第 23 回　適合度検定（12.3）

村澤 康友

2024 年 12 月 20 日

今日のポイント

1. 母集団分布に対する標本の適合度を検定する．
2. 分布の範囲を k 階級に分割したときの母比率の両側検定を χ^2 適合度検定という．2 階級ならベルヌーイ母集団の母比率の両側検定となる．
3. χ^2 適合度検定を応用して 2 変量の独立性を検定できる．

目次

1　**母比率の検定**　1
　1.1　片側検定　1
　1.2　両側検定（p. 250）　1

2　**適合度検定**　2
　2.1　適合度検定問題（p. 245）　2
　2.2　ピアソンの χ^2 適合度検定（p. 246）　2
　2.3　独立性の χ^2 検定（p. 248）　3

3　**今日のキーワード**　4

4　**次回までの準備**　4

1　母比率の検定

1.1　片側検定

母集団分布を $\mathrm{Bin}(1,p)$ とする．次の片側検定問題を考える．

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p > p_0$$

有意水準を 5 %とする．$\mathrm{Bin}(1,p)$ の平均は p，分散は $p(1-p)$．大きさ n の無作為標本の標本比率（＝標本平均）を $\hat{p}$ とすると，中心極限定理より

$$\hat{p} \overset{a}{\sim} \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

標準化すると

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

$H_0 : p = p_0$ を代入すると，検定統計量は

$$Z := \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$$

H_0 の下で

$$Z \overset{a}{\sim} \mathrm{N}(0,1)$$

標準正規分布表より H_0 の下で

$$\Pr[Z \geq 1.65] \approx 0.05$$

したがって近似的な棄却域は $[1.65, \infty)$．

1.2　両側検定（p. 250）

次の両側検定問題を考える．

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0$$

有意水準を 5 %とする．標準正規分布表より H_0 の下で

$$\Pr[|Z| \geq 1.96] \approx 0.05$$

したがって近似的な棄却域は $(-\infty, -1.96] \cup [1.96, \infty)$．

注 1．Z^2 を検定統計量としてもよい．すなわち

$$Z^2 = \frac{n(\hat{p}-p_0)^2}{p_0(1-p_0)}$$

H_0 の下で

$$Z^2 \overset{a}{\sim} \chi^2(1)$$

χ^2 分布表より H_0 の下で

$$\Pr\left[Z^2 \geq 3.84146\right] \approx 0.05$$

したがって近似的な棄却域は $[3.84146, \infty)$．もちろん $1.96^2 \approx 3.84146$ で両検定は同等．

2 適合度検定

2.1 適合度検定問題（p. 245）

母集団分布の cdf を $F(.)$ とする（ノンパラメトリックでもよい）．

定義 1. 母集団分布に対する標本の適合度の検定を**適合度検定**という．

注 2. 適合度検定問題は

$$H_0 : F(.) = F_0(.) \quad \text{vs} \quad H_1 : F(.) \neq F_0(.)$$

k 階級に分割して分布を表すと

階級	$F(.)$	$F_0(.)$
1	p_1	$p_{0,1}$
$\vdots$	$\vdots$	$\vdots$
k	p_k	$p_{0,k}$
計	1	1

次の適合度検定問題を考える（元の問題と同等ではない）．

$$H_0 : \begin{pmatrix} p_1 \\ \vdots \\ p_{k-1} \end{pmatrix} = \begin{pmatrix} p_{0,1} \\ \vdots \\ p_{0,k-1} \end{pmatrix}$$
$$\text{vs} \quad H_1 : \begin{pmatrix} p_1 \\ \vdots \\ p_{k-1} \end{pmatrix} \neq \begin{pmatrix} p_{0,1} \\ \vdots \\ p_{0,k-1} \end{pmatrix}$$

未知母数は $k-1$ 個．$k=2$ なら母比率の両側検定．$k \geq 3$ なら多次元母数の両側検定となる．

例 1. $\mathrm{U}[0,1]$ と $\mathrm{N}(0,1)$ の標本（100 個の乱数）の適合度（図 1）．

2.2 ピアソンの χ^2 適合度検定（p. 246）

大きさ n の無作為標本における第 j 階級の度数を N_j とする．

定義 2. ピアソンの χ^2 適合度検定統計量は

$$\chi^2 := \sum_{j=1}^{k} \frac{(N_j - np_{0,j})^2}{np_{0,j}}$$

注 3. N_j を**観測度数**，$np_{0,j}$ を**期待度数**という．

注 4. 第 j 階級の相対度数を $\hat{p}_j := N_j/n$ とすると，

$$\begin{aligned}\chi^2 &= \sum_{j=1}^{k} \frac{(n\hat{p}_j - np_{0,j})^2}{np_{0,j}} \\ &= \sum_{j=1}^{k} \frac{n(\hat{p}_j - p_{0,j})^2}{p_{0,j}}\end{aligned}$$

