

Fisher の正確確率検定に対応する比率の差の信頼区間

松田 眞一*・佐野 正樹†

E-Mail: matsu@ms.nanzan-u.ac.jp

Fisher の正確確率検定は小標本の 2×2 分割表の独立性の検定でよく用いられる。一方、分割表が 2 標本の比率データととれるときそれぞれの比率の差に関心のある場合がある。Fisher の検定結果と一致するような信頼区間はどのように与えられるべきか議論して、ある信頼区間を提示する。

1 はじめに

分割表の中で一番簡単な 2×2 分割表は様々な分野で用いられる重要なものであり、それぞれの分類の間に関連があるかどうかを調べる独立性の検定はその解析法として中心的なものとなる。独立性の検定でよく用いられるものは χ^2 検定と Fisher の正確確率検定であり、データ数が小さい場合後者の方法が適している。

一方、分割表が 2 標本の比率データと考えられるとき、それぞれの比率の差に関心のある場合がある。たとえば、2 つの薬剤に対して有効か無効かを判定し、薬剤に違いがあるかどうかを検定した上で 2 つの薬剤の有効率の差がどのくらいか見積もるといった状況である。

従来、独立性の検定に Fisher の検定を用いたときでも有効率の差の信頼区間には 2 つの 2 項分布の差から求めるのが普通であった。なぜなら、Fisher の正確確率検定に対応する信頼区間を導出する方法がなかったからである。このような状況は、検定結果と「信頼区間に比率の差 0 が含まれるかどうか」が対応しないという事態を生じさせる。

本論文では、Fisher の検定に対応する信頼区間を構成する際どこに問題があるのかを示し、その解決法の一つを与えて具体的に信頼区間を構成する。

2 2×2 分割表と比率の差の信頼区間の関連

ここで問題を整理するために記号を導入する。 2×2 分割表を次の記号で示すことにする。(分かりやすくするため分類は先ほどの例に従う。)

	有効	無効	計	真の有効率
薬剤 A	x	z	n	p
薬剤 B	y	w	m	q
計	k	ℓ	N	

本論文では $x \sim B(n, p)$, $y \sim B(m, q)$ とそれぞれ 2 項分布に従っている 2 項分布モデルで考える。2 つの比率の差 (薬剤の有効率の差) は $p - q$ であるのだが、この信頼区間は未知である一方の比率 (以下では p で考える) に依存する。すなわち、 p が攪乱母数となる

* 南山大学数理情報学部

† 大鵬薬品工業株式会社

のである。市販の統計パッケージで比率の差の信頼区間を与えているものはこの攪乱母数を $0 \leq p \leq 1$ で動かして区間の上限の最大値と下限の最小値を求める形で処理しているようである。この方法で得られる信頼区間は広くなりがちである。

一方、周辺度数を固定して条件付で考えると x は一般超幾何分布に従うことになり、Fisher の正確確率検定は其中でオッズ比 1 の場合の超幾何分布を用いて行われる。Fisher の検定に対応する信頼区間はこの条件付の下で構成し、無条件の場合と同様に攪乱母数 p を動かして求めればよいように思われるがこの方法ではうまくいかない。後の具体例であげるように無意味な区間となってしまう場合が生じるからである。

3 オッズ比の信頼区間

統計学辞典 (1989) などに示されているように 2×2 分割表の独立性の検定は確率モデルが何であれ、一般超幾何分布によるオッズ比の検定に帰着する。(本論文ではオッズ比を $\lambda = p(1-q)/q(1-p)$ と記しておく。)

前節で述べたように Fisher の検定はその特殊な場合なので一般超幾何分布に基づくオッズ比の信頼区間とは自然と対応している。また、オッズ比の正確な信頼区間は市販の統計パッケージで求めることができる。したがって、比率の差の信頼区間はこのオッズ比の信頼区間から導出するのが適切と思われる。前節の議論より λ が 1 次元であるのに対して $p-q$ は攪乱母数を含む 2 次元の動きをすることが問題の焦点であることが分かる。

4 比率の差の信頼区間の構成法

オッズ比の信頼区間から比率の差の信頼区間を得るための方法を 2 つ提案する。

1. 周辺度数を真値であるかのように扱い、比率に関する次の条件式を設定し、それを基にオッズ比の信頼区間から変換する。

$$np + mq = k$$

(この式は「2 つの群を合わせた場合の真の比率は k/N である」ということを定式化したものである。)

2. オッズ比に対応する比率の組 (p, q) の中で「2 項分布モデルにおいて周辺度数が得られる確率」が最大となるものを求める方法でオッズ比の信頼区間から変換する。

