

データ点数有限のベイズ計測の理論[1] 2024年8月18日(日)

東京大学 大学院新領域創成科学研究科 複雑理工学専攻

片上 舜 岡田真人

©2024 Shun Katakami and Masato Okada

1. はじめに

ベイズ計測(Bayesian Measurement)とは、幅広い領域を持つベイズ推論を計測科学に必要な部分だけ取り出したコンパクトな情報科学的枠組みである[2]。ベイズ計測はベイズ計測三種の神器の(1) パラメータの事後確率分布推定[3] (2) モデル選択[3] (3) ベイズ統合[4, 5]から構成される。(2)のモデル選択では、一つの実験/計測データに対して、複数のモデルが存在する場合に適用される。その実験/計測データを説明するのに相応しいモデルを一つ選ぶ枠組みである。(3)のベイズ統合では、一つの物質に対して、複数の実験/計測データが存在する場合に、どのように、それら複数の実験/計測データの統合を取り扱う枠組みである。

モデル選択とベイズ統合の両方で重要な役割を演ずるのが自由エネルギーと自由エネルギー差である[3]。(2)のモデル選択では、各モデルに対する自由エネルギーを計算し、その中で一番小さな自由エネルギーを持つモデルを正しいモデルとする。つまり、モデル間の自由エネルギー差がモデル選択にとって重要な量である。(3)のベイズ統合ではまず、複数の実験/計測データを統合して、パラメータの事後確率分布や統合した際の自由エネルギーを求める。次に、先ほど求めたベイズ統合した際の自由エネルギーと、複数の実験/計測データを個別にベイズ計測で解析した際の自由エネルギーの和を比較し、その自由エネルギー差が負の場合には、ベイズ統合すべきだと判断する。一方、その自由エネルギー差が正の場合には、ベイズ統合してはいけないと判断する。このように、モデル選択とベイズ統合の両方で重要な役割を演ずるのが自由エネルギー差である。

2. ベイズ統合の具体例

YokoyamaらはNiOのXPS(X線光電子分光法スペクトル)とXAS(X線吸収スペクトル)のベイズ統合を議論した[4]。それまでのデータ統合の通常法では、複数の実験/計測データの誤差関数をまず定義し、それらの線形和からなるデータ統合の誤差関数を定義する。ここで、線形和の係数は研究者が決めたり、CV(交差検証法)で決めたりしていた。従来法では、データ統合をするという立場でデ

ータ解析が始まっており、データ統合しない場合は全く想定されていなかった。一方、ベイズ統合でも、ベイズ統合の誤差関数は複数の実験/計測データの誤差関数の線形和で表されるが、その重みはベイズ統合の理論の中で実験/計測データのみから自動的に決定される。また、前述のように自由エネルギー差を用いて、データ統合すべきかどうかは実験/計測データのみから自動的に決定される。この二つの点が、従来法との著しい相違点であり、それが統一的なベイズ計測の枠組みから自然に導出されるところが、学問としては一段上である。

Yokoyama らの XPS と XAS のベイズ統合の講演会でも、必ず統合して良いのか？統合しなくても良いというケースは無いのかという質問が出ていた。前述のようにデータ統合の通常法では、データ統合するのが前提であったら、実験家を中心として、まず実験/計測データだけから、データ統合すべきかどうかを決めることが必要であるという現場の声を聞いたことになる。まさしくこれに応えるのがベイズ統合である。我々は希土類化合物の比熱と磁化率のベイズ統合を議論する中で[5]、ベイズ統合しなくて良いというケースがないかを探したが、XPS と XAS のベイズ統合の同様に、全てのケースで、自由エネルギー差は負になり、ベイズ統合すべきだという結果が得られた。

3. データ数 N 有限の理論の描像

3-1 ベイズ統合のメカニズムの描像

常にデータ統合すべきかどうかをというベイズ統合にとっては根源的な問題をできれば解析的理論で決着をつけたいと思い、解析計算でベイズ計測を取り扱えるほぼ唯一の例として知られる線形回帰モデル $y=ax+b$ を取り上げた。線形回帰モデル $y=ax+b$ の理論は解析計算がしやすいためだけに導入する Toy モデルの理論ではない。このモデルは、系の線形応答を取り扱うための実際的なモデルであるとともに、大学理系の標準的リテラシーで取り扱われるものであるために、ベイズ計測の習得のためのロールモデルでもある。このように、計測データ数 N 有限での数値実験の解明等実際的な問題から、線形回帰モデル $y=ax+b$ の理論構築が始まるのは、大変健全なことである。

ここでは、ベイズ推論の主に自由エネルギーの理論的枠組みを俯瞰する。我々の知る限り、Schwarz により提案された BIC に代表されるように、全ての理論はデータ数 N が無限大の極限の漸近論しか議論していない。このような漸近論では、ここで議論している N が有限である場合を取り扱うことはできない。

そこで私はまず、計測データ数 N が有限の場合に、最も簡単な一変数の線形回帰モデル $y=ax$ で、統合すべきデータが二組の場合に、何がおこるかを、図 1 と 2 に示すような描像(図)を用いて考察した。

図 1 は、二つのデータの生成モデルが同一である場合をしめす。この場合は、

直線回帰の係数は、二つのデータで同一の値 a_0 である。データ 1 の係数 a の事後確率分布 $p(a|D)$ は、図 1 左のガウス関数のように平均 a^1 のガウス分布に従うことを簡単な計算で示せる[1]。ここで D は、 N 個のデータ 1 の観測データの集合である。また、このガウス分布の分散は、 $\sigma^2 \overline{x^2}/N$ である。ここで、 σ^2 は観測ガウスノイズの分散であり、 $\overline{x^2}$ は入力 x_i の 2 乗平均 $1/N \sum_{i=1}^N x_i^2$ である。一方、データ 2 の係数 a の事後確率分布 $p(a|D)$ は、図 1 右のガウス関数のように、平均 a^2 のガウス分布に従う。ここで D は、 N 個のデータ 2 の観測データの集合である。また、このガウス分布の分散も、 $\sigma^2 \overline{x^2}/N$ である。ここで、 σ_0^2 は、観測ガウスノイズ n_i の分散である。

