

区間推定・母比率の区間推定

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(2024-07-01 Mon)

最終更新: Time-stamp: "2024-07-01 Mon 12:52 JST hig"

今日の目標

- 母比率を区間推定できる 岩薩林 確率・統計 §7.3
- カイ二乗分布を説明できる 岩薩林 確率・統計 p.123
- 母分散を区間推定できる 岩薩林 確率・統計 §7.2



L10-Q1

Quiz 解答: 確率変数としての標本平均値の分布

- ① $E[\bar{X}] = \frac{1}{4} \cdot 4E[X_i] = 10.$
- ② $V[\bar{X}] = \frac{1}{4^2} \cdot 4V[X_i] = 6^2/4.$
- ③ $Z = \frac{\bar{X}-10}{6/\sqrt{4}}$

L10-Q2

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 $\bar{X} = \frac{1}{6}(117 + \dots + 112) = 111\text{g}$ 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 $\overline{X^2} = \frac{1}{6}(117^2 \dots + 112^2) = 37078/3\text{g}^2$ なので, 母期待値は $37078/3\text{g}^2$ と推定できる.
- ③ 不偏標本分散 (の実現値) は, $S^2 = \frac{1}{6-1}[(117-111)^2 + \dots + (112-111)^2] = 46\text{g}^2$ なので, 母分散は 46g^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.

ここまで来たよ

10 母集団と標本・母平均値/母分散/母比率の点推定

11 区間推定・母比率の区間推定

- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 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

点推定 対 区間推定

点推定 岩薩林 確率・統計 §6.1

真の母平均値はわからないが、標本平均値を使って、

「母平均値を A 円と推定する」

それどのくらい正確なの？ 正確さは実は **母分散や標本サイズによる**

区間推定 岩薩林 確率・統計 §6.1

「母平均値が、 B 円以上 C 円以下である '確率' は $1 - \alpha = 0.95$ 」

推定の精度・正確さまで表現

ここで '確率' というのは不誠実. 正しい言葉遣いは, **信頼係数=信頼度**で

「母平均値の**信頼係数** $1 - \alpha = 0.95$ の**信頼区間**は B 円以上 C 円以下」

動く (確率変数である) のは母平均値 μ でなく, B, C のほう.

標本平均値のしたがう分布 (正規母集団, 母分散既知)

岩薩林 確率・統計 p.144

$N(\mu, \sigma^2)$ 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 \bar{X} を計算する.

その意味で, 標本平均値 $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ は確率変数

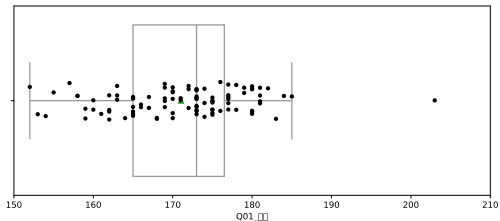
先週チーム課題で計算したのは, 1 標本で得られた標本平均値 (1 試行で得られた確率変数の値).

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

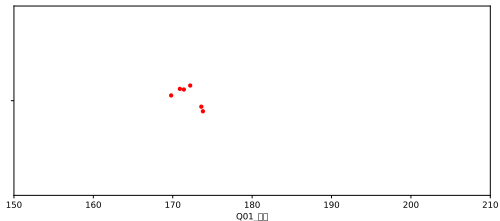
正規分布の再生性 [確率統計 II\(2023\)L??](#) [岩薩林 確率・統計 p.99](#) から. 使わなくても, $n \rightarrow +\infty$ で正しいことは中心極限定理からわかる. 正規母集団でないときも, 標本サイズ n が大きい (30 くらい) なら, 近似的に成立することが多い.

身長の標本平均値の分布と母平均値

母集団



サイズ 10 の標本 6 個の標本平均値

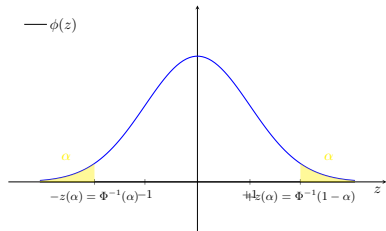


L11-Q1

Quiz(正規分布の上側確率)

$Z \sim N(0, 1^2)$ とする. Z の累積分布関数を $\Phi(z)$ とする.

