《講　座》

統計的推論 (2)

―― 計数データでの出現率を主とした扱い ――

佐久間　昭*

11. 出現率の推論

計数データの代表的な要約値は出現率である．$r=0$ の標識なしと $r=1$ の標識ありの個体が π: $(1-\pi)$ の割合で存在する母集団から，ランダムに n 個体を抽出し，$O_n\{r_i\}$ とすれば，$x=\sum r_i$ は標識ありの個体数で，その平均値，

$$p=\frac{\sum r_i}{n}=\frac{x}{n} \tag{36}$$

を標本での**出現率**という．

(1) 1 個体の結果は 1 か 0 で，
(2) それらの確率が π と $(1-\pi)$ であり，
(3) 毎回の観測が独立して
(4) n 回行なわれる，

という条件で得られる $x=\sum r_i$ は，**二項分布**にしたがい，x の確率を f(x) とするとき，

$$\begin{aligned} &x\in B\{n,\pi\}\\ &f(x)={}_nC_x\pi^x(1-\pi)^{n-x}\end{aligned} \tag{37}$$

と示される．

$\pi\leqq0.5$ としたとき，$n\pi\geqq5$ の条件では，式 (37) の分布に対して正規近似が利用できる．n 中 x の標識ありは $(n-x)$ の標識なしであるから，必要に応じて有効を無効に読みかえて，x を $(n-x)$，π を $(1-\pi)$ にする．

出現率 $p=x/n$ は，母出現率 π を反映するシグナルであるが，p に関するノイズは，

$$SE\{p\}=\sqrt{\frac{\pi(1-\pi)}{n}} \tag{38}$$

で与えられるが，両辺を n 倍し出現数，度数に関しては $SE\{x\}=\sqrt{n\pi(1-\pi)}$ と書ける．

中心極限定理 (→ 節 4) を利用し，式 (13)，

$$t_0=\frac{|\bar{y}-\eta|}{SE\{\bar{y}\}}$$

の形式の SN 比にして，検定を行なうのであるが，p ないし x は，$-\infty$ から $+\infty$ まで分布するわけではなく，また，連続的な値でもないため，連続修正と呼ばれる細工を行なう．

$H_0: \pi=\pi_0$, $H_1: \pi=\pi_1\neq\pi_0$ の検定にあたつて，式 (38) のノイズを利用し，

$$t_0=\frac{|p-\pi_0|-1/2n}{\sqrt{\pi_0(1-\pi_0)/n}}=\frac{|x-n\pi_0|-0.5}{\sqrt{n\pi_0(1-\pi_0)}} \tag{39}$$

の S/N 比をつくる．p なり x なりのきざみの半分を修正している．これによって，$\alpha=0.05$ 付近での議論についての近以が非常によくなる．$t_0\geqq t_\alpha[\infty]$ ならば水準 α で有意とする．$t_\alpha[\infty]$ は正規偏位で，$t_{0.05}[\infty]=1.960$ である．

出現率 π の点推定は $p=x/n$ であるかが，$100(1-\alpha)\%$ の信頼係数では，

$$\pi_L{}^U=\frac{x}{n}\pm\left\{t_\alpha[\infty]\sqrt{\frac{x(n-x)}{n^3}}+\frac{1}{2n}\right\} \tag{40}$$

を，π の信頼限界とする．ノイズとしては，$p=x/n$ をもとに，$SE\{p\}\doteqdot p(1-p)/n$ としている．ときに $\pi_L<0$, $\pi_U>1$ になるが，この場合は $\pi_L=0$, $\pi_U=1$ とおきなおす．

【例 6】

36 例の両腕に注射液 A, B を無作為に用い，6 例では A, B の発赤は同程度，10例で A が，20例で B が強い発赤を示した．6 例を除き，1 標本問題として，n=30 中の標識あり (A が強い) を x=

* 東京医科歯科大学難治疾患研究所臨床薬理学部門

10例とする．標識ありの母出現率 π につき， H_0: $\pi=0.5$ (A, B に差なし)，H_1: $\pi\fallingdotseq 0.5$ として，

$$t_0=\frac{|10-30\times 0.5|-0.5}{\sqrt{30\times 0.5\times(1-0.5)}}=1.643<1.960$$

であるから，5% 水準で有意ではなく，この限りでは A, B に差ありとする証拠は弱い．$t_P=1.643$ から $P\fallingdotseq 0.10$, つまり検定の実質水準は 10% となる．正確な計算では 0.0987 であり，近似はよい．

