6장. 가설 검증론

과학은 가설(hypothesis)의 끊임없는 개선으로 발전합니다. 특히, 실증과학에 있어서는 경험적 증거를 확률적 관점에서 평가하는 일이 중요합니다. 이에 대한 철학 및 방법으로서 피셔(R.A. Fisher, 1890-1962)의 유의성 검증론을 소개하고 네이만(J. Neyman, 1894-1981)과 피어슨(E.S. Pearson, 1895-1980)이 기여한 수리적 최적검증론을 설명하기로 하겠습니다. 즉 가능도(likelihood)의 비(比)가 두 가설을 비교・평가하기 위한 지표로서 왜 유용한가를 다룬 네이만-피어슨 보조정리를 다루고 대역적으로 가장 좋은 검증법인 균일최강력검증(uniformly most powerful test)을 만들어보기로 하겠습니다. 그 밖에 양측검증을 위한 비편향성(unbiasedness) 개념과 대역적으로 최적인 검증법이 존재하지 않는 경우에 국소적으로 최적인 검증법을 개발해 보도록 하겠습니다.

파라미터가 2개 이상인 경우의 검증 방법으로서 가능도 비 검증(likelihood ratio test)을 일반화한 형태를 고려해보는 것은 당연합니다. 이에 대한 대표본 성질을 탐구하고 또한 동등한 다른 두 방법 -왈드 검증(Wald test)과 점수 검증(score test)-을 유도하기로 하겠습니다. 이 중 한 방법은 다항분포에 적용되는 경우 그 유명한 피어슨 카이제곱 검증(Pearson's chi-square test)이 된다는 것을 보이도록 하겠습니다. 그리고 영가설이 복합적인 경우, 즉 파라미터가 완전히 명시되지 않는 경우로 일반화 가능도 비 검증 등을 확장하겠습니다.

차례 :

6.1 피셔의 유의성 검증(Fisher's Significance Testing)

통계학의 아버지라고 할 수 있는 피셔(R.A. Fisher, 1890-1965)가 젊었을 때 그가 일하던 로담스테드(Rothamsted Experimental Station; 영국 런던 교외의 농업연구소)에서의 일입니다. 티 타임에 연구소에 새로 온 젊은 여성 과학자에게 피셔가 그녀를 위해 티를 만들어 권하였습니다. 그러나 그녀는 그 티를 사양하였습니다. 왜냐하면 그녀는, 매우 까다롭게도, 우유를 찻물보다 먼저 넣은 티만 마신다는 겁니다. "우유를 찻물보다 먼저 넣었는가? 아니면 찻물을 우유보다 먼저 넣었는가?"를 그녀가 구분할 수 있는가 테스트하자고 피셔가 제안하였고 그녀가 그것을 수락하여 실험을 하게 되었습니다. 실험 과정과 결과는 이러했습니다.

1) 피셔가 여덟 잔의 티를 준비하였습니다. 그 중 네 잔에는 찻물을 우유보다 먼저 넣었고 나머지 네 잔에는 우유를 찻물보다 먼저 넣었습니다. 그러나 우리는 각 잔에서 찻물과 우유의 순서가 임의로 결정되었다고 가정하겠습니다.
2) 그녀는 여덟 잔 각각을 마셔보고 각 잔에서 찻물이 먼저인가, 아니면 우유가 먼저인가를 답하였습니다.
3) 결과적으로 그녀는 여섯 잔을 맞추었고 두 잔에서 틀렸습니다.

이와 같은 통계적 실험에서는 완벽한 추리가 가능하지 않습니다. 순전히 우연만으로도 여섯 번뿐 아니라 모두 맞출 가능성이 있고, 그 반대로 실력이 상당히 좋은데도 불구하고 두 번이나 실수할 가능성이 있기 때문입니다. 그렇다면 어떻게 생각해야 할까요?

피셔의 기본적인 생각은 반증주의(反證主義, falsificationism) 철학이었습니다. 그녀가 자신의 능력을 인정받기 위해서는 능력이 없을 것이라는 세인(世人)들의 의심을 뒤엎을만한 증거를 스스로 보여야 한다는 것입니다. 그래서 피셔의 유의성 검증(有意性 檢證, significance testing)은 영가설과 대안가설을 세우는 데에서 시작합니다.

