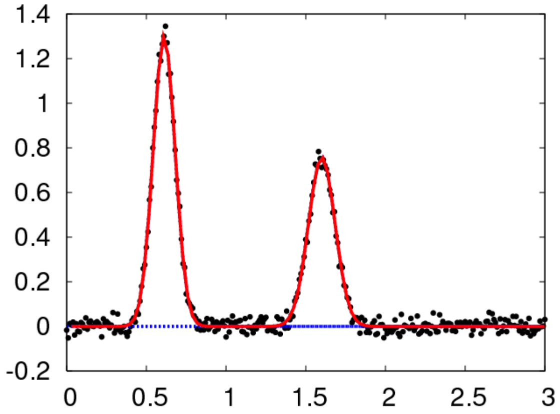
$$E(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i;\theta)\right)^2$$

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，

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

それぞれの出力 y_i が，独立であるとすると，

$$p(Y \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta) \propto \exp\left(-nE(\theta)\right) \qquad Y = \{y_1, \cdots, y_n\}$$

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

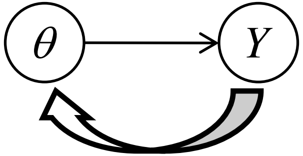
ボルツマン分布

ベイズ推論：因果律を組み込んでデータ解析

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

$$\Downarrow$$

生成（因果律）



＜ベイズの定理＞

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

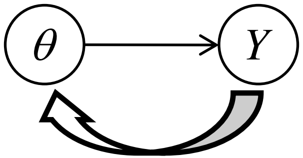
$p(\theta\,|\,Y)$ ：事後確率。データが与えられたもとでの，パラメータの確率.

$p(\theta)$ ：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

ベイズ推論：因果律を組み込んでデータ解析

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

生成（因果律）



＜ベイズの定理＞

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

$p(\theta|Y)$ ：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

事前分布はEMCのソースコードを参考

データからモデルに対する事前知識を考察する。

(何もなければ何もないで良い)

上図を以下の式でフィッティングする。

$$y = \sum_i a_i \exp(-b_i(x - \mu_i)^2/2)$$

データから恐らく

ピーク強度は正の値で 1 程度

ピーク位置は 1.5 程度

逆分散は正の値で 1 程度

```cpp
#define gsl_ran_normal_pdf gsl_ran_gaussian_pdf
    inline double SpectralDecompositionGaussEMC::SamplingFromPrior(const MBHP_i & MBHP_i, const
    static std::gamma_distribution<> prior_0(2,0.5); // ← aに関する事前分布
    if(MBHP_i.P ==0) return prior_0(engine);
    static std::normal_distribution<> prior_1(1.5,0.5); // ← muに関する事前分布
    if(MBHP_i.P ==1) return prior_1(engine);
    static std::gamma_distribution<> prior_2(14,10); // ← bに関する事前分布
    if(MBHP_i.P ==2) return prior_2(engine);
    return 0;
};
inline double SpectralDecompositionGaussEMC::ProbCalculationOnPrior(const MBHP_i & MBHP_i, const
    if (MBHP_i.P == 0) return gsl_ran_gamma_pdf(value,2,0.5); // ← aに関する事前分布
    if (MBHP_i.P == 1) return gsl_ran_gaussian_pdf(value - 1.5,0.5); // ← muに関する事前分布
    if (MBHP_i.P == 2) return gsl_ran_gamma_pdf(value,14,10); // ← bに関する事前分布
    return 0;
};
```

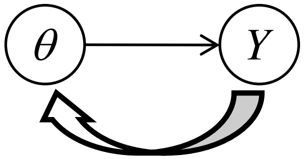
prior.hpp

ベイズ推論：因果律を組み込んでデータ解析

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

$$\Downarrow$$

生成（因果律）



＜ベイズの定理＞

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

$p(\theta|Y)$ ：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

フォワードモデルは
EMCのソースコードを参考

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

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

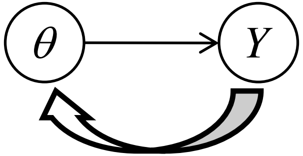
targetfunction.hpp

```
inline double SpectralDecompositionGaussEMC::TargetFunction(const vector<double> & x, const vectorMBHP
    double y = 0;int BaseNum = op_.base_nums_[0];
    const auto a = [&](int B) -> const double& {return MBHPparams[0][(int)B][0][0][0];};
    const auto mu = [&](int B) -> const double& {return MBHPparams[0][(int)B][0][1][0];};
    const auto b = [&](int B) -> const double& {return MBHPparams[0][(int)B][0][2][0];};
    rep(i,BaseNum){
      y += a(i) * exp(- 0.5 * b(i)*pow(x[0]-mu(i),2));
    }
    return y;
};
```

