

構造解析におけるベイズ推定の応用

井田 隆

名古屋工業大学先進セラミックス研究センター
〒507-0071 岐阜県多治見市旭ヶ丘 10-6-29

Application of Bayesian Inference to Structure Analysis

Takashi Ida

Advanced Ceramics Research Center, Nagoya Institute of Technology
10-6-29, Asahigaoka, Tajimi, Gifu 507-0071, JAPAN

This article describes basic concepts of statistical estimation based on experimental data, including (i) Bayesian inference, (ii) maximum *a posteriori* estimation, (iii) maximum likelihood estimation, and (iv) least-squares estimation. It is emphasized that Bayesian inference is nothing but calculation of “conditional probability”, which is written in textbooks of high-school mathematics in Japan. Bayesian approach allows any prejudice, biased ideas, preconception, and never forces us to get rid of them. The basic concept of Bayesian inference explicitly assumes that we can never reach the absolute truth, but we can always approach to the truth at the same time. A simple implementation of Bayesian inference in crystal structure analysis based on X-ray diffraction measurements is proposed.

1. はじめに

研究室では、実験データに基づいて、何らかの物理量や幾何学的な対称性を推定する作業が頻繁に行われる。X線回折実験から得られるデータに基づいて結晶構造を推定しようとする作業は、その典型的な例の一つである。化学組成と温度・圧力が確定すれば、エントロピーも含めた自由エネルギーの観点から「最も熱力学的に安定な結晶構造」は一意に決まるはずだから、X線回折データに基づいて結晶構造を推定することは、まず第一に、この「正しい結晶構造」を求めようとすることであると考えてもよいだろう。

回折実験に基づく「空間群決定」では、伝統的に消滅則というルールが用いられる。運動学的な回折理論 kinematical theory of diffraction に基づけば、結晶構造が特定の対称性を持つ場合に、特定の「偶奇・ゼロ」パターンの指数を付けられた回折斑点が消失しているように見えることになるので、これから候補となる構造の対称性を絞り込める。しかし、「対称性から消滅する」ということと、「観測できなかった」ということは、本来区別することができないはずであることを注意しておこう。

特に粉末試料を用いた回折実験では、単結晶試料を用いた丁寧な回折実験で観測される「弱い回折」を明瞭なピークとしては検出できない場合も少なくなく、この見落としが深刻な誤認とみなされる可能性もある。

純物質の熱力学的に安定な結晶構造を推定することに限

れば、単結晶試料を用いたX線回折実験は、事実上決定的な方法でありつづけてきたと言える。一方で、不純物や構造欠陥を含み、通常は多相混合物として応用される実用材料の直接的な評価が望まれる社会的な要請から、粉末X線回折法による構造推定の重要性も高くなっている。粉末X線回折測定は、単結晶試料を用いた測定に比べると実験的に得られる情報が少ない。もし他の実験（特に電子顕微鏡観察、電子回折実験、X線吸収分光測定など）や、第一原理的、経験的な理論計算の結果を有効に利用すれば、さらに粉末X線回折実験を活用することができるであろう。また、粉末X線回折実験に限っても、条件の異なる測定結果を繰り返し、総合的に構造推定の妥当性を判断できることが望ましい。このために、「ベイズ推定」Bayesian inference と呼ばれる方法を応用しようと考えられる。

その一方で、単結晶試料を用いるか粉末試料を用いるかに関わらず、回折実験の結果に基づく結晶構造推定には長い間「最小二乗推定」という方法が用いられてきた。筆者は最近、確率論的な推定法としては最小二乗法の上位概念にあたる「最尤推定（さいゆうすいてい）」maximum likelihood estimation (ML 推定) という手法を適用することにより、粉末X線回折データから、最小二乗法に基づくリートベルト法 Rietveld method と呼ばれる方法 [1] より、もっともらしい構造を推定できる例が少なくないことを示した [2-4]。しかし、ML 推定は最大事後確率推定 maximum *a posteriori* estimation (MAP 推定) の下位概

