正規分布の式に近づくくふうの方法

飯島光治（埼玉）

はじめに

正規分布の式 $y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ を数学的に導くのは、大変で証明をフォローするのもです。小倉金之助（1885～1962）先生の「統計的研究法」（1925刊）には、実験式を用いて、帰納的に、正規分布の式の近似式を、導いています。これは教育的と思い、ご紹介します。

誤差の法則というのがあり、種々のデータより、誤差の表す曲線は、y 軸に関して山型の対称形になることが知られている。一誤差の少ないデータは多く、誤差の大きいデータは、少なくなることもあり一一方 y 軸に関して山型に対称になる曲線に、$y=2^{-x^2}$ があり一般形は、$y=A\cdot B^{-x^2}$ が考えられる。このyが実験式。

ここで例として、イギリスの男子の身長

19字×20行（3枚で1頁）

8585人のデータがある。（1883年、明治16年）

N＝8585、平均M＝67.46イチ、

σ＝2.57イチ（注 1イチ≒2.5cm）

ここでM＝67.5　σ^2＝6.6とする。

また身長の平均Mを、座標の横軸の基準にとるために、d＝X－M（Xは各人の身長）を考える。

以下「身長の分布が、$f = A \cdot B^{-d^2}$の形で表せるか。もし表せたら定数A、Bをいかに定めるか」が問題となる。

この本には、右表がある。

d	f	d	f
0	1330	±6	81
±1	1227	±7	39
±2	1027	±8	15
±3	658	±9	5
±4	393	±10	2
±5	186		

<表1>

（注）この表で±1と一しで各1ヒトカタとして以下合計すると、8592人です。

d＝0とすると

f＝A＝1330。

ここでNが大きければ、Aも多くなり、σ（標準偏差）が小さければ、Aが多くなると考えられるから

Aは、Nに比例し、σに反比例すると、仮

19字×20行（3枚で1頁）

定すると、即ち $A=\frac{KN}{\sigma}$（K:定数）

とおくと $\frac{A\sigma}{N}=K$ となり定数となる。

この時は、$\frac{A\sigma}{N}=\frac{1330\times 2.57}{8585}=0.398\fallingdotseq 0.4$

より、$A=\frac{0.4N}{\sigma}$ ①

次にBの値を求めるのに、dを$\frac{d}{\sigma}$とした方が便利。$\frac{d}{\sigma}$とすると、2つのデータの分布を比べる時、dのままだと2つのデータの大きさや単位が違う時の問題が、解消される。なぜなら$\frac{d}{\sigma}$は、単位のつかない数（無名数）となり、同一の基準で比較できるから。

$f=A\cdot B^{-d^2}=A\cdot(B^{\sigma^2})^{-(\frac{d}{\sigma})^2}$ として

$C=B^{\sigma^2}$ とおくと、$f=A\cdot C^{-(\frac{d}{\sigma})^2}$ ②

この両辺の10を底とする対数をとると

$\log f=\log A-(\frac{d}{\sigma})^2\log C$ より、

$\log C=(\frac{\sigma}{d})^2(\log A-\log f)$

次に、各dについて下線部分を計算する。

たとえばd=±1のとき $f=1227$（注）より、

$6.6(\log 1330-\log 1227)=6.6\times\log\frac{1330}{1227}$

$=6.6\times\log 1.08=6.6\times 0.0334\fallingdotseq 0.22$

5　10　15　19

19字×20行（3枚で1頁）

以下同様にすると右表になる。

右表で、±9, ±10の時を除いて、下線部分の平均をとると 0.21

d	下線部分	d	下線部分
±1	0.22	±6	0.22
±2	0.19	±7	0.21
±3	0.22	±8	0.20
±4	0.22	±9	0.12
±5	0.23	±10	0.12

<表2>

(注) 0.22は小数第3位四捨五入。以下同じ。

∴ $\log C = 0.21$

$C = 10^{0.21}$ ③

①③③より

$$f = \frac{0.4N}{\sigma} \cdot 10^{-0.21\left(\frac{d}{\sigma}\right)^2}$$

ここで $10^{0.21} = (10^{0.42})^{\frac{1}{2}}$ とし

$a = 10^{0.42}$ とおくと $\log a = 0.42$

対数表より $a = 2.63$

$$f = \frac{0.4N}{\sigma} \times 2.63^{-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2}$$

$0.4 = \frac{4}{10} = \frac{1}{2.5} = \frac{1}{\sqrt{6.25}}$ とすると

$$f = \frac{N}{\sqrt{6.25}\cdot\sigma} \times 2.63^{-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2}$$ Ⓐ

ここで正規分布の本当の式は、

$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ Ⓑ と紹介します。

Ⓐで、$y = \frac{f\sigma}{N}$, $x = \frac{d}{\sigma}$ とおくと

Ⓐは $y = \frac{1}{\sqrt{6.25}} \times 2.63^{-\frac{x^2}{2}}$ となる。

19字×20行（3枚で1頁）

Ⓑと比べると

$\sqrt{2\pi} \fallingdotseq \sqrt{2 \times 3.1416} = \sqrt{6.2832} = 2.5066\cdots$

$e \fallingdotseq 2.7183$ より、$2.7183 - 2.63 \fallingdotseq 0.09$ と誤差は、わずかになります。

よって、Ⓐは本当の式Ⓑの近似式といえます。

（注）表1の元のデータをみると $d=1$ と $d=-1$ の f の計は、2453です。$f=1227$ は、$2453 \div 2$ ですが、この時期数が、Bの形を定める（Cといいかえていますが）なので、下線部分の式の形より $d=1$ の時も、$d=-1$ の時も同じ計算となるので、片一方のみでよい。以下同様です。求める式の形を、理想的な山型の対称形と仮定しているので、平均の数 $f=1227$ でしています。

（追記1）表1、表2でこの本の数字を修正した所があります。この本の執筆時は

電卓は、ないので小倉先生は、手計算でしたのでしょうか。大変です。私は電卓で、計算していて、結果が不安でもあり、楽しみでもありました。

対数を習った高校生に、以上の内容は指導可能に思えます。

（追記2）

$y=\frac{f\sigma}{N}$ とおきかえています。ここでNとσは、定数なので、yはfの定数倍で、fに多いに関連します。

（2009年9月記）

ご質問、ご意見等ありましたら下記まで。

〒344-0041
埼玉県春日部市増富763-1
飯島　光治

19字×20行（3枚で1頁）