核データニュース，No.41 (1992)

解　　説（I）

ピールのパズルについて（1）

（日本原子力研究所）千葉　　敏

1. 初めに

ピールのパズル（Peelle's Pertinent Puzzle、以下 PPP と略す）は、1987 年に ORNL の Robert Peelle 氏が数人の親しい友人に宛てた手紙[1]の中で尋ねている質問に端を発している。これは最小自乗法に関する不可思議な現象であるが、その後多くの専門家の間で激しい議論を巻き起こすことになり、筆者の理解では、現在でも完全なコンセンサスが得られたとは言えない状況にある。Peelle 氏自身もこの分野での大家であるが、彼自身はこの質問を発して以来どういうわけか沈黙を保っており、1990 年に筆者が彼と会った時などは、『こんなパズルを見つけなければ良かった』と言っていた。なお、本稿では、以下敬称は省略させていただく。

このパズルは、明らかにされた直後、わずかな議論がなされただけで未解決のまま 2 年程が過ぎた。たまたま筆者が 1989 年に ANL に滞在中、当事者の一人で滞在先の研究室にいた D.L.Smith よりこの話を聞き、二人でこれに関するレポート[2]の草稿をまとめ Peelle を含めこの問題に関心を持っている専門家数人に送りつけたところ、冒頭で述べたような激しい議論－その後 1 年以上続くことになった－が始まるきっかけとなったわけである。この専門家とは、F.G.Perey(ORNL、退職)、H.K.Vonach(IRK)、F.H.Fröhner(KfK)、W.P.Pönitz(ANL)、D.W.Müir(LANL、IAEA) で、Peelle 自身が例の手紙を送った人たちである。我々の送った草稿は、実験の誤差に関する基本的な考察に基づいているが、Fröhner と Perey からは非難を受け、Vonach からは賞賛され、Pönitz、Müir と Peelle からは無視された。その後、主に手紙と電子メールを通して意見の交換を行い、我々としては一応の理解が得られたと思っている。

2. PPP とは

さて、PPP は最初定義自体にも曖昧さがあり、多少の混乱をきたしたが、最終的には次のように要約することができる。

一つの物理量 D に対してふたつの測定データ $d_1=1.5$ と $d_2=1.0$ があるのでそれらの平均値を求めることにしよう。ただし、それぞれ 10% のランダムな誤差成分と、20% の系統的な誤差成分をもっていることが分かっている。従って、d_1 と d_2 の共分散行列、V、を

次のように計算出来る。

$$V = \begin{pmatrix} d_1^2 \cdot (0.1^2+0.2^2) & d_1 \cdot d_2 \cdot 0.2^2 \\ d_1 \cdot d_2 \cdot 0.2^2 & d_2^2 \cdot (0.1^2+0.2^2) \end{pmatrix} = \begin{pmatrix} 0.1125 & 0.06 \\ 0.06 & 0.05 \end{pmatrix} \qquad (1)$$

ふたつのデータ d_1 と d_2 は、ベクトル d の第1および第2成分とする。すなわち、

$$d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1.0 \end{pmatrix} \qquad (2)$$

このふたつのデータの平均値を推定すべきパラメータとして x とおくと、観測方程式は

$$\begin{cases} d_1 \sim x \\ d_2 \sim x \end{cases} \longrightarrow d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \sim \begin{pmatrix} x \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot x \equiv A \cdot x \qquad (3)$$

となり、これはデザイン行列（design matrix、感度行列とも言う）A を

$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad (4)$$

と取ってパラメータ x を推定する問題となる。最小自乗法によると、 x は、次式、

$$Q^2 = [\ d - A \cdot x\ ]^t \cdot V^{-1} \cdot [\ d - A \cdot x\ ] \qquad (5)$$

で定義される Q^2 を最小にするようなものとして与えられる。この時、Q^2 はカイ２乗と呼ばれる。最小自乗法の解 x はパラメータの不偏推定値かつ最小分散の解（Gauss-Markov の定理）となっている[3]等いろいろ好ましい性質をもっているのでここでもそれに従って x を求めることにしよう。最小自乗の条件から、x は行列の形で、

$$x = [\ A^t \cdot V^{-1} \cdot A\ ]^{-1} \cdot A^t \cdot V^{-1} \cdot d \qquad (6)$$