念にあたり、MAP 推定はベイズ推定の下位概念にあたると通常考えられている。筆者が本稿で主に取り上げようとする「ベイズ推定」に比べると、筆者が実際に演示した ML 推定は、まだ「下位」の「下位」の概念でしかない。リートベルト法で用いられる最小二乗法は、さらにその下位の概念であると言える。しかし、そうだとすると、性急に最小二乗法という方法を軽視する立場は取らない方が良いだろう。この方法は、200 年以上の長い間使われてきたのにも関わらず、その意味が誤解されがちである。むしろベイズ推定と呼ばれる推定法の考え方と関係付けることによって、論理の基盤が明確になり、最小二乗法という方法をどのように位置付けるべきかが、より明確になるのではないかと思う。

そこで、本稿では、はじめにベイズ推定、MAP 推定、ML 推定、最小二乗推定という確率論的な推定の体系について概説する。さらに実験データに基づいて物質の構造を推定する目的で、ベイズ推定と呼ばれる方法がどのように応用しようと考えられるかについての一端を述べることにする。

2. 確率論的な推定

2.1 ベイズ推定

ベイズ推定という考え方がいつ成立したのかは、必ずしも明確でない。おそらく今でも否定的に思われることが少なくないので、ある意味では、まだ成立していないのかもしれない。しかし、この方法の中身は高校の数学で習う「条件付き確率」を導くことに過ぎないので、ベイズ推定が理解できないなら、高校の数学から勉強し直すと良い。ただし、「条件付き確率」は、高校数学の中でも意味が理解しにくいと思われがちな項目のうちの一つであることは確かである。

ベイズ推定の考え方を端的に表す事例としては、法廷で資料として用いられる DNA 鑑定があげられる。DNA 鑑定では、例えば「犯行現場から採取された毛髪が容疑者のものである確率が 99.99999% 以上である」という結果が示される。この毛髪は容疑者のものであるか、そうでないかのどちらかに決まっているのに、そのことを 1 でも 0 でもない中途半端な確率の数値で表現するのは奇妙なことに思われるかもしれない。また、もちろん容疑者の毛髪が犯行現場に遺留されていたとしても、その容疑者が犯人であるとは限らないのだが、それは別として、どのように高精度な科学捜査に基づいたとしても、犯人を誤認して冤罪を着せてしまうことが、例えば 0.00001% 以下の確率だとしても、ゼロではない確率で生じうることも暗示されている。

したがって、「冤罪は決して許されない」「疑わしきは罰せず」としたら、どのような処罰も正当化することができない。「冤罪の可能性があったとしても、社会秩序を維持するために公的な機関による制裁を加えることが有効である」という立場を取れば論理的には整合するが、この立場は冷酷で非情に見え、感情的に受け入れがたく思われるだろう。しかし歴史的に見れば、社会秩序を形成・維持するために、処罰をとまなう法制度が有効であったことも間違いないと考えら

れる。したがって、公的な表明としては許されにくいとしても、「冤罪はありうる」ので「冤罪をなるべく少なくするための努力をしよう」と思っても良いだろう。

たかが「条件付き確率」のことで大げさのようだが、ベイズ推定が理解しがたいと思われることの根の一つは、このように「心情的に許しがたい」という自然な心の動き方にあると推測される。

さて、ベイズ推定は、実験結果 experimental results (X) に基づいて、仮説 hypothesis (H) の妥当性を確率的に評価するものとみなすことができる。日本の高校数学の記法では、この条件付き確率を $P_X(H)$ と表す。

ベイズ推定の考え方を、回折実験に基づく構造推定に適用するとしたら、「実際に回折実験をした結果 X に基づいて仮想構造モデル H の正しさを確率論的に評価する」という意味のことをすることになる。ベイズ推定では、実験結果 X の出現する確率を $P(X)$ とすれば、

$$P(H)P_H(X)=P(X)P_X(H) \quad (1)$$

の関係 (ベイズの定理) があるはずだから、

$$P_X(H)=P(H)P_H(X)/P(X) \quad (2)$$

という式によって $P_X(H)$ を求められるという考え方をする。

ここで $P_H(X)$ は、「仮想構造モデル H に基づけばどのような実験結果 X が予想されるか」を意味し、普通の意味で、教科書に書いてあるような「理論」と呼ばれるものと大きく違わない。ところが、 $P_H(X)$ と $P_X(H)$ とでは、因果の関係が逆転しているように見えるので、このことが混乱の一因になるかもしれない。そこで、以下のように思えば良いと言おう。原因 H のせいで結果 X があるはずなのは変わらない。 $P_X(H)$ という表現は、結果 X から原因 H を推理しようとしているだけのことである。

