「カイ自乗検定法」用法に関する図表と模形実験

上 田 常 吉

奈良医大解剖学教室

DIAGRAMS AND MODEL EXPERIMENTS FOR USE OF THE
TESTING OF GOODNESS OF FIT

TSUNEKICHI UEDA

Nara Medical School

緒 言

生物学には、理論的に計算せられた数列 A と、これに対応し一致する筈の統計的数列 B とを、比較研究することが屢々ある。併し B 数列は現実の数列であるから、必ず誤差を伴なう。その誤差が偶然にも起りうる誤差であるか、又は偶然には決して起り得ず、何か数値を狂わせる原因が存在するのかを判定する必要がある。その方法に適合検査の法（カイ自乗検定法）がある。（Test of goodness of fit, χ^2 -test）。この法は 1900 年 Pearson により考案せられ、Elderton により詳細な数表が発表せられ、其後 Yule, Fisher, Neyman, 西木等により或は計算を補充せられ、或は利用法について論ぜられた。理論数列と統計数列との比較には最もよい完全に近い方法で、遺伝学や一般生物学に利用の向が大きいと思うが、尙お一般に誤解せられている点もあるからそれを実験で正し、又 χ^2 の数表でなく計算図表を呈供して、研究の資料としたいと思う。図表は一度朝鮮医学雑誌に発表せられたのであるが、一般に知られるに至らなかつたので再び印刷に付した次第である。

第 I: χ^2 検定法とその図表の用法

1. カイ自乗

今理論的数列 $m_1 m_2 m_3 \dots m_n$ (A 数列) と、これに対応する統計的数列 $m'_1 m'_2 \dots m'_n$ (B 数列) とがあり、この両者が一致することを期待しているが、両者数値の末まで一致することは先づ無い。併しその差違が偶然にも起りうる程度であるならば一致していると判定し、偶然には決して起らない程の大きな違いならば一致していないと判定する。この判定の爲めに χ^2 (Chi-square) を計算する。 χ^2 とは相対応する理論値及び統計値の差（誤差）の平方を、理論値で除したものの総和であつて、即ち

$$\chi^2 = \sum_i \frac{(m_i - m'_i)^2}{m_i}$$

又は

$$= \frac{(m_1 - m'_1)^2}{m_1} + \frac{(m_2 - m'_2)^2}{m_2} + \dots + \frac{(m_n - m'_n)^2}{m_n}$$

二数列よく一致して居れば χ^2 小であるが、漸次不一致となるに従い χ^2 大となること勿論である。この一致と不一致は「二数列一致する確率 P 」で表わす。 P 大なれば一致する確率が大きいから常識的には一致すると考え、 P 小なばら一致しないと言ひ得る。勿論中間的の値ならば何とも判定出来ないから尙お将来の研究を待つこととなろう。さて P の大きさは χ^2 の大きさに左右せられるのみならず、又自由度 n によつても変る。即ち P は χ^2 と n との函数である。

$$P=f(\chi^2, n)$$

2. 自 由 度

今統計的数列 $m_1', m_2', \dots, m_{n'}'$ に於いて項数は n' 個あるが、そのうち一つの項だけは自由の値を取り得ない。何となれば

$$m_1' + m_2' + m_3' + \dots = N$$

なる式で縛られているからである。この事を自由度 (degree of freedom) n が $n = n' - 1$ なりと云う。さて二数列一致の確率 P は χ^2 と n との函数であるから、

$$P=f(\chi^2, n)$$

若し、 $m_1, m_2, \dots$ 等が 2 個の式で縛られて居る時は、 $n = n' - 2$
 k 個の式で縛られて居れば、 $n = n' - k$
 である。

3. 計 算 図 表

さて χ^2 と n と (又は n') が明になれば、Pearson 又は Fisher-Yates の表を用いて、二数列一致の確率 P を計算できる。

さて判定には人それぞれの癖は多少あるが大体、

$P > 0.1$ ならば P 大で二数列一致すると考え、

$P < 0.01$ ならば P 小で二数列一致しないと考えてよい。若し P がその中間の値ならば何とも断定出来ないの、観測を継続して P が何れかに移動するのを見るのである。そのためには Fisher の表でも Pearson の表でも充分役にたつけれども、私は西木と計算図表を作製してその用法を一層簡約にした。それが附図 I と II とである。