と書くことができる[4,5]。今の場合は２行２列の行列の逆を容易に求めることができて、最終的には

$$x = \frac{(V_{22}-V_{12}) \cdot d_1 + (V_{11}-V_{12}) \cdot d_2}{(V_{22} - V_{12}) + (V_{11} - V_{12})} = \frac{(0.05-0.06) \cdot d_1 + (0.1125-0.06) \cdot d_2}{(0.05 - 0.06) + (0.1125 - 0.06)}$$

$$= -0.235 \cdot d_1 + 1.235 \cdot d_2 = 0.882 \qquad (7)$$

となる。x の誤差(標準偏差) Δx は誤差伝搬の法則より、

$$\Delta x = \sqrt{[\ A^t \cdot V^{-1} \cdot A\ ]^{-1}} = 0.218 \qquad (8)$$

また、カイ２乗は $x = 0.882$ を（5）式に代入して 5.882 となる。

この解法では、1.5 と 1.0 というふたつのデータの平均値が 0.882 となった訳で、このため Peelle が友人に助けを求めたのが冒頭に述べた手紙である。以上が PPP の実質的内容である。

我々がこれは大問題だと感じたのは、核データの評価の方法として、最小自乗法を適用したものの中に往々にして測定データに対して低めの評価値となっているものがあることを知っていたからである。例えば、図１は $^{115}In(n,n')^{115m}In$ 反応断面積であるが、最小自乗法を使った結果はカーブＡとなり、明らかに小さすぎる値となった。積分実験の結果と比較しても過小評価となっていることが分かった。一方カーブＢは後述する我々が提案した方法で得られたものである。

さて、読者の皆さんは 0.882 というこの答についてどういう感想をお持ちであろうか。筆者が今まで出会ったさまざまな意見は大別すると以下のようなものである。

① 最小自乗法を信じる限り、一見矛盾しているように見えるこの答は実は正しい。
② この問題では、結果の誤差が大きいので、一見変な結果には見えるが、少なくとも x は小さい方のデータ d_2 と誤差範囲内では一致しているので問題ではない。
③ 結果のカイ２乗が１より非常に大きいので用いたデータは明らかに矛盾しており、この様なデータを用いて統計的推論をやるのはおかしい。より信頼のおけるデータが測定されるまで待つべきである。
④ 何かが間違っているに違いないがよく分からない。
⑤ 最小自乗法が正しくない場合もあることを示している。

① を主張する人の言い分はこうである。すなわち、この問題ではデータに正の相関があるので、データ d_1 と d_2 は、真値に対して、同時により高い値、あるいは低い値になっている可能性がある。すなわち偏差がある。それがどちら側に出ているかは神のみぞ知るところであるが、最小自乗法は、神様しか知らないはずのこの偏差をデータの共分散行列より読みとって適切に －この場合低い側に－ 補正を行った、というわけである。これは一見もっともな意見のように思えるかも知れないが、最小自乗法で扱える誤差相関とは、複数のデータに<u>共通な偶然誤差</u>（common random error）のことを指しており、いかに最小自乗法と言えども、真に系統的な誤差（systematic error）、例えば常に数え落としをする測定器を使っていながらそれに気づかない場合の誤差など、は扱いようのないもので、それを補正することは不可能である。また、PPP の場合のように、この共通な偶然誤差に

関する情報が一つしかなければ、最小自乗法をもってしても、その共通な成分に対して、より良い推定値を与えることが不可能なことは明かである。このことは後ほど Zhao-Perey の方法の所で示す問題の置き換えをすることによっておいおい説明する。ただ、理解しておかなければいけないことは、この問題ではデータ数が少なくしかも強い相関があるので、真値が区間［1.0, 1.5］の外にある確率を無視することは勿論できない。

② については、確かにこの場合は推定値 x の誤差は大きい。しかし、仮にデータの持つ誤差、10% のランダム成分と 20% の共通成分をそれぞれ 1% と 2% に変えたとしよう。データの共分散行列を計算しなおして x を再び（6）式に従って計算すると答は今度も 0.882 である。Δx は 0.022 と小さくなって、② の主張はもはや成立しないことになる。