最初の方法は実際に信頼区間を求める方法が代数的計算となり簡便であるのだが、安易な方法である印象を受ける。一方、2 番目の方法は最尤法に似た考え方で受け入れやすいが求め方が複雑となる。ところが、次の定理 1 によりこの 2 つの構成法で得られる信頼区間は一致することが分かり、求め方は簡便な最初の方法を、意味は受け入れやすい 2 番目の方法をと考えることができるようになる。

定理 1 2 項分布モデル

$$f(x, y) = {}_n C_x p^x (1-p)^{n-x} \cdot {}_m C_y q^y (1-q)^{m-y}$$

においてオッズ比 λ を固定した下で次の 2 つの条件は同値となる。

$$i) \quad np + mq = k$$

$$ii) \quad \sum_{x=\max(0, k-m)}^{\min(k, n)} f(x, k-x) \text{ を最大にする } (p, q)$$

(証明) 条件 ii から i が導かれることを示す。

$\lambda = p(1-q)/q(1-p)$ より $f(x, y)$ は次のように変形される。

$$f(x, y) = {}_n C_x \left(\frac{q\lambda}{1-q+q\lambda} \right)^x \left(\frac{1-q}{1-q+q\lambda} \right)^{n-x} \cdot {}_m C_y q^y (1-q)^{m-y}$$

ゆえに, $x = 0$ などの特殊な場合を除き

$$\frac{\partial f(x, y)}{\partial q} = {}_n C_x \left(\frac{q\lambda}{1-q+q\lambda} \right)^x \left(\frac{1-q}{1-q+q\lambda} \right)^{n-x} \cdot {}_m C_y q^y (1-q)^{m-y} \\ \times \left\{ \frac{x+y}{q(1-q)} - \frac{n\lambda + m(1-q) + mq\lambda}{(1-q)(1-q+q\lambda)} \right\}$$

が成り立つ。(特殊な場合は形式的に同じ形に帰着する。)

したがって, $\sum_{x=\max(0, k-m)}^{\min(k, n)} f(x, k-x)$ を最大にする必要十分条件は

$$\frac{k}{q(1-q)} - \frac{n\lambda + m(1-q) + mq\lambda}{(1-q)(1-q+q\lambda)} = 0$$

であることが分かる。(各項の増減が同調することに注意する。)

すなわち, $np + mq = k$ が必要十分条件となる。

□

5 比率の差の信頼区間の計算式

前節の構成法 1 によりオッズ比から比率の差を計算する式を求める。

オッズ比の信頼区間 $\lambda_1 \leq \lambda \leq \lambda_2$ が与えられたとき, それに対応する比率の差の信頼区間は次の定理を用いて $p(\lambda_1) - q(\lambda_1) \leq p - q \leq p(\lambda_2) - q(\lambda_2)$ で得られる。

定理 2 オッズ比 λ に対して $np + mq = k$ を満たす (p, q) は以下の式のように一意に与えられる。

$$p(\lambda) = \begin{cases} \frac{k}{m+n} & \lambda = 1 \text{ のとき} \\ \frac{-k + m + k\lambda + n\lambda - \sqrt{(k-m-k\lambda-n\lambda)^2 - 4kn\lambda(\lambda-1)}}{2n(\lambda-1)} & \lambda \neq 1 \text{ のとき} \end{cases}$$

$$q(\lambda) = \frac{k - np(\lambda)}{m}$$

(証明) $\lambda = p(1-q)/q(1-p)$ と $np + mq = k$ を連立させると以下の 2 次方程式となる。

$$g(p) = Ap^2 + Bp + C = 0$$

ただし、

$$\begin{aligned} A &= n(\lambda - 1) \\ B &= k - m - k\lambda - n\lambda \\ C &= k\lambda \end{aligned}$$

である。

ゆえに、 $\lambda = 1$ のときは自明。

$\lambda \neq 1$ のとき、2 次方程式の解は次の 2 つになる。

$$\begin{aligned} p^+ &= \frac{-k + m + k\lambda + n\lambda + \sqrt{(k - m - k\lambda - n\lambda)^2 - 4kn\lambda(\lambda - 1)}}{2n(\lambda - 1)} \\ p^- &= \frac{-k + m + k\lambda + n\lambda - \sqrt{(k - m - k\lambda - n\lambda)^2 - 4kn\lambda(\lambda - 1)}}{2n(\lambda - 1)} \end{aligned}$$

したがって、この 2 つの解のうち p^- のみが適していることを示せばよい。

$\lambda < 1$ のとき、 $g(0) = k\lambda > 0$ で $g(1) = k - m - n \leq 0$ となることから $p^+ < 0 < p^- \leq 1$ が分かる。

$\lambda > 1$ のとき、 $g(0) = k\lambda > 0$ で $g(1) = k - m - n \leq 0$ となる。ゆえに、 $k \neq m + n$ のときは、 $0 < p^- < 1 < p^+$ が分かる。 $k = m + n$ のときは $p^- = 1$ となり、 $p^+ = (n+m)\lambda/n(\lambda-1) > 1$ となる。

