제 4장 검정

§4.1 서론 및 용어

1장에서는 통계학의 기본적인 틀을 언급했고, 2장에서는 통계학에 필요한 확률이론을 다루었다. 그리고, 3장부터 정식으로 통계학을 논하고 있는데, 3장에서는 미지의 모수에 대한 추정을 다루었고 4장에서는 미지의 모수에 대한 가설을 검정(檢定)한다.

<비고 4.1.1> 더러 가설검정 (test of hypothesis)을 가설검증으로 표현하기도 하는데, 공인된 용어는 가설검정이다.

통계학에 필요한 확률이론을 2장에서 다루었다고 했으나 실제로 3장에서 쓰인 것은 2장 내용의 일부에 지나지 않는다. 이는 통계학 과목의 전반부에서 다루는 확률이론이 후반부만을 위한 것이 아니라 기타 관련과목 (예: OR, 신뢰성 공학, 생산관리 등)을 위한 준비과정의 역할도 담당하고 있기 때문이다. 4장은 2장보다 오히려 3장과 관계가 깊다. 예를 들어, 3장에서 신뢰구간을 얻을 때 사용했던 PQ(pivotal quantity)가 이제 4장에서는 검정통계량(test statistic)의 역할을 한다. 또한, 3장에서와 같이 4장에서도 LF(likelihood function: 우도함수)가 핵심적인 역할을 한다. §1.6에서 이미 보았듯이, LF가 3장에서는 MLE 및 MVUE에 대한 근거가 되었고, 4장에서는 LRT(likelihood ratio test: 우도비검정)의 근거가 된다.

배심원 평결을 예로 들어서 가설검정의 배경에 깔린 기본적인 틀을 설명하고, 아울러 앞으로 사용할 용어를 정의한다. 첫째로, 가설은 두 가지인데, 귀무(null)가설은 H_0로 대립(alternative)가설은 H_a로 표기한다. 배심원 평결에서 H_0와 H_a는 다음과 같다.

$$\begin{aligned} &H_0 : (\text{피고는})\ \text{무죄} \\ &H_a : (\text{피고는})\ \text{유죄} \end{aligned} \tag{4.1.1}$$

<비고 4.1.2> 고등학교 과정에서는 흔히 귀무가설 하나만 등장하는데, 이때 대립가설 은 □ H_a : $Not\ H_0$ □이다.

<비고 4.1.3> 책에 따라 H_a를 H_1으로 표기하기도 한다.

재판은 피고 (또는 피고측 변호사)가 피고의 무죄를 입증하기 위해서 하는 것이 아니라, 원고 (또는 검사)가 피고의 유죄를 입증하기 위해서 하는 것이다. 따라서, 재판이 없으면 유죄판결도 있을 수 없으며, 또한 최종 판결이 내려질 때까지는 피고를 죄인취급하지 않는다. 일반적으로, 비용과 시간을 들여서 표본을 추출하고 이를 근거로 가설검정을 하는 이유는 귀무가설을 입증하기 위한 것이 아니라 대립가설을 입증하기 위한 것이다. 즉, 구태여 검정을 하지 않더라도 통념적으로 (또는 보편적으로) 받아들여지는 것이 귀무가설이다. 예를 들어, 품질관리에서 기계의 정상적인 가동상태는 귀무가설로, 비정상 가동상태는 대립가설로 설정한다.

배심원 평결의 결과는 □무죄평결□과 □유죄평결□의 두 가지 인데, 무죄평결을 □귀무가설을 채택□(accept)한다고 하고 유죄평결을 □귀무가설을 기각□(reject)한다고 표현한다.

<비고 4.1.4> □대립가설을 채택□한다는 표현은 잘 쓰이지 않음.

배심원 평결은 물론 100% 정확한 것이 아니다. 두 가지의 오류가 가능한데, 제 1종 오류(type Ⅰ error)라 불리는 것은 □무죄인 피고에게 유죄평결을 내리는 것□이고, 제 2종 오류(type Ⅱ error)라 불리는 것은 □유죄인 피고에게 무죄평결을 내리는 것□이다. 그리고 이러한 오류를 저지를 확률을 각각 α와 β로 표기한다. 즉,

$$\begin{aligned} \alpha &= p(reject\ H_0 \mid H_0\ in\ true) \\ \beta &= p(accept\ H_0 \mid H_0\ in\ false) \end{aligned} \tag{4.1.2}$$

인데, 이때 α를 유의수준(significance level)이라 한다.

α와 β는 작을수록 좋다. 배심원 평결에서는 물적 증거와 증언이 많을수록 그리고 가설검정에서는 표본의 크기 n이 클수록 α와 β가 작아진다. 그러나 표본의 크기가 일정할 때에는 α를 감소시키면 β가 증가하고 반면에 β를 감소시키면 α가 증가한다. 극단적으로, 모든 피고에게 무죄평결을 내리면 $\alpha=0$이지만 이때 β는 최대가 된다. 반면에, 모든 피고에게 유죄평결을 내리면 $\beta=0$이지만 이때 억울한 사람이 가장 많아진다. 따라서, 표본의 크기가 일정할 때에는 α와 β간의 절충이 필요한데, 보통 α를 더 작게 잡는다. (예:

10명의 죄인에게 무죄평결을 내리는 한이 있더라도 한명의 억울한 사람은 없도록 한다는 것이 현대 민주국가의 법철학임.)

그런데, 일반적인 가설검정의 상황에서는 β를 계산하기 어렵거나 또는 β를 딱부러지게 정의하기가 어려운 경우가 대부분이다. 따라서, 관행상 α를 먼저 책정하는데, 주로 1%, 5%, 10%의 세가지 중에서 하나를 선택한다.

<비고 4.1.5> 유의수준으로 $\alpha = 0.05$가 가장 많이 쓰이는데, 이는 신뢰수준으로 95%가 가장 많이 쓰이는 것과 같은 맥락임.

증거 및 증언은 표본정보에 해당된다고 했는데, 이제 관계법규 및 판례에 해당되는 검정통계량이 필요하다. 또한, 예산과 시간적인 제약 때문에 효율적인 재판의 진행이 요구되는데, 이것이 바로 우리가 사용할 LRT에 해당된다. 즉, LRT방법은 필요한 검정통계량을 제공해 줄 뿐만 아니라 주어진 n과 α에 대해서 β를 최소가 되게하는 효율적인 검정법으로 알려져 있다.

<비고 4.1.6> "$1-\beta$"를 검정력(power of test)이라 한다. 즉, 검정력은 "틀린 귀무가설을 기각할 확률"인데, 주어진 n과 α에 대해서 검정력이 최대(most powerful)인 검정법이 바로 LRT이다.

마지막으로, 유의할 점은 검정결과에 대한 해석이다. 피고가 무죄인지 유죄인지는 (본인 외에는?) 아무도 모른다. 다만, 관계법규 및 판례에 비추어 볼 때 증거 및 증언이 피고에게 충분히 (또는, 결정적으로) 불리하면 유죄판결을 내린다. 반면에, 증거 및 증언이 피고에게 (유리한 경우뿐만 아니라) 다소 불리하더라도 "증거불충분"인 경우에는 무죄판결을 내린다. 마찬가지로, 귀무가설이 참인지 아닌지는 아무도 모른다. 따라서, 귀무가설을 기각할 때에는 "귀무가설을 기각할 만한 충분한 근거가 (표본정보에 담겨) 있음"이라 하고, 귀무가설을 채택할 때에는 "귀무가설을 기각할 만한 충분한 근거가 없음"이라 표현한다.

§4.2 LRT

β를 계산하기 쉬운 문제를 예로 들어서 LRT를 설명한다. 다음날 종합주가지수가 올라갈지 내려갈지를 잘 알아맞히기로 소문난 점쟁이가 있는데, 본인의 주장에 의하면 적중률이 70%라고 한다. 이를 확인하기 위해서 앞으로 열흘간에 걸쳐서 몇번을 맞추는지 알아보기로 하자. 적중률에 대한 가설은 다음과 같이 설정한다.

$$\begin{aligned} H_0 &: \ p=0.5 \\ H_a &: \ p=0.7 \end{aligned} \tag{4.2.1}$$

확률변수 Y_i를 다음과 같이 정의한다 ($i=1,\cdots,10$).

$$Y_i = \begin{cases} 1\ , & \text{if } i\text{번째 날에 맞춤} \\ 0\ , & \text{if } i\text{번째 날에 못 맞춤} \end{cases}$$

$Y_1,\cdots,Y_{10}$은 독립이 아닐 가능성이 다분히 있지만, 편의상 독립으로 간주한다. 그러면, $X=\sum_{i=1}^{10} Y_i$는 이항분포를 따르므로, LF는 $P(X=x)=\binom{10}{x}p^x q^{10-x}$이다 (단, $q=1-p$). 이제, L_0와 L_a를 다음과 같이 정의한다.

$$\begin{aligned} L_0(x) &= P(X=x|\,H_0 \ is\ true) = \binom{10}{x}(0.5)^{10} \\ L_a(x) &= P(X=x|\,H_a \ is\ true) = \binom{10}{x}(0.7)^x(0.3)^{10-x} \end{aligned} \tag{4.2.2}$$

아래의 표는 모든 가능한 x값에 대해서 $L_0(x)$와 $L_a(x)$ 그리고 이들의 비율인 $L_0(x)/L_a(x)$를 구한 것이다.

