

§ 5.2 非復元抽出と復元抽出

母集団から標本を抽出するのに次の 2 通りの仕方を考える：

復元抽出：一度選び出した事物を含めて次の事物を選び出す
(一つの事物が何回も選ばれる可能性がある)；

非復元抽出：一度選び出した事物を除いて次の事物を選び出す
(一つの事物が二回以上選ばれることはない)。

例えば、袋に入った 100 個の物の中から 10 回物を取り出すとする。1 回取り出した物は除いて(袋の中に戻さずに)残りの物の中から以後の物を取り出すと非復元抽出になる。また、1 回取り出した物も袋の中に戻していつも最初の 100 個の物の中から取り出すと復元抽出になる。

大きさ N の母集団から大きさ n の標本を抽出するとき、抽出される可能性がある標本の総数は、復元抽出とすると

$$\underbrace{N \times N \times N \times \cdots \times N}_{n \text{ 個の積}} = N^n$$

であり、非復元抽出とすると

$${}_N P_n = \underbrace{N(N-1)(N-2)(N-3)\cdots(N-n+1)}_{n \text{ 個の積}}$$

である。

例 大きさ 3 の母集団に属す 3 個の事物の各々の属性値が 0, 2, 10 であるとする。母平均 μ は、

$$\mu = \frac{0+2+10}{3} = 4 \text{ .}$$

母標準偏差 σ は、

$$\sigma^2 = \frac{(0-4)^2 + (2-4)^2 + (10-4)^2}{3} = \frac{16+4+36}{3} = \frac{56}{3} \text{ .}$$

この母集団から無作為に抽出された大きさ 2 の標本の標本確率変数 X_1, X_2 の平均 $\overline{X} = \frac{X_1 + X_2}{2}$ について考える。標本の抽出の仕方として非復元抽出と復元抽出とがある。

復元抽出のときを考える。抽出される可能性がある標本として次の 9 通りがある；無作為抽出なので各々の標本について抽出される確率は総て $\frac{1}{9}$ である。

X_1 の実現値	X_2 の実現値	標本平均 $\overline{X}$ の実現値	確率
0	0	$(0+0)/2 = 0$	1/9
0	2	$(0+2)/2 = 1$	1/9
0	10	$(0+10)/2 = 5$	1/9
2	0	$(2+0)/2 = 1$	1/9
2	2	$(2+2)/2 = 2$	1/9
2	10	$(2+10)/2 = 6$	1/9
10	0	$(10+0)/2 = 5$	1/9
10	2	$(10+2)/2 = 6$	1/9
10	10	$(10+10)/2 = 10$	1/9

この表より、

$$P(X_1 \leq 0, X_2 \leq 0) = P(X_1 = 0, X_2 = 0) = \frac{1}{9} \text{ ,}$$

$$P(X_1 \leq 0) P(X_2 \leq 0) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \text{ ,}$$

よって $P(X_1 \leq 0, X_2 \leq 0) = P(X_1 \leq 0) P(X_2 \leq 0)$ 。また、

$$P(X_1 \leq 0, X_2 \leq 2) = P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 2) = \frac{2}{9} \text{ ,}$$

$$P(X_1 \leq 0) P(X_2 \leq 2) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \text{ ,}$$

よって $P(X_1 \leq 0, X_2 \leq 2) = P(X_1 \leq 0) P(X_2 \leq 2)$ 。また、

$$\begin{aligned} &P(X_1 \leq 0, X_2 \leq 10) \\ &= P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 2) + P(X_1 = 0, X_2 = 10) \\ &= \frac{3}{9} = \frac{1}{3} \text{ ,} \end{aligned}$$

$$P(X_1 \leq 0) P(X_2 \leq 10) = \frac{1}{3} \cdot 1 = \frac{1}{3} \text{ ,}$$

よって $P(X_1 \leq 0, X_2 \leq 10) = P(X_1 \leq 0) P(X_2 \leq 10)$ 。このように計算すると、 X_1 と X_2 とは互いに独立であることが分かる。