ベイズ推定が応用されるとき、実際には上述の関係だけではなく、さらに、

$$P(X)=\sum_H P(H)P_H(X) \quad (3)$$

の関係も用いられる。このことに関しても、違和感が持たれるのは自然であり当然であろう。この式 (3) が意味することは、「想定したこと H とその理論的な帰結 $P_H(X)$ しか考えないことにする」という宣言でもあるからである。しかし、このような態度をとっても許されるのであれば、確かに $P(H)$ を決めれば、式 (2)、(3) から、実験結果 X に基づいて仮想構造モデル H の正しさを確率論的に評価できることになる。

仮想構造モデル H には複数の対称性の異なる結晶構造モデル A, B, C, \dots が候補として含まれていても良いし、個々の結晶構造モデルは連続変数で表されるような未知のパラメータを含んでも良い。ベイズ推定に基づけば、結晶構造モデル A, B, C, \dots のそれぞれが正しい構造である確率を計算できるし、個々の結晶構造モデルが含むパラメータを $\alpha_1, \alpha_2, \dots; \beta_1, \beta_2, \dots; \gamma_1, \gamma_2, \dots$ として、 $(\chi_j)_{\min} < \chi_j < (\chi_j)_{\max}$ ($\chi = \alpha, \beta, \gamma, \dots; j = 1, 2, \dots$) という関係を満たす確率が、任意の $(\chi_j)_{\min}, (\chi_j)_{\max}$ の設定に対して求められ

るはずである。このことは、例えば 30 年前には先端的な研究に従事する科学者でも「事実上不可能」と思っていたことだったかもしれないが、現在はかなりの水準で実現するし、近い将来には「実現できるのが当然」ということになると思われる。

ベイズ推定の文脈の中で用いられる確率 $P(H)$ は、実際に実験をする前に予想された確率だという意味で事前確率 *prior probability* と呼ばれ、条件付き確率 $P_X(H)$ は実験結果あるいは経験 X に基づいて事前の予想を修正した確率の値を意味するから、事後確率 *posterior probability* と呼ばれる。ベイズ推定の考え方では、事前確率 $P(H)$ は任意である。この意味で「主観的な確率」とも呼ばれる。極端な言い方をすれば、先入観でも良いし、偏見でも良い。これが、ベイズ推定の考え方の中で、最も奇妙で理解したいポイントではないかと思われる。

実証主義・科学的な考え方では「先入観を持つてはいけない」「客観的に判断するべきだ」と思われそうなのに対して、ベイズ推定のこのような論理は、それを真つ向から否定しているように聞こえるかもしれない。しかし、この考え方は、「はじめにどのような偏見を持っていたとしても、実験や経験、学習を積み重ねることによって、誤った偏見は改められて、正しい考えに近づいていくことが期待される」という意味を持つ。そして、経験や学習を繰り返して真実に近づこうとするこのような振る舞いを「ベイズ改訂」*Bayesian update* と呼ぶ。

二つの対立する仮説 A, B があつたとして、どちらが正しいかまったくわからなかったとしたら、これを数値で表すと、 $P(H=A) = 0.5, P(H=B) = 0.5$ である。このことを前提として実験 X_1 を行い、事後確率として $P_{X_1}(H=A) = 0.7, P_{X_1}(H=B) = 0.3$ を得たとする。さらに実験 X_1 とは独立とみなせる実験 X_2 を行って、「実験 X_1 の結果」と「実験 X_2 の結果」を両方とも考慮に入れた推定結果が、例えば $P_{(X_1 \text{ and } X_2)}(H=A) = 0.9, P_{(X_1 \text{ and } X_2)}(H=B) = 0.1$ になったとする。

これは事前確率として $P(H=A) = 0.7, P(H=B) = 0.3$ を仮定して、実験 X_2 の結果に基づく事後確率が $P_{X_2}(H=A) = 0.9, P_{X_2}(H=B) = 0.1$ になったということと同じである。つまり、実験 X_2 の解析を行う前に、「 $P(H=A) = 0.7, P(H=B) = 0.3$ 」という先入観をもつということに等しい。