この例のように $\pi_0=0.5$ の二項検定をとくに**符号検定** sign test といい，式 (39) は，

$$t_0=\frac{|2x-n|-1}{\sqrt{n}}=\frac{|a-b|-1}{\sqrt{a+b}} \tag{41}$$

となる．第2の表現では標識あり，なしの例数を a, b としている．a+b=n である．

つぎに，勝負ありの n=30 をもとに，p=10/30, 33.3% の出現率を認めたが，

$$\pi_L{}^U=\frac{10}{30}\pm\left\{1.960\sqrt{\frac{10(30-10)}{30^3}}+\frac{1}{2\times 30}\right\}$$
$$=0.3333\pm 0.1854=0.148;\ 0.519$$

で，π の 95% 信頼区間は [14.8%, 51.9%] となり，A, B の勝ち負け半々という 50% を含み，検定結果と矛盾しないが，式 (39), (40) のノイズは等しくなく，ときに矛盾する．正確な計算では [17.3%, 52.8%] となる．

観測対象が n=200～300 と大きくなり，ここで少数の標識あり x を認めたとき，この度数の $100(1-\alpha)$% の信頼限界の近似は，

$$\mu_L{}^U\fallingdotseq\{\sqrt{x+0.25}\pm 0.5t_\alpha[\infty]\}^2 \tag{42}$$

で与えられる．

12. McNemar の検定

対応のある2群比較では $O_n\{x_i, y_i\}$ とペアーのデータを得るが，x_i, y_i とも 1, 0 ないし，+, − のデータであるとき，{1, 1}, {0, 0} の勝負なしの結果と，{1, 0}, {0, 1} の勝負ありの結果に整理できる．

g 例の左右の腕について，たとえば注射液 A, B で 5 mm 以上を発赤あり + と約束して，

	B+	B−	
A+	c	a	a+c
A−	b	d	b+d
	b+c	a+d	g

のような度数表ができる．いわば A 勝ちの a ペアーと B 勝ちの b ペアーをもとに A, B を比べる方法を，**McNemar の検定**という． A 勝ちの出現率を π とすれば， H_0: $\pi=0.5$, H_1: $\pi\fallingdotseq 0.5$ の符号検定で，式 (41) により，

$$t_0=\frac{|a-b|-1}{\sqrt{a+b}}\geqq t_\alpha[\infty] \tag{43}$$

のとき有意となる．

つぎに A, B での + の出現率を π_A, π_B とし，差を δ とおけば，それぞれの推定値は，

$$p_A=\frac{a+c}{g},\quad p_B=\frac{b+c}{g},\quad \hat{\delta}=\frac{a-b}{g}$$

となるが，$\hat{\delta}$ のノイズは，

$$SE\{\hat{\delta}\}\fallingdotseq\frac{1}{g}\sqrt{(a+b)-\frac{(a-b)^2}{g}}$$

で与えられ，出現率の差の信頼限界は，

$$\delta_L{}^U=\frac{a-b}{g}\pm\frac{t_\alpha[\infty]}{g}\sqrt{(a+b)-\frac{(a-b)^2}{g}} \tag{44}$$

となる．検定では，勝負ありの n=a+b のみを用いたが，推定では，g も関係する．

【例 7】

注射液 A, B の比較で，発赤ありを 5 mm 以上とし，それぞれ g=43, 68, 34 につき，3回の別々の試験を行なった．前出の表と同じ形式で，表内度数のみを示す．

T_1	B+	B−	T_2	B+	B−	T_3	B+	B−
A+	15	4	A+	30	4	A+	15	2
A−	14	10	A−	14	20	A−	7	10

たとえば T_3 につき，a=2, b=7, g=34.

$$t_0 = \frac{|2-7|-1}{\sqrt{2+7}} = 1.333 < 1.960$$

実質水準は $t_P=1.333$ で $P \doteqdot 0.19$. 正確には 0.180 であり, $n=a+b=9$ は少ないが, 一応はよい近似になっている.

$$\delta_L{}^U = \frac{2-7}{34} \pm \frac{1.960}{34}\sqrt{(2+7)-\frac{(2-7)^2}{34}}$$
$$= -0.1471 \pm 0.1657 = -0.313;\ 0.019$$

で差 δ の 95% 信頼間は [−31.3%, 1.9%].

以上の 3 回の成績をまとめると

—	1	2	3
t_0	2.121*	2.121*	1.333
$\hat{\delta}$	−0.233	−0.147	−0.147
δ_L	−0.413	−0.264	−0.313
δ_U	−0.052	−0.030	0.019

これらの成績と原表とをつきあわせてみると, 検定と推定のちがい, 勝負のあり, なし, および全体の度数(ペアー)の影響に一応の見通しがつくであろう. * は 5% で有意を示す.