一方、 ③ はしごく正当的な主張である。但し、このようなことを言う人の多くは純粋の統計学者で、実際にはこんなことを言っていたら現実世界で起きる多くの問題に対して最小自乗法を適用することはできなくなってしまう。特に核データの評価においては、自由度あたりのカイ２乗が１以下になる場合などめったにあることではなく、我々は明日にでも職業安定所にいかなくてはいけないはめになる。逆に、データにばらつきがあるからこそこのような統計的推論をやることが必要になるわけで、我々がやろうとしているのは、データが必ずしも整合的でなくとも、そこから少しでも情報を読みとろうという努力である。今はふたつのデータがあり、そのおかげで D が大体１程度の大きさであることが分かった訳であるから、このふたつのデータのもつ価値は非常に大きい。定量的には例えばシャノンの相互情報量[6]（情報伝達度[7]）を計算するとこのデータが有意であることが明かとなる。また、データの誤差を 25% と 50% に変えたとしよう。この場合も x は 0.882 となるが、カイ２乗は今や 0.941 となる。従って ③ の主張もここで完全に挫折する事になる。従って ③ の立場は、現実的でないという意味と、せっかくの情報を無駄にしているという意味、さらに誤差の大きさが変わると成り立たなくなる点から退けられるべきである。

一般にこの問題ではランダム成分と共通成分の誤差の比を 1:2 に保つ限り 1.5 と 1.0 の平均値 x は 0.882 となることが分かっている。問題をオリジナルの PPP だけに限れば ② と ③ の主張は正しいと言っても良いので、PPP はパズルではないという人もいないわけではないが、我々の求めているものはもう少し一般的な場合にも通用する解釈なのである。

この問題で我々が最も疑問に感じたのは、(7) でデータ d_1 に負の重みがかかっていることである。一体これはどういうことであろうか。また、d_1 の重みは (V_{22} -V_{12}) に比例するので、たまたま $V_{22} = V_{12}$であるとすると、答は常に（d_1 の値によらずに）$x = d_2$ となる。例えば $d_1 = 1.5$ であろうが、 $d_1 = 6.02\times10^{23}$ であろうが、それと $d_2 = 1.0$ を平均したものは 1.0 になってしまう。こんなことが合理的であるはずがない。というわ

けで、我々は ④ あるいは ⑤ の立場から出発した。

3. 我々の提唱した方法

我々が最初考えたのは非常に単純なことであった。つまり、今の場合データの誤差は % で与えられているが、核データの測定等ででてくる例では、その意味するものは、真値の回りのぼやけ具合のことである。例えば、 Poisson 分布の幅とか、検出系の分解能である。それならば、この %-誤差がかかる相手は、測定値そのものではなくて、今から推定しようとしている平均値（こちらの方が生のデータよりも真値に近いはずなので）そのものであるはずである。

この考えが正しいとして、データベクトル d の共分散行列を計算してみよう。平均値は x であるから（勿論今は未知の値である）、

$$V = \begin{pmatrix} x^2 \cdot (0.1^2+0.2^2) & x \cdot x \cdot 0.2^2 \\ x \cdot x \cdot 0.2^2 & x^2 \cdot (0.1^2+0.2^2) \end{pmatrix} = \begin{pmatrix} 0.05 & 0.04 \\ 0.04 & 0.05 \end{pmatrix} \cdot x^2 \qquad (9)$$

となる。これは x が未知の量だから計算できないじゃないかと言われるかも知れないが、この問題では好運なことにこれらの値を（7）式に代入すると未知数 x は分母と分子でキャンセルして、結果は

$$x = 0.5 \cdot d_1 + 0.5 \cdot d_2 = 1.25 \qquad (10)$$

と求まるので、この値を（9）式に代入して

$$V = \begin{pmatrix} 0.0781 & 0.0625 \\ 0.0625 & 0.0781 \end{pmatrix} \qquad (11)$$

と求めることができる。この値を（8）式に代入すると Δx も求めることが出来て、

$$\Delta x = \sqrt{[\, A^t \cdot V^{-1} \cdot A \,]^{-1}} = 0.265、 \qquad (12)$$

一方カイ２乗は 8.00 となる。カイ２乗が１より大きいのでデータが良くないことは前と同じだが、最小自乗法を適用した結果は 1.25 ± 0.27 で、どちらのデータともコンシステントな値である。カイ２乗が大きくなったのは、単にデータに与えられている誤差が過小評価されていることを示している。

さて、我々のやり方の本質は %-誤差を平均値にかけたことである。これに対して Fröhner と Perey は痛烈な批判をしてきた。我々としても、このやり方が常に正しいと言っていたわけではない。少なくとも 0.882 よりはまともな答を出せるという意味で実用的