先入観と言うと悪く聞こえるだろうが、それは、多くの場合、それまでの知識や経験に基づく。

このような問題への処し方が許されれば、例えば粉末 X 線回折実験だけで決定的な解答を導くことができなかつたとしても、その実験は決して無意味なことではなく、なんらかの意味を持たせることが可能である。

ベイズ推定が構造解析に正しく適用された場合、それは、「唯一の正しい構造」を求めるものではないし、「一番もっともらしい構造」を求めるだけのものでもない。「一見

もっともらしく見える構造が、どの程度信用できるか」「how much can it be believed」ということを求めようとしているのである。

2.2 最大事後確率推定

個人的なレベルでベイズ推定の考え方に少しは好感が持てたとしても、なんらかの形で「意志」を決定すること *decision* を要求される場合がある。「結局のところ正しいことは完全にはわかりません」ということが、ベイズ推定の根本にある真実だとしても、社会的な生活を送る限り、「そうだと、どうするつもりなのか？」と問われることになるだろう。

なんらかの意志決定を要求された場合に使える（あるいは使わざるを得なくなる）方法の一つが、最大事後確率推定 *maximum a posteriori estimation* であるとも言える。これは、前節で示したベイズ推定の考え方の中に含まれる事後確率 *posterior probability* を最大にする解を選択することであり、このように決めてしまえば $P_X(H)$ を最大にするように仮説 H を調整し終われば、それで無限の連鎖から脱出できることになる。

回折実験が物質の構造を推定するために決定的な方法であるのだとしても、MAP 推定を用いることにすれば、最終的な結論を導く前に、例えば元素分析や部分構造の情報、電子回折像、その他のいろいろなことを「どの程度信頼するか」考慮しながら取り入れることができるはずである。MAP 推定では先入観を持つことが許容されるので、それまでの知識や経験を活用できるということはベイズ推定と同じことである。その一方で MAP 推定を用いた時点で、その結果はもう改善の余地がなくなってしまうから、この点ではベイズ推定とまったく異なる性格も持っている。

なんらかの意思決定が要求されたとして、本当に「事後確率最大の解」だけを選抜する MAP 推定が必要なのかということは、よく考え直して見た方が良いでしょう。このことについては 3 節でももう少し詳しく検討することにする。

また、もし MAP 推定を使うと決めるなら、その前に「どうして確率を最大にする解（最頻値、モード）を選択しようとするのか？確率を最大にする解より、偏りの少ないと思われる解（不偏推定解）を選択したほうが良いかもしれないし、ロバスト解（「異常値」の影響を受けにくい解）を選択したり、ミニマックス解（もっとも深刻な被害を想定して、それを最小限にする解）を選択したほうが良いかもしれないのではないか？」という問いに対する回答を準備しておいた方が良いでしょう。

2.3 最尤推定

最尤推定 *maximum likelihood (ML) estimation* という方法は、20 世紀初頭に英国の生物統計学者である R. Fisher が提案したとされるが [5]、ここではベイズ推定、MAP 推定の文脈の中で位置付けることを試みる。

ベイズ推定の中の理論 $P_H(X)$ が H を明示的に含む数

式で表現されていれば、 $P_H(X)$ を H の関数とみなして、この関数を最大にするように H を調整するだけで良からうというのが ML 推定の考え方と思っても良い。MAP 推定では $P_X(H)$ を最大にするように H を調整しなければいけなかったのだが、ML 推定では、 $P_H(X)$ を最大にするよう H を調整すれば良いだけなのである。そう思えば、MAP 推定の含んでいた紛らわしさがなくなり、特に、初心者にとっては親しみやすくなるだろう。

MAP 推定ではなく、ML 推定を用いることにすれば、事前確率 $P(H)$ はもはや不要である。MAP 推定で最大化する $P_X(H)$ を計算するためには、理論 $P_H(X)$ だけでなく知識や経験、あるいは主観や先入観、偏見のような意味を持つ事前確率 $P(H)$ が必要であったことを、もう一度、式 (2), (3) で確認しよう。これに対して、ML 推定では、知識や経験を持たない「いたいけな初心者」であっても、熟練者に対して何の気兼ねも恐れの気持ちも抱く必要がない。