また、標本確率変数の平均 $\overline{X}$ は次のような離散型確率変数である： $\overline{X}$ が取り得る値は 0, 1, 2, 5, 6, 10 であり、

$$\begin{aligned} P(\overline{X} = 0) &= \frac{1}{9} \text{ ,} & P(\overline{X} = 1) &= \frac{2}{9} \text{ ,} & P(\overline{X} = 2) &= \frac{1}{9} \text{ ,} \\ P(\overline{X} = 5) &= \frac{2}{9} \text{ ,} & P(\overline{X} = 6) &= \frac{2}{9} \text{ ,} & P(\overline{X} = 10) &= \frac{1}{9} \text{ .} \end{aligned}$$

これより、標本確率変数の平均 $\overline{X}$ の平均値 $E[\overline{X}]$ は

$$\begin{aligned} E[\overline{X}] &= 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 2 \cdot \frac{1}{9} + 5 \cdot \frac{2}{9} + 6 \cdot \frac{2}{9} + 10 \cdot \frac{1}{9} = \frac{2+2+10+12+10}{9} \\ &= 4 \text{ .} \end{aligned}$$

標本確率変数の平均 $\overline{X}$ の分散 $V[\overline{X}]$ は

$$\begin{aligned} V[\overline{X}] &= (0-4)^2 \frac{1}{9} + (1-4)^2 \frac{2}{9} + (2-4)^2 \frac{1}{9} + (5-4)^2 \frac{2}{9} + (6-4)^2 \frac{2}{9} + (10-4)^2 \frac{1}{9} \\ &= \frac{16+18+4+2+8+36}{9} = \frac{28}{3} \text{ .} \end{aligned}$$

次に非復元抽出のときを考える。このとき抽出される可能性がある標本は次の 6 通りがある；無作為抽出なので各々の標本が抽出される確率は総て $\frac{1}{6}$ である。

X_1 の実現値	X_2 の実現値	標本平均 $\overline{X}$ の実現値	確率
0	2	$(0+2)/2 = 1$	1/6
0	10	$(0+10)/2 = 5$	1/6
2	0	$(2+0)/2 = 1$	1/6
2	10	$(2+10)/2 = 6$	1/6
10	0	$(10+0)/2 = 5$	1/6
10	2	$(10+2)/2 = 6$	1/6

この表より、

$$P(X_1 \leq 0, X_2 \leq 0) = 0 \text{ ,} \quad P(X_1 \leq 0) = \frac{1}{3} \text{ ,} \quad P(X_2 \leq 0) = \frac{1}{3} \text{ ,}$$

よって $P(X_1 \leq 0, X_2 \leq 0) \neq P(X_1 \leq 0) P(X_2 \leq 0)$ なので、 X_1 と X_2 とは互いに独立ではない。また、標本確率変数の平均 $\overline{X}$ は次のような離散型確率変数である： $\overline{X}$ が取り得る値は 1, 5, 6 であり、

$$P(\overline{X} = 1) = \frac{1}{3} \text{ ,} \quad P(\overline{X} = 5) = \frac{1}{3} \text{ ,} \quad P(\overline{X} = 6) = \frac{1}{3} \text{ .}$$

これより、標本確率変数の平均 $\overline{X}$ の平均値 $E[\overline{X}]$ は

$$E[\overline{X}] = 1 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} + 6 \cdot \frac{1}{3} = \frac{1+5+6}{3} = 4 \text{ .}$$

標本平均を表す確率変数 $\overline{X}$ の分散 $V[\overline{X}]$ は

$$V[\overline{X}] = (1-4)^2 \cdot \frac{1}{3} + (5-4)^2 \cdot \frac{1}{3} + (6-4)^2 \cdot \frac{1}{3} = \frac{9+1+4}{3} = \frac{14}{3} \text{ .}$$

大きさ 3 の母集団に属す 3 つの事物の各々の属性値が 0, 2, 10 であるとき、この母集団から無作為抽出した大きさ 2 の標本の標本確率変数 X_1, X_2 の平均 $\overline{X}$ について次のようになる：抽出が復元抽出のとき、 X_1 と X_2 とは互いに独立であり、

$$E[\overline{X}] = 4 \text{ ,} \quad V[\overline{X}] = \frac{28}{3} \text{ ;}$$

抽出が非復元抽出のとき、 X_1 と X_2 とは互いに独立でなく、

$$E[\overline{X}] = 4 \text{ ,} \quad V[\overline{X}] = \frac{14}{3} \text{ .} \quad \text{[終]}$$