<u>영가설</u>(零假說, null hypothesis) H_0 : 정의
효과가 없다, 차이가 없다, 다르지 않다는 등 세인들의 의심이 그 내용입니다.
<u>대안가설</u>(代案假設, alternative hypothesis) H_1 : 정의
효과가 있다, 차이가 있다, 다르다는 등 연구자의 주장을 내용으로 합니다.

따라서 피셔의 티 실험에서 영가설은 그녀가 찻물-우유 순서를 구분할 아무 능력이 없다는 것이고 대안가설은 그녀가 찻물-우유 순서를 구분할 능력이 어느 정도 있

다는 것입니다. 즉 개별시행에서 그녀의 맞출 확률을 θ라고 하면 영가설과 대안가설이 각각

$$\mathrm{H}_0 : \theta = 1/2 , \qquad \mathrm{H}_1 : \theta > 1/2$$

입니다. (이 여성 연구자의 주장이 “나는 능력이 있다”는 것이므로 대안가설을 양측 방향을 모두 포함하는 $\mathrm{H}_1 : \theta \neq 1/2$ 로 설정하는 것은 논센스입니다.)

만약 영가설에서 연역되는 관측사건에 대한 포괄적인 개연성이 극히 미미하다면 영가설을 부인해야 할 입장에 서게될 것이라는 것이 반증주의적 논리입니다. 그렇지 않다면 영가설을 부인해야 할 실제적 증거가 충분하다고 볼 수 없을 것입니다. 그러므로 피셔의 검증론은 다음과 같이 정의되는 p 값에 결정적으로 의존합니다.

<u>p 값</u>(p-value) : 정의

실제 관측된 사건, 또는 이에 못지 않게 영가설에 반대되는 증거로 볼 수 있는 사건들이 영가설 하에서 파생될 확률.

앞의 티 실험에서는 맞춘 회수 X가 실제 6번이었는데 7번이나 8번이었다면 더욱 영가설에 반대되는 증거가 될 수 있었을 터이므로

$$\text{p-값} = P\{X \geq 6; \mathrm{H}_0\}$$

가 됩니다. 이 사례에서는 X가 이항분포 $\mathrm{B}(n, \theta)$를 따른다고 볼 수 있으므로 (n은 시행 수 8), 구체적으로

$$\text{p-값} = P\{X \geq 6; \mathrm{H}_0\} = \sum_{x=6}^{8} \binom{n}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} = 0.145$$

입니다.

실제 유의성 검증에서 p-값이 어느 정도 작아야 영가설을 기각하고 대신 대안가설에 심증을 두는 것이 합리적일까요? 이것은 근본적으로 영가설이 잠정적인 기준값인가, 아니면 얼마나 단단한가에 달려 있습니다. 영가설 H_0가 수세기에 걸쳐 모든 과학자들이 아무 의심 없이 믿어왔던 것이라면 p-값이 0.0001보다 작더라도 “확률이 무척 작은 사건이 관측되었구나” 또는 “어디엔가 관측에 오류가 있었구나”라고 생각하면서 영가설 H_0를 기각하지 않게 될 것입니다. 그러나 그것이 잠정적인 기준값에 불과하다면, p-값이 일정 값 예컨대 0.05 보다 작은 경우에 한하여 대안가설 H_1에 대한 통계적 유의성(statistical significance)을 인정하는 것이 관례입니다. 이렇듯 p-값의 통계적 유의성을 가름하는 임계값을 유의수준(significance level) α라고 합니다.

티 실험 사례에서는 p-값이 0.145이므로 썩 느슨한 유의수준인 $\alpha = 0.10$에서도 통계적 유의성을 인정할 수 없습니다 (즉, 영가설 H_0를 기각할 수 없습니다). 즉 이

여성 연구자의 주장에 대한 통계적 증거가 불충분하다는 것이 결론입니다. 그렇지만 만약 그녀가 7번을 맞추었다면

$$\text{p-값} = P\{X \geq 7; \mathrm{H}_0\} = \sum_{x=7}^{8} \binom{n}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} = 0.035$$

이므로 $\alpha = 0.05$에서 영가설 H_0를 기각할 수 있습니다. 따라서 이 여성 연구자의 능력에 심증을 둘 수밖에 없을 것입니다.

일반적으로 유의성 검증 절차는 다음과 같습니다.

