

応用数理 E 第 9 回目

永幡幸生
新潟大学工学部

5 月 13 日

この授業で行う統計においては

「何らかの理由で分布の形は分かっているが、そのパラメータが分からない。一方で数多くのサンプルを得ることができる。」ことが前提になる。この時に、このパラメータがどのような値なのかを知りたい。

○ 例 製品のサイズ

ある工場で作っている製品のサイズの分布は正規分布 $N(\mu, \sigma^2)$ に従うが、設定によって、 μ が変化する。現在の設定での μ の値を知りたい。

(理想の μ から大きく外れていたら、設定を再調整する。)

○ 例 コイン

どう見ても偏った形をしたコインがあり、コイン投げをすれば、表が出やすそうに見える。このとき表が出る確率 p を知りたい。

今までの授業で扱っていた確率では、分布に加えて、そのパラメータも分かっている時に、統計量（特性量）、例えば期待値などがどのようなになるかを考えてきた。この授業で取り扱う推定、検定では、分布と統計量（のサンプル）が与えられたときにパラメータをどのように推定するかを逆向きに考える。

○ 基本的な設定

$X_1, X_2, \dots, X_n$ を独立同分布でその分布関数は $F_\theta(x)$ とする。

この $X_1, X_2, \dots, X_n$ を母集団、 F_θ を母集団分布と呼ぶ。

この実現値を $x_1, x_2, \dots, x_n$ とする。

注意

分かりにくいと思うが、大文字 X と書いている間は、実験を行う前に、どのようなことが起こりそうかを考えている段階で確率変数、小文字 x に変わったところで、実際に実験を行って、その結果として得られたサンプルを取り扱っている。

すなわち実現値 $x_1, x_2, \dots, x_n$ は $X_1, X_2, \dots, X_n$ の取りうる値の一つで、 $F_\theta(x)$ からの無作為標本と呼ぶ。

F_θ の中にある θ が分からない（知りたい）量で、母数またはパラメータと呼ぶ。

この母数は平均、分散、相関係数であることが多く、これらの場合母平均、母分散、母相関係数と呼ばれる。

用語の整理

母集団： $X_1, X_2, \dots, X_n$

母集団分布： F_θ

実現値、無作為標本： $x_1, x_2, \dots, x_n$

母数、パラメータ： θ

母平均、母分散、母相関係数

点推定

○ 母平均の点推定

多くの分布において、例えばポアソン分布、指数分布、正規分布は平均がパラメータとなっている。それ以外でも平均は重要な統計量であるので、その値を知りたい。

大数の法則： $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ に関して、任意の $\varepsilon > 0$ に対して、

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| > \varepsilon) = 0$$

但し μ は X_k の平均値。

であったので、(μ の値を知らない場合、) 実現値 $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ は n が大きければ μ をよく近似しているはずである。

点推定

前ページの例では

定義

$\bar{X}$: 母平均の推定量

$\bar{x}$: 母平均の推定値
と呼ぶ。

一般には

定義

母数（を近似することが予想されるもの）に対して

$T = T(X_1, X_2, \dots, X_n)$: $(X_1, \dots, X_n)$ の関数) を推定量

$t = T(x_1, x_2, \dots, x_n)$: を推定値
と呼ぶ。

推定量は母数を近似していることが予想されているが、なにか「良い」推定量であることの指標が欲しい。以下に挙げるのはそのいくつかであり、同時にいくつかの指標を満たしていてもおかしくない。

定義

母数 θ に対する推定量 T に関して次のような指標がある：

(1) 一致性： n を大きくすれば一致する。

$$\forall \varepsilon, \lim_{n \rightarrow \infty} P(|T - \theta| < \varepsilon) = 1$$

(2) 不偏性： 期待値はいつも一致する

$$E[T] = \theta$$

(3) 最尤性： 得られたデータをとる確率を最大化する。

点推定

自然な推定量として

$$\text{母平均: } \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

$$\text{母分散: } V = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

$$\text{母共分散: } \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})(Y_k - \bar{Y})$$

例 コイン投げ

ある偏っているコインを 20 回ずつ 500 セット投げた時、表が出た回数の度数分布表が次のようになった。

表	0	1	2	3	4	5	6	7	8	9	10
回数	0	0	0	0	0	0	0	0	1	3	16
表	11	12	13	14	15	16	17	18	19	20	計
回数	26	43	83	99	110	62	37	15	4	1	500

これは二項分布 $B(20, p)$ を並べ替えたものと思えるので、母平均 $20p$ を点推定してみると

$$\bar{x} = \frac{1}{500} \{0 \times 0 + 1 \times 0 + \cdots + 7 \times 0 + 8 \times 1 + 9 \times 3 + 10 \times 16 + \cdots + 19 \times 4 + 20 \times 1\} = \frac{1}{500} 7158 = 14.3160$$

と推定でき、 p も $p = 0.7158$ と推定できる。

点推定

前ページの例と同じように別のコインで同様なことを行おうと

表	0	1	2	3	4	5	6	7	8	9	10
回数	0	0	0	1	1	1	10	15	40	58	74
表	11	12	13	14	15	16	17	18	19	20	計
回数	76	95	69	31	18	8	3	0	0	0	500

であった。

問題

この度数分布表に基づいて、このコインが表になる確率 p を点推定せよ。また同じくこの度数分布表に基づいて、この二項分布 $B(20, p)$ の母分散を点推定せよ。

(電卓等使用可)

表計算ソフトなどを使っても構わないが、どのような計算をするのかや、途中経過を書き入れるように。