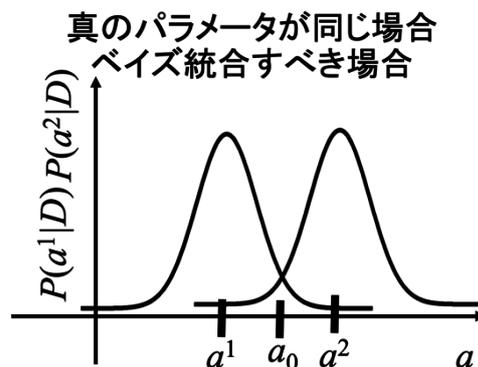


図 1: 真のパラメータが同じ場合で、ベイズ統合すべき場合

図 1 の左右二つのガウス分布の中心である a^1 と a^2 は、観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、 a_0 に近づく。さらに、左右二つのガウス分布の分散も $\sigma^2 \overline{x^2}/N$ であるので、観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、二つのガウス分布はシャープになり、 δ 関数に近づいていく。その結果、データ 1 とデータ 2 の係数 a の事後確率分布 $p(a|D)$ は、 δ 関数が重なったようになり、 a^1 と a^2 の推定値は、係数 a の真の値 a_0 に収束していく。これにより、データ 1 とデータ 2 をベイズ統合すべきである確率がほぼ 1 になるであろう。

次に先ほどとは逆の、観測ガウスノイズの分散 σ_0^2 が大きい場合や、観測データ数 N が小さい場合は、 $\sigma_0^2 N$ が大きくなるので、図 1 の左右二つのガウス分布の中心である a^1 と a^2 は、係数 a の真の値 a_0 から大きく外れ出す。さらに、左右二つのガウス分布の分散も $\sigma_0^2 N$ であるので、二つのガウス分布はブロードになる。以上二つの効果から、データ 1 とデータ 2 をベイズ統合しない方が良いという結果が得られる場合が多くなるであろう。

図 2 は、二つのデータの生成モデルが異なっている場合をしめす。この場合は、直線回帰の係数の真の値は、データ 1 に対しては a^1_0 であり、データ 2 に対しては a^2_0 である。データ 1 の係数 a の事後確率分布 $p(a|D)$ は、図 2

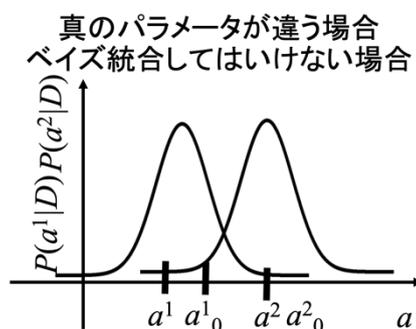


図 1: 真のパラメータが違う場合で、ベイズ統合してはいけない場合

左のガウス関数のように、平均 a^1 のガウス分布に従う [1]。このガウス分布の分散は、 $\sigma_0^2 \overline{x^2}/N$ であり、データ 2 の係数 a の事後確率分布 $p(a|D)$ は、図 1 右のガウス関数のように、平均 a^2 のガウス分布に従う。図 1 と同様に、これらのガウス分布の分散は、 σ_0^2/N である。

図 2 の左右二つのガウス分布の分散は $\sigma_0^2 \overline{x^2}/N$ であるので、観測ガウスノイズの分散 σ_0^2 が小さい場合や、観測データ数 N が大きい場合、二つのガウス分布はシャープになり、 a^1_0 を中心とする δ 関数と a^2_0 を中心とする δ 関数の 2 本の δ 関数に近づいていく。その結果、 a^1 と a^2 の推定値は、それぞれ別の係数 a の真の値 a^1_0 と a^2_0 に収束していく。これにより、データ 1 とデータ 2 をベイズ統合すべきでない確率がほぼ 1 になるであろう。

次に先ほどとは逆の、観測ガウスノイズの分散 σ_0^2 が大きい場合や、観測データ数 N が小さい場合は、 σ_0^2/N が大きくなるので、図 2 の左右二つのガウス分布の中心である a^1 と a^2 は、それぞれのデータ生成のモデルの係数の真の値 a^1_0 と a^2_0 から大きく外れ出す。さらに、左右二つのガウス分布の分散も σ_0^2/N であるので、二つのガウス分布はブロードになる。以上二つの効果から、データ 1 とデータ 2 をベイズ統合した方が良いという結果が偶然得られる場合が多くなるであろう。

以上まとめると、常にデータ統合すべきかどうかをというベイズ統合にとっては根源的な問題をできれば解析的理論で決着をつけるために線形回帰モデル $y=ax+b$ を取り上げた。その結果、ベイズ統合では、真のパラメータが同じ場合で、データ統合すべき場合と真のパラメータが違う場合で、データ統合してはいけない場合のそれぞれに対して、観測ガウスノイズの分散 σ_0^2 が小さい場合や、観測データ数 N が大きい場合、統合するかしないかに対して正しく判断できることを示した。また、観測ガウスノイズの分散 σ_0^2 が小さい場合や、観測データ数 N が大きい場合、データ統合に関して間違っただけの判断をする場合があることがわかった。また、それぞれの場合の確率を、定量的に議論できるのではないかという可能性があることがわかった。ここでまとめたように、データ統合に関して、データ統合するかしないかの判断ができるのは、パラメータの確率分布の推定を行っているベイズ統合だけで、通常データ統合する前提の手法では、当然それは無理であり、今後のデータ統合に関して、ベイズ統合は必須の情報数理的枠組みであることがわかった。