な方法であると主張したわけである。しかしながら読者のみなさんの中にも余りにも安易ではないかとおっしゃる方がおられるかも知れないので、我々が考えた得た理由付け（I～V）を以下に示すことにしたい。

I．まず、最小自乗の条件に関する（5）式であるが、これは一体どうやって導かれた式だろうか。我々はこの Q^2 を最小にする解が不偏かつ最小分散の解であることは知っているが、それにしてもいきなりこの式が与えられるのはいかにも唐突である。この式の意味を理解するために、もう少し基本的な Bayes の定理に戻って考えてみよう。 Bayes の定理によると、データ d を得た場合のパラメータ x に関する知識は事後分布 $p(x \mid d)$ に集約される。 ところで、これは

$$p(x \mid d) = C \cdot p(d \mid x) \cdot p(x) \tag{13}$$

のようにデータ発生モデル $p(d \mid x)$ と x に関する事前分布 $p(x)$ の積で与えられる[4,6,7]。C は単なる規格化定数である。$p(d \mid x)$ は x が所与の場合の d の分布を表しているが、もろもろの理由から正規分布とするのが最も一般的な選び方である[5-8]。 つまり $p(d \mid x) \sim \exp(-\frac{1}{2}Q^2)$ となる。ただし、この時点では d は確率変数である。一方、正規分布をする確率変数に対する事前分布 $p(x)$ は、ジェフリーズのルールに従うと一様分布となる[7]が、これは最大エントロピーの分布[6]でもあって十分妥当である。 従って、今や $p(x \mid d)$ は $p(d \mid x)$ に比例するものになる。最小自乗の条件（5）式は、$p(x \mid d)$ が x の関数として最大値を取るような値（モード）を x の推定値とするという意味を持っているわけである。これは $p(d \mid x)$ を尤度関数とみなして最尤推定法を採用することとも一致している（読者のみなさんはこんなことはとうにご存知かも知れませんが）。ところで、データ発生モデルというのは、データの振る舞いをある確率分布で近似させるモデルである[7]。具体的には、未知のパラメータが全て分かっているとして、データを繰り返し測定した場合にそれがどんな分布をするかを表すモデルである。従って、ここで考えるべき分布の幅は、真値の回りの真の分布幅であるべきである。通常はこの幅の推定値としてデータの誤差を代入しているわけだが、本来はそれは正しくないわけである。

II．計数実験の場合は元々の分布がある幅を持った分布、 Poisson 分布、となっている。この場合には、通常はデータの平方根をデータの誤差と呼んでいるが、これも実は正しくない。何度も繰り返すように、データの誤差とは、データが属している母分布の標準偏差のことを指しているのである。従って、同一の Poisson 分布からサンプリングしたデータは、個々の値は当然ばらつきを示しているだろうが、全て同一の誤差を持っていると考えるべきである。これが文献８の主題である。Poisson 分布に従うデータの重みを、よくや

られるようにデータの逆数と取ってしまうと、最小自乗法によってあてはめをした結果は、積分値において常にカイ２乗の分だけデータより小さくなってしまう[8]。この場合も、データの重みとしては真の分布の幅（ Poisson 分布では真の平均値に等しい ）を取るべきである。もちろん真の分布の幅、あるいは平均値など最初からは分かっておらず、それを求めるのが最小自乗法の目的であることが多いわけだが、文献８のやり方では、平均値として最初ある初期値を仮定して、それを用いて最尤推定法をすることを推奨している。ただし、非線形問題での繰り返し計算においてこの初期値を固定して使っている。

Ⅲ. 核データの測定で良く出てくる例として、例えば検出効率の誤差が 5% というときはどういうことを指しているのだろう。本来であれば、検出効率を複数回測定して、その分布の幅をもって誤差と呼ぶべきであるが、実際にこれが行われることは少ない。通常 5% の誤差が意味するものは、過去の経験から得られた再現性の程度であるとか、検出器の分解能のことを指している、あるいは測定者の主観的な判断のときもある。このように、誤差のうちの多くは、絶対値でなく、初めから相対値で、しかも多くの場合単に信頼性の目安として与えられている。従って、それをそのままデータにかけると、たとえ同程度の信頼性を持つ母集団からのサンプリングを行ったとしても、常に小さめのデータの誤差がより小さくなって、大きな重みを持ってしまうことになる。