McNemar の検定の形式の表で,

$$(1)\quad \begin{array}{c|cc|} & B+ & B- \\ \hline A+ & 0 & a \\ A- & b & 0 \\ \hline \end{array} \qquad (2)\quad \begin{array}{c|cc|} & B+ & B- \\ \hline A+ & c & 0 \\ A- & 0 & d \\ \hline \end{array}$$

と, 2 つの極端な場合を考えてみよう. (1) では, A+ なら常に B−, A− なら常に B+, (2) では, A+ なら常に B+, A− なら常に B− であるが, (1) では負の関連, (2) では正の関連がある.

したがって, 本来の検定で $H_0: \pi=0.5$ が捨てられず, 正の関連が強ければ, A, B はますます似たものと考えられる. 負の関連が強ければ, もう少し掘りさげた検討が必要となろう.

McNemar の検定で, A, B に差がなければ,

$$p_m = \frac{a+b+2c}{2g}, \quad p = p_m{}^2+(1-p_m)^2$$

A, B に差があれば,

$$p_A = \frac{a+c}{g}, \quad p_B = \frac{b+c}{g},$$

$$p = p_A p_B + (1-p_A)(1-p_B)$$

とする. A, B に関連性がないとしたとき, 勝負なしの度数 (c+d) の出現率が p と予想される. ついで式 (39) で,

$$t_0 = \frac{|c+d-gp|-0.5}{\sqrt{gp(1-p)}} \geqq t_\alpha[\infty] \tag{45}$$

となれば, 関連性ありとし, (c+d)>gp では正, (c+d)<gp では負の関連性と考える.

【例 8】

前出の T_2 について, A, B に差があったから,

$$p_A = \frac{19}{43} = 0.4419, \quad p_B = \frac{29}{43} = 0.6744$$

$$p = 0.4419 \times 0.6744 + 0.5581 \times 0.3256$$
$$= 0.4797$$

$$t_0 = \frac{|15+10-43\times 0.4797|-0.5}{\sqrt{43\times 0.4797\times 0.5203}}$$
$$= 1.182 < 1.960$$

したがって, A, B には, とくに関連性はなく, 1 例ごとの A+ と B+ はランダムである.

有効性に開して, A が B よりも優れるという結果につき, 正の関連があれば, B の出る幕はないが, 負の関連があれば, B は A と守備範囲が異なるので一応の価値はある.

13. 2 つの出現率

対応のない 2 群比較で, 2 処理 A, B が, 全体に対してランダムに割りつけられ, $O_m\{x_i\}$, $O_n\{y_j\}$ の独立な 2 標本が得られているが, いずれも, 1, 0 の 2 枝択 1 のデータとする. A, B の標識ありの出現率を π_A, π_B とすれば, それぞれの標識ありの度数 a, b から

$$p_A = \frac{a}{m}, \quad p_B = \frac{b}{m}$$

の推定値を得る.

この種の計数データの2標本問題は，いわゆる2×2分割表，**4分表**として整理されるが，1標本問題の McNemar の検定の場合と，まったく別の意味になっている．

	A	B	
+	a	b	s
−	c	d	f
	m	n	g

$H_0: \pi_A = \pi_B$, $H_1: \pi_A \neq \pi_B$ については，差 $\delta = \pi_A - \pi_B$ の推定値 $\hat{\delta} = p_A - p_B$ のノイズを求める必要がある．H_0 のもとには，A, B に共通の，

$$p_m = \frac{s}{g}$$

を推定したうえ，

$$SE\{\hat{\delta}\} \doteqdot \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\frac{sf}{g^2}} = \sqrt{\frac{sf}{gmn}}$$

とする．

さらに連続修正 (Yates の補正) g/2 mm を用いて，S/N 比を，

$$t_0 = \frac{|a/m - b/n| - g/2\,mn}{\sqrt{sf/gmn}} \tag{46}$$

とする．$t_0 \geqq t_\alpha[\infty]$ で，水準 α で有意とし，$\pi_A \neq \pi_B$ と考える．

差 δ の推定にあたっては，それぞれの p_A, p_B をもとにノイズを，

$$SE\{\hat{\delta}\} \doteqdot \sqrt{\frac{ac}{m^3} + \frac{bd}{n^3}}$$

としたうえ，信頼限界を，

$$\delta_L{}^U = \left(\frac{a}{m} - \frac{b}{n}\right) \pm \left\{t_\alpha[\infty]\sqrt{\frac{ac}{m^3} + \frac{bd}{n^3}} + \frac{g}{2\,mn}\right\} \tag{47}$$

とする．

ところで，4分表の形式のデータは，いくつかの方法で集められる．

1) 横断研究 cross-sectional study では，ある時点で，g のみを指定して標本を抽出し，(A, B), (+, −) でクロス分類を行なう，この際，表内度数 a, b, c, d, 周辺度数 m, n, s, f を g で割った頻度は，それぞれが現実的な意味をもっている．