3-2 モデル選択のメカニズムの描像

ベイズ統合のメカニズムの描像が獲得できると、モデル選択に関しても描像を構築できるような気がしてきた。我々は [1] の中で、入力 x の平均値を 0 とおいた。これは x 軸の原点の取り方の変更に対応するので、この操作で一般性は失わない。線形回帰モデル $y=ax+b$ をモデル 1 とおこう。この x の平均値を 0 の仮

定から、線形回帰モデル $y=ax+b$ の a と b の事後確率分布 $p_1(a,b|D)$ は $p_1(a|D) p_1(b|D)$ と書ける[1]。線形回帰モデル $y=ax$ をモデル 2 としよう。このモデル 2 の係数 a の事後確率分布 $p_2(a|D)$ は、モデル 1 の事後確率分布 $p_1(a,b|D)=p_1(a|D) p_1(b|D)$ の $p_1(a|D)$ と同じになることを示すことができる。この解析的な結果から、線形回帰モデル $y=ax+b$ と $y=ax$ のモデル選択は、 $p_1(b|D)$ がのように自由エネルギーに寄与するかを考えれば良いことがわかる。

$p_1(b|D)$ は平均 b_0 分散 σ^2/N のガウス分布に従う。ここで b_0 はモデル 1 の切片の真の値である。観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、 $p_1(b|D)$ は分散 σ^2/N は小さくなるので、 $p_1(b|D)$ は b_0 を中心とする δ 関数に収束する。データの生成モデルがモデル 1 の $y=ax+b$ である場合、 b の推定値は非ゼロである b_0 に収束し、モデル選択の結果はモデル 1 の $y=ax+b$ となる確率が 1 に近くなるであろう。一方、観測ガウスノイズの分散 σ^2 が大きい場合や、観測データ数 N が小さい場合、 $p_1(b|D)$ の分散 σ^2/N は大きくなるので、 $p_1(b|D)$ は b_0 を中心とするブロードなガウス分布になる。その結果、推定値 b の値が 0 になる場合が生じるので、モデル選択の結果はモデル 2 の $y=ax$ となる場合が生じ、モデル選択の結果が誤ることが生じる。

次に、データの生成モデルがモデル 2 の $y=ax$ である場合を議論する。観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、 $p_1(b|D)$ の分散 σ^2/N は小さくなるので、 $p_1(b|D)$ は 0 を中心とする δ 関数に収束する。その場合、モデル選択の結果はモデル 2 の $y=ax$ となる確率が 1 に近くなるであろう。一方、観測ガウスノイズの分散 σ^2 が大きい場合や、観測データ数 N が小さい場合、 $p_1(b|D)$ の分散 σ^2/N は大きくなるので、 $p_1(b|D)$ は平均 0 のブロードなガウス分布になる。その結果、推定値 b の値が非 0 になる場合が生じるので、モデル選択の結果はモデル 1 の $y=ax+b$ となる場合が生じ、モデル選択の結果が誤ることが生じる。

以上まとめると、モデル選択では、観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、統合するかしないかに対して正しく判断できることを示した。また、観測ガウスノイズの分散 σ^2 が小さい場合や、観測データ数 N が大きい場合、モデル選択に関して間違った判断をする場合があることがわかった。さらにベイズ統合と同様に、それぞれの場合の確率を、定量的に議論できるのではないかという可能性があるとわかった。

4. データ数 N 有限の理論とその結果のまとめ[1]

4.1 データ数 N 有限の理論の導出の概説

我々は 3. で述べた考察を、線形回帰モデル $y=ax+b$ に関して具体化する理論を

構築することに成功した[1]. まず入力 x の平均値をとした. その結果, 3.で述べたように係数 a と切片 b の事後確率分布は $p_1(a,b|D)=p_1(a|D)p_1(b|D)$ となる. つぎに, 計測データ数 N が有限のケースを取り扱うために, i 番目のデータ x_i i 番目のデータに重畳された平均 0 分散 σ_0^2 の観測ノイズ n_i の三つの統計量を導入した.

$$\tau_1 = \frac{1}{N\bar{x}^2} \sum_{i=1}^N x_i n_i$$

$$\tau_2 = \frac{1}{N} \sum_{i=1}^N n_i$$

$$v = \frac{1}{N} \sum_{i=1}^N n_i^2$$

ここで x_i は i 番目のデータであり, n_i は i 番目のデータに重畳された平均 0 分散 σ_0^2 の観測ノイズである. また, \bar{x}^2 は入力の 2 乗平均である

$\bar{x}^2 = \frac{1}{N} \sum_{i=1}^N x_i^2$. また, v は χ 二乗分布に従う確率変数である. 我々は, x_i と n_i から

定義される統計量をメゾ変数と定義した. 我々は, ベイズ計測三種の神器の (1) パラメータの確率分布の推定 (2) モデル選択 (3) ベイズ統合は, これら三つのメゾ変数の閉じた形で表現できることを示したので, この理論体系をベイズ計測のメゾ理論と呼ぶことにした.

ベイズ統合の描像から, ベイズ統合の自由エネルギー ΔF は τ_1 と τ_2 に依存するので, ΔF を τ_1 と τ_2 の二変数で表せて, 2次元ヒートマップで可視化できると予想していた. しかし, 上記のように ΔF は τ_1 と τ_2 だけでなく v にも依存するので, 次元ヒートマップで可視化できと考えるようになった. 苦肉の策としては, v をある値に固定して, その上で, ΔF を τ_1 と τ_2 の二変数で表し τ_1 と τ_2 の 2次元ヒートマップで可視化しようかとも思っていた.

