§5.3　無限母集団と有限母集団

標本調査において，母集団の大きさに対する標本の大きさの割合がきわめて小さいとき，その母集団を**無限母集団** (infinite population) といいます．また，母集団の大きさに対する標本の大きさの割合がきわめて小さいわけではないとき，その母集団を**有限母集団** (finite population) といいます．

例えば母集団の大きさが 100 万で標本の大きさが 1000 であるとき，標本を抽出する前と全て抽出した後とでも 1000 分の 1 しか変わりません．このように，無限母集団では，母集団の大きさに対する標本の大きさの割合がきわめて小さいので，抽出する前の状態と抽出した後の状態とはほとんど同じです．それ故，非復元抽出であっても復元抽出とほとんどど同じようになります．

例えば大きさが 100 万である母集団から無作為抽出した大きさが 1000 である標本の標本平均を表す確率変数を $\overline{X}$ とおきます．母標準偏差を σ とおくと，$\overline{X}$ の分散 $\mathrm{V}[\overline{X}]$ は次のようになります：非復元抽出のとき，

$$\mathrm{V}[\overline{X}] = \frac{1000000-1000}{1000000-1}\frac{\sigma^2}{1000} = \frac{\sigma^2}{1001} \ ;$$

復元抽出のとき，

$$\mathrm{V}[\overline{X}] = \frac{\sigma^2}{1000} \ .$$

ここで

$$\frac{1}{1001} \fallingdotseq 0.999000999 \times 10^{-3} \fallingdotseq 10^{-3} = \frac{1}{1000} \ ;$$

非復元抽出のときの $\mathrm{V}[\overline{X}]$ の値と復元抽出のときの $\mathrm{V}[\overline{X}]$ の値との違いは誤差の範囲に属します．

このように，母集団の大きさに対する標本の大きさの割合がきわめて小さいとき，つまり無限母集団のとき，標本平均を表す確率変数 $\overline{X}$ の確率分布は，非復元抽出のときと復元抽出のときとで近似的に等しいといえます．

中心極限定理から導かれた定理 5.2.3 を思い起こして下さい．

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

無限母集団において，標本平均を表す確率変数の確率分布は，非復元抽出のときと復元抽出のときとで近似的に等しいですから，次の定理が成り立ちます．

定理 5.3　母平均が μ であり母標準偏差が σ である無限母集団から無作為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 $\mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従う．

つまり，無限母集団においては，非復元抽出であっても，無作為抽出された標本の大きさが大きいとき標本平均を表す確率変数は近似的に正規分布に従います．

例題 5.3　母平均が 59 であり母標準偏差が 7 である大きさ約 37 万の母集団から無作為に非復元抽出された大きさ 100 の標本の標本平均を表す確率変数 $\overline{X}$ について，確率 $\mathrm{P}(\,\overline{X} \geq 60\,)$ を近似的に求める．

母集団の大きさ約 37 万に対する標本の大きさ 100 の比率は非常に小さいので，母集団を無限母集団とみなす．確率変数 $\overline{X}$ の分散 $\mathrm{V}[\overline{X}]$ は

$$\mathrm{V}[\overline{X}] = \frac{7^2}{100} \ .$$

確率変数 $\overline{X}$ は近似的に正規分布 $\mathrm{N}\left(59, \dfrac{7^2}{100}\right)$ に従うので，確率変数

$$Z = \frac{\overline{X}-59}{\sqrt{\dfrac{7^2}{100}}} = \frac{10}{7}(\overline{X}-59)$$

は近似的に標準正規分布 $\mathrm{N}(0,1)$ に従う．

$$\overline{X} \geq 60 \iff \frac{10}{7}(\overline{X}-59) \geq \frac{10}{7}(60-59) \iff Z \geq \frac{10}{7} \ .$$

標準正規分布の累積分布関数を $\Phi(x)$ とおく．

$$\begin{aligned}\mathrm{P}(\,\overline{X} \geq 60\,) &= \mathrm{P}\left(\,Z \geq \frac{10}{7}\,\right) \fallingdotseq 1-\Phi\left(\frac{10}{7}\right) \fallingdotseq 1-\Phi(1.43) \fallingdotseq 1-0.9236 \\ &= 0.0764 \ .\end{aligned}$$

問題 5.3　母平均が 378 であり母分散が 162 である大きさ約 74 万の母集団から無作為に非復元抽出された大きさ 200 の標本の標本平均を表す確率変数 $\overline{X}$ について，確率 $\mathrm{P}(\,377 \leq \overline{X} \leq 380\,)$ を近似的に求めなさい．

母平均を μ とおき，母標準偏差を σ とおきます．前節の定理 5.2 より次のことが分かります：無限母集団から無作為抽出された大きさ n の標本の標本平均を表す確率変数 $\overline{X}$ について

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

大きさ N の有限母集団から無作為に非復元抽出された大きさ n の標本の標本平均を表す確率変数 $\overline{X}$ について，

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

このように，無限母集団から無作為抽出された標本の標本平均を表す確率変数の分散に $\dfrac{N-n}{N-1}$ を掛けると，有限母集団から無作為に非復元抽出された標本の標本平均を表す確率変数の分散になります．標本の大きさ n が充分大きいとき，無限母集団では，非復元抽出でも近似的に復元抽出とみなして中心極限定理を適用することができます．しかし有限母集団では本来はそのようなことができません．しかしながら，数学的には無理があることを承知で，便宜的に，標本平均を表す確率変数の確率分布を正規分布 $\mathrm{N}\left(\mu, \dfrac{N-n}{N-1}\dfrac{\sigma^2}{n}\right)$ で近似することがあります．