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 L10(2025-06-23 Mon)

最終更新: Time-stamp: "2025-07-21 Mon 18:01 JST hig"

今日の目標

- 母平均値を近似的に区間推定できる 久保川 統計学入門 §8.3
- 母比率を区間推定できる 久保川 統計学入門 §8.3
- カイ二乗分布を説明できる 久保川 統計学入門 §8.1

L09-Q1

Quiz 解答: 確率変数としての標本平均値の分布

1. $\mathrm{E}[\overline{X}] = \frac{1}{4} \cdot 4\mathrm{E}[X_i] = 10.$
2. $\mathrm{Var}(\overline{X}) = \frac{1}{4^2} \cdot 4\mathrm{Var}(X_i) = 6^2/4.$
3. $Z = \frac{\overline{X}-10}{6/\sqrt{4}}$

L09-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 $\overline{X} = \frac{1}{6}(118 + \cdots + 113) = 111.5\mathrm{g}$ なので, 母平均値を $111.5\mathrm{g}$ と推定する.
2. 標本期待値 (の実現値) は $\overline{X^2} = \frac{1}{6}(118^2 \cdots + 113^2) = 12469.5\mathrm{g}^2$ なので, 母期待値を $12469.5\mathrm{g}^2$ と推定する.

3. 不偏標本分散 (の実現値) は,
$S^2 = \frac{1}{6-1}[(117-111.5)^2 + \cdots + (112-111.5)^2] = 37.25\mathrm{g}^2$ なので,
母分散を $37.25\mathrm{g}^2$ と推定する.
4. 標本比率 (の実現値) は, $\hat{p} = \frac{1}{6}[1+1+1+1+0+1] = 0.83$ なので,
母比率を 0.83 と推定する.

ここまで来たよ

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

点推定 対 区間推定

点推定 久保川 統計学入門 §7.2 「1 個の推定値を求めること」
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 久保川 統計学入門 §8.2 「信頼係数に応じた信頼区間を求めること」

「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

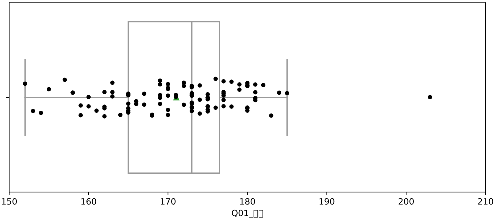
ここで '確率' というのは不誠実. 正しい言葉遣いは, 信頼係数=信頼度で

「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」

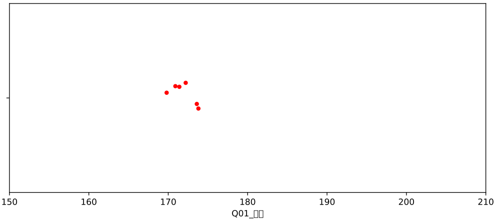
動く (確率変数である) のは母平均値 μ でなく, B, C のほう.

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



大標本からの母平均値特に母比率の近似的信頼区間

未知 母集団の母平均値 μ, 母分散 σ^2

$\downarrow$

既知 標本 $X_1, \ldots, X_n \rightsquigarrow$ 標本平均値 $\overline{X}$, 不偏標本分散 V^2.

$\boxed{\text{近似 1}}$ サンプルサイズ n が大きいとき，中心極限定理により，近似的に.

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

$$\text{標準化} \quad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1^2)$$

$\boxed{\text{近似 2}}$ σ のかわりに $\sqrt{V^2}$ を使っても近似的になりたつ.

$$\frac{\overline{X} - \mu}{\sqrt{V^2/n}} \sim N(0, 1^2)$$

近似=誤差がある．後から，一定の条件下で，これらの近似なしに，より正確に評価できるハッピーなケースを扱う

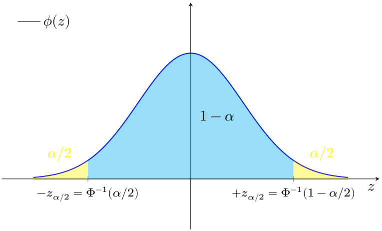
標準正規分布の計算の準備 I

L10-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_α と書く．
3. Z の確率密度関数 $\phi(z)$ 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. 「母平均値 0 から遠い確率」 $P(|Z-0| > u) = P(Z < -u) + P(u < Z) = \alpha$ となる u を α と $z_\bullet$ 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_α 久保川 統計学入門 公式 5.9,(8.5), 図 8.4)

$Z \sim N(0,1^2)$ のとき，上側確率 $P(Z \geq z_{\alpha/2}) = \alpha/2$ となる境い目を $z_{\alpha/2}$ と定める．
Φ を累積分布関数，Φ^{-1} をその逆関数とするとき，
$z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$．偶関数だから $z_{1-\alpha/2} = -z_{\alpha/2}$．
$P(-z_{\alpha/2} < Z < +z_{\alpha/2}) = \alpha$．