しかし, 理論を深く展開してみると, 3.1 のように, ベイズ統合の ΔF は v に依存せず, τ_1 と τ_2 だけに依存することがわかり, τ_1 と τ_2 の 2次元ヒートマップで可視化できるとことがわかった. さらに, モデル選択についても, 3.1 のようにモデルの ΔF は τ_2 だけに依存することがわかった.

結局のところ, 3-1 と 3-2.で説明した描像は, 理論の裏付けがあるものとなった.

4.1 データ数 N 有限の理論の数値計算結果の概要

この節では、メゾ理論の数値計算の結果の概要を紹介する. 4.1.1 ではモデル選択について紹介し, 4.1.2 ではベイズ統合について紹介する.

4.1.1 モデル選択

$y=ax+b$ が真のモデルである場合

まず $y=ax+b$ が真のモデルであり, その係数 $a_0=1.0$, 切片 $b_0=1.0$, 観測ガウス分布の分散 $\sigma^2=1.0$ である場合を考える. 図 3 は, モデル選択に関する自由エネルギー差 ΔF の確率分布であり, 横軸は ΔF であり, 縦軸はその ΔF での確率分布である. 自由エネルギー差の確率分布をもとめるための積分をサンプル数 100,000 サンプルングで近似している. 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 理論曲線である. 図 3 (a), (b), (c) を比較すると, (a) の $N=5$ では, 自由エネルギー差の確率分布 ΔF は負の部分を持ち, 真のモデルが $y=ax+b$ であるにも関わらず,

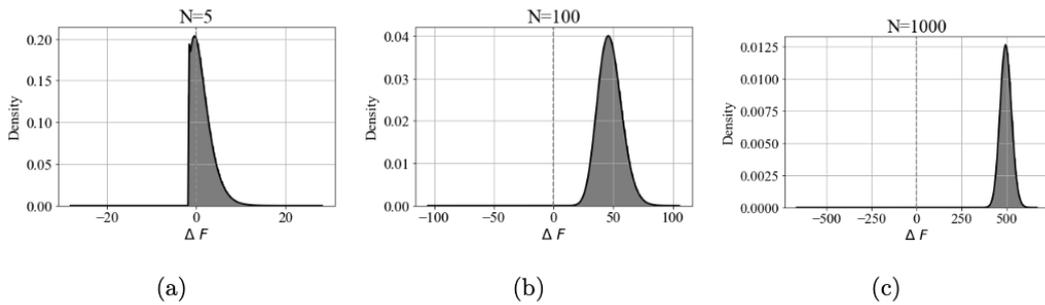


図 3: サンプル数 100,000 でのモデル選択に関する自由エネルギー差の確率分布である. $a_0=1.0$, $b_0=1.0$, $\sigma^2=1.0$ で, 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 自由エネルギー差の確率分布の理論曲線である.

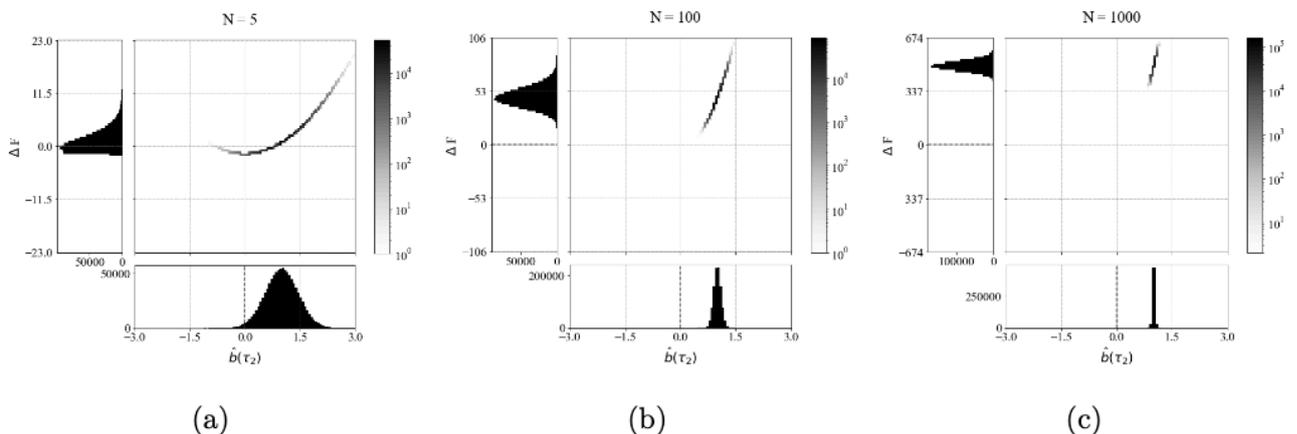


図 4: $a_0=1.0$, $b_0=1.0$, $\sigma^2=1.0$ であり, 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 縦軸は, サンプル数 100,000 でのモデル選択に関する自由エネルギー差 ΔF の確率分布の 2 次元グラフである. 横軸は, $\widehat{b}(\tau_2)$ の分布であり, ガウス分布に従う.

モデル選択の結果は $y=ax$ と間違えている. 図 3 (b) $N=100$ になると, 自由エネルギー差 ΔF の確率分布はすべて正のところのみ値を持ち, モデル選択の結果は $y=ax+b$ と正しく判断されている. ここから, 実際の実験計測でも, 計測点数が 100 以上であれば, 正しくモデル選択できることがわかる.

図 4 の縦軸は, サンプル数 100,000 でのモデル選択に関する自由エネルギー差 ΔF の確率分布の 2 次元グラフである. 横軸は, $\widehat{b}(\tau_2)$ の分布であり, ガウス分布に従う. 図 3 と図 4 の (a), (b), (c) がそれぞれ対応する. 図 4 (a) の $N=5$ では, $\widehat{b}(\tau_2)$ ガウス分布が $\tau_2=0$ を横切っているのがわかる. それにより, 自由エネルギー差 ΔF は負の部分を持ち, 真のモデルが $y=ax+b$ であるのも関わらず, モデル選択の結果は $y=ax$ と間違えている. 図 4 (b) の $N=100$ になると, 自由エネルギー差 ΔF はすべて正の値を持ち, モデル選択の結果は $y=ax+b$ と正しく判断される. 図 4 は, τ_2 を積分消去していない分, 図 3 よりも, モデル選択の結果を定性的に理解しやすくなっている. このため, 図 4 は 3-2 モデル選択のメカニズムの描像との対応が非常に明白で分かりやすい.