$L_0(x)$는 $x=5$일 때 최대이고, $L_a(x)$는 $x=7$일 때 최대이다. 반면에 LR(likelihood ratio)인 “$L_0(x)/L_a(x)$”는 x가 증가함에 따라 계속 감소한다.

LRT는 다음과 같다.

$$\textit{Reject } H_0 \quad \text{if} \quad \frac{L_0(x)}{L_a(x)} < k \tag{4.2.3}$$

k는 α값에 의해 정해지는 양의 상수인데, 이러한 판정법(decision rule)을 따르면 β가 최소가 된다고 알려져 있다. 그런데, 이 문제에서는 x가 증가함에 따라 LR이 계속 감소하므로 판정법은 결국 x가 어떤 기준치보다 크면 귀무가설을 기각하고, 기준치 이하이면 채택하라는 것이 된다. 사실 이는 상식적으로 납득이 가는 당연한 결과이다.

x	0	1	2	3	4	5	6	7	8	9	10
$L_0(x)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$L_a(x)$	.000	.000	.002	.009	.036	.103	.200	.267	.234	.121	.028
$\frac{L_0(x)}{L_a(x)}$	165	70.9	30.4	13.0	5.58	2.39	1.02	.439	.188	.081	.035

<비고 4.2.1> 식 (4.2.2)에서 $\binom{10}{x}$를 누락시키더라도 식 (4.2.3)에는 변화가 없음. 즉, 복원추출시 추출하는 순서를 따지든 안 따지든 LRT의 결과와 동일함 (<비고 3.1.2> 참조).

그러나, 문제가 복잡해지면 일일이 위와 같은 표를 만들어서 LR의 값을 직접 눈으로 확인하기가 어렵다. 따라서, 일반적인 문제에서는 식 (4.2.3)로부터 구체적인 판정법을 얻어야 되는데, 다음과 같이 시범을 보인다. (표본의 크기를 n, 귀무가설을 H_0: $p=p_0$, 대립가설을 H_a: $p=p_a$라 함. 단, $p_0<p_a$이고 $q_0=1-p_0$, $q_a=1-p_a$ 임.)

$$\frac{L_0(x)}{L_a(x)} = \frac{\binom{n}{x} p_0^x\, q_0^{n-x}}{\binom{n}{x} p_a^x\, q_a^{n-x}} = \left(\frac{p_0 q_a}{p_a q_0}\right)^x \left(\frac{q_0}{q_a}\right)^n < k$$

$$\rightarrow \left(\frac{p_0 q_a}{p_a q_0}\right)^x < k' = \frac{k}{(q_0/q_a)^n}$$

$$\rightarrow x \ln\left(\frac{p_0 q_a}{p_a q_0}\right) < k'' = \ln k'$$

$$\rightarrow x > k''' = \frac{k''}{\ln\left(\frac{p_0 q_a}{p_a q_0}\right)} \tag{4.2.4}$$

(비고: 식 (4.2.4)에서 부등호의 방향이 바뀌는 이유는 $p_0<p_a$, $q_a<q_0$ 이므로 $(p_o q_a/p_a q_0)<1$ 이고 따라서 $\ln(p_0 q_a/p_a q_0)<0$ 이기 때문임.) 즉, x값이 상수 k'''보다 크

면 귀무가설을 기각한다는 판정법이 LRT로부터 도출된다. 아울러, (4.2.4)의 x에 대응되는 확률변수 X가 검정통계량이 되는데, 이는 바로 최소충분통계량이다 (§3.4 참조).

<비고 4.2.2> 식 (4.2.4)를 기각역(rejection region)이라 함.

이제, 식 (4.2.4)의 k'''값을 변화시킴에 따라 α와 β가 어떻게 달라지는지 알아보자. 식 (4.1.2)에 의해서

$$\begin{aligned} \alpha &= P(X > k''' \mid p=0.5) = \sum_{x=k'''+1}^{10} \binom{10}{X} (0.5)^{10} \\ \beta &= P(X \le k''' \mid p=0.7) = \sum_{x=0}^{k'''} \binom{10}{X} (0.7)^{x} (0.3)^{10-x} \end{aligned} \tag{4.2.5}$$

인데, 아래의 표는 $k'''=0,\cdots,10$에 대해서 α와 β를 계산한 것이다.

k'''	0	1	2	3	4	5	6	7	8	9	10
α	.999	.989	.945	.828	.623	.377	.172	.055	.011	.001	.000
β	.000	.000	.002	.011	.047	.150	.350	.617	.851	.972	1.00

예를 들어, “$x>6$”을 기각역으로 사용하면 (비고 <4.2.2> 참조) $\alpha=0.172$이고 $\beta=0.350$이다. 그러나, α를 더 줄이기 위해서 “$x>7$”을 기각역으로 사용하면 $\alpha=0.055$이지만 $\beta=0.617$이 된다.

역시, 문제가 복잡해지면 위와 같은 표를 만들기가 어렵다. 예를 들어, 위의 문제에서 $n=100$이라 하자. 즉, 앞으로 100일 간에 걸쳐서 점쟁이가 몇 번을 맞히는지 알아본다고 하자. 이 경우, 중심극한정리를 이용해서 X의 분포(인 이항분포)를 정규분포로 근사하면

$$Z \equiv \frac{X-np}{\sqrt{npq}} \overset{A}{\sim} N(0,1) \tag{4.2.6}$$

이 된다. 일반적인 문제에서는 α값을 먼저 책정한다고 했다. 이제, $\alpha=0.05$로 했을 때 (<비고 4.1.5> 참조) 기각역이 어떻게 되는지 알아보자. 식 (4.2.5)와 (4.2.6)에 의해서

$$\begin{aligned} 0.05 &= P(X > k''' \mid p=0.5) \\ &= P\left(Z > \frac{k'''-(100)(0.5)}{\sqrt{(100)(0.5)(0.5)}}\right) \end{aligned}$$

인데, $P(Z>1.645)\approx 0.05$ 이므로 $k''\approx 58.225$ 를 얻는다. 따라서, 유의수준 5% 에서의 기각역은 $x\geq 59$ 이다. 즉, 100일 중에 59번 이상을 맞추면 귀무가설을 기각하고, 58번 이하를 맞추면 귀무가설을 채택한다. 이때 β값은 식 (4.2.5)와 (4.2.6)에 의해서 다음과 같다.

$$
\begin{aligned}
\beta &= P(X\leq 58 \mid p=0.7) \\
&= P\left(Z\leq \frac{58-(100)(0.7)}{\sqrt{(100)(0.7)(0.3)}} = -2.62\right) \\
&\approx 0.0044
\end{aligned}
$$

만약, 유의수준 α를 1% 로 책정했더라면 기각역은 $x\geq 62$가 되고 β값은 2.5% 가 된다.

§4.3 검정의 종류

먼저 모수공간(parameter space)과 표본공간(sample space)을 정의한다. 모수공간이란 미지의 모수가 가질 수 있는 모든 값들의 집합이다. 예를 들어, 식 (4.2.1)에서는 모수가 p 하나인데, 그나마 $p=0.5$ 와 $p=0.7$ 두가지의 값만 고려하였으므로 모수공간은 $\{0.5, 0.7\}$ 이다. 그러나, 일반적인 비율 검정 문제에서는 모수공간이 $\{p: 0 \le p \le 1\}$ 이다. 또한, 정규 모분포의 경우 모수공간은 일반적으로 2차원 공간인 $\{\mu, \sigma^2: -\infty < \mu < \infty, \sigma^2 > 0\}$ 이다.

반면에, 표본공간은 n 차원 공간인데, 관찰된 표본 $\{y_1, \cdots, y_n\}$ 은 n 차원 공간에서 하나의 점에 해당된다. 그리고, 표본공간을 두 부분으로 나누어서 그 중 하나를 기각역으로 사용하는 것이다. 그런데, 이 n 차원 공간을 좌표변환하면 기각역이 간단히 표현되기도 하는데, 예를 들어 §4.2에서는 기각역이 $x = \sum_{i=1}^{n} y_i$ 하나로만 표현되었다 (§3.4의 “최소충분통계량의 의미” 참조).

모수공간의 정의에 따라 검정의 종류가 달라지는데, 이를 §4.2의 예제로 설명한다. 귀무가설은 “$p=0.5$”이고 대립가설은 “$p=0.7$”인데, 이러한 가설을 단순(simple) 가설이라 한다. 만약에, 대립가설이 “$p>0.5$”였다면 이를 복합(composite) 대립가설이라 하는데, 이 경우 모수공간은 $\{p: p \ge 0.5\}$ 가 된다. 이때 유의할 점은 다음과 같다. 식 (4.2.4)는 $p_a = 0.7$ 뿐만아니라 0.5 보다 큰 모든 p_a값에 대해서 성립한다. 따라서, 대립가설이 “$p>0.5$”이었더라도 기각역은 여전히 동일하다. (예: $n=100$, $\alpha=0.01$ 이면 기각역은 $x \ge 62$.) 다만, β값을 계산하기가 애매해진다. 대립가설이 단순가설일 때에는 β가 하나의 값이지만, 이제는 $p_a \in H_a$인 모든 p_a에 대해서 β값이 하나씩만 있으므로 β는 p_a의 함수가 된다. 이를 $\beta(p_a)$로 표현하면, 예를 들어 $n=100$, $\alpha=0.01$ 일 때 $\beta(0.7)=0.025$ 이다.

