

統計学 01

早稲田大学政治経済学部

第13回

西郷 浩

本日の目標

■ 区間推定

- 母集団比率(成功確率)についての区間推定
 - 教科書では、二項母集団に関する推定(pp. 229-230)と表現されている。
- 母分散についての区間推定
 - 使用頻度は低い。

成功確率の推定(1)

- ある(いかさまの?)サイコロを $n = 100$ 回振ったところ、偶数の目が40回しか出なかった。このサイコロの出目が偶数である確率はどの程度か？
- ポイントの整理
 - 成功確率(出目が偶数である確率) p の独立なベルヌーイ試行が $n(100)$ 回おこなわれる。
 - $n(100)$ 回の試行のうち偶数の目が出る回数 X は、成功確率 p のベルヌーイ確率変数 X_i の和になる。
 - $X = \sum_i X_i$

成功確率の推定(2)

- 標本からえられる(経験的な)成功比率: \hat{p}

$$\hat{p} = \frac{\text{偶数が出た回数}}{\text{試行回数}} = \frac{X}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

- つまり、標本からえられる成功比率は、一種の標本平均である。
 - 試行回数(サンプルサイズ n)が大きければ、中心極限定理によって、 \hat{p} の標本分布は正規分布で近似できる。

成功確率の推定(3)

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p$$

$$V(\hat{p}) = V\left(\frac{X}{n}\right) = \frac{V(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

中心極限定理によって、

$$\frac{\hat{p} - E(\hat{p})}{\sqrt{V(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0, 1)$$

成功確率の推定(4)

したがって、

$$P\left(-1.96 \leq \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq 1.96\right) \approx 0.95$$

カッコ内を p について解く（二次不等式）ことも可。

しかし、通常は以下の簡便法を使う。カッコ内を変形すると、

$$\hat{p} - 1.96\sqrt{p(1-p)/n} \leq p \leq \hat{p} + 1.96\sqrt{p(1-p)/n}$$

試行回数が多ければ p と \hat{p} とのずれは小さく、 $p(1-p)$ と $\hat{p}(1-\hat{p})$ とのずれはさらに小さい。

成功確率の推定(5)

この置き換えによる影響は小さいとみなして、

$$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 0.95$$

したがって：

成功確率 p に関する信頼係数0.95の信頼区間

$$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

成功確率の推定(6)

先の問題について、

偶数が出る確率 p に関する

信頼係数0.95の信頼区間

$$\begin{aligned}\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} &= 0.4 \pm 1.96\sqrt{0.4 \times 0.6/100} \\ &= 0.4 \pm 0.096 = 0.304, 0.496\end{aligned}$$

「公正なサイコロよりも偶数の目が出にくい
ようだ」といえる。

母集団における比率の推定(1)

■ 母集団における比率の推定

□ 例:

世論調査にもとづく有権者全体の政党支持率の推定

□ 結論からいえば

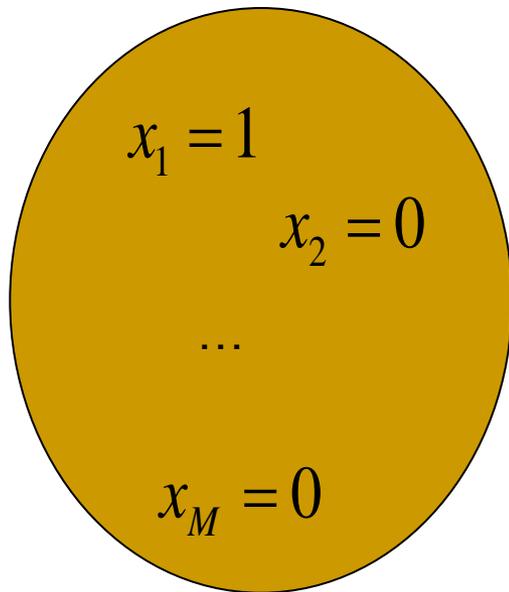
- 標本を「(近似的に)独立なベルヌーイ試行の結果である」とみなして、成功確率 p と同じように推定する。
- 母集団のサイズが極端に小さくない限り、上記の近似の悪影響は小さい。