$y=ax$ が真のモデルである場合

次に $y=ax$ が真のモデルであり, その係数 $a_0=1.0$, 観測ガウス分布の分散 $\sigma_0^2=1.0$ である場合を考える. 図 5 は, モデル選択に関する自由エネルギー差 ΔF の確率分布であり, 横軸は ΔF で確率分布である. 自由エネルギー差 ΔF の確率分布をもとめるための積分をサンプル数 100,000 サンプルングで近似している. 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 理論曲線である. 図 5 (a), (b), (c) を比較すると, 図 5 (a) の $N=5$ では, 自由エネルギー差の確率分布は正の部分を持ち, 真のモデルが $y=ax$ であるのも関わらず, モデル選択の結果は $y=ax+b$ と間違えている. 5 (b) の $N=100$ や (c) の $N=1000$ になると, 自由エネルギー差 ΔF の確率分布はすべて負のところのみ値を持ち, モデル選択の結果は $y=ax$ と正しく判断されている.

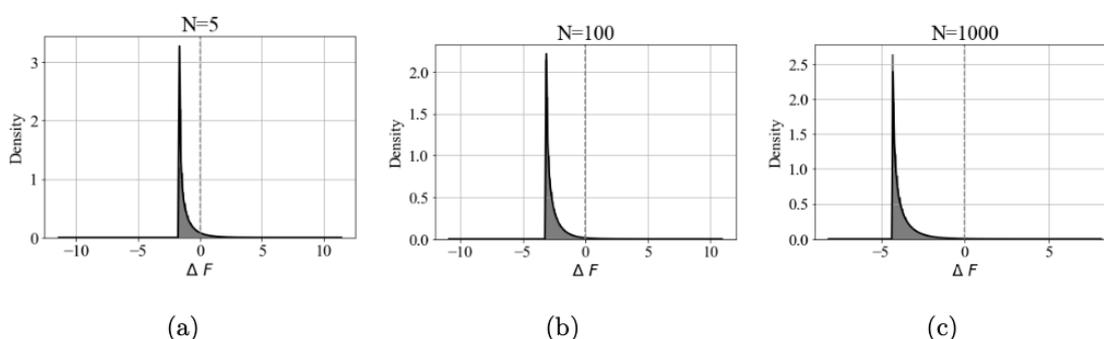


図 5: サンプル数 100,000 でのモデル選択に関する自由エネルギー差 ΔF の確率分布である. $a_0=1.0, \sigma_0^2=1.0$ で, 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 自由エネルギー差の確率分布の理論曲線である.

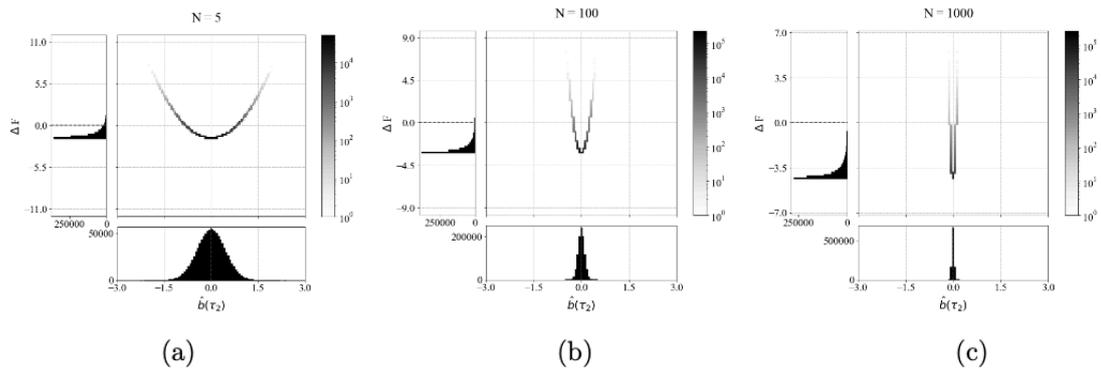


図 6: $a_0=1.0$, $\sigma_0^2=1.0$ であり, 観測データ数 N は(a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 縦軸は, サンプル数 100,000 でのモデル選択に関する自由エネルギー差 ΔF の確率分布の 2 次元グラフである. 横軸は, $\widehat{b}(\tau_2)$ の分布であり, ガウス分布に従う. ここから, 実際の実験計測でも, 計測点数が 100 以上であれば, 正しくモデル選択できることがわかる.

図 6 の縦軸は, サンプル数 100,000 でのモデル選択に関する自由エネルギー差 ΔF の確率分布の 2 次元グラフである. 横軸は, $\widehat{b}(\tau_2)$ の分布であり, ガウス分布に従う. 図 5 と図 6 の(a), (b), (c) がそれぞれ対応する. 図 6(a) の $N=5$ では, の, $\widehat{b}(\tau_2)$ のガウス分布が両端で $\tau_2=0$ を横切っているのがわかる. それにより, 自由エネルギー差 ΔF は正の部分を持ち, 真のモデルが $y=ax$ であるのも関わらず, モデル選択の結果は $y=ax+b$ と間違えている. 図 6(b) の $N=100$ になると, 自由エネルギー差 ΔF はすべて負の値を持ち, モデル選択の結果は $y=ax$ と正しく判断される. 図 6 は, τ_2 を積分消去していない分, 図 5 よりも, モデル選択の結果を定性的に理解しやすくなっている. また, 図 4 は 3-2 モデル選択のメカニズムの描像との対応が非常に明白で分かりやすい.

