頻數曲線の變換について

小河原, 正巳
中央氣象臺調査課

出版情報：統計数理研究. 1 (2), pp.91-97, 1942-03-15. Research Association of Statistical Sciences

頻數曲線の變換について

小河原　正　巳（中央氣象臺調査課）

（昭和十六年九月廿五日受理）

1. 緒論. 或る量 X を觀測測定した値 x の度數，又は X に關聯した他の性質 Y の測定値 y に對しては，種々の曲線を當嵌めたり或は Charlier の級數などで近似して偶然の誤差を除去し，所謂統計的な法則を解析的に $y=F(x)$ なる形で表現することが普通に行はれる．このうち頻度分布として最も基本的で又理論的に取扱易いものは Gauss の正規分布であらう．

Gauss の分布では x は $(-\infty,\infty)$ の範圍で起り得るのであるが，實際には測定値 x の範圍は制限されてゐるのが普通である．例へば身長・年齢・溫度・質量など，大抵の場合には x の範圍は左に有界な區間 (a,∞) である．又雲量は $(0,10)$ の間で測定し試驗成績の評點の區間は $(0,100)$ などが普通であつて，之等は有界な區間の例である．從つて通例は Gauss の分布はあくまで近似的な意味を失はない．

しかし一般に觀測測定の結果が正規分布にならない場合にも X を測る物指の各部分を適當に伸縮すれば，即ち任意の區間 (a,b) の變數 x に或る位相的變換

$$\xi=f(x) \tag{1.1}$$

を施すことによつて，ξ が $(-\infty,\infty)$ に於て近似的に正規分布に從ふ樣にすることができると考へられる．逆に言へば，正規曲線

$$\eta=\Phi(\xi)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(\xi-m)^2}{2\sigma^2}} \tag{1.2}$$

に (1.1) なる變換をした

$$y=\Phi(f(x))f'(x) \tag{1.3}$$

を實驗公式とする方法も考へられる．

頻數曲線の變換の一例として對數變換 $\xi=\log x$ がある．これは $\log x$ そのものが正規分布に從ふとき，これを $(0,\infty)$ に於ける $x=e^{\xi}$ の頻數函數に變換するものであつて，それを母函數とする Charlier 級數が人口統計の Graduation などに用ひられてゐる.(1)

又 R.A. Fisher(2) は小標本の場合の相關係數 r の有義性を判定するために，所謂 Fisher の z として $z=\frac{1}{2}\{\log(1+r)-\log(1-r)\}$ なる變換を用ひた．即ちこれによれば，區間 $(-1,1)$ の r は $(-\infty,\infty)$ の z に位相的に變換されて，標本の數が少い場合にも又原集團の相關係數が ± 1 に近い樣な場合にも z の頻數分布は近似的に正規型になることに因つて，r の誤差を z の誤差として判定することができる．

