母平均の区間推定と標準化（2）

岩村 英之

標本平均を用いた母平均の推定

前回と同じ，無限に数字のつまった壺を考えよう（分布は正規分布）．
この壺から取り出した n 個の数字の標本平均 $\overline{X}$ は，どの数字を取り出すかによって異なる．すなわち，$\overline{X}$ 自体が確率分布する（＝いろいろな可能性を持っており，やってみるまで何が出るかわからない）．

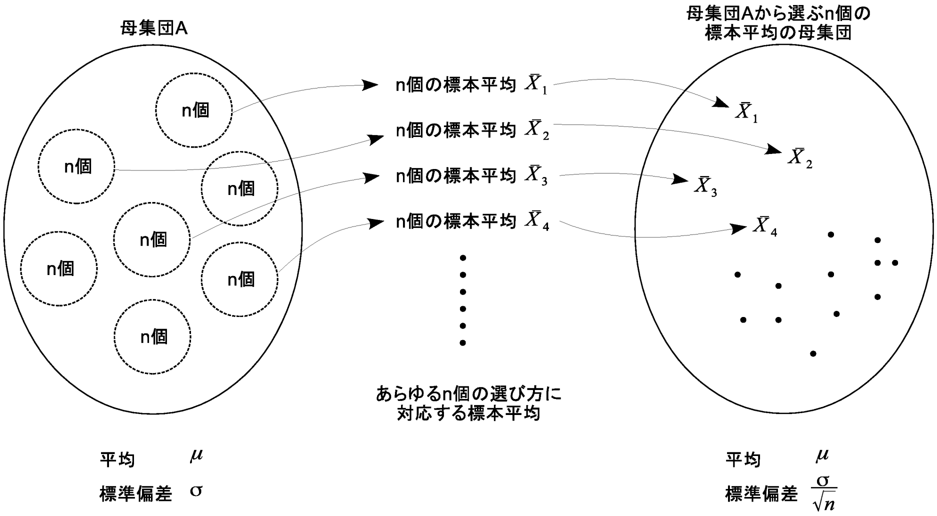

図 1: 「標本平均の母集団」という考え方

ところで，小島（2006）の第 14 講ですでに学んだように，この $\overline{X}$ の母平均はもとの壺の母平均と同じ μ である．したがって，$\overline{X}$ の母平均を推定することで，もとの壺の母平均の推定の代用とすることとができる．
さらに，$\overline{X}$ は正規分布するので，前回とまったく同じ方法でこの標本平均 $\overline{X}$ の母平均を推定することができる．ただし，この $\overline{X}$ の母標準偏差は σ ではなく $\frac{\sigma}{\sqrt{n}}$ であることに気を付けなければならない．

今，実際に n 個の数値を取り出してみたところ，$\overline{X_1}$ という標本平均が得られたとする．前回の手順に従えば，次の式を満たす μ が，$\overline{X_1}$ だけを証拠として浮かび上がってくる母平均の容疑者達ということになる．

$$-1.96 \leq \frac{\overline{X_1} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$$

真ん中の項の分母が σ ではなく $\frac{\sigma}{\sqrt{n}}$ になっていることに注意すること．この連立不等式を μ について解けば，容疑者の範囲がよりはっきりする．

$$\overline{X}_1 - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_1 + 1.96\frac{\sigma}{\sqrt{n}}$$

やはり，1.96 にかかっているのが σ ではなく $\frac{\sigma}{\sqrt{n}}$ になっていることに注意すること．

母集団の標準偏差が未知の場合の母平均の推定

母標準偏差がわかっていない場合に,「観測値を標準化したものが-1.96 から 1.96 の範囲に入るかどうか」というチェック方法が使えなくなる．なぜなら，

$$\frac{観測した標本平均 - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{1}$$

のように標準化したくとも，分母の σ の部分に入れる数値がないのだから．

本物の σ の代わりに，σ の推定値を使ってみてはどうか．σ の推定値としてすぐに思い浮かぶのは，次の標本標準偏差であろう．

$$s = \sqrt{\frac{1}{n}\left[\left(X_1 - \overline{X}_1\right)^2 + \left(X_2 - \overline{X}_1\right)^2 + \cdots + \left(X_n - \overline{X}_1\right)^2\right]}$$

しかし，(1) の σ のところにこの s を入れても，残念ながら標準正規分布に従わないことが知られている．これは，(1) のルールによって変えられた背番号においては，95 パーセントを占めるのが-1.96 から 1.96 ではないことを意味する．-1.96 から 1.96 が 95 パーセントを占めるのは，標準正規分布に限った話である．
次に，s に似ているが少しだけ違う，次の $\hat{\sigma}$ を用いることを考えてみよう．

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\left[\left(X_1 - \overline{X}_1\right)^2 + \left(X_2 - \overline{X}_1\right)^2 + \cdots + \left(X_n - \overline{X}_1\right)^2\right]}$$

n（標本数）ではなく $n-1$ で割っているところが，標本標準偏差との違いである．本物の標準偏差の代わりにこの $\hat{\sigma}$ を使って標準化すると，本物を使ったときと同じチェック方法が使えるだろうか.

$$t = \frac{\overline{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \tag{2}$$

本物の標準偏差の代わりにこの $\hat{\sigma}$ を使って標準化された (2) 式も，やはり標準正規分布に従わないのだが，標準正規分布とは違う別の，比較的扱いやすい分布に従うことが知られている．それはどのような分布だろうか．また，その分布において 95 パーセントを占める数字の範囲はどうなっているのであろうか.

t 分布

(2) 式によって変換された背番号たちは,「t 分布」と呼ばれるパターンをとることが知られている．
t 分布は標準正規分布によく似ていて，母平均は標準正規と同じで 0 である．ただし，標準偏差は標準正規分布より少し大きい．したがって，範囲を少し広くとらないと 95 パーセントが入ってこない．具体的には，95 パーセントが入る範囲は「標本数（厳密には，標本数-1=自由度）」に依存する．たとえば，10 枚のカードをとって得られた標本平均ならば, (2) 式に従って変換された変数は「自由度 9 の t 分布」に従う．そして，自由度 9 の t 分布に従う変数たちの 95 パーセントが入る範囲は，-2.2622 から 2.2622 までとなる．標準正規分布の「-1.96 から 1.96 まで」と比べると，範囲が広くなっていることに注意したい.

t 分布を用いた推定

σ の代わりに $\hat{\sigma}$ を用いる点と，± 1.96 の代わりに適当な t 分布の値を用いる点を除けば，母平均の区間推定はこの資料の前半とまったく同じである．すなわち, (2) 式によって標準化された値が適切な t 分布の区間に入ってくるような μ を見つけてやれば，それが母平均のあり得る区間である．具体的には，そのような μ は次の条件を満たす．

$$-t_{n-1} \leq \frac{\overline{X}_1 - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t_{n-1}$$

ただし，t_{n-1} は自由度 $n-1$ の t 分布において 95 パーセントが入る区間の上限を表している．
この連立不等式を μ について解けば，μ のあり得る区間がはっきりとする．

$$\overline{X}_1 - t_{n-1}\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{X}_1 + t_{n-1}\frac{\hat{\sigma}}{\sqrt{n}}$$