ベイズ推論：因果律を組み込んでデータ解析

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

生成（因果律）



⇩

＜ベイズの定理＞

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

$p(\theta\mid Y)$ ：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

サンプリング設定は
EMCのソースコードを参考

execution.ipynb

```
data = {
    # ピーク数
    "base_num": 3,
    "replica_num": 36,
    "gamma": 1.4,
    "sample_num": 1000,
    "burnin_num": 1000,
    "C d": [
        [0.5, 0.6],
        [0.5, 0.6],
        [20, 0.9]
    ]
}
updateConfig(data)
```

サンプリング設定は
EMCのソースコードを参考

execution.ipynb

ステップサイズ調整パラメータ

```
data = {
    # ピーク数
    "base_num": 3,
    "replica_num": 36,
    "gamma": 1.4,
    "sample_num": 1000,
    "burnin_num": 1000,
    "C d": [
        [0.5, 0.6],
        [0.5, 0.6],
        [20, 0.9]
    ]
}
updateConfig(data)
```

サンプリング設定は
EMCのソースコードを参考

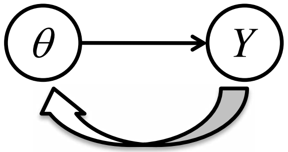
execution.ipynb

```
data = {
    # ピーク数
    "base_num": 3,
    "replica_num": 36,
    "gamma": 1.4,
    "sample_num": 1000,
    "burnin_num": 1000,
    "C d": [
        [0.5, 0.6],
        [0.5, 0.6],
        [20, 0.9]
    ]
}
updateConfig(data)
```

ベイズ推論：因果律を組み込んでデータ解析

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

生成（因果律）



＜ベイズの定理＞

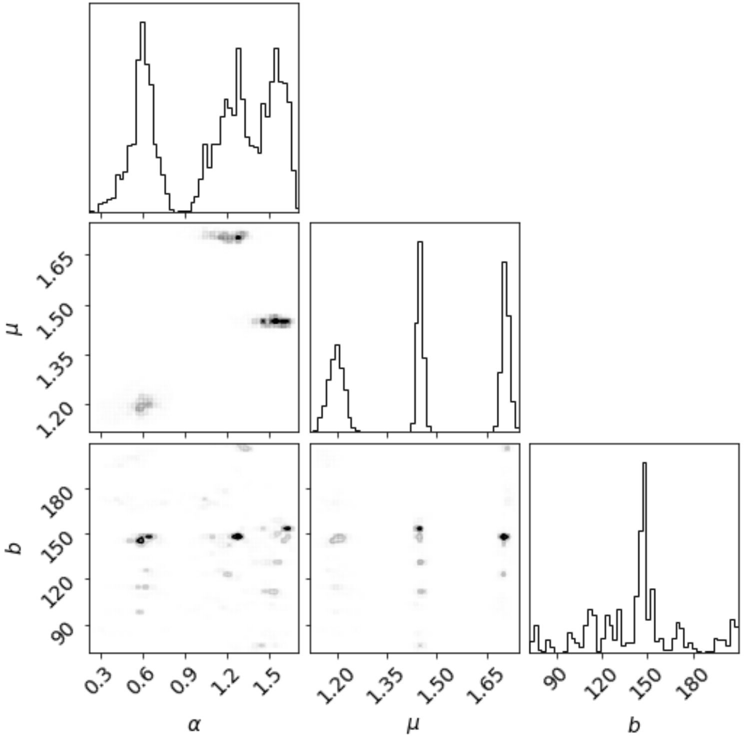
$$p(\theta \mid Y) = \frac{p(Y \mid \theta)p(\theta)}{p(Y)} \propto \exp\left(-nE(\theta)\right)p(\theta)$$

$p(\theta \mid Y)$ ：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.

