

# 母比率の区間推定（大標本）の公式の導出と 推定の計算に関する一考察

石 川 忠 孝

## 1 はじめに

母比率の区間推定（大標本）の公式の導出には

- ① 厳密性を重視した方法
- ② 統計的な考え方を利用した方法

の2通りがある。①の方法はめんどろな計算があるためか、ほとんどのテキストは②の方法を用いている。この方法では

$$\hat{p} - z\left(\frac{\alpha}{2}\right)\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right)\sqrt{\frac{p(1-p)}{n}}$$

から、根号の中の  $p$  を  $\hat{p}$  でおきかえるとき「 $\hat{p}$  は  $p$  に近いとみなしてよいから」と述べて

$$\hat{p} - z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

を導いている。初学者はそれなら最初から  $p$  を出すのに  $\hat{p}$  と同じでいいではないかと疑問を持ち、十分な納得がいかないものがある。そこで、明確な理解が得られるような公式の導出と推定の計算について考察する。

## 2 母比率の区間推定（大標本）の公式の導出

- (1) 厳密性を重視した方法

母比率  $p$  の事象が、 $n$  個の独立標本中に現れる標本比率を  $\hat{p} = X/n$  とする

と

$P\left(\frac{X}{n} = \frac{x}{n}\right) = P(X = x) = {}_nC_x p^x (1-p)^{n-x}$  であり,  $\hat{p} = \frac{X}{n}$  の分布は二項分布であるから, 標本数  $n$  が十分大きいときは, 標本比率  $\hat{p} = X/n$  の分布は  $N(p, p(1-p)/n)$  に従う。すなわち,  $z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$  は近似的に  $N(0, 1)$  に従う。

〈注〉 標本比率の分布について詳述すると次のようになる。

大きさ  $n$  の標本をとるとき, 確率変数を  $X_1, X_2, \dots, X_n$  とする。 $i$  番目にある特性を持つものを抽出したとき  $X_i=1$ , それ以外のものを抽出したとき  $X_i=0$  とする。このとき, 標本比率を  $\hat{p}$  とすると

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{X_1}{n} + \frac{X_2}{n} + \dots + \frac{X_n}{n}$$

である。母比率が  $p$  であれば

$$P(X_1 = 1) = P(X_2 = 1) = \dots = P(X_n = 1) = p$$

であるから,  $X = X_1 + X_2 + \dots + X_n$  は二項分布  $B(n, p)$  に従う。このとき

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_n) = np$$

$$V(X) = V(X_1) + V(X_2) + \dots + V(X_n) = np(1-p)$$

したがって

$$\begin{aligned} E(\hat{p}) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} \{E(X_1) + E(X_2) + \dots + E(X_n)\} \\ &= \frac{1}{n} \cdot E(X) = p \end{aligned}$$

$$V(\hat{p}) = \frac{1}{n^2} \{V(X_1) + V(X_2) + \dots + V(X_n)\} = \frac{1}{n^2} \cdot V(X) = \frac{p(1-p)}{n}$$

よって,  $n$  が十分に大きいとき,  $\hat{p}$  の分布は正規分布  $N(p, \frac{p(1-p)}{n})$  で近似できる, すなわち,  $z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$  は  $N(0,1)$  で近似できる。

したがって,

$$P\left(\left|\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right| \leq z\left(\frac{\alpha}{2}\right)\right) = 1-\alpha$$

$P$  の中味の不等式より

$$|p - \hat{p}| \leq z\left(\frac{\alpha}{2}\right) \sqrt{\hat{p}(1-\hat{p})/n} \quad (1)$$

$$p^2 - 2\hat{p}p + \hat{p}^2 \leq z\left(\frac{\alpha}{2}\right)^2 \hat{p}(1-\hat{p})/n$$

$$p^2 + \frac{z\left(\frac{\alpha}{2}\right)^2}{n} p^2 - 2\hat{p}p - \frac{z\left(\frac{\alpha}{2}\right)^2}{n} p + \hat{p}^2 \leq 0$$

