

分子量と BCF の関係についての一考察

平成 16 年 3 月 18 日

近畿大学農学部

教授 米虫節夫

0. 検討に用いたデータ

以下のデータを用いて統計学的な解析を行った。

- ・平成 12 年までの新規化学物質届出データ (1503 物質)

1. データの分割表形式による二次元集計

原データを横軸に分子量 MW、縦軸に logBCF を取ってプロットすると、測定点の分布は一見してわかるように、分子量が小さい部分に集中しており、かつ分子量が大きくなると logBCF が 3 を超える値は出ていない。これを分割表形式に示すと表 1 のようになる。

なお、原データの logBCF 値には -2 と記載されたデータが多いが、その他の負の値のものと合併し、全てを "0" 未満の値としてまとめた。これは、logBCF=0 の時には濃縮度が 1 であり (環境と同じ濃度)、負の値の時には濃縮度が 1 未満 (環境中の濃度よりも低い) のものであり、さらに "-2" と標記されているものは濃縮度がみられないということなので、今般検討を要する化合物が logBCF 値の大きなものという観点から考え、低濃縮度のものとして一つにまとめた。logBCF が正の値については、0 ~ 2 未満、2 ~ 3 未満、3 ~ 3.5 未満、3.5 ~ 3.7 未満、3.7 以上に区分した。log 値の 3.5 は通常数値の約 3200、log 値の 3.7 は約 5000 である。

表 1. 分割表形式のデータ集計

MW	LogBCF						Total
	負	0~1.9	2~2.9	3~3.49	3.5~3.69	3.7~	
0~100	15	12	0	0	0	0	27
101~200	179	86	20	9	0	0	294
201~300	164	114	48	7	1	0	334
301~400	118	84	33	4	0	0	239
401~500	86	90	17	1	3	0	197
501~600	57	58	14	0	0	0	129
601~700	54	26	6	0	0	0	86
701~800	39	10	0	0	0	0	49
801~900	33	16	0	0	0	0	49
901~1000	23	7	0	0	0	0	30
1000~	49	17	3	0	0	0	69
total	817	520	141	21	4	0	1503

表 1 の分子量に対する物質の出現頻度を MW=100 刻みにしてヒストグラムの的に見ると、MW が 201 ～ 300 をピークとして、MW が大きい方に裾を引く単峰性の分布を示している。MW=0 以下の物質は存在せず、MW の極端に大きな高分子物質が存在することから考え、この分布は現場データとして良く現れる形をして居る。故に、特に意図的に作られた分布、大きくバイアスのかかった分布と考えられないので、以下の推計学的解析を行う。

原データである 1503 物質中、98%に当たる 1478 個の物質は、 $\log BCF < 3$ すなわち濃縮率 1000 倍以下であった。 $\log BCF > 3$ の物質は 25 個存在したが、5000 倍以上に濃縮される物質 ($\log BCF > 3.7$) は無かった。

また、分子量 MW>500 では、 $\log BCF > 3$ を示す物質はなかった。分子量 MW>700 では $\log BCF > 2$ を示すものは 3 個のみであった。概して、分子量の大きな物質の $\log BCF$ 値は小さいといえる。

2. 二項分布による率の区間推定

表 1 から明らかなように、分子量 MW>500 では $\log BCF$ が 3 以上の物質はない。しかし、この” 0 ” という値は与えられた原データから得られたものなので、その” 0 ” がどの程度の分布をものであるかを二項分布の精密計算による区間推定という方法で考察した。計算はかなり複雑であるので文献 1) の表から、幾つかの値を抜粋し表 2 に示した。

表 2. 二項分布の精密計算による区間推定値

N	X	P	95%		99%	
			PL	PU	PL	PU
20	0	0.00	0.00	16.84	0.00	23.27
27	0	0.00	0.00	12.77	0.00	17.82
30	0	0.00	0.00	11.57	0.00	16.57
49	0	0.00	0.00	7.25	0.00	10.25
50	0	0.00	0.00	7.11	0.00	10.05
100	0	0.00	0.00	3.45	0.00	4.92
130	0	0.00	0.00	2.80	0.00	3.99
150	0	0.00	0.00	2.43	0.00	3.47
200	0	0.00	0.00	1.83	0.00	2.61
250	0	0.00	0.00	1.46	0.00	2.10
300	0	0.00	0.00	1.22	0.00	1.75
400	0	0.00	0.00	0.92	0.00	1.32
500	0	0.00	0.00	0.74	0.00	1.05
1000	0	0.00	0.00	0.37	0.00	0.53