1) 영가설 H_0와 대안가설 H_1을 설정합니다.
2) 검증통계량 T를 정합니다. 대강의 요령은 T가 큰 값을 취할수록 영가설 H_0로부터 멀리 떨어진 증거가 되도록 하는 것입니다.
3) 표본자료로부터 T가 취하는 값 t를 산출합니다 (즉, $T(x_1, \cdots, x_n) = t$).
4) t에 대한 p-값은 $P\{\mathrm{T} \geq \mathrm{t}; \mathrm{H}_0\}$ 입니다. 이것과 유의수준 α와 비교하여 검증 결과를 냅니다.

예를 들기로 하겠습니다. $X_1, \cdots, X_n$을 오염정규분포

$$0.9\,\mathrm{N}(\theta, \sigma^2) + 0.1\,\mathrm{N}(\theta, k^2\sigma^2), \quad k \geq 1$$

로부터의 임의표본이라고 합시다 (5.2절 참조). 여기서 σ^2과 k는 알려진 상수라고 가정하겠습니다 ($\sigma^2 = 1$, $k = 3$). 표본크기는 $n = 10$이라고 합시다. 이 때, 분포의 중심인 θ에 대한 유의성 검증을 하기로 합니다.

1) 연구자가 $\theta > 10$인 것을 보이고 싶다고 합시다. 그러면 영가설과 대안가설이 각각 다음과 같습니다.

$$\mathrm{H}_0 : \theta = 10 \quad \text{vs.} \quad \mathrm{H}_1 : \theta > 10.$$

2) 표본중위수 $\tilde{X}$를 검증통계량 T로 하겠습니다. 즉

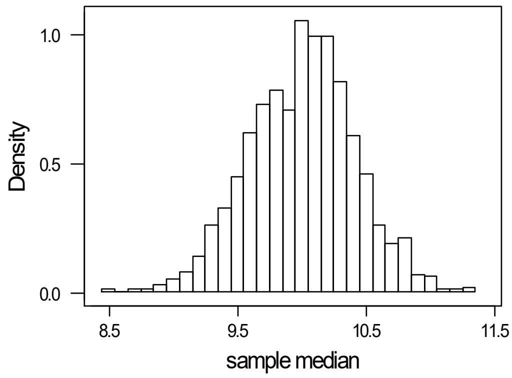
$$T = \mathrm{median}(X_1, \cdots, X_n).$$

T의 값 t가 클수록 θ가 크다는 증거이므로 영가설 $\mathrm{H}_0 : \theta = 10$으로부터 멀어지게 됩니다.

3) 크기 $n = 10$인 임의표본이 다음과 같이 관측되었다고 합시다 (순서정렬).

8.95, 9.49, 9.65, 10.28, 10.42, 10.58, 10.63, 10.65, 11.10, 12.60.

따라서 표본에서의 중위수는 $t = 10.50$입니다.



<그림 1> 영가설 $H_0 : \theta = 10$ 하에서 표본 중위수의 표집분포

4) 이제 p-값 = $P\{\, T \geq 10.50\,;\theta = 10\,\}$을 계산해야 합니다. T의 분포를 정확히 계산하는 것이 어려워 보이므로 몬테칼로 계산을 해보기로 합시다 (연습문제 6.2). <그림 1>은 영가설 $H_0 : \theta = 10$ 하에서 표본 중위수 T의 표집분포입니다 (반복수 $N = 1{,}000$). 한 결과로서 p-값 = 0.102가 나옵니다. 따라서 $\alpha = 0.05$에서 영가설 H_0를 기각할 수 없습니다.

유의성 검증의 창안자라고 할 수 있는 피셔의 다음 글을 음미함으로써 이 절을 마치겠습니다. p-값이란 실제의 확률이 아니라 *가설적인* 확률이라고 강조한 부분을 주목하십시오.

In general, tests of significance are based on *hypothetical* probabilities calculated from the null hypothesis. They do not generally lead to any probability statements about the real world, but to a rational and well-defined measure of reluctance to the acceptance of the hypothesis they test. (Fisher, 1956, Chapter III.1; Statistical Methods and Scientific Inference).

6.2 네이만-피어슨 이론(Neyman-Pearson Theory)

피셔의 유의성 검증에서 미진한 부분은 검증통계량의 선정에 관한 부분입니다. 피셔 자신이 그러했듯이 우리들도 많은 경우에서 무엇을 검증통계량으로 해야 할 것인가를 직관적으로 알 수 있습니다. 예컨대, 티 실험에서 맞춘 잔의 수를 검증통계량으로 하는 것은 너무나 당연하지 않습니까? 그러나 당연한 듯하면서도 그렇지 않을 수가 있고 어떤 경우엔 뾰족한 생각이 나지 않을 수 있기 때문에 검증에 관한 체계적 이론이 필요합니다.

