

統計学のいろいろな富士山曲線 ーローレンツ曲線とガウス曲線ー

西川 利男

今年から始まった休日、山の日の8月11日に考えた。統計学では、いろいろな分布関数の山形曲線ー正規分布、t分布、 χ^2 分布、F分布などが出てくる。

一方、もう20年以上も前には、筑波の国立研究所で、私は化学の磁気共鳴の一研究者として、スペクトル曲線のシミュレーションとして、似たような富士山型曲線ーローレンツ曲線とガウス曲線とで格闘していた。

ところが最近では、統計学の正規分布、t分布などで、再会し、頭を悩ましていた。日頃の素朴な疑問を、真夏の夢の雑言として、お話してみたい。

1. 富士山曲線とは

統計学に限らず、富士山型の曲線は自然界のいろいろなところに出てくる。その特徴は、次のようであろうか。

- ・真ん中に最高点がある。
- ・そこから左右対称になだらかに低くなってゆく。
- ・左右、無限に遠くなると、0へと近づいていく。

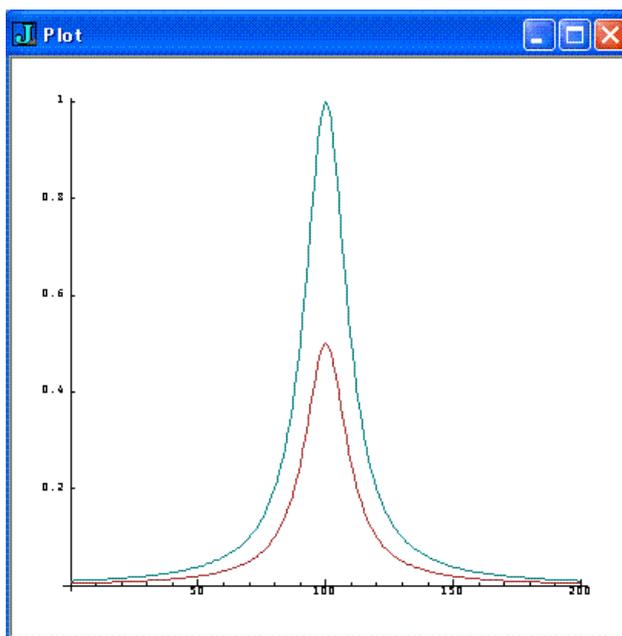
このような、曲線を数学の関数で表現するとどうなるだろうか。最も簡単には、次のようになるであろう。

$$f(x) = \frac{1}{1+x^2}$$

$$f(A, B; x) = \frac{A}{1+Bx^2}$$

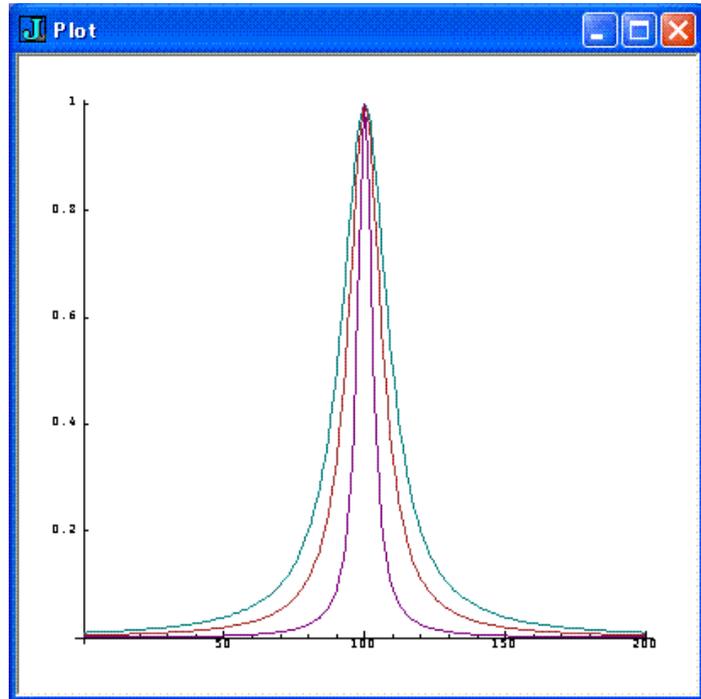
パラメータ A を変えると高さが変化する。

```
L1 =: (1, 1) lorentz 10  
L21 =: (0.5, 1) lorentz 10  
plot L1, , : L21
```



