

§ 7.1 区間推定

未知の母数（母平均や母分散など） θ の区間推定とは、感覚的にいうと、例えば 99% という確率が与えられたとき、 θ が区間 $[\bigcirc, \square]$ に属す確率が 99% になるような“感じ”の区間 $[\bigcirc, \square]$ を求めることである。

実数 α について $0 < \alpha < 1$ とする。未知の母数 θ について、2 つの標本統計量 X_1 と X_2 とがあって、

$$P(X_1 \leq \theta \leq X_2) = \alpha$$

とする。標本調査を行う度に標本統計量 X_1, X_2 の各々の実現値 x_1, x_2 が決まるが、仮にこのような標本調査を繰り返し繰り返し繰り返し行くと、そのうち $x_1 \leq \theta \leq x_2$ となる回数の割合は α に収束する。そこで、この区間 $[x_1, x_2]$ を、母数 θ の**信頼度** (confidential degree) α の**信頼区間**¹⁾ (confidence interval) という。信頼度のことを信頼係数ともいう。区間推定 (interval estimation) とは、信頼度が与えられたときにその信頼度の信頼区間を求めることである。与えられた信頼度の信頼区間は短いほど望ましい。

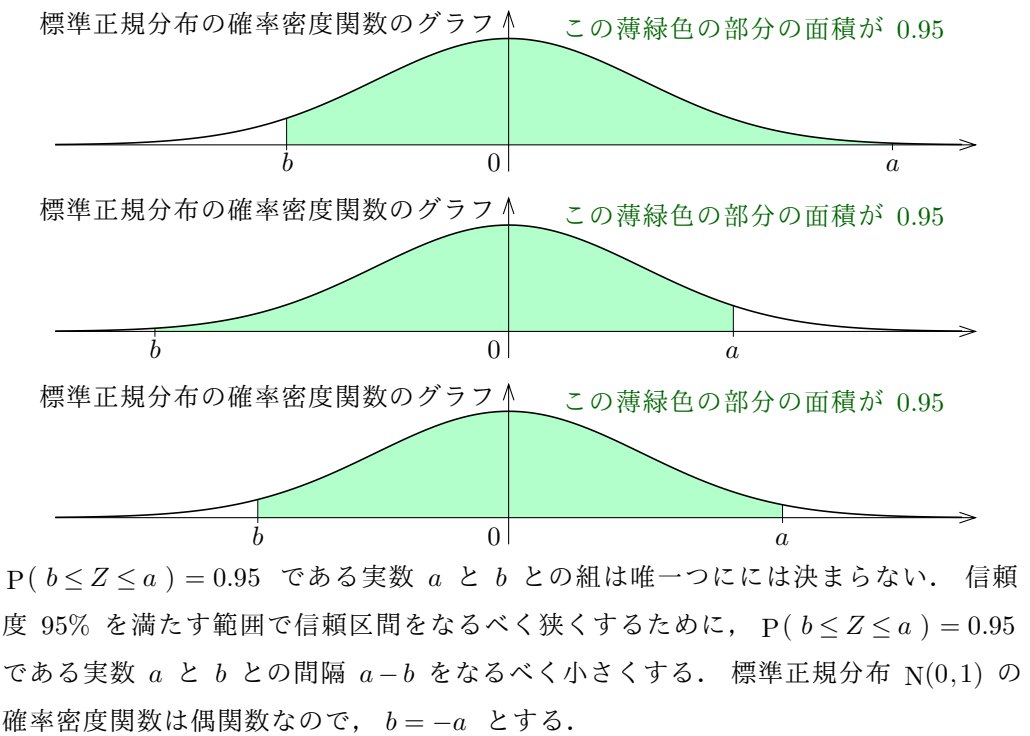
【例】 母標準偏差が 2 である正規母集団から無作為抽出した一つの標本確率変数 X の実現値が 70 であったとする。母平均の値は分からない。母平均の信頼度 95% の信頼区間を求める。母平均を μ とおく。確率変数 X は正規分布 $N(\mu, 2^2)$ に従うので、 X を標準化した確率変数 $Z = \frac{X - \mu}{2}$ は標準正規分布 $N(0, 1)$ に従う。標準正規分布 $N(0, 1)$ の累積分布関数を $\Phi(x)$ とおく：