<비고 4.3.1> 식 (4.2.4)는 p_0보다 큰 모든 p_a에 대해서 성립하는데, 각각의 p_a에 대해 LRT에 의한 $\beta(p_a)$가 최소이므로 (<비고 4.1.6> 참조), 복합 대립가설 “H_a: $p > p_0$”의 경우에는 LRT를 UMPT(uniformly most powerful test)라 부른다. 이때 “uniformly”는 (p_0보다 큰) “모든 p_a에 대해서 골고루” 라는 뜻이다.

식 (4.2.4)와 같이 기각역이 “ $x>c$” 형태인 경우를 UTT(upper-tail test)라 하는데, 이는 검정통계량인 X (또는 식 (4.2.6)의 Z)의 분포에서 우측 꼬리부분이 기각역에 해당되기 때문이다. 반면에, 만약 귀무가설은 여전히 “ $p=p_0$ ”이나 대립가설이 “ $p<p_0$”였다면 기각역은 “ $x<c$” 형태가 되는데, 이 경우는 LTT(lower-tail test)라 한다. 그리고, UTT와 LTT를 합쳐서 OTT(one-tail test)라 한다.

지금까지 등장한 OTT는 UMPT라는 장점이 있다 (<비고 4.3.1> 참조). 반면에, 모수공간이 모든 가능한 모수값을 포함하지 않는다는 단점이 있다. 예를 들어, UTT의 모수공간은 $\{p: p\geq p_0\}$이고 LTT의 모수공간은 $\{p: p\leq p_0\}$이다. 모수공간이 $\{p: 0\leq p\leq 1\}$가 되는 대표적인 경우는 두가지가 있다. 첫째는 귀무가설을 복합가설로 설정하는 경우인데, 예를 들면 “ H_0: $p\leq p_0$”이고 “ H_a: $p>p_0$”인 경우이다. 이때, 등호 “ = ”는 반드시 귀무가설에 포함시키는데, 이는 α를 계산할 때 p_0값을 사용할 수 있게 하기 위해서이다. 물론, (복합 대립가설의 경우에 β가 H_a에 속한 p값들의 함수이듯이) 복합 귀무가설의 경우에는 α도 H_0에 속한 p값들의 함수가 된다. 그러나, $p=p_0$일 때 α값이 최대가 되는데, 이를 기준치로 사용한다. 예를 들어, H_0: $p\leq 0.5$, H_a: $p>0.5$, $n=100$인 경우에 $\alpha(0.5)$를 0.01로 책정하면 기각역은 $x\geq 62$가 되는데, 0.5보다 작은 p값에 대해서는 $\alpha(p)<0.01$이다.

<비고 4.3.2> 복합 귀무가설 “ H_0: $p\leq p_0$”와 복합 대립가설 “ H_a: $p>p_0$”의 경우에도 (최소한 이 책에 등장하는 모분포에 대해서는) 식 (4.2.4)의 LRT가 UMPT이다 (<비고 4.3.1> 참조). 따라서, UTT 경우 “ H_0: $p=p_0$”를 “ H_0: $p\leq p_0$”로 간주해도 별로 무리가 없다.

모수공간을 $\{p: 0<p<1\}$이 되게하는 두번째 경우는 “ H_0: $p=p_0$”이고 “ H_a: $p\neq p_0$” 인 경우이다. 이는 가장 많이 사용되는 형태의 가설검정인데 (<비고 4.1.2> 참조), 결론부터 언급하면 기각역이 “ $x>c_1$ 또는 $x<c_2$” 형태라서 이를 TTT(two-tailed test)라 한다. 즉, 검정통계량의 분포에서 양쪽 꼬리부분이 모두 기각역에 해당된다.

TTT역시 LRT로 시행한다. 편의상, H_a에 포함된 p값을 p_a라 부르자. 즉, $p_a\in H_a$인데 H_a가 “ $p\neq p_0$”이므로 p_a는 p_0보다 클수도 있고 작을수도 있다. 그런데, 식 (4.2.4)에서 $p_a>p_0$이면 기각역은 $x>c_1$의 형태이고 $p_a<p_0$이면 기각역은 $x<c_2$의 형태가 된다. 따라서, 식 (4.2.5) 대신에 (단, $c_1>c_2$)

$$\alpha = P(X > c_1 \text{ 또는 } X < c_2 \mid p = p_0)$$
$$= P(X > c_1 \mid p = p_0) + P(X < c_2 \mid p = p_0)$$

를 얻는데, 관행상 α를 이등분하여

$$\frac{\alpha}{2} = P(X > c_1 \mid p = p_0) = P(X < c_2 \mid p = p_0) \qquad (4.3.1)$$

를 사용한다. 예를 들어, $p_0 = 0.5$이고 $n = 100$일 때 α를 10%로 책정하면 ($\alpha/2$는 앞에서와 같이 5%가 되므로) $c_1 \approx 50 + 8.225 = 58.225$이고 $c_2 \approx 50 - 8.225 = 41.775$이다. 따라서, 기각역은 "$x \geq 59$ 또는 $x \leq 41$"이다.

TTT는 가장 많이 사용되는 검정법이지만 UMPT는 아니다 (<비고 4.3.1>, <비고 4.3.2> 참조). 위의 예에서 보았듯이, TTT에서는 표본공간을 세 부분으로 나누어서 그 중 둘을 기각역으로 사용하는데, 이때 α를 이등분해서 기각역 별로 $\alpha/2$씩 할당한다. 그러니까, OTT와 동일한 검정력을 얻기 위해서는 α값을 두배로 책정해야 된다. 따라서, (동일한 n과) 동일한 α값을 사용하면 TTT의 검정력은 OTT보다 못하다.

그러나, 이와 같이 TTT를 OTT와 비교하는 것은 사실 공평하지 못하다. 이들 두 방법은 엄연히 서로 다른 상황에 사용되는 서로 다른 방법일 뿐이지, 각각이 사용되는 각각의 상황에서는 각각 나름대로 최적(optimal)의 방법이라고 할 수 있다.

선거철에 각 정당에서 발표하는 여론조사 결과를 예로 들자. 만약에 정당마다 자기에게 유리한 결과를 발표한다면 이는 의심을 받을만하다. 그 이유는, 여론조사 결과가 마음에 들면 이를 발표하고, 마음에 들지 않으면 이를 묵살할 수도 있기 때문이다. 이와 유사한 현상이 TTT와 OTT간에도 발생할 수 있다. 가설검정의 목적과 상황에 비추어 TTT가 적합하다고 하자. 예를 들어, 보편적으로 받아들여지고 있는 사회적인 통념을 "H_0: $p = 0.5$"라 하고, 이 통념이 옳지 않다는 것을 보이기 위해서 "H_a: $p \neq 0.5$"인 TTT를 한다고 하자. 앞의 예에서와 같이 $n = 100$, $\alpha = 0.10$이면 기각역은 $x \geq 59$ 또는 $x \leq 41$이다. 그런데, 만약 표본을 먼저 추출해서 x값을 미리 확인한 다음에 x가 50보다 크면 대립가설을 "$p > 0.5$"로 고치고, x가 50보다 작으면 대립가설을 "$p < 0.5$"로 고친다고 하자. 즉, 관찰된 표본정보에 따라서 TTT를 UTT 또는 LTT로 바꾸는 것인데, 이렇게 하면 (동일한 검정력에 대해서는) α가 절반인 0.05로 줄거나 또는 (동일한 $\alpha = 0.10$에 대해서는) 검정력이 증가한다. 그러나, 이는 정당한 방법이 아니라 일종의 사기라고 할 수 있다.

§4.4 정규 모분포 관련

4.4.1 진짜 LRT

지금까지 이 책에서 LRT라 부른 것을 (대부분의) 다른 책에서는 LRT라 부르지 않는다. 이 책에서는 LF(likelihood function)의 비율인 LR(likelihood ratio)에 근거한 검정법은 모두 LRT라 부르고 있지만, 공인된 용어에 따른 진짜 LRT는 §4.4.2에 처음으로 등장한다.

진짜 LRT는 주로 (정규 모분포 경우와 같이) 모수공간이 이차원(이상)인 경우에 사용하는데, LR의 형태는 식 (1.6.5)와 같다. 그러나, 이를 모수공간이 일차원인 경우에도 적용할 수 있다. 예를 들어, §4.2와 §4.3에 등장한 예제에서 식 (4.2.3) 대신에 식 (1.6.5) 형태의 LR를 사용해도 동일한 기각역을 얻는다.

