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 L11(2025-06-30 Mon)

最終更新: Time-stamp: "2025-07-06 Sun 14:36 JST hig"

今日の目標

- カイ二乗分布を説明できる 久保川 統計学入門 §8.1
- 母分散を区間推定できる 久保川 統計学入門 §8.2
- t 分布を説明できる 久保川 統計学入門 §8.1

L10-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. $\Phi^{-1}(1-\frac{0.05}{2}) = \Phi^{-1}(0.975) = \texttt{.ppf}(0.975) = 1.96$.

Quiz 解答: 区間推定の性質

1

L10-Q2

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, $\Phi^{-1}(1-\frac{0.05}{2})=z_{0.05/2}=1.96$ 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」こののりで判定を繰り返すと，20 回に 1 回に間違えるということ）

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

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

$$0.53 < p < 0.87$$

命題 (復習：母平均値の近似的信頼区間（大標本） 久保川 統計学入門 公式 8.10 確率統計 I(2025)L10)

母平均値 μ, 母分散 σ^2 のなにかの分布にしたがう母集団の, 母平均値 μ の 信頼係数 $1-\alpha$ の近似的信頼区間 ($(1-\alpha)$ 信頼区間) は, n が大きいとき, $\overline{X}$ を標本平均値, V^2 を不偏標本分散として,

$$\overline{X} - z_{\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{V^2/n}.$$

長所

- どんな母集団（分布）でも可

短所

- 近似 1 標本サイズ N が大きい（大標本）ことを使った，中心極限定理に頼る近似
- 近似 2 母分散 σ^2 のかわりに不偏標本分散 V^2 を使う近似

ここまで来たよ

10 区間推定

11 母分散・母平均値の区間推定
- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

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?

	の推定値 (点推定)	の信頼区間 (区間推定)
母平均値 μ	標本平均値 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	左右対称 $\overline{X} - \boxed{\text{正}}\sqrt{\quad} < \mu < \overline{X} - \boxed{\text{負}}\sqrt{\quad}$
母分散 σ^2	不偏標本分散 $V^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots]$	左右対称でない $V^2 \times \boxed{\text{小}} < \sigma^2 < V^2 \times \boxed{\text{大}}$

母集団が正規分布にしたがうとき

- 標本平均値の分布 (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 $N(0, 1^2)$
- 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

カイ二乗分布 久保川 統計学入門 §8.1

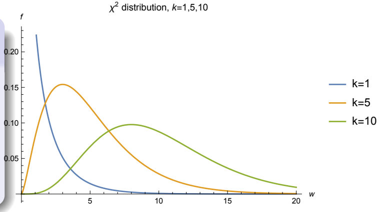
定義 (カイ二乗分布 久保川 統計学入門 定義 8.1(p.170))

$Z_1, \dots, Z_k$, i.i.d. $\sim N(0, 1^2)$ 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

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



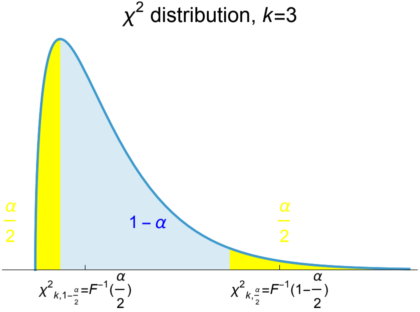
久保川 統計学入門 公式 8.2 $W_k \sim \chi_k^2$ に対して,
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{Var}(W_k) = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

```
rvx2=stats.chi2(df=k) # k は自由度
```

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

久保川 統計学入門 数値表 p.292 久保川 統計学入門 数値表 p.290 と p.292 はフォーマットが違う



$\chi^2_{k,\alpha/2}$ の定義

$\alpha/2 = P(W > \chi^2_{k,\alpha/2})$.

$\chi^2_{k,\alpha/2} = F^{-1}_{\mathrm{w}k}(1-\alpha/2)$.