2) 後向き研究 retrospective study では，結末 (+, −) に着目し，+から s，−から f の抽出を行ない，これを原因 (A, B) で分類する．普通は s≑f が都合がよい．

3) 前向き研究 prospective study では，原因 (A, B) に着目し，A の m，B の n の抽出を行ない，これを結末 (+, −) で分類する．普通は m≑n が都合がよい．

4) 比較研究 controlled study では，g を定め，(A, B) を無作為に割りつけて，m≑n の群をつくり，時間的にも前向きに，(+, −) を追跡観測する．

特殊な場合として，第1例では A, B を無作為に用い，これが+なら同じもの，−なら違うものを第2例に用い，以下同様に逐次的に続ける，勝ち馬に賭ける方式 play the winner rule がある．裏がえした，負け馬に賭ける方式 play the loser rule もあるが，いずれも標本抽出を積極的に偏らせている．

一般的に言えることであるが，(1)～(4) のいずれによっても，g 例が結論をあてはめたい集団を十分に**代表**していることが第一に要求される．ついで，(A, B) での (+, −) の割合，あるいは (+, −) での (A, B) の割合を比べるにあたり，群形成が**公平**で，比較可能性をそなえていなければならない．(4) では一般に，公平性が確率的に保証される．公平性が切実な問題になるなら，マッチングなどの工夫が折りこまれるであろう．

本節の最初で，(4) の形式の，A, B の独立な2群の出現率の比較という観点からのデータを扱った．一方 (1) のデータにつき，4分表内の度数を，(A, B) の分類と (+, −) の分類に関連性があるかどうかという観点からみることもできる．H_0: 分類相互に関連性なし，の条件のもとに，表内度数は，

$$\hat{a} = \frac{sm}{g}, \quad \hat{b} = \frac{sn}{g}, \quad \hat{c} = \frac{fm}{g}, \quad \hat{d} = \frac{fn}{g}$$

と予測されるが，これらが3～5をこえるならば，4項目の和，

$$t_0^2=\frac{(a-\hat{a})^2}{\hat{a}}+\cdots+\frac{(d-\hat{d})^2}{\hat{d}}$$

を求めると，t_0^2 は，自由度 [1] の**カイ 2 乗分布**の実現値となる，ここに，

$$\{t_\alpha[\infty]\}^2=\chi_\alpha^2[1] \tag{48}$$

の性質がある，実際上は連続修正を加え，計算しやすく変形し，

$$\chi_0^2=\frac{g\{|ad-bc|-0.5g\}^2}{mnsf}\geqq\chi_\alpha^2[1] \tag{49}$$

のとき水準 α で有意とし，H_1: 関連性あり，を採用する．俗に，χ^2 検定と呼んでいるがまぎらわしい名称である．

出現率の差の検定の式(45)は，機械的な変形で式(49)と一致する．こうした事情があるので，出現率の差の検定であっても，**関連性の検定の式**(49) を用いることが多い．データの集め方 (1)～(4) について，細かいことをいえば検定手法が少しちがってくるが，漸近的には，すべて式(49)で処理してよい．

χ_0^2 は，関連性の有無の判断に使う統計量であるが，**関連性の大きさ**の指標ではない．関連性の大小の尺度には，いくつかのものがあるが，その一つに**交差積比** cross-product ratio, odds-ratio がある．これを ϕ とするが，$\ln\phi=\log_e\phi$ を用いることも多い．

実際のデータからは，

$$\hat{\phi}=\frac{ad}{bc},\quad SE\{\hat{\phi}\}=\hat{\phi}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

$$\ln\hat{\phi}=\ln\frac{a}{c}-\ln\frac{b}{d}, \tag{50}$$

$$SE\{\ln\hat{\phi}\}=\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

と求める．しかし，表内度数に0があると不都合なため，すべての表内度数に0.5を加えたものを，あらためて a, b, c, d として，式 (50) を定義することが多い．

出現率 p_A について，$y_A=\ln\{p_A/(1-p_A)\}=\ln\{a/c\}$ を**ロジット** logit という．a+0.5, c+0.5 をあらためて a, c とすることもあるが，

$$\ln\hat{\phi}=y_A-y_B \tag{51}$$

となり，交差積比と結びつく．ロジットは，計数データの複雑な解析を単純化したり，正規分布の近以，用量一反応率曲線の解析，ラジオイムノアッセーでの検量曲線などに，広く利用される．