먼저, “ H_0: $p=0.5$, H_a: $p\neq 0.5$ ”이고 “ $n=10$ ”인 경우를 다룬다. LF는 $P(X=x)=\binom{10}{x}p^x q^{10-x}$인데, $L(p_0)$와 $L(\hat{p})$을 다음과 같이 정의한다 (식 (4.2.2) 참조).

$$\begin{aligned} L(p_0)&=P(X=x \mid p=p_0)=\binom{10}{x} p_0^x\, q_0^{10-x} \\ L(\hat{p})&=P(X=x \mid p=\hat{p})=\binom{10}{x}\, \hat{p}^x\, \hat{q}^{10-x} \end{aligned} \qquad (4.4.1)$$

식 (4.4.1)에서 p_0는 귀무가설 하에서의 p값인 0.5이고 ($q_0=1-p_0=0.5$), $\hat{p}$은 p에 대한 최우추정치인 $x/10$이다 ($\hat{q}=1-\hat{p}$). 아래의 표는 모든 가능한 x값에 대해서 $L(p_0)$와 $L(\hat{p})$ 그리고 이들의 비율을 구한 것이다.

x	0	1	2	3	4	5	6	7	8	9	10
$L(p_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$\hat{p}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$L(\hat{p})$	1	.387	.302	.267	.251	.246	.251	.267	.302	.387	1
$\frac{L(p_0)}{L(\hat{p})}$	.001	.025	.146	.439	.818	1	.818	.439	.146	.025	.001

최우추정치 $\hat{p}$는 (정의상) LF를 최대가 되게하는 p값이므로, LR인 $L(p_0)/L(\hat{p})$의 최대치는 1이다 (최소치는 0). 그리고, LR이 1에 가까울수록 귀무가설의 설득력이 강하고, 0에 가까울수록 약하다. 따라서, 기각역의 형태는

$$\frac{L(p_0)}{L(\hat{p})} \le k \qquad (4.4.2)$$

가 되는 것이 바람직한데, 이는 §4.3에서 다룬 TTT와 동일한 것이다. 예를 들어, $\alpha = 2 \cdot (0.172) = 0.344$ 이면 기각역은 " $x \ge 7$ 또는 $x \le 3$ "이고, $\alpha = 2 \cdot (0.055) = 0.11$ 이면 기각역은 " $x \ge 8$ 또는 $x \le 2$ "가 된다 (§4.2의 표 참조).

다음, " H_0: $p = 0.5$, H_a: $p > 0.5$ "이고 " $n = 10$ "인 경우를 다룬다. 이때 유의할 점은 다음과 같다. 모수공간은 $\{p\colon p \ge 0.5\}$ 인데, 이는 마치 " $p < 0.5$ "가 불가능하다는 전제조건과 같다. 이에 따라, p에 대한 최우추정치인 $\hat{p}$에 대해서도 " $\hat{p} \ge 0.5$ "만 고려대상이 된다. 구체적으로

$$\hat{p} = \max\left(\frac{x}{10}, 0.5\right) = \begin{cases} \dfrac{x}{10} & \text{if } x \ge 5 \\ 0.5 & \text{if } x < 5 \end{cases} \qquad (4.4.3)$$

인데, 이를 식 (4.4.1)에 대입하면 아래의 표를 얻는다.

x	0	1	2	3	4	5	6	7	8	9	10
$L(p_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$\hat{p}$	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.7	0.8	0.9	1.0
$L(\hat{p})$	.001	.010	.044	.117	.205	.246	.251	.267	.302	.387	1
$\frac{L(p_0)}{L(\hat{p})}$	1	1	1	1	1	1	.818	.439	.146	.025	.001

따라서, 식 (4.4.2)를 적용하면 자동적으로 UTT가 된다. 예를 들어, $\alpha = 0.172$ 이면 기각역은 " $x \ge 7$ "이고, $\alpha = 0.055$ 이면 기각역은 " $x \ge 8$ "이 된다.

마지막으로, " $H_0 \le 0.5$, $H_a > 0.5$ "이고 " $n = 10$ "인 경우를 다룬다. 모수공간은 다시 TTT에서과 같이 $\{p\colon 0 \le p \le 1\}$ 이므로, 최우추정치 역시 $\hat{p} = x/10$ 이다. 그런데, 귀무가설이 복합가설이라서 $L(p_0)$를 계산하기가 애매해진다. §4.3에서는 p_0값으로 복합 귀무가설의 경계치인 0.5를 사용한다고 했는데 (<비고 4.3.2> 참조), 이는 어디까지나 단순 귀무가설의 틀에 맞추어 보려는 임시변통이었을 뿐이다. 사실상 이 예제는 §4.4.2에 등장할 진짜 LRT의 범주에 속하는 것인데, 이를 위해서 식 (4.4.2)를 다음과 같이 확장시킨다.

$$\frac{L(\widehat{p_0})}{L(\hat{p})} \le k \qquad (4.4.4)$$

식 (4.4.4)에서 $\widehat{p_0}$은 복합 귀무가설 하에서의 최우추정치로서 다음과 같다 (식 (4.4.3) 참조).

$$\widehat{p_0} = \min\left(\frac{x}{10}, 0.5\right) = \begin{cases} \dfrac{x}{10} & \text{if } x \le 5 \\ 0.5 & \text{if } x > 5 \end{cases} \tag{4.4.5}$$

<비고 4.4.1> 단순 귀무가설 H_0: $p = p_0$ 경우 $\widehat{p_0} = p_0$ 인데, 이는 귀무가설 하에서 p가 취할 수 있는 값이 p_0하나 뿐이기 때문이다. 따라서, 식 (4.4.2)는 식 (4.4.4)의 특수한 경우이다.

아래의 표는 모든 가능한 x값에 대해서 $L(\widehat{p_0}) = \binom{10}{x} \widehat{p_0}^{\,x}\ \widehat{q_0}^{\,10-x}$와 $L(\hat{p}) = \binom{10}{x} \hat{p}^{x}\ \hat{q}^{10-x}$ 그리고 이들의 비율을 구한 것이다.

x	0	1	2	3	4	5	6	7	8	9	10
$\widehat{p_0}$	0.0	0.1	0.2	0.3	0.4	0.5	0.5	0.5	0.5	0.5	0.5
$L(\widehat{p_0})$	1	.387	.302	.267	.251	.246	.205	.117	.044	.010	.001
$\hat{p}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$L(\hat{p})$	1	.387	.302	.267	.251	.246	.251	.267	.302	.387	1
$\frac{L(\widehat{p_0})}{L(\hat{p})}$	1	1	1	1	1	1	.818	.439	.146	.205	.001

그런데, 유의할 점은 LR인 "$L(\widehat{p_0})/L(\hat{p})$"의 값이 단순 귀무가설 "$H_0$: $p = 0.5$, H_a: $p > 0.5$" 경우의 LR인 "$L(p_0)/L(\hat{p})$"의 값과 동일하다는 점이다. 따라서, 동일한 UTT가 된다 (<비고 4.3.2> 참조).

4.4.2 μ에 대한 검정

정규 모분포의 평균 μ에 대한 검정을 T-test라 하는데, 이는 검정통계량인

$$T_{n-1} = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}} \tag{4.4.6}$$

가 (자유도가 $n-1$인) t분포를 따르기 때문이다 (§2.15.5 참조). §4.4.1에서 언급했듯이, 검정통계량 T_{n-1}은 식 (4.4.4) 형태의 LRT로부터 얻는다.

먼저, "H_0: $\mu=\mu_0$, H_a: $\mu\neq\mu_0$" 경우를 다룬다. 이때 유의할 점은, 정규 모분포의 모수공간이 이차원이므로 (§4.3 참조), 사실상 "$\sigma^2>0$"가 H_0와 H_a에 포함되어 있는 것이나 마찬가지라는 점이다. 따라서, 귀무가설은 단순가설이 아니라 복합가설 "H_0: $\mu=\mu_0$ and $\sigma^2>0$"인 셈이다. 귀무가설 하에서의 μ와 σ^2에 대한 최우추정치는 다음과 같다 (<비고 4.4.1>과 식 (3.2.5) 참조).

$$\widehat{\mu_0}=\mu_0,\quad \widehat{\sigma_0^2}=\frac{\sum_{i=1}^{n}(y_i-\mu_0)^2}{n} \tag{4.4.7}$$

반면에, "귀무가설 하에서"라는 제약(constraint) 없이 구한 최우추정치는 다음과 같다 (§ 3.2.1 참조).

$$\widehat{\mu}=\overline{y},\quad \widehat{\sigma^2}=\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{n} \tag{4.4.8}$$

식 (4.4.7)과 (4.4.8)을 LF인 식 (3.2.1)에 대입하면

$$L(\widehat{\mu_0},\widehat{\sigma_0^2})=\left(\frac{1}{\sqrt{2\pi}}\right)^n\left(\frac{1}{\widehat{\sigma_0^2}}\right)^{\frac{n}{2}}\exp\left(-\frac{n\widehat{\sigma_0^2}}{2\widehat{\sigma_0^2}}\right)$$
$$L(\widehat{\mu},\widehat{\sigma^2})=\left(\frac{1}{\sqrt{2\pi}}\right)^n\left(\frac{1}{\widehat{\sigma^2}}\right)^{\frac{n}{2}}\exp\left(-\frac{n\widehat{\sigma^2}}{2\widehat{\sigma^2}}\right)$$

를 얻으므로, 식 (4.4.4) 형태의 LRT는 다음과 같다.

$$\lambda\equiv\frac{L(\widehat{\mu_0},\widehat{\sigma^2})}{L(\widehat{\mu},\widehat{\sigma^2})}=\left(\frac{\widehat{\sigma^2}}{\widehat{\sigma_0^2}}\right)^{\frac{n}{2}}\leq k \tag{4.4.9}$$

<비고 4.4.2> 식 (4.4.9)에서 "$0\leq\lambda\leq1$"이다 ($0<k<1$). 그리고, 표본 관찰치 y_i를 이에 대응되는 확률변수 Y_i로 대체하면 λ는 확률변수가 되는데, 이를 (그대로 또는 손질해서) 검정통계량으로 사용한다.