- ① $P(u < Z)$ を u と Φ で表そう.
- ② $P(u < Z) = \alpha$ となる u を α と Φ で表そう. この u のことをよく $z(\alpha)$ と書く.
- ③ Z の確率密度関数が偶関数であることから,
 $P(u < Z) = P(Z < -u)$ であることを説明しよう.
- ④ $P(Z < -u) + P(u < Z) = \alpha$ となる u を α と $z()$ で表そう.
- ⑤ $\alpha = 0.05$ のとき, 上のような u を求めよう.



定義 (標準正規分布の $z(\alpha)$)

$Z \sim N(0, 1^2)$ のとき, **上側確率** $P(Z \geq z(\alpha)) = \alpha$ となる境い目を $z(\alpha)$ と定める. 岩薩林 確率・統計付表 1 下

$z(\alpha) = \Phi^{-1}(1 - \alpha)$. 偶関数だから $z(1 - \frac{\alpha}{2}) = -z(\frac{\alpha}{2})$.

標本平均値 \bar{X} が母平均値 μ から大きく外れない確率は大きい (ここでは $1 - \alpha = 1 - 0.05$) という式を書くと…

$$P\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) < \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

$$P\left(\mu + \Phi^{-1}\left(\frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n} < \bar{X} < \mu + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n}\right) = 1 - \alpha.$$

μ について不等式を解くと,

$$P\left(\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n} < \mu < \bar{X} - \Phi^{-1}\left(\frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n}\right) = 1 - \alpha.$$

$$P\left(\bar{X} - z\left(\frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n} < \mu < \bar{X} - (-z\left(\frac{\alpha}{2}\right)) \times \sqrt{\sigma^2/n}\right) = 1 - \alpha.$$

業界の習慣で, しばしば $\alpha = 0.05, 0.01$

$$\alpha = 0.05 \rightsquigarrow P\left(\bar{X} - 1.96 \times \sqrt{\sigma^2/n} < \mu < \bar{X} + 1.96 \times \sqrt{\sigma^2/n}\right) = 1 - 0.05.$$

命題 (母平均値 (正規母集団, 母分散既知) の信頼区間

岩薩林 確率・統計 定理 6.1(p.18)

$N(\mu, \sigma^2)$ 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, \bar{X} を標本平均値として,

$$\bar{X} - z\left(\frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n} < \mu < \bar{X} + z\left(\frac{\alpha}{2}\right) \times \sqrt{\sigma^2/n}.$$

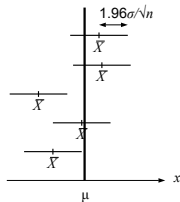
何回も標本抽出して何個も信頼区間を求めたとき, 信頼区間が μ を含む確率は, 信頼係数 $1 - \alpha$. 推定が外れる確率 α .

切りがいい α の $z(\alpha)$ は [岩薩林 確率・統計 付表 1 下 \(p.227\)](#)

$$z\left(\frac{0.05}{2}\right) = 1.96, z\left(\frac{0.01}{2}\right) = 2.58.$$

ゆるゆるな [高校 数学 B](#) では, $z\left(\frac{0.05}{2}\right) = 1.96$ の場合のみ.

$a < \mu < b$ でなく, 閉区間の記号 $[a, b]$ で.



[岩薩林 確率・統計 例題 6.4\(p. 145\)](#) [岩薩林 確率・統計 問題 3\(p.146\)](#) [岩薩林 確率・統計 第 6 章練習問題 1](#)

推定が正確であるとは 信頼区間が **自分の言葉で** であること.

Quiz(区間推定の性質)

標本からの母平均値の区間推定について, 正しいのはどれ?

- ① 母分散が大きいほど, 信頼区間は大きくなる
- ② 標本サイズが大きいほど, 信頼区間は大きくなる
- ③ 母平均値が大きいほど, 信頼区間は小さくなる
- ④ 信頼係数が大きいほど, 信頼区間は小さくなる

標本平均値が大きい \Rightarrow 信頼区間は **平行移動する**

母分散が大きい \Rightarrow 信頼区間は **大きい**

標本サイズ n が小さい 信頼区間は **小さい**

ここまで来たよ

10 母集団と標本・母平均値/母分散/母比率の点推定

11 区間推定・母比率の区間推定

- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 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

母比率の信頼区間

母比率は母平均値の一種。

近似1 サンプルサイズが大きいとき、中心極限定理により、標本平均値は近似的に正規分布 $N(p, p(1-p))$ に従う。

$$P\left(p - z\left(\frac{\alpha}{2}\right) \times \sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z\left(\frac{\alpha}{2}\right) \times \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha.$$

