2024/5/2 第 1 版
2025/5/17 第 2 版

標準誤差－非復元抽出の場合

野津

1. はじめに

要素数N，標準偏差σの母集団からn個の標本を抽出する際の標本平均がどの程度ばらつくかを表す指標として標準誤差（厳密には標本平均の標準誤差）がある．復元抽出の場合，標準誤差を表す式は

$$SE = \frac{\sigma}{\sqrt{n}} \tag{1}$$

であり，その証明は難しくないが，非復元抽出の場合，標準誤差を表す式は

$$SE = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}} \tag{2}$$

であり，その証明は少し難しい．本稿では数学的帰納法による式(2)の証明を示す．

2. 標本平均の期待値

ここでは，要素数N，平均μ，標準偏差σの母集団は確定的に与えられているものとし，そこからn個の標本を抽出する際の標本平均$\overline{x(n)}$は確率変数であると考える．

標本平均のばらつきについて検討する前に，先ず，標本平均の期待値について検討する．標本平均の期待値は直感的にはμとなりそうである．これは次のように数学的帰納法で示すことができる．

まず，$n = 1$のとき，標本平均は選択した要素の値に等しいので$E\left(\overline{x(1)}\right) = \mu$である．次に，

$$E\left(\overline{x(k)}\right) = \mu \tag{3}$$

を仮定し

$$E\left(\overline{x(k+1)}\right) = \mu \tag{4}$$

を導く．いま，N個の要素から既にk個の要素が選ばれている状況で，新たに 1 要素を選択することを考える（**図-1**）．既に選ばれている要素を$x_i\ (i = 1,\cdots,k)$とし，追加で選択する 1 要素を$y_j\ (j = 1,\cdots,N-k)$とする．この場合

$$\overline{x(k+1)} = \frac{y_j + k\overline{x(k)}}{k+1} \tag{5}$$

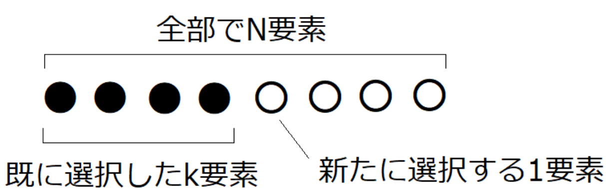

図-1　既にk個の要素が選ばれている状況で新たに 1 要素を選択

である．まず，x_i $(i=1,\cdots,k)$が選ばれているという条件の下での$\overline{x(k+1)}$の条件付き期待値を検討すると

$$E'\left(\overline{x(k+1)}\right)=\frac{1}{N-k}\sum_{j=1}^{N-k}\frac{y_j+k\overline{x(k)}}{k+1}=\frac{1}{(N-k)(k+1)}\left(\sum_{j=1}^{N-k}y_j+(N-k)k\overline{x(k)}\right) \tag{6}$$

となる．ここで右辺括弧内第 1 項は全要素の和から既に選択されているk個の要素の和を引けば良いので

$$E'\left(\overline{x(k+1)}\right)=\frac{1}{(N-k)(k+1)}\left(N\mu-k\overline{x(k)}+(N-k)k\overline{x(k)}\right)=\frac{1}{(N-k)(k+1)}\left(N\mu+\left(N-(k+1)\right)k\overline{x(k)}\right) \tag{7}$$

となる．ここでx_iについてもすべての組み合わせを考えた場合の期待値をとると

$$E\left(\overline{x(k+1)}\right)=\frac{1}{(N-k)(k+1)}\left(N\mu+\left(N-(k+1)\right)kE\left(\overline{x(k)}\right)\right) \tag{8}$$

であり，式(3)を用いると

$$E\left(\overline{x(k+1)}\right)=\frac{1}{(N-k)(k+1)}\left(N\mu+\left(N-(k+1)\right)k\mu\right)=\mu \tag{9}$$

であり，式(4)が正しいことが示された．以上により，数学的帰納法により標本平均の期待値はμであることが示された．

3.　標本平均のばらつき

次に標本平均の分散が

$$E\left(\left(\overline{x(n)}-\mu\right)^2\right)=\frac{N-n}{N-1}\frac{\sigma^2}{n} \tag{10}$$

