

## Developing of Exact Tests for Order-Restrictions in Categorical Data

Jusun Nam<sup>a</sup> · Seung-Ho Kang<sup>b,1</sup>

<sup>a</sup>Department of Statistics, Ewha Womans University, Biologics Division, NIFDS

<sup>b</sup>Department of Applied Statistics, Yonsei University

(Received March 15, 2013; Revised May 20, 2013; Accepted August 14, 2013)

---

### Abstract

Testing of order-restricted alternative hypothesis in  $2 \times k$  contingency tables can be applied to various fields of medicine, sociology, and business administration. Most testing methods have been developed based on a large sample theory. In the case of a small sample size or unbalanced sample size, the Type I error rate of the testing method (based on a large sample theory) is very different from the target point of 5%. In this paper, the exact testing method is introduced in regards to the testing of an order-restricted alternative hypothesis in categorical data (particularly if a small sample size or extreme unbalanced data). Power and exact  $p$ -value are calculated, respectively.

Keywords: Exact test, categorical data, order-restricted alternative hypothesis.

---

### 1. 서론

통계학에서 통계적 검정 방법론은 많은 경우 대표본 이론에 근거하여 개발되었다. 마찬가지로 많은 경우 범주형 자료에서의 통계적 검정 방법론도 대표본 이론에 근거하여 개발되었다. 하지만 표본크기가 작거나 표본크기가 매우 불균등한 경우 대표본이론에 근거한 귀무가설 하에서 검정통계량의 근사분포로의 근사의 정확도는 상당히 떨어진다는 것이 일반적으로 알려져 있다. 이처럼 표본크기가 작거나 표본크기가 매우 불균등한 경우, 범주형 자료에서는 검정통계량의 근사분포에 의존하지 않고 정확분포를 사용하여 정확검정을 개발하는 것이 가능한 경우들이 존재한다. 최초로 개발된 정확검정은 Fisher에 의해 개발된 두 개의 독립인 이항분포에서 두 성공확률이 동일한지를 검정하는 Fisher's exact test이다. Fisher's exact test에서는 귀무가설 하에서의 장애모수(nuisance parameter)를 충분통계량을 사용한 조건부 분포를 구하는 방법으로 제거하게 되고, 이러한 방법으로 얻어진 정확검정을 조건부 정확검정이라고 한다 (Kang, 2002). 반면에 귀무가설 하에서 장애모수의 모든 가능한 경우를 고려한 후 최대값을 취하는 방식으로 장애모수를 없애서 정확검정을 만드는 것도 가능한데, 이러한 방법으로 얻어진 정확검정을 무조건부 정확검정이라고 한다 (Barnard, 1979). 정확검정은 이렇게 20세기 초기에 개발되

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2010-0009224).

<sup>1</sup>Corresponding author: Professor, Department of Applied Statistics, Yonsei University, 262 Seongsanno Seodaemun-ku, Seoul 120-749, Korea. E-mail: [seungho@yonsei.ac.kr](mailto:seungho@yonsei.ac.kr)

있음에도 불구하고, 분할표의 차원이 커지거나 자료의 크기가 커지면 계산량이 기하급수적으로 늘어나는 문제점이 존재하여 20세기 말까지 별다른 추가적인 발전이 이루어지지 않았다. 20세기 말에 급속하게 발전한 컴퓨터의 계산능력에 향상에 힘입어 다양한 정확검정들이 다양한 범주형 자료의 가설 검정 문제들에서 개발되었고, 현재 정확검정만을 전문으로 하는 상업용 프로그램까지 개발되어 판매되고 있다 (Mehta와 Patel, 1983; Mehtal 등, 1985; Mehta 등, 1998; Cytel, 2006).

이처럼 매우 다양한 범주형 자료의 가설 검정 문제들에서 정확검정이 개발되었음에도 불구하고, 순서화된 대립가설을 검정하는 정확검정은 아직 개발되지 않은 실정이다. 그래서 본 논문에서는 순서화된 대립가설을 검정하는 정확검정의 개발의 필요성과, 정확검정을 개발하는 방법, 그리고 정확검정으로 부터 얻어지는 검정력에 대하여 살펴볼 것이다.

## 2. 순서화된 대립가설

본 논문에서 연구할 순서화된 대립가설은 다음과 같다. 확률변수  $X_i (i = 1, 2, \dots, k)$ 들은 서로 독립이면서 각각 이항분포  $B(n_i, p_i)$ 를 따른다고 가정하자. 검정하고 싶은 귀무가설은  $H_0 : p_1 = p_2 = \dots = p_k = p, 0 < p < 1$ 이고, 대립가설은

$$H_1 : p_1 \leq p_2 \leq \dots \leq p_k$$

이고, 단 대립가설에 주어진 부등식들 중에서 적어도 하나는 등호 없이 엄격하게 크다고 주어진다. 이러한 가설을 검정하는 대표본 이론에 근거한 두 가지의 검정방법들이 개발되어 있다. 먼저 대립가설 하에서  $p_i$ 의 최대우도 추정량은 다음과 같이 주어진다.

$$\hat{p}_i = \max_{u \leq i} \min_{v \geq i} \hat{p}_{uv}, \quad \hat{p}_{uv} = \sum_{j=u}^v X_j / \sum_{j=u}^v n_j.$$

첫 번째 검정방법은 Bartholomew가 제안한  $\bar{X}^2$  검정으로서, 그 검정통계량은 다음과 같이 주어진다 (Bartholomew, 1959).

$$\bar{X}^2 = \sum_{i=1}^k \frac{n_i (\hat{p}_i - \hat{p})^2}{\nu}, \quad \hat{p} = \frac{R}{n}, \quad \nu = \frac{n\hat{p}(1-\hat{p})}{n-1}, \quad R = \sum_{i=1}^k X_i, \quad n = \sum_{i=1}^k n_i.$$

두 번째 검정방법은 우도비검정으로서, 그 검정통계량은 다음과 같이 주어진다 (Robertson 등, 1988, p.167).

$$\bar{G}^2 = 2 \sum_{i=1}^k \left[ X_i \log \left( \frac{\hat{p}_i}{\hat{p}} \right) + (n_i - X_i) \log \left( \frac{1 - \hat{p}_i}{1 - \hat{p}} \right) \right].$$

$\bar{X}^2$ 와  $\bar{G}^2$ 는 귀무가설 하에서 표본크기가 충분히 클 때,  $\chi^2$ 분포로 수렴함이 알려져 있다.  $\chi^2$ 분포는  $\chi^2$ 분포에 순위 확률(level probability)의 가중합으로 만들어진 분포로 다음처럼 주어진다.

$$P[\bar{X}^2 > C | H_0] = \sum_{l=2}^k P(l, k) \Pr[\chi^2_{k-1} > C], \quad C > 0,$$

$$P[\bar{X}^2 = 0 | H_0] = P(1, k),$$

여기서 순위확률인  $P(l, k)$ 은  $p_i$ 의 등위회귀함수(isotonic regression function)가 정확하게  $l$ 개의 구분 값(distinct values)을 가질 확률을 의미하고 다음과 같이 구해지며, 이는  $\sum_{l=1}^k P(l, k) = 1$ 의 특성을 가

지고 있다 (Barlow 등, 1972).

$$\begin{aligned}
 P(1, k) &= \frac{1}{k}, & P(k, k) &= \frac{1}{k!}, \\
 P(l, k) &= \frac{1}{k}P(l-1, k-1) + \frac{k-1}{k}P(l, k-1), & (l = 2, 3, \dots, k-1).
 \end{aligned}
 \tag{2.1}$$

$\bar{\chi}^2$  분포의 상위 1% 백분위수와 상위 5% 백분위수는 Bartholomew의 논문 표 2와 표 3에 제시되어 있다 (Bartholomew, 1959).

