ノイズ、S／N、そして平滑化　I

福島　整
科学技術庁　無機材質研究所
（1995年2月13日受け付け）

実測スペクトルに必ず付随するノイズについて、取り扱うためのモデル、S／Nの問題、平滑化の基礎とそれに伴う問題を取り上げて数回に分けて解説する。今回は、総ての議論の基礎となるノイズを取り扱うためのモデルについて論ずる。

ノイズの実用的なモデルに重要な仮定は、真の信号とノイズ（変動）が分離できる事、分離されたノイズ成分の分布が決められる事である。この点からモデルに要求される、加法的雑音の性質と、ノイズ成分の分布としてのポアソン分布とその取り扱いについて解説する。

1.　序

表面分析に限らず、ありとあらゆる計測には変動が伴う。この変動は、計測結果の解釈や利用にとって大きな影響をもたらすのが常であると言って良い。スペクトロスコピーを応用する場合であれば、この変動はいわゆる「ノイズ」という言葉で表現される。しかし、この「ノイズ」も様々な意味を含んでおり、しばしば混同される。その結果、計測結果の解釈、利用のための高度な処理等において無用の混乱を生じたり、誤った結論さえ導きかねない原因となる。

「ノイズ」というと、計測系が持っているものと、計測対象を観察するとき本質的に逃れえないものの二通りに大別できよう。ある部分ではこの二つは分けることが出来るが、ある部分では本質的に不可分である。すなわち、「ノイズ」というものは実際に明確に定義し区別することが意外に困難であり、その影響を正しく評価することも困難であり、結果として多くの場合曖昧にしたままとなってしまうこととなる。

過去の表面分析の分野におけるノイズに関連した議論としては、装置の検出方式の違いによるS/Nの取り扱いについて論じたSeahらの報告[1,2,3]がある。また、測定されたスペクトルに対する平滑化処理に対するSeahの経験的な評価[4]やBrombaらの荷重移動平均法一般に関する理論的な提案[5]もある。しかし、前者は測定されたスペクトルのS/Nの評価というよりはS/Nの数値化による装置の性能比較が目的であり、後者は理想的なものであってノイズに対する実用的的な議論とは言いがたい。

今日の分析機器はほぼ完全にコンピュータ化されており、計測された分析データは一連の離散変数のセットで与えられる。このデータを解釈するために、そのままの形だけでなく、ファクターアナリシスなどのさらに高度な処理も容易に利用出来るようになってきた。しかし、これらの高度な数値処理は一般にデータに本質的に付随している統計変動（計測系以外に起因しているノイズ）に強く影響される場合が多く、この影響の評価、あるいは除去の重要性が改めて問題になりつつある。

では、それらのニーズに答えうる様なノイズ除去の手段が存在するかというと、表面分析の分野では、実はその為の議論すらあまり活発になされてこなかったのが現状であると言い切っても良かろう。ノイズ除去は、平滑化あるいはスムージングと呼ばれるジャンルであり、ほぼ一般的かつ日常的に用いられている代表的な方法もいくつか存在する。しかし、これほどその結果が与える影響をほとんど考慮されずに用いられている手法も珍しいと言って良い。

ノイズ除去に関する議論は、通信工

学の分野では高度に発達を遂げている。通信工学では情報の正確な伝達が目的であるから、伝達過程に存在する擾乱による影響を正確に除く事が必用である。その為の、いわゆる信号復元、あるいは波形の再構成に関する研究が常に重要なテーマとなっている。

この通信工学の分野（信号伝達論）と分光データ処理では、その出発点に本質的に大きな違いがある。すなわち、通信工学はノイズの無い真の波形が常に与えられているのに対し、分光データ処理ではほとんどの場合真の波形が未知である点である。この為、通信工学の様々な理論が直接適用できず、したがって分光学上の実用的な議論も大きく遅れてしまったものと考える事も出来よう。

以上の簡単な総括から容易に理解できるとおり、ノイズとS/Nおよび平滑化の議論は密接不可分といえる。本稿は、表面分析研究会の実用電子分光法講座の一環として行わせていただいたS/Nに関する一連の講義ノートを基に、講義では触れなかった数式もおりまぜて、ノイズ論の基礎、S/Nの評価法、およびディジタルフィルタ論を含む平滑化論についての簡単な議論を紹介させていただく。

2．ノイズ

2.1　ノイズを取り扱う為の基本的な仮定

これから展開する一連の議論における「ノイズ」とは、計測系起因による擾乱以外の、観測対象が本質的に持つ統計的変動を指すものとする。もっとも、計測系を考慮せずしては観測対象の信号特性も議論できないわけで、その意味では厳密な線引きを行う事は簡単ではない。

