§5.2　非復元抽出と復元抽出

母集団から標本を抽出する仕方に次の2通りがあります：

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回物を取り出すとします． 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ります． 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ります．

大きさ N の母集団から大きさ n の標本を抽出するとき，標本の選び方の総数は，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，非復元抽出とすると

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

です．

例として簡単な場合を考えます． 大きさ3の母集団に属す3個の事物の各々の属性値が $0,2,10$ であるとします． 母平均 μ は，

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

母標準偏差 σ は，

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

この母集団から無作為抽出された大きさ2の標本の標本確率変数 X_1,X_2 の平均

$$\overline{X}=\frac{X_1+X_2}{2}$$

について考えます． 標本の抽出の仕方として非復元抽出と復元抽出とがあります．

まず復元抽出のときを考えます． このとき抽出される標本の可能性として次の9通りがあります；無作為抽出ですから各々の標本が抽出される確率は総て $\frac{1}{9}$ です．

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，

$$\mathrm{P}(\,X_1\le 0\,,X_2\le 0\,)=\mathrm{P}(\,X_1=0\,,X_2=0\,)=\frac{1}{9}\ ,$$

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

よって $\mathrm{P}(\,X_1\le 0\,,X_2\le 0\,)=\mathrm{P}(\,X_1\le 0\,)\,\mathrm{P}(\,X_2\le 0\,)$ ． また，

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

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

よって $\mathrm{P}(\,X_1\le 0\,,X_2\le 2\,)=\mathrm{P}(\,X_1\le 0\,)\,\mathrm{P}(\,X_2\le 2\,)$ ． また，

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

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

よって $\mathrm{P}(\,X_1\le 0\,,X_2\le 10\,)=\mathrm{P}(\,X_1\le 0\,)\,\mathrm{P}(\,X_2\le 10\,)$ ． 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わかります． 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す：$\overline{X}$ が取り得る値は $0,1,2,5,6,10$ であり，

$$\mathrm{P}\big(\,\overline{X}=0\,\big)=\frac{1}{9}\ ,\qquad \mathrm{P}\big(\,\overline{X}=1\,\big)=\frac{2}{9}\ ,\qquad \mathrm{P}\big(\,\overline{X}=2\,\big)=\frac{1}{9}\ ,$$

$$\mathrm{P}\big(\,\overline{X}=5\,\big)=\frac{2}{9}\ ,\qquad \mathrm{P}\big(\,\overline{X}=6\,\big)=\frac{2}{9}\ ,\qquad \mathrm{P}\big(\,\overline{X}=10\,\big)=\frac{1}{9}\ .$$

従って，標本確率変数の平均 $\overline{X}$ の平均値 $\mathrm{E}\big[\overline{X}\big]$ は

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

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

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

次に非復元抽出のときを考えます． このとき抽出される標本の可能性として次の6通りがあります；無作為抽出ですから各々の標本が抽出される確率は総て $\frac{1}{6}$ です．

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，

$$\mathrm{P}(\,X_1\le 0\,,X_2\le 0\,)=0\ ,\qquad \mathrm{P}(\,X_1\le 0\,)=\frac{1}{3}\ ,\qquad \mathrm{P}(\,X_2\le 0\,)=\frac{1}{3}\ ,$$

よって $\mathrm{P}(\,X_1\le 0\,,X_2\le 0\,)\ne\mathrm{P}(\,X_1\le 0\,)\,\mathrm{P}(\,X_2\le 0\,)$ なので，X_1 と X_2 とは互いに独立ではありません． 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す：$\overline{X}$ が取り得る値は $1,5,6$ であり，

$$\mathrm{P}\big(\,\overline{X}=1\,\big)=\frac{1}{3}\ ,\qquad \mathrm{P}\big(\,\overline{X}=5\,\big)=\frac{1}{3}\ ,\qquad \mathrm{P}\big(\,\overline{X}=6\,\big)=\frac{1}{3}\ .$$

従って，標本確率変数の平均 $\overline{X}$ の平均値 $\mathrm{E}\big[\overline{X}\big]$ は

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

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

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

こうして次のことが分かります：大きさ3の母集団に属す3個の事物の各々の属性値が $0,2,10$ であるとき，この母集団から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，

$$\mathrm{E}\big[\overline{X}\big]=4\ ,\qquad \mathrm{V}\big[\overline{X}\big]=\frac{28}{3}\ ;$$

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，

$$\mathrm{E}\big[\overline{X}\big]=4\ ,\qquad \mathrm{V}\big[\overline{X}\big]=\frac{14}{3}\ .$$

一般的に次の定理が成り立ちます（証明は略します）．

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 n である標本の標本平均を表す確率変数 $\overline{X}$ の平均値 $\mathrm{E}\big[\overline{X}\big]$ 及び分散 $\mathrm{V}\big[\overline{X}\big]$ は次のようになる：復元抽出のとき，

$$\mathrm{E}\big[\overline{X}\big]=\mu\ ,\qquad \mathrm{V}\big[\overline{X}\big]=\frac{\sigma^2}{n}\ ;$$

非復元抽出のとき，

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

例題 5.2　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 $\overline{X}$ 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 $\mathrm{E}\big[\overline{X}\big]$ と分散 $\mathrm{V}\big[\overline{X}\big]$ と標準偏差 $\sigma\big[\overline{X}\big]$ とを求める．

復元抽出のとき，

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

非復元抽出のとき，

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

終

問題 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 $\mathrm{E}\big[\overline{X}\big]$ と分散 $\mathrm{V}\big[\overline{X}\big]$ と標準偏差 $\sigma\big[\overline{X}\big]$ とを求めなさい．

4.3節で述べた中心極限定理を思い起こして下さい．

> 自然数 n は充分大きいとする． 互いに独立な n 個の確率変数 $X_1,X_2,X_3,\cdots,X_n$ が総て同じ確率分布に従い，その確率分布の平均値 μ と標準偏差 σ とがあるとき，$X_1,X_2,X_3,\cdots,X_n$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_n}{n}$ は近似的に正規分布 $\mathrm{N}\Big(\mu,\dfrac{\sigma^2}{n}\Big)$ に従う．

中心極限定理を標本調査に応用します． 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 $X_1,X_2,X_3,\cdots,X_n$ は，互いに独立で，総て母集団分布に従い，平均値は母平均 μ で標準偏差は母標準偏差 σ です． 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，標本確率変数 $X_1,X_2,X_3,\cdots,X_n$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_n}{n}$ は近似的に正規分布 $\mathrm{N}\Big(\mu,\dfrac{\sigma^2}{n}\Big)$ に従います．

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 $\mathrm{N}\Big(\mu,\dfrac{\sigma^2}{n}\Big)$ に従う．