$X_1, \cdots, X_n$ 이 확률(밀도)함수 $f(x;\theta)$ 로부터의 iid 확률변수라고 합시다. 영가설과 대안가설이 단순한 경우를 생각해 봅시다 (어떤 문제든지 가장 간단한 경우부터 해결하는 것이 좋습니다).

$$\mathrm{H}_0 : \theta = 0 \quad \text{대} \quad \mathrm{H}_1 : \theta = 1.$$

무엇을 검증통계량으로 하는 것이 가장 좋을까요? 여기서 '좋음'을 어떻게 정할까요?

네이만(J. Neyman, 1894-1981)과 피어슨(E.S. Pearson, 1895-1980)은 1920년대 후반 각 가설 하에서의 가능도를 상대적으로 비교하자는 아이디어를 냈습니다. 즉

$$L(\theta; x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

라고 하고, 가능도의 비(比, ratio)

$$\lambda(x_1, \cdots, x_n) = \frac{L(1; x_1, \cdots, x_n)}{L(0; x_1, \cdots, x_n)}$$

를 보자는 것이지요. 만약 $\lambda(x_1, \cdots, x_n)$ 이 1보다 크다면 가설 H_1이 가설 H_0보다 가능성이 크다는 증거라고 할 수 있을 것입니다. 반대로 $\lambda(x_1, \cdots, x_n)$ 이 1보다 작다면 가설 H_1이 가설 H_0보다 가능성이 작습니다. 그러므로 가능도비 $\lambda(x_1, \cdots, x_n)$ 이 일정 수준을 넘어가는 경우 영가설 H_0를 기각하는 방법을 생각할 수 있습니다.

$$\text{영가설 } \mathrm{H}_0\text{를 기각} \quad \Leftrightarrow \quad \lambda(x_1, \cdots, x_n) \geq k, \quad k > 0 \text{ 은 상수.}$$

이런 검증방법을 가능도 비 검증(可能度比 檢證; likelihood ratio test, LRT)이라고 합니다. 즉 가능도 비 자체가 검증통계량이 되는 것입니다.

예를 들어 볼까요. $X_1, \cdots, X_n$ 이 정규분포 $\mathrm{N}(\theta, \sigma^2)$, σ^2 이 기지(既知)인 경우를 생각해봅시다 ($\mathrm{H}_0 : \theta = 0$ 대 $\mathrm{H}_1 : \theta = 1$). 가능도비가

$$\lambda(x_1, \cdots, x_n) = \frac{L(1; x_1, \cdots, x_n)}{L(0; x_1, \cdots, x_n)}$$

$$= \frac{\exp\left\{-\sum_{i=1}^{n}(x_i-1)^2/2\sigma^2\right\}}{\exp\left\{-\sum_{i=1}^{n}x_i^2/2\sigma^2\right\}} = \exp\left\{\left(\sum_{i=1}^{n}x_i - n/2\right)/\sigma^2\right\} \geq k$$

인 경우에 영가설 H_0를 기각하자는 것이 LRT입니다. 즉, LRT의 기각역은

$$\sum_{i=1}^{n} x_i \geq n/2 + \sigma^2 \log_e k, \quad \therefore \bar{x} \geq c\ (=\text{상수})$$

입니다. 다시 말하자면, 평균을 검증통계량으로 한다는 것이 되겠습니다.

직관적으로 가능도비 검증이 타당해 보이지만 왜 가능도비 검증이 가장 좋은 검증이라는 것일까요? 그러한 주장에 앞서 '가장 좋은' 검증의 의미를 명확히 할 필요가 있겠습니다.

