

$$E(X^2) - \{E(X)\}^2 = \sum_{1 \leq j < k \leq n} (a_j - a_k)^2 p_j p_k \geq 0$$

かみ や だし
神谷 正

§1. はじめに

参考文献[1] p.61 の6行目に『注意 一般に、 $E(X^2) = \{E(X)\}^2$ は成り立たない。』と書かれている。これをきっかけに、本稿表題の計算結果を得た。もう少し詳しく、教科書の流れを説明する。

1個のさいころを投げて出る目を X とするとき、p.59 の例3で『確率変数 X の期待値は $E(X) = \frac{7}{2}$ 』であることを求め、次に p.61 の例5で『確率変数 X^2 の期待値は $E(X^2) = \frac{91}{6}$ 』を求め、その直後に『注意 一般に、 $E(X^2) = \{E(X)\}^2$ は成り立たない。』と書かれている。

そして、p.61 の後半から p.62 で、確率変数 X の分散 $V(X)$ が

『 $V(X) = E((X - m)^2) = \sum_{k=1}^n (x_k - m)^2 p_k$ 』の式で定義され、最後に、p.63 で

『 $V(X) = E(X^2) - \{E(X)\}^2$ 』が成り立つことが示されている。

このように、 $V(X) = E(X^2) - \{E(X)\}^2$ を使わずに、 $E(X^2) \neq \{E(X)\}^2$ を示さなければならない状況に立たされた。これは、授業の予習をしていたときのことであったが、生徒に授業中に話す話さないの如何に関わらず、不明瞭なことを放置したままでは、自信を持って授業ができない。

とりあえず、 $E(X^2) - \{E(X)\}^2$ を $n=2, 3, 4$ の場合で具体的に計算してみた。 $n=2$ では上手くいった。 $n=3$ でも上手くいった。これはいけるかも。 $n=4$ でも上手くいった。もう大丈夫だ。後は、一般の n の場合の証明を書くだけだ。

さて、本稿では、§2 で $n=2, 3, 4$ の場合の証明を書いた。これは、高校生でもできる。次に、§3 で一般の n の場合の証明を書いた。これは、 Σ の計算に堪能な高校生でなければ分からないかもしれない

い。ここまでの結果を踏まえ、§4 では、 $E(X^2) \neq \{E(X)\}^2$ や $E(X^2) = \{E(X)\}^2$ となるための必要十分条件を書いた。

本稿の内容をご存じの方々もおられるかもしれないが、啓蒙の目的で投稿する。

§2. $n=2, 3, 4$ の場合

【命題1】 確率変数 X の確率分布が

X	a	b	計
P	p	q	1

で与えられるとき

$$E(X^2) - \{E(X)\}^2 = (a - b)^2 pq$$

が成り立つ。

【証明】

$$\begin{aligned} E(X^2) - \{E(X)\}^2 &= (a^2 p + b^2 q) - (ap + bq)^2 \\ &= (a^2 p + b^2 q) - (a^2 p^2 + 2abpq + b^2 q^2) \\ &= a^2 p(1 - p) + b^2 q(1 - q) - 2abpq \\ &= a^2 pq + b^2 qp - 2abpq \\ &= (a - b)^2 pq \end{aligned}$$

よって $E(X^2) - \{E(X)\}^2 = (a - b)^2 pq$

【終】

【命題2】 確率変数 X の確率分布が

X	a	b	c	計
P	p	q	r	1

で与えられるとき

$$\begin{aligned} E(X^2) - \{E(X)\}^2 &= (a - b)^2 pq + (a - c)^2 pr + (b - c)^2 qr \\ &\text{が成り立つ。} \end{aligned}$$

