母分散・母平均値の区間推定

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I L12(2024-07-08 Mon)

最終更新: Time-stamp: "2024-07-08 Mon 11:26 JST hig"

今日の目標

- 岩薩林 確率・統計 p.123 カイ二乗分布を説明できる
- 岩薩林 確率・統計 §7.2 母分散を区間推定できる
- 岩薩林 確率・統計 p.127 t 分布を説明できる
- 岩薩林 確率・統計 §7.1 母平均値を区間推定できる

L11-Q1

Quiz 解答: 正規分布の上側確率

1. $1 - \Phi(u)$.
2. $z(\alpha) = \Phi^{-1}(1-\alpha)$.
3. Z の確率密度関数が偶関数であることからいえる.
4. 上より，$P(Z < -u) = P(u < Z) = \alpha/2$ となる u を考えて，$z(\frac{\alpha}{2}) = \Phi^{-1}(1-\frac{\alpha}{2})$.
5. $\Phi^{-1}(1-\frac{0.05}{2}) = \Phi^{-1}(0.975) =$ .ppf(0.975) $= 1.96$.

L11-Q2

Quiz 解答: 区間推定の性質

1

L11-Q3

Quiz 解答: 母比率の区間推定

A 候補に投票したを $X=1$, しなかったを $X=0$ とする.

1. 標本比率は $\hat{p} = \frac{35}{50} = 0.7$. 母比率 p を 0.7 と推定する.
2. 母比率 p の信頼係数 $1 - \alpha = 0.95$ の信頼区間は,
   $\Phi^{-1}(1 - \frac{0.05}{2}) = z(\frac{0.05}{2}) = 1.96$ より

$$\frac{7}{10} - 1.96 \times \sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}} < p < \frac{7}{10} + 1.96 \times \sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}$$
$$0.7 - 0.13 < p < 0.7 + 0.13$$
$$0.57 < p < 0.83$$

$0.5 < 0.57$ なので，信頼係数 0.95 では当選ってことですね (放送用語「当選確実」で, 多くの選挙区で判定したとき, 後で社長があやまらなきゃいけない確率は 0.05).

3. 母比率 p の信頼係数 $1-\alpha=0.99$ の信頼区間は, $z(\frac{0.01}{2})=2.58$ より

$$\frac{7}{10}-2.58\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}} < p < \frac{7}{10}+2.58\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}$$

$$0.7-0.17 < p < 0.7+0.17$$

$$0.53 < p < 0.87$$

正規母集団からの標本抽出

サイズ n の標本.
$Z_1, \ldots, Z_n \sim \mathrm{N}(0, 1^2)$ 独立同分布.
$\mathrm{E}[Z_i] = 0, \mathrm{V}[Z_i] = 1^2, \mathrm{E}[Z_i^2] = \mathrm{V}[Z_i] + \mathrm{E}[Z_i]^2 = 1.$
和 $S = Z_1 + \cdots + Z_n$. 実は $\sim \mathrm{N}(0, n)$.
標本平均値 $\overline{Z} = \frac{1}{n}(Z_1 + \cdots + Z_n)$.
$\mathrm{E}[\overline{Z}] = \frac{1}{n}(\mathrm{E}[Z_1] + \cdots + \mathrm{E}[Z_n]) = 0.$
$\mathrm{V}[\overline{Z}] = \frac{1}{n^2}(\mathrm{V}[Z_1] + \ldots \mathrm{V}[Z_n]) = \frac{1}{n}.$
標本分散ぽいもの $\frac{1}{n}W = \frac{1}{n}((Z_1 - 0)^2 + \cdots + (Z_n - 0)^2)$.

カイ二乗分布 岩薩林 確率・統計 p.123

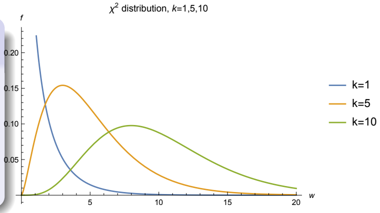
カイ二乗分布

$Z_1, \ldots, Z_k \sim \mathrm{N}(0, 1^2)$, iid のとき, 確率変数 $W = Z_1^2 + \cdots + Z_k^2$ は, 自由度 k のカイ二乗分布 χ_k^2 にしたがう.