$k = 2$ 인 경우에는 식 (2.1)에 의하면,  $P(1, 2) = 1/2$ 이고  $P(2, 2) = 1/2$ 이므로,  $\bar{\chi}^2$  분포는 다음처럼 주어지게 된다.

$$\begin{aligned}
 P[\bar{X}^2 > C|H_0] &= \frac{1}{2} \times \Pr[\chi^2_1 > C], & C > 0, \\
 P[\bar{X}^2 = 0|H_0] &= \frac{1}{2}.
 \end{aligned}$$

즉 0보다 큰 곳에서는 자유도 1인 카이제곱 분포에 (1/2)를 곱한 형태이고, 한 점 0에는 (1/2)의 확률이 집중된 형태이다.

한 점 0에 (1/2)의 확률이 집중된 것은 다소 이해하기 어려운 현상인데, 검정통계량  $\bar{X}^2$ 의 형태를 살펴보면 이해할 수 있다.  $k = 2$ 인 경우 검정통계량  $\bar{X}^2$ 의 다른 형태를 유도하기 위하여 다음의 준비 단계를 거쳐보자. 설명의 편의상  $n_1 = n_2 = n$ 을 가정하자.

$$\begin{aligned}
 \hat{p}_1 &= \max_{u \leq 1} \min_{v \geq 1} \hat{p}_{uv} = \min_{v \geq 1} \hat{p}_{1v} = \min\left(\frac{X_1}{n}, \frac{X_1 + X_2}{2n}\right), \\
 \hat{p}_2 &= \max_{u \leq 2} \min_{v \geq 2} \hat{p}_{uv} = \max_{u \leq 2} \hat{p}_{u2} = \max\left(\frac{X_1 + X_2}{2n}, \frac{X_2}{n}\right), \\
 \hat{p} &= \frac{X_1 + X_2}{2n}.
 \end{aligned}$$

만일  $X_1 > X_2$ 이 만족된다면,  $\hat{p}_1 = \hat{p}_2 = \hat{p}$ 가 성립하여  $\bar{X}^2 = 0$ 이 되게 된다. 귀무가설 하에서 표본크기가 무한대로 증가하는 경우,  $X_1 > X_2$ 가 성립할 확률은 1/2로 수렴하게 되어 한 점 0에 (1/2)의 확률이 집중되게 된다.

### 3. $\bar{X}^2$ 검정과 우도비 검정의 제 1종 오류

$\bar{X}^2$  검정과 우도비 검정  $\bar{G}^2$ 는 귀무가설 하에서 표본크기가 충분히 큰 경우  $\bar{\chi}^2$  분포로 수렴함이 알려져 있다. 하지만 표본크기가 작거나 불균등한 경우  $\bar{\chi}^2$  분포로의 수렴의 정확도가 떨어져 실제 제 1종 오류가 유의수준 5%가 매우 달라지는 경우가 발생할 수도 있다. 이를 확인해보기 위하여 다음의 식을 이용하여  $\bar{\chi}^2$  검정의 정확한 제 1종 오류 확률을 계산해 보았다. 우도비 검정  $\bar{G}^2$ 에 대해서도 비슷한 방법으로 정확한 제 1종 오류 확률을 계산할 수 있다.

$$P(\bar{X}^2 \geq \bar{\chi}^2_\alpha | H_0) = \sum_{(x_1, \dots, x_k) \in \Gamma} \prod_{i=1}^k \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i} I_{[\bar{X}^2 \geq \bar{\chi}^2_\alpha]}, \tag{3.1}$$

여기서  $\Gamma$ 은  $\{(x_1, x_2, \dots, x_k) | 0 \leq x_i \leq n_i, i = 1, 2, \dots, k\}$ 으로  $(x_1, \dots, x_k)$ 의 전체 표본공간을 의미한다. 이 식에 의하면 모든 가능한 경우를 모두 나열하여 검정통계량  $\bar{X}^2$ 의 값이  $\bar{\chi}^2$  분포의 상위  $100\alpha$  백분위수인  $\bar{\chi}^2_\alpha$ 보다 큰 경우들의 정확한 확률들을 합하여 계산하므로, 제 1종 오류 확률을 정확하게 평가할 수 있게 된다.

**Table 3.1.** Type 1 error rate using exact formula:  $\bar{X}^2$  test,  $\bar{G}^2$  test ( $k = 3, \alpha = 5\%$ )

$\bar{X}^2$ test ( $k = 3, \alpha = 5\%$ )												
$p_1$	$n_1, n_2, n_3$											
$= p_2$	50, 10,	10, 10,	100,	10, 100,	10, 10,	10, 50,	10, 100,	100,	50, 100,	200,	200,	200,
$= p_3$	10	50	10, 10	10	100	100	50	50, 10	10	50, 50	100, 50	100, 50
0.01	6.2	<0.1	10.5	6.8	<0.1	<0.1	3.0	7.3	6.5	9.2	8.1	8.1
0.05	8.2	<0.1	8.4	5.3	<0.1	1.5	2.5	7.7	6.4	6.3	6.0	6.0
0.1	6.9	0.8	6.9	4.3	0.5	2.3	3.1	6.7	6.3	5.7	5.9	5.9
0.2	6.2	3.2	6.3	4.2	2.5	3.0	3.3	6.1	5.6	5.5	5.5	5.5
0.3	5.8	4.2	5.9	4.8	4.1	4.4	4.4	5.6	5.3	5.3	5.3	5.3
0.4	5.4	4.5	5.3	4.7	4.7	4.8	4.7	5.4	5.1	5.1	5.1	5.1
0.5	5.0	5.0	5.0	4.6	5.0	5.1	4.9	5.1	4.9	4.9	4.9	4.9
0.6	4.5	5.4	4.7	4.7	5.3	5.4	5.1	4.8	4.7	4.9	4.9	4.9
0.7	4.2	5.8	4.1	4.8	5.9	5.6	5.3	4.4	4.4	4.7	4.7	4.7
0.8	3.1	6.2	2.5	4.2	6.3	6.1	5.6	3.0	3.3	4.5	4.5	4.5
0.9	0.8	6.9	0.5	4.3	6.9	6.7	6.3	2.3	3.1	4.0	4.0	4.0

$\bar{G}^2$ test ( $k = 3, \alpha = 5\%$ )												
$p_1$	$n_1, n_2, n_3$											
$= p_2$	50, 10,	10, 10,	100,	10, 100,	10, 10,	10, 50,	10, 100,	100,	50, 100,	200,	200,	200,
$= p_3$	10	50	10, 10	10	100	100	50	50, 10	10	50, 50	100, 50	100, 50
0.01	0.1	<0.1	0.7	<0.1	<0.1	<0.1	<0.1	0.3	0.2	2.4	2.0	2.0
0.05	3.4	<0.1	4.7	0.4	<0.1	0.2	0.4	4.2	2.1	5.1	4.9	4.9
0.1	5.0	<0.1	4.5	1.0	<0.1	0.9	1.0	4.8	4.7	4.8	5.2	5.2
0.2	5.2	1.2	5.0	2.0	1.2	1.8	1.9	5.0	5.1	5.0	4.8	4.8
0.3	5.2	3.4	5.0	3.2	3.1	2.9	2.9	5.0	5.3	5.0	5.0	5.0
0.4	5.2	5.3	5.1	5.2	5.3	5.3	5.3	5.2	5.2	5.0	5.0	5.0
0.5	5.5	5.5	5.3	5.4	5.3	5.4	5.3	5.4	5.3	5.1	5.2	5.2
0.6	5.3	5.2	5.3	5.2	5.1	5.2	5.2	5.3	5.3	5.2	5.1	5.1
0.7	3.4	5.2	3.1	3.2	5.0	5.0	5.3	2.9	2.9	5.2	5.2	5.2
0.8	1.2	5.2	1.2	2.0	5.0	5.0	5.1	1.8	1.9	5.5	5.4	5.4
0.9	<0.1	5.0	<0.1	1.0	4.5	4.8	4.7	0.9	1.0	5.4	5.2	5.2