例えば、ディスクリミネータのレベルを調整することによりカットできる低い電圧の細かい変動は、観測対象の本質的な変動と明らかに分離することができる。また、ケーブルが拾ういわゆる外部からのノイズの類も、様々な工夫によって軽減あるいは閉め出す事が可能である。これらが、擾乱による変動である。

一方、様々な電子部品の中を流れる信号電流は一連の電子の流れであるから、媒体中を移動する電子の性質に伴った統計的（確率的）変動を信号成分から除く事は本質的に不可能である。したがって、いかに充分注意を払って作製された計測機器で変動の全く存在しない信号を計測したとしても、結果として検出される信号から統計変動を取り除く事はおそらく不可能である。

さて、計測系が与える統計的変動は数式の上ではどう与えられるであろうか。これは、まさに通信工学におけるもっとも基礎的な信号伝達論のモデルがそのまま用いられるケースであり、統計的変動は次の式（モデル）によって扱われる。すなわち、測定対象の変動のない信号をx、ある定数（例えば計測系の増幅率など）をc、計測系のxに対する出力信号をyとすると、ある統計的な性質をもつ乱数wを用いて

$$y = c \cdot x + w \qquad \ldots(1)$$

と表される事が多い。このwがノイズに相当することは、容易に想像がつくであろう。

しかし(1)は、実は問題を非常に単純化したモデルである。すなわち、この式では計測系の作用c（系の伝達関数に相当する）とノイズwが切り離されている。したがって、実際の信号がこのモデルに従うのであれば、ノイズを完全に取り払って真の波形を推定することが可能であることになる。現実には、分光学的なデータにおけるその様な保証はまったく無い。これに対して、信号伝達系において(1)の様なモデルが良い近似で成り立つ場合が多いというのが、通信理論でこの様な式が基本として現れる所以であろう。

さて、測定系が(1)の様な伝達関数cを持つような性質の良い系（線形応答を示す系）であるとしよう。この系で、平均値$\overline{x}$、分散σ_x^2であるような任意の確率分布に従う様な信号xを観測したとする。また、ノイズに対しても、平均値$\overline{w}$、分散σ_w^2であるような任意の確率分布にしたがうものとし、xとwは独立に変化する（互いに影響しあわない）ものとする。この段階で、仮定が少なくとも二つ入ったことに注意されたい。

この様なxを観測した場合の出力yの平均値yは、統計でよく用いられる平均値を与える演算子Eを用いて、$\overline{x} = Ex$、$\sigma_x^2 = E(x - \overline{x})^2$であることから、

$$\bar{y} = Ey = E(c \cdot x + w)$$
$$= c \cdot Ex + Ew = c \cdot \bar{x} + \bar{w} \quad ...(2)$$

また、分散については

$$\sigma_y{}^2 = E(y - \bar{y})^2$$
$$= E(c \cdot (x - \bar{x}) + (w - \bar{w}))^2$$
$$= c^2 \cdot E(x - \bar{x})^2 + E(w - \bar{w})^2$$
$$+ 2c \cdot E(x - \bar{x})(w - \bar{w})$$

ここで、xとwの変化が互いに独立であるという仮定から、$2c \cdot E(x - \bar{x})(w - \bar{w})=0$が成立し（互いに独立な変化であるので、共分散が0となる）、結局

$$\sigma_y{}^2 = c^2 \cdot E(x - \bar{x})^2 + E(w - \bar{w})^2$$
$$= c^2 \cdot \sigma_x{}^2 + \sigma_w{}^2 \quad ...(3)$$

となる。(1,3)は、c=1としてみると良く見慣れた誤差の伝搬則の式そのものであることに気がつかれると思う。

すなわち、(1,1)のモデルを用いることができて、観測する信号の変動と系が与えるノイズの変動が無関係（互いに独立）であるのならば、観測の結果として得られた信号の平均と分散も、信号成分とノイズ成分のそれに簡単に分離できることが示されるのである。この様なモデルで扱えるノイズを、「加法的雑音」と呼ぶ。信号伝達論ではノイズの大きさをその分散で評価することで（平均値は通常0と仮定される）S/Nの議論を進める場合が多いことから、このモデルがいかに都合がよいかも理解できるであろう。

一方、理想的な内部雑音のない分光器をあるエネルギー値にセットして信号強度を観測した場合、得られた信号強度はある分布を示す。したがって、その分布からある値（ほとんどの場合は平均値）i_oとそれからの変動分ω（信号強度の分布関数で決まる乱数となる）を定義する事ができて、観測される信号強度iは

$$i = i_o + \omega \quad ...(4)$$

と記述できる。これはすなわち、全く統計変動を持たない信号源を、単にノイズをつけ加えるだけのシステムを通して観測するモデルを採用する事と対応する。したがって、(4)は(1)のモデルと同等に扱うことができ、ωは加法的雑音としての取り扱いが可能となる。つまり、元々の信号強度に分散を考えていないから、観測された信号強度の分散がそのままノイズの性質として扱えるわけである。