これを整理すると

$$\frac{n + z\left(\frac{\alpha}{2}\right)^2}{n} p^2 - \left(2\hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{n}\right) p + \hat{p}^2 \leq 0 \quad (2)$$

$p$  の 2 次方程式

$$\frac{n + z\left(\frac{\alpha}{2}\right)^2}{n} p^2 - \left(2\hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{n}\right) p + \hat{p}^2 = 0 \text{ を解くと}$$

$$\begin{aligned} P &= \frac{\hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} \pm \sqrt{\left(\hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n}\right)^2 - \frac{n + z\left(\frac{\alpha}{2}\right)^2}{n} \cdot \hat{p}^2}}{\frac{n + z\left(\frac{\alpha}{2}\right)^2}{n}} \\ &= \frac{n}{n + z\left(\frac{\alpha}{2}\right)^2} \left( \hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} \pm z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z\left(\frac{\alpha}{2}\right)^2}{4n^2}} \right) \end{aligned}$$

よって、②の解は

$$\begin{aligned} \frac{n}{n + z\left(\frac{\alpha}{2}\right)^2} \left( \hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z\left(\frac{\alpha}{2}\right)^2}{4n^2}} \right) &\leq p \leq \\ \frac{n}{n + z\left(\frac{\alpha}{2}\right)^2} \left( \hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z\left(\frac{\alpha}{2}\right)^2}{4n^2}} \right) &\quad (3) \end{aligned}$$

$n$  が十分大きいから

$$\frac{z\left(\frac{\alpha}{2}\right)^2}{2n} \doteq 0, \frac{z\left(\frac{\alpha}{2}\right)^2}{4n^2} \doteq 0, \frac{n}{n+z\left(\frac{\alpha}{2}\right)^2} \doteq 1 \text{ とし, } \frac{1}{n} \text{ も小さい数だが, これ}$$

は  $\hat{p}(1-\hat{p})$  の係数となっているから残すと

$$\hat{p} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (4)$$

①から④を求めるには、理論的には簡単に求められるわけであるが、実際には必ずしも簡単ではない。また、③から④への変形は次のようにもできる。

③から

$$\frac{n}{n+z\left(\frac{\alpha}{2}\right)^2} \left( \hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} - z\left(\frac{\alpha}{2}\right) \frac{1}{\sqrt{n}} \sqrt{\hat{p}(1-\hat{p}) + \frac{z\left(\frac{\alpha}{2}\right)^2}{4n}} \right) \leq p \leq$$

$$\frac{n}{n+z\left(\frac{\alpha}{2}\right)^2} \left( \hat{p} + \frac{z\left(\frac{\alpha}{2}\right)^2}{2n} + z\left(\frac{\alpha}{2}\right) \frac{1}{\sqrt{n}} \sqrt{\hat{p}(1-\hat{p}) + \frac{z\left(\frac{\alpha}{2}\right)^2}{4n}} \right)$$

$n$  は十分大きいから、 $n / \left( n + z\left(\frac{\alpha}{2}\right)^2 \right) \doteq 1$ ,  $\frac{1}{\sqrt{n}}$  より小さい  $\frac{1}{n}$  程度の微小量

$z\left(\frac{\alpha}{2}\right)^2 / n \doteq 0$  として

$$\hat{p} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

このほうがいくぶん分かりやすい。

## (2) 統計的な考え方を利用した方法

統計の知識を実際に使う場合には、簡単な手続で手軽に結論が得られるというのも大切な点である。

母比率  $p$  の事象が、 $n$  個の独立標本中に現れる標本比率を  $\hat{p} = X/n$  とする

と

$P\left(\frac{X}{n} = \frac{x}{n}\right) = P(X = x) = {}_nC_x p^x (1-p)^{n-x}$  であり,  $\hat{p} = \frac{X}{n}$  の分布は二

項分布であるから, 標本数  $n$  が十分大きいときは, 標本比率  $\hat{p} = X/n$  の分布は

$N(p, p(1-p)/n)$  に従う。すなわち,  $z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$  は近似的に  $N(0, 1)$  に従

う。

したがって,

$$P\left(\left|\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right| \leq z\left(\frac{\alpha}{2}\right)\right) = 1-\alpha$$

