

§6.1 点推定

標本調査とは、母集団の中から一部分だけを抽出してそれらの属性値を調べることです。標本調査では母集団から抽出した一部の標本の属性値だけを調べるわけですから、母平均や母分散などの母数は推測するしかありません。ですから、標本調査をするとき、得られた標本のデータから母数をどのように推測するかが問題になります。

大きさ n の標本に属す1番めの事物の属性値 x_1 , 2番めの事物の属性値 x_2 , 3番めの事物の属性値 x_3, \dots, n 番めの事物の属性値 x_n に対して、それらの関数 $f(x_1, x_2, x_3, \dots, x_n)$ の値が母数 θ に近いであろうと推測するとします。このとき、標本確率変数 $X_1, X_2, X_3, \dots, X_n$ の関数 $f(X_1, X_2, X_3, \dots, X_n)$ が標本統計量であるならば、この標本統計量 $f(X_1, X_2, X_3, \dots, X_n)$ を θ の**推定量** (estimator) といい、推定量の実現値を**推定値** (estimate) といいます。

例えば、大きさ4の標本に属す1番めの事物の属性値 x_1 , 2番めの事物の属性値 x_2 , 3番めの事物の属性値 x_3 , 4番めの事物の属性値 x_4 が分かったとき、母平均は $\frac{x_1 + x_2 + x_3 + x_4}{4}$ に近いであろうと見当を付けます。このとき、標本確率変数 X_1, X_2, X_3, X_4 の関数 $\bar{X} = \frac{X_1 + X_2 + X_3 + X_4}{4}$ は標本統計量であり、母平均の推定量です。この推定量の実現値が推定値です。

推定量を決めて未知の母数の推定値を求めることを**点推定** (point estimation) といいます。点推定では推定量をどのように定めるかが問題になります。どのような標本統計量が推定量であるか、はっきりとした規準がある訳ではありません。しかし、未知の母数を推定量の実現値で推定するので、推定量として望ましい性質がいくつかあります。ここでは、推定量が持つべき性質として、不偏性、有効性、一致性、の3つを説明します。

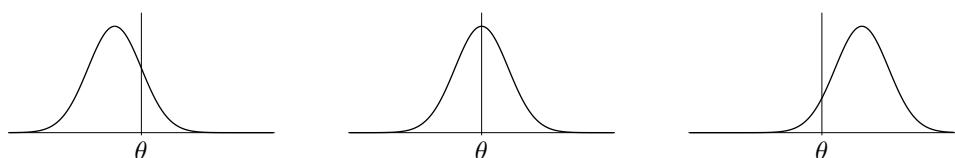
まず不偏性について述べます。

定義 標本統計量 T が母数 θ の不偏推定量 (unbiased estimator) であるとは、

$$E[T] = \theta$$

となることである。

統計量 T が母数 θ の不偏推定量であるとは次のことを意味します：推定量 T による母数 θ の推定値は、 θ より大きいことも小さいこともあるが、平均すると θ になる。 T の確率密度関数のグラフで図解すると例えば次のようになります。

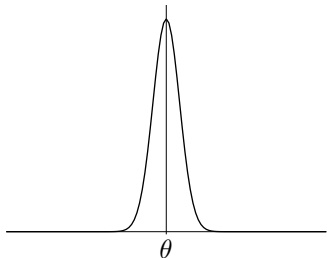


不偏でない ($E[T] < \theta$) 不偏である ($E[T] = \theta$) 不偏でない ($E[T] > \theta$)

次に有効性について述べます。

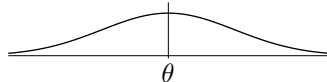
母数 θ の不偏推定量 T_1 と T_2 とについて、確率密度関数のグラフが例えば下図のようになるとします (各々の座標系の座標軸の尺度は同じであるとします)。

T_1 の確率密度関数のグラフ



分散が小さい

T_2 の確率密度関数のグラフ



分散が大きい

このようなとき、 T_1 の実現値は θ に近い値になる確率が大きいのにに対して、 T_2 の実現値は θ から離れた値になる確率が小さくありません。ですから θ の値を推定するためには T_2 より T_1 の方が望ましいことになります。

母数 θ の不偏推定量 T_1 と T_2 とについて、 $V[T_1] < V[T_2]$ のとき、 T_2 より T_1 の方が母数 θ に近い実現値になりやすい、つまり、 T_1 の方が T_2 より母数 θ に近い推定値になる確率が大きくなります。

定義 母数 θ の不偏推定量 T_1 と T_2 とについて、 T_1 が T_2 より有効 (efficient) であるとは

$$V[T_1] < V[T_2]$$

となることである。

最後に一致性について述べます。大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \dots, X_n$ の関数 $T_n = T_n(X_1, X_2, X_3, \dots, X_n)$ が母数 θ の一致推定量 (consistent estimator) であるとは、標本の大きさ n を大きくしていくと統計量 T_n の実現値が母数 θ の値に近い値になる確率が大きくなることです (正確な定義は省きます)；つまり標本の大きさ n が大きいほど統計量 T_n の実現値が母数 θ の推定値として信頼性が高いということです。