この図表も亦、 χ^2, n', P の表であつて根本的には Pearson, Fisher と異なつてはいないけれど、それに比し次の利点と特徴がある。

(i) χ^2 を横軸に取つた。その数字は χ^2 である。又 P を縦軸に取つた。その数は一致の確率である。今 $n' = 2$ と定めて、 χ^2 と P との関係を図示すれば、附図 I の曲線 2 となる。 $n' = 3, n' = 4, \dots$ と順次取れば、同図上の種々の曲線が得られる。

(ii) 今 n' が定まり、 χ^2 が定まれば附図の上の点が定まり、従つて P の値を縦線の目盛りで、読みとる事が出来る。

(iii) 全図表に於て $n'X^2$ が移動するに従い、 P 点が移動することを一目して理解できる。

(iv) P の値を知るために、比例部分の法や補間法を用いる必要がない。数値は手軽に迅速に読むことが出来る。

(v) $X^2=20$ までは第 I 附図で、 X^2 それ以上の場合には図表 II を用いる。但し第 II 図表では X^2 の目盛りは少し狭くなっているから注意されたい。

〔例 1〕 次の実験で $n'=9$ $X^2=23.46$ を得た。これに対する P の値を求めよ。

図表の用法： 図表 II で横軸で $X^2=23.5$ に対する $n'=9$ の曲線上の点を求め、その点の縦軸の目盛りを読むに極めて小で、

$$P < 0.01$$

である。Fisher の表では $0.01 > P > 0.001$ 、Pearson Table では $0.003 > P > 0.002$ となり、多少の精粗はあるが何れも正しい。

第 II. 用法についての批判と骰子による実験

カイ自乗検定法を多くの人々は正確に理解していない。その主なる二つをあけると、第一論者は云う。

(i) この方法は誤りで好い方法ではない。何となれば、今実験回数を増して 10 倍にしたとすれば、 $\frac{(m_i - m'_i)^2}{m_i}$ の分母は 10 倍し、分子は略々 100 倍となるから X^2 は概して 10 倍となり、従て P 甚だしく小となる。即ち P の値は試行回数 N によつて、著しく変るので用をなさないと云う。

(ii) 第二は反対に試行回数 N に無頓着の人がある。 P が小になりさえすれば、二数列一致せずと云い、 P が大なれば二数列一致せりと考え、 N を考慮に入れないのである。

この二種の考え方は共に極端で、誤りである。その批判をする前に骰子で実験を行つて見る。

4. 実験用の骰子

エボナイトで各辺約 1 cm の骰子 100 個を作らせた。製作には必ずしも正しい立方体を要求しなかつた。

〔第一実験に用いた骰子〕 作つて来た骰子のうち、なるべく角度の直角でない不正な骰子 10 個を選び、何れも、最も鋭い立体角を挟んだ三面を赤く塗り、この 10 個を同時に投じ、赤の面の出たものを成功とし、成功せる骰子を数えた。この実験を 2000 回行い、100 回毎に成功の分布表を作り、理論数と比較した。理論値は赤の出る確率 $P = \frac{1}{2}$ として、二項分布で計算した。第 1 表の如く略一致するが、細かく見ると一致していないので、 X^2 を計算し、更に P を計算した。(第 1 表、第 2 表)。