$P$  の中味の不等式より

$$-z\left(\frac{\alpha}{2}\right) \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z\left(\frac{\alpha}{2}\right)$$

$$\hat{p} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{p(1-p)}{n}}$$

$n$  が十分大きいので, 母比率  $\doteq$  標本比率 すなわち  $p \doteq \hat{p}$  と考え,  $\sqrt{\quad}$  の中の  $p$  を  $\hat{p}$  で代用して

$$\therefore \hat{p} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$p$  を  $\hat{p}$  でおきかえるとき, 「 $\hat{p}$  は  $p$  に近いとみなしてよいから」とだけ説明すると, 「それなら最初から  $p$  を出すのに  $\hat{p}$  と同じでいいではないか」という疑問を持つ者がいる。しかし, それならば,  $p$  は  $\hat{p}$  の近くというだけで, どれくらい近くかということが明確に表示できない。そこで

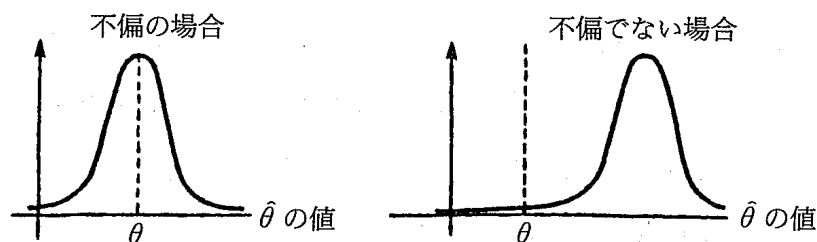
$$\left[ \hat{p} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

なる不等式が成立しているから,  $p$  の範囲を出すために,  $p$  に十分近い  $\hat{p}$  を左右の項に代入すると上の不等式より少し荒くなるかも知れないが

$$\hat{p} - z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

という不等式が成立して、 $p$ の範囲が表示できる。」と説明してもいまひとつすっきりとはしない。 $\hat{p}$ は標本ごとに値が変わるのに、それを真の値 $p$ の代わりに使用することが理解できにくいようである。厳密性を重視した方法や $\hat{p}$ が $p$ の最も望ましい点推定量であることが分かっている者は納得できるが、初めて学ぶ者には何か奇異にさえ感じられる。手元の文献を調べるとほとんどは(31冊中26冊)は「 $p$ を $\hat{p}$ で代用する」とだけ説明している。他には「 $p$ はその不偏推定量 $\hat{p}$ でおきかえる」(3冊)とか「 $p$ はその一致推定量 $\hat{p}$ でおきかえる」(2冊)、「根号 $\sqrt{\quad}$ の中にも $p$ があるが、これを第1近似 $\hat{p}$ で置き換えてもそれほど悪くないだろう」(1冊)がある。

不偏性、有効性、十分性は $n$ が有限のときの点推定量の性質であり、一致性は $n$ が極めて大きい場合の点推定量の性質であるから、「 $p$ をその一致推定量でおきかえる」という説明では不完全である。また、推定量が不偏であるということは誤差に規則的な偏りがないというだけのことであるから、「 $p$ をその不偏推定量でおきかえる」という説明だけではこれまた不完全である。 $\hat{p}$ を $p$ に代用するためには、 $n$ が十分大きいときには、 $p \doteq \hat{p}$ でなければならない。そのた



図—1  $\hat{\theta}$ の標本分布と不偏推定量

左の場合 $\hat{\theta}$ は真の値の回りに散布しており、好ましい推定量と考えられるが右の場合は真の値から外れており好ましい推定量とは言えない。たて軸は $\hat{\theta}$ の確率密度関数の値、なお、不偏性は平均的な原理であって、具体的な現実の $\hat{\theta}$ の現実の値が $\theta$ に等しいことを言っているのではない。

めには  $\hat{p}$  が  $p$  の不偏推定量でありかつ一致推定量でなければならない。不偏性、一致性は推定量が最小限満たす必要のある2つの性質である。後述するように  $\hat{p}$  は  $p$  を推定するための最適な推定量である。区間推定法の前に点推定法について説明しておればよいが、テキストによっては点推定法を取り扱っていないものもある。したがっ