$k=2$ なら

$$\begin{aligned}\chi^2 &= \frac{n(\hat{p}_1 - p_{0,1})^2}{p_{0,1}} + \frac{n(\hat{p}_2 - p_{0,2})^2}{p_{0,2}} \\ &= \frac{n(\hat{p}_1 - p_{0,1})^2}{p_{0,1}} + \frac{n[(1-\hat{p}_1) - (1-p_{0,1})]^2}{1-p_{0,1}} \\ &= \frac{n(\hat{p}_1 - p_{0,1})^2}{p_{0,1}} + \frac{n(\hat{p}_1 - p_{0,1})^2}{1-p_{0,1}} \\ &= \frac{(1-p_{0,1})n(\hat{p}_1 - p_{0,1})^2 + p_{0,1}n(\hat{p}_1 - p_{0,1})^2}{p_{0,1}(1-p_{0,1})} \\ &= \frac{n(\hat{p}_1 - p_{0,1})^2}{p_{0,1}(1-p_{0,1})}\end{aligned}$$

すなわち母比率の検定統計量と一致する．

定理 1. H_0 の下で

$$\chi^2 \overset{a}{\sim} \chi^2(k-1)$$

証明.「統計学入門」の水準を超えるので略． □

例 2 (p. 245，メンデルの法則)**.** えんどう豆の形質の遺伝に関する実験結果：

階級	N_j	$\hat{p}_j$	$p_{0,j}$
黄・丸	315	0.5665	0.5625
黄・しわ	101	0.1817	0.1875
緑・丸	108	0.1942	0.1875
緑・しわ	32	0.0576	0.0625
計	556	1.0000	1.0000

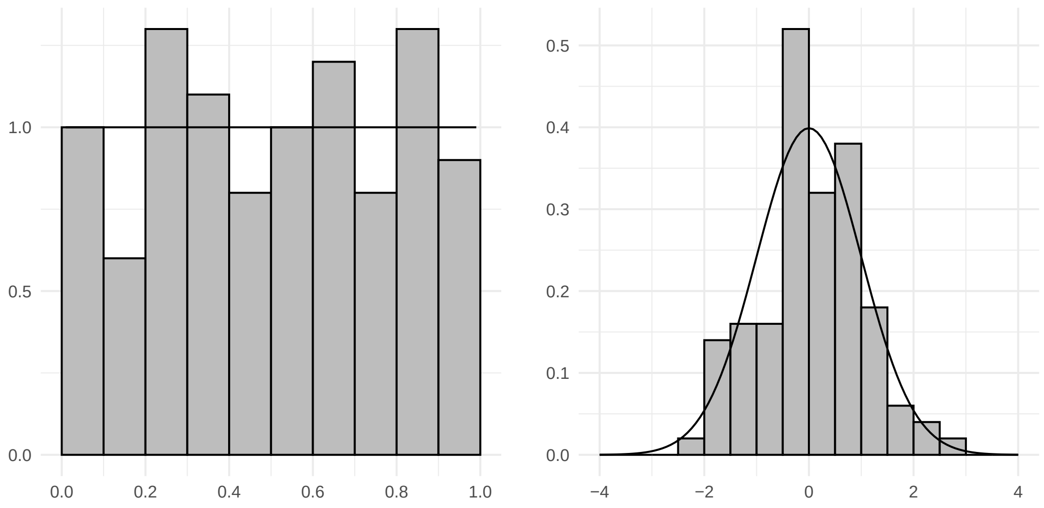

図 1　U[0, 1] と N(0, 1) の標本（100 個の乱数）の適合度

適合度検定問題は

$$
\begin{aligned}
&H_0 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} p_{0,1} \\ p_{0,2} \\ p_{0,3} \end{pmatrix} \\
\text{vs}\quad &H_1 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \neq \begin{pmatrix} p_{0,1} \\ p_{0,2} \\ p_{0,3} \end{pmatrix}
\end{aligned}
$$

有意水準を 5 %とする．H_0 の下で

$$\chi^2 \sim \chi^2(3)$$

χ^2 分布表より H_0 の下で

$$\Pr\left[\chi^2 \geq 7.81473\right] \approx 0.05$$

したがって近似的な棄却域は $[7.81473, \infty)$．$\chi^2 = 0.47$ となるので H_0 は棄却されない（ただし捏造の疑いあり？）．

2.3　独立性の χ^2 検定（p. 248）

2 変量母集団分布を $F_{X,Y}(.,.)$，その周辺分布を $F_X(.), F_Y(.)$ とする．独立性の検定問題は

$$
\begin{aligned}
&H_0 : F_{X,Y}(.,.) = F_X(.)F_Y(.) \\
\text{vs}\quad &H_1 : F_{X,Y}(.,.) \neq F_X(.)F_Y(.)
\end{aligned}
$$

$k \times l$ 分割表で分布を表すと

階級	1	…	l	計
1	$p_{1,1}$	…	$p_{1,l}$	$p_{1,.}$
⋮	⋮		⋮	⋮
k	$p_{k,1}$	…	$p_{k,l}$	$p_{k,.}$
計	$p_{.,1}$	…	$p_{.,l}$	1