Table 3.1과 Table 3.2에 의하면 1종 오류 확률이 유의수준 5%를 초과하는 경우들이 발견된다. 예를 들어,  $(n_1, n_2, n_3) = (100, 10, 10)$ 이고  $(p_1, p_2, p_3) = (0.01, 0.01, 0.01)$ 인 경우  $\bar{X}^2$  검정의 제 1종 오류 확률은 10.5%로서 유의수준을 2배 이상 넘는다. 반면에 Table 3.1과 Table 3.2에 의하면 1종 오류 확률이 유의수준 5%보다 훨씬 작은 경우들도 발견되었다. 예를 들어,  $(n_1, n_2, n_3) = (10, 10, 50)$ 이고  $(p_1, p_2, p_3) = (0.01, 0.01, 0.01)$ 인 경우  $\bar{X}^2$  검정의 제 1종 오류 확률은 0.1%에 불과하였다.

$\bar{X}^2$  검정통계량은 다음처럼 표현이 가능한데

$$\bar{X}^2 = \sum_{i=1}^k \frac{n_i(\hat{p}_i - \hat{p})^2}{\nu} = \sum_{i=1}^k n_i/n \frac{(\hat{p}_i - \hat{p})^2}{\hat{p}(1 - \hat{p})/(n - 1)}.$$

이는 표준화된  $\hat{p}_i$ 들의 가중합이다. 각각의 표준화된  $\hat{p}_i$ 는 귀무가설 하에서 표본크기가 클 때, 이항분포의 정규근사에 의해 표준정규분포를 따른다. 알려진 대로, 이 정규근사는  $p_i$ 의 값이 0.5에 가까울수록 정확하므로  $p_i$ 가 0 또는 1에 매우 가까울 때는 대표본 이론에 의한  $\bar{X}^2$ 의 근사가 부정확해지게 된다. 또 다른 이유는 정규근사가 정확하기 위해서는 표본크기가 충분히 커야 하는데,  $(n_1, n_2, n_3) =$

**Table 3.2.** Type 1 error rate using exact formula:  $\bar{X}^2$  test,  $\bar{G}^2$  test ( $k = 4, \alpha = 5\%$ )

$\bar{X}^2$ 검정 ( $k = 4, \alpha = 5\%$ )									
$p_1 = p_2 = p_3 = p_4$	$n_1, n_2, n_3, n_4$								
	20, 20, 20, 20	30, 30, 30, 30	10, 50, 50, 100	50, 50, 10, 100	30, 50, 100, 50	30, 100, 50, 100	100, 100, 50, 50	150, 50, 50, 80	
0.01	0.1	1.6	0.1	0.7	2.0	1.6	6.7	5.0	
0.05	3.5	4.0	1.9	4.1	4.0	3.6	5.9	5.6	
0.1	4.2	4.6	2.7	4.5	4.3	4.3	5.5	5.4	
0.2	5.1	4.9	3.3	4.9	4.9	5.0	5.3	5.3	
0.3	5.0	5.0	4.1	4.9	5.1	5.1	5.2	5.2	
0.4	4.8	5.0	4.5	5.0	5.2	5.3	5.1	5.1	
0.5	4.9	4.9	4.9	5.1	5.2	5.4	4.9	5.1	
0.6	4.8	5.0	5.2	5.1	5.3	5.5	5.0	5.0	
0.7	5.0	5.0	5.6	5.2	5.3	5.6	4.8	4.9	
0.8	5.1	4.9	5.9	5.4	5.4	5.8	4.7	4.9	
0.9	4.2	4.6	6.7	5.5	5.6	6.1	4.5	4.7	

$\bar{G}^2$ 검정 ( $k = 4, \alpha = 5\%$ )									
$p_1 = p_2 = p_3 = p_4$	$n_1, n_2, n_3, n_4$								
	20, 20, 20, 20	30, 30, 30, 30	10, 50, 50, 100	50, 50, 10, 100	30, 50, 100, 50	30, 100, 50, 100	100, 100, 50, 50	150, 50, 50, 80	
0.01	1.2	2.1	0.7	3.0	2.2	3.2	4.4	6.5	
0.05	6.5	6.7	5.1	7.7	6.6	6.7	5.2	5.3	
0.1	7.7	6.6	4.9	6.0	7.0	7.3	5.2	5.1	
0.2	5.7	5.5	8.2	5.6	5.7	5.9	5.1	5.1	
0.3	5.6	5.2	6.1	5.3	5.5	5.8	5.1	5.1	
0.4	5.3	5.2	5.7	5.3	5.5	5.7	5.1	5.1	
0.5	5.2	5.2	5.3	5.3	5.4	5.6	5.1	5.1	
0.6	5.3	5.2	5.2	5.2	5.4	5.5	5.2	5.1	
0.7	5.6	5.2	5.1	5.1	5.4	5.5	5.2	5.2	
0.8	5.7	5.5	5.0	5.2	5.4	5.4	5.3	5.4	
0.9	7.7	6.6	4.9	5.3	5.6	5.5	5.8	5.6	

(100, 10, 10)의 경우에는  $n_2$ 와  $n_3$ 가 충분히 크지 않아서 일수도 있다.

반면에  $p_i$ 가 0이나 1 근처의 값을 가지지 않으며, 각 군의 표본수가 적당히 큰 경우에도 표본 불균형이 있는 경우에는 제 1종 오류 확률이 5.0%를 초과하는 경우가 발생할 수 있다. 예를 들어, Table 3.2의  $\bar{X}^2$  검정의 경우를 보면  $(n_1, n_2, n_3, n_4) = (30, 100, 50, 100)$ 이고,  $p_1 = p_2 = p_3$ 이 0.7, 0.8인 경우, 제 1종 오류 확률은 각각 5.6%, 5.8%로 5.0%를 상당히 초과함을 알 수 있다.

이처럼 표본크기가 작거나 불균등한 경우 귀무가설 하에서  $\bar{X}^2$  검정과 우도비 검정  $\bar{G}^2$ 의  $\bar{\chi}^2$  분포로의 수렴의 정확도가 떨어져, 실제 제 1종 오류가 유의수준 5%보다 커지거나 작아지는 경우 모두 바람직하지 않은 현상들이다. 실제 제 1종 오류가 유의수준 5%보다 지나치게 작아지는 경우는 검정력이 떨어지게 되어 바람직하지 않으며, 실제 제 1종 오류가 유의수준 5%보다 커지는 경우는 제 1종 오류가 제 2종 오류보다 심각한 오류이므로 매우 바람직하지 않은 경우이다. 이와 같은 경우에 대안으로 사용할 수 있는 방법이 바로 정확검정이다. 정확검정의 가장 큰 장점이 바로 제 1종 오류 확률을 어떤 경우에도 유의수준 5% 이하로 통제하는 것을 보장해준다는 점이다 (Kang, 2002).