ただし、 $P(H)$ が不要になるということを別の言い方で表現すれば、「 $P(H)$ として常に一様分布を仮定する」ことに等しい。

単純化して、 $H=A, B$ の二つの対立する仮説があるとしよう。事前分布が一様ということは、 $P(H=A) = P(H=B) = 0.5$ であるということと同じである。このとき、式 (2) から、

$$P_X(H=A) = 0.5 P_{H=A}(X) / P(X) \quad (4)$$

$$P_X(H=B) = 0.5 P_{H=B}(X) / P(X) \quad (5)$$

となるから、 $P_X(H)$ を最大にする H は、 $P_H(X)$ を最大にする H と同じことになる。

つまり ML 推定では、選択の余地なく特定の先入観（一様事前分布）が強制され、過去の経験がまったく生かされない。ML 推定は、初心者には向いているが、熟練者にとっては実は厳しい面がある。

さて、理論 $P_H(X)$ を H の関数とみなすということを明示するために、 $L(H) = P_H(X)$ と書くことにする。このとき、関数 $L(H)$ は、仮説 H の「もっともらしさ」を表し、尤度（ゆうど）関数 likelihood estimator と呼ばれる。「最尤推定」とは、尤度関数を最大化する推定法であるとも言える。

ここで、前節の最後に示した「問い」に対する回答のヒントになりそうなことについて述べる。

実験データ、あるいは観測されたデータが $\mathbf{y} \equiv \{y_1, y_2, \dots, y_N\}$ という複数の数値からなり、それら複数の数値に基づいて一つの代表値 z を決める問題を考える。複数の数値の組と仮想的な一つの数値 z とから、ノルム norm と呼ばれる値が³、

$$L_v = \|\mathbf{y} - \mathbf{z}\|_v \equiv \left(\sum_{j=1}^N |y_j - z_j|^v \right)^{1/v} \quad (5)$$

という式で定義される。ただし、ここでは、

$$z_1 = z_2 = \dots = z_N = z \quad (6)$$

とする。このノルムという値は、なんらかの文脈で「代表値

z が観測値 $\mathbf{y} \equiv \{y_1, y_2, \dots, y_N\}$ からどのように隔たっているか」を意味する。

特に $v=1$ として計算される値 L_1 のことは、最近まで「マンハッタン・ノルム」とか「イエローキャブ・ノルム」と呼ばれた。米国 New York 市の中心部であるマンハッタン Manhattan 島の中には、京都や札幌と同じように、道路網が縦横に張り巡らされている。黄色に塗られたタクシー (yellow cab) で移動するときの「移動距離」あるいは「料金」は、このようなノルムで評価される。最近と同じことが「タクシーキャブ taxi cab・ノルム」とも呼ばれたり、もっと直接的に L_1 (エル・ワンあるいはエル・サブ・ワン) ノルムとも呼ばれる。この L_1 ノルムは、指紋認証などの画像解析にも使われる。中央値 (メジアン) を選ぶことは、 L_1 ノルムを最小化する典型的な例の一つである。

一般的に、 $v < 2$ の L_v ノルムを最小化することをロバスト (robust; 頑丈なという意味) 推定と呼ぶ。こうすると、数値データの中に、突拍子もなくはずれた数値、明らかに異常な値が何かの間違いで紛れ込んでしまったとしても、 L_2 ノルムに比べて、その影響を受けにくくなるからである。

最も頻繁に使われる $v=2$ のノルムのことはユークリッド Euclid・ノルムと呼ばれる。もちろん L_2 (エル・ツーあるいはエル・サブ・ツー) ノルムと呼んでもよい。これは中学校の数学で習う普通の意味での「距離」を拡張したものであると考えてよい。後述する最小二乗法とは、ユークリッド・ノルムを最小化する推定法と言える。最も単純な例が平均をとることである。

$v \rightarrow \infty$ の極限でのノルムを最小化する最適化法のことをミニマックス mini-max 法と呼ぶ。危機管理 risk management の目的でも使われるが、理系の分野であれば、コンピューターの「数値計算ライブラリ」を使って関数の近似計算をするとき、ほとんどのライブラリでは最小二乗近似解ではなく、ミニマックス近似解が用いられていること [6] を知っておいても良いだろう。