식 (4.4.9)를 손질해서 아래의 식 (4.4.10)을 얻는 과정은 (식 (4.2.4)를 얻는 과정과 유사하지

만 훨씬) 복잡하므로 생략한다 (문헌 [9] 참조).

$$t^2 \equiv \left(\frac{\overline{y}-\mu_0}{\sqrt{\frac{\sum(y_i-\overline{y})^2}{n-1}}/\sqrt{n}} \right)^2 \geq k' = (n-1)(k^{-\frac{2}{n}}-1) \tag{4.4.10}$$

따라서, LRT는 TTT로써 기각역은 $t \geq \sqrt{k'}$ 또는 $t \leq -\sqrt{k'}$이다. 그리고, 식 (4.4.10)의 t에서 y_i를 Y_i로 대체하면 식 (4.4.6)을 얻는다.

<비고 4.4.3> 식 (4.4.6)의 T_{n-1}은 "$\mu=\mu_0$"일 때 자유도가 $n-1$인 t분포를 따른다.

예를 들어, $n=10$일 때 $0.025 = P(T_{n-1} > 2.262) = P(T_{n-1} < -2.262)$이므로, 유의수준 5%에서의 기각역은 $t > 2.262$ 또는 $t < -2.262$이다.

다음, "H_0: $\mu=\mu_0$, H_a: $\mu > \mu_0$" 경우를 다룬다. 이때 유의할 점은 모수공간이 $\{\mu, \sigma^2 \colon \mu \geq \mu_0, \sigma^2 > 0\}$이므로 $\hat{\mu} = \max(\mu_0, \overline{y})$라는 점이다 (식 (4.4.3) 참조). 이에 따라, 식 (4.4.9)와 (4.4.10) 대신에

$$\lambda = \begin{cases} (\widehat{\sigma^2}/\widehat{\sigma_0^2})^{\frac{n}{2}} & \text{if } \overline{y} \geq \mu_0 \\ 1 & \text{if } \overline{y} < \mu_0 \end{cases} \tag{4.4.11}$$

$$t = \frac{\overline{y}-\mu_0}{\sqrt{\frac{\sum(y_i-\overline{y})^2}{n-1}}\Big/\sqrt{n}} \geq \sqrt{k'}\ (>0) \tag{4.4.12}$$

를 얻는데, 이는 UTT의 기각역이다. 그리고, 식 (4.4.12)의 y_i를 Y_i로 대체하면 역시 식 (4.4.6)을 얻는다 (<비고 4.4.3> 참조). 예를 들어, $n=10$일 때 $0.05 = P(T_{n-1} > 1.833)$이므로, 유의수준 5%에서의 기각역은 $t > 1.833$이다.

마지막으로, "H_0: $\mu \leq \mu_0$, H_a: $\mu > \mu_0$" 경우를 다룬다. 이때 달라지는 것은 귀무가설 하에서의 μ에 대한 최우추정치인데, 이는 식 (4.4.5)와 유사한 형태인 $\widehat{\mu_0} = \min(\mu_0, \overline{y})$가 된다. (비고: 식 (4.4.8)은 유효함.) 그런데, 결과론적으로 식 (4.4.11)과 (4.4.12)는 여전히 유효하다. (비고:§4.4.1의 마지막 부분에서도 $L(\widehat{p_0})/L(\hat{p}) = L(p_0)/L(\hat{p})$ 이었음.) 즉, 식 (4.4.6)이 여전히 검정통계량이고 또한 UTT가 된다. 다만, 유의수준과 검정통계량은 귀무가설 "$\mu \leq \mu_0$"에 속한 모든 μ 값들 중에서 경계치인 μ_0를 기준으로 삼은 것으로 간주하면

된다 (<비고 4.3.2> 위의 문단과 <비고 4.4.3> 참조).

4.4.3 기타 정규 모분포 관련

진짜 LRT보다 먼저 등장한 UMPT(<비고 4.3.1> 참조)와 식 (4.2.4)까지 모두 LRT라 하면(§4.4.1 참조), 사실상 모든 검정이 LRT라 해도 과언이 아니다. 그런데, LR인 λ를 손질해서(<비고 4.4.2> 참조) 식 (4.2.6)의 Z와 식 (4.4.6)의 T_{n-1} 같이 눈에 익은 검정통계량을 얻어(내고 아울러 기각역의 형태를 알아)내는 과정은 앞에서 보았듯이 제법 복잡하다. 따라서, 앞으로는 LRT의 결과만을 언급하기로 한다.

첫째로, σ^2이 알려진 경우에는 μ에 대한 검정통계량으로 "$Z = (\overline{Y} - \mu_0)/(\sigma/\sqrt{n})$"을 사용하는데, 이는 물론 $\mu = \mu_0$일 때 $N(0,1^2)$분포를 따른다. 그런데, §4.4.2의 T-test가 이제 Z-test로 바뀌었으므로 t분포의 확률표 대신 $N(0,1^2)$분포의 확률표를 사용하는 것만 달라질 뿐이지, 가설의 형태에 따라 검정의 종류 및 기각역의 형태가 달라지는 양상은 T-test 때와 동일하다.

<비고 4.4.4> $N(0,1^2)$의 확률표는 t분포의 확률표에서 자유도가 ∞인 경우에 해당됨 (<비고 2.6.1> 참조).

다음, σ^2에 대한 검정통계량으로는 μ가 알려진 경우에는 식 (2.15.10)을 그리고 μ를 모르는 경우에는 식 (2.15.11)을 사용하는데 (단, σ^2을 귀무가설의 σ_0^2으로 대체함), 이들은 $\sigma^2 = \sigma_0^2$일 때 자유도가 각각 n과 $n-1$인 카이제곱분포를 따른다. 그러나, 가설의 형태에 따라 검정의 종류 및 기각역의 형태가 달라지는 양상은 μ에 대한 검정때와 동일하다. 예를 들어, μ를 모르는 경우에 $n=10$이라 하자. 만약 귀무가설이 "$\sigma^2 = \sigma_0^2$" 또는 "$\sigma^2 \le \sigma_0^2$"이고, 대립가설이 "$\sigma^2 = \sigma_a^2$" (단, $\sigma_a^2 > \sigma_0^2$) 또는 "$\sigma^2 > \sigma_0^2$"이면, 이는 UTT이고 유의수준 5%에서의 기각역은 $c_{n-1} \ge 16.919$이다. 또한, 귀무가설이 "$\sigma^2 = \sigma_0^2$"이고 대립가설이 "$\sigma^2 \ne \sigma_0^2$"이면, 이는 TTT이고 유의수준 5%에서의 기각역은 $c_{n-1} \ge 19.0228$ 또는 $c_{n-1} \le 2.7004$이다. (비고: c_{n-1}은 검정통계량에서 확률변수 Y_i를 표본 관찰치 y_i로 대체한 것임. 즉, c_{n-1}은 카이제곱분포를 따르는 검정통계량이 구현된(realized) 값을 의미함.)

<비고 4.4.5> 지금 등장한 카이제곱 검정은 UTT, LTT, TTT가 모두 가능하다. 그러나, §

4.6에 등장하는 카이제곱 검정은 모두 UTT이다.

결국 §3.7에서 신뢰구간을 얻을 때 사용되었던 PQ(pivotal quantity)가 이제는 LRT에 따른 검정통계량으로 쓰이고 있는 셈인데 (단, PQ의 μ와 σ^2을 각각 μ_0와 σ_0^2으로 대체함), 이는 나머지 PQ에 대해서도 성립한다.

두 개의 정규 모분포가 있을 때 귀무가설 "$\sigma_1^2=\sigma_2^2$"를 검정하는 통계량은 식 (2.15.14)인데, 이는 귀무가설 하에서 (단, μ_1과 μ_2를 모를 때)

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$$

이다. 그런데, 귀무가설 "$\sigma_1^2=\sigma_2^2$"은 "$(\sigma_1^2/\sigma_2^2)=1$"과 동일하므로, 대립가설이 "$\sigma_1^2>\sigma_2^2$" 또는 "$(\sigma_1^2/\sigma_2^2)>1$"이면 UTT가 되고 대립가설이 "$\sigma_1^2\neq\sigma_2^2$" 또는 "$(\sigma_1^2/\sigma_2^2)\neq 1$"이면 TTT가 되는 것이 쉽게 수긍이 간다. 물론, 대립가설이 "$\sigma_1^2<\sigma_2^2$" 또는 "$(\sigma_1^2/\sigma_2^2)<1$"이면 LTT가 된다. 그리고, 예를 들어 $n_1=10, n_2=5, \alpha=0.05$일 때 TTT의 기각역은 8.90이상 또는 $(4.72)^{-1}$ 이하이다 (§3.7.2 참조).

<비고 4.4.6> 지금 등장한 F-test는 UTT, LTT, TTT가 모두 가능하다. 그러나, 5장 이후에 등장하는 F-test는 모두 UTT이다. (<비고 4.4.5> 참조.)