식 (3.1)을 이용하여 계산되는 제 1종 오류 확률은 각 그룹의 표본크기가 동일한 경우 ( $n_1 = n_2 = \dots = n_k$ ),  $(p_1, p_2, p_3)$ 의 결과와  $(1 - p_1, 1 - p_2, 1 - p_3)$ 의 결과가 같아지는 성질이 존재한다. 그 이유는

$$\Gamma = \{(x_1, x_2, \dots, x_k) | 0 \leq x_i \leq n, i = 1, 2, \dots, k\}$$

에 속하는 임의의 원소  $(x_1, x_2, \dots, x_k)$ 에 대하여 다음의 성공확률  $p$ 인 이항분포 확률을 계산하면

$$\prod_{i=1}^k \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}.$$

이는  $\Gamma$ 에 속하는  $(n - x_1, n - x_2, \dots, n - x_k)$ 에 대하여 성공확률  $1 - p$ 인 이항분포 확률과 동일해지기 때문이다. 이러한 성질은 Table 3.2에 제시된 결과들 중에 표본크기가 동일한 경우들에 대한 결과들을 통하여서도 확인할 수 있다. 하지만 표본크기가 동일하지 않으면, 위 성질은 성립하지 않으며 Table 3.1에서 보듯이  $(p_1, p_2, p_3)$ 의 결과와  $(1 - p_1, 1 - p_2, 1 - p_3)$ 의 결과는 다르게 된다.

#### 4. 조건부 정확검정

4절에서는 조건부 정확검정의 기각역을 구하는 방법과 검정력을 계산하는 방법에 대해 살펴볼 것이다. 우선 귀무가설 하에서  $(X_1, X_2, \dots, X_k)$ 의 우도함수는 다음처럼 주어져서 미지의 장애모수  $p$ 에 의존함을 알 수 있다.

$$P(X_1 = x_1, \dots, X_k = x_k | H_0) = \prod_{i=1}^k \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i}. \quad (4.1)$$

이 장애모수  $p$ 를 제거하지 않고는 정확검정을 만들 수 없기 때문에 조건부 정확검정에서는 (4.1)에서 주어진 분포 대신, 귀무가설 하에서 장애모수인  $p$ 의 충분통계량이 주어졌을 때의 조건부 분포를 이용하여 조건부 정확검정을 만들게 된다. 충분통계량의 정의에 의하여 귀무가설 하에서 장애모수인  $p$ 의 충분통계량이 주어졌을 때의 조건부 분포는 더 이상  $p$ 에 의존하지 않게 된다.

귀무가설  $H_0 : p_1 = p_2 = \dots = p_k = p$  하에서 장애모수인  $p$ 의 충분통계량은  $R = \sum_{i=1}^k X_i$ 로 주어지므로,  $R = \sum_{i=1}^k X_i$ 이 주어졌을 때의 조건부 분포를 이용하여 조건부 정확검정을 만들게 되는데, 그 조건부 분포는 다음처럼 주어진다.

$$P(X_1 = x_1, \dots, X_k = x_k | R = r, H_0) = \prod_{i=1}^k \binom{n_i}{x_i} / \binom{n_1 + n_2 + \dots + n_k}{r}. \quad (4.2)$$

식 (4.2)에 주어진 분포는 다변량 초기하분포로서 더 이상 미지의 장애모수  $p$ 에 의존하지 않는다.

$T$ 를 표본공간  $S = \prod_{i=1}^k S_i$ ,  $S_i = \{0, 1, 2, \dots, n_i\}$  위에 정의된 큰 값일수록 더 대립가설에 대한 증거를 나타내는 검정통계량이라고 가정하자. 2절에서 살펴본  $\bar{X}^2$ 와  $\bar{G}^2$ 는 모두  $T$ 의 예가 될 수 있다.  $(x_1^0, x_2^0, \dots, x_k^0)$ 를  $(X_1, X_2, \dots, X_k)$ 의 관찰된 벡터라고 하고,  $r = \sum_{i=1}^k x_i^0$ 라고 하자. 조건부 정확검정은  $R$ 의 특정한 관측값  $r$ 이 주어졌을 때, 다음의  $2 \times k$  분할표를 표본공간으로 삼게 된다.

$$\Gamma_r = \left\{ (x_1, x_2, \dots, x_k) \mid \sum_{i=1}^k x_i = r, 0 \leq x_i \leq n_i, i = 1, 2, \dots, k \right\}.$$

이는 다시 말하면 조건부 정확검정은 관측된  $2 \times k$  분할표의 행의 합과 같은 행의 값을 갖는  $2 \times k$  분할표들만을 표본공간으로 고려하게 되는 것이다. 그 이유는 이렇게 함으로써 미지의 장애모수인  $p$ 를 제거할 수 있기 때문이다.

검정통계량  $T$ 는 큰 값일수록 대립가설에 대한 증거를 나타내므로, 주어진  $r$ 값에 대하여, 조건부 정확검정의 기각역은 다음과 같은 형태를 갖게 된다.

$$\Gamma_r(t) = \{(x_1, \dots, x_k) \in \Gamma_r | T(x_1, \dots, x_k) \geq t\}.$$

조건부 정확검정의 기각역을 완벽하게 결정하기 위해서는, 주어진  $r$ 에 대하여 귀무가설 하에서  $t$ 의 모든 가능한 값에 대하여, 다음의 귀무가설 하에서의  $T$ 의 분포를 먼저 구해야만 한다.

$$P(T \geq t | r, H_0) = \sum_{(x_1, \dots, x_k) \in \Gamma_r(t)} \prod_{i=1}^k \binom{n_i}{x_i} / \binom{n_1 + n_2 + \dots + n_k}{r}. \quad (4.3)$$

$\alpha$ 가 주어진 유의수준인 경우, 주어진  $r$ 에 대해, 식 (4.3)에서 구한 분포를 이용하여  $t_\alpha(r)$ 를 다음을 만족시키는 가장 작은  $t$ 값이라고 정의하자.

$$P(T \geq t | r, H_0) \leq \alpha.$$

$t_\alpha(r)$ 가 조건부 정확검정의 기각역을 정의하는 값이 되므로, 주어진  $r$ 에 대하여, 만일 검정통계량  $T$ 의 값이  $t_\alpha(r)$ 보다 크거나 같다면, 조건부 정확검정은 귀무가설을 기각하게 된다.

$n = \sum_{i=1}^k n_i$ 라고 할 때, 조건부 정확검정의 기각역이 정해졌으므로, 조건부 정확검정의 검정력을 계산할 수 있는데 조건부 정확검정에는 두 가지 종류의 검정력이 존재한다. 첫째는 주어진  $r$ 의 값에 대해서만 계산하는 conditional power인데 식으로는 다음과 같이 표현된다.

$$P(T \geq t_\alpha(r) | r, (p_1, \dots, p_k)).$$

실제 조건부 정확검정을 실시할 때는 특정한  $r$ 값을 관측한 후 조건부 정확검정을 실시하는 것이어서, conditional power가 보다 현실에서 의미 있는 검정력이라고 볼 수 있다.

둘째는 unconditional power이다. conditional power는  $r$ 값에 따라 매번 달라지기 때문에, conditional power를 사용한다면, 조건부 정확검정과 무조건부 정확검정의 검정력을 비교해볼 수가 없게 된다. 그래서 가능한 모든  $r$ 값을 고려하는 unconditional power를 계산하게 되는데, 다음의 식 (4.4)에 주어진 검정력이 바로 unconditional power이다. 식 (4.4)에 주어진 unconditional power는 conditional power의 가중평균이라고 볼 수 있다.

$$\begin{aligned} \beta^c(p_1, \dots, p_k) &= \sum_{r=0}^n P(T \geq t_\alpha(r) | r, (p_1, \dots, p_k)) P(R = r | p_1, \dots, p_k) \\ &= \sum_{(x_1, \dots, x_k) \in \Gamma_\alpha} \prod_{i=1}^k \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i} \end{aligned} \quad (4.4)$$

이고,  $\Gamma_\alpha$ 는 주어진  $r$ 에 대한 조건부 정확검정의 기각역들을 모두 포함하는 집합으로서 다음처럼 표현된다.