<u>수준 α의 기각역</u>(棄却域, critical region of level α) : 정의

영가설 H_0 아래서 영가설을 기각하게 되는 경우에 대한 확률이 α 이하인 표본영역 C를 수준 α의 기각역이라고 합니다. 즉

$$P\{(X_1, \cdots, X_n) \in C;\ H_0\} \leq \alpha.$$ ■

<u>최량 기각역</u> (最良 棄却域, best critical region) : 정의

단순가설 H_0 대 단순가설 H_1의 검증에서 수준 α의 어떤 기각역 C^*에 대하여도

$$P\{(X_1, \cdots, X_n) \in C;\ H_1\} \geq P\{(X_1, \cdots, X_n) \in C^*;\ H_1\}$$

을 만족하는 수준 α의 기각역 C를 수준 α의 최량 기각역이라고 합니다. 여기서 '최량'을 이렇게 정의하는 이유는, 영가설 H_0 하에서는 일정수준 이하의 빈도로만 영가설 H_0를 기각하되 대안가설 H_1 하에서는 되도록 자주 영가설 H_0가 기각되도록 하자는 것입니다. 그리고, 이런 기각역 C에 의한 검증을 최강력 검증(most powerful test)이라고 합니다. ■

<u>네이만-피어슨 보조정리</u> (Neyman-Pearson Lemmma) :

단순가설 $H_0: \theta = 0$ 대 단순가설 $H_0: \theta = 1$의 검증에서

(i) $\lambda(x_1, \cdots, x_n) = \dfrac{L(1; x_1, \cdots, x_n)}{L(0; x_1, \cdots, x_n)} \geq k \quad \text{for } (x_1, \cdots, x_n) \in C,$

(ii) $\lambda(x_1, \cdots, x_n) = \dfrac{L(1; x_1, \cdots, x_n)}{L(0; x_1, \cdots, x_n)} \leq k \quad \text{for } (x_1, \cdots, x_n) \in \overline{C}$

이면서 (단, $k > 0$)

(iii) $$P\{(X_1, \cdots, X_n) \in C; \theta = 0\} = \alpha$$

인 기각역 C가 수준 α의 최량 기각역입니다. 그러므로 가능도비 검증(LRT)보다 더 좋은 검증방법은 없다는 것이 이 보조정리의 핵심 내용입니다.

기각역 C가 최량 기각역의 조건을 만족하는지 꼼꼼히 검토하다보면 저절로 증명이 됩니다. C^*를 수준이 α인 임의의 기각역이라고 합시다. 즉

$$P\left\{(X_1, \cdots, X_n) \in C^*; \theta = 0\right\} \leqq \alpha.$$

이 가정하에서, 우리가 보여야 할 것은

$$P\{(X_1, \cdots, X_n) \in C; \theta = 1\} - P\left\{(X_1, \cdots, X_n) \in C^*; \theta = 1\right\} \geq 0$$

입니다. 그런데

$$\begin{aligned}
&P\{(X_1, \cdots, X_n) \in C; \theta = 1\} - P\left\{(X_1, \cdots, X_n) \in C^*; \theta = 1\right\} \\
&= \int_C L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n - \int_{C^*} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&= \int_{C \cap C^*} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n + \int_{C \cap \overline{C^*}} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&\quad - \left(\int_{C \cap C^*} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n + \int_{\overline{C} \cap C^*} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n\right) \\
&= \int_{C \cap \overline{C^*}} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n - \int_{\overline{C} \cap C^*} L(1; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \equiv \bigstar
\end{aligned}$$

입니다. 그런데

$C \cap \overline{C^*}$ 에서는 $L(1; x_1, \cdots, x_n) \geq k \cdot L(0; x_1, \cdots, x_n)$,

$\overline{C} \cap C^*$ 에서는 $L(1; x_1, \cdots, x_n) \leq k \cdot L(0; x_1, \cdots, x_n)$

이므로 $\bigstar \geq$

$$\begin{aligned}
&\int_{C \cap \overline{C^*}} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n - \int_{\overline{C} \cap C^*} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&= \int_{C \cap C^*} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n + \int_{C \cap \overline{C^*}} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&\quad - \left(\int_{C \cap C^*} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n + \int_{\overline{C} \cap C^*} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n\right) \\
&= \int_C k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n - \int_{C^*} k \cdot L(0; x_1, \cdots, x_n)\, dx_1 \cdots dx_n \\
&= k \cdot P\{(X_1, \cdots, X_n) \in C; \theta = 0\} - k \cdot P\left\{(X_1, \cdots, X_n) \in C^*; \theta = 0\right\}
\end{aligned}$$