두 개의 정규 모분포가 있을 때 귀무가설 "$\mu_1=\mu_2$"를 검정하는 통계량으로는 σ_1^2과 σ_2^2이 알려져 있으면 식 (3.7.1)을 그리고 σ_1^2과 σ_2^2을 모르는 경우에는 식 (3.7.4)를 사용한다. 즉, 식 (3.7.1)과 (3.7.4)에 "$\mu_1-\mu_2=0$"를 대입하면 귀무가설 하에서 각각 $N(0,1^2)$과 자유도가 n_1+n_2-2인 t분포를 따른다. 그리고, 대립가설이 "$\mu_1<\mu_2$"이면 UTT, "$\mu_1>\mu_2$"이면 LTT, "$\mu_1\neq\mu_2$"이면 TTT가 된다. 물론, 대립가설이 "$\mu_1<\mu_2$" 또는 "$\mu_1>\mu_2$"일 때에는 귀무가설이 "$\mu_1\geq\mu_2$" 또는 "$\mu_1\leq\mu_2$"이더라도 각각 UTT와 LTT가 된다. 또한, 기각역을 구하는 방법은 앞에서 등장한 Z-test와 T-test의 경우와 동일하다.

<비고 4.4.7> T-test이고 TTT인 "H_0: $\mu_1=\mu_2$, H_a: $\mu_1\neq\mu_2$" 경우는 §5.2에서 상세히 다룬다.

사실, 가설이 다르더라도 검정통계량이 귀무가설 하에서 $N(0,1^2)$분포를 따르기만 하면 기각역은 동일하다. 예를 들어, 모든 Z-test에서 $\alpha=0.05$이고 UTT이면 기각역은 $z\geq 1.645$이다. 이는 또한 T-test, 카이제곱 검정, F-test에서도 성립한다. 예를 들어, 모든 T-test에서 자유도가 9, $\alpha=0.05$, 그리고 UTT이면 기각역은 $t\geq 1.833$이다.

§4.5 중심극한정리와 검정

중심극한정리(central limit theorem)를 CLT라 부르자. CLT는 §2.5에서 정규분포를 소개할 때 처음 등장했는데, 이를 협의의 CLT라 하자. 즉, $Y_1, \cdots, Y_n$이 *iid* 확률변수이면 $\overline{Y} = \sum_{i=1}^{n} Y_i / n$ 은 점근적으로 정규분포를 따른다는 것이 협의의 CLT인데(§2.15.2 참조), 이는 물론 Y의 분포와 무관하게 (그리고, Y의 분포를 모를 때에도) 성립한다.

반면에, 광의의 CLT는 다음과 같다. 첫째로, §3.5.2에서 모든 최우추정량은 점근적으로 정규분포를 따른다고 했다. 둘째도 이와 같은 맥락인데, LRT에서 LR를 λ라 하면 (<비고 4.4.2> 참조)

$$-2\ln\lambda \overset{A}{\sim} \chi^2(d) \tag{4.5.1}$$

이라고 알려져 있다. 즉, "$-2\ln\lambda$"는 점근적으로 카이제곱분포를 따른다는 것인데, 이때 유의할 점은 검정의 대상이 무엇이든지 불문하고 성립한다는 점과 또한 모집단의 분포와 무관하다는 점이다. 그리고, 자유도 d 는 귀무가설에 포함된 제약식의 개수인데, 예를들면 하나의 모집단의 경우에 "$\mu = \mu_0$", "$\sigma^2 = \sigma_0^2$" 그리고 두 개의 모집단의 경우에 "$\mu_1 = \mu_2$", "$\sigma_1^2 = \sigma_2^2$ " 등이 제약식이다. 그런데, 식 (4.4.9)에서 $0 \le \lambda \le 1$이고 기각역은 $\lambda \le k$이므로, 식 (4.5.1)에서는 $0 \le -2\ln\lambda < \infty$ 이고 기각역은 $-2\ln\lambda \ge k' (= -2\ln k)$가 된다. 즉, 광의의 CLT에 의한 카이제곱 검정은 항상 UTT이다 (<비고 4.4.5> 참조). 예를 들어 자유도가 1이고 유의수준이 5%이면 기각역은 $-2\ln\lambda > 3.84146$ 이다.

협의의 CLT의 대표적인 활용사례는 모비율에 대한 추정 및 검정이다. §3.1에서는 모비율에 대한 점 추정량 및 신뢰구간을 구했고, §4.2에서는 모비율에 대한 검정을 예로 들어서 LRT를 설명했다.

광의의 CLT의 대표적인 활용사례는 §4.6에 등장하는데, 이 역시 모비율에 대한 검정이다. 다만, 차이점은 이항분포 대신에 다항분포가 등장하고 또한 Z-test 대신에 카이제곱 검정이 되는 것이다.

<비고 4.5.1> 이항분포는 다항분포의 특수한 경우이고, (Z를 제곱하면 자유도가 1인 카이제곱 확률변수가 되므로 : §2.6.1 참조) Z-test는 카이제곱 검정의 특수한 경우이다(§4.6.2 참조).

§4.6 분할표 분석

4.6.1 일차원 분할표

주사위 하나를 600번 굴려서 아래의 결과를 얻었다고 하자. (비고: 이 예제는 문헌[9]의 연습문제 14-2임.)

i	1	2	3	4	5	6	합계
n_i	89	113	98	104	117	79	600(= n)

이와 같은 표를 일차원 분할표 (contingency table)라 한다. (비고:이차원 분할표는 §4.6.4에 등장함.) 즉, 전체 600을 주사위 눈의 수에 따라 여섯 부분으로 분할해 놓은 표이다.

이 예제에서 관심사는 주사위가 과연 대칭인가 (또는, balanced 인가) 하는 것이다. 이를 가설검정의 틀로 표현하면

$$\begin{aligned} &H_0 : p_i = \frac{1}{6} \ , \ i=1,\cdots,6 \\ &H_a : Not\ H_0 \end{aligned} \tag{4.6.1}$$

인데, p_i는 주사위를 한번 굴릴 때 i개의 눈이 나올 확률이다.

n_i에 대응하는 확률변수를 N_i라 하자. 즉, 600번 굴릴 때 i개의 눈이 나오는 횟수가 N_i인데, 귀무가설 하에서 $\{N_1, \cdots, N_6\}$는 다항분포를 따른다(§3.1.2 참조). 아래의 표는 위의 표에 $E(N_i)$와 관련된 몇가지를 추가한 것이다. (비고: $E(N_i)$ = np_i)

i	1	2	3	4	5	6	합계
n_i	89	113	98	104	117	79	600
$E(N_i)$	100	100	100	100	100	100	600
$n_i - E(N_i)$	-11	13	-2	4	17	-21	0
$\{n_i - E(N_i)\}^2$	121	169	4	16	289	441	
$\frac{\{n_i - E(N_i)\}^2}{E(N_i)}$	1.21	1.69	0.04	0.16	2.89	4.41	10.40

□contingency table□을 □분할표□라 부르기는 하지만, □contingency□는 □우발성□이라 부르는데 이는 위의 표에서 $\{n_i - E(N_i)\}$를 일컫는 표현이다. 귀무가설이 참이

라고 해서 반드시 $n_i = 100(=E(N_i))$이 되는 것은 아니다. 확률적으로 또는 □우발적□으로 n_i는 100보다 크기도 하고 작기도 한 것이다.

결론부터 언급하면, 이는 카이제곱 검정인데 UTT이고(<비고 4.4.5> 참조) 자유도는 5이다. 따라서, α =0.05에 의한 기각역은 11.0705 이상이다. 그런데, 위의 표에서

$$\sum_{i=1}^{6} \frac{\{n_i - E(N_i)\}^2}{E(N_i)} = 10.40 < 11.0705 \qquad (4.6.2)$$

이므로 귀무가설을 채택(accept)한다. 즉, 주어진 표본정보는 유의수준 5%에서 귀무가설을 기각할 만한 충분한 근거가 되지 못한다.

이 검정 역시 LRT에 의한 것인데, 2-단계 근사라고 할 수 있다(<비고 2.15.3>참조). 즉, 1차적으로는 광의의 CLT인 식 (4.5.1)에 의한 근사이고, 2차적으로는 $-2\ln\lambda$를 멱급수(power series)로 전개하여 식(4.6.2)만 남기고 나머지는 무시함에 따른 근사이다. (비고 : □$-2\ln(1-x) = 2\sum_{k=1}^{\infty} x^k/k$□에서 □$2x + x^2$□만 남긴 것임.) 참고로, LF는 식(3.1.1) 형태인 $K\prod_{i=1}^{6} p_i^{n_i}$인데, p_i에 대한 최우추정치는 귀무가설 제약 하에서는 1/6 이고(<비고 4.4.1> 참고) 제약없이는 n_i/n이다. 따라서, $\lambda = \prod_{i=1}^{6}\{(1/6)/(n_i/n)\}^{n_i}$이다. (이후 과정은 생략함.)

이제, 자유도가 5인 이유를 설명한다. 식(4.6.1)의 귀무가설에 포함된 제약식의 개수는 6개이지만, 이중에서 5개가 주어지면 남은 하나는 자동적으로 결정되므로(비고 : $\sum_{i=1}^{6} p_i = 1$) 사실상 5개인 셈이다. 이는 또한 정보의 개수라는 관점으로 해석할 수도 있다(§2.15.4 참조). 분할표에 있는 정보는 $\{n_1 = 89, \cdots, n_6 = 79\}$라고 할 수 있는데, 정보의 개수가 5인 이유는 $\sum_{i=1}^{6} n_i = 600$이기 때문에 $n_1, \cdots, n_6$중에서 5개만 주어지면 남은 하나는 자동적으로 결정되기 때문이다.