一般的に、母集団から無作為抽出した標本について、復元抽出のときは標本確率変数は互いに独立であり、非復元抽出のときは標本確率変数は互いに独立でない。また、母集団から無作為抽出した標本の標本平均を表す確率変数について、非復元抽出のときの分散は復元抽出のときの分散より小さい。

次の定理が成り立つ（証明は省く）。

[定理 5.2.1] 母集団から無作為に復元抽出した標本の標本確率変数は互いに独立である。

[定理 5.2.2] 母平均が μ であり母標準偏差が σ である大きさ N の母集団から無作為抽出された大きさが n である標本の標本平均を表す確率変数 $\overline{X}$ の平均値 $E[\overline{X}]$ 及び分散 $V[\overline{X}]$ は次のようになる：復元抽出のとき、

$$E[\overline{X}] = \mu \text{ ,} \quad V[\overline{X}] = \frac{\sigma^2}{n} \text{ ;}$$

非復元抽出のとき、

$$E[\overline{X}] = \mu \text{ ,} \quad V[\overline{X}] = \frac{N-n}{N-1} \frac{\sigma^2}{n} \text{ .}$$

例 母平均が 130 であり母標準偏差が 20 である大きさ 1000 の母集団から無作為抽出された大きさ 200 の標本の標本平均を表す確率変数 $\overline{X}$ の平均値 $E[\overline{X}]$ と分散 $V[\overline{X}]$ と標準偏差 $\sigma[\overline{X}]$ とを求める。

復元抽出のとき、

$$\begin{aligned} E[\overline{X}] &= 130 \text{ ,} \\ V[\overline{X}] &= \frac{20^2}{200} = 2 \text{ ,} \\ \sigma[\overline{X}] &= \sqrt{V[\overline{X}]} = \sqrt{2} \approx 1.414 \text{ .} \end{aligned}$$

非復元抽出のとき、

$$\begin{aligned} E[\overline{X}] &= 130 \text{ ,} \\ V[\overline{X}] &= \frac{800}{999} \cdot \frac{20^2}{200} = \frac{1600}{999} \approx 1.602 \text{ ,} \\ \sigma[\overline{X}] &= \sqrt{V[\overline{X}]} = \sqrt{\frac{1600}{999}} \approx 1.266 \text{ .} \quad \text{[終]} \end{aligned}$$

問 5.2 母平均が 70 であり母分散が 100 である大きさ 2000 の母集団から無作為抽出した大きさ 400 の標本の標本平均を表す確率変数 $\overline{X}$ について、復元抽出のとき及び非復元抽出のときの、平均値 $E[\overline{X}]$ と分散 $V[\overline{X}]$ と標準偏差 $\sigma[\overline{X}]$ とを求めよ。

4.3 節で次の定理を述べた。

[中心極限定理] 自然数 n は充分大きいとする。ある一つの分布に従う n 個の確率変数 $X_1, X_2, X_3, \dots, X_n$ が互いに独立でありそれらの平均値 μ と標準偏差 σ とがあるならば、 $X_1, X_2, X_3, \dots, X_n$ の平均を表す確率変数 $\frac{1}{n} \sum_{k=1}^n X_k$ は近似的に正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う。

中心極限定理を標本調査に応用する。母平均が μ であり母標準偏差が σ である母集団から無作為に復元抽出された大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \dots, X_n$ は、互いに独立であり、総て母集団分布に従い、平均値は母平均 μ であり標準偏差は母標準偏差 σ である。故に、中心極限定理より、標本の大きさ n が充分大きいとき、 $X_1, X_2, X_3, \dots, X_n$ の平均を表す確率変数 $\frac{1}{n} \sum_{k=1}^n X_k$ は近似的に正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う。

[定理 5.2.3] 母平均が μ であり母標準偏差が σ である母集団から無作為に復元抽出された標本の標本平均を表す確率変数は、標本の大きさ n が充分大きいとき、近似的に正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う。