< 母集団と標本 >

ある集団を調査したり、ある物質の性質を調べる実験をする場合、あらゆる可能性を全て調査 (実験) をする事を**全数調査**と呼ぶことにする。この全数調査をするかわりに、いくつかの標本を抽出し、その標本を調べることによって全体を推測することを**統計的推測**という。推測すべき全体を**母集団**と呼ぶ。

例 1 10000 人の市民の内閣支持率を調べるために 100 人を選んで調査した。このとき 100 人の選び方に 2 通りある。

①＜復元抽出＞ 10000 人の中から (乱数表などを使って) 無作為に 1 人選び、支持するかしないかを聞き、記録する。これを 100 回くり返す。ただし同じ人が 2 回以上選ばれる可能性がある。

②＜非復元抽出＞ 10000 人の中から無作為に 1 人選び、支持するかしないかを聞き、記録する。次に残りの 9999 人の中から無作為に 1 人選び、同じ事をする。このようにして、一度選んだ人は除外し、残りの中から順に 100 人選ぶ。

この例の場合、無作為に選ぶ事が確率を必要とする理由である。
①の場合 X_k を

$$X_k = \begin{cases} 1: & k \text{ 番目の人が支持すれば } 1 \\ 0: & \quad〃 \quad \text{支持しなければ } 0 \end{cases} \qquad (k = 1, 2, \cdots, 100)$$

とすると、X_k は確率が

$$P(X_k = 1) = \frac{\text{支持する人数}}{10000} \quad , \quad P(X_k = 0) = \frac{\text{支持しない人数}}{10000} \qquad (k = 1, 2, \cdots, 100)$$

である確率変数と考えられ、しかも $X_1, X_2, \cdots, X_{100}$ は独立である。この場合、和 $X_1 + X_2 + \cdots + X_{100}$ は成功確率 $p = \frac{\text{支持する人数}}{10000}$ の二項分布 $B(100, p)$ に従っていると考えられる。
②の場合 X_k を①と同様に定めると、和 $X_1 + X_2 + \cdots + X_{100}$ は 10000 人から 100 人を非復元抽出で選んだときの超幾何分布 $H(10000, 100, p)$ $(p = \frac{\text{支持する人数}}{10000})$ に従う確率変数と考えられる。