Ⅳ. 核データの測定においては、最終的な物理量、例えば断面積は、いくつかの独立な因子の積あるいは商の形で与えられることが多い。例えば、散乱断面積（σ）の例を取ると、

$$\sigma = \frac{A_1 \cdot A_2 \cdot \ldots\ldots \cdot A_n}{B_1 \cdot B_2 \cdot \ldots\ldots \cdot B_m} \tag{14}$$

ここで A_1 は計数率、A_2 は多重散乱の補正係数、A_n は標準断面積、B_1 は検出効率、B_2 はモニターの計数率、B_m は標準サンプルの計数率、 ..... という具合に書ける。誤差伝搬の法則に従うと、A_1 の誤差を ΔA_1 等、σ の誤差を $\Delta\sigma$ とすると、次のようになる。

$$\frac{\Delta\sigma}{\sigma} = \frac{\Delta A_1}{A_1} + \frac{\Delta A_2}{A_2} + \ldots \frac{\Delta A_n}{A_n} - \frac{\Delta B_1}{B_1} - \frac{\Delta B_2}{B_2} - \ldots \frac{\Delta B_m}{B_m} \tag{15}$$

要するに、伝搬していくのは相対誤差であって絶対誤差ではない。従って、最終的に求められる σ の誤差も、信用できるのは相対誤差である。

Ｖ. 以上のことから、少なくとも核データの分野で行われる測定の多くの場合に、信用す

べきものは相対誤差であり、また、最小自乗法で用いるべき誤差が真の分布幅であるならば、誤差の絶対値を求めるときに、相対誤差がかかるべき相手は、生のデータそのものではなくて、真値、あるいは平均値であるべきである。

以上が我々が考えた理由付けである。通常、真値、あるいは平均値は最小自乗法を適用して初めて求めることができるものなので、最初はある値を仮定して、繰り返し計算によってそれを逐次改良していく必要がある。このため、いわゆる線形の最小自乗法においても繰り返し計算が必要となる。図1のカーブBはこのようにして求めた例である（詳細は参考文献2および次号を参照のこと）。PPP の場合のように求めるべきパラメータが線形でしかも一つしかない場合にのみ繰り返し計算は必要ではない。これが我々が最初提案した方法である。考え方としては文献8のものに近いと言えなくもないが、 Poisson 分布だけでなくてより一般的な分布に対しても適用できる（適用しようとしたと言った方が適切かもしれない）点と、繰り返し計算を行う点で改良が行われている。

以上の説明から、通常のやり方で最小自乗法を適用すると、結果はほとんどの場合において、小さいデータにより大きな重みがかかる傾向にあることが分かる。言ってみれば、PPP はこれが極端になって、大きい方のデータに負の重みがかかってしまった例である。言葉を変えて言うと、最小自乗法は、データベクトルからパラメータベクトルへの線形変換（(6) 式）と考えることが出来るが、その変換行列がデータの共分散行列(と感度行列)に依存しているので、それが正しく与えられていなければ最小自乗法が誤った答を出すことがあるのは無理からぬことである。非線形の最小自乗問題では、感度行列 A が未知のパラメータの関数となるので、誤った初期値を用いると悪い解に集束する事がある。これも誤った A を用いる事により線形変換の行列が正しく与えられなかったためと考えれば、PPP と共通の現象だと言う事ができる。ただし、これらは最小自乗法が悪いのではなく、入力データが悪いのである（Garbage-in, Garbage-out の法則）。従って、我々としては意見 ⑤ は成り立たないと思っている。

さて、以上が PPP に対する我々の最初の解釈と、可能な対策を述べたものであるが、Fröhner などはこれに全く賛同せず、『最小自乗法を知らない奴が ENDF の評価をやっている』とまで手紙に書いてきたものである。『最小自乗法を知らない奴が JENDL の評価をやっている』と書いてこなかったのは単に彼が筆者のことをよく知らなかったからにすぎない。ちなみに、彼は最初は 0.882 という答を正しいと信じきっており、これがなぜ合理的かを示すために一見実に見事な説明までしてきた。Perey は 0.882 という値はおかしいと感じていたようであるが、かと言って我々の方法は気に入らなかったようである。一方 Vonach は、『この方法に完全に同意する』と言ってくれた唯一の味方であった。繰り返し

