3.5.3 ベンフォードの法則

「ベンフォードの法則」ってご存じですか？ ある数値の最初の桁の数は1が最も多く、その先頭の桁の数が増えるごとに出現割合は徐々に減っていき、9を先頭とする数が最も少ないという法則です。確かめてみましょう。

問. 2^n において桁数を m とするとき最初の桁の数が1になるときの区間を m を用いて表しなさい。またこのときの n の範囲を m を用いて表しなさい。

$$1 \times 10^{m-1} \leq 2^n < 2 \times 10^{m-1}$$

底が10の対数をとると

$$\log(1 \times 10^{m-1}) \leq \log 2^n < \log(2 \times 10^{m-1})$$

$$\log 10^{m-1} \leq n \log 2 < \log 2 + \log 10^{m-1}$$

$$\frac{m-1}{\log 2} \leq n < \frac{\log 2 + m-1}{\log 2}$$

上の式から最初の桁の数が1になる区間幅は桁数 m とは関係なく1になることがわかります。同様にして最初の桁の数が2になる区間幅はどうなるのでしょうか？

$$2 \times 10^{m-1} \leq 2^n < 3 \times 10^{m-1}$$

底が10の対数をとると

$$\log(2 \times 10^{m-1}) \leq \log 2^n < \log(3 \times 10^{m-1})$$

$$\log 2 + \log 10^{m-1} \leq n \log 2 < \log 3 + \log 10^{m-1}$$

$$\frac{\log 2 + m-1}{\log 2} \leq n < \frac{\log 3 + m-1}{\log 2}$$

最初の桁の数が2の区間幅は $\frac{\log 3 - \log 2}{\log 2}$ になります。最初の桁が a になるときの区間幅を x_a としてまとめてみると

区間	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
区間幅	1	$\frac{\log 3 - \log 2}{\log 2}$	$\frac{\log 4 - \log 3}{\log 2}$	$\frac{\log 5 - \log 4}{\log 2}$	$\frac{\log 6 - \log 5}{\log 2}$	$\frac{\log 7 - \log 6}{\log 2}$	$\frac{\log 8 - \log 7}{\log 2}$	$\frac{\log 9 - \log 8}{\log 2}$	$\frac{1 - \log 9}{\log 2}$

区間幅の合計が $\frac{1}{\log 2}$ になることから、それぞれの区間の割合を求めてみると

区間	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
割合	$\log 2$	$\log 3 - \log 2$	$\log 4 - \log 3$	$\log 5 - \log 4$	$\log 6 - \log 5$	$\log 7 - \log 6$	$\log 8 - \log 7$	$\log 9 - \log 8$	$1 - \log 9$
近似値	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

3^n について同様にまとめ割合を求めると

区間	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
区間幅	$\frac{\log 2}{\log 3}$	$\frac{\log 3 - \log 2}{\log 3}$	$\frac{\log 4 - \log 3}{\log 3}$	$\frac{\log 5 - \log 4}{\log 3}$	$\frac{\log 6 - \log 5}{\log 3}$	$\frac{\log 7 - \log 6}{\log 3}$	$\frac{\log 8 - \log 7}{\log 3}$	$\frac{\log 9 - \log 8}{\log 3}$	$\frac{1 - \log 9}{\log 3}$

区間幅の合計が $\frac{1}{\log 3}$ になることから、割合は 2^n のときと同じということがわかる。

こんな問題を解かせると対数に対する理解が深まるんじゃないかなあ～。

3.5.3.1 ベンフォードの法則を感じるための数値データ

授業で扱う場合には生徒が持っている理科や社会の資料集の中の数値データを最初に扱うと生徒の興味関心が増すだろう。実際 2012 年の数学セミナーには理科年表の基礎物理定数や加工食品の生産量、OECD 内の国内総生産等の具体的な数値データが載っています。

何かの数値データの先頭の数分布を調べさせた後、その分布の様子(最上位の数)が現在学習している対数を用いた数で表される驚き、そしてその法則を自分の力で証明することができることいろいろな可能性をもつ教材です。

例えば下のデータは私が (m,k) -完全数の m をみつけるために必要になった約 20000 個の 7 桁 ~ 229 桁の素数の先頭の数です。

先頭数字	個数	割合	理論値	誤差
1	6046	0.3102	0.3010	0.0092
2	3510	0.1801	0.1761	0.0040
3	2386	0.1224	0.1249	-0.0025
4	1847	0.0948	0.0969	-0.0021
5	1532	0.0786	0.0792	-0.0006
6	1278	0.0656	0.0669	-0.0013
7	1056	0.0542	0.0580	-0.0038
8	971	0.0498	0.0512	-0.0014
9	866	0.0444	0.0458	-0.0014
合計	19492	1.0001	1	0.0001

もっと基本的なデータとして初項 1 公比 2 の等比数列(整数列大辞典: A000244)がある。下のデータはこの数列の 100 項までと 1000 項までをまとめた表である。

先頭数字	第 100 項まで				第 1000 項まで			
	個数	割合	理論値	誤差	個数	割合	理論値	誤差
1	30	0.3	0.30	0	301	0.301	0.301	0
2	17	0.17	0.18	-0.01	176	0.176	0.176	0
3	13	0.13	0.12	0.01	125	0.125	0.125	0
4	10	0.1	0.10	0	97	0.097	0.097	0
5	7	0.07	0.08	-0.01	79	0.079	0.079	0
6	7	0.07	0.07	0	69	0.069	0.067	0.002
7	6	0.06	0.06	0	56	0.056	0.058	-0.002
8	5	0.05	0.05	0	52	0.052	0.051	0.001
9	5	0.05	0.05	0	45	0.045	0.046	-0.001
合計	100	1	1.01	-0.01	1000	1	1	0

この法則は意外とすごいと感じました。数は 1 に支配されているということなのか…。

(参考文献 数学セミナー 2012 年 8 月号, 2013 年 1 月号)