$\geqq k\alpha - k\alpha = 0$

입니다. 그러므로 증명이 끝났습니다. ■

계속되는 예로서 $X_1, \cdots, X_n$이 정규분포 $\mathrm{N}(\theta, \sigma^2)$, σ^2이 알려진(旣知인) 경우,

$$\mathrm{H}_0 : \theta = 0 \quad \text{대} \quad \mathrm{H}_1 : \theta = 1$$

에 대한 가능도비 검증의 기각역

$$\overline{X} \geq c$$

는 수준 $\alpha = P\{\overline{X} \geq c ; \theta = 0\} = 1 - \Phi(\sqrt{n}\, c / \sigma)$에서 최량입니다. 여기서, 주어진 α에 대하여 임계값 c가

$$c = z_\alpha \frac{\sigma}{\sqrt{n}}$$

으로 결정된다는 것을 알 수 있습니다 (단, $1 - \Phi(z_\alpha) = \alpha$). 예컨대, $\alpha = 0.05$ (즉 $z_\alpha = 1.645$), $\sigma = 1$, $n = 9$이면 $c = 0.548$입니다. $\mathrm{H}_1 : \theta = 1$하에서 이 검증은

$$P\{\overline{X} \geq z_\alpha \sigma / \sqrt{n} ; \theta = 1\}$$

$$= P\left\{\frac{\overline{X} - 1}{\sigma / \sqrt{n}} \geq \frac{z_\alpha \sigma / \sqrt{n} - 1}{\sigma / \sqrt{n}} ; \theta = 1\right\}$$

$$= 1 - \Phi(z_\alpha - \sqrt{n} / \sigma) = \Phi(\sqrt{n} / \sigma - z_\alpha)$$

라는 기각확률을 갖습니다. 예컨대, $\alpha = 0.05$ (즉 $z_\alpha = 1.645$), $\sigma = 1$, $n = 9$이면 이 기각확률은 0.912입니다. 대안가설 하에서의 기각확률을 특별히 <u>검증력</u>(檢證力, testing power)이라고 정의합니다. 이 예에서, 네이만-피어슨 보조정리를 따라, 수준 0.05의 검증 중에서 0.912보다 더 큰 검증력을 갖는 검증은 없습니다.

네이만-피어슨 보조정리를 테스트해보기 위하여[1] 앞의 예에서 중위수

$$\widetilde{X} = \mathrm{median}(X_1, \cdots, X_n)$$

에 의한 5% 수준의 검증을 세워보고 또 검증력을 계산하여 봅시다 ($\sigma = 1$, $n = 9$). 기각역을 $\widetilde{X} \geq \tilde{c}$라고 합시다. 그러면 이 검증의 수준은

$$P\{\widetilde{X} \geq \tilde{c} ; \theta = 0\} = \sum_{i=0}^{4} \binom{9}{i} \Phi(\tilde{c})^i (1 - \Phi(\tilde{c}))^{9-i}$$

가 됩니다. 이 값이 0.05가 되는 $\tilde{c}$를 시행착오로 찾아보면 대략 $\tilde{c} = 0.67$이 됩니다. 즉 $\widetilde{X} \geq 0.67$일 때 영가설 $\mathrm{H}_0 : \theta = 0$을 기각하는 검증의 수준은 0.05입니다. 그리고

1) 이것은 이미 수리적으로 증명이 되었기 때문에 소용이 없는 짓이기는 하지만 때로는 바보짓도 필요하지 않을까요?

이 검증의 검증력은

$$P\{\tilde{X} \geq \tilde{c}\,;\theta = 1\} = \sum_{i=0}^{4}\binom{9}{i}\Phi(\tilde{c}-1)^{i}(1-\Phi(\tilde{c}-1))^{9-i}$$

입니다. 이것은, $\tilde{c} = 0.67$일 때, 0.79로 계산됩니다. 평균 $\overline{X}$에 의한 5% 수준의 검증이 0.912의 검증력을 갖는 데 비교하여 물론 작습니다.