【例 9】

いずれも $p_A-p_B=0.187$ の結果になる2つの4分表がある．

T_1

	A	B	
+	17	11	28
−	13	18	31
	30	29	59

T_2

A	B	
51	33	84
39	54	93
90	87	177

T_1 について，式 (46), (49) で，

$$t_0=\frac{|17/30-11/29|-59/(2\times30\times29)}{\sqrt{(28\times31)/(59\times30\times29)}}$$
$$=\frac{0.1534}{0.1300}=1.180$$

$$\chi_0^2=\frac{59\{|17\times18-11\times13|-0.5\times59\}^2}{30\times29\times28\times31}$$
$$=1.392=1.180^2$$

$t_0^2=\chi_0^2=1.392<1.960^2$ で 5% 水準で，出現率の差は有意ではない．

$$\delta_L{}^U=\left(\frac{17}{30}-\frac{11}{29}\right)$$
$$\pm\left\{1.960\sqrt{\frac{17\times13}{30^3}+\frac{11\times18}{29^3}+\frac{59}{2\times30\times29}}\right\}=0.1874\pm0.2842$$
$$=-0.097;\ 0.472$$

で，出現率の差の 95% 信頼限界は [9.7%, 47.2%] となり，検定結果と矛盾しない．

表内度数に +0.5 の補正をしたうえ，

$$\hat{\phi}=\frac{17.5\times18.5}{13.5\times11.5}=2.085$$

$$\mathrm{SE}\{\hat{\phi}\}=2.085\sqrt{\frac{1}{17.5}+\cdots+\frac{1}{18.5}}=1.088$$

T_2 についても同様に計算し,

T	1	2
t_0	1.180	2.345*
δ_L	−0.097	0.032
δ_U	0.472	0.343
$\hat{\phi}$	2.085	2.121
$\mathrm{SE}\{\hat{\phi}\}$	1.088	0.647

T_1 から T_2 に，標本の大きさが $g=59$ から，177 と 3 倍になると，検定は5% 水準で有意 * となるが，原表の内容からみて，関連性が大きくなったわけではない．$[\delta_L, \delta_U]$ の区間幅と，$\mathrm{SE}\{\hat{\phi}\}$ は例数を増すと明らかに小さくなることは，ノイズに例数が逆数的に影響しているという，一般的な性質による．$+0.5$ の補正をしないと $\hat{\phi}_0=2.140$ で，T_1 と T_2 で完全に等しいが，ここの $\hat{\phi}$ でも，準不変性が認められる．

14. 多分類の計数データ

対応のある 2 群比較の McNemar の検定の形式で，カテゴリーが，(+, −) の 2 種類から増して (+, ±, −) の $c=3$ 種類になった場合を考えてみる．もちろん，同じ趣向で $c=4$ 種類の場合にも拡張できるであろう．

表記法を少しかえて，度数を n とし，

	B+	B±	B−	
A+	n_{11}	n_{12}	n_{13}	$n_{1.}$
A±	n_{21}	n_{22}	n_{23}	$n_{2.}$
A−	n_{31}	n_{32}	n_{33}	$n_{3.}$
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

対称性の検定として知られているものは，n_{11}, n_{22}, n_{33} の主対角線に関して対称の位置にある，n_{ij} と n_{ji} を 1 対として比べる．全部で $k=c(c-1)/2=3$ 対あるが，

$$\chi_0^2=\sum^{k}\frac{(n_{ij}-n_{ji})^2}{n_{ij}+n_{ji}}$$

$$\chi_0^2=2\Big[\sum^{k}(n_{ij}\ln n_{ij}+n_{ji}\ln n_{ji}) + \sum^{k}(n_{ij}+n_{ji})\ln 2 - \sum^{k}(n_{ij}+n_{ji})\ln(n_{ij}+n_{ji})\Big] \qquad (52)$$

を求める．$\sum$ の k は，k 対ないし k 個の和を意味する．第 1 式と第 2 式の χ_0^2 は少し異なるが，第 2 式は尤度比検定法と呼ばれるものである．H_0: データの対称性，のもとに $\chi_0^2 \geqq \chi_\alpha^2[k]$ となれば有意で，H_1: 非対称，を採用する．

McNemar 型の $c=2$ の場合には，$n_{11}, n_{12}, n_{21}, n_{22}$ の表内度数になるが，この場合に限り，$n_{12}=n_{21}$ に近づくように，$n_{12}\pm0.5$, $n_{21}\mp0.5$ と修正をしたものを n_{12}, n_{21} とみて式 (51) を用いると $k=2(2-1)/2=1$ で，自由度は 1 になり，式 (52) の第 1 式は式 (43) と同等になる．