近似2 平方根の中の $p(1-p)$ を $\hat{p}(1-\hat{p})$ と近似して p について解くと、

母比率の信頼区間 (母分散未知) 岩薩林 確率・統計 §7.3

X のサイズ n の標本で、標本比率 $\hat{p} = k/n$ のとき、母比率の **信頼係数** $1 - \alpha$ の **信頼区間** ($(1 - \alpha)$ **信頼区間**) は、

$$\hat{p} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + \Phi^{-1}\left(\frac{\alpha}{2}\right) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

$$\hat{p} - z\left(\frac{\alpha}{2}\right) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z\left(\frac{\alpha}{2}\right) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

L11-Q2

Quiz(母比率の区間推定)

選挙で出口調査をしたところ、50人中35人がA候補に投票したと答えた。母集団を投票した人2500人とする。そのうちA候補に投票した人の母比率(得票率)を考える。

- ① A候補の得票率を、(点)推定しよう
- ② A候補の得票率を、信頼係数 $1 - \alpha = 0.95$ で区間推定しよう。
- ③ A候補の得票率を、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 7.6(p.170)

岩薩林 確率・統計 問題 7(p.171)

岩薩林 確率・統計 第7章練習問題 2(2)

注: 下限, 上限が $0, 1$ を越えるときは, $0, 1$ に直してしまってもいい。

ここまで来たよ

10 母集団と標本・母平均値/母分散/母比率の点推定

11 区間推定・母比率の区間推定

- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 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

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? $\sqrt{\text{母分散}}$ $\xleftarrow{\text{点推定}}$ $\sqrt{\text{不偏標本分散}}$

- 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 $\bar{X} = \frac{1}{n}[X_1 + \dots]$	$\bar{X} - \boxed{\text{正}}\sqrt{\quad} < \mu < \bar{X} - \boxed{\text{負}}\sqrt{\quad}$
母分散 σ^2	不偏標本分散 $S^2 = \frac{1}{n-1}[(X_1 - \bar{X})^2 + \dots]$	$S^2 \times \boxed{\text{小}} < \sigma^2 < S^2 \times \boxed{\text{大}}$

母集団が正規分布にしたがうとき

- 標本平均値の分布 (**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

岩薩林 確率・統計 p.123

カイ二乗分布

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

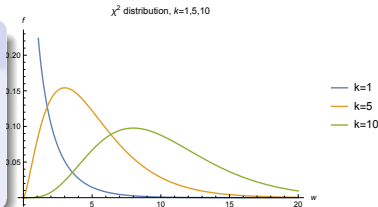
言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

χ_k^2 の確率密度関数

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

$W_k \sim \chi_k^2$ に対して,

$$E[W_k] = E[Z_1^2 + \dots + Z_k^2] = k, V[W_k] = 2k, E[(W_k)^\ell] = \text{簡単じゃない.}$$



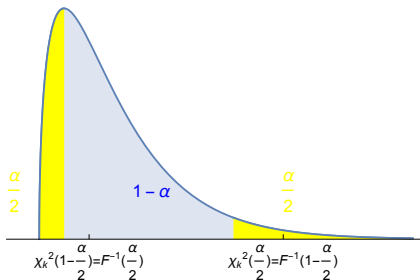
カイ二乗分布の確率密度関数と累積分布関数

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

上側確率 $\alpha = P(W > \chi_k^2(\alpha))$ となる境い目 $\chi_k^2(\alpha) = F^{-1}(1 - \alpha)$.
 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

χ^2 distribution, $k=3$



$\chi_k^2(\alpha)$ の定義

岩薩林 確率・統計 例題 5.7

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

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

L11-Q3

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 $F(w)$ とする．

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

ここまで来たよ

10 母集団と標本・母平均値/母分散/母比率の点推定

11 区間推定・母比率の区間推定

- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 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

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

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

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_{n-1}^2 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 $\frac{W}{n-1}$. ($W \sim \chi_{n-1}^2$)

証明じゃないけど説明

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

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

は自由度 n のカイ二乗分布 χ_n^2 にしたがう.

μ を \bar{X} に替えた,

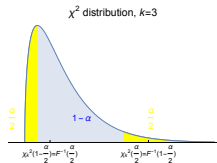
$$W = \left(\frac{X_1 - \bar{X}}{\sigma}\right)^2 + \dots + \left(\frac{X_n - \bar{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ_{n-1}^2 にしたがう.

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

母分散の区間推定

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



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 $(n-1)/W \simeq 1$, $W \sim \chi_{n-1}^2$.

L11-Q4

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 72g^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)