よって、題意は示された。

□

6 検定の種類による信頼区間の違い

検定に対応する信頼区間を与えるためには検定自体をどのように行うかが影響してくる。通常片側検定に対応する信頼区間を構成することはない。従来は片側 5% の検定に対して信頼区間の有意水準を 90% とすることが多かったが、最近では両側検定との違いをなくすため片側検定を 2.5% で行い信頼区間を 95% で構成することが増えているようである。

さてその際にこの信頼区間を両側検定に対応するように構成したならそれは片側検定の結果と一致するのであろうか。ここで問題にしている一般超幾何分布の場合、両側検定で左右に割り振る確率がほとんどの場合 2.5% ずつとはならないため一致しない。

そのような両側検定の特性を合理的でないと考える立場から左右に 2.5% ずつ分けて考える検定法が提案されている。(椿 (1982), 佐藤 (1993) 参照) それを片側 $\times 2$ 検定と呼ぶ。この検定に対応するように信頼区間を構成すると片側検定の結果とも一致して都合がよい。本論文でも 2×2 分割表での検定として片側 $\times 2$ 検定を推奨する。次節の具体例ではすべて片側 $\times 2$ 検定で行う。

7 例

提案する方法による信頼区間の様子を調べるため具体例をいくつかあげる。ここで示す数値計算結果は自作のプログラミングによるものか統計パッケージによるものである。(浜田 (1997) 参照)

7.1 実例での導出

制酸剤の効果を調べたある試験において胃液の pH 値に関する以下のような結果を得た。

	pH \geq 2.5	pH $<$ 2.5
薬剤 F	118	1
薬剤 R	111	4

Fisher の片側検定の p 値は 0.207 となり, 有意な差は見られない。

一方, オッズ比の正確な信頼区間は

$$0.4103 \leq \lambda \leq 211.15$$

となる。ゆえに, 5 節の計算式より

$$A = -70.1743, \quad 25007.85$$

$$B = -28.7844, \quad -73366.2$$

$$C = 93.9587, \quad 48353.35$$

となり, 提案する方法に基づく信頼区間 (以下では「提案する区間」と呼ぶ) は

$$-0.01743 \leq p - q \leq 0.04304$$

となる。

7.2 信頼区間の違いを示す例

以下の例は信頼区間の違いを示すための模造例である。ここで比較するのは無条件の下で攪乱母数を動かして求めた信頼区間 (以下では「無条件での区間」と呼ぶ) と条件付の下で攪乱母数を動かして求めた信頼区間 (以下では「条件付での区間」と呼ぶ) と提案する区間である。

	有効	無効
処理 1	9	1
処理 2	5	5

Fisher の片側検定の p 値は 0.0704 となる。したがって, 有意水準 86% の信頼区間を作ればぎりぎり棄却されないため比率差 0 を含む区間になるはずである。それぞれの方法による結果は下記の通りである。

無条件での区間	[0.022, 0.697]
条件付での区間	[-0.0006, 0.855]
提案する区間	[-0.0006, 0.582]

無条件での区間は 0 を含まない区間となっており, Fisher の検定結果と食い違う。また, 条件付での区間は上限が非常に大きくなっている。

一方, 有意水準 80% の信頼区間を作ると比率差 0 を含まない区間となるはずであるが,

無条件での区間	[0.066, 0.668]
条件付での区間	(0, 0.827]
提案する区間	[0.037, 0.576]

となり, 条件付での区間は片側が開区間となる。棄却される場合の条件付での区間はこのように 0 を含まないこと以外下限について何もいえないものになってしまう。また, データのよっては上限についても 1 未満となってしまう場合がある。これが先に述べた条件付での区間が使用に耐えない理由である。

8 おわりに

Fisher の正確確率検定に対応する比率の差の信頼区間を提案するのが本論文の目的であったわけだが, 実用性の高い方法を提案できたと思う。

プログラミングによって信頼区間を計算していた当初は 2 つの構成法が同一の結果を出すとは思っていなかった。定理の形できれいにまとめたことは思った以上の収穫であったといえる。

謝辞

適切な助言をいただいた神戸大学の稲葉太一講師に感謝いたします。

参考文献

- 佐藤俊哉 (1993): 3BISO の質疑応答より, “医薬安全性研究会会報”, **38**, 30-32.
 竹内啓 他 (1989): “統計学辞典”, 東洋経済新報社.
 椿広計 (1982): 2×2 分割表に関する 2,3 の話題, *Clin. Eval.*, **10**, 3-17.
 浜田知久馬 (1997): SAS による正確な検定, “第 16 回日本 SAS ユーザー会研究発表会論文集”, 17-34.