て、「標本比率  $\hat{p}$  は母比率  $p$  を推定するために最適な推定量であることが知られているから、 $p$  を  $\hat{p}$  でおきかえる」と説明するのがよい。これならば厳密性を重視した方法や点推定法を学習しなくても納得できるだろう。なお明確な理解が必要ならばそのときに厳密性を重視した方法や点推定法を指導すればよい。

例題1 表1は1995年国勢調査の結果による就業者の所属産業を、1%抽出集計（速報集計）と全数集計（確定集計）との結果について比較したものである。抽出された就業者総数  $n = 639041$  のうち、サービス業就業者数は  $x = 156541$  であった。したがって全国就業者のなかのサービス業就業者比率の推定値は

$$\hat{p} = \frac{x}{n} = \frac{156541}{639041} = 0.2450$$

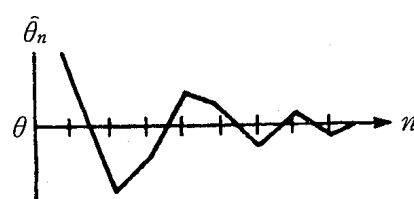
であって、推定値の標準誤差は  $p = \hat{p}$  とおいて

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.2450 \times 0.7550}{639041}} = 0.000538$$

となる。ところで全数集計の結果によると、サービス業就業者の比率は

$$p = \frac{15932490}{64141544} = 0.2484$$

であって、標本比率との差  $0.2450 - 0.2484 = -0.0034$  の絶対値は標準誤差の3倍より大である。これは国勢調査の抽出集計では抽出作業の便宜上個人単位で



図—2  $\hat{\theta}_n$  が一致推定量である場合  
不偏性と異なり、一致性は  $\hat{\theta}$  の値が  
現実に  $n$  とともに  $\theta$  に近づいていく  
ことを要求している ( $\hat{\theta}_n$  は  $n$  によるこ  
とを示す)。

表一 産業別就業者数 (1995年)

産 業	1%抽出集計		全数集計	
	就業者数	千分比	就業者数	千分比
農 業	3,498,900	54.8	3,426,497	53.4
林 業	84,600	1.3	85,824	1.3
漁 業	311,900	4.9	307,528	4.8
鉱 業	58,300	0.9	60,597	0.9
建設業	6,617,500	103.6	6,630,578	103.4
製造業	13,376,500	209.3	13,556,253	211.3
電気ガス・水道業	348,500	5.5	364,183	5.7
運輸・通信業	3,904,200	61.1	3,890,110	60.6
卸売・小売業	14,761,700	231.0	14,618,405	227.9
金融・保険業	1,972,100	30.9	1,974,508	30.8
不動産業	739,600	11.6	707,149	11.0
サービス業	15,654,100	245.0	15,932,490	248.4
公務	2,146,700	33.6	2,155,214	33.6
分類不能の産業	429,600	6.7	432,208	6.7
総 計	63,904,100	1,000.0	64,141,544	1,000.0

なく、一般世帯では世帯を抽出単位として標本を抽出しているため、標本は就業者単位の単純確率標本になっていないからだと思われる。なお正確な母集団比率を使つてのサービス業就業者の標本比率の標準誤差も計算してみると、全数集計結果のサービス業就業者の比率は  $p = 0.2484$  であるから

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2484 \times 0.7516}{639041}} = 0.000541$$

となる。標本比率を代用して計算した上記の結果との差はきわめてわずかである。なお標本比率  $\frac{x}{n} = 0.2450$  との差  $0.2450 - 0.2484 = -0.0034$  の絶対値は上記標準誤差の3倍以上になっておる。これは前述のように、抽出方法が就業者一人ひとりの抽出単位としての抽出でなく、世帯を抽出単位として抽出が行われたためと思われる。



