

# 有限混合モデルを用いた検定について

東京理科大学 下川 朝有  
東京理科大学 黒澤 匠雅  
東京理科大学 宮岡 悦良

## はじめに

層別化された分割表を用いて二群間の説明変数の目的変数に対する効果の比率の検定を行うには、コクラン・マンテル・ヘンツェル (CMH) 検定が広く用いられる。CMH 検定は母集団内の各層の比率を考慮せず、層別にサンプリングを行う事で、すなわち共変量に関してコントロールを行う事で得られる。一方で層別にサンプリングを行う事が出来ない場合、すなわち、直接母集団からのランダムサンプリングにより得られた標本を用いて同様の検定を行うには、母集団における各層の比率を考慮する必要があると考えられる。そこで本研究では、このような問題に対処するため、有限混合モデルを用いて2群間の効果の比率の検定について考察する。

## 記述・モデル

二値の応答を持つ独立な二群に対し、得られた標本による  $2 \times 2$  表は Table 1 の形式で与えられるとする。すなわち、 $a, b$  は有効な結果を返した群 1, 群 2 に含まれる標本数をそれぞれ表し、 $c, d$  は無効な結果を返したそれを表すとする。また、 $m_1, m_2$  はそれぞれ全標本に対する有効及び無効な結果を返した標本数を表すとし、 $n_1, n_2$  はそれぞれ群 1, 群 2 に含まれる標本数を表すとする。ここで Table 1 で与えられる標本は、共変量に従い一定の水準で  $K$  個の層に分割されるとし、第  $k$  層 ( $k = 1, 2, \dots, K$ ) に含まれる標本は Table 2 で与えられるとする。

Table 1:  $2 \times 2$  表

	群 1	群 2	
+	$a$	$b$	$m_1$
-	$c$	$d$	$m_2$
	$n_1$	$n_2$	$N$

Table 2:  $k$  層における  $2 \times 2$  表

	群 1	群 2	
+	$a_k$	$b_k$	$m_{1k}$
-	$c_k$	$d_k$	$m_{2k}$
	$n_{1k}$	$n_{2k}$	$N_k$

ここで、層  $k$  に含まれる群 1 の標本が有効な結果を返す真の確率を  $\pi_{1k}$ 、群 2 のそれを  $\pi_{2k}$  とし、帰無仮説  $H_0: \pi_{1k} = \pi_{2k}, \forall k$  の検定を考えるとする。層別にサンプリングが行われる場合、すなわち各層において  $n_{1k}$  及び  $n_{2k}$  が固定される場合、コクラン検定では  $a_k, b_k$  がそれぞれ従う分布としてパラメータ  $n_{1k}, \pi_{1k}$  の二項分布及び、パラメータ  $n_{2k}, \pi_{2k}$  の二項分布を考える。帰無仮説の下、 $a_k$  の期待値  $E(a_k)$  及び  $a_k - E(a_k)$  の分散を推定し、各層が独立であることを利用して以下の形で与えられる統計量を用いたカイ二乗検定を行う：

$$\frac{[\sum_k (a_k - \hat{E}(a_k))]^2}{\widehat{\text{Var}}\{\sum_k (a_k - \hat{E}(a_k))\}} \quad (1)$$

他方、マンテル・ヘンツェル検定では、 $m_{1k}$  の値も固定する事により、超幾何分布を用いて同様の検定を行う。

一方で直接母集団からのランダムサンプリングにより標本が得られる場合、 $n_{1k}$  及び  $n_{2k}$  は母集団における層  $k$  の比率  $w_k$  に従う確率変数と見なす必要がある。即ち、 $n_{1k}$  及び  $n_{2k}$  はそれぞれパラメータ  $n_1, w_k$  及び、パラメータ  $n_2, w_k$  の二項分布に従うと考えられ、 $a_k$  及び  $b_k$  は有限混合モデルによって与えられる事となる。本研究ではこの仮定の下、 $\sum_k (a_k - \hat{E}(a_k))$  及び、その分散の推定量を導出し、上記 (1) 式の形で与えられる検定統計量について考察する。さらにシミュレーションを用いて得られた検定統計量について考察し、CMH 検定との比較を行う。