統計学2及び演習

母比率の検定

創域理工学部
Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年5月24日

Contents

- 二項分布
  - 期待値，分散
  - 二項分布の正規分布への近似
- 母比率の検定の考え方
- 母比率の検定

手元にあるコインはいかさまコインかどうか

- 表が出る確率 π は1/2かどうか
- 実際にコインをn回投げて，確かめる実験を考える

表 表 表 表 表 表 … 表　x 回

裏 裏 裏 裏 裏 裏 … 裏　$n-x$ 回

- 表が出た割合　$p=\dfrac{x}{n}$

ベルヌーイ試行

次の3条件を満たす試行系列

- 独立性
  - 各回の試行は，他の回の試行に影響を及ぼさない
- 定常性
  - 成功の確率・失敗の確率が各試行を通じて一定
- 二値性
  - ある事象が生じるか否か（成功・失敗）に関心がある

二項分布

- 各試行

$$X_i = \begin{cases} 1 & \text{成功} \\ 0 & \text{失敗} \end{cases}$$

- 試行回数 n
- 成功回数 $X = X_1 + \cdots + X_n$
- 成功確率 $\Pr(X_i = 1) = \pi$

確率関数

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \quad x = 0, 1, \dots, n$$

$n = 1$ のときはベルヌーイ分布

二項分布の性質

- 期待値 $E[X] = n\pi$
- 分散 $V[X] = n\pi(1 - \pi)$
- 平均と分散が1つの母数$\boldsymbol{\pi}$で表現される
- $\boldsymbol{n}$ が十分に大きいときは正規分布で近似できる

$E[X]$と$V[X]$の導出（パターン1）

$$E[X_i] = 1 \times \Pr(X_i = 1) + 0 \times \Pr(X_i = 0) = \pi$$

$$\begin{aligned} V[X_i] &= E[X_i^2] - (E[X_i])^2 \\ &= 1^2 \times \Pr(X_i = 1) + 0^2 \times \Pr(X_i = 0) - \pi^2 = \pi(1 - \pi) \end{aligned}$$

独立性の条件より

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n\pi$$

$$V[X] = V[X_1 + X_2 + \cdots + X_n] = V[X_1] + \cdots + V[X_n] = n\pi(1 - \pi)$$

$E[X]$と$V[X]$の導出（パターン2）

$$E[X] = \sum_{x=0}^{n} x \binom{n}{x} \pi^{x}(1-\pi)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} \pi^{x}(1-\pi)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!\,(n-x)!} \pi^{x}(1-\pi)^{n-x}$$

$$= n\pi \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-1-(x-1))!} \pi^{x-1}(1-\pi)^{n-1-(x-1)}$$

$$y = x - 1$$

$$= n\pi \sum_{y=0}^{n-1} \frac{(n-1)!}{y!\,(n-1-y)!} \pi^{y}(1-\pi)^{n-1-y} = n\pi$$

$E[X]$と$V[X]$の導出（パターン2）

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\binom{n}{x}\pi^x(1-\pi)^{n-x} = \sum_{x=2}^{n} x(x-1)\binom{n}{x}\pi^x(1-\pi)^{n-x}$$

$$= \sum_{x=2}^{n} x(x-1)\frac{n!}{x!\,(n-x)!}\pi^x(1-\pi)^{n-x}$$

$$= n(n-1)\pi^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,(n-2-(x-2))!}\pi^{x-2}(1-\pi)^{n-2-(x-2)}$$

$y = x - 2$

$$= n(n-1)\pi^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y!\,(n-2-y)!}\pi^{y}(1-\pi)^{n-2-y} = n(n-1)\pi^2$$

$E[X]$と$V[X]$の導出（パターン2）

$$E[X^2] = E[X(X-1)] + E[X] = n(n-1)\pi^2 + n\pi = n\pi(n\pi - \pi + 1)$$

$$V[X] = E[X^2] - (E[X])^2 = n\pi(n\pi - \pi + 1) - (n\pi)^2$$

$$= n\pi(1-\pi)$$

仮説の設定

- 「表が出る確率 π は1/2かどうか」を検証するために2つの仮説を設定する
  - 帰無仮説 $H_0 : \pi = 1/2$
  - 対立仮説 $H_1 : \pi \neq 1/2$

- 表が出る回数 X
- 帰無仮説が真であると仮定すると
  - 表が出る確率

$$P(X = x) = {}_nC_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{n-x} \qquad (x = 1, \ldots, n)$$