4.1.2 ベイズ統合

二つのモデルが同一でデータ統合した方が良い場合

ここでは, 3-1 ベイズ統合のメカニズムの描像との対応を取るために, 線形回帰 $y=ax$ について議論する.

まず, 二つのモデルが同一でデータ統合した方が良い場合を議論しよう. 係数の真の値は $a_0=2.0$ であり, 観測ガウスノイズの平均は 0 で分散は $\sigma_0^2 = 1.0$ であるとする.

図 7 は, ベイズ統合に関する自由エネルギー差 ΔF の確率分布のであり, 横軸は ΔF であり, 縦軸はその ΔF での確率分布である. 自由エネルギー差 ΔF の確率分布をもとめるための積分をサンプル数 100,000 サンプルングで近似している. 観測データ数 N は(a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 理論曲線である.

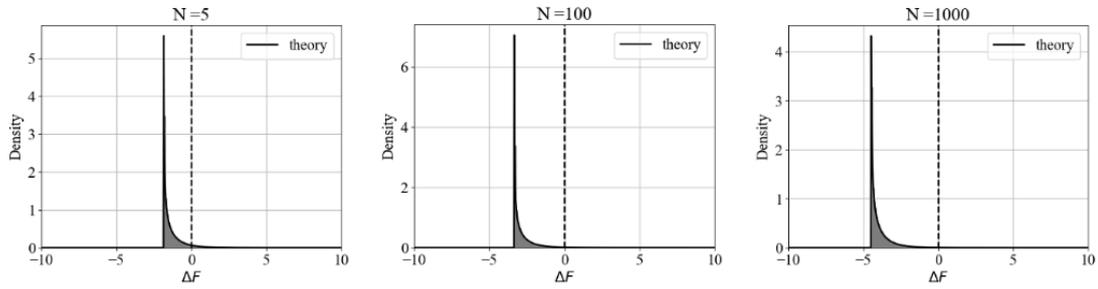


図 7: サンプル数 100,000 でのベイズ統合に関する自由エネルギー差 ΔF の確率分布である。 $a_0=2.0$, $\sigma_0^2 = 1.0$ で、観測データ数 N は(a) $N=5$, (b) $N=100$, (c) $N=1000$ である。黒の実線は、自由エネルギー差の確率分布の理論曲線である。

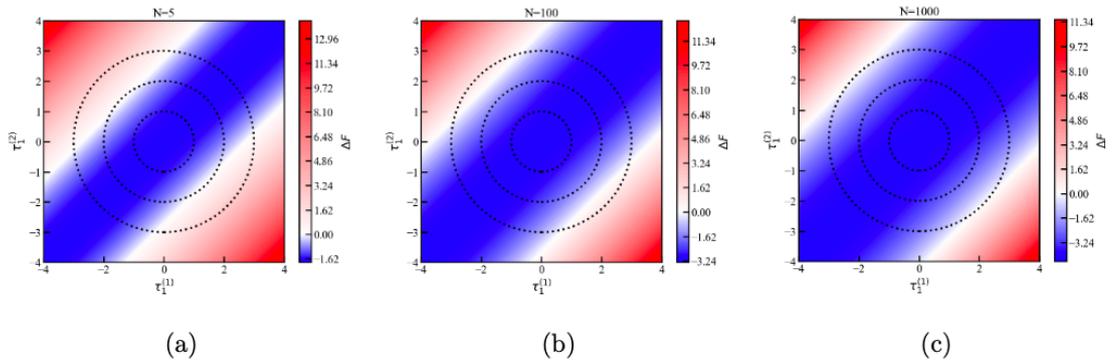


図 8: (a) $N=5$, (b) $N=100$, (c) $N=1000$ に関する、 $a^{(1)}=2.0$ と $a^{(2)}=2.0$ の時の自由エネルギー差 ΔF の τ_1 と τ_2 に対するヒートマップである。中央の半径 1, 2, 3 の三つの円は、 τ_1 と τ_2 の二次元ガウス分布を表す。 τ_1 と τ_2 の半径 1 の円は 39%を表し、径 1 の円は 86%を表し、半径 3 の円は 98%を表す。

図 7 (a), (b), (c)を比較すると、(a)の $N=5$ では、自由エネルギー差 ΔF の確率分布は若干正の部分を持ち、データ統合した方が良い場合であるにも関わらず、ベイズ統合の結果はデータ統合してはいけないと判断する場合が若干ある。図 7 (b) $N=100$ や (c) $N=1000$ になると、自由エネルギー差 ΔF の確率分布はすべて負のところのみ値を持ち、ベイズ統合の結果は、データ統合せよと正しく判断されている。ここから、実際の実験計測でも、計測点数が 100 以上であれば、正しくデータ統合するか否かが判断モデル選択できることがわかる。

図 8 は、ベイズに統合に関する、 τ_1 と τ_2 に依存する自由エネルギー差 ΔF の確率分布を理論的に求めた 2 次元ヒートマップである。図 8 (a), (b), (c)とも、ヒートマップが斜めになっているのは、3-1 ベイズ統合のメカニズムの描像の図 1 の二つのパラメータの事後分布が同時に同じ方向に動いた時に、ベイズ統合の自由エネルギー差 ΔF が変わらないことを表している。図 7 と図 8 の(a), (b), (c)がそれ

ぞれ対応する. 図 8 (a)の $N=5$ では, 中央の青い帯が $N=100$ や $N=1000$ に比べて狭いので, 自由エネルギー差 ΔF は正の部分を持ち, データ統合をするべき場合であるが, しないように間違えて判断する場合が生じる. これが図 7 (a)で, 自由エネルギー差 ΔF が正の確率を若干持つことに対応する. 図 8 (b)の $N=100$ や(c)の $N=1000$ になると, 自由エネルギー差 ΔF はすべて負の値を持ち, データ統合すべきと正しく判断している. 図 8 は, τ_1 と τ_2 を積分消去していない分, 図 7 よりも, モデル選択の結果を定性的に理解しやすくなっている. このため, 図 8 は 3-1 のベイズ統合メカニズムの描像との対応が非常に明白で分かりやすい.