일차원 분할표 분석에 대한 별칭은 □chi-square test of GOF(goodness of fit)□이다. GOF란 주어진 표본의 분포가(<비고 1.5.1>참조) 귀무가설 하의 분포와 얼마나 잘 맞는가를 가늠하는 척도인데, □$H_0 : p_i = 1/6$□은 이산 uniform 분포를 의미한다(§2.1.5 참조).

<비고 4.6.1> 귀무가설 하의 분포에 미지의 모수가 포함되어 있으면 이를 MLE로 대체한다. 아울러, MLE로 대체된 모수의 개수만큼 자유도가 감소하는데, 이는 식(2.15.10)

의 μ를 MLE인 $\overline{Y}$로 대체함에 따라 식 (2.15.11)에서는 자유도가 하나 감소한 이유와 같다.

GOF 검정에 대해서 마지막으로 짚고 넘어갈 것은 유의수준이다. §4.1에서 검정의 목적은 귀무가설을 기각하기 위한 것이라고 했다. 그리고, 귀무가설이 기각되었을 때에는 α 값이 작을수록 설득력이 강하다. 즉, α 값을 작게 책정함으로써 어지간하면 귀무가설이 채택되도록 했음에도 불구하고 귀무가설이 기각되었다는 것은 표본정보가 귀무가설에 결정적으로 불리했다는 것을 의미한다. 그러나, GOF 검정은 예외이다. 그 이유는 GOF 검정이 귀무가설을 (기각이 아니라) 채택하기 위해서 사용되는 경우가 많기 때문이다. 이는 주어진 표본의 분포가 귀무가설에서 주장하는 분포와 잘 맞는다는 것을 보이기 위한 검정을 의미하는데, 이 경우 귀무가설이 채택되었을 때에는 α 값이 클수록 오히려 설득력이 강해진다. 위의 예에서 $\alpha=0.05$로 귀무가설을 채택했는데, 만약 $\alpha=0.10$이었다면 기각역은 9.23635 이상이 되어 귀무가설을 (채택하지 못하고) 기각하게 된다.

4.6.2 Z-Test 와의 관계

§4.6.1의 예제에서 p_1 하나만 관심의 대상이라고 하자. 이에 따라, 식 (4.6.1)을

$$H_o: p_1 = 1/6,\quad H_a: p_1 \neq 1/6 \tag{4.6.3}$$

로 고치면, 식(4.6.2) 대신

$$\frac{(89-100)^2}{100} + \frac{(511-500)^2}{500} = 1.452 \tag{4.6.4}$$

를 얻는다. 이 역시 카이제곱 검정이고 UTT지만 자유도는 1이 된다. 따라서, $\alpha=0.10$에 의한 기각역은 2.70554 이상이므로 ($\alpha=0.05$에 의한 기각역은 3.84146 이상), 유의수준 10%에서 조차 귀무가설을 채택하게 된다.

식 (4.6.3)은 협의의 CLT에 의한 Z-Test로 처리할 수도 있다(<비고 4.5.1> 참조). 귀무가설하에서 검정통계량인 식 (3.1.7)은 점근적으로 $N(0,1^2)$분포를 따르는데, 이에 표본관찰치를 대입하면

$$z = \frac{\widehat{p_1} - p_1}{\sqrt{p_1(1-p_1)/n}} = \frac{\frac{89}{600} - \frac{1}{6}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}/600}} = -1.205 \tag{4.6.5}$$

를 얻는다. 그런데, 이는 TTT 이므로 $\alpha = 0.10$에 의한 기각역은 $z \ge 1.645$ 또는 $z \le -1.645$ 이다. 따라서, 유의수준 10%에서 귀무가설을 채택한다.

위의 두가지 검정이 동일함을 수치적으로 보인다. 첫째, 식 (4.6.5)을 제곱하면 식 (4.6.4)가 된다. 둘째로, $\alpha = 0.10$에 의한 기각역 “$z \ge 1.645$ 또는 $z \le -1.645$”는 “$z^2 \ge 2.70554$”와 동일하다(<비고 4.5.1> 참조>). 따라서, §4.6.1의 GOF 검정은 모비율에 대한 Z-Test를 확장한 것이라 할 수 있다.

4.6.3 p-value

§4.6.2의 예제에서 $\alpha = 0.10$으로 H_0를 채택했다. 만약 $\alpha < 0.10$ 이었더라도 (기각역은 오히려 더 좁아지므로) 여전히 H_0를 채택했을 것이다. 그러나, α값을 점점 증가시키면 (기각역이 점점 넓어져서) 언젠가는 H_0를 기각하게 되는데, 이 때 경계치를 p-value 라 부른다. 구체적으로, TTT의 기각역을 “$z \ge 1.205$ 또는 $z \le -1.205$”가 되게하는 α값이 바로 p-value 인데, 이는 $P(Z \ge 1.205$ 또는 $Z \le 1.205) = 2(0.1141) = 0.2282$이다. 즉, $\alpha < 0.2282$였다면 H_0를 채택하고 $\alpha \ge 0.2282$였다면 H_0를 기각한다.

일반적으로, 모든 검정에서 $\alpha <$ p-value 이면 H_0를 채택하고 $\alpha \ge$ p-value 이면 H_0를 기각한다. 통계 관련 패키지 (package 또는 s/w)는 대부분 p-value를 제공하는데, 이는 사용자들이 각자가 선호하는 α 값과 비교해서 H_0를 채택 또는 기각할 수 있게하기 위함이다.

4.6.4 이차원 분할표

§4.6.1에 등장한 일차원 분할표는 전체 n을 한가지 요인(factor)에 따라 $n_1, n_2, \cdots$로 분할한 것이다 (단, $n = \sum_i n_i$). 반면에, 이차원 분할표는 전체 n을 두가지 요인에 따라 아래의 형태로 분할한 것이다. (비고: $r_i = \sum_j n_{ij}$, $c_j = \sum_i n_{ij}$, $n = \sum_i r_i = \sum_j c_j$)

$i \backslash j$	1	2	$\cdots$	J	row sum
1	n_{11}	n_{12}	$\cdots$	n_{1J}	r_1
2	n_{21}	n_{22}	$\cdots$	n_{2J}	r_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	r_I
column sum	c_1	c_2	$\cdots$	c_J	n

일차원 분할표의 분석을 “chi-square test of GOF”라고 부르듯이, 이차원 분할표의 분석은 “chi-square test of independence (또는, homogeneity)”라고 부른다. 여기에서, “independence”란 두가지 요인이 서로 영향을 미치지 않는다는 의미이다. 예를 들어, 담배를 피우는가($i=1$) 안 피우는가($i=2$)하는 것이 교회에 다니는가($j=1$) 안 다니는가($j=2$)하는 것과 상관이 있는지를 알아보는 것이다. 또한, “homogeneity”라는 표현은 다음과 같은 형태의 “동질성”을 의미한다.

$$\frac{n_{1j}}{r_1} \approx \frac{n_{2j}}{r_2} \approx \cdots \approx \frac{n_{Ij}}{r_I}\left(\approx \frac{c_j}{n}\right),\ j=1,\cdots,J \tag{4.6.6}$$

<비고 4.6.2> 요인 i와 요인 j는 교환가능(interchangeable)하다. 따라서, 식(4.6.6)은 “ $(n_{ij}/c_j) \approx (r_i/n)$, $i=1,\cdots,I$, $j=1,\cdots,J$ ”와 같다.

이차원 분할표의 분석을 간단히 “독립성 검정”이라 부르자. 독립성 검정은 GOF 검정을 확장시킨 것이다. 구체적으로, GOF 검정에는 다항분포 하나가 등장하지만 독립성 검정에는 I개의 (또는, <비고 4.6.2>에 의해서 J개의) 다항분포가 등장한다. 예를 들어, 지난 총선 때 여론조사가 I번 시행되었는데, 그 중 i번째에서 j번째 정당을 지지한 유권자의 수를 n_{ij}라 하자. 이 경우, 귀무가설은 “시간이 흘러도 정당별 지지율은 변하지 않음”이다. 가설을 수학적으로 표현하기 위해서, n_{ij}에 대응하는 확률변수를 N_{ij}라 하면 $\{N_{i1},\cdots,N_{ij}\}$는 다항분포를 따르는데, 이의 모수를 $\{p_{i1},\cdots,p_{ij}\}$라 하자. 그러면 H_0와 H_a는 다음과 같다.

$$H_0 : \ p_{11} = p_{21} = \cdots = p_{I1} \qquad\qquad H_a : \ Not\ H_0$$

$$
\begin{aligned}
&p_{12} = p_{22} = \cdots = p_{I2} \\
&\qquad \vdots \\
&p_{1J} = p_{2J} = \cdots = p_{IJ}
\end{aligned}
\tag{4.6.7}
$$

독립성 검정 역시 광의의 CLT에 의한 카이제곱 검정이고 UTT인데, 자유도는 $(I-1)(J-1)$이다. 즉, H_0에 포함된 제약식 중에서 자동적으로 결정되는 것을 제외하면 모두 $(I-1)(J-1)$개다. 구체적으로, “$p_{1j} = p_{2j} = \cdots = p_{Ij}$”는 “$p_{1j} = p_{2j}$, $p_{2j} = p_{3j}, \cdots$, $p_{I-1,j} = p_{Ij}$”와 같으므로 $(I-1)$개의 제약식에 해당된다. (비고: “$p_{Ij} = p_{1j}$”는 자동적으로 결정됨.) 그리고, 이같이 $(I-1)$개의 제약식에 해당되는 것들은 모두 J개가 있으나 이중에서 $(J-1)$개가 주어지면 남은 하나는 자동적으로 결정된다 (이유 ?). 이는 자유도가 “사용된 정보의 개수”라는 관점으로도 설명이 가능하다. 모두 IJ개의 n_{ij}값들이 있으나 일차적으로 $r_1, \cdots, r_I$ 에 의해서 I개는 자동적으로 결정된다. (비고: r_i는 여론조사에서 표본의 크기에 해당됨.) 남은 $I(J-1)$개에서 이차적으로 $(J-1)$개를 빼는데 이는 귀무가설 하에서 필요한 MLE의 개수이다 (<비고 4.6.1> 참조). 즉, $I(J-1)$개의 가용 정보 중에서 $(J-1)$개는 미지의 모수를 추정하는데 쓰이고 나머지 $(I-1)(J-1)$개 만이 검정에 쓰이는 셈이다.