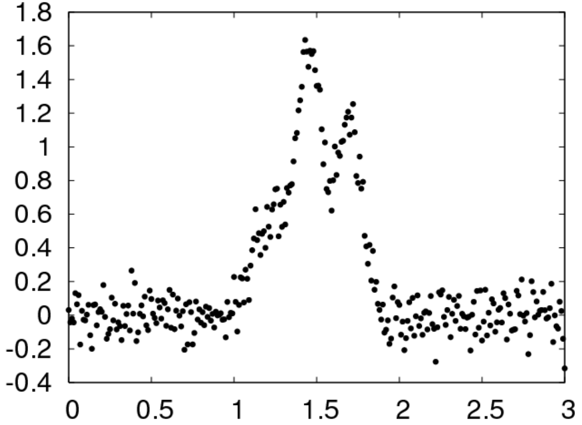
$p(\theta)$ ：事前確率。あらかじめ設定しておく必要がある。
これまで蓄積されてきた科学的知見

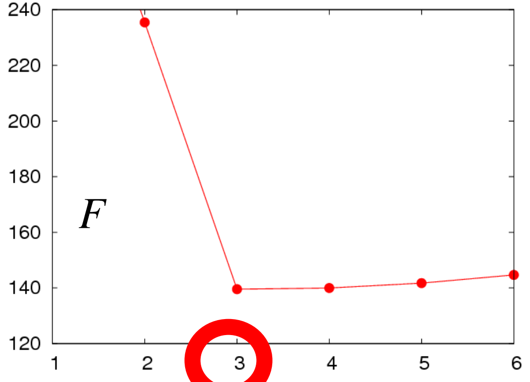
事後分布は
EMCのソースコードを参考

execution.ipynb

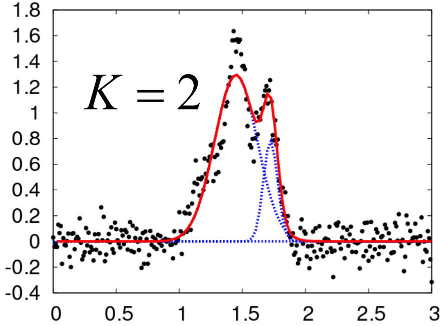


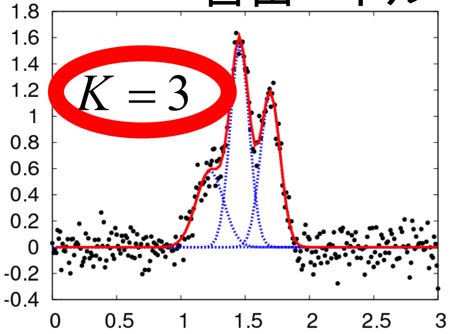
スペクトル分解

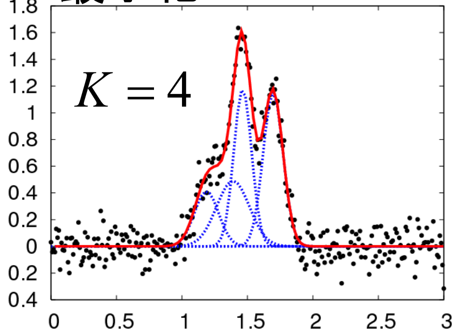




最適なKをデータだけから決める
自由エネルギー最小化







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

EMCを使った際の仕事のプロセス

- 皆さんのラボのメンバーが具体的にプロジェクトを動かす
- EMCを使って、$y=ax+b$とスペクトル分解のトレーニングをする
- 計測のフォワードモデルを構築
  - フィッティングの式があればOK
- フォワードモデルを組み込んだベイズ計測の枠組みを構築
  - まずはフォワードモデルから生成した人工データをベイズ計測する
  - VMA(Virtual Measurement Analysis)
  - こちらの具体的な作業を皆さんのラボのメンバーの方に担していただく
- ベイズ計測により得られる知見を従来のフォワードモデルによる解析と比較して，ベイズ計測のアドバンテージを明確化
- 論文を書く