こゝでは Fisher の變換を一般の場合に擴張して，有限な區間に於ける任意の頻數分布を Gauss の正規分布から導くことに就いて考へ見よう．

2. 變換式の一般形. 區間 $(-1,1)$ を $(-\infty,\infty)$ に位相的に移す變換式 $f(x)$ は一般に $(-1,1)$ に

於ける狹義の單調增加函數で

(2.1) $$f(-1+0)=-\infty,\quad f(1-0)=+\infty$$

であればよい．これに $f(x)$ の開區間 $(-1,1)$ に於ける微分可能性を附加すれば (2.1) の他に

(2.2) $$f'(x)\geqq 0$$

ならばよいことになる．こゝに等號は有限個の點だけで許されるものである．從つて $f'(x)$ は次の形でこれを表はすことができる．

(2.3) $$f'(x)=\frac{\varphi(x)}{(1-x)^m(1+x)^n}.$$

こゝに

(2.4) $$m,n\geqq 1$$

(2.5) $$\varphi(x)\geqq 0\quad(-1<x<1)$$

(2.6) $$\varphi(-1+0)>0,\quad \varphi(1-0)>0.$$

何となれば (2.1), (2.2) から

$$f'(-1+0)=f'(1-0)=+\infty.$$

故に $(1-x)^m f'(x)=\varphi_1(x)$ とおくとき $\varphi_1(1-0)>0$ なる如き $m\geqq 0$ が存在する．次に $(1+x)^n f'(x)=\dfrac{\varphi(x)}{(1-x)^m}$ とおけば $\displaystyle\lim_{x\to-1+0}\frac{\varphi(x)}{(1-x)^m}=\frac{1}{2^m}\varphi(-1+0)>0$，從つて $\varphi(-1+0)>0$ なる如き $n\geqq 0$ が存在する．然るに $\varphi(x)=(1+x)^n\varphi_1(x)$ であるから $\varphi(1-0)=2^n\varphi_1(1-0)>0$ であつて，この $\varphi(x)$ が (2.5) を滿足することは明かである．さて吾々の假定によれば，$f'(x)$ は $(-1+\varepsilon,1-\varepsilon)$ (但し $\varepsilon>0$) で有界從つて積分可能であるが，(2.1) が成立つためには $f'(x)$ の積分が $(-1,x_1)$, $(x_1,1)$ (但し $-1<x_1<1$) で發散することが必要十分である．その條件は $x=1$ の附近では $f'(x)>A(1-x)^{-\lambda}$, $A>0$, $\lambda\geqq 1$ なる λ が存在することであり，又 $x=-1$ の附近でも同じことが言はれるから，結局 (2.5), (2.6) を滿足する $\varphi(x)$ と (2.4) なる m,n によつて $f'(x)$ は (2.3) なる形で表はされるのである．

故に吾々の變換式の一般形は (2.2) を積分することによつて得られるのであるが，この式で $\varphi(x)$ は有限個の點を除いては正で (2.6) を滿足するものであれば，全く任意の一價函數でよい．$\varphi(x)$ が零となる點では變換された頻數曲線 (1.3) は零となり，この點で曲線が分離される．

吾々は (2.3) に於ける最も簡單な次の樣な場合を採つて見よう．即ち

$$m=n=1,\quad \varphi(x)=a>0$$

とおけば

(2.7) $$f(x)=\frac{a}{2}\log c\frac{1+x}{1-x}$$

となり，之によつて (1.2) から變換された頻數曲線は

(2.8) $$y=\frac{a}{\sqrt{2\pi}(1-x^2)\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left(\frac{a}{2}\log c\frac{1+x}{1-x}-m\right)^2\right\}$$

となる．この變換は $(-1,1)$ の點を射影的に $(0,\infty)$ の點に移し，更に $(0,\infty)$ の點を對數變換によつて $(-\infty,\infty)$ の點に移すものであつて，所謂 Fisher の Z 變換に外ならない．

3. **常數の決定．** (2.8) に於て $\log c-\dfrac{2m}{a}=-\log\kappa$, $\dfrac{2\sigma}{a}=q$ とおけば，(2.8) は (1.2) に於て m

$=0$, $\sigma=1$ としたものに變換

$$\xi=\frac{1}{q}\log\frac{1}{\kappa}\frac{1+x}{1-x} \tag{3.1}$$

を施したものとなり，(2.8) は

$$y=\frac{2}{\sqrt{2\pi}(1-x^2)q}\exp\left\{-\frac{1}{2q^2}\left(\log\frac{1}{\kappa}\frac{1+x}{1-x}\right)^2\right\} \tag{3.2}$$

となる．又 (1.2) に於て $m=\log\kappa$, $\sigma=q$ ととつておけば

$$\xi=\log\frac{1+x}{1-x} \tag{3.1$'$}$$

によつて (1.2) から (3.2) が得られる．何れにしても變換された頻數曲線は常數 κ と q を決定することによつて定まる．尚 (3.1) に於て $x=p$ のとき $\xi=0$ とすれば $\kappa=\dfrac{1+p}{1-p}$ となる．