二つのモデルが異なり, データ統合しない方が良い場合

次に, 二つのモデルが異なり, データ統合しない方が良い場合を議論しよう. 二つのデータを生成した線形回帰モデル $y=ax$ の係数の真の値を $a^{(1)}=2.0$ と $a^{(2)}=2.0$ とし, 観測ガウスノイズの平均は 0 で分散は $\sigma^2 = 1.0$ であるとする.

図 9 は, ベイズ統合に関する自由エネルギー差 ΔF の確率分布のであり, 横軸は ΔF であり, 縦軸はその ΔF での確率分布である. 自由エネルギー差 ΔF の確率分布をもとめるための積分をサンプル数 $100,000$ サンプルングで近似している. 観測データ数 N は(a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 理論曲線である. 図 9 (a), (b), (c)を比較すると, (a)の $N=5$ では, 自由エネルギー差 ΔF の確率分布は負の部分を持ち, データ統合しない方が良い場合である関わらず, ベイズ統合の結果はデータ統合すると判断する場合がある. 図 9 (b) $N=100$ や(c) $N=1000$ になると, 自由エネルギー差 ΔF の確率分布はすべて正のところのみ値を持ち, ベイズ統合の結果は, データ統合してはいけないと正しく判断されている. ここから, 実際の実験計測でも, 計測点数が 100 以上であれば, 正しくデータ統合するか否かが判断できることがわかる.

図 10 は, ベイズに統合に関する, τ_1 と τ_2 に依存する自由エネルギー差 ΔF の確率分布を理論的に求めた 2 次元ヒートマップである. 図 10 (a), (b), (c)とも, ヒートマップが斜めになっているのは, 3-1 ベイズ統合のメカニズムの描像の図 2 の二つのパラメータの事後分布が同時に同じ方向に動いた時に, ベイズ統合の自由エネルギー差 ΔF が変わらないことを表している. 図 9 と図 10 の(a), (b), (c)がそれぞれ対応する. 図 10 (a)の $N=5$ では, 青い帯が左上にある. これは, τ_1 が正でと τ_2 が負の場合に, 3-1 ベイズ統合のメカニズムの描像の図 2 の二つのガウス分布が重なり合って, 自由エネルギー差 ΔF が負になり, 同一データで統合すべきだとだと誤って判断してしまうことを表現している. 図 10 (b)の $N=100$ や(c)の $N=1000$ では, 青い帯が図では見えないぐらいに原点より離れるので, 自由エネルギー差 ΔF は常に正の部分を持ち, データ統合をすべきでない判断する.

図 10 は, τ_1 と τ_2 を積分消去していない分, 図 9 よりも, モデル選択の結果を定性的

に

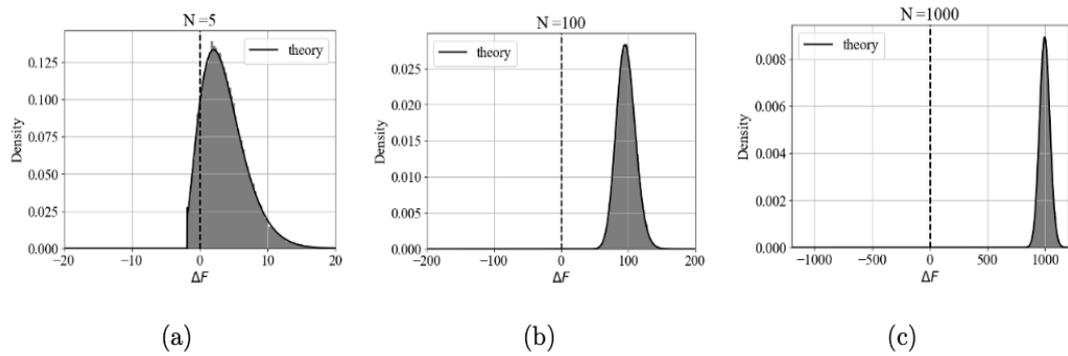


図 9: サンプル数 100,000 でのベイズ統合に関する自由エネルギー差 ΔF の確率分布である. $a^{(1)}=2.0, a^{(2)}=4.0, \sigma_0^2 = 1.0$ で, 観測データ数 N は (a) $N=5$, (b) $N=100$, (c) $N=1000$ である. 黒の実線は, 自由エネルギー差の確率分布の理論曲線である.