【証明】

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= (a^2p + b^2q + c^2r) - (ap + bq + cr)^2 \\
 &= (a^2p + b^2q + c^2r) \\
 &\quad - (a^2p^2 + b^2q^2 + c^2r^2 + 2abpq + 2acpr + 2bcqr) \\
 &= a^2p(1-p) + b^2q(1-q) + c^2r(1-r) \\
 &\quad - 2abpq - 2acpr - 2bcqr \\
 &= a^2p(q+r) + b^2q(p+r) + c^2r(p+q) \\
 &\quad - 2abpq - 2acpr - 2bcqr \\
 &= (a^2 + b^2 - 2ab)pq + (a^2 + c^2 - 2ac)pr \\
 &\quad + (b^2 + c^2 - 2bc)qr \\
 &= (a-b)^2pq + (a-c)^2pr + (b-c)^2qr
 \end{aligned}$$

よって

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= (a-b)^2pq + (a-c)^2pr + (b-c)^2qr
 \end{aligned}$$

【終】

【命題 3】 確率変数 X の確率分布が

X	a	b	c	d	計
P	p	q	r	s	1

で与えられるとき

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= (a-b)^2pq + (a-c)^2pr + (a-d)^2ps \\
 &\quad + (b-c)^2qr + (b-d)^2qs + (c-d)^2rs
 \end{aligned}$$

が成り立つ。

【証明】

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= (a^2p + b^2q + c^2r + d^2s) - (ap + bq + cr + ds)^2 \\
 &= (a^2p + b^2q + c^2r + d^2s) \\
 &\quad - (a^2p^2 + b^2q^2 + c^2r^2 + d^2s^2) \\
 &\quad - 2(abpq + acpr + adps + bcqr + bdqs + cdrs) \\
 &= a^2p(1-p) + b^2q(1-q) + c^2r(1-r) + d^2s(1-s) \\
 &\quad - 2(abpq + acpr + adps + bcqr + bdqs + cdrs) \\
 &= a^2p(q+r+s) + b^2q(p+r+s) \\
 &\quad + c^2r(p+q+s) + d^2s(p+q+r) \\
 &\quad - 2(abpq + acpr + adps + bcqr + bdqs + cdrs) \\
 &= (a^2 + b^2 - 2ab)pq + (a^2 + c^2 - 2ac)pr \\
 &\quad + (a^2 + d^2 - 2ad)ps + (b^2 + c^2 - 2bc)qr \\
 &\quad + (b^2 + d^2 - 2bd)qs + (c^2 + d^2 - 2cd)rs \\
 &= (a-b)^2pq + (a-c)^2pr + (a-d)^2ps \\
 &\quad + (b-c)^2qr + (b-d)^2qs + (c-d)^2rs
 \end{aligned}$$

よって

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= (a-b)^2pq + (a-c)^2pr + (a-d)^2ps \\
 &\quad + (b-c)^2qr + (b-d)^2qs + (c-d)^2rs
 \end{aligned}$$

【終】

§3. 一般の n の場合

【命題 4】 確率変数 X の確率分布が

X	a_1	a_2	a_3	$\cdots$	a_n	計
P	p_1	p_2	p_3	$\cdots$	p_n	1

で与えられるとき

$$E(X^2) - \{E(X)\}^2 = \sum_{1 \leq j < k \leq n} (a_j - a_k)^2 p_j p_k$$

が成り立つ。

【証明】

$$\begin{aligned}
 & E(X^2) - \{E(X)\}^2 \\
 &= \left(\sum_{j=1}^n a_j^2 p_j \right) - \left(\sum_{j=1}^n a_j p_j \right)^2 \\
 &= \left(\sum_{j=1}^n a_j^2 p_j \right) - \left(\sum_{j=1}^n a_j p_j \right) \left(\sum_{k=1}^n a_k p_k \right) \\
 &= \left(\sum_{j=1}^n a_j^2 p_j \right) - \left(\sum_{1 \leq j \leq n, 1 \leq k \leq n} a_j a_k p_j p_k \right) \\
 &= \left(\sum_{j=1}^n a_j^2 p_j \right) - \left(\sum_{1 \leq j=k \leq n} a_j a_k p_j p_k \right) \\
 &\quad - \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) - \left(\sum_{1 \leq k < j \leq n} a_j a_k p_j p_k \right) \\
 &\quad \dots\dots ①
 \end{aligned}$$