〔第二実験に用いた骰子〕 第二実験ではやや不正形の骰子 10 個を選び、その最も鋭

第 1 表. 第 I 実験 100 回目の分布比較

成功数	最初の 100 回目			
	m'	m	$m'-m$	$\frac{(m'-m)^2}{m}$
0	0	0.1	-3.5	2.23
1	1	1.0		
2	1	4.4		
3	11	11.7	-0.7	0.04
4	19	20.5	-1.5	0.11
5	22	24.6	-2.6	0.27
6	20	20.5	-0.5	0.01
7	17	11.7	-5.3	2.40
8	6	4.4	-3.5	2.23
9	3	1.0		
10	0	0.1		

$$n' = 7 \quad \chi^2 = 7.23$$

$$P = 0.297$$

第 2 表. 第 I 実験 各回の χ^2 と P

試行 回数 N	実験 I	
	χ^2	(P)
100	7.23	0.297
200	7.97	0.241
300	10.92	0.091
400	14.75	0.023
500	20.38	0.002
600	21.09	0.002
700	21.04	0.002
800	19.25	0.004
900	21.65	0.001
1000	23.64	0.0006
1200	32.60	0.00001
1400	32.46	0.00001
1600	35.86	< 0.00001
1800	37.08	< 0.00001
2000	35.15	< 0.00001

第 3 表. 第 II 及 第 III 実験

試行 回数 N	実験 II		実験 III	
	χ^2	(P)	χ^2	(P)
500	4.53	(0.6063)	2.85	(0.8255)
1000	8.31	(0.2181)	4.42	(0.6209)
1500	6.59	(0.3628)	4.21	(0.6089)
2000	8.62	(0.1981)	7.30	(0.2960)
2500	12.49	(0.0527)	9.38	(0.1550)
3000	15.77	(0.0153)	10.99	(0.0887)
3500	10.82	(0.0949)	10.46	(0.1080)
4000	11.82	(0.0667)	8.91	(0.1794)
4500	11.12	(0.0852)	4.93	(0.5531)
5000	10.47	(0.1076)	3.73	(0.7124)
5500	14.46	(0.0253)	4.96	(0.5491)
6000	14.10	(0.0287)	4.20	(0.6901)
7000	20.77	(0.0021)		
8000	28.19	(0.00009)		
9000	40.80	(< 0.00001)		
10000	51.42	(< 0.00001)		

い立体角を挟む三面を赤く塗りて、之を投じ赤い面の出た骰子を成功とし、その数を数えた。この実験を 10,000 回行い、500 回毎に成功数の分布表を作り、理論分布と χ^2 -検定法で比較した。

〔第三実験に用いた骰子〕 第三実験には、なるべく正しい立方体の骰子を選び、その三面を赤く塗り、赤い面の出た回数の度数分布表を同様に作製し χ^2 検定法を行つて比較した。

5. 第 I 実験の結果

第 1 表はその 100 回目の度数分布表で、 m' は実験値、 m は理論値である。可なりよく一致している。又 χ^2 を計算すれば 7.23、可なり小で、従て P は大で 0.297 であつた。

第 2 表では 100 回ごとに、その χ^2 と P とを記入したのであるが、回を重ねるごとに χ^2 大となり、 P が小となつていく。300 回目では既に 0.1 より小となり、500 回目では P が 0.01 より小となり、1000 回目で 0.001 より小となつていく。

この事実を何う見るか。元来実験に用いた骰子は不正なものだから、理論値と一致しないのは初めから判つてゐる。然るに実験回数が少いとき (100 回又は 200 回のとき) は χ^2 まだ小で、従て一致の確率 P がまだ大であつた。表で明確に N を大とすれば、 χ^2 も大となり、従て P 小となり、漸次一致しない事が立証せられて来る。故に「 N 小なる間は不一致が証明できないが、 N 大となれば不一致を明に立証できるのである」。第 I 実験では 500 回投じて、 P 0.01 より小となり不一致が証明できた。 N 大となれば益々 P 小となつた。

6. 第 II 実験第 III 実験

第 II 実験はやや不良な骰子を用いたのであるから、一致の確率が大きいのは当然予想されたが、果して 500 回の試験でも尙お $P=0.606$ 、 N を大とするに従い動揺は多少あるが、結局 $N=3000$ に至り尙お P が 0.01 より大であつた。直に P 小となつたと云えるのは、 $N=7000$ 回以後であつた。故に「統計的数列が理論的数列と不一致の筈であつても、その程度が弱ければ、 P の値の降下極めて緩徐で、 N 甚だ大とならねば不一致が立証できない。それでも忍耐して N を大とすれば結局は立証に成功することも分つた。第 III 実験では、出来るだけ正しい骰子を選んだのであるが、 χ^2 が容易に大とならない。 $N=3000$ 回のとき一度 $\chi^2=10.99$ まで増大したけれども、回を重ねたら又反対に χ^2 が小となつた。遂に $N=6000$ 回で忍耐が盡きて、実験を中止し遂に不一致を立証できなかった。