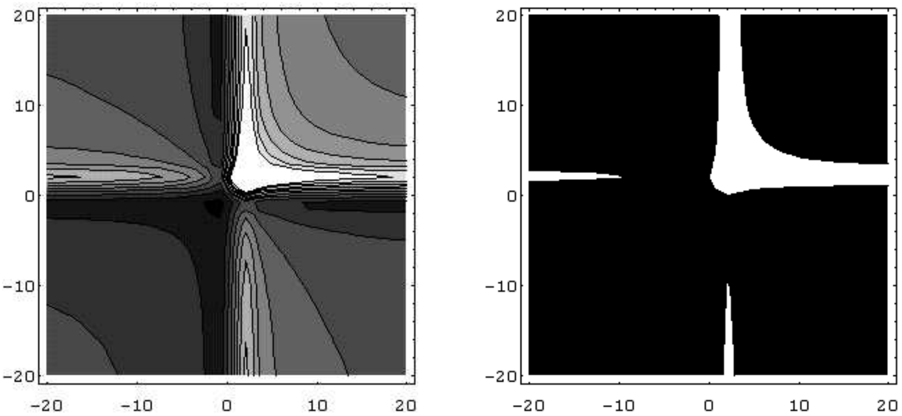
네이만-피어슨 보조정리를 활용하는 예 하나를 더 들어보기로 하겠습니다. X_1과 X_2를 코쉬분포 Cauchy(θ)로부터의 iid 확률변수라고 합시다. 따라서

$$L(\theta\,;x_1,x_2) = \frac{1}{\pi^2}\frac{1}{1+(x_1-\theta)^2}\frac{1}{1+(x_2-\theta)^2}$$

입니다. 이 때, $\mathrm{H}_0 : \theta = 0$ 대 $\mathrm{H}_1 : \theta = 1$에 대하여 검증을 해봅시다. 어떻게 하는 것이 좋습니까? 한 방법은 가능도비 검증을 유도해보는 것입니다. 이 경우, 가능도비는

$$\lambda(x_1,x_2) = \frac{L(1\,;x_1,x_2)}{L(0\,;x_1,x_2)} = \frac{1+x_1^2}{1+(x_1-1)^2}\frac{1+x_2^2}{1+(x_2-1)^2}$$

입니다. <그림 2>는 가능도비 $\lambda(x_1,x_2)$에 대한 등고선 그림(contour plot)입니다. 적어도 3개의 봉우리를 갖는 등 특이한 패턴을 볼 수 있습니다. 또한 가능도비에 대한 어떤 임계값 $k > 0$에 대하여는 기각역이 하나로 연결되지 않습니다. 그럼에도 불구하고 이것이 네이만-피어슨 정리에 따라 해당하는 수준의 최강력 검증이 됩니다.



<그림 2> 코쉬 가능도비에 대한 등고선 그림 (밝은 색일수록 큰 값을 갖음) :
오른쪽 그림은 $k = 2$인 경우의 기각역 (밝은 부분)

6.3 균일최강력검증(Uniformly Most Powerful Test)

앞 절에서는 단순가설 대 단순가설의 검증, 예컨대

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta = \theta_1$$

을 다루었습니다만 실제 검증 문제에서는 이와 같이 진짜 단순한 경우는 거의 없습니다. 대신 1절에서 본 바와 같이 대안가설이 복합적인 것이 일반적입니다. 그러므로 실제로는

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta > \theta_0$$

에 대한 검증을 다루어야 합니다. 이런 경우에서 진짜 바람직한 검증은 균일하게 힘이 좋은 검증방법입니다.

<u>균일 최강력 기각역</u> (均一最强力棄却域, uniformly most powerful critical region) : 정의

단순가설 H_0 대 복합가설 H_1의 검증에서 어떤 수준 α의 기각역 C^*에 대하여도

$$P\{(X_1, \cdots, X_n) \in C ; \theta_1\} \geq P\{(X_1, \cdots, X_n) \in C^* ; \theta_1\}, \text{ for } all\ \theta_1 > \theta_0$$

을 만족하는 수준 α의 기각역 C를 수준 α의 균일최강력기각역이라고 합니다. 그리고, 이런 기각역 C에 의한 검증을 균일최강력검증(uniformly most powerful test; 약어로 UMP test)이라고 합니다. ■

<u>검증력 함수</u> (檢證力 函數, power function) : 정의

영가설 H_0의 기각확률을 파라미터 θ에 대한 함수로 표현한 것을 검증력 함수 $\beta(\theta)$라고 합니다. 즉 앞의 검증 문제에서 기각역이 C인 검증에 대한 검증력 함수는

$$\beta(\theta) = P\{(X_1, \cdots, X_n) \in C ; \theta\} \quad \text{for } \theta \geq \theta_0$$