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt .$$

各実数 a について、

$$\begin{aligned} P(Z \leq a) &= \Phi(a) , \\ \Phi(-a) &= 1 - \Phi(a) . \end{aligned}$$

信頼度 α に対して次のような実数 a, b を考える： $P(b \leq Z \leq a) = 0.95$ 。標準正規分布 $N(0, 1)$ の確率密度関数のグラフを考えると次の図のようになる。



$P(b \leq Z \leq a) = 0.95$ である実数 a と b との組は唯一つには決まらない。信頼度 95% を満たす範囲で信頼区間をなるべく狭くするために、 $P(b \leq Z \leq a) = 0.95$ である実数 a と b との間隔 $a - b$ をなるべく小さくする。標準正規分布 $N(0, 1)$ の確率密度関数は偶関数なので、 $b = -a$ とする。

標準正規分布の確率密度関数のグラフ↑ この薄緑色の部分の面積が 0.95

$P(-a \leq Z \leq a) = 0.95$ である実数 a を求める。

$$\begin{aligned} P(-a \leq Z \leq a) &= P(Z \leq a) - P(Z \leq -a) = \Phi(a) - \Phi(-a) = \Phi(a) - \{1 - \Phi(a)\} \\ &= 2\Phi(a) - 1 . \end{aligned}$$

よって

$$\begin{aligned} P(-a \leq Z \leq a) = 0.95 &\iff 2\Phi(a) - 1 = 0.95 \iff \Phi(a) = 0.975 \\ &\iff a = \Phi^{-1}(0.975) = 1.960 , \end{aligned}$$

故に

$$P(-1.960 \leq Z \leq 1.960) = 0.95 .$$

標準正規分布の確率密度関数のグラフ↑ この薄緑色の部分の面積が 0.95

$Z = \frac{X - \mu}{2}$ なので

$$P\left(-1.960 \leq \frac{X - \mu}{2} \leq 1.960\right) = 0.95 .$$

ここで

$$\begin{aligned} -1.960 \leq \frac{X - \mu}{2} \leq 1.960 &\iff -3.920 \leq X - \mu \leq 3.920 \\ &\iff 3.920 \geq \mu - X \geq -3.920 \\ &\iff -3.920 \leq \mu - X \leq 3.920 \\ &\iff X - 3.920 \leq \mu \leq X + 3.920 . \end{aligned}$$

よって

$$P(X - 3.920 \leq \mu \leq X + 3.920) = 0.95 .$$

このように、確率変数 X について、

$$X - 3.920 \leq \mu \leq X + 3.920$$

となる確率は 0.95 である。いま、確率変数 X の実現値が 70 であるので、 $\mu = E[X]$ の信頼度 95% の信頼区間は

$$[70 - 3.920, 70 + 3.920] = [66.080, 73.920]$$

である。 [終]

$0 < \alpha < 1$ となる実数 α が与えられたとき、未知の母数 θ の信頼度 α の信頼区間を求めることを母数 θ の区間推定という。信頼度は、多くの場合、 $0.95 = 95\%$ あるいは $0.99 = 99\%$ のどちらかに設定される。更に高い確実性が要求されるときは信頼度を更に 1 に近い値にする。

大きさが充分大きい標本を**大標本**といい、大きさがあまり大きくない標本を**小標本**という。区間推定の方法は、大標本のときと小標本のときとは異なるのが普通である。

¹⁾ 区間 $[x_1, x_2]$ が母数 θ の信頼度 α の信頼区間であることは、数学的には、 $x_1 \leq \theta \leq x_2$ となる確率が α になるということではない； θ も x_1 も x_2 も定数なので、そのような確率はない。