さて、これらのことを踏まえて、次の式で定義される関数 $\rho_v(x)$:

$$\rho_v(x) = \frac{\text{vexp}(-|x|^v)}{2\Gamma(1/v)} \quad (7)$$

$$\Gamma(\alpha) \equiv \int_0^\infty t^{\alpha-1} e^{-t} dt \quad (8)$$

を考えてみよう。この関数は、

$$\int_{-\infty}^\infty \rho_v(x) dx = 1 \quad (9)$$

という関係を満たすから、確率変数を x とする確率密度関数とみなすことができる。そして、測定や調査などから、互いに独立とみなせる数値の組 $\{y_1, y_2, \dots, y_N\}$ が得られたとする。この実験データの出現確率がこの仮想的な確率密度関数でモデル化することができれば、このデータセットがまると同じように出現する確率は、代表値を z とすると、

$$L_\nu(y_1, \dots, y_N; z) = \prod_{j=1}^N \rho_\nu(y_j - z) \\ = \exp \left[- \sum_{j=1}^N |y_j - z|^\nu + N \ln \frac{\nu}{2\Gamma(1/\nu)} \right] \quad (10)$$

に比例するはずである。この式は、確率に比例する値を表現していると考えても良いし、何かの尤もらしさを表現するものと考えて尤度関数 likelihood estimator と呼んでも良い。そして、尤度の対数を対数尤度 logarithmic likelihood と呼ぶ。式 (10) に対応する対数尤度は次式で表される。

$$\ln L_\nu(y_1, \dots, y_N; z) = - \sum_{j=1}^N |y_j - z|^\nu + N \ln \frac{\nu}{2\Gamma(1/\nu)} \quad (11)$$

この対数尤度を最大化することは、次数を固定すればノルムを最小化することと同じであり、ロバストでも最小二乗、ミニマックスでも好きなものを選択できる形になる。

さらに、ML 推定では、実験データに基づいて、この尤度関数モデルが含む次数 ν の値まで形式的には最適化する。ただし、このような操作はやや作為的にすぎるかもしれない。本当にこのような処置に意味があるかは、注意深く検討するべきであろう。

前節の最後に示した問いにどのように回答するかは、このようなことを含めて考えてみると良いだろう。

2.4 最小二乗推定

最小二乗法は、18 世紀にドイツの数学・物理学・天文学者として知られる C. F. Gauss がまだ十代のときに考案し、準惑星ケレスの軌道を計算するのに用いたのが発端と言われる [5]。それ以来、データ解析の最も基本的な手法として、自然科学や工学を中心とした広い分野で中心的な役割を担い続けてきた。ただし、小学校 5 年の算数で習う「平均をとること」は、最小二乗法の一つである。このことについては後で述べよう。

最小二乗法は、実験データの確率分布として正規分布を仮定した最尤推定法であるとも言える。前節の式 (7) に $\nu = 2$ を代入した形式が、正規分布に相当する。したがって、正規分布を仮定した場合の対数尤度は

$$\ln L_2(y_1, \dots, y_N; z) = - \sum_{j=1}^N |y_j - z|^2 - \frac{N}{2} \ln \pi \quad (12)$$

のように表される。この値を最大にするということは、「残差の二乗和」を最小にすることと同じである。そのような解を求めることが最小二乗法と呼ばれるものに他ならない。

このような背景があるので、最小二乗法は、以下の条件が満たされる場合にのみ (最尤推定として) 正当化されると思われるがちのようである。

- (i) データが含む数値は互いに独立で、相関が無視しうる
- (ii) 測定値の統計的な誤差あるいは変動の大きさが既知

とみなせること

- (iii) 測定値の真値からのずれ (偏差) の分布が、真値を中心とした正規分布 (Gauss 分布) に従うこと

しかし、このうち (i) と (iii) は必ずしも重要なことではない。測定値に相関があったとしても、共分散行列が求まっていれば、この逆行列を重み行列として計算に取り入れるか、あらかじめ行列を対角化してしまえば良いだけである [5]。また、(iii) の要請は一見もつともらしく聞こえるかもしれないが、話を単純化しすぎである。そもそも最尤推定であることが常に重要とは限らない。このことを以下に示そう。