さてこの κ (又は p) と q の値を與へられた觀測値から定めるために Maximum Likelihood 法を採用するならば，觀測値 $x_i\,(i=1,2,\cdots,N)$ に對して

$$\Pi y_i=\left(\frac{2}{\sqrt{2\pi}}\right)^N\frac{1}{q^N}\left(\Pi\frac{1}{1-x_i^2}\right)\exp\left\{-\frac{1}{2q^2}\sum\left(\log\frac{1}{\kappa}\frac{1+x_i}{1-x_i}\right)^2\right\}$$

を Maximnm ならしめればよい．

$$S=\sum\log y_i$$

とおけば，$\dfrac{\partial S}{\partial\kappa}=0$ から

$$\log\kappa=\frac{1}{N}\sum\log\frac{1+x_i}{1-x_i}. \tag{3.3}$$

又 $\dfrac{\partial S}{\partial q}=0$ から

$$\begin{aligned}q^2&=\frac{1}{N}\sum\left(\log\frac{1+x_i}{1-x_i}-\log\kappa\right)^2\\&=\frac{1}{N}\sum\left(\log\frac{1+x_i}{1-x_i}\right)^2-(\log\kappa)^2,\end{aligned} \tag{3.4}$$

この $\log\kappa$ 及 q の値は x_i を (3.1′) で變換した ξ_i の頻數分布を Maximum Likelihood 法で (1.2) で近似したときの m 及 σ の値と一致する．即ち Maximum Likelihood はこの變換によつて不變であることも分る．

尚實用的には次の樣な方法も考へられる．即ち $x=p$ は $\xi=0$ に對應し，又吾々の變換が相位的であることから $\Pr.[x\leqq p]=\Pr.[\xi\leqq 0]=\frac{1}{2}$．故に p を x_i の中央値にとることができる．次に p_1, p_2 を x_i の上下の四分の一限界即ち $\Pr.[x\leqq p_2]=\Pr.[x\leqq p_1]=\frac{1}{4}$ とすれば，$x=p_1$, $x=p_2$ は夫々 $\xi=0.6745$, $\xi=-0.6745$ に對應するから (3.1) によつて

$$q_1=\frac{1}{0.6745}\log\frac{1}{\kappa}\frac{1+p_1}{1-p_1},\quad q_2=-\frac{1}{0.6745}\log\frac{1}{\kappa}\frac{1+p_2}{1-p_2}.$$

これ等の平均を以て q とすれば

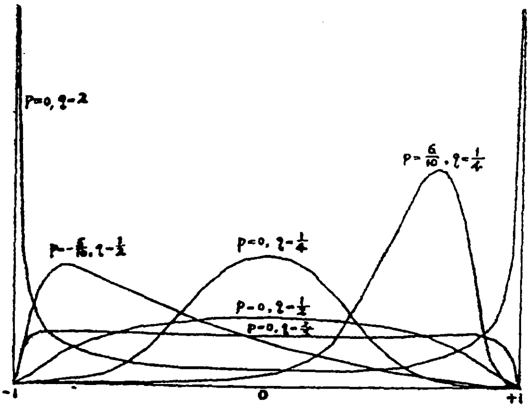
$$q=\frac{1}{1.35}\log\frac{1+p_1}{1-p_1}\frac{1-p_2}{1+p_2}. \tag{3.5}$$

觀測値に對應する ξ_i が完全に正規型である場合には，こゝに求めた p, q は (3.3), (3.4) から求めれらるその値と一致する．このことは先づ p については明かである．次に q についても (3.3), (3.4) から求められる $\log\kappa, q$ を夫々 m, σ とすれば，$\Pr.[\xi\leqq\xi_0]=\Pr.[x\leqq x_0]$ なる ξ_0 と x_0 には常に (3.1′) なる關係があるから

$$m+0.6745\,\sigma=\log\frac{1+p_1}{1-p_1}\,,\quad m-0.6745\,\sigma=\log\frac{1+p_2}{1-p_2}\,.$$