【例 10】

A, B の施設で 100 枚のレ線写真を独立に判定し，(+, ±, −) に分類した結果を T_0 に示す．

T_0	B +	±	−	
A +	19	17	7	43
±	7	26	5	38
−	3	12	4	19
	29	55	16	100

式 (52) で，

$$\chi_0^2=\frac{(7-17)^2}{24}+\frac{(3-7)^2}{10}+\frac{(12-5)^2}{17}=8.649$$

$$\chi_0^2=2[116.5693+35.3505-147.4638]=8.912$$

で，いずれも $\chi^2_{0.05}[k=3]=7.815$ より大きく，判定は非対称的である．

＋とその他，±とその他，－とその他の 2 カテゴリーの表につくりなおすと，

T_1	+	他
+	19	24
他	10	47

T_2	±	他
±	26	12
他	29	33

T_3	−	他
−	4	15
他	12	69

$\chi_0^2=$ 4.971　　6.244　　0.148

となる．この場合，McNemar の検定になるが，節9の，くり返し検定の観点から，それぞれの χ_0^2 を自由度 [1] ではなしに，自由度 [c−1=2] の $\chi_\alpha^2[2]$ と比べる．$\chi^2_{0.05}[2]=5.991$ をこえる（±，他）の分類での T_2 が有意で，29 と 12 のバランスがくずれ，B で±の判定が多い．T_1 は，A で＋の判定が多い傾向を示している．

同じ程度の的中率をもった r 人の評価者が，n 例の対象を，ときには不明などを含む c 個のカテゴリーに分類するとき，r 人の分類が完全に一致した例を n* とする．このとき，真の**的中率**ないし**信頼性** π につき，

$$p=\frac{n^*}{n}=\pi^r+\frac{(1-\pi)^r}{(c-1)^{r-1}} \qquad (53)$$

の式が成り立つ．直接に π を解くことは面倒であるが，c, r が大になれば，$\hat{\pi}\fallingdotseq p^{1/r}$ となる．図4に，$\hat{\pi}$ を求めるグラフを示すが，r が小さいところでは，c=1, 2, ···, 6 などを区別して示す．

T_0 につき，p=(19+26+4)/100=0.49，A, B を r=2 の評価者とみて，c=3 の曲線から，$\hat{\pi}\fallingdotseq 0.66$，つまり信頼性は 66% で，＋，±，− の分類は 100 枚のうち 66 枚で的中している．もっとも，対象の分布様式で，[$\hat{\pi}$ は変化する．2分類につぶした T_1, T_2, T_3 について，r=2, c=2 とし，$p_1=0.66$, $p_2=0.59$, $p_3=0.73$ で，それぞれ $\hat{\pi}_1=0.78$, $\hat{\pi}_2=0.71$, $\hat{\pi}_3=0.84$ となり，−か否かでの分類については，信頼できそうである．

対応のない比較の 2×2 分割表を拡張し，A_1, ···, A_i, ···, A_a の多群につき，C_1, ···, C_j, ···, C_c のカテゴリーによる，**a×c 分割表**ができる．

