応用数理 E 第 11 回目

永幡幸生
新潟大学工学部

2022 前期

区間推定：母分散が未知で正規分布に従う場合

前回の話で分かるように、区間推定では与えられた α に対して
$1 - \alpha = P(T_1 \leq \theta \leq T_2)$
となるような T_1, T_2 をうまく作るという話であった。
今回は母分散も未知で正規分布に従う母平均の区間推定を考える。
同時に考えることも可能であるが、サンプル数のサイズの大小で、分けた方が考えやすく、正規分布に従っているとは思えない分布の区間推定（ただしサンプルサイズは十分大きいとする）とも関連する。
上では等号で書いたが、必要であれば近似を行い
$1 - \alpha \cong P(T_1 \leq \theta \leq T_2)$
とする必要もある。

区間推定：母分散が未知で正規分布に従う場合

◦ 母分散も未知で正規分布に従う母平均の区間推定

(1) サンプルサイズが大きいとき

とりあえずどのくらいからサンプルサイズが大きいと思えるかは考えないことにする。

前回の母分散 σ^2 が既知の場合は

$[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}]$

と推定した。この σ をサンプルから点推定して代入すると、サンプルサイズが大きい仮定の下で、

◦ σ に関して、真の値と、点推定した値の差が小さいはず。

◦ $\frac{1}{\sqrt{n}}$ で σ を割っていて、n が大きいからこれがかかっている部分の寄与が小さくなるはず。

区間推定：母分散が未知で正規分布に従う場合

と自然に思えるので、σ^2 を分散の不偏推定量

$V = \dfrac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$ の実現値 $v = \dfrac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$ で近似する。

結果として、$[\bar{x} - \dfrac{\sqrt{v}}{\sqrt{n}} z_{\alpha/2}, \bar{x} - \dfrac{\sqrt{v}}{\sqrt{n}} z_{\alpha/2}]$

と推定される。

注意

通常 V は不偏推定量

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

を使い、$n-1$ で割っていることに注意する。

区間推定：母分散が未知で正規分布に従う場合

(2) サンプルサイズが大きくないとき
前回と同様にうまく確率変数 T_1, T_2 を見つければよい。
実際に次の事実から T_1, T_2 を見つけることができた。
◦ 自由度 n の t-分布は独立で標準正規分布に従う X と自由度 n の χ^2 分布を使って、

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

であったが、そもそも自由度 n の χ^2 分布は独立同分布で標準正規分布に従う $Z_1, Z_2, \ldots, Z_n$ に対して $\chi^2 = \sum_{k=1}^n Z_k^2$ の分布である。

区間推定：母分散が未知で正規分布に従う場合

少し議論が必要ではあるが、独立同分布で正規分布 $N(\mu, \sigma^2)$ に従う $X_1, X_2, \ldots, X_n$ に対して、

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

とおくと、$\dfrac{\bar{X} - \mu}{\sqrt{\frac{V}{n}}}$ は自由度 $n-1$ の t-分布に従う。

この事実と、前回と同様の議論で、$t(n-1; \alpha/2)$ を t-分布の上側 $\alpha/2$ 点として、与えられた信頼係数 $1-\alpha$ と、$\bar{X}, V$ の実現値 $\bar{x}, v$ に対して、

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

と区間推定できる。

区間推定：母分散が未知で正規分布に従う場合

問題

「前回と同様の議論で」と省略した部分を与えよ。
特に式変形でどの不等式と、どの不等式が対応しているかを与えよ。

区間推定：母分散が未知で正規分布に従う場合

∘ サンプルサイズに関して、また整合性に関して、
t-分布の上側 α 点の付表 3 を見ると、n に関して、1 から 40 まではあるが、その先は $60, 120, \infty$ しかない。
一方で ∞ での値 $t(\infty; \alpha)$ と標準正規分布の上側 α 点の付表 2 z_α を比べると一致している。また 120 くらいでそもそも $t(120, \alpha) \cong z_\alpha$ として構わない。
このことから、$n \cong 100$ 程度まではサンプルサイズが大きくないと思って、このスライドでは (2) とした、t-分布を用いた推定を行う。
それより大きい場合は、(1) で行った正規分布の推定で、分散を点推定により近似した推定を行う。
これにより (1)(2) に分けて考えるのは問題ない（整合性がある）ことが分かる。
なお、n が 40 から 120 までの間は、この付表 3 で抜けている値に関しては、線形補間をすることもある。

区間推定：母分散が未知で正規分布に従う場合

◦ 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：母分散が未知で正規分布に従う場合

新潟の 8 月の最高気温に関して、2017 年から 2021 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、σ^2 も未知として各年の母平均を信頼係数 95% で区間推定する。

データ数 $=$ 日数 $n = 31$、であり、各年

標本平均 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ の実現値 $\bar{x}$ を計算すると、

$\bar{x}_{17} = 29.8, \quad \bar{x}_{18} = 30.7, \quad \bar{x}_{19} = 31.7 \quad \bar{x}_{20} = 31.5 \quad \bar{x}_{21} = 30.8$

となり、$V = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X})^2$ の実現値 v を計算すると、

$v_{17} = 5.1, \quad v_{18} = 13, \quad v_{19} = 13, \quad v_{20} = 6.7, \quad v_{21} = 13$

になる。信頼係数 95% としたので教科書付表 3 より $t(30, 0.025) = 2.042$ を得る。$n = 31$ より 4 ページ前の結果に代入することで、2016 年から順に

$[29.0, 30.7], \quad [29.4, 32.1], \quad [30.4, 33.0], \quad [30.5, 32.4], \quad [29.55, 32.1]$

と与えられる。

区間推定：母分散が未知で正規分布に従う場合

問題

最高気温と同様に最低気温に関しても 5 年分信頼係数 95% で区間推定をせよ。但し σ^2 も未知とする。（電卓等使用可、付表を使う）
前回と同様、表計算ソフトなどを使って構わないが、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。