京 城 談 話 會 大 會

Seisin NISIKI 1942. Über die unrichtig gebrauchten χ^2 -Prüfung. *Jap. Jour. Genet.* 18: 142-145.

Chi-Square-Test の誤られたる使用について

西 木 世 振

京城帝國大學醫學部解剖學教授室
京城女子醫學專門學校解剖學教室

人體に關する統計をやつて居ると完全な綜合比較法が無くて悩むのである。カイ自乗試験法（又は適合試験法、 χ^2 -test, test of goodness of fit）は多くの綜合比較法中でも最も完全に近いものではあるが、それでも使用法を誤つて不正の結論に達してゐるのを見ることがある。今何等かの方法で理論的度数分布が得られ、且つこの理論数に一致する筈の實驗的分布が得られたとする。この二數列を比較するに普通完全には一致しない。その理由の第一は實驗度数が偶然の影響で不規則となるからである。第二には實驗数が理論数に一致しない何等かの理由が他に存在する爲である。然らば如何なる程度までが偶然起り得る誤差であり、何れ以上が偶然起り得ない即ち何か他の理由によつて起るのであるかを判定する必要がある。 χ^2 -試験法はこの目的のために Pearson が案出したものである。その計算法は既に我が國でも紹介せられてゐるから詳細は省略するが、要點は二數列より χ^2 を計算し Pearson's Table 又は Fisher's Table を用ひて、 χ^2 より二數列一致の確率 P を計算して、P の大小を以て二數列が一致せりや否やを判定するのである。所がこの法が誤つて用ひられてゐるのを屢々發見するのでそれを訂正したいと思ふ。

χ^2 -試験法の普通の用法に従へば P は二數列が一致せりと認め得る確率であるから、P 大であれば二數列が一致してをり、二數列間に存する誤差は多分偶然の爲めに起るのであらう。又 P 小ならば二數列は一致せず、二つは別個の數列であるか、若しく實驗値を狂はせる何かの理由が存在するのであらうと判定する。

何れからを「P 大なり」又は「P 小なり」と定めるかは、使用する人によつて多少差異があるが、0.1 より大であれば何人でも P 大と認めてゐる様であり、0.01 より小なれば P 小なりと考へてゐる様である。而して P が 0.1 より小、0.01 より大なる場合には人により批判を異にしてゐるが、この場合は何れとも決定しがたいといふのが誤りのない用法と考へられる。所が余は最近ある論文を讀んで以上の普通使用法の誤りを直感したので、先づ骰子について實驗を試み、余の直感の誤りでない事を確めた。此の法の誤用は二通りあるから順次これを述べやう。

I. P 大であつても輕々しく二數列一致せりと信ずべからず。

實驗 I (a) 骰子の三面を赤く塗つて置いてこれを投ずるに若しその骰子が正しく正立方形で且つ重心が正しく中心にあれば、赤の出る確率が $\frac{1}{2}$ である。故に斯の如き 10 個の骰子を投

ずれば各回赤の出る数 (即ち成功数) は 0 より 10 まであり得るが, 成功数の確率分布は $(q+p)^{10}$ の展開で求め得る。即ち

$$(q+p)^{10} = q^{10} + 10q^9p + \frac{10 \cdot 9}{1 \cdot 2} q^8p^2 + \dots$$

右邊の第一項は 10 個の骰子共に失敗する確率, 第二項は 1 個の骰子のみが成功する確率, ……である。故に 1000 回の実験を行へば理論的成功数分布は次の通りとなる。

成 功 数	0	1	2	3	4	5	6	7	8	9	10	計
理論的度数	1.0	9.8	43.9	117.2	205.1	246.1	205.1	117.2	43.9	9.8	1.0	1000.1

若し不正の骰子を投ずればその通りとはならないはずである。余は実験 I (a) にはエボナイトで作つた 100 個の骰子の内, 特に出来損つたイビツな骰子を 10 個擇び, 最も鋭い立體角を圍む三面を赤く塗つた。多分成功確率 (赤の出る確率) p は $\frac{1}{2}$ より大であらう。さてこの骰子 10 個を同時に 2000 回投じて各回の成功数を記録し, 100 回をます毎に成功数の分布表を作り, χ^2 -試験法を用ひて理論的度数分布と比較した。