右端の $\sum$ の項で、 j と k の記号を書きかえると

$$\begin{aligned}
 ① &= \left(\sum_{j=1}^n a_j^2 p_j \right) - \left(\sum_{j=1}^n a_j^2 p_j^2 \right) \\
 &\quad - \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) - \left(\sum_{1 \leq j < k \leq n} a_k a_j p_j p_k \right) \\
 &= \sum_{j=1}^n a_j^2 p_j (1 - p_j) - 2 \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) \\
 &= \sum_{j=1}^n a_j^2 p_j \left(\sum_{k \neq j, 1 \leq k \leq n} p_k \right) - 2 \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) \\
 &= \left(\sum_{k \neq j, 1 \leq j \leq n, 1 \leq k \leq n} a_j^2 p_j p_k \right) \\
 &\quad - 2 \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) \\
 &= \left(\sum_{1 \leq j < k \leq n} a_j^2 p_j p_k + \sum_{1 \leq k < j \leq n} a_j^2 p_j p_k \right) \\
 &\quad - 2 \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) \dots\dots ②
 \end{aligned}$$

2 番目の $\sum$ の項で, j と k の記号を書きかえると

$$\begin{aligned} ② &= \left(\sum_{1 \leq j < k \leq n} a_j^2 p_j p_k + \sum_{1 \leq j < k \leq n} a_k^2 p_k p_j \right) \\ &\quad - 2 \left(\sum_{1 \leq j < k \leq n} a_j a_k p_j p_k \right) \\ &= \sum_{1 \leq j < k \leq n} (a_j^2 p_j p_k + a_k^2 p_k p_j - 2 a_j a_k p_j p_k) \\ &= \sum_{1 \leq j < k \leq n} (a_j - a_k)^2 p_j p_k \end{aligned}$$

よって

$$E(X^2) - \{E(X)\}^2 = \sum_{1 \leq j < k \leq n} (a_j - a_k)^2 p_j p_k \quad \text{終}$$

§4. $E(X^2) \neq \{E(X)\}^2$ である必要十分条件

【命題4】から, 次の【命題5】が導かれる。

【命題5】 確率変数 X の確率分布が

X	a_1	a_2	a_3	$\cdots$	a_n	計
P	p_1	p_2	p_3	$\cdots$	p_n	1

で与えられるとき

$$E(X^2) \neq \{E(X)\}^2$$

$\iff p_1, p_2, p_3, \dots, p_n$ の中に, 0 でないものが 2 個以上ある

$$\iff 1 \in \{p_1, p_2, p_3, \dots, p_n\}$$

証明

『 $j \neq k \implies a_j \neq a_k$ 』, $p_k \geq 0$, $\sum_{k=1}^n p_k = 1$ であるから

$$E(X^2) \neq \{E(X)\}^2$$

$$\iff E(X^2) - \{E(X)\}^2 \neq 0$$

$$\iff E(X^2) - \{E(X)\}^2 = \sum_{1 \leq j < k \leq n} (a_j - a_k)^2 p_j p_k > 0$$

$\iff j \neq k$ で, $p_j p_k > 0$ となるものが, 少なくとも 1 つ存在する

$\iff p_1, p_2, p_3, \dots, p_n$ の中に, 0 でないものが 2 個以上ある

$$\iff 1 \in \{p_1, p_2, p_3, \dots, p_n\} \quad \text{終}$$

【命題5】から, 次の【命題6】が導かれる。

【命題6】 確率変数 X の確率分布が

X	a_1	a_2	a_3	$\cdots$	a_n	計
P	p_1	p_2	p_3	$\cdots$	p_n	1

で与えられるとき

$$E(X^2) = \{E(X)\}^2$$

$$\iff 1 \in \{p_1, p_2, p_3, \dots, p_n\}$$

§5. 結び

本稿の計算は, §1 で書いたように, $n=2, 3, 4$ の場合で調べることから始めたが, その動機は別の公式での成功体験である。それは次のものである。

母平均 m , 母標準偏差 σ の母集団から大きさ n の無作為標本を復元抽出する場合,

$$E(\bar{X}) = m, \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

によく分からなくて, 『大きさが $N=3$ の母集団 $\{a, b, c\}$ から, $n=2$ の無作為標本を抽出する』場合に, 『復元抽出の場合』と『非復元抽出の場合』で, 具体的にすべての場合を書き出して計算してみた。それが次のものである。

