§6.3　母分散の点推定

母分散の不偏推定量を考えます．

簡単な例で考えます．大きさ 3 の母集団の事物の各々の属性値が $0, 2, 10$ であるとします．母平均は

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

で，母分散は

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

です．この母集団から無作為に非復元抽出された大きさ 2 の標本の標本確率変数を X_1, X_2 とおきます．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	標本分散	確率
0	2	$(0+2)/2=1$	$\{(0-1)^2+(2-1)^2\}/2=1$	1/6
0	10	$(0+10)/2=5$	$\{(0-5)^2+(10-5)^2\}/2=25$	1/6
2	0	$(2+0)/2=1$	$\{(2-1)^2+(0-1)^2\}/2=1$	1/6
2	10	$(2+10)/2=6$	$\{(2-6)^2+(10-6)^2\}/2=16$	1/6
10	0	$(10+0)/2=5$	$\{(10-5)^2+(0-5)^2\}/2=25$	1/6
10	2	$(10+2)/2=6$	$\{(10-6)^2+(2-6)^2\}/2=16$	1/6

この表より，標本確率変数の分散の平均値は

$$1\cdot\frac{1}{3}+16\cdot\frac{1}{3}+25\cdot\frac{1}{3} = \frac{42}{3} = 14\ .$$

このように，母分散は $\frac{56}{3}$ で，標本確率変数の分散の平均値は 14 です．標本確率変数の分散の平均値は母分散とは通常は異なるので，

標本確率変数の分散は母分散の不偏推定量ではない

ことになります．その理由は後で説明しますが，それではどのような統計量が母分散の不偏推定量になるかということが問題になります．この問題に答えるのが以下の 2 つの定理です．その証明は後にします．

定理 6.3.1　母集団から無作為に非復元抽出した大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \ldots, X_n$ の平均を $\overline{X}$ とおく．母集団の大きさ N が分かるならば，確率変数

$$\frac{N-1}{N(n-1)}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

は母分散の不偏推定量である．

この定理において有限母集団の大きさ N を $N\to\infty$ とすると，無限母集団に関する結果が導けます．

定理 6.3.2　無限母集団から無作為に抽出した大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \ldots, X_n$ の平均を $\overline{X}$ とおく．統計量

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

は母分散の不偏推定量である．

上の定理で述べたように，無限母集団から無作為に抽出した大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \ldots, X_n$ の平均を $\overline{X}$ とおくと，統計量 $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ は母分散の不偏推定量です；この統計量を**不偏分散**といいます．

$\boxed{\text{例題 6.3}}$　大きさ 40 の有限母集団から無作為に非復元抽出した大きさ 4 の標本の属性値が次のようになったとする：

73,　59,　83　65.

標本分散と，母分散の不偏推定値とを求めよ．

標本平均の実現値は

$$\frac{73+59+83+65}{4} = 70\ .$$

標本分散は

$$\frac{(73-70)^2+(59-70)^2+(83-70)^2+(65-70)^2}{4} = 81\ .$$

母分散の不偏推定値は

$$\frac{39}{40}\frac{(73-70)^2+(59-70)^2+(83-70)^2+(65-70)^2}{3} = 105.3\ .$$

$\boxed{\text{終}}$

$\boxed{\text{問題 6.3}}$　大きさ 50 の有限母集団から無作為に非復元抽出した大きさ 4 の標本の属性値が次のようになったとします：

60,　47,　38　55.

標本分散と，母分散の不偏推定量による推定値とを求めなさい．

標本分散の統計量 $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ の平均値は母分散より小さくなります．その理由を簡単な例で説明します．例えば，大きさ 2 の標本の実現値が 3 と 5 とであったとします．このとき，標本平均は

$$\frac{3+5}{2} = 4\ ,$$

標本分散は

$$\frac{1}{2}\{(3-4)^2+(5-4)^2\} = 1\ .$$

ところが，母平均は，丁度 4 であることはまずなく，むしろ 4 からいくらかずれているのが普通です．仮に母平均が 4.2 であったとすると，母分散の推定値は

$$\frac{1}{2}\{(3-4.2)^2+(5-4.2)^2\} = 1.04$$

とすべきでしょう．また，仮に母平均が 3.9 であったとすると，母分散の推定値は

$$\frac{1}{2}\{(3-3.9)^2+(5-3.9)^2\} = 1.01$$

とすべきでしょう．このように，母平均が標本平均より大きくても小さくても，母分散の推定値は標本分散よりも大きくなります．ですから，母分散は標本分散の平均値より大きくなります．

なお，標本確率変数の不偏分散は母分散の一致推定量です（証明は省略します）．

定理　無限母集団から無作為に抽出した大きさ n の標本の標本確率変数を $X_1, X_2, X_3, \ldots, X_n$ とおくと，不偏分散を表す統計量

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

は母分散の一致推定量である．

――――　定理 6.3.1 の証明

定理 6.3.1 を証明します．大きさ N の有限母集団から無作為に非復元抽出した大きさ n の標本の標本確率変数を $X_1, X_2, X_3, \ldots, X_n$ とおき，標本平均を表す統計量を $\overline{X}$ とおきます．母平均を μ とおき，母分散を σ^2 とおきます．$i=1,2,3,\ldots,n$ について

$$\begin{aligned}\left(X_i-\overline{X}\right)^2 &= \{(X_i-\mu)-(\overline{X}-\mu)\}^2\\ &= (X_i-\mu)^2-2(\overline{X}-\mu)(X_i-\mu)+(\overline{X}-\mu)^2\ ;\end{aligned}$$

よって，

$$\begin{aligned}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 &= \sum_{i=1}^{n}\left\{(X_i-\mu)^2-2(\overline{X}-\mu)(X_i-\mu)+(\overline{X}-\mu)^2\right\}\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)\left(\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\mu\right)+\sum_{i=1}^{n}(\overline{X}-\mu)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)(n\overline{X}-n\mu)+n(\overline{X}-\mu)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-2n(\overline{X}-\mu)^2+n(\overline{X}-\mu)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\ .\end{aligned}$$

定理 5.2 より　$\mathrm{V}[\overline{X}] = \frac{N-n}{N-1}\frac{\sigma^2}{n}$　ですから，

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] &= \mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]\\ &= \sum_{i=1}^{n}\mathrm{E}[(X_i-\mu)^2]-n\mathrm{E}[(\overline{X}-\mu)^2] = \sum_{i=1}^{n}\mathrm{V}[X_i]-n\mathrm{V}[\overline{X}]\\ &= \sum_{i=1}^{n}\sigma^2-n\frac{N-n}{N-1}\frac{\sigma^2}{n} = n\sigma^2-\frac{N-n}{N-1}\sigma^2\\ &= \frac{N(n-1)}{N-1}\sigma^2\ .\end{aligned}$$

故に

$$\begin{aligned}\mathrm{E}\left[\frac{N-1}{N(n-1)}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] &= \frac{N-1}{N(n-1)}\mathrm{E}\left[\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] = \frac{N-1}{N(n-1)}\frac{N(n-1)}{N-1}\sigma^2\\ &= \sigma^2\ .\end{aligned}$$

つまり統計量 $\frac{N-1}{N(n-1)}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ は母分散 σ^2 の不偏推定量です．