第 1 表はその χ^2 と P との値を記入したものである。この実験結果を通覽するに試験回数 N が小なる間は P が可なり大であつて, N が 100 や 200 の間は P は 0.30 又は 0.24 となつて

第 1 表 不正なる 10 個の骰子を投じ 100 投毎に成功分布表より χ^2 と P とを計算し比較せる表

試験回数 (N)	χ^2	P
100	7.29	0.2969
200	7.97	0.2406
300	10.92	0.0913
400	14.75	0.0226
500	20.38	0.0024
600	21.09	0.0018
700	21.04	0.0018
800	19.25	0.0038
900	21.65	0.0014
1000	23.64	0.00062
1100	32.17	0.00001
1200	32.60	0.00001
1300	36.28	<0.00001
1400	32.46	0.00001
1500	31.35	0.00002
1600	35.86	<0.00001
1700	36.58	<0.00001
1800	37.08	<0.00001
1900	36.21	<0.00001
2000	35.15	<0.00001

第 2 表 やゝ不正な骰子 10 個を以てせる実験

試験回数	χ^2	P
500	4.53	0.6063
1000	8.31	0.2181
1500	6.59	0.3628
2000	8.62	0.1981
2500	12.49	0.0527
3000	15.77	0.0153
3500	10.82	0.0949
4000	11.82	0.0667
4500	11.12	0.0852
5000	10.47	0.1976
5500	14.46	0.0253
6000	14.10	0.0287
6500	17.10	0.0090
7000	20.77	0.0021
7500	22.79	0.00088
8000	28.19	0.00009
9000	40.80	<0.00001
10000	51.42	<0.00001

第 3 表 正しい正方形と思はせる 10 個の骰子を以てせる実験

N	χ^2	P
500	2.85	0.8255
1000	4.42	0.6209
1500	4.51	0.6089
2000	7.30	0.2960
2500	9.38	0.1550
3000	10.99	0.0887
3500	10.46	0.1080
4000	8.91	0.1794
4500	4.93	0.5531
5000	3.73	0.7124
5500	4.96	0.5491
6000	4.20	0.6501

居る。即ち P 大であつて數例は一致すると認めてよい様である。所が實驗を進め 300, 500 と回數を増すに従ひ P 漸次小となり, 500 投以上では一致せりとは何うしても認められない。1000 投以上に至ると一層甚だしく遂に P が 0.00001 以下となり, 二數列は全く一致せず, 各骰子の成功確率 $\frac{1}{6}$ ならずと斷定してよい迄になつた (第 1 表)。

即ち同一の實驗でも, 實驗回數を小に取りたる實驗と大にした實驗とでは結果がちがふのである。故に只「P 大なるが故に一致せり, 又 P 小なるが故に一致せず」とは云ひ得ない。元來この實驗では不正な骰子を用ひたのであるから一致しないのが當然である。然るに N 小なる間は P 大であつて, 通念では二數列一致せりと考へられ, N 大となつて始めて二數列の不一致を確證出来る。故に χ^2 試驗法でも觀測回數 N を無視して, 單に P の値のみから批判するのは誤りで, N を批判の重要要素としなければならぬ。

實驗 I (b) 次にやゝ不正なる別の 10 個の骰子を選び, 同様の實驗を 10000 回繰り返したが今度は P の小となり方が極めて緩徐で, 5000 の實驗でもまだ P が充分小なりとは云へなかつた。それにかゝらずそれ以後では甚だしく小となつてゐる (第 2 表)。

實驗 I (c) 次に 100 個の骰子の内正しい立方形と思はれる 10 個を選定し各三面を赤く塗り同様の實驗を行つた。而して 500 投ごとに成功回數の分布表を作り 6000 投に達した。この分布表から χ^2 と P とを計算したのが第 3 表である。

第 3 表から P を通覽するにこれまでの實驗に比し P 極めて大である。これは當然であつて p は $\frac{1}{6}$ に近いから, 二數列が一致するのが當然である。又 N 大となつても P 小とならない。但し余は N=6000 で實驗を止めたが, 若し數萬回の實驗を繰り返して行ふならばいつかは P が小となるに相違ない。又この場合も P が大ではあるが可なりの動搖を示し, N=2500 に於ては P=0.155, N=3000 に於て P=0.089, N=3500 に於て P=0.108 とまでなつたが, N=4000 回以後回復して P が再び大となつてゐる。これは無論偶然の現象だと思はれる。要するに以上の實驗から

