

検出限界を考慮した分布モデル

広島大学大学院 岩瀬晃盛

平成13年10月2日 於統計数理研究所

ダイオキシン類の例で述べれば或る物質の検出下限は $0.02\text{pg}/\text{m}^3$ となっている。ここでの $0.02\text{pg}/\text{m}^3$ はこれに満たない濃度は検出できない(またはその必要がない)ことを意味するのであるけれども、この濃度を越える濃度の記述に於いてはこの濃度よりも細かい記述はないことになる。つまりデータの記述の観点から言えば、データの最小の刻みの幅は $0.02\text{pg}/\text{m}^3$ であることになる。検出下限と述べたときに、この量以下(または未満)は計測できないけれども、これ以上の量は望むだけの有効桁で計測できる場合が想定されていることが多い。象徴的な例としては、液体の熱膨張を利用した普通見かける温度計。或る温度以下とか以上は計測できない。しかし上記のダイオキシン類の計測の場合は検出下限とデータの刻みの幅は一致している。最小の刻み幅で記述されたデータは、離散量である。従ってこれらデータが抽出されたであろう想定されるべき母集団分布は離散分布となる。この考え方は通常見かける考え方と異なる。多く見かける考え方は母集団分布に連続分布を想定して、データはその連続分布に従う確率変数の実現値であると見なす。しかし連続分布に従う確率変数の実現値であるならば無限桁でデータは表現されるはずである。勿論無限桁で表現されたデータなど見られるはずもない。そこで本来は無限桁であるけれども適当な最小の刻み幅で打ち切って「近似的」に表現していると解釈する。この考え方は最初に母集団分布があって、これに従属した形でデータが考えられている。これに反して、先に述べたような「最小の刻み幅で記述されたデータは離散量である。従ってこれらのデータが抽出されたであろう想定されるべき母集団分布は離散分布である」との考え方は、データが在ってそれに伴って母集団分布が想定されていることに注目されるべきである。さて、データの記述の最小の刻み幅は量である。量とは数と単位との組である。先のダイオキシン類の場合では単位は pg/m^3 である。従って、ここでの離散分布とは量の離散分布であることに注意されなければならない。計測装置の検出能力が高まって且つその必要があればデータの最小の刻み幅は更に小さくなる。仮に無限に検出能力が高まってデータの最小の刻み幅が無限に小さくなっていく状況を想定すれば、これらのデータが抽出されるであろう想定されるべき母集団分布は連続分布となるべきであろう。この理由によって、データの最小の刻み幅を限りなく小さくしていく操作に対応して、当初想定された離散分布が連続分布に収束していくことが示さなければならない。このことが実現してはじめて収束した連続分布に従う確率変数の実現値として「近似的」な離散的なデータの記述が解釈されるものである。

1. 離散分布に於いても量を扱えるように表現すべきであるとの考え方をいくつかの分布で具体的に展開をした。最初に二項分布の確率関数を尺度母数を導入して与えた。見かけ上は3母数であるけれども実際は2母数である。データの最小の刻み幅をどのようにとるかは現実のいくつかの要請に基づいて定まるものであり、この制約条件が加わるので2母数となる。この下で分布の中心

と拡がりを変化させないでデータの最小の刻み幅を変化させると尺度母数が入ったポアソン分布が導出される．二項分布の場合にはデータの刻みの幅を分布の中心も拡がりも変化させないでゼロにすることはできない．

2．次に Consul(1989) が定義した generalized Poisson distribution の特別な場合を考え，尺度母数を導入した一般化されたポアソン分布を提示した．この分布に於いては分布の中心も拡がりも変化させないで刻みの幅を変化できるのは二つの場合であり，一つはポアソン分布に収束し，一つは刻みの幅を無限小にする場合となり連続分布である逆ガウス型分布に収束することが示せる．

3．尺度母数を入れた負の二項分布を提示し，分布の中心も拡がりも変化させないで刻みの幅を無限小にすることにより連続分布であるガンマ分布に収束することが示された．

4．更に，尺度母数を入れた離散一様分布を提示し，分布の中心も拡がりもある有限な値に収束する条件で刻みの幅を無限小にすることにより連続一様分布に収束することが示された．

5．拡張された負の二項分布を新たに定義し，刻みの幅を無限小にすることにより拡張されたベータ分布を導出した．

6．この拡張されたベータ分布から母数の極限操作によってべき逆ガウス型分布が導出され，結果として逆ガウス型分布や対数正規分布が示された．

これらの議論で経験的に言えることは，尺度母数を入れた離散分布のにおいて，分布の中心とか拡がりとしてどのような特性量を採用するかは自由度は残されているにせよ，これらが或る一定量に収束すると言う条件の下でデータの刻み幅を無限小にすることに対応する母数の極限操作をすることにより，連続分布が導出されるということである．本研究では分布の中心として母算術平均を採用し，分布の拡がりとして母変動係数を採用して極限操作を行った．