次の適合度検定問題を考える（元の問題と同等ではない）．

$$
\begin{aligned}
H_0 : &\begin{bmatrix} p_{1,1} & \cdots & p_{1,l-1} \\ \vdots & & \vdots \\ p_{k-1,1} & \cdots & p_{k-1,l-1} \end{bmatrix} \\
= &\begin{bmatrix} p_{1,.}p_{.,1} & \cdots & p_{1,.}p_{.,l-1} \\ \vdots & & \vdots \\ p_{k-1,.}p_{.,1} & \cdots & p_{k-1,.}p_{.,l-1} \end{bmatrix} \\
\text{vs}\quad H_1 : &\begin{bmatrix} p_{1,1} & \cdots & p_{1,l-1} \\ \vdots & & \vdots \\ p_{k-1,1} & \cdots & p_{k-1,l-1} \end{bmatrix} \\
\neq &\begin{bmatrix} p_{1,.}p_{.,1} & \cdots & p_{1,.}p_{.,l-1} \\ \vdots & & \vdots \\ p_{k-1,.}p_{.,1} & \cdots & p_{k-1,.}p_{.,l-1} \end{bmatrix}
\end{aligned}
$$

未知母数は $(k-1)(l-1)$ 個．大きさ n の無作為標本における相対度数を $\hat{p}_{i,j}, \hat{p}_{i,.}, \hat{p}_{.,j}$ などとする．

定義 3. **独立性の χ^2 検定統計量は**

$$\chi^2 := \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{n(\hat{p}_{i,j}-\hat{p}_{i,\cdot}\hat{p}_{\cdot,j})^2}{\hat{p}_{i,\cdot}\hat{p}_{\cdot,j}}$$

定理 2. H_0 の下で

$$\chi^2 \overset{a}{\sim} \chi^2((k-1)(l-1))$$

証明.「統計学入門」の水準を超えるので略. □

例 3 (pp. 248–250)**.** 2 つの試験の成績（$n=42$）

成績	A	B	C	計
A	0.10	0.05	0.07	0.21
B	0.19	0.10	0.14	0.43
C	0.14	0.07	0.14	0.36
計	0.43	0.21	0.36	1.00

独立なら

成績	A	B	C	計
A	0.09	0.05	0.08	0.21
B	0.18	0.09	0.15	0.43
C	0.15	0.08	0.13	0.36
計	0.43	0.21	0.36	1.00

独立性の検定問題は

$$\begin{aligned} H_0 &: \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix} = \begin{bmatrix} p_{1,\cdot}p_{\cdot,1} & p_{1,\cdot}p_{\cdot,2} \\ p_{2,\cdot}p_{\cdot,1} & p_{2,\cdot}p_{\cdot,2} \end{bmatrix} \\ \text{vs} \quad H_1 &: \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix} \neq \begin{bmatrix} p_{1,\cdot}p_{\cdot,1} & p_{1,\cdot}p_{\cdot,2} \\ p_{2,\cdot}p_{\cdot,1} & p_{2,\cdot}p_{\cdot,2} \end{bmatrix} \end{aligned}$$

有意水準を 5 %とする．H_0 の下で

$$\chi^2 \sim \chi^2(4)$$

χ^2 分布表より H_0 の下で

$$\Pr\left[\chi^2 \geq 9.48773\right] \approx 0.05$$

したがって近似的な棄却域は $[9.48773, \infty)$．$\chi^2 = 0.19$ となるので H_0 は棄却されない（ただし捏造の疑いあり？）．

例 4. 男女の相性は血液型で決まるとの俗説がある．その真偽を科学的に検証したい．そこで無作為に選んだ 117 組の夫婦の血液型を調べたところ，次表の結果が得られた（数値は百分率を四捨五入）．

夫＼妻	A	O	B	AB	計
A	0.15	0.14	0.06	0.07	0.41
O	0.10	0.07	0.10	0.03	0.30
B	0.08	0.09	0.04	0.01	0.22
AB	0.04	0.00	0.03	0.00	0.07
計	0.37	0.30	0.23	0.10	1.00

独立なら

夫＼妻	A	O	B	AB	計
A	0.1517	0.1230	0.0943	0.0410	0.41
O	0.1110	0.0900	0.0690	0.0300	0.30
B	0.0814	0.0660	0.0506	0.0220	0.22
AB	0.0259	0.0210	0.0161	0.0070	0.07
計	0.37	0.30	0.23	0.10	1.00

有意水準を 5 %とする．H_0 の下で

$$\chi^2 \sim \chi^2(9)$$

χ^2 分布表より H_0 の下で

$$\Pr\left[\chi^2 \geq 16.919\right] \approx 0.05$$

したがって近似的な棄却域は $[16.919, \infty)$．$\chi^2 = 14.2309624$ となるので H_0 は棄却されない．

3 今日のキーワード

母比率の片側検定，母比率の両側検定，適合度検定，ピアソンの χ^2 適合度検定，観測度数，期待度数，独立性の χ^2 検定

4 次回までの準備

復習 教科書第 12 章 3 節，復習テスト 23

予習 教科書第 3 章 4 節，第 13 章 1–2.1 節