(i) 同一の實驗も回を重ね N を大にすれば, 結局は P 極めて小となり, 實驗數列と理論數列とが一致せずと云ひうる。これ當然で骰子の確率を完全に $\frac{1}{6}$ とし得ないからである。

(ii) N を大にすれば結局二數列一致しない事を證明し得るが, N 小なる間は常に P 大であつて二數列の不一致を證明し得ない。

(iii) 骰子の不正形が甚だしければ, N が小なりともその歪形なることを證明しうる。

實驗 II 次に實驗 I (a) に用ひた不正なる骰子を用ひて, 40, 60, 80 投の實驗をそれぞれ 5 回

第 4 表

N と χ^2 , P 實驗回次	N=40 回の實驗		N=60 回の實驗		N=80 回の實驗	
	χ^2	P	χ^2	P	χ^2	P
第一回の實驗	5.63	0.4678	6.93	0.3280	11.10	0.0857
第二回の實驗	6.64	0.3577	6.88	0.3434	1.68	0.9408
第三回の實驗	7.65	0.2671	3.32	0.7666	6.99	0.3219
第四回の實驗	3.66	0.7216	2.09	0.9097	6.90	0.3311
第五回の實驗	2.07	0.9119	10.43	0.1091	4.24	0.6448

宛行つて見た。P 何れも大であるが、その値に可なりの動揺がある。かく N 小なる場合は P の値は相當の動揺を示すが常に大である (第 4 表)。

以上の實驗によつて明な如く「P 大なるが故に二數列一致せり」とは斷言し得ない。

骰子の確率を完全に $\frac{1}{6}$ となし得ないと同じく元來理論的分布と實驗的分布とが完全に一致する事は寧ろないと云つてよい。 χ^2 試驗法を用ひても二數列の間の差甚だしければ僅少の實驗回數によつても直ちにこれを證明し得るが、差少ければ多數の實驗を行はねばこれを證明し得ないのである。

故に P の判定にも常に觀測數 N を考慮に入れねばならぬ。例へば N=100 に於て既に P=0.001 となつたとすれば唯 100 回の觀測で二數列の不一致を證明し得た程大なる差異があると判定する。又實驗 I (a) では N=1000 に至つて始めて P=0.001 以下となつて不一致が證明できた。實驗 I (c) では比較的正しい骰子を用ひたから、實に N=6000 になつても尙、P=0.65 と云ふ大きな値を取り不一致を證明し得なかつた。それ程數列が一致してゐる。併しこの場合も數萬回の實驗を行つたならば P が小なつたかも知れぬが、困難だから實驗を中止したのである。

要するに「N が云々……なる實驗でも P 小とならず、二數列の不一致を證明し得なかつた」又は「N が云々……の實驗に於て P=0.001 以下となり、二數列の一致しない事を證明し得た」又は「N=1000 に於て P=0.001 であつた、二數列の差異はその程度である」等の如く常に N を考慮して批判の料に供すべきである。

II. f 列及び f' 列は度數でなければならぬ。

χ^2 試驗法で比較する爲に二數列は度數分布表でなければならぬ。余は度數分布でない二數列を χ^2 -試驗法を用ひて比較し、一致するとかしないとか論じてあるのを見た事があるからその誤りを常識的に指摘して置く。

二數列が度數でなく或る單位を有する數列とす。例へば cm. 單位で表はしたるものなる時、これを他の單位で表はしても差支ない筈である。かりに mm. に換算すれば f も f' もすべての數は 10 倍となるを以て、

$$\chi^2 = \sum \frac{(f - f')^2}{f}$$

の代りに、單位をかへた時新しきカイ自乗を χ_1^2 とせば

$$\chi_1^2 = \sum \frac{(10f - 10f')^2}{10f} = 10 \sum \frac{(f - f')^2}{f} \quad \text{即ち} \quad \chi_1^2 = 10\chi^2$$

cm. 單位を mm. 單位に代へたばかりに、 χ^2 が 10 倍になつて了つた。これを以つて見ても χ^2 -試驗法が用ひられるのは理論的及び實驗的の度數分布表の比較に限ること明らかである。