言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

χ_k^2 の確率密度関数

$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (\text{他}) \end{cases}$$



$W \sim \chi^2_k$ に対して,
$\mathrm{E}[W] = \mathrm{E}[Z_1^2] + \cdots + \mathrm{E}[Z_k^2] = k.$
$\mathrm{V}[W] = \mathrm{V}[Z_1^2] + \cdots + \mathrm{V}[Z_k^2] = ?$
$\mathrm{V}[Z^2] = \mathrm{E}[(Z^2)^2] - \mathrm{E}[Z^2]^2 = \mathrm{E}[Z^4] - 1^2.$

$$
\begin{aligned}
\mathrm{E}[Z^4] &= \int_{-\infty}^{+\infty} z^3 \cdot \left(z \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \right) \mathrm{d}z \\
&= \left[z^3 \left(-\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \right) \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} 3z^2 \left(-\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \right) \mathrm{d}z \\
&= 0 + 3\mathrm{E}[Z^2] = 3.
\end{aligned}
$$

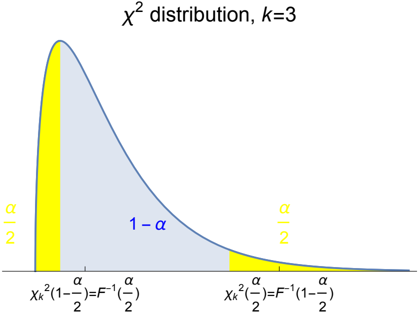
モーメント $\mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

```
rvw=stats.chi2(df=k)  # k は自由度
```

上側確率 が $\alpha = P(W > w)$ となる境い目 $w = \chi_k^2(\alpha) = F^{-1}(1-\alpha)$.
F はカイ二乗分布の累積分布関数.

岩薩林 確率・統計 付表 3 付表 3 は付表 1 とフォーマットが違う



$\chi_k^2(\alpha)$ の定義

岩薩林 確率・統計 例題 5.7

$\alpha = P(W > \chi_k^2(\alpha))$.
$\chi_k^2(\alpha) = F^{-1}(1-\alpha)$.

L12-Q1

Quiz(カイ二乗分布)

自由度 2 のカイ二乗分布 χ^2_2 にしたがう確率変数 W を考える．累積分布関数を $F_{\mathrm{W}}(w)$ とする．

1. $P(W < 0.1)$ を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．
2. $P(W > 0.9)$ を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．
3. $P(W < w) = 0.1$ となる w を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．
4. $P(W > w) = 0.1$ となる w を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．

L12-Q2

Quiz(カイ二乗分布の確率と $\chi_k^2(\alpha)$)

標準正規分布にしたがう確率変数 $Z \sim \mathrm{N}(0, 1^2)$ と，自由度 $k = 1$ のカイ二乗分布にしたがう確率変数 $W \sim \chi_1^2$ を考える．各分布の累積分布関数の逆関数 Φ^{-1}, F^{-1} を使って答え，さらに Python, 数表を使って数値にしよう．

1. 確率 $P(Z > x_0) = 0.025$ となる $x_0 = z(0.025)$ を求めよう．
2. 確率 $P(Z > x_0) = 1 - 0.025$ となる $x_0 = z(1 - 0.025)$ を求めよう．
3. 確率 $P(W > w_0) = 0.05$ となる $w_0 = \chi_1^2(0.05)$ を求めよう．
4. 確率 $P(W > w_0) = 1 - 0.05$ となる $w_0 = \chi_1^2(1 - 0.05)$ を求めよう．

ここまで来たよ

11 区間推定・母比率の区間推定

12 母分散・母平均値の区間推定
- 不偏標本分散のちらばりとカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散未知)

ちらばり (不偏標本分散) のちらばりを考えたい

標本データのちらばりって? $\sqrt{\text{母分散}} \overset{\text{点推定}}{\longleftarrow} \sqrt{\text{不偏標本分散}}$

- 母分散の点推定の精度って?

	の点推定	の区間推定
母平均値 /比率 μ	標本平均値/比率 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	$\overline{X} - \boxed{\text{正}}\sqrt{\quad} < \mu < \overline{X} - \boxed{\text{負}}\sqrt{\quad}$
母分散 σ^2	不偏標本分散 $S^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots]$	$S^2 \times \boxed{\text{小}} < \sigma^2 < S^2 \times \boxed{\text{大}}$

母集団が正規分布にしたがうとき

- 標本平均値/比率の分布 (正規分布) をうまく平行移動, 拡大縮小すると標準正規分布 $\mathrm{N}(0, 1^2)$
- 不偏標本分散の分布をうまく拡大縮小するとカイ二乗分布 χ_k^2

確率統計 II(2023)L??