$$\Gamma_\alpha = \cup_{r=0}^n \Gamma_r(t_\alpha(r)).$$

### 5. 무조건부 정확검정

무조건부 정확검정에서는 다음의  $2 \times k$  분할표들을 표본공간으로 삼게 된다.

$$\Gamma = \cup_{r=0}^n \Gamma_r = \{(x_1, \dots, x_k) | 0 \leq x_i \leq n_i, i = 1, \dots, k\}.$$

무조건부 정확검정의 기각역을 결정하기 위해서, 우선 0과 1 사이에 주어진 각각의  $p$ 값에 대하여, 모든 가능한  $t$ 값에 대한 귀무가설 하에서의  $T$ 의 분포를 다음처럼 구한다.

$$P(T \geq t|p, H_0) = \sum_{(x_1, \dots, x_k) \in \Gamma} \prod_{i=1}^k \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i} I_{[T(x_1, \dots, x_k) \geq t]}.$$

$\alpha$ 가 주어진 유의수준인 경우, 0과 1사이의 각각의  $p$ 값에 대하여  $t_\alpha(p)$ 를 다음을 만족시키는 가장 작은  $t$ 값으로 정의하자.

$$P(T \geq t|p, H_0) \leq \alpha.$$

이렇게 얻어진 기각역을 나타내는 값인  $t_\alpha(p)$ 는 미지의 장애모수인  $p$ 값에 의존하게 되어 실제로는 사용이 불가능한 값이다. 그러므로 무조건부 정확검정에서는 미지의 장애모수  $p$ 를 없애기 위해 다음의 방법을 사용하여 무조건부 정확검정의 기각역을 나타내는 값인  $t_\alpha$ 을 정하게 된다.

$$t_\alpha = \sup_{0 < p < 1} t_\alpha(p). \quad (5.1)$$

즉 무조건부 정확검정에서는 검정통계량  $T$ 의 값이  $t_\alpha$ 보다 크거나 같으면 귀무가설을 기각하게 된다.

무조건부 정확검정의 검정력을 나타내는 식은 다음과 같이 주어진다.

$$\begin{aligned} \beta^u(p_1, \dots, p_k) &= P(T \geq t_\alpha | p_1, \dots, p_k) \\ &= \sum_{(x_1, \dots, x_k) \in \Gamma} \prod_{i=1}^k \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i} I_{[T(x_1, \dots, x_k) \geq t_\alpha]}. \end{aligned} \quad (5.2)$$

## 6. $p$ -value 비교

$(x_1^0, x_2^0, \dots, x_k^0)$ 를  $(X_1, X_2, \dots, X_k)$ 의 관찰된 벡터라고 하고,  $r = \sum_{i=1}^k x_i^0$ 라고 하자. 이 경우 조건부 정확검정의 exact  $p$ -value는 행합이  $r$ 과 같은 모든  $2 \times k$  분할표에 대해 검정통계량의 값이 관측된 검정통계량의 값보다 크거나 같은 조건부 확률의 합으로서 다음과 같이 주어지게 된다.

$$\text{조건부 exact } p\text{-value} = \sum_{(x_1, \dots, x_k) \in \Gamma_r} \prod_{i=1}^k \binom{n_i}{x_i} / \binom{n_1 + n_2 + \dots + n_k}{r} I_{[T(x_1, \dots, x_k) \geq T(x_1^0, \dots, x_k^0)]}.$$

반면에 무조건부 정확검정의 exact  $p$ -value는 각 열의 합이  $(n_1, \dots, n_k)$ 인 모든  $2 \times k$  분할표에 대해 검정통계량의 값이 관측된 검정통계량의 값보다 크거나 같은 무조건부 조건부 확률의 합을 구한 후, 장애모수 공간 상에서 장애모수에 대해 최대값을 취한 값으로서 다음과 같이 주어지게 된다.

$$\text{무조건부 exact } p\text{-value} = \max_{0 < p < 1} \sum_{(x_1, \dots, x_k) \in \Gamma} \prod_{i=1}^k \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i} I_{[T(x_1, \dots, x_k) \geq T(x_1^0, \dots, x_k^0)]}.$$

반면에  $\bar{X}^2$  검정과 우도비 검정  $\bar{G}^2$ 의 대표본 이론에 근거한 근사  $p$ -value는  $\bar{\chi}^2$  분포의 꼬리 확률을 평균함으로써 구할 수 있는데 현재까지는 알려진 수치적 도구가 존재하지 않는다. 하지만 세 군의 경우에 분포의 꼬리 확률을 비교적 간단히 계산할 수 있으며, 본 논문에서도 SAS 9.2를 이용하여 각 확률을 계산하였다.

Table 6.1에는 근사  $p$ -value와 exact  $p$ -value들이 여러 가지 표본크기에 대하여 제시되어 있다. 이 값들을 비교해보면 많은 경우 유의수준 5%에서 동일한 결론을 나타냈으나, 다른 결론을 내는 경우도 종종



**Table 6.1.** Exact  $p$ -value(%) for conditional and unconditional exact test method

$n_1, n_2, n_3$	$x_1$	$x_2$	$x_3$	asymptotic test		Unconditional Exact test		conditional Exact test	
				$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$
5, 5, 5	0	0	2	3.8	8.0	4.1	5.7	9.5	9.5
	0	1	3	3.5	5.9	3.9	3.0	4.0	4.0
	0	2	3	4.7	6.0	5.5	3.6	5.0	5.0
	0	3	3	3.2	3.5	3.3	2.7	4.3	7.8
	0	4	1	6.0	7.3	7.2	4.4	9.2	9.2
	1	4	4	3.2	7.5	3.3	5.4	7.8	4.7
	2	2	5	3.2	3.5	3.7	2.7	4.3	7.8
	2	3	5	4.7	6.0	5.5	3.6	5.0	5.0
	2	4	5	3.5	5.9	4.0	3.0	4.0	4.0
	2	5	4	4.6	12.5	4.9	6.4	7.7	7.7
	3	1	5	3.2	3.5	3.7	2.7	4.3	7.8
	3	2	5	6.0	7.3	7.3	4.5	9.2	9.2
	4	1	5	6.0	7.3	7.3	4.5	9.2	9.2
	5	0	5	6.0	7.3	7.3	4.4	5.1	9.2
10, 5, 5	0	0	2	1.2	4.3	1.9	2.5	5.3	5.3
	0	1	2	4.1	7.5	5.3	4.4	5.3	5.3
	0	2	2	2.9	3.8	3.4	2.0	4.3	6.4
	1	1	3	3.7	11.9	4.0	6.1	5.2	5.2
	1	3	3	2.2	5.1	2.5	3.0	2.9	5.1
	2	0	3	4.0	13.6	4.8	8.7	8.1	8.1
	2	2	4	2.9	7.5	3.4	4.4	3.0	3.0
	2	3	4	2.3	5.5	2.6	3.2	2.4	2.4
	2	5	2	2.8	7.1	3.1	4.0	4.0	4.0
	3	2	5	1.2	1.2	1.2	0.6	0.9	1.8
	3	3	5	1.2	1.4	1.9	0.9	1.0	1.0
	3	4	4	2.8	7.1	2.8	4.4	5.6	5.6
	4	0	4	3.8	10.4	4.3	5.4	6.3	6.3
	6	0	5	2.3	2.6	2.6	1.7	3.1	3.1
	6	1	5	3.8	4.5	4.4	2.5	5.6	5.6
6	5	5	2.9	3.8	3.1	2.0	4.3	4.3	
7	0	5	3.8	4.5	4.4	2.5	5.6	5.6	
10, 10, 10	0	0	2	4.2	9.6	4.1	5.5	10.3	10.3
	0	1	3	4.8	8.9	4.8	4.2	5.1	5.1
	0	5	1	5.6	6.1	6.4	3.7	7.0	12.1
	1	1	5	1.8	5.7	1.6	2.8	5.7	2.6
	1	2	5	3.9	11.0	3.9	6.7	4.6	4.6
	1	5	5	3.6	7.4	3.5	3.8	4.2	4.5
	5	1	7	4.1	11.1	4.1	6.7	5.3	5.3
	5	2	8	2.3	5.8	2.5	2.8	4.5	3.4
	5	6	9	5.1	11.3	4.9	6.7	6.0	6.0
	5	10	7	4.4	13.4	4.8	7.1	7.1	7.1
	7	1	8	4.2	10.4	4.5	6.6	5.3	5.3
	7	8	10	6.9	9.0	7.7	4.9	6.9	6.9
7	9	10	4.8	8.9	4.8	4.2	5.1	5.1	

