

# 母比率の点推定と区間推定

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L11(2014-12-25 Thu)

## 今日の目標

- 標本から母集団の比率を点推定できる
- 標本から母集団の比率を区間推定できる



<http://hig3.net>

## L11-S2

## Quiz 解答:母平均値の区間推定 (母分散未知)

- ① 重さの標本平均値は  $m = 50\text{g}$ . 不偏標本分散は  $s^2 = \frac{1}{4-1} \cdot 14$ . 自由度  $k = n - 1 = 3$  の t 分布表を参照して, 95%信頼区間は

$$50 - 3.182 \times \sqrt{\frac{1}{3} \frac{14}{3}} < \mu < 50 + 3.182 \times \sqrt{\frac{1}{3} \frac{14}{3}}.$$

- ② 同様に,

$$50 - 5.841 \times \sqrt{\frac{1}{3} \frac{14}{3}} < \mu < 50 + 5.841 \times \sqrt{\frac{1}{3} \frac{14}{3}}.$$

## L11-S3

## Quiz 解答:母平均値の区間推定 (母分散未知, 大標本)

- ① 大標本なので, 正規分布で考えてよい. 95%信頼区間は

$$51 - 1.96 \times \sqrt{\frac{4}{400}} < \mu < 51 + 1.96 \times \sqrt{\frac{4}{400}}.$$

- ② 同様に,

$$51 - 2.58 \times \sqrt{\frac{4}{400}} < \mu < 51 + 2.58 \times \sqrt{\frac{4}{400}}.$$

# ここまで来たよ

## 1 略解:中心極限定理と区間推定

## 2 母比率の点推定と区間推定

- 比率とは
- 母比率の点推定
- 母比率の区間推定

## 質的変数

これまで扱ってた変数 or データ  $X$

量的変数 例  $X \in \mathbb{R}, X \in \mathbb{Z}$ .

今日ちょっとのぞき見るもの:

質的変数 = カテゴリカル変数

例  $Y \in \{A \text{ 型}, B \text{ 型}, AB \text{ 型}, O \text{ 型}\}$ . 順序なし

例  $Y \in \{\text{優}, \text{良}, \text{可}, \text{不可}\}$ . 順序あり

例  $Y \in \{\dots \text{である}, \text{それ以外}\}$ . 2 値  $\rightarrow \dots$  の比率

例  $Y \in \{\text{身長 } 165\text{cm 未満}, \text{身長 } 165\text{cm 以上}\}$ . 2 値  $\rightarrow$  身長 165cm 未満の人の比率

例  $Y \in \{\text{サイコロの目が } 1, \text{サイコロの目が } 2 \text{ 以上}\}$ . 2 値  $\rightarrow$  サイコロの目 1 の出る比率

## 比率=ratio

2 値の場合, 次が考えられる.

### 母比率

$$\dots\text{の母比率} = \frac{\text{「}\dots\text{」であるデータの個数}}{\text{母集団サイズ}}$$

例  $Y \in \{ \text{身長 165cm 未満, 身長 165cm 以上} \}$ .

$$\text{母比率 } p = \frac{\text{身長 165cm 未満の人の数}}{\text{母集団サイズ}}.$$

例  $Y \in \{ \text{サイコロの目が 1, サイコロの目が 1 以上} \}$ .

$$\text{母比率 } p = \text{サイコロの目の 1 がでる確率}$$

確率と言いきうになるけど, 推定や検定の文脈では比率という

## やりたいこと:母比率の推定

- クラスの中で、血液型 A 型の人々の比率は？  $n$  人に質問しただけで推定したい。
- 候補者 A の得票率は何%？  $n$  人に質問しただけで推定したい。
- 工場から出荷する製品のうち、何% が不良品？  $n$  個だけ抜き出して調査したい。
- このコインの表が出る確率は？  $n$  回投げるだけで推定したい。

## ここまで来たよ

### 1 略解:中心極限定理と区間推定

### 2 母比率の点推定と区間推定

- 比率とは
- 母比率の点推定
- 母比率の区間推定



## 母比率の点推定

### 母比率の点推定

サンプルのデータ  $n$  個中  $k$  個が…に該当するとき,

$$\text{'標本比率'} \hat{p} = \frac{k}{n}$$

が…の母比率  $p$  のよい推定値になっている。

## 量的変数に置きかえ

$X$	$Y$	確率
1	A型である	$p$
0	A型でない	$1 - p$

$X$  はベルヌーイ分布に従う

$X$  の母平均値  $E[X] = 1 \times p + 0 \times (1 - p) = p$ .

$X$  の母分散  $V[X] = (1 - p)^2 \times p + (0 - p)^2 \times (1 - p) = p(1 - p)$ .

母比率を推定したい = 母平均値を推定したい

母平均値の推定

サイズ  $n$  の標本中  $k$  個が該当するとき,

母平均値  $E[X]$  の推定値 = 標本平均値  $\bar{X}$

$$= \frac{1}{n} \left[ \underbrace{1 + \cdots + 1}_k + \underbrace{0 + \cdots + 0}_{n-k} \right]$$

$$= \frac{k}{n} = \hat{p}.$$

## L11-Q1

## 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。母比率、すなわち有権者全体でのA候補の得票率を考える。

- 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95%で区間推定しよう。
- 3 A候補の得票率を、信頼係数99%で区間推定しよう。

# ここまで来たよ

## 1 略解:中心極限定理と区間推定

## 2 母比率の点推定と区間推定

- 比率とは
- 母比率の点推定
- 母比率の区間推定

## 母比率の区間推定

### 母比率の区間推定

母比率の信頼係数 95% の信頼区間は

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p})}$$

母比率の信頼係数 99% の信頼区間は

$$\hat{p} - 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p})} < p < \hat{p} + 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p})}$$

注: 上の計算で下限, 上限が 0,1 を越えるとき (おかしいことではない) は, 0,1 に直してしまってもいいでしょう.

## 導出

標本平均値  $\frac{K}{n}$  は,  $n$  が大きいとき, 中心極限定理から, 母平均値  $p$ , 母分散  $\frac{1}{n}p(1-p)$  の正規分布にしたがう. よって,

$$p - 1.96 \times \sqrt{\frac{1}{n}p(1-p)} < \hat{p} < p + 1.96 \times \sqrt{\frac{1}{n}p(1-p)}$$

これを,  $p$  に関する条件に書き替えたい

**極小標本** 中心極限定理はやめて, 2項分布と思って扱え

**小標本**  $p$  についての2次方程式を解け

**大標本**  $p$  と  $\hat{p}$  はかなり近いはず. 最左右辺の  $p(1-p)$  は  $\hat{p}(1-\hat{p})$  に置きかえちゃっていい.

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}$$

## 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。母比率、すなわち有権者全体でのA候補の得票率を考える。

- 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 95% で区間推定しよう。
- 3 A 候補の得票率を、信頼係数 99% で区間推定しよう。

## 連絡

- 配布資料は 1-503 向かいの引出, <http://hig3.net> で再配布しています.
- Quiz の略解は授業終了後に <http://hig3.net> で配布しています.
- 2014-11-17 から チューターは月火水木昼 (1-614).
- 2015-01-09 金 09:00 までに予習問題.
- 2015-01-09 金 2 の非参照 Quiz は L10,L11 両方を出題します.
- すみません冬休み中にプチテスト返却します.
- いつか補講もう 1 回