故にこの σ は (3.5) の q に等しい．

さて變換された頻數曲線 (3.2) は p, q の二三の値に對して第一圖の樣になる．これ等の曲線の何れに於ても $\lim\limits_{x\to-1+0}y=\lim\limits_{x\to1-0}y=0$ であつて，各曲線は互ひに連續的に他に移り得るものである．例へば $p=0,\ q=2$ の如き曲線も $x=\pm1$ に極めて近い所で極大になる．之等の極大極小は (3.2) から直ちに分る樣に，$y_p=\log_e\dfrac{1}{\kappa}\dfrac{1+x}{1-x}$ なる曲線と $y_q=q^2x$ なる直線との交點によつて圖計算で求めることができる．



第 1 圖

4. 應用例. 昭和16年3月中央氣象台附屬氣象技術官養成所入學試驗の東京に於ける受驗者114名の成績は次の表の通りであつた．各課目の頻數分布は第二圖の細線に見る樣に，その大體の趨勢は

點數＼課目	數學	物理	化學	地理	英語	作文
0 —— 4	17	12	5	—	7	—
5 —— 9	9	5	4	1	24	—
10——14	9	4	5	1	10	—
15——19	11	3	4	—	17	—
20——24	7	4	3	3	6	—
25——29	4	6	3	11	9	—
30——34	8	7	6	14	6	—
35——39	13	2	9	11	5	—
40——44	4	8	6	14	6	—
45——49	2	6	8	13	5	—
50——54	6	6	5	12	5	1
55——59	4	5	4	9	3	—
60——64	4	9	7	7	3	7
65——69	3	4	3	8	4	12
70——74	7	5	11	6	2	13
75——79	1	11	7	3	2	13
80——84	2	5	5	1	—	27
85——89	2	4	6	—	—	25
90——94	1	1	6	—	—	15
95——100	—	7	7	—	—	1
計	114	114	114	114	114	114
平均	33	49	54	47	26	81
中央値	30	50	54	46	19	82
標準偏差	24.6	28.3	28.0	15.7	20.7	9.5

物理と化學が稍似てゐる外夫々異つた型を示してゐる.

今受驗者の眞學力は實數 ξ を以て測定し得るもので，受驗者の集團及試驗課目に從つて，原點と測定の單位とを適當にとるならば，ξ の頻數分布は Gauss の正規曲線 $\eta=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\xi^2}$ に從ふものと假定すると，眞學力の評價點 x は，試驗問題によつて ξ を或る有限な區間 $(-1,1)$ に位相的に變換したものと考へることができよう．斯様な變換式として，上に述べた最も簡單な場合即ち (3.1) を採用して見よう．