標本平均値と母平均値の差 $\overline{X}-\mu$ が 0 に近い確率は，正規分布が真ん中で盛り上がっているために大きい (ここでは $1-\alpha$) という式を書くと…

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

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

μ について不等式を解くと，

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

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

$$\boxed{\text{近似 2}} P\left(\overline{X}-z_{\alpha/2}\times\sqrt{V^2/n} < \mu < \overline{X}+z_{\alpha/2}\times\sqrt{V^2/n}\right) = 1-\alpha.$$

業界の習慣で，しばしば $\alpha = 0.05, 0.01$

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

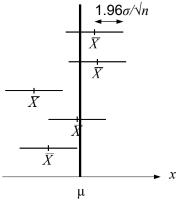
命題 (母平均値の近似的信頼区間 久保川 統計学入門 公式 8.10)

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,

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

$a < \mu < b$ でなく, 閉区間で $\mu \in [a, b]$ とかくこともある.「信頼区間は $[a, b]$」のように答えることもある.

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 久保川 統計学入門 公式 5.9,(8.6)

$z_{0.05/2} = 1.96$, $z_{0.01/2} = 2.58$.

高校 数学 B では, $z_{0.05/2} = 1.96$ の場合のみ.

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

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

母比率の信頼区間 久保川 統計学入門 例題 8.11

母比率 p は，ベルヌーイ分布 $Ber(p)$ の母平均値 p. 母分散は $p(1-p)$. 標本比率は標本平均値 $\overline{X} = \hat{p} = k/n$.

$\boxed{\text{近似 1}}$ サンプルサイズが大きいとき，中心極限定理により，標本比率は近似的に正規分布 $N(p, p(1-p)/n)$ に従う．

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

$\boxed{\text{近似 2'}}$ 母分散 σ^2 を，不偏標本分散 V^2 のかわりに $\hat{p}(1-\hat{p})$ で近似する

母比率の近似的信頼区間 久保川 統計学入門 例題 8.11

$X \sim Ber(p)$ 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，

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

$$\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

L10-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.

注: 下限, 上限が 0,1 を越えるときは, 近似 1,2 が正確でないことを意味しているが, 実用的には 0,1 に直してしまっていい.

ここまで来たよ

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

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? $\sqrt{母分散} \overset{点推定}{\leftarrow} \sqrt{不偏標本分散}$

- 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 $\overline{X} - \boxed{正}\sqrt{\quad} < \mu < \overline{X} - \boxed{負}\sqrt{\quad}$
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 $V^2 \times \boxed{小} < \sigma^2 < V^2 \times \boxed{大}$

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

確率統計 II(2023)L??

カイ二乗分布 久保川 統計学入門 §8.1

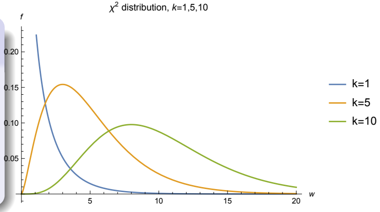
定義 (カイ二乗分布 久保川 統計学入門 定義 8.1(p.170))

$Z_1, \dots, Z_k$, 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,

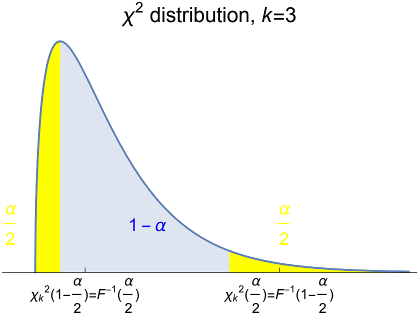
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{Var}(W_k) = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

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 $\chi^2_{k,\alpha/2} = F_W^{-1}(1-\alpha/2)$.
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
$\alpha/2 = P(W > \chi^2_{k,\alpha/2})$.
$\chi^2_{k,\alpha/2} = F_W^{-1}(1-\alpha/2)$.

L10-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 $F_{\mathrm{W}}(w)$ ，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

$X_i, \dots, X_n$, 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^2_n にしたがう．
μ を $\overline{X}$ に替えた，

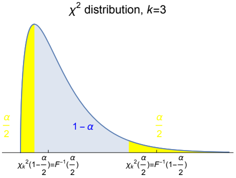
$$W=\left(\frac{X_1-\overline{X}}{\sigma}\right)^2+\cdots+\left(\frac{X_n-\overline{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう．

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

母分散の区間推定

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



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

$$\frac{n-1}{F_W^{-1}(1-\alpha/2)} \times V^2 < \sigma^2 < \frac{n-1}{F_W^{-1}(\alpha/2)} \times V^2$$

$$\frac{n-1}{\chi^2_{n-1,\alpha/2}} \times V^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\alpha/2}} \times V^2$$

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

L10-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 $80\mathrm{g}$, 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.