そこで次の事がいえる。

7. 実験の批評と結語

以上の実験は χ^2 検定法に対する第一の論者に答え、又第二の一般使用者に対する警告の意味で行つたのである。

(i) 理論と実験方法とに隔りがあつて、統計が理論値に一致しない様な場合、 N を大にすれば χ^2 は益々大となり、速にその不一致が立証できる。第一の論者がこれを不都合と見たのは考えが足りないのである。而して理論と実験とのクヒチガイが甚しければそれだけ早く、そのクヒチガイを発見しうる。

(ii) 併し実験が完全に（又は完全に近く）理論通り行われて居れば、その不一致を発見するのが容易でない。第 III 実験の場合でも人の造つた骰子だから、必ず多少理論とのクヒチガヒはあつたのだから。けれど遂にそのクヒチガヒを立証できなかつた。この時は N を極めて大にする必要があつたのだから。この点でも第一の論者は誤つてゐる。 N を 10 倍にすれば χ^2 は 10 倍になると云う事であつたが、それは実験と理論とのクヒチガヒが大であれば、略その通りになるだろうが、クヒチガヒが無く完全に理論通り実験が試行せられるか、又はそれに近い時は N を大にしても、 χ^2 は必ずしも大とならぬ事は第 III 実験で見られる通りである。

(iii) 元来実験は完全に理論通り行いうる事はない筈である。だから結局は χ^2 大となり、 P 極めて小となるのが当り前である。唯そのクヒチガヒの大なる場合は比較的容易に之を立証し得られ、クヒチガヒ小なる場合はその立証が困難なだけの違いである（実験 I II III）。故に吾々はその実験回数 N を常に考慮に入れなければならぬ。「今 $P < 0.01$ になつた場合に不一致を証明し得たと考えるならば、第 I 実験では $N=500$ のとき $P=0.002$ となり、2 数列の不一致を立証し得、実験 II では $N=7000$ に至りて始めて不一致が立証せられ、第 III 実験では $N=6000$ でも遂に不一致が証明できず、 P 大であつた。此分なら $N=20,000$ 回位では未だ目的を達せられなかつたであらう」。

N を忘れて「二数列一致する」とか「しない」とか云うのは、無意義である事が判るであらう。

文 献

- 1) 統計学研究会編—統計数値表, 昭和 19.
- 2) Fischer-Yates, 1938—Statistical Tables for Biological Agricultural and Medical Research.
- 3) Pearson, K. 1924—Table for Statisticians and Biometritians. London.
- 4) Elderton, W. P. 1902—Tables for Testing goodness of fit, Biomet. vol. I.
- 5) Fisher, R. A. 1924—The Conditions under which χ^2 measures the discrepancy between observation and hypothesis, J. Roy. stats. Soc. 87.
- 6) Fisher, R. A. 1950—Statistical Methods for Research Workers, Edinb. and London.
- 7) Neyman, J. & Pearson, E. 1931—Further Notes on the χ^2 -distribution, Biom. vol. 22.
- 8) Pearson, K. 1900—On the Criterion that a given system of deviations etc. Philos. Mag. 50.
- 9) Yule, G. U. & Kendal—An Introduction to the theory of Statistics.
- 10) 上田常吉 昭 10—生物統計学
- 11) 西木世振 昭 17— χ^2 試験法に於ける P の計算表並に計算図表に就て—朝鮮医, 第 32 第巻 5 号

図表 I. 横軸に χ^2 , 縦軸に P を取った. 曲線に示した数字はそれぞれ曲線の n' (項数) であつて, 自由度 n でない.

図表 II. I は X^2 が 20 まで, II は 20 以上の場合, この二ツで完全な表となる.

但し II では X^2 の目盛りをやや細かくした.