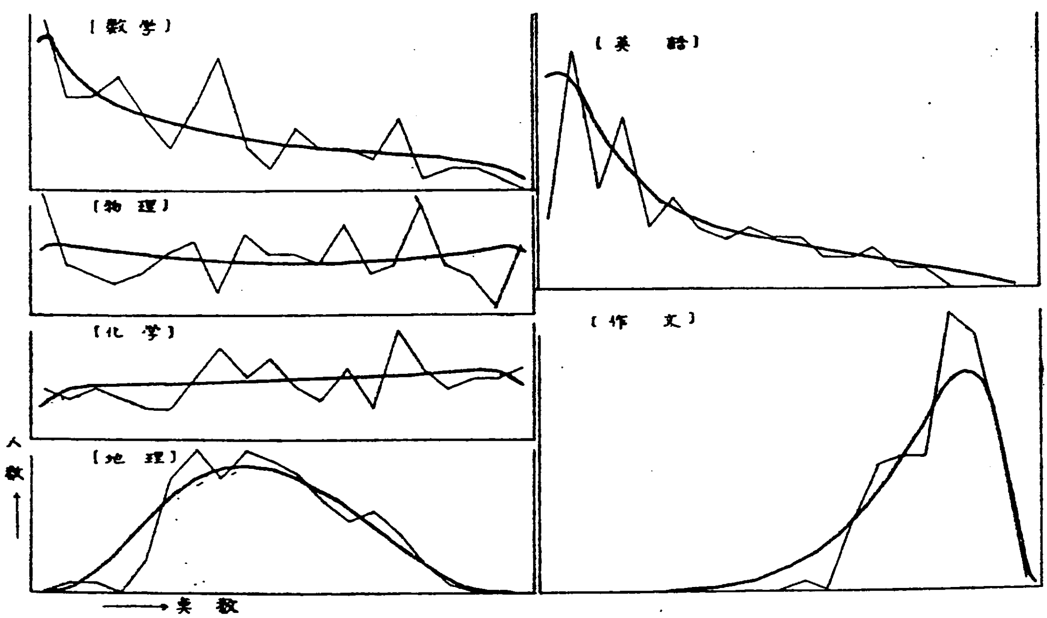
上の表に於て點數の全區間 (0, 100) を $(-1,1)$ にとり，中央値を p とし，q を (3.5) で計算すれば p,q の値は次の様になる．

	數學	物理	化學	地理	英語	作文
p	−0.4	0.0	0.1	−0.1	−0.6	0.6
q	0.77	0.77	0.67	0.33	0.69	0.34

これに對應する曲線が第2圖の太線である．これに對して Pearson の χ^2 驗査を行へば

	數學	物理	化學	地理	英語	作文
χ^2	14.7	10.2	8.1	3.2	5.0	10.9
P	0.04	0.25	0.42	0.53	0.52	0.05
自由度	8	9	9	5	7	6

となり，數學と作文は稍惡いが實際との差は大體に於て小標本に因る誤差の影響と看做すことができよう．



第 2 圖

尚この場合に於いて常數 p,q の演ずる役割は (3.2) からも明かな様に，p の大小は曲線の右左への片寄りを表はし，又 q が大きい課目では滿點に近い者或は零點に近い者が可成あり得るが，q の小さい課目では反對に兩端は低く中程に山ができる．從つてこの例の如き問題に於ては，p で受驗者に對する試驗問題の相對的な難易を測り，q の大小で問題の相對的常識性とでも言ふべきもの ゝ

少い多いを測ることができよう.

5. **補遺**. 吾々は (2.3) に屬する最も簡單な一つの變換式について述べたのであるが, 前の例の如き問題では大體之でも間に合ふ樣に思はれる.

有限區間といふ點を固執して Pearson 型の曲線を選ぶならば Type I, II, VIII, IX, XII 等[3] であるが, 何れにしても Pearson の曲線も凹凸の點から見れば矢張り單純なものであるから, 四次の moment まで考慮する割合に適合度が著しくよいとは考へられない.

例へば前の例の英語に Pearson 型の曲線を當嵌めて見るならば $\beta_1=0.76$, $\beta_2=2.65$ となり, Type I 曲線の J 字型に屬する. 即ちその頻數曲線は

$$y_P=\frac{114\,\Gamma(1.8)}{3.8\times 21\,\Gamma(0.5)\,\Gamma(1.3)}\left(\frac{1.1+t}{3.8}\right)^{-0.5}\left(\frac{2.7-t}{3.8}\right)^{0.3} \tag{5.1}$$

$$t=\frac{x-26}{21}\quad(0<x<100)$$

となる. 又之を Poison-charlier 級數で近似すれば

$$y_c=114\,[\psi(t)+0.9\,\Delta^2\psi(t)] \tag{5.2}$$

が得られる. こゝに

$$\psi(t)=\frac{e^{-2.6}(2.6)^t}{t!}$$

$$t=\frac{x}{10}\quad(0<x<100)$$

であつて, 又 $\Delta^2\psi(t)$ は $\psi(t)$ の二次の階差即ち

$$\Delta^2\psi(t)=\psi(t)-2\psi(t-1)+\psi(t-2)$$

である.

