χ^2検定

岡本正芳

χ^2分布を用いた検定

* 適合度検定
  * 関係性を評価する方法
* 独立性検定
  * 無関係を調査する方法

真逆

証明は多項分布によるので省略

適合度検定

標本をk個の階級に分けて、各階級iに入った標本数を観測度数y_i(現実)とし、その合計を総度数nとする。各階級iに入る確率p_i(理論)とすると各階級に入ると期待される期待度数(理想)はnp_iとなる。総度数nが十分に大きいとき、確率変数

$$x = \sum_{i=1}^{k} \frac{\left(y_i - np_i\right)^2}{np_i}$$

は自由度k–1のχ^2分布に従う。有意水準αで

$$x > \chi^2\left(k-1,\alpha\right)$$ 棄却条件

のとき、各階級iに入る確率はp_iではない。(理論が破綻)

理論を標本が保証するのが望み

適合度検定の例

* メンデルの遺伝実験（エンドウ豆)
* 子葉が黄色の豆をn=8023個蒔いたところ発芽した子葉は黄色と緑色が6022:2001であった。メンデルの遺伝法則3:1は成立するかどうか。有意水準0.05で検定すると。

標本結果の整理

	観測値 y_i	理論値 np_i	差 $y_i - np_i$
黄色 $i=1$	6022	6017.25	4.75
緑色 $i=2$	2001	2005.75	–4.75
合計	8023	8023	0

* 確率変数は

$$x = \frac{(y_1 - np_1)^2}{np_1} + \frac{(y_2 - np_2)^2}{np_2} = \frac{4.75^2}{6017.25} + \frac{4.75^2}{2005.75} = 0.015$$

結果

* χ^2分布表より

$$\chi^2(2-1, 0.05) = 3.841$$

* 大小関係は

$$x = 0.015 < 3.841 = \chi^2(k-1, \alpha)$$

* 棄却条件は成立せず、メンデルの法則は成立している。

独立性検定

仮説H_0:事象AとBは独立　　対立仮説H_1:AとBは非独立

総度数nの標本を事象Aに関してk_a階級に、事象Bに関してk_b階級に分類する。(A_i,B_j)階級に属する観測度数y_{ij}より確率変数

$$x = \sum_{i=1}^{k_a}\sum_{j=1}^{k_b}\frac{\left(y_{ij}-\frac{y_{i\bullet}y_{\bullet j}}{n}\right)^2}{\frac{y_{i\bullet}y_{\bullet j}}{n}} \qquad y_{i\bullet}=\sum_{j=1}^{k_b}y_{ij},\quad y_{\bullet j}=\sum_{i=1}^{k_a}y_{ij}$$

は自由度$(k_a-1)(k_b-1)$のχ^2分布に従う。有意水準αで

$$x > \chi^2\left((k_a-1)(k_b-1),\alpha\right)$$ 棄却条件

のとき、H_0が棄却され、事象AとBは独立ではない。

独立性検定の例

* 朝食をとることと疲労の関係
* 朝食をとった100人中45人が疲労を感じており、朝食をとっていない400人中140人が疲労を感じていると報告された。この結果から調査。

仮説の宣言

* 仮説H_0:朝食をとることと疲労には関連性がない。
* 対立仮説H_1: 朝食をとることと疲労には関連性がある。

標本結果の整理

B \ A	朝食摂取	朝食非摂取	合計
疲労	45 $=y_{11}$	140 $=y_{21}$	185 $=y_{\bullet 1}$
疲労無	55 $=y_{12}$	260 $=y_{22}$	315 $=y_{\bullet 2}$
合計	100 $=y_{1\bullet}$	400 $=y_{2\bullet}$	500 $=n$

* 確率変数は

$$x=\frac{\left(y_{11}-\frac{y_{1\bullet}y_{\bullet 1}}{n}\right)^2}{\frac{y_{1\bullet}y_{\bullet 1}}{n}}+\frac{\left(y_{12}-\frac{y_{1\bullet}y_{\bullet 2}}{n}\right)^2}{\frac{y_{1\bullet}y_{\bullet 2}}{n}}+\frac{\left(y_{21}-\frac{y_{2\bullet}y_{\bullet 1}}{n}\right)^2}{\frac{y_{2\bullet}y_{\bullet 1}}{n}}+\frac{\left(y_{22}-\frac{y_{2\bullet}y_{\bullet 2}}{n}\right)^2}{\frac{y_{2\bullet}y_{\bullet 2}}{n}}$$

確率変数の計算

$$x=\frac{\left(y_{11}-\frac{y_{1\bullet}y_{\bullet1}}{n}\right)^2}{\frac{y_{1\bullet}y_{\bullet1}}{n}}+\frac{\left(y_{12}-\frac{y_{1\bullet}y_{\bullet2}}{n}\right)^2}{\frac{y_{1\bullet}y_{\bullet2}}{n}}+\frac{\left(y_{21}-\frac{y_{2\bullet}y_{\bullet1}}{n}\right)^2}{\frac{y_{2\bullet}y_{\bullet1}}{n}}+\frac{\left(y_{22}-\frac{y_{2\bullet}y_{\bullet2}}{n}\right)^2}{\frac{y_{2\bullet}y_{\bullet2}}{n}}$$

$$=\frac{\left(45-\frac{100\cdot185}{500}\right)^2}{\frac{100\cdot185}{500}}+\frac{\left(55-\frac{100\cdot315}{500}\right)^2}{\frac{100\cdot315}{500}}+\frac{\left(140-\frac{400\cdot185}{500}\right)^2}{\frac{400\cdot185}{500}}+\frac{\left(260-\frac{400\cdot315}{500}\right)^2}{\frac{400\cdot315}{500}}$$

$$=\frac{(45-37)^2}{37}+\frac{(55-63)^2}{63}+\frac{(140-148)^2}{148}+\frac{(260-252)^2}{252}$$

$$=\frac{8^2}{37}+\frac{8^2}{63}+\frac{8^2}{148}+\frac{8^2}{252}=3.432$$

結果

* χ^2分布表より

$$\chi^2\left((2-1)(2-1),0.05\right)=3.841$$

* 大小関係は

$$x=3.432<3.841=\chi^2\left(k-1,\alpha\right)$$

* H_0が棄却されず、朝食をとるかどうかと疲労には繋がりがあるとは言えない。