偶然に出る可能性のある「表の回数」の範囲

- いかさまコインではない（帰無仮説 $H_0: \pi = 1/2$）と仮定

表の回数	確率
0	0.1%
1	0.98%
2	4.39%
3	11.72%
4	20.51%
5	24.51%
6	20.51%
7	11.72%
8	4.39%
9	0.98%
10	0.1%

確率の計算式 $${}_{10}C_x \left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{10-x}$$

表の回数3から7である確率は85%以上

表の回数が1以下，または9回以上である確率は2.16%

表の回数が2以下，または8回以上である確率は10.94%

母比率の統計的仮説検定の手順

- Step1
  - 帰無仮説と対立仮説を設定し，有意水準を定める
  - 慣例的には， $\alpha = 0.05$を用いることが多い
- Step2
  - 検定統計量$T = t(X_1, X_2, \ldots, X_n)$を定める
- Step3
  - 棄却域$W = \{(x_1, x_2, \ldots, x_n) \mid t(x_1, x_2, \ldots, x_n) \in R\}$を求める
- Step4
  - 検定統計量Tの実現値$t^* = t(x_1, x_2, \ldots, x_n)$
  - $t^* \in R$ならば帰無仮説を棄却， $t^* \notin R$ならば帰無仮説を採択

誤り確率の制御

- 第1種の過誤確率
  - 検定の大きさ $\sup_{\pi \in \Theta_0} P\big((X_1, X_2, \ldots, X_n) \in W\big)$
  - 検定の大きさがα $(0 \le \alpha \le 1)$より小さい検定を，有意水準(significance level) αの検定という

$$\sup_{\pi \in \Theta_0} P\big((X_1, X_2, \ldots, X_n) \in W\big) \le \alpha$$

この式を満たすようにWを定める

- 第2種の過誤確率
  - 検定の段階では制御していない

棄却域Wの設定

- 有意水準 $\alpha = 0.05$

- 10回コインを投げた結果
  - 表の回数が1以下，または9回以上である確率は2.16%
  - 表の回数が2以下，または8回以上である確率は10.94%

- 棄却域W
  - 表の回数が1以下，または9回以上

計算の大変さ

- 10回のコイン投げであれば，計算は容易
- 試行回数$\boldsymbol{n}$が増えた場合はどうだろうか？

- 棄却域Wの設定

$$P(X \geq w) = \sum_{x=w}^{n} {}_{n}C_{x} \left(\frac{1}{2}\right)^{x} \left(1 - \frac{1}{2}\right)^{n-x} \leq \frac{\alpha}{2}$$

$$P(X \leq w) = \sum_{x=1}^{w} {}_{n}C_{x} \left(\frac{1}{2}\right)^{x} \left(1 - \frac{1}{2}\right)^{n-x} \leq \frac{\alpha}{2}$$

