母分散・母平均値の区間推定

樋口さぶろお　https://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習　L13(2020-12-23 Wed)

最終更新: Time-stamp: "2021-01-22 Fri 12:14 JST hig"

今日の目標

- 母分散を区間推定できる　岩薩林 確率・統計 §7.2
- 母平均値を区間推定できる　岩薩林 確率・統計 §7.1

L12-Q1

Quiz 解答:母平均値, 母分散, 母比率の点推定　　在庫のフライドチキンの重さを X とすると,

1. 標本平均値 (の実現値) は $\overline{X} = \frac{1}{6}(117 + \cdots + 112) = 111\mathrm{g}$ なので, 母平均値は 111g と推定できる.
2. 標本期待値 (の実現値) は $\overline{X^2} = \frac{1}{6}(117^2 \cdots + 112^2) = 37078/3\mathrm{g}^2$ なので, 母期待値は $37078/3\mathrm{g}^2$ と推定できる.
3. 不偏標本分散 (の実現値) は,
   $S^2 = \frac{1}{6-1}[(117-111)^2 + \cdots + (112-111)^2] = 46\mathrm{g}^2$ なので, 母分散は $46\mathrm{g}^2$ と推定できる.
4. 標本比率 (の実現値) は, $p = \frac{1}{6}[1+0+0+1+0+1] = 0.5$ なので, 母比率は 0.5 と推定できる.

L12-Q2

Quiz 解答:母平均値, 母分散, 母比率の点推定

1. $\mathrm{E}[X] = 2 \cdot \frac{2}{3} = \frac{4}{3}$.
2. $\mathrm{V}[X] = 2 \cdot \frac{2}{3}\frac{1}{3} = \frac{4}{9}$.
3. $\overline{X} = \frac{1}{6}(0+0+0+1+1+2) = \frac{2}{3}$.
4. $S^2 = \frac{1}{6-1}((0-\frac{2}{3})^2 + \cdots + (2-\frac{2}{3})^2) = \frac{7}{15}$.

L12-Q3

Quiz 解答:母平均値の区間推定 (母分散既知)

1. 重さの標本平均値は $m = 50\mathrm{g}$. よって, 信頼係数 0.95 の信頼区間は

$$50 - 1.96 \times \sqrt{\frac{9}{4}} < \mu < 50 + 1.96 \times \sqrt{\frac{9}{4}}.$$

   すなわち, $47.06 < \mu < 52.94$.
2. 同様に,

$$50 - 2.58 \times \sqrt{\frac{9}{4}} < \mu < 50 + 2.58 \times \sqrt{\frac{9}{4}}.$$

   すなわち, $46.13 < \mu < 53.87$.

L12-Q4

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. X の母分散は $0.7 \times (1 - 0.7) = 0.21$ と推定する.
   母比率 p の信頼係数 $1 - \alpha = 0.95$ の信頼区間は,

$$
\begin{aligned}
0.7 - 1.96 \times \sqrt{\frac{1}{50} \cdot 0.7 \cdot (1 - 0.7)} <& p < 0.7 + 1.96 \times \sqrt{\frac{1}{50} \cdot 0.7 \cdot (1 - 0.7)} \\
0.7 - 0.13 <& p < 0.7 + 0.13 \\
0.57 <& p < 0.83
\end{aligned}
$$

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 0.99 の信頼区間は,

$$
\begin{aligned}
0.7 - 2.58 \times \sqrt{\frac{1}{50} \cdot 0.7 \cdot (1-0.7)} <& p < 0.7 + 2.58 \times \sqrt{\frac{1}{50} \cdot 0.7 \cdot (1-0.7} \\
0.7 - 0.17 <& p < 0.7 + 0.17 \\
0.53 <& p < 0.87
\end{aligned}
$$

信頼係数 0.99 のほうが慎重な判断基準ですが, それでも当選ってことですね.

ここまで来たよ

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 μ	標本平均値 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	$\overline{X} - \boxed{}\sqrt{} < \mu < \overline{X} + \boxed{}\sqrt{}$
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

- 標本平均値の分布 ($\boxed{\text{正規分布}}$) 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^2_k

確率統計☆演習 I(2020)L11

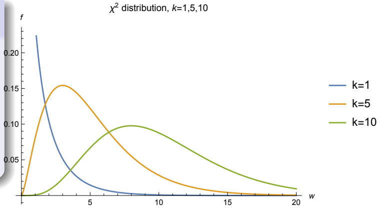
カイ二乗分布 岩薩林 確率・統計 p.123

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.

χ_k^2 の確率密度関数

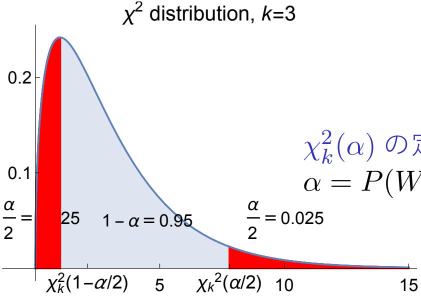
$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (\text{他}) \end{cases}$$



$W_k \sim \chi_k^2$ に対して,
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{V}[W_k] = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

χ^2 分布表 岩薩林 確率・統計 付表 3

上側確率 $\alpha = P(W > \chi_k^2(\alpha))$ となる $\chi_k^2(\alpha)$ の表.
付表 3 は付表 1 とフォーマットが違う
付表 1 横軸縦軸 座標 z, セル内 確率 $I(z)$
付表 3 縦軸 自由度 k, 横軸 確率 α, セル内 座標 $w = \chi_k^2(\alpha)$
付表 2 縦軸 自由度 k, 横軸 確率 α, セル内 座標 $t = t_k(\alpha)$



$\chi_k^2(\alpha)$ の定義 岩薩林 確率・統計 例題 5.7
$\alpha = P(W > \chi_k^2(\alpha))$.