例えば、ポアソン Poisson 分布と呼ばれる確率分布に従うデータに対して最小二乗法を適用することを考えてみよう。光の強さを、光子計数 photon counting 法で測定する場合に、測定値はポアソン分布に従うとみなせる場合が多い。一定の計数時間 t (s) の間に、平均計数率 ν (s^{-1}) の光子を測定するとする。何回も測定を繰り返して平均をとれば、その平均値は測定を繰り返すごとに νt という値に近づくはずである。そして、以下のように、観測値の平均をとることは、最小二乗法に他ならない。

測定を N 回繰り返して得られた観測値 (カウント数) が $\{n_1, n_2, \dots, n_N\}$ だったとする。ここでカウント数に関する推定値を \bar{n} としよう。すると、残差の二乗和は

$$S = (n_1 - \bar{n})^2 + (n_2 - \bar{n})^2 + \dots + (n_N - \bar{n})^2 \quad (13)$$

となる。この S の値を最小にする解は、

$$\bar{n} = \frac{n_1 + n_2 + \dots + n_N}{N} \quad (14)$$

である。

ポアソン分布では、個々の測定値 (カウント数) n の出現する確率は、

$$P(n) = \frac{(\nu t)^n}{n!} e^{-\nu t} \quad (15)$$

と表されるから、平均値と最頻値 (最尤推定値) は常に異なったものになる。たとえば $\nu t = 0.1$ のとき、1 回の測定では 0 カウントとなる確率が 90 % 以上である。測定を繰り返して複数の数値からなるデータに対して最尤推定法をほどこせば、最頻値は平均値に近くなることを期待できるが、有限の試行回数の場合、最頻値は常に平均値より小さい値になるはずである。

それに対して平均値の期待値は常に 0.1 であり、例えば測定を 100 回繰り返して観測された値の平均をとれば、十分に 0.1 に近い値が得られるだろう。ところが、この場合、平均値は最小二乗解であるが、最尤推定解とは言えない。この「平均をとる」という操作は、「正規分布とはまったく異なるポアソン分布」に従っているはずのデータに対して、無理やり正規分布のモデルを仮定した最尤推定を適用してしまったのと同じことだからである。

しかし、ここでもう一度よく考えてみよう。我々が知っていた $\nu t = 0.1$ という値が、最小二乗法 (平均をとること)

で得られるのだとしたら、「この方法が最尤推定になっていないから正当化できない」と言うのは、いかにも了見の狭いことのように思われまいだろうか？

何らかの測定値が統計的に変動し、その「真の平均」が y であるとする。そして、 N 回の繰り返し測定で $\{y_1, y_2, \dots, y_N\}$ という結果が得られたとする。ここで、

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_N}{N} \quad (15)$$

を計算しよう。この \bar{y} の値は「標本平均」と呼ばれる。測定値（標本）が統計的に変動するのだから、この標本平均も「真の平均」 y とは異なる値になる。しかし、その一方で、この「標本平均の期待値」は、 N の値に関わらず、常に「真の平均」に一致する。つまり、このようにして計算される標本平均は、期待値がどちらかに偏ることがないという意味で「不偏推定量」と呼ばれる。これは、測定値の統計的な変動が、どのような確率分布であるかとはまったく関係無い。つまり、ある意味で最小二乗法は「不偏推定」の性格を持っていると言える。

「正規分布を仮定した最尤推定法は最小二乗法になる」ということは正しい。しかし「正規分布でない場合でも最小二乗法は有効である」というのが筆者の見解である。

3. ベイズ推定へのアプローチ

ここで、ベイズ推定を粉末X線回折測定に基づく構造解析に応用する具体的な方法について考えてみよう。

考え方の基本は、複数の構造モデルをたてて、唯一無二の正解を導こうとはしないことである。ここでは単純に A という構造モデルと B という構造モデルを候補として立ててみよう。そして、粉末X線回折測定実験を実施する。この結果が X_1 となったとする。 $P_{H=A}(X_1)$ と $P_{H=B}(X_1)$ の定式化は、それほど困難ではない。実際に行う計算は、最小二乗法に基づくリートベルト解析の計算の一部とほとんど変わらないはずである。より具体的には、筆者が ML 推定に用いた形式 [2-4] が概ねそのまま使える。ML 推定とは $P_H(X)$ を最大化する H を求めることであるが、このことは、既に実現できているのだということを強調しておきたい。