Table 6.1 (Continue)

$n_1, n_2, n_3$	$x_1$	$x_2$	$x_3$	asymptotic test		Unconditional Exact test		conditional Exact test	
				$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$
15, 15, 15	1	1	5	2.3	7.8	2.4	4.4	6.8	6.8
	1	2	6	1.8	5.7	1.8	2.8	2.8	2.8
	2	0	5	2.3	7.8	2.4	4.4	6.8	6.8
	2	2	7	1.6	5.6	1.6	2.7	2.4	2.4
	2	4	8	1.8	5.3	1.8	2.6	2.1	2.1
	3	10	7	2.2	5.6	2.2	2.7	2.6	2.6
	3	11	6	2.2	5.6	2.2	2.7	2.3	2.3
	4	0	7	1.6	5.6	1.6	2.7	1.3	1.3
	4	9	10	2.1	5.8	2.1	2.8	1.8	1.8
	5	2	9	1.8	5.5	1.7	2.6	2.7	2.7
	5	4	10	2.1	6.3	2.1	2.8	2.6	2.6
	7	6	12	2.2	5.6	2.2	2.7	2.3	2.3
	7	9	13	2.2	5.3	2.2	2.6	1.9	1.9
	7	11	13	1.8	5.3	1.8	2.6	2.1	2.1
	9	13	14	1.8	5.7	1.7	2.8	2.8	1.5
	10	3	12	2.2	5.6	2.2	2.7	2.6	2.6
10	14	14	2.3	7.8	2.4	4.4	6.8	3.2	
11	2	12	2.2	5.6	2.2	2.7	2.6	2.6	
11	15	14	2.1	7.3	2.1	3.9	3.6	3.6	
30, 5, 5	0	0	1	0.8	11.4	5.2	6.5	12.5	12.5
	0	1	1	1.3	5.3	6.4	3.0	5.8	5.8
	0	2	0	1.3	5.3	6.4	3.0	5.8	5.8
	2	0	2	1.8	13.9	6.7	7.8	8.9	8.9
	2	1	2	3.1	15.6	6.8	9.5	5.0	5.0
	2	2	2	1.1	5.9	5.3	3.2	3.6	3.6
	4	1	3	1.6	9.4	6.5	5.2	3.4	3.4
	4	2	3	1.3	6.0	5.3	3.5	1.9	1.9
	7	3	3	3.2	11.1	6.8	6.4	6.1	6.1
	7	4	3	0.8	3.0	5.2	1.7	1.4	1.4
	9	3	4	2.1	6.5	6.7	3.8	2.1	2.1
	9	4	3	2.6	8.3	6.7	4.9	3.5	3.5
	12	3	5	1.2	1.5	5.3	0.8	1.0	1.0
	12	4	4	2.8	8.0	6.7	4.7	4.8	4.8
	16	3	5	4.6	5.9	7.2	3.5	5.0	5.0
	16	4	5	2.9	3.9	6.7	2.3	2.2	2.2
	18	0	5	3.9	4.8	6.8	2.8	5.9	5.9
	18	1	5	4.8	6.1	7.2	3.8	7.7	7.7
	19	1	5	5.9	7.7	7.9	4.5	8.1	9.9
	19	5	5	2.5	2.2	6.7	1.2	2.4	2.4
21	5	5	4.7	5.0	7.2	2.8	5.2	5.2	
22	5	5	6.3	7.3	7.9	3.9	7.6	7.6	

있다. 대표본 이론에 근거한 근사 검정방법 안에서도  $\bar{X}^2$  검정과 우도비 검정  $\bar{G}^2$ 의 결과가 다른 경우가 존재하고, 마찬가지로 무조건부 정확검정과 조건부 검정 각각의 안에서도  $\bar{X}^2$  검정과 우도비 검정  $\bar{G}^2$ 의

**Table 7.1.** Power for conditional and unconditional exact test method

$n_1, n_2, n_3$	$p_1$	$p_2$	$p_3$	Conditional exact test		Unconditional exact test	
				$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$
50,10,10	0.1	0.2	0.5	84.2	84.3	75.4	78.2
	0.1	0.3	0.6	95.3	95.5	90.7	92.6
	0.2	0.2	0.6	74.8	74.8	62.2	66.8
	0.3	0.3	0.7	71.2	71.2	54.3	63.5
	0.3	0.4	0.8	90.0	90.0	77.8	85.3
10,50,10	0.1	0.2	0.5	64.7	59.1	49.7	60.2
	0.1	0.3	0.6	73.6	73.4	57.7	70.8
	0.2	0.2	0.6	73.0	64.0	62.0	66.7
	0.3	0.3	0.7	68.5	65.0	54.2	63.4
	0.3	0.4	0.8	77.4	76.5	60.6	74.2
100,10,10	0.1	0.2	0.5	87.3	87.7	84.6	81.0
	0.1	0.3	0.6	97.1	97.2	95.9	94.9
	0.2	0.2	0.6	76.6	77.4	71.0	68.9
	0.3	0.3	0.7	73.8	74.2	64.4	65.3
	0.3	0.4	0.8	92.6	92.8	87.0	87.8
10,50,100	0.1	0.2	0.5	99.4	99.2	96.4	98.6
	0.1	0.3	0.6	99.7	99.6	98.1	99.2
	0.2	0.2	0.6	99.9	99.9	99.6	99.8
	0.3	0.3	0.7	99.9	99.9	99.4	99.7
	0.3	0.4	0.8	99.9	99.9	99.8	99.9
10,100,50	0.1	0.2	0.5	98.5	97.3	94.3	97.2
	0.1	0.3	0.6	99.0	98.8	95.7	98.1
	0.2	0.2	0.6	99.8	99.8	99.3	99.7
	0.3	0.3	0.7	99.8	99.7	99.0	99.6
	0.3	0.4	0.8	99.9	99.9	99.6	99.8
100,50,10	0.1	0.2	0.5	89.6	89.9	82.1	84.1
	0.1	0.3	0.6	98.9	99.1	97.4	98.2
	0.2	0.2	0.6	77.3	76.3	65.7	67.9
	0.3	0.3	0.7	73.2	73.5	56.8	64.4
	0.3	0.4	0.8	92.9	93.1	82.3	88.0

결과가 다른 경우가 존재한다. 또한 같은 검정통계량을 사용하는 경우에도 대표본 이론에 근거한 근사 검정방법, 무조건부 정확검정, 그리고 조건부 정확검정의 결과가 많은 경우 유의수준 5%에서 동일한 결론을 냈지만, 다른 결론을 내는 경우도 종종 발생하였다. 어느 검정방법이 가장 적절한지에 대한 논의는 7절과 9절에서 계속할 것이다.