で与えられることを数学的帰納法で示す．

まず，$n=1$のとき，標本平均は選択した要素の値に等しいので，式(10)の左辺は$E\left(\left(\overline{x(1)}-\mu\right)^2\right)=\sigma^2$であり，式(10)の右辺に等しい．次に，

$$E\left(\left(\overline{x(k)}-\mu\right)^2\right)=\frac{N-k}{N-1}\frac{\sigma^2}{k} \tag{11}$$

を仮定し

$$E\left(\left(\overline{x(k+1)}-\mu\right)^2\right)=\frac{N-(k+1)}{N-1}\frac{\sigma^2}{k+1} \tag{12}$$

を導く．いま，N個の要素から既にk個の要素が選ばれている状況で，新たに 1 要素を選択することを考える（図-1）．既に選ばれている要素を$x_i\ (i = 1, \cdots, k)$とし，追加で選択する 1 要素を$y_j\ (j = 1, \cdots, N - k)$とする．この場合，式(5)より

$$\left(\overline{x(k+1)} - \mu\right)^2 = \left(\frac{y_j + k\overline{x(k)}}{k+1} - \mu\right)^2 \tag{13}$$

である．まず，$x_i\ (i = 1, \cdots, k)$が選ばれているという条件の下での$\left(\overline{x(k+1)} - \mu\right)^2$の条件付き期待値を検討すると

$$\begin{aligned} E'\left(\left(\overline{x(k+1)} - \mu\right)^2\right) &= \frac{1}{N-k}\sum_{j=1}^{N-k}\left(\frac{y_j + k\overline{x(k)}}{k+1} - \mu\right)^2 = \frac{1}{(N-k)(k+1)^2}\sum_{j=1}^{N-k}\left((y_j - \mu) + k\left(\overline{x(k)} - \mu\right)\right)^2 \\ &= \frac{1}{(N-k)(k+1)^2}\left(\sum_{j=1}^{N-k}(y_j - \mu)^2 + 2k\left(\overline{x(k)} - \mu\right)\sum_{j=1}^{N-k}(y_j - \mu) + (N-k)k^2\left(\overline{x(k)} - \mu\right)^2\right) \end{aligned} \tag{14}$$

となる．ここで右辺括弧内の和（2 箇所）は全要素の和から既に選択されているk個の要素の和を引けば良いので

$$\begin{aligned} E'\left(\left(\overline{x(k+1)} - \mu\right)^2\right) &= \frac{1}{(N-k)(k+1)^2}\left(N\sigma^2 - \sum_{i=1}^{k}(x_i - \mu)^2 - 2k\left(\overline{x(k)} - \mu\right)\sum_{i=1}^{k}(x_i - \mu) + (N-k)k^2\left(\overline{x(k)} - \mu\right)^2\right) \\ &= \frac{1}{(N-k)(k+1)^2}\left(N\sigma^2 - \sum_{i=1}^{k}(x_i - \mu)^2 - 2k^2\left(\overline{x(k)} - \mu\right)^2 + (N-k)k^2\left(\overline{x(k)} - \mu\right)^2\right) \\ &= \frac{1}{(N-k)(k+1)^2}\left(N\sigma^2 - \sum_{i=1}^{k}(x_i - \mu)^2 + (N-k-2)k^2\left(\overline{x(k)} - \mu\right)^2\right) \end{aligned} \tag{15}$$

となる．ここでx_iについてもすべての組み合わせを考えた場合の期待値をとると

$$E\left(\left(\overline{x(k+1)} - \mu\right)^2\right) = \frac{1}{(N-k)(k+1)^2}\left(N\sigma^2 - k\sigma^2 + (N-k-2)k^2 E\left(\left(\overline{x(k)} - \mu\right)^2\right)\right) \tag{16}$$

であり，式(11)を用いると

$$E\left(\left(\overline{x(k+1)} - \mu\right)^2\right) = \frac{1}{(N-k)(k+1)^2}\left((N-k)\sigma^2 + (N-k-2)k^2\frac{N-k}{N-1}\frac{\sigma^2}{k}\right) = \frac{N-(k+1)}{N-1}\frac{\sigma^2}{k+1} \tag{17}$$

であり，式(12)が正しいことが示された．以上により，数学的帰納法により標本平均の分散は式(10)で与えられることが示された．式(10)の平方根をとったものが式(2)である．

なお，$n = N$のときは標本の抽出方法は一通りしかないため式(10)の左辺はゼロである．よって式(10)は$n = N$の場合にも正しい．