また、(4)から様々に設定されたエネルギー値に対応した$\bar{x}$の与える曲線が真のスペクトル形状を与えることになるのであるが、これもここでの議論の範囲では明らかに仮定の産物であることに注意しなければならない。なぜなら、$\bar{x}$はあくまでも平均値であって、推定量だからである。ただし統計の立場から、このモデルがおそらくもっとも真実に近い結果を与えてくれるであろう事が示されるが、それは後ほどふれる。

もう一つ、ノイズを扱う多くの議論において、一般的に（あるいは暗黙の了解として）用いられる重要な仮定として、そのエルゴード的性質がある。

自然現象の中には様々な不規則性が観測できるが、それらは、一般に定常的なものと非定常的なものに分類される。この定常的な不規則過程は、さらにエルゴード的（ergodic）と非エルゴード的な過程に分類される。

「エルゴード的」とは、任意のサンプリングによる集合平均と任意の観測時間による時間平均が一致する事を言う。「起きる確率が分かっているから、何回かやっていればそのうち起きるであろう。（やってりゃそのうちあたるさ）」という考え方である。数式による説明は冗長であるので、参考書[6]等を参照されたい。

ある分光計をあるエネルギー値にセットして信号強度を観測する場合を考えよう。まず、1台の分光計で測定をn回繰り返し、得られた信号から平均値を得る（時間平均）。これとは独立に、その分光計とまったく同じものをn台準備し、総ての装置で同一条件下で1回だけ測定を行いn個のデータを得る。このn個の信号の平均値（サンプル平均。アンサンブル平均とも呼ばれる）が、先のn回繰り返しにより求められた平均値と統計的に一致しているとみなすことが出来る場合、観測している過程がエルゴード的であるとするのである。

これは、一見当たり前のように見える仮定であるが、実際の信号においてこの性質が成立することを証明するのは大変困難である。しかし、例えばホワイトノイズの定義やある伝達関数を作用させたときのノイズの変化など、ノイズの周波数特性を理論的に議論するときなどには不可欠な仮定となる。

またこの仮定を採用することで、計測法や信号の発生過程等によらず、どの様な観測で得られた信号でもそれに含まれているノイズの性質は全く同じであるという仮定を採用することができるのである。

この様に、実用性を主眼としてノイズを簡単な数学で取り扱おうとするためには、問題を単純化するために様々な仮定が実は必要であることがお分かりになって頂けたかと思う。

現実には、例えば分光器を走査しながら検出粒子の計数を行うとき、隣り合ったデータ点それぞれの計数値に伴う統計変動が完全に互いに独立であるという保証すら無い。もし各データ点の統計変動が互いに何らかの相関を持っていた場合であると、たとえエルゴード性が保証されたとしても、加法的でないため、真のスペクトル形状に対する仮定が複雑になってしまう。一方加法的でもエルゴード的でなければ、ノイズの取り扱いに対する一般化が困難となり、やはり問題が複雑になる。

したがって、多くの場合、極めて単純化されたモデルによってノイズの取り扱いが議論されていることを常に念頭におく必要がある。

2.2 ノイズのモデル （ポアッソン分布と正規分布）

先の議論から、以後の検討におけるノイズのモデルとして、(4)で示されるエルゴード的な加法的雑音を考慮する事にする。

信号iに対して、(4)の形から、平均値i_0とそれからの変動ωを与える分布関数を考えることが出来る。すなわち、iが従う確率過程が与える分布関数をモデルにもちいるわけである。

一般に光子や電子などの量子計測の場合、検出する数が大きくない場合の信号強度分布はポアッソン分布に従うことが知られている。この分布関数は上に凸の山形の形状を示し、次の様な式で表される。

$$f(x) = (\lambda^{x} / x!) \cdot e^{-\lambda} \qquad \ldots(5)$$

このポアッソン分布は、あまり頻繁に起こらない事象の確率分布としてよく用いられる。例えば、ミスプリントが生ずる平均確率λが分かっているときの、ある本のページ毎のミスプリントの確率はポアッソン分布から計算できる。実際に、電子線照射時に生成されるあるエネルギーを持つ二次電子の強度がポアッソン分布に従うことも良く知られている。

ポアッソン分布を導出するための仮定は、次の3点である。

1) 単位時間あたりに観測される事象の平均回数（λ）はいつも一定である。
2) δtあたりに事象が1回観測される確率は$\lambda \cdot \delta t$である。
3) δtあたりに事象が2回以上観測される確率は、δt^2あるいはδt^3等に依存する。