表 2 をわかりやすく説明すると、次のように云えよう。

表には N 個のサンプル（新規物質）中に該当する性質（今回の問題では、 $\log BCF > 3$ と

なるような物性)を持つものの個数 $x = 0$ の時、その性質を持つものの出現率の点推定値は $P = 0$ であるが、その”0”がどの程度の広がり(分布)を持つかを確率計算した値が示されている。確率としては 95%信頼区間と、99%信頼区間の下限値と上限値が示されている。下限値はもちろん 0%であるので、ここでは上限値が問題とされる。今データ総数が $N = 50$ 、 $X = 0$ とするとき、出現率の点推定値では 0%であるが、95%信頼区間を考えるとときにはその上限は 7.11%となり、99%信頼区間を考えるとときにはその上限は 10.05%になることを示している。ただし、「区間推定」という値は、その範囲内に母集団の母平均値が一定の確率でその区間内にはいるという概念であり、かなりわかりにくい概念である。そこでこの上限値という値を正確ではないが一般的にわかるように言い換えると、そのような値以上の値の出現する確率は 95%信頼区間の時は 20 回に 1 回、99%信頼区間の時には 100 回に 1 回であると認識してもらって良いであろう。

3. 分子量が一定値以上の物質における $\log BCF > 3$ の出現予測値の検討

表 1 と表 2 から、次のように考えられる。

例えば分子量を $MW > 500$ とするとき 412 の物質が示されており、これら物質中には $\log BCF > 3$ のものは存在しない(0 個である)。 $n=400$ として表 2 を見ると、95%信頼区間の上限値は 0.92%、99%信頼区間の上限値では 1.32%である。言い換えると、今後 $MW > 500$ の物質が作られたとき、期待値的には $\log BCF > 3$ の物質は 0%である。しかし、信頼区間の概念からは、 $MW > 500$ の物質の中で $\log BCF > 3$ のものが現れる確率は、95%の信頼率では多くても 0.92%程度くらい、99%信頼率では 1.32%程度と見積もれることを示している。

しかしながら、得られている原データの数 $n=1503$ の内、 $MW > 500$ の化合物は 412 個であり、その出現頻度は $412 \div 1503 = 0.2741$ と示せる。すなわち、検討対象物質の内 $MW > 500$ は、約 27%であり、その内の 0.92%が、 $\log BCF > 3$ なのであるから、

$$412 \div 1503 \times 0.0092 = 0.0025218$$

なる計算をしたとき、「 $MW > 500$ でかつ $\log BCF > 3$ のものが出現する確率」は、全体に対して約 0.25%となり、十分に小さな数値といえる。

表 1 の数値の現れ方および原データの散布図を見ると、明らかに分子量が大きくなるに連れて $\log BCF > 3$ の出現する傾向は小さくなっている。故に、 $MW > 700$ の $\log BCF > 3$ の出現する確率は、上記の様に $MW > 500$ の物質とする時よりも、さらに低いものと見られる。

しかしながら、単純確率計算では、 n の数が大きいほど計算で得られる信頼区間の幅は狭くなり、逆に n の数が少なければ信頼区間の幅は広くなる。その為、上記の $MW > 500$ の物質に対する推論が $n=412$ であるのに対して、 $MW > 700$ とすると $n=197$ なので表 2 の $n=200$ の所を見ることになり、信頼区間の幅はより広がっている。

ここで問題とされている $MW > 700$ の物質が $\log BCF > 3$ を示す確率は、本来、表 1 のような数値の現れ方の傾向(2次元分布を考えたときの出現率的な値)と表 2 の二項分布の信頼区間とを組み合わせたものとして算出されるものであり、そのような値が実際現場における大きな分子量の物質における $\log BCF > 3$ の出現予測値となるべきであろう。しかし、寡聞にして現行の統計的推論において、そのような複合確率を簡単に計算する手法はない。

そこで、先に述べたような確率計算により、2つの確率値を組み合わせる方法を採用した。

故に、ここでは上記のような推論と、MW>700の物質数から、先と同じ計算を行ない

$$197 \div 1503 \times 0.0183 = 0.0023986$$

なる値が得られる。しかし、この値は表2との関連で述べたごとく、「大きめに」評価された値と考えられる。

さらに、安全サイドをとり、MW>800や、MW>1000なども検討できるが、この安全側の余裕は表1データから得られるMW>500の値より考え、MW=700またはMW=800程度で十分であろう。

私見としては、得られた値をも参考にするとき、MW=700を用いて「MW>700の物質がlogBCF>3の値を示す確率は十分少ない」と結論したい。

文献1) K.diem & C Lentner: Scientific tables, 7th ed., CIBA-Geigy Ltd., 1971