L13-Q1

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.

1. 標準正規分布の数表から確率 $P(Z > x_0) = 0.025$ となる $x_0 = z(0.025)$ を求めよう.
2. 標準正規分布の数表から確率 $P(Z > x_0) = 1 - 0.025$ となる $x_0 = z(1 - 0.025)$ を求めよう.
3. カイ二乗分布の数表を用いて, 確率 $P(W > w_0) = 0.05$ となる $w_0 = \chi_1^2(0.05)$ を求めよう.
4. カイ二乗分布の数表を用いて, 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

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, 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,

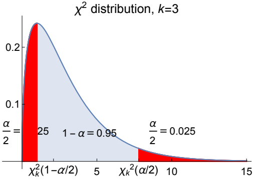
$$n \times \frac{1}{n}\left[\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2\right]$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
不偏標本分散 S^2 に対して,

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

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.
$-\mu$ でなく $-\overline{X}$ であるため自由度 $n-1$.

母分散の区間推定



$\chi_k^2(\alpha)$ の定義 岩薩林 確率・統計 例題 5.7

$\alpha = P(W > \chi_k^2(\alpha))$.

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

不等式を σ^2 について解いて次の結果を得る.

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

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

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

L13-Q2

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.

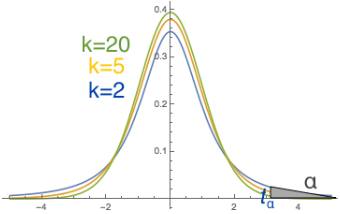
岩薩林 確率・統計 例題 7.3(p163), 問題 4(p.164), 練習問題 7.1(4)(p.173)

ここまで来たよ

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.

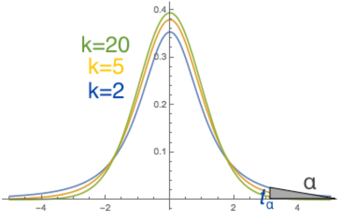


t 分布

自由度 k の t 分布 t_k 岩薩林 確率・統計 §5.5

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

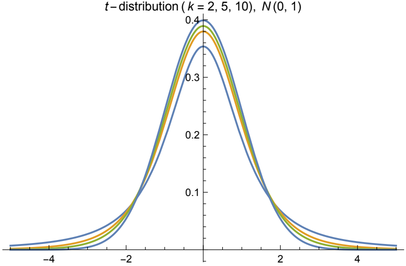
$$\text{確率密度関数 } f_k(x) = A_k \cdot \left(1 + \frac{1}{k}x^2\right)^{-\frac{k+1}{2}}.$$

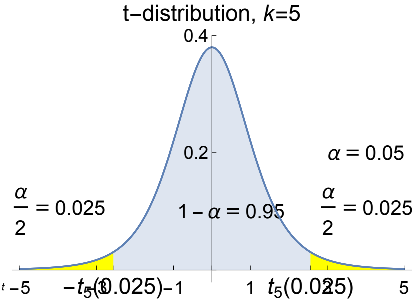


t 分布表 岩薩林 確率・統計 付表 2

上側確率 $\alpha = P(T > t_k(\alpha))$ となる $t_k(\alpha)$ の値の表.
t 分布の確率密度関数は偶関数なので, $t_k(1-\alpha) = -t_k(\alpha)$.
付表 2 は付表 1 とフォーマットが違う
付表 1 横軸縦軸 座標 z, セル内 確率 $I(z)$
付表 2 縦軸 自由度 k, 横軸 確率 α, セル内 座標 $t = t_k(\alpha)$
`https://www.geogebra.org/classic#probability`
$\mathrm{t}_k \to \mathrm{N}(0, 1^2), t_k(0.025) \to 1.960, t_k(0.005) \to 2.576 \ (k \to +\infty)$.

岩薩林 確率・統計 図 5.8(p.128)





L13-Q3

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.

1. 標準正規分布の数表から, 確率 $P(Z > z_0) = 0.025$ となる $z_0 = z(0.025)$ を求めよう.
2. 標準正規分布の数表から, 確率 $P(Z > z_0) = 1 - 0.025$ となる $z_0 = z(1 - 0.025)$ を求めよう.
3. t 分布の数表から, 確率 $P(T > w_0) = 0.025$ となる $w_0 = t_{40}(0.025)$ を求めよう.
4. t 分布の数表から, 確率 $P(T > w_0) = 1 - 0.025$ となる $w_0 = t_{40}(1 - 0.025)$ を求めよう.

ここまで来たよ

12 母集団と標本・点推定・区間推定

13 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
- 母分散の区間推定
- t 分布
- 母平均値の区間推定 (正規母集団, 母分散未知)

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

母分散既知の区間推定 確率統計☆演習 I(2020)L12 では, 標本平均値を標準化した.
$Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$.
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
$T = \frac{\overline{X}_{(n)}-\mu}{\sqrt{S^2/n}} = \frac{\frac{\overline{X}_{(n)}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$ 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

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

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 確率統計☆演習 I(2020)L12 と比べて,

$\sigma^2 \rightsquigarrow S^2$

$z(\frac{\alpha}{2}) \rightsquigarrow t_{n-1}(\frac{\alpha}{2})$

L13-Q4

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.
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)