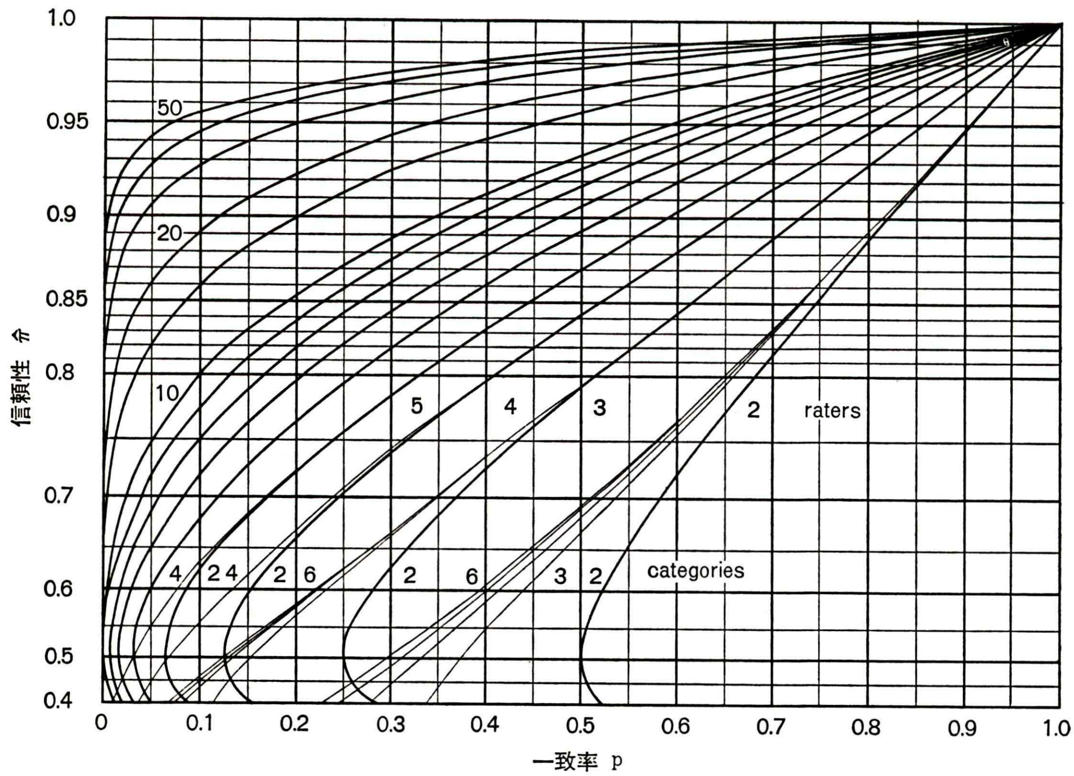

図 4 評価の信頼性

$$\begin{array}{c|ccc|c} & C_1 & \cdots\ C_j\ \cdots & C_c & \\ \hline A_1 & n_{11} & \cdots\cdots & n_{1c} & n_{1\cdot} \\ \vdots & & & & \vdots \\ A_i & & n_{ij} & & n_{i\cdot} \\ \vdots & & & & \vdots \\ A_a & n_{a1} & \cdots\cdots & n_{ac} & n_{a\cdot} \\ \hline & n_{\cdot 1} & \cdots\ n_{\cdot j}\ \cdots & n_{\cdot c} & n_{\cdot\cdot} \end{array}$$

実際には，表内の度数が小さい場合，

$$\hat{n}_{ij}=\frac{n_{i\cdot}\times n_{\cdot j}}{n_{\cdot\cdot}}\geqq 3\sim 5$$

となるようにカテゴリーを設定しなおす必要もでてくる．この条件に満足しないと，近似は悪い．

H_0: A と C に関連性なし，H_1: A と C に関連性あり，とおいて，関連性を吟味するとき，

$$\chi_0^2=\sum_i\sum_j\frac{(n_{ij}-\hat{n}_{ij})^2}{\hat{n}_{ij}}$$
$$=n_{\cdot\cdot}\left\{\sum_i\sum_j\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}}-1\right\} \tag{54}$$

を求め，$\chi_0^2\geqq\chi_\alpha^2[(a-1)(c-1)]$ ならば有意で，関連性ありと考える．

尤度比検定では，すべての度数を $n \ln n = n\log_e n$ におきかえて，

$$\chi_0^2=2[(\text{表内の和})+(\text{総度数})-(\text{周辺の和})] \tag{55}$$

を求める．これは式 (54) と近い値になる．

ここで，有意となったことは，$A_1, \cdots, A_a$ の少なくとも1つでは，$C_1, \cdots, C_n$ の度数の並び方が他とは異なったパタンをもっているということである．分類Cが，著効，有効，無効などと，順序をもったものであれば，おそらく，$A_1, \cdots, A_a$ の良さの比較に強い関心がもたれるであろうが，式 (54), (55) には，良さを比べる性質は含まれていない．

良さを比べるには，たとえば**順位検定**がある．$C_1, \cdots, C_c$ が順序に並んでいるとしたとき，順序分類尺度のデータになるが，まず，C_1 の $n_{\cdot 1}$ には，本来は $1, 2, \cdots, n_{\cdot 1}$ の順位をつけるべきところ，同位になったと考える．形式的に $n_{\cdot 0}=0$ とおき，第 j カテゴリーの C_j に含まれる $n_{\cdot j}$ に対しては，

$$r_j=\sum_{m=0}^{j-1}n_{\cdot m}+\frac{n_{\cdot j}+1}{2} \tag{56}$$

と計算した同じ順位を与える．$n_{ij}\times r_j$ が，A_i についての C_j のセルが稼いだ順位になる．すべてのセルの順位を計算し，ヨコに加えて，$T_1, \cdots, T_i, \cdots, T_a$ と各群の和を求め，念のため，これらをすべて加えた T_G が，

$$T_G=\frac{1}{2}n_{\cdot\cdot}(n_{\cdot\cdot}+1)$$

になることを確かめる．

2群 a=2 のときは，第1群に着目し，

$$A=\frac{n_{1\cdot}}{n_{\cdot\cdot}}T_G,\qquad B=n_{\cdot\cdot}^3-n_{\cdot\cdot}$$

$$\kappa=1-\frac{\sum_{j=1}^{c}(n_{\cdot j}^3-n_{\cdot j})}{B}$$

を求めたうえ，

$$t_0=\frac{|T_1-A|-0.5}{\sqrt{(\kappa/6)n_2.A}}\geqq t_\alpha[\infty] \tag{57}$$