### 7. 검정력의 비교

식 (4.4)와 식 (5.2)에 주어진 식을 사용하여 조건부 정확검정과 무조건부 정확검정의 검정력을 계산하는 SAS macro를 만들어서 여러 가지 조합에 대해 검정력을 계산한 결과가 Table 7.1에 제시되어 있다. 정확검정에서는 어떤 검정통계량을 사용해야 가장 적절한지에 대해서는 알려진 바 없으므로, 통상적으로 대표본이론에 근거한 검정방법에 사용되는 검정통계량이 흔히 사용된다. 그러므로 본 논문에서는 2절에서 주어진  $\bar{X}^2$  검정통계량과 우도비 검정의 검정통계량이 사용되었다. 조건부정확검정의 경우

**Table 7.2.** Exact  $p$ -value(%) for conditional and unconditional exact test method  $(n_1, n_2, n_3) = (50, 10, 10)$   $\alpha = 5\%$ .

$p_1$	$p_2$	$p_3$	Conditional exact test		Unconditional exact test	
			$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$
0.1	0.1	0.1	3.91	3.94	2.67	2.74
0.2	0.2	0.2	4.34	4.50	2.09	2.70
0.3	0.3	0.3	4.54	4.67	1.81	2.74
0.4	0.4	0.4	4.30	4.40	1.50	2.83
0.5	0.5	0.5	4.32	4.45	1.22	3.07
0.6	0.6	0.6	4.22	4.40	1.06	3.14
0.7	0.7	0.7	3.94	4.11	0.73	3.77
0.8	0.8	0.8	3.31	3.23	0.31	4.79
0.9	0.9	0.9	0.85	0.84	0.01	2.96

는 사용된 검정통계량에 따라 검정력이 큰 차이를 보이지 않았다. 하지만, 무조건부 정확검정의 경우에는 소표본의 경우 우도비 검정의 검정통계량을 사용하는 경우가 더 높은 검정력을 보였다.  $\bar{X}^2$  검정과 우도비 검정에 대하여 조건부 검정의 검정력이 무조건부 검정의 검정력에 비하여 대체로 더 높았다. 하지만 이런 결과는 Table 7.1에서 조사한 표본크기와  $(p_1, p_2, p_3)$ 의 조합들에 한하여 해당되는 내용이며, 다른 조합들에 대해서는 다른 결과가 나올 수도 있다.

일반적으로 같은 검정통계량을 사용하는 경우 조건부 검정과 무조건부 정확검정의 검정력 비교는 두 검정이 얼마나 보수적이냐를 판단해보면 알 수 있다 (Kang과 Ahn, 2008). 조건부 정확검정은 표본크기가 극단적으로 작은 경우 검정통계량의 누적분포의 함수가 계단 모양이 되는데 각 계단의 높이가 너무 커서 주어진 5%의 유의수준을 다 사용하지 못하면서 발생하게 된다. 반면에 무조건부 정확검정에서 기각역은 식 (5.1)에서 주어진 것처럼  $t_\alpha = \sup_{0 < p < 1} t_\alpha(p)$ 처럼 얻어지는데, 이렇게 구한  $t_\alpha$ 와 장애모수  $p$ 미지의 참값에서 구한  $t_\alpha(p)$ 의 차이가 클수록 더 보수적이게 된다. Table 7.1에서 주어진 경우들에서는 조건부 정확검정이 보수적일 수 있는 상황보다 무조건부 정확검정이 보수적일 수 있는 상황이 더 심하게 발생하여, 무조건부 검정검정이 더 보수적이고 그래서 검정력이 더 낮다고 판단되어진다.

조건부 정확검정과 무조건부 정확검정의 검정력이 차이를 좀 더 심도 있게 이해하기 위하여  $(n_1, n_2, n_3) = (50, 10, 10)$ 의 경우 제 1종 오류 확률들을 계산하여 표5에 제시하였다. Table 7.2의 결과들을 보면 조건부 정확검정의 제 1종 오류 확률들이 무조건부 정확검정의 제 1종 오류확률들보다 대체로 5% 유의수준에 더 가깝다는 것을 알 수 있다. 즉 조건부 정확검정은 무조건부 정확검정에 비하여 주어진 5% 유의수준을 더 많이 사용하였고 그래서 검정력이 더 높게 되는 것이다. 무조건부 정확검정의 제 1종 오류 확률들이 매우 낮아서 보수적이 되는 것을 이해하기 위하여,  $(n_1, n_2, n_3) = (50, 10, 10)$ 의 경우  $p$ 값과  $t_\alpha(p)$ 의 관계를 Figure 7.1에 제시하였다. 무조건부 정확검정에서는 기각역을  $t_\alpha = \sup_{0 < p < 1} t_\alpha(p)$ 로 구하게 되는데, 이렇게 구한  $t_\alpha$ 와 장애모수  $p$ 미지의 참값에서 구한  $t_\alpha(p)$ 의 차이가 크기 때문에 무조건부 정확검정이 보수적이 된 것이다.

## 8. 예제

생물학적제제의 경우 그 물질의 역가를 단독으로 설정할 수가 없다. 이에 물질의 특성과 그 역가를 명확히 알고 있는 표준물질이 필요하게 된다. 이에 WHO는 생물학적제제의 국제표준품을 설정하여 분양을 하고 있고, EU, 일본 등 제약선진국에서 또한 그 지역표준품 또는 국가표준품을 설정하여 분양하고 있다. 우리나라 또한 1990년대 국가표준품 확립의 필요성을 인식하고, 소마트로핀(인성장호르몬)을 시작

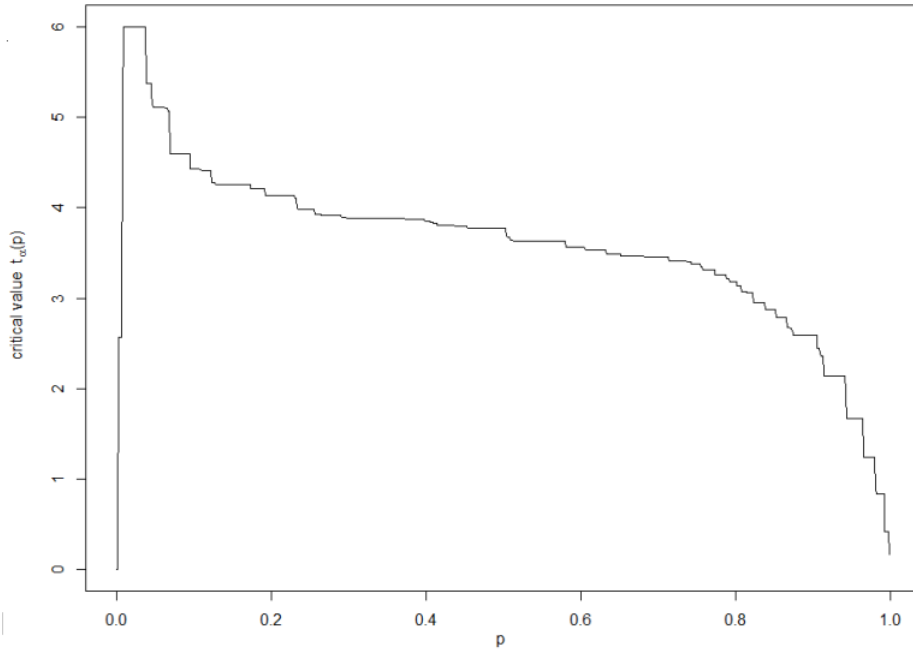


Figure 7.1. Relation between  $p$  and  $t_\alpha(p)$  ( $\alpha = 0.05, n_1 = 50, n_2 = 10, n_3 = 10$ )

Table 8.1. Potency test result for national standard materials of Human Tetanus Immunoglobulin: number of survival mouse per 6 number of group

		Dose		
		Dose1 (0.95 IU/mL)	Dose2 (1 IU/mL)	Dose3 (1.05IU/mL)
기관1	후보물질	0	1	4
	국제표준품	1	4	6
기관2	후보물질	0	1	6
	국제표준품	0	0	6
기관3	후보물질	0	0	6
	국제표준품	0	5	6
가상자료	후보물질	3	4	6
	국제표준품	3	3	6

으로 국가표준품을 확립하여 국내 제약업계에 분양하고, 국가출하승인을 위한 실험에도 사용하고 있다. 이에 매년 추가 확립을 위한 연구사업 뿐 아니라 보관하고 국가표준품의 역가를 확인하여 재확립의 역가의 추세를 확인하고 있다.