### 3 推定量の条件と最尤法

#### (1) 推定量の条件

##### a 不偏性

標本比率  $\hat{p} = \frac{X}{n}$  は母比率  $p$  を点推定するために望ましい条件を満たすことを述べる。

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$$

ゆえに  $\hat{p} = \frac{X}{n}$  は  $p$  の不偏推定量である。

##### b 有効性

クラメル・ラオの不等式

$$V(\hat{\theta}) \geq \frac{1}{nE\left[\left\{\frac{\partial}{\partial\theta}\log f(X; \theta)\right\}^2\right]} = \frac{1}{n\int_{-\infty}^{\infty}\left\{\frac{\partial\log f(x; \theta)}{\partial\theta}\right\}^2 f(x; \theta)dx}$$

で、等号が成り立つならば、その推定量は有効推定量である。

クラメル・ラオの不等式の左辺は

$$V(\hat{p}) = V(X/n) = \frac{1}{n^2}V(X)$$

$$= \frac{1}{n^2} \cdot npq = \frac{pq}{n}$$

一方、右辺は  $f(X; p) = p^X(1-p)^{1-X}$

( $X=0,1$ )より

$$\begin{aligned} \frac{1}{nE\left[\left\{\frac{\partial}{\partial p}\log f(X; p)\right\}^2\right]} &= \frac{1}{nE\left[\left\{\frac{\partial}{\partial p}\log p^X(1-p)^{1-X}\right\}^2\right]} \\ &= \frac{1}{nE\left[\left\{\frac{\partial}{\partial p}(X\log p + (1-X)\log(1-p))\right\}^2\right]} = \frac{1}{nE\left[\left\{\frac{X}{p} - \frac{1-X}{1-p}\right\}^2\right]} \end{aligned}$$

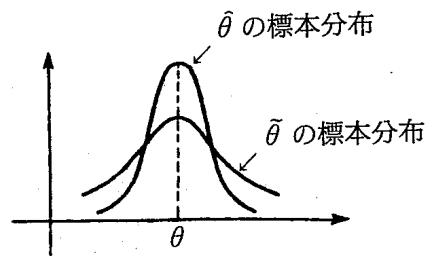


図-3 不偏かつ一致推定量の比率

二つの推定量  $\hat{\theta}$  と  $\tilde{\theta}$  を比較する場合、 $\hat{\theta}$  の標本分布の方が  $\theta$  の回りに集中しており、 $\hat{\theta}$  の方が望ましい推定量である。これを、 $\hat{\theta}$  は  $\tilde{\theta}$  より有効である、という。

$$= \frac{1}{nE\left[\left\{\frac{X-p}{p(1-p)}\right\}^2\right]} = \frac{p^2q^2}{nE[\{X-p\}^2]} = \frac{p^2q^2}{nV(X)} = \frac{pq}{n}$$

となって等式が成り立つ。

ゆえに  $\hat{p} = \frac{X}{n}$  は  $p$  の有効推定量である。

c 十分性

$X_1, X_2, \dots, X_n$  の同時確率分布が,  $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$  より

$$\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = p^{n\hat{p}}(1-p)^{n(1-\hat{p})} \times 1 = g(p, \hat{p}) \times h(x_1, x_2, \dots, x_n)$$

と書けるので十分性が示された。

ゆえに  $\hat{p} = \frac{X}{n}$  は  $p$  の十分推定量である。

d 一緻性

チェビシェフの定理より

$$P\{|X-m| < k\sigma\} \geq 1 - \frac{1}{k^2} \quad (k \text{ は任意の正数})$$

において,  $X$  のかわりに  $X/n$  を代入すると

$$m \rightarrow E(X/n) = p \quad \sigma \rightarrow \sigma(X/n) = \sqrt{pq/n}$$

となり, 上の定理は次のように変形される。

$$P\left\{\left|\frac{X}{n} - p\right| < k\sqrt{\frac{pq}{n}}\right\} \geq 1 - \frac{1}{k^2}$$

$k$  は任意の正数であるから

$$k\sqrt{\frac{pq}{n}} = \varepsilon \text{ すなわち } \frac{1}{k^2} = \frac{pq}{n\varepsilon^2} \text{ とおくと}$$

$$P\left\{\left|\frac{X}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{pq}{n\varepsilon^2}$$