となれば，H_0: A_1 と A_2 での中心位置は同じ，を捨てて，H_1: A_1 と A_2 での中心位置はずれている，を採用する．これを **Wilcoxon の 2 標本検定**というが，Mann-Whitney の U 検定，順位和検定などとも呼ばれる．$t_0>2$ のときは，実質水準を知るには，分子の 0.5 を省いて t_0^* を求めなおす．

多群 a>2 のときは，

$$\chi_0^2=\frac{6}{\kappa}\left\{\frac{1}{T_G}\sum_{i=1}^{a}\left(\frac{T_i^2}{n_{i\cdot}}\right)-\frac{T_G}{n_{\cdot\cdot}}\right\}$$
$$\geqq\chi_\alpha^2[a-1] \tag{58}$$

のとき有意とし，$A_1, \cdots, A_a$ の少なくとも1つでは，他と中心位置がずれていると判断する．これは，**Kruskal-Wallis の検定**と呼ばれ，a=2 の場合に用いると，χ_0^2 は，式 (57) で 0.5 を省いた計算の t_0^{*2} と等しくなる．

これらの順位検定は分布にとらわれず，頑健な性質をもっている．しかも，正規性のある分布に

利用しても，かなり効率がよく，**ノンパラメトリック手法**とか，分布によらない手法と呼ばれるものの，一つの代表である．本来の順位データや計量データを相手にする場合も，対応のない a 群の全データを 1 列に並べて順位を与え，第 k 番から始まる t 個の同位には，$m=k+(t-1)/2=k-1+(t+1)/2$ の平均順位を与えたうえ，各群ごとの和を求め，あちらこちらにある同位については，それぞれ (t^3-t) を計算することで，式 (57), (58) をそのまま利用できる．

【例 11】

A_1, A_2 の治療効果は次表のようであった．

C	−	±	+	
A_1	12	8	15	35
A_2	6	26	8	40
	18	34	23	75

C には順序があるが，A_1 と A_2 での −, ±, + の度数の並び方のパタンを比べてみる．予測値 $\hat{n}_{ij}$ と $(n_{ij}-\hat{n}_{ij})^2/\hat{n}_{ij}$ は，つぎの通りである．

8.4000	15.8667	10.7333
9.6000	18.1333	12.2667

1.5429	3.9003	1.6961
1.3500	3.4128	1.4841

式 (54) の χ_0^2 の成分につき，±でのくいちがいが大きいことが，右側の表からわかる．これらを加えると，$\chi_0^2=13.386$ で，第 2 の表現で求めたものと等しい．

$$\chi_0^2=75\left\{\frac{12^2}{35\times18}+\cdots+\frac{8^2}{40\times23}-1\right\}$$
$$=13.386$$

式 (55) の n ln n の方法では，

$$\chi_0^2=2[199.1718+323.8116-516.0317]$$
$$=13.903$$

となり，いずれも $\chi_{0.05}[1\times2]=5.991$ より大で，A_1, A_2 の度数の並び方は，明らかに異なる．なお，±の列，$8+26=34$ と＋の列 $15+8=23$ とを，そっくり入れかえても，χ_0^2 は変わらないことに注意されたい．

つぎに，良さの比較をするために，−, ±, + にかけて順位をつけると，− の 18 例にはすべて $(18+1)/2=9.5$，± の 34 例には $18+(34+1)/2=35.5$，＋の 23 例には $18+34+(23+1)/2=64$ の平均順位が与えられる．これから，$T_1=12\times9.5+8\times35.5+15\times64\times1358$，同様に $T_2=1492$，$T_G=1358+1492=2850$ で，$75\times76/2$ と等しい．

$$A=\frac{35}{75}\times2850=1330$$

$$B=75^3-75=74\times75\times76=421800$$

$$\kappa=1-\frac{57228}{421800}=0.8643$$

$$t_0=\frac{|1358-1330|-0.5}{\sqrt{(1/6)\times0.8643\times40\times1330}}=0.314$$

で，$t_{0.05}[\infty]=1.960$ よりも小さく，有意ではないため，A_1, A_2 の中心位置にずれはないという意味で，良さが異なるとは考えにくい．

この結果と，関連性の結果をあわせてみると，A_1 では−か＋に集中し，A_2 では±に集中しているが，平均的な効果としては同程度と考えられる．したがって，賭けに対する情熱が強ければ，A_1 をまず用いるだろう．

式 (58) では，

$$\chi_0^2=\frac{6}{0.8643}\left\{\frac{1}{2850}\left(\frac{1358^2}{35}+\frac{1492^2}{40}\right)-\frac{2850}{75}\right\}=0.1023$$

となるが，t_0 の 0.5 を省いた $t_0^*=0.3198$ を求めると，$0.3198^2=0.1023$ となる．