て言っておくが、我々もこのやり方が常に正しいと言うつもりはない。一番の弱点は、計数実験の誤差（計数を N としたときの $N^{1/2}$）は、N が十分大きくない限り絶対値としても相対値としても正しくないことである。従って（14）の形の式中の因子の中に計数実験の結果を含む核データの分野においては、一般的に成り立つ厳密な正当化ができないことは承知している。しかし、最終データが（14）式の形に書ける時に、しかも絶対誤差よりも相対誤差に信頼がおける因子が多い場合には、従来のように "データの誤差 = %-誤差×データ" とするよりは最小自乗法で用いるべき本来の意味の誤差に近く、従って PPP のようなへんな結果とはならないと言っているわけである。厳密にやるためには、個々の因子の正しい値と正しい誤差を予め求めておかなければならないが、現実問題としてその様にはなされてこなかったし、今後もされることは希であろう。

参考文献

1）R. W. Peelle, "Peelle's Pertinent Puzzle", private communication (1987).

2）S. Chiba and D. L. Smith、"A SUGGESTED PROCEDURE FOR RESOLVING AN ANOMALY IN LEAST-SQUARES DATA ANALYSIS KNOWN AS "PEELLE'S PERTINENT PUZZLE" AND THE GENERAL IMPLICATIONS FOR NUCLEAR DATA EVALUATION", to be published as ANL/NDM-121.

3）S. ブラント、データ解析の方法、みすず書房（1975）。

4）B. R. Martin、Statistics for Physicists, Academic Press (1971).

5）D. L. Smith、"Probability, Statistics, and Data Uncertainties in Nuclear Science and Technology"、Neutron Physics and Nuclear Data in Science and Technology, Vol. 4, OECD, American Nuclear Society (1991).

6）井原俊輔、確率過程とエントロピー、岩波書店(1984)。

7）繁桝算男、ベイズ統計入門、東京大学出版会(1985)。

8）栗屋隆、データ解析、学会出版センター(1983)。

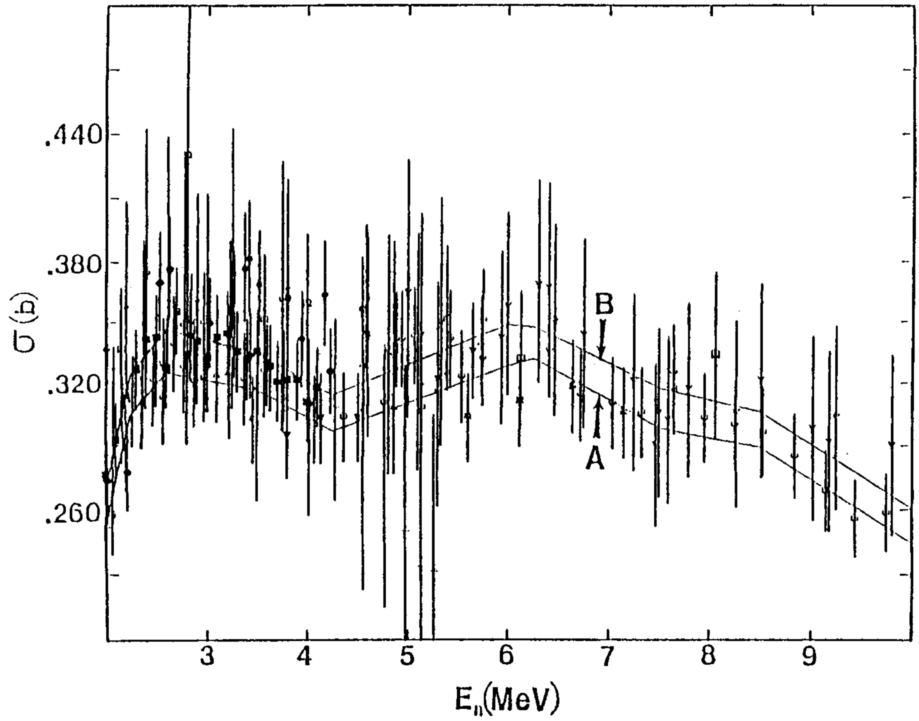

図１　^{115}In(n, n')^{115m}In 反応断面積

カーブＡは通常の最小自乗法の結果

カーブＢは我々の提案した方法による結果

（詳細は本文および次号を参照されたし）