$n \rightarrow \infty$  とすれば, 右辺  $\rightarrow 1$  となる。

$$\therefore \lim_{n \rightarrow \infty} P\left\{\left|\frac{X}{n} - p\right| < \varepsilon\right\} = 1$$

よって  $\hat{p} = \frac{X}{n}$  は  $n \rightarrow \infty$  のとき  $p$  に確率収束する。

ゆえに  $\hat{p} = \frac{X}{n}$  は  $p$  の一致推定量である。

以上のように,  $\hat{p} = \frac{X}{n}$  は  $p$  の不偏推定量, 有効推定量, 一致推定量, 十分推定量であるから,  $\hat{p}$  は  $p$  の最適な推定量である。

## (2) 最尤法

推定統計量に対して求められる性質には, 不偏性, 有効性, 一致性, 十分性がある。これらの望ましい性質をもつような推定統計量の求め方が最尤法である。

母比率を  $p$  とする二項母集団を考える。この二項母集団からの標本を  $X_1, X_2, \dots, X_n$ , データ (実現値) を  $x_1, x_2, \dots, x_n$  とする。

尤度関数  $L(p)$  は

$$L(p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}$$

となる。 $\sum_{i=1}^n x_i$  は標本和  $S_n$  のデータ (実現値) と考えればよいから,  $S_n$  (データの和) と書くと

$$L(p) = p^{S_n} (1-p)^{n-S_n}$$

$$\log L(p) = S_n \log p + (n - S_n) \log(1-p)$$

これを  $p$  で微分すると

$$\frac{d}{dp} \log L(p) = \frac{S_n}{p} - \frac{n - S_n}{1-p} = \frac{(1-p+p)S_n - np}{p(1-p)} = \frac{S_n - np}{p(1-p)}$$

$$= \frac{-n(p - \frac{S_n}{n})}{p(1-p)}$$

よって  $p = \frac{S_n}{n}$  のとき  $L(p)$  は極大か

$p$	0		$\frac{S_n}{n}$		1
$\frac{d}{dp} \log L(p)$		+	0	-	
$L(p)$		↗	極大	↘	

つ最大であるから、母比率  $p$  の最尤推定量は

$$\hat{p} = s_n/n = \sum_{i=1}^n x_i/n = \bar{x}$$

しかるに  $\sum_{i=1}^n x_i$  は  $n$  回中の事象  $A$  の起こる回数を表しているから

$$\hat{p} = \frac{X}{n} \text{ (標本比率)}$$

#### 4 2つの解法の比較

母比率の区間推定の求め方には

- ① 公式を利用する解法
- ② 厳密性を重視した2次不等式による解法

の2通りがある。次にこの2通りで問題を解いて考察する。

例題2 ある市で100人を選んで、内閣を支持するか否かを質問したところ65人から支持の回答を得た。信頼度95%で支持率の信頼区間を求めよ。

解1 公式を利用する解法

まず、正規分布への近似条件  $np \geq 5$ ,  $nq \geq 5$  を満たしているかを調べる。

$$np = 100 \times 0.65 = 65 \geq 5, \quad nq = 100 \times 0.35 = 35 \geq 5$$

よって、正規分布へ近似可能であるといえる。

$$\hat{p} = 0.65 \text{ より}$$

$$0.65 - 1.96\sqrt{\frac{0.65 \times 0.35}{100}} < p < 0.65 + 1.96\sqrt{\frac{0.65 \times 0.35}{100}}$$

$$0.65 - 0.093 < p < 0.65 + 0.093$$

$$\therefore 0.557 < p < 0.743$$

解2 厳密性を重視した2次不等式による解法

まず、正規分布への近似条件  $np \geq 5$ ,  $nq \geq 5$  を満たしているかを調べる。

$$np = 100 \times 0.65 = 65 \geq 5, \quad nq = 100 \times 0.35 = 35 \geq 5$$

よって、正規分布へ近似可能であるといえる。

$n$  が大きいとき、信頼度 95% の母比率の信頼区間は次の不等式で定められる。

$$-1.96 < \frac{\frac{x}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} < 1.96 \quad (\text{ただし標本比率 } \hat{p} = \frac{x}{n})$$