不偏標本分散のしたがう分布

不偏標本分散のしたがう分布 岩薩林 確率・統計 定理 5.6

正規分布 $\mathrm{N}(\mu, \sigma^2)$ からサイズ n の標本 $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, iid を取り出すとき, 不偏標本分散

$$S^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

から定めた

$$W = (n-1) \times \frac{S^2}{\sigma^2}$$

は, 自由度 $k = n - 1$ の カイ二乗分布 χ^2_{n-1} にしたがう.

比 $\frac{\text{不偏標本分散}}{\text{母分散}}$ は 1 に近いところに分布するが,
実は，この比は確率変数 $\frac{W}{n-1}$. $(W \sim \chi^2_{n-1})$

証明じゃないけど説明

独立な $X_i \sim \mathrm{N}(\mu, \sigma^2)$ $(i = 1, \ldots, n)$ に対して,

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
μ を $\overline{X}$ に替えた,

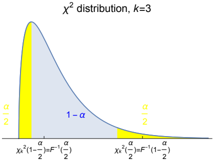
$$W = \left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.

$$W = (n-1) \times \frac{S^2}{\sigma^2}.$$

母分散の区間推定

$$P(F^{-1}(\tfrac{\alpha}{2}) < (n-1)\frac{S^2}{\sigma^2} < F^{-1}(1-\tfrac{\alpha}{2})) = 1-\alpha$$



不等式を σ^2 について解いて次の結果を得る.

$\chi_k^2(\alpha)$ の 定 義

岩薩林 確率・統計 例題 5.7

$\alpha = P(W > \chi_k^2(\alpha))$.
$\chi_k^2(\alpha) = F^{-1}(1-\alpha)$.

母分散の信頼区間 岩薩林 確率・統計 定理 7.3

標本の不偏標本分散が S^2 のとき, 母分散 σ^2 の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{n-1}{F^{-1}(1-\frac{\alpha}{2})} \times S^2 < \sigma^2 < \frac{n-1}{F^{-1}(\frac{\alpha}{2})} \times S^2$$

$$\frac{n-1}{\chi_{n-1}^2(\frac{\alpha}{2})} \times S^2 < \sigma^2 < \frac{n-1}{\chi_{n-1}^2(1-\frac{\alpha}{2})} \times S^2$$

だいたい S^2 だけど,「かける」補正係数 $(n-1)/W \simeq 1$, $W \sim \chi_{n-1}^2$.

L12-Q3

Quiz(母分散の区間推定)

あるファーストフードチェーンのポテトフライ S の重さは正規分布に従うという.
お店で 9 個のポテトフライ S サイズを買って重さを量り, サイズ 9 の標本とした.
このとき標本平均値は 80g, 不偏標本分散は $72\mathrm{g}^2$ だった.
母分散を信頼係数 $1-\alpha=0.95$ で区間推定しよう.

岩薩林 確率・統計 例題 7.3(p163), 問題 4(p.164), 練習問題 7.1(4)(p.173)

ここまで来たよ

11 区間推定・母比率の区間推定

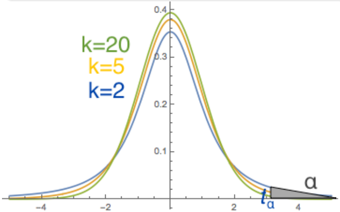
12 母分散・母平均値の区間推定

- 不偏標本分散のちらばりとカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散未知)

t 分布

t 分布

$Z \sim \mathrm{N}(0, 1^2)$, $W \sim \chi^2_k$, Z と W が独立,
のとき連続型確率変数 $T = \dfrac{Z}{\sqrt{W/k}}$ のしたがう分布を自由度 k の (スチューデントの, またはゴセットの)t 分布 t_k という.



自由度 k が小さいとき, $\mathrm{N}(0, 1^2)$ より低く広い.
自由度 $k \to +\infty$ で $\mathrm{N}(0, 1^2)$ に一致する.

t 分布の確率密度関数と累積分布関数 岩薩林 確率・統計 付表 2

Python scipy.stats.t(df=k)

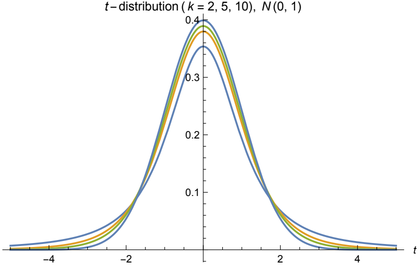
岩薩林 確率・統計 付表 2 は付表 1 とフォーマットが違う

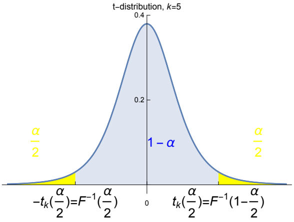
上側確率が $\alpha = P(T > t)$ となる $t = F_{\mathrm{t}}^{-1}(1-\alpha)$ を $t = t_k(\alpha)$ とかく.

t 分布の確率密度関数は偶関数なので, $t_k(1-\alpha) = -t_k(\alpha)$.