今 (3.2) による y と (5.1) による y_p 及 (5.2) による y_c とを比較すれば

點　數	0—9	10—19	20—29	30—39	40—49	50—59	60—69	70—	計	χ^2	P	自由度
實測度數	31	27	15	11	11	8	7	4	114			
y	33.8	24.1	15.2	11.0	8.4	6.8	5.5	9.2	114	5.0	0.66	7
y_P	46.9	18.7	13.2	10.4	8.4	6.9	5.5	4.0	114	10.7	0.16	7

點　數	0	10	20	30	40	50	60	70以上	計	χ^2	P	自由度
實測度數	7	34	23	15	11	10	6	8	114			
y	16.6	30.6	19.1	12.9	9.6	7.5	6.0	11.7	114	9.3	0.24	7
y_c	16.1	26.6	22.4	15.4	11.7	9.2	6.3	6.1	114	8.0	0.34	7

こゝに y_c との比較に於ては區間の分割が變つたために y の適合度の檢査の結果は稍惡くなつた. この一例では y_p は一番惡くなつてゐるが, これによつても適合度の點では何れも大差がない樣に思はれる.

標本の數が多く, しかもその頻數分布が單純な型を示してゐない樣なものに對しては, それに對應する複雜な曲線で近似する必要が起る. その爲めには, 變換式を更に常數の多いものにとるか, 變換する元の曲線をもつと一般的なものにとつておくか, 或は又, 變換された (3.2) の如き函數を

母函数とする charlier 型級數等で近似するといふ樣な方法も考へられる．但し，最後の方法では moment の計算が複雜になるかも知れない．

今 (2.3) を比較的簡單な函數にとつて得られる二三の變換式を擧げれば

$f'(x)$	$f(x)$
$\dfrac{a}{(1-x^2)^{\frac{3}{2}}}\quad(a>0)$	$\dfrac{ax}{\sqrt{1-x^2}}-\dfrac{ap}{\sqrt{1-p^2}}$
$\dfrac{a+b(1-x^2)}{1-x^2}\quad(\lvert b\rvert<a)$	$\dfrac{a}{2}\log\dfrac{1-p}{1+p}\dfrac{1+x}{1-x}+b(x-p)$
$\dfrac{a}{(1+x)(1-x)^2}\quad(a>0)$	$\dfrac{a}{4}\log\dfrac{1-p}{1+p}\dfrac{1+x}{1-x}+2a\left(\dfrac{1}{1-x}-\dfrac{1}{1-p}\right)$
$a_1\sec^2\frac{\pi}{2}x\left(=\dfrac{(1-x^2)^2\sec^2\frac{\pi}{2}x}{(1-x^2)^2}\right)\quad(a_1>0)$	$a(\tan\frac{\pi}{2}x-\tan\frac{\pi}{2}p)$
等	等

要するに，同種の觀測から得られる頻度曲線を，一定の基本曲線から變換し，その變換式の補助變數の値でそれを特徴づけ，比較することもできる樣に思はれる．又逆に，一般の曲線を正規曲線に變換して論ずれば，その解析的取扱ひが便利になる，變換式 (3.1) の場合には (3.3),(3.4) を補助變數とする正規分布を考へればよい．

最後に，この小論を草するに當り，北川博士並に増山元三郎氏に多大の御教示を賜つたことを茲に記して謝意を表したい．

註

(1) A. Fisher: The mathematical theory of probabilities, Vol. I, 1936 Chap. XVI 參照.

(2) R. A. Fisher: Statistical methods for research workers, 6th ed. 200頁, P. R. Rider: A survey of the theory of small samples, Annals of Math. Vol. 31, No. 4, 1930, 589頁.

(3) 成實清松: 數理統計學，岩波講座，30頁參照.