パラメータ B を変えると
裾の広がりが増える。

```
L1 =: (1, 1) lorentz 10
L2 =: (1, 2) lorentz 10
L10 =: (1, 10) lorentz 10
plot L1, L2, : L10
```



以上のグラフは、次の J プログラムで plot ルーチンにより簡単に表示される。

```
NB. plot test =====
X_axis =: 3 : '(, } : (i: y.) +/ (i.10) % 10), y.'
NB. plot *: X_axis 4

NB. Lorentz =====
lorentz =: 3 : 0
:
'A B' =. x.
X =. X_axis y.
A % (1 + B * X^2)
)
```

2. Lorentzian と Gaussian の 2 つの曲線—理工学者の考え方

スペクトル曲線のシミュレーションという観点からは、次のように考えている。一般に分光器のスリットは平行な矩形なので、スペクトルは線（輝線）になるはずである。しかし実際には、光の干渉、回折などがあるため、線ではなくバンドとして広がってしまう。それが Lorentzian である。

ところで、さまざまな状態の光源が混じり合っている場合は、統計的な確率現象の結果として、つまり多数の Lorentzian の集合となり、これが Gaussian として観測される。

私が筑波で、その昔行っていた電子スピン磁気共鳴 (Electron Spin Resonance) の世界では、均一な溶液などのスペクトルでは Lorentzian、固体の粉末などでは Gaussian

として、スペクトルのシミュレーションを行ってきた。

3. 統計学における正規分布と t 分布の曲線

統計学では、統計をとる元となる集団は母集団とよばれる。その個数が非常に多数ときには、集団の分布は平均値を中心にしてある広がりを持って分布している場合がきわめて多い。これが正規分布であり、これはガウスの誤差曲線からでてくる、つまり Gaussian である。これについては、先月の例会で扱った。[1]

$$f(m, \sigma; x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

[1] 西川利男、「超幾何関数(Hypergeometric Function)と J プログラム
—誤差関数、正規分布関数、累積密度関数を目指して—」JAPLA 研究会資料 2016/8/6

一方、実際にサンプリングした標本は、母集団の全体の数に比べれば、ずっと少ない集合である。このようなときには、ゴセットによる t 分布として、次の式が使われている。

$$f(n; x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}, \quad \text{ここで } n \text{ は自由度}$$

上の式からわかるように、これはガンマ関数の係数を含んでいるが、先あげた Lorentzian にほかならない。

統計学の教科書では、正規分布と t 分布とが並べてあげてあり、どちらも富士山型の曲線ではあるが、性質が違う曲線である。また、t 分布で自由度が大きくなったときは、正規分布と同じような形になるとしているが、これはたまたまの一致としたほうがよいのではないか。

正規分布は母集団の分布であり、t 分布は採取した標本の分布の性質であり、まったく違う。形が似ているからといって、このことをはっきりと区別してとらえ、説明することが大切であると思う。

統計学の基本の問題だが、私の素朴な考えに対して諸兄のご意見をたまわりたい。

付記

私がボランティアとして出ているお台場の日本科学未来館では、海外からも含めて多くのノーベル賞受賞者の来館がある。そして、それぞれ各国語で（手書きも含めて）、若い来館者に向けていろいろなメッセージが残されている。

そのなかの、米国オバマ大統領(Balak Obama)のメッセージの中に、“critical thinking”という語があり、私としては、大きな感銘をうけた。いろいろな分野で、真に大切なのは、覚えるのではなく、自分の頭で考えることだと思う。