$\mathrm{t}_k \to \mathrm{N}(0, 1^2), t_k(0.025) \to 1.960, t_k(0.005) \to 2.576$ ($k \to +\infty$).

岩薩林 確率・統計 図 5.8(p.128)





L12-Q4

Quiz(t 分布の確率と $t_k(\alpha)$)

標準正規分布にしたがう確率変数 $Z \sim \mathrm{N}(0, 1^2)$ と, 自由度 $k = 40$ の t 分布にしたがう確率変数 $T \sim \mathrm{t}_{40}$ を考える. 各分布の累積分布関数の逆関数 Φ^{-1}, F^{-1} を使って答え, さらに Python を使って数値にしよう.

1. 確率 $P(|Z| > z) = 0.05$ となる z を求めよう.
2. 確率 $P(|T| > w) = 0.05$ となる w を求めよう.

母平均値の区間推定 (正規母集団, 母分散未知) 岩薩林 確率・統計 §7.1

母分散 σ^2 既知なら，標本平均値を標準化 $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$.
ふつう, μ がわからないときは σ^2 もわかってない.
σ^2 のかわりに不偏標本分散 S^2 (確率変数) を使った, 標準化もどき

$T = \frac{\overline{X}_{(n)}-\mu}{\sqrt{S^2/n}} = \frac{\frac{\overline{X}_{(n)}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{S^2}{\sigma^2}}}$ でやっちゃいたい.

分子 $\sim \mathrm{N}(0,1^2)$, 分母 $= \sqrt{W/(n-1)}, W \sim \chi^2_{n-1}$ より, $T \sim \mathrm{t}_{n-1}$.

T の分布

母集団 $\mathrm{N}(\mu,\sigma^2)$ から, サイズ n の標本 $X_1,\cdots,X_n$ を取り出したとき,

$$T = \frac{\overline{X}_{(n)} - \mu}{\sqrt{S^2/n}}$$

は, 自由度 $n-1$ の Student の t 分布にしたがう.

母集団が厳密に正規分布にしたがわなくても近似的に正しいことが多い.

母平均値の信頼区間 (母分散未知) 岩薩林 確率・統計 定理 7.1(7.3)

(母分散未知の) 正規分布 $\mathrm{N}(\mu, ?^2)$ にしたがう母集団から, サイズ n の標本を得たとき, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間は

$$\overline{X} - F_{\mathrm{t}}^{-1}(1-\tfrac{\alpha}{2}) \times \sqrt{S^2/n} < \mu < \overline{X} - F_{\mathrm{t}}^{-1}(\tfrac{\alpha}{2}) \times \sqrt{S^2/n}.$$

$$\overline{X} - t_{n-1}(1-\tfrac{\alpha}{2}) \times \sqrt{S^2/n} < \mu < \overline{X} + t_{n-1}(\tfrac{\alpha}{2}) \times \sqrt{S^2/n}.$$

ただし, $\overline{X}$: 標本平均値, S^2: 不偏標本分散, n: 標本サイズ, $t_{n-1}(\frac{\alpha}{2})$: 自由度 $n-1$ の t 分布の上側確率が $\frac{\alpha}{2}$ となる点.

母分散既知と比べて,
$\sigma^2 \rightsquigarrow S^2$
$z(\frac{\alpha}{2}) \rightsquigarrow t_{n-1}(\frac{\alpha}{2})$

L12-Q5

Quiz(母平均値の区間推定 (母分散未知))

あるドーナツ製造マシンが製造するドーナツの重さ Xg は, 正規分布にしたがう確率変数である
製造された 4 個のドーナツの重さを測定したところ, 次のようだった.
$51\text{g}, 52\text{g}, 47\text{g}, 50\text{g}.$

1. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう. まず, $t_k(\alpha)$ または F_{t}^{-1} で書き, 小数に直そう.
2. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.99$ で区間推定しよう.

岩薩林 確率・統計 例題 7.1(p.158), 問題 1(p.158) 第 7 章練習問題 1(2)