참고로, LF는 식 (3.1.1)를 확장한 형태인 $\prod_{i=1}^{I} K_i \prod_{j=1}^{J} p_{ij}^{n_{ij}}$인데, p_{ij}에 대한 최우추정치는 귀무가설 제약 하에서는 $\widehat{p}_j = c_j/n$ 이고 (식 (4.6.6) 참조), 제약없이는 $\widehat{p}_{ij} = n_{ij}/r_i$ 이다. 따라서,

$$
\lambda = \frac{\prod_{i=1}^{I}\prod_{j=1}^{J}(c_j/n)^{n_{ij}}}{\prod_{i=1}^{I}\prod_{j=1}^{J}(n_{ij}/r_i)}
\tag{4.6.8}
$$

이다. 그리고, GOF 검정에서와 같이, $-2\ln\lambda$를 멱급수로 전개하여 3차항 이상을 무시하면 (비고: 1차항은 0이 되므로 사실상 2차항만 남기는 것임), 기각역으로

$$
Reject\ H_0 \text{ if } \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\{n_{ij}-\widehat{E}(N_{ij})\}^2}{\widehat{E}(N_{ij})} \geq k
\tag{4.6.9}
$$

를 얻는데, 이는 식 (4.6.2)을 확장한 형태이다. 단, $\widehat{E}(N_{ij})$는 귀무가설 하에서의 “$E(N_{ij}) \equiv r_i p_{ij}$”에 대한 MLE 이다. 즉,

$$\widehat{E}(N_{ij}) = r_i \widehat{p}_j = \frac{r_i c_j}{n} \tag{4.6.10}$$

이다(§3.5.1에 등장한 MLE의 invariance 속성 참조).

<비고 4.6.3> 식 (4.6.9)를 3-차원으로 확장하면 (카이제곱 검정의) 기각역은

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\{n_{ijk} - \widehat{E}(N_{ijk})\}^2}{\widehat{E}(N_{ijk})} \ge k$$

가 되고, 자유도는 $(I-1)(J-1)(K-1)$이다.

4.6.5 독립성 검정 예제

100명에게 설문한 결과 담배를 피우는 사람은 40명이고 안 피우는 사람은 60명인데, 그 중에서 교회에 다니는 사람은 각각 8명과 22명이라 하자. 이를 분할표로 표현하면 다음과 같다.

$i \backslash j$	1	2	
1	$n_{11}=8$	$n_{12}=32$	$n_1 \equiv r_1 = 40$
2	$n_{21}=22$	$n_{22}=38$	$n_2 \equiv r_2 = 60$
	$c_1=30$	$c_2=70$	$n=100$

편의상, $p_i \equiv p_{i1}$, $q_i \equiv p_{i2}$ 라 하면 H_0와 H_a는

$$\begin{aligned} &H_0 : p_1 = p_2 \ \text{(비고: “}p_1 = p_2\text{”이면 자동적으로 “}q_1 = q_2\text{”)} \\ &H_a : p_1 \neq p_2 \ \text{(비고: “}p_1 \neq p_2\text{”는 “}Not\ H_0\text{”와 같음)} \end{aligned} \tag{4.6.11}$$

인데, 이는 식 (4.6.7)의 특수한 경우 $(I=J=2)$일 뿐만 아니라 식 (3.7.5)에 대응되는 검정이기도 하다. 돌이켜보면, §3.7에서 신뢰구간용 PQ로 사용되었던 통계량들이 4장에서는 (LRT에 의한) 검정통계량으로 재등장하고 있는데, 아직 하나 남은 것이 식 (3.7.5)에 대한 PQ인

$$Z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \tag{4.6.12}$$

인 것이다.

식 (4.6.12)의 배경에 깔린 가정은 다음과 같다. p_1은 담배를 피우는 사람들의 모집단에서 교회에 다니는 사람의 모비율이고, p_2는 담배를 안 피우는 사람들의 모집단에서 교회에 다니는 사람의 모비율이다. 표본의 크기는 각각 $n_1 = 40$과 $n_2 = 60$인데, n_{i1}에 대응하는 확률변수를 N_i라 하면 N_1과 N_2는 서로 독립이고 (이유: 서로 다른 모집단에서 추출), 각각 이항분포를 따른다 (이유: 복원추출로 가정). 이에, 협의의 CLT를 적용하면 $N_i \overset{A}{\sim} N(n_i p_i, n_i p_i q_i)$이고 또한 $\widehat{p_i} \equiv (N_i / n_i) \overset{A}{\sim} N(p_i, p_i q_i / n_i)$이다. 그리고, N_1과 N_2가 독립이므로 $\widehat{p_1}$과 $\widehat{p_2}$도 독립이다 (<비고 2.11.2> 참조). 따라서, 식 (4.6.12)는 점근적으로 $N(0, 1^2)$분포를 따른다.

귀무가설 하에서 식 (4.6.12)에 "$p_1 = p_2$"를 대입하면, 분자의 "$p_1 - p_2$"는 "0"이 되지만 분모에는 여전히 미지의 모수(인 p_1과 p_2)가 남는다. (비고: 식 (4.6.5)에서는 이러한 문제가 발생하지 않았는데, 이는 식 (4.6.3)의 귀무가설에서 구체적인 수치 ($p_1 = 1/6$)가 주어졌기 때문임.) 따라서, 2차적인 근사로 (1차는 CLT에 의한 근사), 미지의 모수를 최우추정치로 대체한다. 그런데, 유의할 점은 귀무가설에서 구체적인 수치가 주어지지는 않았으나, "$p_1 = p_2$"라고 했으므로 p_1과 p_2를 따로따로 추정하지 않고 한꺼번에 묶어서

$$\widehat{p} = \frac{8 + 22}{40 + 60} = \frac{30}{100} \left(= \frac{c_1}{n} \right) \tag{4.6.13}$$

로 추정한다는 점이다. 즉, $\widehat{p}$은 소위 pooled 추정치인데 (식 (3.7.3) 참조), 이는 물론 식 (4.6.8)의 (c_j / n)에 해당된다.

Z-test의 판정결과를 얻기 위해서 식 (4.6.12)의 "$\widehat{p}_i = N_i / n_i$"를 이에 대응하는 관찰치인 "$n_{ij} / n_i$"로 대체하고, 또한 p_1과 p_2를 식 (4.6.13)으로 대체하면

$$z = \frac{\left(\frac{8}{40} - \frac{22}{60} \right) - 0}{\sqrt{\left(\frac{30}{100} \right) \left(\frac{70}{100} \right) \left(\frac{1}{40} + \frac{1}{60} \right)}} = -1.782 \tag{4.6.14}$$

를 얻는다. 그런데, 이는 TTT이므로 $\alpha=0.05$에 의한 기각역은 $z \geq 1.96$ 또는 $z \leq -1.96$이다. 따라서, 유의수준 5%에서 귀무가설은 채택된다. 반면에, $\alpha=0.10$에 의한 기각역은 $z \geq 1.645$ 또는 $z \leq -1.645$이므로, 유의수준 10%에서는 귀무가설을 기각할 수 있다. 참고로, p-value는 $P(Z > 1.782$ 또는 $Z \leq -1.782) \approx 0.075$이다(§4.6.3 참조).

이제 식 (4.6.11)의 가설을 식 (4.6.9)로 검정한다. 카이제곱 검정에서 자유도는 $(2-1)(2-1)=1$이고 UTT이므로, 기각역은 (§4.6.2에서와 동일하게) $\alpha=0.05$ 일 때는 3.84146이상이고 $\alpha=0.10$일 때는 2.70554이상이다. 그런데, 식 (4.6.9)에 n_{ij}값과 식 (4.6.10)을 대입하면

$$\frac{(8-12)^2}{12}+\frac{(32-28)^2}{28}+\frac{(22-18)^2}{30}+\frac{(38-42)^2}{42}=3.1746 \qquad (4.6.15)$$

을 얻는데, 이는 $\alpha=0.05$일 때는 기각역에 속하지 않고 $\alpha=0.10$일 때는 기각역에 속한다. 또한, §4.6.2에서와 같이, 위의 두가지 검정은 동일하다. (비고: 식 (4.6.14)를 제곱하면 식 (4.6.15)를 얻음.) 따라서, 카이제곱 검정의 p-value 역시 (Z-test와 같이) 약 0.075 (인데, 이는 카이제곱 확률표에는 잘 등장하지 않는 값)이다.