分布の再生性

命題 (カイ二乗分布の再生性 久保川 統計学入門 公式 8.2)

$X \sim \chi_n, Y \sim \chi_m$, X, Y 独立のとき，$X + Y \sim \chi_{n+m}$

定義 (再生性)

一般に，1 パラメタ θ を持つ分布の族 $D(\theta)$ があり，
$X \sim D(\theta_1), Y \sim D(\theta_2), Y \sim \chi_m$, X, Y 独立のとき，$X + Y \sim D(\theta_3)$ となるとき，分布の族 $D(\theta)$ は再生性 (reproductive property) を持つという．

例 (二項分布の再生性 久保川 統計学入門 公式 6.21)

$X \sim Bin(n, p), Y \sim Bin(m, p)$, X, Y 独立のとき，
$X + Y \sim Bin(n + m, p)$.

証明：式でも示せるが，コイン投げの意味的には明らか．

L11-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 $F_{\mathrm{W}}(w)$ ，逆関数を $F_{\mathrm{W}}^{-1}(p)$ とする．
次の量を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．

1. $P(W < 0.1)$
2. $P(W > 0.9)$
3. $P(W < w) = 0.1$ となる w
4. $P(W > w) = 0.1$ となる w

不偏標本分散のしたがう分布 久保川 統計学入門 §8.2.2

不偏標本分散のしたがう分布

正規分布 $N(\mu, \sigma^2)$ からサイズ n の標本 $X_1, X_2, \dots, X_n$, i.i.d. $\sim N(\mu, \sigma^2)$ を取り出すとき, 不偏標本分散

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

から定めた

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

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

$X_1, \ldots, X_n$, i.i.d. $\sim N(\mu, \sigma^2)$ 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_n^2 にしたがう.
μ を $\overline{X}$ に替えた,

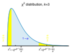
$$W=\left(\frac{X_1-\overline{X}}{\sigma}\right)^2+\cdots+\left(\frac{X_n-\overline{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ_{n-1}^2 にしたがう.

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

母分散の区間推定

$$P\left(F^{-1}_{\mathrm{w}n-1}(\alpha/2) < (n-1)\frac{V^2}{\sigma^2} < F^{-1}_{\mathrm{w}n-1}(1-\alpha/2)\right) = 1-\alpha$$



久保川 統計学入門 図 8.6

不等式を σ^2 について解いて次の結果を得る.

正規母集団の母分散の信頼区間 久保川 統計学入門 公式 8.8

$X \sim N(\mu, \sigma^2)$ のサイズ n の標本の不偏標本分散が V^2 のとき, 母分散 σ^2 の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{n-1}{F^{-1}_{\mathrm{w}n-1}(1-\alpha/2)} \times V^2 < \sigma^2 < \frac{n-1}{F^{-1}_{\mathrm{w}n-1}(\alpha/2)} \times V^2$$
$$\frac{n-1}{\chi^2_{n-1,\alpha/2}} \times V^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\alpha/2}} \times V^2$$

n 大という条件なし，近似なしだけど，X が正規分布にしたがう（正規母集団 久保川 統計学入門 §8.1）という超強い条件下.

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

L11-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 σ^2 を信頼係数 $1-\alpha=0.95$ で区間推定しよう.

久保川 統計学入門 例題 8.9

ここまで来たよ

10 区間推定

11 母分散・母平均値の区間推定

- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

標本平均値のしたがう分布 (正規母集団, 母分散既知)

久保川 統計学入門 §8.2

$N(\mu, \sigma^2)$ にしたがう母集団 (正規母集団) の, サイズ n の標本を何回も取り出して, 毎回, 標本平均値 $\overline{X}$ を計算する.

標本平均値 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ は確率変数

$U_n = \overline{X} \sim N(\mu, \sigma^2/n)$. $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0, 1^2)$.

命題 (正規分布の再生性 久保川 統計学入門 公式 6.21)

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, X, Y 独立のとき,
$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$n \to +\infty$ しなくても最初から正規分布

超強い仮定として，正規母集団を仮定すれば，$\boxed{\text{近似 1}}$を取り除ける!

