応用数理 E 第 12 回目

永幡幸生
新潟大学工学部

5 月 24 日

区間推定：母分散等

○ 正規分布に従う母分散の区間推定

$X_1, X_2, \ldots, X_n$ が独立同分布で $N(\mu, \sigma^2)$ に従うとすると、
$S = \sum_{k=1}^{n} (X_k - \bar{X})^2$ に対して、$\dfrac{S}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う。

$\chi^2(n-1, 1-\alpha/2)$ を χ^2 分布の上側 $1-\alpha/2$ 点 (下側 $\alpha/2$ 点)
$\chi^2(n-1, \alpha/2)$ を χ^2 分布の上側 $\alpha/2$ 点とすれば
母平均の区間推定の時と同様にして、信頼係数 $1-\alpha$ に対して、

$$1-\alpha = P\left(\chi^2(n-1, 1-\alpha/2) \le \frac{S}{\sigma^2} \le \chi^2(n-1, \alpha/2)\right)$$
$$= P\left(\frac{S}{\chi^2(n-1, \alpha/2)} \le \sigma^2 \le \frac{S}{\chi^2(n-1, 1-\alpha/2)}\right)$$

と式変形できるので、S の実現値 s を使って、

$$\left[\frac{s}{\chi^2(n-1, \alpha/2)}, \frac{s}{\chi^2(n-1, 1-\alpha/2)}\right]$$

を得る。

区間推定：母分散等

表計算ソフトなどを使う場合、S に対応した関数がなく、定義通りに、二乗和を計算する手もあるが、分散（の不偏推定量）に対応する $V = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}S$ から計算する方法もある。

区間推定：母分散等

○ 新潟の 8 月の気温の例
前回も書いたが、第 8 回目の資料にあるように、新潟の 8 月の気温（最高気温、最低気温共に）、ヒストグラムを見た感じでは、正規分布に従っているとは思えないので、例としてはあまりふさわしくないが、適応してみる。
Book1.csv にデータをコピーしてあるので、適宜参照するように。

区間推定：母分散等

新潟の 8 月の最高気温に関して、2016 年から 2020 年までの 5 年分のデータ 31×5 日分ある。これらのデータは、年ごとに $N(\mu, \sigma^2)$ に従うが、この σ^2 を信頼係数 95% で区間推定する。
データ数 ＝ 日数 $n = 31$、であり、各年

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

の実現値 v を計算すると、

$$v_{16} = 6.5, \quad v_{17} = 5.1, \quad v_{18} = 13, \quad v_{19} = 13, \quad v_{20} = 6.7$$

になる。

$V = \dfrac{1}{n-1} S$ であるのでその実現値に関して $v = \dfrac{1}{n-1} s$ を満たし、付表 4 より

$$\chi^2(n-1, 0.025) = \chi^2(30, 0.025) = 46.98, \chi^2(30, 0.975) = 16.791$$

3 ページ前の結果に代入することで 2015 年から順に

$$[4.14, 11.59], \quad [3.28, 9.16], \quad [8.30, 23.2], \quad [8.28, 23.2], \quad [4.25, 11.9]$$

と与えられる。(2018 年と、2019 年の違いは、実際は $v_{18} = 13.0041, \quad v_{19} = 12.9699$ を使用しているため。)

区間推定：母分散等

∘ 比率の推定

選挙の速報で、ものすごい時間で当選確実などと出てくるが、その基礎となる話。速報自体は推定というよりは、検定の話。

$X_1, X_2, \ldots, X_n$ は独立同分布で

$$X_k = \begin{cases} 1 & \text{確率 } p \\ 0 & \text{確率 } 1-p \end{cases}$$

のように 2 つの値しかとらない場合に 1 が出る確率 p を推定したい。

$X = \sum_{k=1}^{n} X_k$ は二項分布 $B(n,p)$ に従う。従って　$E[X] = np$ なので $\frac{X}{n}$ は p の良い推定量である。さらに n が十分大きければ、

$\frac{X-np}{\sqrt{np(1-p)}}$ は標準正規分布で近似できる。

区間推定：母分散等

$z_{\alpha/2}$ を標準正規分布の上側 $\alpha/2$ 点とすると

$$1-\alpha \cong P\left(-z_{\alpha/2} \leq \frac{X-np}{\sqrt{np(1-p)}} \leq z_{\alpha/2}\right)$$

と計算できるが、括弧の中身を p に関して解くのは大変なので、
分母の $\sqrt{\ }$ の中に入っている p は $\frac{X}{n}$ で近似することで

$$1-\alpha \cong P\left(\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}} \leq p \leq \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}}\right)$$

と計算できるので、p は

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

と推定できる。

区間推定：母分散等

なお、近頃では、コンピュータ自体の性能が上がったことに加え、プログラムを組む方法論も発達したため、二項分布の確率などは（表計算ソフトなどでも）簡単に一つの関数で出してくれる。
（$\dfrac{n!}{k!(n-k)!}$ は結果的にはあまり大きい数にはならないが、$n!$ 自体は大きい数になり、単純計算ではオーバーフローしてしまうことは良く起きる。）
この値を用いて区間推定できなくはないが、やはり近似を使わないと推定できず、計算が煩雑になるため、あまり意味を持たない。

区間推定：母分散等

第 9 回目の例：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

第 9 回目では $B(20, p)$ を 500 回行ったものを並べなおしたものとみなしたが、
一方で $20 \times 500 = 10,000$ 回コインを投げて、
$0 \times 0 + \cdots + 8 \times 1 + 9 \times 3 + 10 \times 16 + \cdots + 19 \times 4 + 20 \times 1 = 7158$
回表が出たともみなせる。
このとき $\bar{x} = \frac{7158}{10000} = 0.7158$ なので付表 2 から $z_{0.025} = 1.96$ を使えば、2 ページ前の結果に代入して、
このコインに関して、表が出る確率 p は信頼係数 0.95 で $[0.7069, 0.7246]$ と区間推定できる。

区間推定：母分散等

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 を信頼係数 0.95 で区間推定せよ。
(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、途中経過を書きいれるように。