Table 8.1에는 식품의약품안전처를 포함한 3개의 기관에서 항과상품 면역글로불린 국가표준품을 확립하기 위한 시험자료의 일부가 제시되어있다. 각 용량별로 6마리의 마우스를 이용해 시험하고 생존한 마우스수로 생물학적인 활성도인 역가(Potency)를 결정하는 시험이다. 일반적으로 후보물질의 역가는 역가를 알고 있는 물질을 이용하여 후보물질의 상대적인 역가의 비를 이용하여 계산되어진다. 이러한 자료에서 중요한 관심 사항 중의 하나는 용량이 증가함에 따라 생존할 확률이 증가하는지의 여부이다.

Table 8.2에는 Table 8.1에 제시된 자료들을 본 논문에서 언급한 다양한 방법으로 검정한 결과들이 제

**Table 8.2.** Exact  $p$ -value(%) for conditional and unconditional exact test method for Table 6.2 data

		Asymptotic test		Unconditional exact test		Conditional exact test	
		$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$	$\bar{X}^2$	$\bar{G}^2$
기관1	후보물질	1.01	1.63	1.04	0.79	1.12	1.12
	국제표준품	0.45	0.41	0.33	0.30	0.32	0.32
기관2	후보물질	0.03	0.01	0.01	0.01	0.02	0.02
	국제표준품	0.01	<0.01	<0.01	<0.01	<0.01	<0.01
기관3	후보물질	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	국제표준품	0.03	0.01	0.01	0.01	0.02	0.02
가상자료	후보물질	5.45	7.01	6.32	4.15	5.67	5.67
	국제표준품	3.95	4.33	4.60	2.80	9.29	5.17

시되어 있다. 일반적인 역가시험의 자료의 경우, 독소량이 다름에 따라 생존추세가 명확하게 나타나 모든 분석법에서 유의한 결과를 나타냈다. 하지만 결과가 명확한 추세가 나타나지 않았을 경우, 분석방법에 따라 결과가 달라질 수 있다. 결과가 명확한 추세로 나타나지 않는 가상의 자료를 만들어 분석해 보았을 때, 유의수준 0.05에서 검정방법에 따라 다른 결과를 보이는 것을 확인할 수 있었다.

## 9. 맺음말

범주형 자료에서 순서화된 대립가설 검정하는 경우 만일 표본크기가 작거나 불균등하다면 대표본이론에 근거한 근사검정방법보다는 정확검정이 더 권고된다.

그 첫 번째 이유는 정확검정은 어떠한 경우에도 제 1종 오류 확률을 유의수준 5% 이하로 통제하는 것을 보장해 주기 때문이다 (Kang, 2002). 반면에 대표본 이론에 근거한 근사검정방법의 제 1종 오류는 5%보다 커지는 경우가 발생할 수 있다. 물론 이런 경우에는 근사검정방법이 더 많은 제 1종 오류를 범하므로, 근사검정방법의 검정력이 정확검정의 검정력보다 더 높을 가능성이 크다. 하지만 두 통계검정 방법을 통계학 이론 측면에서 비교할 때 우선 순위를 두어야 할 것은 검정력이 아니라 제 1종 오류의 유의수준 이하 통제여부이다.

두 번째 이유는 대표본 이론에 근거한 근사검정방법의 제 1종 오류는 5%보다 매우 작아지는 경우가 발생할 수도 있는데, 이 경우 정확검정은 제 1종 오류 확률을 유의수준 5% 이하로 통제하는 것을 보장해 줄 뿐만 아니라, 정확검정의 검정력이 근사검정방법의 검정력보다 더 높아지게 되어 더 바람직하기 때문이다.

정확검정에는 조건부 정확검정과 무조건부 정확검정이 있으며 또한 사용되는 검정통계량에 따라 매우 다양한 정확검정을 만들어 낼 수 있다. 이처럼 많은 정확검정들 중에 어느 정확검정이 가장 높은 검정력을 가지는지는 표본크기와 대립가설에 속한 모수 값들에 따라 달라질 것이므로, 이에 대한 추가적인 연구가 필요하다고 사료된다.

## References

- Barlow, R. E., Bartholow, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, Wiley, New York.
- Barnard, G. A. (1979). In Contradiction to J. Berkson's Dispraise: Conditional tests can be more efficient, *Journal of Statistical Planning and Inference*, **3**, 181-188.
- Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives, *Biometrika*, **46**, 36-48.

- Cytel (2006). StatXact, version 6.0, Software for Exact Nonparametric Statistical Inference with Continuous or Categorical Data. Cytel Software: Cambridge, MA
- Kang, S. H. (2002). Investigation on exact tests, *The Korean Journal of Applied Statistics*, **15**, 187–199.
- Kang, S. H. and Ahn, C. (2008). Tests for the homogeneity of two binomial proportions in extremely unbalanced  $2 \times 2$  contingency tables, *Statistics in Medicine*, **27**, 2524–2535.
- Mehta, C. R. and Patel, N. R. (1983). A network algorithm for performing Fisher's exact test in  $r \times c$  contingency tables, *Journal of the American Statistical Association*, **78**, 427–434.
- Mehta, C. R., Patel, N. R. and Gray, R. (1985). Computing an exact confidence intervals for the common odds ratio in several  $2 \times 2$  contingency tables, *Journal of the American Statistical Association*, **80**, 969–973.
- Mehta, C. R., Patel, N. R. and Senchaudhuri, P. (1998). Exact power and sample size computations for the Cochran-Armitage trend tests, *Biometrics*, **54**, 1615–1621.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

# 범주형 자료에서 순서화된 대립가설 검정을 위한 정확검정의 개발

남주선<sup>a</sup> · 강승호<sup>b,1</sup>

<sup>a</sup>이화여자대학교 통계학과, 식품의약품안전평가원 바이오생약심사부 생물제제과

<sup>b</sup>연세대학교 응용통계학과

(2013년 3월 15일 접수, 2013년 5월 20일 수정, 2013년 8월 14일 채택)

---

## 요약

범주형 자료에서 순서화된 대립가설을 검정하는 경우는 의학 사회학 경영학 등 다양한 응용분야에서 발생한다. 이러한 검정 방법은 대부분 대표본이론에 근거하여 개발되었다. 하지만 표본크기가 작거나 표본크기가 매우 불균등한 경우 대표본이론에 근거한 검정방법의 제 1종 오류 확률은 목표로 하는 5%와 멀어지는 경우가 많이 발생한다. 본 논문에서는 범주형 자료에서 순서화된 대립가설을 검정하는 경우 표본크기가 작거나 표본크기가 매우 불균등한 경우에 사용될 수 있는 정확검정방법을 소개하고 이에 대한 검정력 및 정확  $p$ -value를 제시할 것이다.

주요용어: 정확검정, 범주형 자료, 순서화 가설.

---

---

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0009224). 본 논문의 내용은 저자들의 개인적인 의견으로 우리나라 심사기관의 공식의견을 반영하는 것은 아니다.

<sup>1</sup>교신저자: (120-749) 서울 서대문구 성산로 262, 연세대학교 응용통계학과, 교수.

E-mail: seungho@yonsei.ac.kr