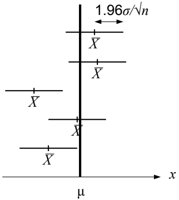
$\boxed{\text{近似 2}}$ (σ^2 を V^2 で置き換えちゃう) は使わずに書いてみる

命題 (母平均値 (正規母集団, 母分散既知) の信頼区間 久保川 統計学入門 公式 8.6(1))

$N(\mu, \sigma^2)$ にしたがう母集団の, σ^2 がわかっているとき, サイズ n の標本から区間推定すると, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間 ($(1-\alpha)$ 信頼区間) は, $\overline{X}$ を標本平均値として,

$$\overline{X} - z_{\alpha/2} \times \sqrt{\sigma^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{\sigma^2/n}.$$

何回も標本抽出して何個も信頼区間を求めたとき, 信頼区間が μ を含む確率は, 信頼係数 $1-\alpha$. 推定が外れる確率 α.



切りがいい α の z_α は $z_{\frac{0.05}{2}} = 1.96, z_{\frac{0.01}{2}} = 2.58.$

ここまで来たよ

10 区間推定

11 母分散・母平均値の区間推定

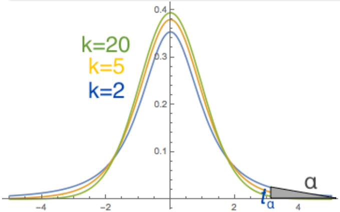
- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

t 分布

あとは，近似 2 をなんとかしたい．

定義 (t 分布 久保川 統計学入門 定義 8.3**)**

$Z \sim N(0,1^2)$, $W \sim \chi^2_k$, Z と W が独立,
のとき連続型確率変数 $T = \frac{Z}{\sqrt{W/k}}$ 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, $N(0,1^2)$ より低く広い.
自由度 $k \to +\infty$ で $N(0,1^2)$ に一致する.

t 分布の確率密度関数と累積分布関数 久保川 統計学入門 p.171

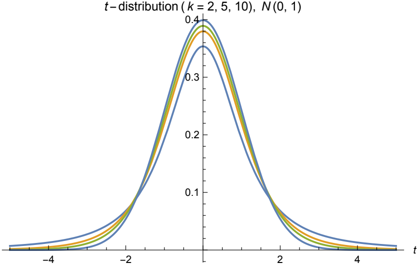
Python scipy.stats.t(df=k)

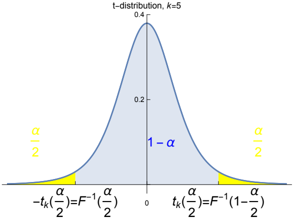
t 分布の確率密度関数は偶関数なので, $t_{k,1-\alpha} = -t_{k,\alpha}$.

久保川 統計学入門 数値表 p.291 上側確率 $\alpha = P(T > t_{k,\alpha})$ となる $t_{k,\alpha}$ の値.

久保川 統計学入門 数値 p.291 と p.290 はフォーマットが違う

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





L11-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 $Z \sim N(0, 1^2)$ 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 F_{t}^{-1} を使って答え, さらに数値にしよう.

1. 確率 $P(Z > z) = 0.025$ となる z を求めよう.
2. 確率 $P(Z > z) = 1 - 0.025$ となる z を求めよう.
3. 確率 $P(T > w) = 0.025$ となる w を求めよう.
4. 確率 $P(T > w) = 1 - 0.025$ 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)

母分散既知の区間推定では, 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 V^2 (確率変数) を使った, 標準化もどき

$T = \frac{\overline{X}-\mu}{\sqrt{V^2/n}} = \frac{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)V^2}{\sigma^2}/(n-1)}}$ でやっちゃいたい. 分子 $\sim N(0,1^2)$, 分母
$= \sqrt{W/(n-1)}, W \sim \chi^2_{n-1}$，実は W と X は独立，より, $T \sim \mathrm{t}_{n-1}$.

命題 (T の分布 久保川 統計学入門 公式 8.4)

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

$$T = \frac{\overline{X}-\mu}{\sqrt{V^2/n}}$$

は, 自由度 $n-1$ の Student の t 分布にしたがう.

母平均値の信頼区間 (母分散未知) 久保川 統計学入門 公式 8.6(2)

(母分散未知の) 正規分布 $N(\mu, ?^2)$ 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

$$\overline{X} - t_{n-1,\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + t_{n-1,\alpha/2} \times \sqrt{V^2/n}.$$

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

母分散既知 確率統計 I(2025)L10 と比べて,

$\sigma^2 \rightsquigarrow V^2$

$z_{\alpha/2} \rightsquigarrow t_{n-1,\alpha/2}$

L11-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. 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.