さて、はじめの段階では先入観を持たないつもりになって（あるいは一様分布を仮定する先入観を持つことにして） $P(H=A) = 0.5, P(H=B) = 0.5$ としよう。そして、以下の式に従った計算をしてみる。

$$P_{X_1}(H=A) = P(H=A) P_{H=A}(X_1) / P(X_1) \quad (16)$$

$$P_{X_1}(H=B) = P(H=B) P_{H=B}(X_1) / P(X_1) \quad (17)$$

$$P(X_1) = P(H=A) P_{H=A}(X_1) + P(H=B) P_{H=B}(X_1) \quad (18)$$

ここで $P_{X_1}(H=A)$ は「A が正しい確率」、 $P_{X_1}(H=B)$ は「B が正しい確率」として勝手に思い込んでいた「どちらも 0.5」という先入観が、実験結果 X_1 に基づいて修正された値になっているはずである。そして、A と B のどちらか一つを選ぶのではなく、両方とも（有力か泡沫かの違いが歴然とし

ていても）候補として残す。

次の実験結果を解析する前に、 $P(H=A) = P_{X_1}(H=A)$ と $P(H=B) = P_{X_1}(H=B)$ という先入観を持ちなおすことにする。そして次の実験結果 X_2 に基づいて、その先入観をもう一度修正しよう。あとは満足するまで（あるいは嫌になるまで）これを繰り返す。「どちらの構造モデルが正しいか決めよ」という強い要求をされたら MAP 推定（A と B のどちらかを選ぶ）を実施しても良いと思うが、むしろ MAP 推定は実施せず、A, B の両候補を残したままで良い場合が多いと思う。

おそらく事後確率の最大になる構造モデルが、正しい構造モデルなのであろう。かりに構造モデル A が正しい構造で、解析の結果も $P_X(H=A) > P_X(H=B)$ になったとする。しかし「一番もつともらしい構造は A である」とだけしか言われたいのと比べたら、例えば「一番有力な構造 A の正しい確率が 99.97 %、二番目に有力な構造 B の正しい確率が 0.03 %」のように言ってもらえる方が、格段に説得力があると筆者は思う。

4. おわりに

この記事では、何かと誤解されやすい「確率論的な推定」について解説することを試みた。

ベイズ推定法は、「我々は決して絶対的な真実に到達することができない」ということ、それと同時に、「いつでも真実に近づく努力をすることができる」ということを明示的に仮定する。そして、得られる結果は「どの仮説が正しいのか」ではなく「それぞれの仮説がどの程度信頼できるのか」である。ベイズ推定は、どのような偏見や先入観も許容し、その先入観を捨て去ることも要求しない。そのかわり、経験を積み重ねることによって、誤った先入観が修正されて、より正しい考え方に近づくことを期待する。ここにあるのは何か「希望」のようなものに近い。

このように扇情的な表現をすることは本来慎むべきだろうが、ベイズ推定の基本は高校数学の教科書に書いてあることなのに、理解しにくく、心理的に受け入れ難いという面が確実にあると思われるので、あえてこのような表現をとるものと理解されたい。本稿では、一例として構造解析にベイズ推定を応用する考え方の一端を示したのみであるが、この方法はさらに広く展開・応用することが可能であろう。

参考文献

- [1] H. M. Rietveld, *J. Appl. Cryst.* **2**, 65-71(1969).
- [2] T. Ida & F. Izumi, *J. Appl. Cryst.* **44**, 921-927(2011).
- [3] T. Ida & F. Izumi, *Powder Diffr.* **28**, 124-126(2013).
- [4] T. Ida, *Solid State Phenomena*, **203-204**, 3-8(2013).
- [5] 中川徹・小柳義夫「最小二乗法による実験データ解析」東京大学出版会(1982).
- [6] W. H. Press, S. A. Teukolsky, W. T. Vetterling & B. P. Flannery, “*Numerical Recipes 3rd ed.*,” Chapter 5, Cambridge University Press(2007).