母集団における比率の推定(2)

- 有権者全体(母集団、 M 人)から無作為抽出で $n=3000$ 人を選び、政党Aの支持者数を調べたところ、1000人であった。有権者全体における政党Aの支持率はどのくらいか。
- 記号・考え方
 - 母集団に所属する有権者全員について、以下のような変数 x_j を定める。
 - $x_j=1$ (有権者 j が政党A支持者のとき) or $=0$ (otherwise)
 - 標本に選ばれた有権者の x の値を確率変数 X_j と記す。

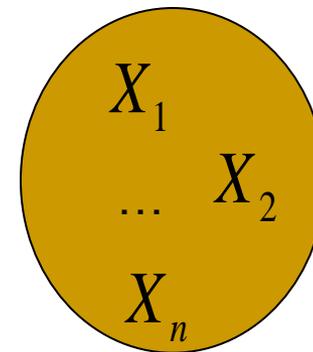
母集団における比率の推定(3)

母集団(有権者全体)



標本(調査される人)

標本抽出
(くじ引き)



母集団
における
比率

$$p = \frac{1}{M} \sum_{j=1}^M x_j$$

標本に
おける
比率

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

母集団における比率の推定(4)

- X_i は(近似的に)成功確率 p の独立なベルヌーイ確率変数とみなせる。
 - 最初にくじ引きで M 人からひとりを無作為に引き抜く。
 - X_1 は、確率 p で1、確率 $1-p$ で0になる。
 - つぎに、残りの $M-1$ 人から一人を無作為に引き抜く。
 - X_2 は、確率 p' で1、確率 $1-p'$ で0になる。 p' は、厳密に言えば、 p と若干異なる。
 - つぎに、残りの $M-2$ 人から一人を無作為に引き抜く。
 - X_3 は、確率 p'' で1、確率 $1-p''$ で0になる。 p'' は、厳密に言えば、 p'' と若干異なる。
 - M が非常に大きければ、 $p \doteq p' \doteq p'' \doteq \dots$

母集団における比率の推定(5)

- したがって、母集団における比率を「成功確率」とみなして、成功確率に関する推定をそのまま適用できる。
- 先の例題については、
 - 信頼係数0.95の政党A支持率の信頼区間

$$\hat{p} = 1000 / 3000 = 1/3$$

$$\begin{aligned}\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} &= 1/3 \pm 1.96\sqrt{\frac{1}{3}\left(1-\frac{1}{3}\right)/3000} \\ &= 1/3 \pm 0.017 = 0.316, 0.350\end{aligned}$$

おまけ: 母分散の区間推定(1)

$X_i \sim N(\mu, \sigma^2) (i = 1, 2, \dots, n)$ かつ相互に独立のとき、

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{自由度 } n-1 \text{ のカイ二乗分布}$$

$$\text{ただし、 } s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

したがって、

$$P\left(\chi_{0.975}^2(n-1) \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{0.025}^2(n-1)\right) = 0.95$$

ただし、 $\chi_{0.975}^2(n-1) = (\chi^2(n-1)$ の上側0.975点)

おまけ: 母分散の区間推定(2)

このことから、

母分散 σ^2 についての信頼係数0.95の信頼区間

$$\left[\frac{(n-1)s^2}{\chi_{0.025}^2(n-1)}, \frac{(n-1)s^2}{\chi_{0.975}^2(n-1)} \right]$$

問題練習

- 以下の問題を解け。

- 11.7 ii) p. 231

- (ホーエル)

1974年6月11日付けの Los Angeles Times に載った記事によれば、過去5年間にオレンジ郡において自動車事故で死んだ2000人のドライバーのうち、40%は酒気帯び運転であった。この5年間でオレンジ郡の自動車事故の状況を代表していると考えて、自動車死亡事故における酒気帯び運転の比率について、信頼係数0.95の信頼区間を構成せよ。