§7.1　区間推定

未知の母数（母平均, 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 [○,□] に属す確率が 99% になるような区間 [○,□] を求めることです．

実数 α について $0<\alpha<1$ とします．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，

$$\mathrm{P}(\,X_1 \le \theta \le X_2\,) = \alpha$$

となるとします．標本調査を行うと標本統計量 X_1, X_2 の各々の実現値 x_1, x_2 が決まりますが，仮にこのような標本調査を何回も何回も行うと，そのうち $x_1 \le \theta \le x_2$ となる回数の割合は α に収束します．そこで，この区間 $[x_1, x_2]$ を，母数 θ の**信頼度** (confidencial degree) α の**信頼区間** [1] (confidence interval) といいます．信頼度のことを信頼係数ともいいます．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す．与えられた信頼度の信頼区間は短いほど望ましくなります．

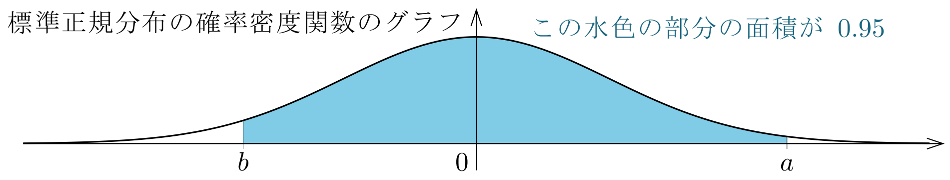
簡単な例を述べます．母標準偏差が 2 である正規母集団から無作為抽出した一つの標本確率変数 X の実現値が 70 であったとします．母平均の値は分かりません．母平均の信頼度 95% の信頼区間を求めます．母平均を μ とおきます．確率変数 X は正規分布 $\mathrm{N}(\mu, 2^2)$ に従いますから，X を標準化した確率変数 $Z = \dfrac{X-\mu}{2}$ は標準正規分布 $\mathrm{N}(0,1)$ に従います．標準正規分布 $\mathrm{N}(0,1)$ の累積分布関数を $\Phi(x)$ とおきます：

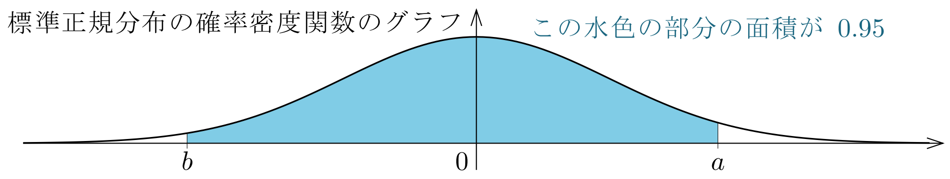
$$\Phi(x) = \mathrm{P}(\,Z \le x\,) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt\ .$$

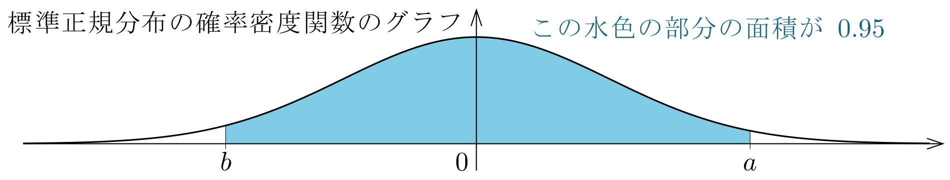
各実数 a について，

$$\mathrm{P}(\,Z \le a\,) = \Phi(a)\ ,$$
$$\Phi(-a) = 1 - \Phi(a)\ .$$

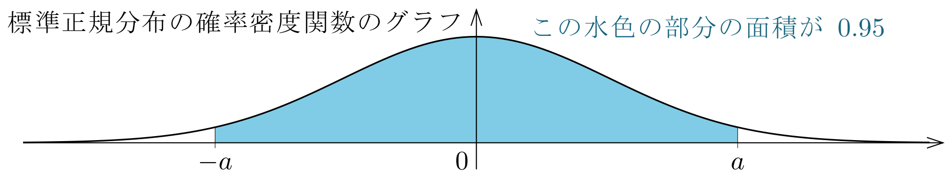
信頼度 α に対して次のような実数 a, b を考えます： $\mathrm{P}(\,b \le Z \le a\,) = 0.95$ ．標準正規分布 $\mathrm{N}(0,1)$ の確率密度関数のグラフを考えると次の図のようになります．







信頼区間をなるべく狭くするために， $\mathrm{P}(\,b \le Z \le a\,) = 0.95$ である実数 a と b との間隔 $a-b$ をなるべく小さくします．標準正規分布 $\mathrm{N}(0,1)$ の確率密度関数は偶関数なので， $b=-a$ とします．



$\mathrm{P}(\,-a \le Z \le a\,) = 0.95$ である実数 a を求めます．

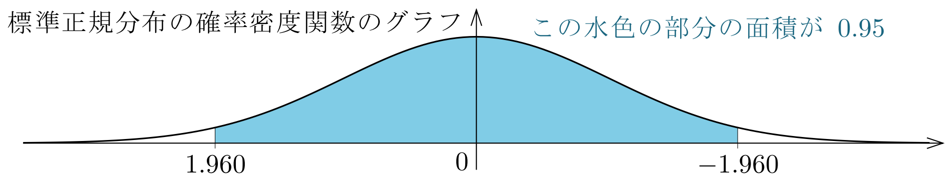
$$\begin{aligned}\mathrm{P}(\,-a \le Z \le a\,) &= \mathrm{P}(\,Z \le a\,) - \mathrm{P}(\,Z \le -a\,) = \Phi(a) - \Phi(-a) = \Phi(a) - \{1-\Phi(a)\} \\ &= 2\,\Phi(a) - 1\ .\end{aligned}$$

よって

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

故に

$$\mathrm{P}(\,-1.960 \le Z \le 1.960\,) = 0.95\ .$$



$Z = \dfrac{X-\mu}{2}$ なので

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

ここで

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

よって

$$\mathrm{P}(\,X-3.920 \le \mu \le X+3.920\,) = 0.95\ .$$

このように，確率変数 X について，

$$X - 3.920 \le \mu \le X + 3.920$$

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

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

です．

$0<\alpha<1$ 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 の区間推定といいます．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ます．更に高い確実性が要求されるときは信頼度を更に 1 に近い値にします．

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**といいます．区間推定の方法は，大標本のときと小標本のときとでは異なるのが普通です．

---

[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， $x_1 \le \theta \le x_2$ となる確率が α になるということでありません； θ も x_1 も x_2 も定数なので，そのような確率はありません．