この不等式は次のように書き直すことができる。

$$\left| \frac{x}{n} - p \right| < 1.96 \sqrt{\frac{p(1-p)}{n}}$$

$$\left( \frac{x}{n} - p \right)^2 < 1.96^2 \frac{p(1-p)}{n}$$

これを整理すると

$$p^2(n + 1.96^2) - p(2x + 1.96^2) + \frac{x^2}{n} < 0$$

$n = 100, x = 65$  のとき

$$p^2(100 + 3.84) - p(2 \times 65 + 3.84) + \frac{65^2}{100} < 0 \quad (\text{ただし } 1.96^2 = 3.84)$$

$$10384p^2 - 13384p + 4225 < 0 \quad \text{①}$$

$$10384p^2 - 13384p + 4225 = 0 \quad \text{②}$$

より

$$p = \frac{6692 \pm \sqrt{6692^2 - 10384 \times 4225}}{10384} = 0.6445 \pm 0.0919$$

$$\therefore 0.553 < p < 0.736$$

②は  $p$  の 2 次方程式であるから、理論的には簡単に求めることができるわけであるが、具体的な数値計算の場合には必ずしも簡単でない。したがって、 $p$  の 2 次不等式①を解いて信頼区間を求めるのは、厳密であるが手間がかかりめんどうである。もちろん理論的にはこのほうが、 $p$  の代わりに  $\hat{p}$  を代用して求める（公式の利用の解法）よりもよいわけであるが、いずれにしても近似計算であるから一長一短である。解 1 による信頼区間は解 2 による信頼区間より少し高い方向へずれているが、解 1 は二項分布の正規分布による近似と、真の  $p$  を

推定値  $\hat{p}$  で代用することの2つの近似をしているわけで、これが誤差の原因である。解2は二項分布の正規分布による近似だけである。したがって、めんどくでも2次不等式を解く解2の解法のほうが厳密でよいということになるわけである。しかし、解1と解2による信頼区間はほとんど等しく、違いは極めてわずかである。なお、手元の文献を調査すると解答にあたり正規分布による近似条件を満たしているかを調べているのは、1冊だけである。もちろん、そこで取り扱っているのは正規分布による近似条件が成り立つ問題ばかりだから、省略しているのであろう。本稿では厳密さと、その後の考察のため各例題とも正規分布による近似条件を満たしているかを調べた。普通に使用されている正規分布による近似条件  $np \geq 5$ ,  $nq \geq 5$  が成立しても、数学的に成り立つことが証明される近似条件  $9/(n+9) < p < n/(n+9)$  が成り立たないときもある。

例題3 サイコロを100回ふって、1の目が10回出たとする。このとき、このサイコロの1の目が出る確率はいくらか。信頼度95%で推定せよ。

解1 公式を利用する解法

まず、正規分布への近似条件  $np \geq 5$ ,  $nq \geq 5$  を満たしているかを調べる。

$$np = 100 \times \frac{10}{100} = 10 \geq 5, \quad nq = 100 \times \frac{90}{100} = 90 \geq 5$$

よって、正規分布へ近似可能であるといえる。

$$\hat{p} = \frac{10}{100} = 0.1 \text{ より}$$

$$0.1 - 1.96 \sqrt{\frac{0.1 \times 0.9}{100}} \leq p \leq 0.1 + 1.96 \sqrt{\frac{0.1 \times 0.9}{100}}$$

$$0.1 - 0.0588 \leq p \leq 0.1 + 0.0588$$

$$\therefore 0.0412 \leq p \leq 0.1588$$

解2 厳密性を重視した2次不等式による解法

まず、正規分布への近似条件  $np \geq 5$ ,  $nq \geq 5$  を満たしているかを調べる。

$$np = 100 \times \frac{10}{100} = 10 \geq 5, \quad nq = 100 \times \frac{90}{100} = 90 \geq 5$$

よって、正規分布へ近似可能であるといえる。

$$-1.96 \leq \frac{\frac{x}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.96 \quad (\text{ただし, 標本比率 } \hat{p} = \frac{x}{n})$$