まず, 母平均 m と母標準偏差 σ は

$$m = a \cdot \frac{1}{3} + b \cdot \frac{1}{3} + c \cdot \frac{1}{3} = \frac{a+b+c}{3}$$

$$\sigma = \sqrt{E(X^2) - \{E(X)\}^2}$$

$$= \sqrt{a^2 \cdot \frac{1}{3} + b^2 \cdot \frac{1}{3} + c^2 \cdot \frac{1}{3} - \left(\frac{a+b+c}{3}\right)^2}$$

= (途中式略)

$$= \sqrt{\frac{(a-b)^2 + (b-c)^2 + (c-a)^2}{3^2}}$$

となる。

『復元抽出の場合』には, 次のように $3^2=9$ 通りの抽出方法があり

X_1	a	a	a	b	b	b	c	c	c
X_2	a	b	c	a	b	c	a	b	c

標本平均 $\bar{X} = \frac{X_1 + X_2}{2}$ の期待値 $E(\bar{X})$ と標準偏差

$\sigma(\bar{X})$ は

$$E(\bar{X}) = \sum \frac{X_1 + X_2}{2} \cdot \frac{1}{3^2}$$

$$= (\text{途中式略}) = \frac{a+b+c}{3} = m$$

$$\sigma(\bar{X}) = \sqrt{E(\bar{X}^2) - \{E(\bar{X})\}^2}$$

$$= \sqrt{\sum \left(\frac{X_1 + X_2}{2}\right)^2 \cdot \frac{1}{3^2} - \left(\frac{a+b+c}{3}\right)^2}$$

= (途中式略)

$$= \sqrt{\frac{(a-b)^2 + (b-c)^2 + (c-a)^2}{2 \cdot 3^2}}$$

$$= \frac{\sigma}{\sqrt{2}}$$

となり, 公式通りの結果が出た。

『非復元抽出の場合』には、次のように ${}_3P_2=6$ 通りの抽出方法があり

X_1	a	a	b	b	c	c
X_2	b	c	a	c	a	b

$$E(\bar{X}) = \sum \frac{X_1 + X_2}{2} \cdot \frac{1}{{}_3P_2}$$

$$= (\text{途中式略}) = \frac{a+b+c}{3} = m$$

$$\sigma(\bar{X}) = \sqrt{E((\bar{X})^2) - \{E(\bar{X})\}^2}$$

$$= \sqrt{\sum \left(\frac{X_1 + X_2}{2} \right)^2 \cdot \frac{1}{{}_3P_2} - \left(\frac{a+b+c}{3} \right)^2}$$

$$= (\text{途中式略})$$

$$= \sqrt{\frac{(a-b)^2 + (b-c)^2 + (c-a)^2}{2^2 \cdot 3^2}}$$

$$= \sqrt{\frac{1}{2 \cdot 2}} \sigma$$

となり、公式通りの結果が出た。非復元抽出の場合、

$$E(\bar{X}) = m, \quad \sigma(\bar{X}) = \sqrt{\frac{N-n}{n(N-1)}} \cdot \sigma \quad \text{であることは、}$$

参考文献〔2〕p.133 (詳細は p.151～p.153) に載っている。これは、上記の計算をした後で見つけた。

$$\text{上記の母分散 } \sigma^2 = \frac{(a-b)^2 + (b-c)^2 + (c-a)^2}{3^2}$$

は、【命題2】で $p=q=r=\frac{1}{3}$ の場合の式である。

この σ^2 を求めるときの計算体験が、§2の命題を計算するときに役立った。

数学には、問題を見つける面白さがある。その意味で、本稿 §1 の『注意』を教科書に書いてくださった執筆者に感謝する。

《参考文献》

- 〔1〕「NEXT 数学B」数研出版
- 〔2〕「NEXT 数学B 〔数B 715〕 教授資料」数研出版

(愛知県立東海南高等学校)