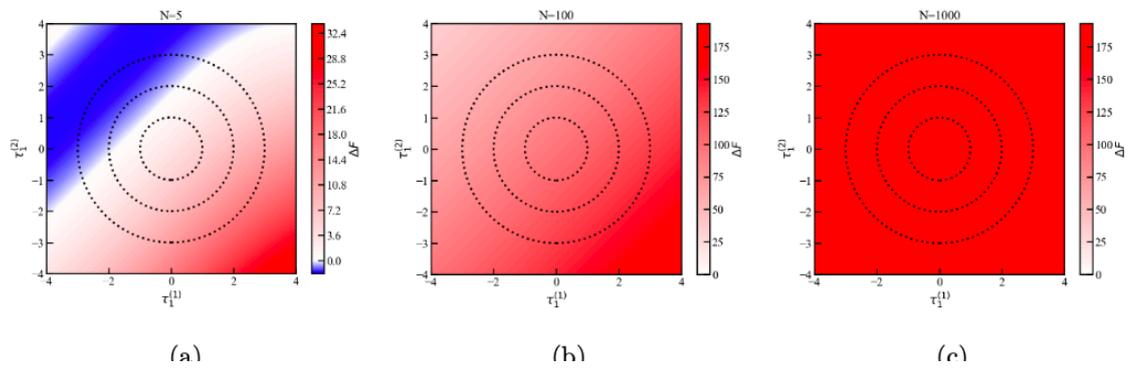


図 10: (a) $N=5$, (b) $N=100$, (c) $N=1000$ に関する, $a^{(1)}=2.0, a^{(2)}=4.0$ の時の自由エネルギー差 ΔF の τ_1 と τ_2 に対するヒートマップである. 中央の半径 1, 2, 3 の三つの円は, τ_1 と τ_2 の二次元ガウス分布を表す. τ_1 と τ_2 の半径 1 の円は 39%を表し, 径 1 の円は 86%を表し, 半径 3 の円は 98%を表す.

理解しやすくなっている. このため, 図 10 は 3-1 のベイズ統合メカニズムの描像との対応が非常に明白で分かりやすい.

5. まとめ

ベイズ計測(Bayesian Measurement)とは、幅広い領域を持つベイズ推論を計測科学に必要な部分だけ取り出したコンパクトな情報科学的枠組みである[2]. ベイズ計測はベイズ計測三種の神器の(1) パラメータの事後確率分布推定[3] (2) モデル選択[3] (3) ベイズ統合[4, 5]から構成される. モデル選択とベイズ統合の両方で重要な役割を演ずるのが自由エネルギーと自由エネルギー差である[3].

そこで私は常にデータ統合すべきかどうかをというベイズ統合にとっては根源的な問題をできれば解析的理論で決着をつけたいと思い、解析計算でベイズ計測を取り扱えるほぼ唯一の例として知られる線形回帰モデル $y=ax+b$ を取り上げた. 線形回帰モデル $y=ax+b$ の理論は解析計算がしやすいためだけに導入する Toy モデルの理論ではない. このモデルは、系の線形応答を取り扱うための実際的なモデルであるとともに、大学理系の標準的リテラシーで取り扱われるものであるために、ベイズ計測の習得のためのロールモデルでもある. このように、計測データ数 N 有限での数値実験の解明等実際的な問題から、線形回帰モデル $y=ax+b$ の理論構築が始まるのは、大変健全なことである.

我々は線形回帰モデル $y=ax+b$ に関して具体化する理論を構築することに成功した[1]. つぎに、計測データ数 N が有限のケースを取り扱うために、 i 番目のデータ x_i と i 番目のデータに重畳された平均 0 分散 σ_0^2 の観測ノイズ n_i の三つの統計量 τ_1 , τ_2 , ν を導入した. τ_1 と τ_2 は平均 0 分散 1 の独立なガウス分布に従う. ν は χ 二乗分布に従う. τ_1 , τ_2 , ν の三統計量で閉じた形で、ベイズ計測三種の神器の (1) パラメータの事後確率分布推定 (2) モデル選択 (3) ベイズ統合が記述できるので、我々はこの理論をメゾスコピック理論と名づけた.

このメゾスコピック理論を用いると、(2) モデル選択 (3) ベイズ統合を記述する際に重要になる自由エネルギー差 ΔF が τ_1 と τ_2 だけで記述することがわかった. この知見を用いると、観測点数 N が有限の場合でも、(2) モデル選択 (3) ベイズ統合の自由エネルギー差 ΔF の確率分布や、自由エネルギー差 ΔF の τ_1 , τ_2 依存性をように解析的に求めることができる.

この理論は、BIC[6]に代表される N 無限大の漸近論に終始してきたベイズ推論の停滞した状況をブレークアウトする画期的な理論体系である. 使用している統計学的なツールは、我々にとっては数理統計学の再発見であったが、既知の数学的事実を巧妙に組み合わせることで、ベイズ推論の新たな理論的枠組みの幕開けに到達した. 我々は、今後この理論をさらに発展させるアイデアを持っている. その結果は、いずれかの場所で公開され、実験/計測にダイレクトに役立つベイズ推論の新たな知見を積み重ねていく所存である.

参考文献

- [1] Katakami, Kashiwamura, Nagata, Mizumaki and Okada, “Mesoscopic Bayesian Inference by Solvable Models”
<https://arxiv.org/abs/2406.02869>
- [2] Nagata, Muraoka, Mototake, Sasaki and Masato Okada, “Bayesian Spectral Deconvolution Based on Poisson Distribution: Bayesian Measurement and Virtual Measurement Analytics (VMA)”, *Journal of the Physical Society of Japan*, 88(4) 044003 - 044003 (2019)
- [3] Nagata, Sugita and Okada, “Bayesian spectral deconvolution with the exchange Monte Carlo method”, *Neural Networks*, **28**, 82-89 (2012)
- [4] Yokoyama, Uozumi, Nagata, Okada, and Mizumaki, “Bayesian Integration for Hamiltonian Parameters of X-ray Photoemission and Absorption Spectroscopy”, *Journal of the Physical Society of Japan*, **90**, 034703, (2021)
- [5] Nishimura, Katakami, Nagata, Mizumaki and Okada, "Bayesian Integration for Hamiltonian Parameters of Crystal Field", *Journal of the Physical Society of Japan* **93**, 034003, (2024)
- [6] Schwarz, “Estimating the dimension of a model”, *Ann. Stat.*, **6**, 461, (1978)

謝辞

[1]の理論を共同で研究して下さった東京大学大学院理学系研究科物理学専攻博士課程2年柏村周平氏, 国立研究開発法人物質・材料研究機構(NIMS)主任研究員永田賢二博士, 熊本大学理学部物理学科水牧仁一朗教授に深く感謝する。