このうち、1)の仮定は統計的釣り合いを示すものであり、ポアッソン分布を決める上でもっとも重要な仮定である。また、δt^2やδt^3は極めて小さな値であることから、3)の仮定は2回以上事象が観測される場合が無視できる事と対応する。ポアッソン分布の導出の仕方には色々あるが、二項分布からの誘導がもっとも分かり易いと思われる。導出過程自体は、参考書等を参照されたい[7)]。

ポアッソン分布が仮定できると、平均値（ポアッソン分布に従う確率変数の期待値）を指定すれば分散も自動的に決まってしまう。すなわち、(5)から平均値$\bar{x}$も分散σ^2も

$$\bar{x} = \lambda, \quad \sigma^2 = \lambda \qquad \ldots(6)$$

である。したがって、ある固定されたエネルギー位置で測定された真の平均強度Iの信号（もちろん、実際の測定ではそんなものはわからないのだが）に対しては、$\lambda = I$と置き換えれば良いことになる。そして、その時のノイズによる強度のばらつきも、$\sigma^2 = I$とできることになる。

しかし、実際には真の平均強度を知ることは不可能であるから、Iにも推定量を用いることになり、通常は時間平均の値が用いられる。この推定量が実は最善の策であることも、後にふれる。

またポアッソン分布の持つ性質から、二つの確率変数x_1, x_2に対して定数λ_1, λ_2を与えたときの、$x = x_1 + x_2$の与える確率分布も

$$f(x_1 + x_2) = ((\lambda_1 + \lambda_2)^{(x_1 + x_2)} / (x_1 + x_2)!) \cdot e^{-(\lambda_1 + \lambda_2)}$$

$$\bar{x} = \lambda_1 + \lambda_2, \quad \sigma^2 = \lambda_1 + \lambda_2 \qquad \ldots(7)$$

であるようなポアッソン分布となることを示すことができる。これはポアッソン分布の再生性と言われる性質で、これを用いると、積算をかけたときのノイズの低減を評価する事ができる。すなわち、比 $R=\sigma/\bar{X}=I^{-1/2}$ を考えると、ある時間での積算強度Iに対して、その倍の時間をかけて積算した場合の強度は2･Iで与えられるから、比の値は積算時間を倍にすると$2^{1/2}$だけ小さくなる事になる。スペクトルのS/Nが信号強度の平方根に比例して良くなるという一般論は、例えばこの様な話で説明できるわけである。

一方、統計学の理論から、どんな分布に従う統計値でもサンプル数が非常に大きな数であると、その統計値は正規分布で扱うことができる事が示される。これは、中心極限定理と呼ばれており、サンプル数が多い場合に正規分布をモデルとて用いてデータ解析できる一つの保証となっている。

実際のスペクトルにおいて、信号自体はその生成過程での物理的な意味によって、ポアッソン分布に従うような確率過程で扱われる。しかし、現実の測定強度値（カウント数等）は大変大きな値であり、これはサンプル数が大変大きな場合に相当すると言っても良い。したがって中心極限定理により、ばらつき自体を正規分布で扱っても実はほとんど問題ないのである。言い換えれば、(4)のωを、平均値0、分散i。で与えられる正規分布を確率分布としてもつ乱数で置き換えられるわけである。

これは、極めて便利な仮定であり、正規分布を基本とした多くの統計的な処理がそのまま適用できる事もこれからわかる。また、実際に平滑化の効果を評価する場合などに用いられるシミュレーションノイズの発生にも利用できるのである。

3．終わりに

以上、ノイズに対する実用的な議論をするために必要ないくつかの重要な仮説について、まとめてみた。実用というにはかなり抽象的な議論だとお感じになった方も多いであろうが、色々な本等で議論されているスペクトルに対するノイズの処理の記述においては、多くの場合、少なくともここに述べた事柄が前提として用いられているのである。

次回は、Seahの一連の報告[1,2,3]で論じられている装置の比較を目的とするS/Nについて、まず紹介する。次に、信号伝達論におけるS/Nの考え方を（できるだけ数式無しで）紹介する。最後に、研究会でも提示させていただいた実測スペクトルのみから求めるS/Nについてまとめさせていただく。

参考文献

1) M.P.Seah and P.J.Cumpson, *J. Elect. Spect.*, 61, 291 (1993)
2) M.P.Seah and C.P.Hunt, *Rev. Sci. Instrum.*, 59 2, 217 (1988)
3) M.P.Seah and C.P.Hunt, *NPL report*, DMM(A)101 (1993)
4) M.P.Seah and W.A.Dench, *J. Elect. Spect.*, 48, 43 (1989)
5) M.U.A.Bromba and H. Ziegler, *Anal. Chem.*, 55, 648 (1983)
6) 例えば南茂夫，喜利元貞，桜井捷海，「機器分析のたるのコンピュータ入門」（講談社サイエンティフィック，1982）のp.135からの説明が分かり易い。
7) 林周二，「統計学講義」第二版，丸善 (1976)