この不等式は次のように書き直すことができる。

$$\left| \frac{x}{n} - p \right| \leq 1.96 \sqrt{\frac{p(1-p)}{n}}$$

$$\left( \frac{x}{n} - p \right)^2 \leq 1.96^2 \frac{p(1-p)}{n}$$

これを整理すると

$$p^2(n + 1.96^2) - p(2x + 1.96^2) + \frac{x^2}{n} < 0$$

$n = 100, x = 10$  のとき

$$p^2(100 + 3.84) - p(2 \times 10 + 3.84) + \frac{10^2}{100} \leq 0 \quad (\text{ただし, } 1.96^2 = 3.84)$$

$$10384p^2 - 2384p + 100 \leq 0$$

$$10384p^2 - 2384p + 100 = 0$$

を解くと

$$p = \frac{1192 \pm \sqrt{1192^2 - 10384 \times 100}}{10384} = 0.114792 \pm 0.059557$$

$$= 0.174349, 0.055235$$

$$\therefore 0.055 \leq p \leq 0.174$$

$n$  をもっと大きく、例えば  $n = 10,000, x = 1,000$  の場合には  $p$  の 95% の信頼区間は、解 1 によれば、 $0.0941 \leq p \leq 0.1059$  となり、また解 2 によれば  $0.0943 \leq p \leq 0.1060$  となって、両者の差異は極めて小さくなることが分かる。

## 5 ま と め

母比率の区間推定（大標本）の公式の導出については、厳密性を重視した方

法は理論的には簡単に求められるわけであるが、実際には必ずしも簡単ではない。統計的な考え方を利用した方法は、統計的な知識を使用すると簡単な手続きで手軽に公式が得られる。しかし、「 $p$ を $\hat{p}$ で代用する」ところが理解できにくい。「標本比率 $\hat{p}$ は母比率 $p$ を推定するために最適な推定量であることが知られているから、 $p$ を $\hat{p}$ でおきかえる」と説明するのがよい。点推定法を学習すれば理解が深まるし、また厳密性を重視した方法を説明すれば、根号の中の $p$ を $\hat{p}$ におきかえるとよいことが明確になる。

推定の計算では、厳密性を重視した2次不等式による解法は数値計算がめんどうであるが、公式を利用する解法は非常に簡単に推定値が求まる。両者で求めた推定値の差異は極めてわずかであり、いずれにしても近似計算である。したがって、厳密性を特に要する場合は別として、一般的には計算が簡単である公式の利用による解法でよい。

#### 参 考 文 献

- 1) 石井博昭・塩出省吾・新森修一共著「確率統計の数理」裳華房, 1995年, P 60~P 68, P 177
- 2) 石川忠孝著「二項分布の正規分布による近似条件と比率の差の検定(大標本)について」松山大学論集第11巻第3号, 松山大学学術研究会, 1999年, P 56~P 60
- 3) 木村 等・大藪和雄・石川 浩共著「統計学入門」実教出版, 1996年, P 139~P 140
- 4) 小寺平治著「新統計入門」裳華房, 1998年, P 88~P 90
- 5) 森田優三・久次智雄共著「新統計概論改訂版」日本評論社, 1993年, P 235~P 237, P 261~P 267
- 6) 長坂建二著「統計学」放送大学教育振興会, 1996年, P 182~P 184, P 193~P 202
- 7) 西山 茂著「楽しい統計学セミナー—証拠を科学する方法—」同文館, 平成10年, P 41~P 46
- 8) 岡田泰栄著「統計」共立出版, 昭和52年, P 107~P 116, P 127~P 131
- 9) 高橋陸男・伊藤清三・山口清他14名共著「高等学校 確率・統計教授資料」数研出版, 昭和58年, P 201, P 207
- 10) 東京大学教養学部統計学教室編「統計学入門」東京大学出版会, 1997年, P 213~P 224, P 229~P 230
- 11) 宇田川正友著「教師のための初等数学講座, 確率と統計」岩崎書店, 昭和34年, P 140~



P 145

- 12) 宇喜多義昌著「統計数学入門」教育出版, 1980 年, P 132～P 135