ピアソンの定理

鈴木 譲

2017 年 5 月 23 日

$j = 0, 1, \cdots, \alpha-1 (\alpha \geq 2)$ のいずれかの値をとる確率変数 X について、n 回中の各頻度を $W_j (\sum_{j=0}^{\alpha-1} W_j = n)$、$p_j = P(X=j)$ とおくとき、$n \to$ とともに、

$$\sum_{j=0}^{\alpha-1} \frac{(W_j - np_j)^2}{np_j}$$

の分布は、自由度 $\alpha - 1$ の χ^2 分布に収束する。

証明

確率変数 $X_1, \cdots, X_n$ は、独立で $P(X_k = j) = p_j$, $j = 0, 1, \cdots, \alpha-1, k = 1, \cdots, n$ であるものとする。以下では、事象 A が成立すれば $I(A) = 1$、しなければ $I(A) = 0$ とかくものとする。$I(X_k = j)$ も確率変数であって、平均は p_j、分散は $np_j(1-p_j)$ となる。また、$W_j = \sum_{k=1}^n I(X_k = j)$ と中心極限定理より

$$\frac{W_j - np_j}{\sqrt{np_j(1-p_j)}}$$

の分布は、標準正規分布に収束する。したがって、確率変数 $Z_j := \dfrac{W_j - np_j}{\sqrt{np_j}}$ は、漸近的に平均 0 分散 $1 - p_j$ の正規分布にしたがう。ところで、分散共分散は

$$EZ_iZ_j = \begin{cases} 1 - p_j, & i = j \\ -\sqrt{p_ip_j}, & i \neq j \end{cases}$$

となることを示すことができる。実際、$i \neq j$ として、以下の 2 式が成立する。

$$E[\frac{W_i - np_i}{\sqrt{np_i}} \cdot \frac{W_j - np_j}{\sqrt{np_j}}] = \frac{EW_iW_j - n^2p_ip_j}{n\sqrt{p_ip_j}}$$

$$\begin{aligned} EW_iW_j &= E[\sum_{k=1}^n I(X_k = i) \sum_{l=1}^n I(X_l = j)] = E\sum_{k \neq l} I(X_k = i, X_l = j) \\ &= \sum_{k=1}^n \sum_{l \neq k} EI(X_k = i, X_l = j) = n(n-1)p_ip_j \end{aligned}$$

この共分散行列 Σ の固有空間は、固有値が 0 の次元が 1 の空間と、固有値が 1 で次元が $\alpha-1$ の空間にわかれる。前者の固有ベクトルは

$$u_0 := [\sqrt{p_0}, \cdots, \sqrt{p_{\alpha-1}}]^T$$

である。実際、共分散行列の各行と内積をとると、すべて 0 になる。また、共分散行列から対角成分から 1 を引いた行列を A とおくとすべての行が他の行の定数倍になるので、A の行列式は 0 になる。したがって、Σ は固有値 1 をもつ。また、A の i 行目を $\sqrt{p_i}$ でわると、すべての行が $[\sqrt{p_0}, \cdots, \sqrt{p_{\alpha-1}}]$ となるので、Σ の固有値 1 の固有空間の次元は、$\alpha-1$ になる。そして、大きさが 1 で相互に直交する固有ベクトルを

$$u_1 = [u_{1,0}, \cdots, u_{1,\alpha-1}]^T, \cdots, u_{\alpha-1} = [u_{\alpha-1,0}, \cdots, u_{\alpha-1,\alpha-1}]^T$$

とおいて直交行列 $U = [u_0, \cdots, u_{\alpha-1}]$ を構成する。このとき、$U^{-1}\Sigma U$ は最初が 0 でそれ以外が 1 の対角行列になる。したがって、$V_i := \sum_{j=0}^{\alpha-1} u_{i,j} Z_j$, $j = 1, \cdots, \alpha-1$ の分散は 1, 共分散は 0 となる。また、U は直交行列であるので、$Z := [Z_0, \cdots, Z_{\alpha-1}]$ に対して、

$$||Z||^2 = ||U^T Z||^2 = \sum_{i=1}^{\alpha-1} V_i^2$$

も成立する。

標準正規分布にしたがう $\alpha-1$ 個の確率変数の和であるから、自由度 $\alpha-1$ の χ^2 分布にしたがう。

分割表の場合

$W_{i,j} := \sum_{k=1}^{n} I(X_k = i, Y_k = j)$, $p_i := P(X = i)$, $q_j := P(Y = j)$ として、X, Y とそれぞれ同じ分布をもつ独立な $X_1, \cdots, X_n$ および $Y_1, \cdots, Y_n$ について、

$$Z_{ij} := \frac{W_{i,j} - np_i q_j}{\sqrt{np_i q_j}}$$

に関して、

$$\sum_{i=0}^{\alpha-1} \sum_{j=0}^{\beta-1} Z_{i,j}^2$$

が自由度 $(\alpha-1)(\beta-1)$ の χ^2 分布にしたがうことを示す。

まず、1 次元の X に関して構成した $\alpha-1$ 次元のベクトル空間と同様に、Y に関しても $\beta-1$ 次元のベクトル空間を構成する。

2 次元の X, Y に関しても、$(\sqrt{p_i q_j})_{i,j}$ と直交するベクトルからなる $\alpha\beta-1$ 次元の空間を設定することができる。

ここで、この 3 個のベクトル空間のいずれにおいても、各次元方向の変動が標準正規分布にしたがい、異なる次元どうし独立になっている。

ここで、$\alpha\beta-1$ 次元のベクトル空間のうち、$\alpha-1$ 次元、$\beta-1$ 次元のベクトル空間と直交する $\alpha\beta-1-(\alpha-1)-(\beta-1) = (\alpha-1)(\beta-1)$ 次元の部分空間を取り出すことができる。そして、直交行列をかけたとしても、$(\alpha-1)(\beta-1)$ 次元の各基底は標準正規分布にしたがい、異なる次元どうし独立になるようにできる。