計算が簡便になる方法を検討する必要がある

二項分布の正規分布への近似

中心極限定理

- 母集団分布に関わらず，和 $X_1 + \cdots + X_n$ の確率分布は，n が十分大きいときは正規分布に近似する

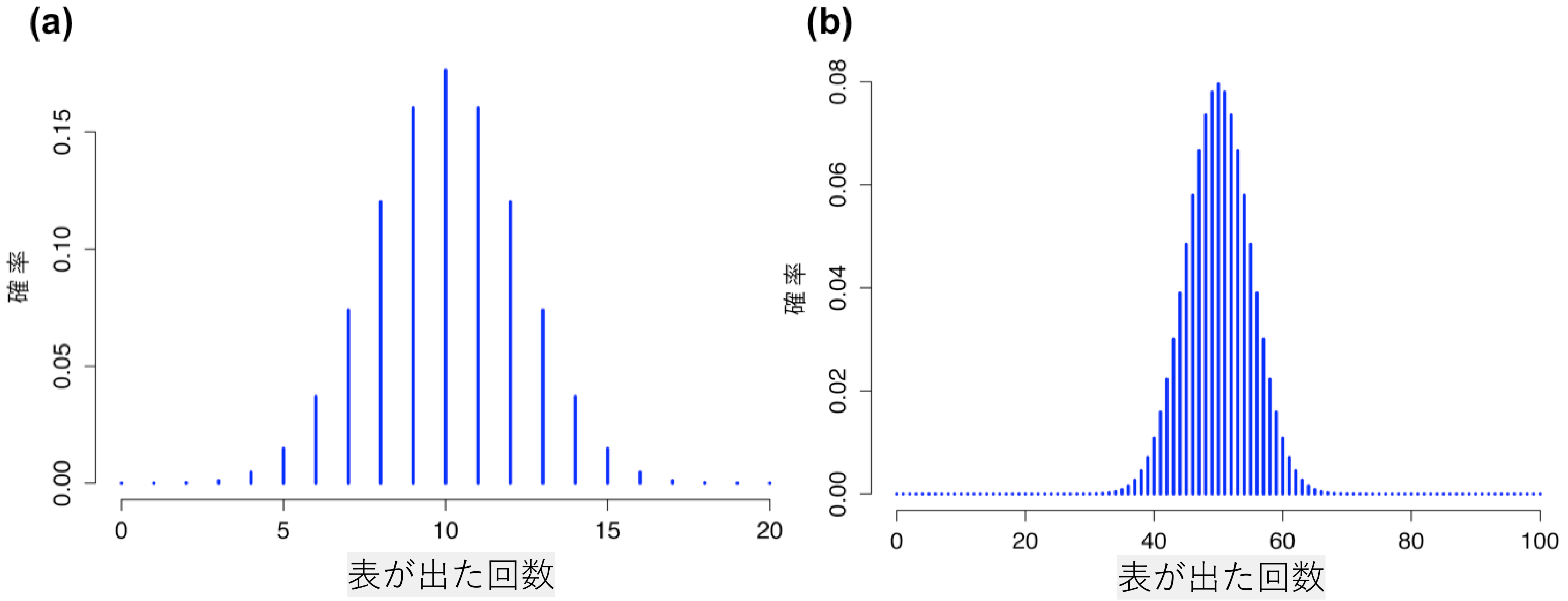
中心極限定理

- $X_1, X_2, \ldots, X_n$ は独立で同一な分布に従う確率変数列
- $E(X_i) = \mu, V(X_i) = \sigma^2 \ (i = 1,2, \ldots, n)$

このとき，確率変数Yは標準正規分布に弱収束する

$$\lim_{n \to \infty} \Pr\left(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leq y\right) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

二項分布の正規分布への近似



母比率の両側検定

- 母集団
  - パラメータπのベルヌーイ母集団
- 無作為標本$X_1, X_2, \ldots, X_n$
- 仮説
  - $H_0 : \pi = \pi_0$ vs $H_1 : \pi \neq \pi_0$

この統計的仮説検定に対する一様最強力不偏検定を構成する

一様最強力不偏検定の構成

帰無仮説 $H_0 : \pi = \pi_0$　　対立仮説 $H_1 : \pi \neq \pi_0$

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

中心極限定理より，n が十分大きいとき

$$Z \equiv \frac{\sqrt{n}(p-\pi)}{\sqrt{\pi(1-\pi)}} \approx N(0,1)$$

$$X \approx N(n\pi, n\pi(1-\pi))$$

$$p \approx N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

一様最強力不偏検定の棄却域W^*は

$$W^* = \left\{ (x_1, x_2, \dots, x_n) \,\middle|\, \bar{x} < \pi_0 - \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \text{ or } \bar{x} > \pi_0 + \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \right\}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \sigma = \sqrt{\pi_0(1-\pi_0)}$$

問1

帰無仮説 $H_0 : \pi = \frac{1}{2}$　　　対立仮説 $H_1 : \pi \neq \frac{1}{2}$

一様最強力不偏検定の棄却域W^*

$$W^* = \left\{ (x_1, x_2, \ldots, x_n) \,\middle|\, \bar{x} < \pi_0 - \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \text{ or } \bar{x} > \pi_0 + \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \right\}$$

コインを投げたとき表が出る確率をπ，$n = 10$,　$\alpha = 0.05$
としたとき，棄却域W^*に含まれる表が出る回数を答えよ。

問2

帰無仮説 $H_0: \pi = \frac{1}{2}$　　対立仮説 $H_1: \pi \neq \frac{1}{2}$

棄却域W

$$P(X \geq w) = \sum_{x=w}^{n} {}_nC_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{n-x} \leq \frac{\alpha}{2}$$

$$P(X \leq w) = \sum_{x=1}^{w} {}_nC_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{n-x} \leq \frac{\alpha}{2}$$

コインを投げたとき表が出る確率をπ，$n = 5$,　$\alpha = 0.05$としたとき，棄却域Wに含まれる表が出る回数を答えよ。

問3

帰無仮説 $H_0: \pi = \frac{1}{2}$　　対立仮説 $H_1: \pi \neq \frac{1}{2}$

一様最強力不偏検定の棄却域W^*

$$W^* = \left\{ (x_1, x_2, \dots, x_n) \,\middle|\, \bar{x} < \pi_0 - \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \text{ or } \bar{x} > \pi_0 + \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right) \right\}$$

コインを投げたとき表が出る確率をπ，$n = 5$,　$\alpha = 0.05$としたとき，棄却域W^*に含まれる表が出る回数を答えよ。