이며 $\beta(\theta_0) = \alpha$ 입니다. UMP 검증의 검증력 함수 $\beta(\theta)$는 수준 α의 다른 어떤 검증력함수 $\beta^*(\theta)$보다 모든 θ에서 작지 않습니다. 즉

$$\beta(\theta) \geq \beta^*(\theta), \text{ for } all\ \theta \geq \theta_0.$$

예를 들어보기로 하겠습니다. $X_1, \cdots, X_n$을 지수분포 Exponential(θ)로부터의 iid 확률변수라고 합시다. 그리고

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta > \theta_0$$

를 검증하여 보겠습니다. 문제가 만약

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{대} \quad \mathrm{H}_1 : \theta = \theta_1$$ (여기서 θ_1 은 θ_0 보다 큰 정해진 수)

이라면 가능도비에 의하여 최량기각역이 유도될 것입니다. 즉

$$\lambda(x_1, \cdots, x_n) = \frac{L(\theta_1; x_1, \cdots, x_n)}{L(\theta_0; x_1, \cdots, x_n)} = \left(\frac{\theta_0}{\theta_1}\right)^n \frac{\exp\{-\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}/\theta_1\}}{\exp\{-\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}/\theta_0\}}$$
$$= \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \cdot \sum_{i=1}^n x_i\right\}$$

이므로, $\theta_1 > \theta_0$ 인 한에 있어서는 (즉, $\frac{1}{\theta_0} > \frac{1}{\theta_1}$), θ_1 에 관계없이 최량기각역이

$$\sum_{i=1}^n x_i \geq c$$

의 형태가 됩니다. $\mathrm{H}_0 : \theta = \theta_0$ 하에서 $\sum_{i=1}^n X_i / \theta_0$ 가 감마분포 Gamma$(n, 1)$ 을 따르므로 (또는 $2\sum_{i=1}^n X_i / \theta_0$ 가 카이제곱분포 $\chi^2(2n)$ 을 따르므로), 검증의 수준을 α 로 하기 위해서는

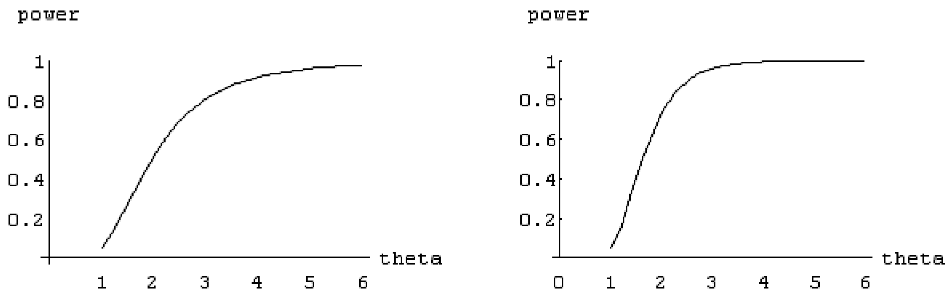
$$c = \theta_0 \, \Gamma_{n,1,\alpha} = \theta_0 \, \chi^2_{2n,\alpha} / 2$$

로 두어야 합니다 (여기서 $\Gamma_{n,1,\alpha}$ 는 Gamma$(n,1)$ 분포의 상위 α 분위수, $\chi^2_{2n,\alpha}$ 는 $\chi^2(2n)$ 분포의 상위 α 분위수를 표기함). 이 기각역은 모든 $\theta_1 (> \theta_0)$ 에 대하여 최량이므로 균일최강력기각역이라고 할 수 있습니다.

앞의 문제에서 $\theta (\geqq \theta_0)$ 에서 검증력 함수 $\beta(\theta)$ 를 구해보기로 합시다. 그것은

$$\beta(\theta) = P\left\{\sum_{i=1}^n X_i \geq \theta_0 \, \Gamma_{n,1,\alpha} ; \theta\right\}$$
$$= P\left\{\sum_{i=1}^n X_i / \theta \geq (\theta_0/\theta)\, \Gamma_{n,1,\alpha} ; \theta\right\}$$
$$= P\{G_{n,1} \geq (\theta_0/\theta)\, \Gamma_{n,1,\alpha}\} \left(= \int_{(\theta_0/\theta)\Gamma_{n,1,\alpha}}^{\infty} \frac{1}{\Gamma(n)} u^{n-1} e^{-u} \, du\right)$$

가 됩니다 (여기서 $G_{n,1}$ 은 Gamma$(n,1)$ 분포를 따르는 확률변수입니다). <그림 3>의 왼쪽 그림은 $\theta_0 = 1, \alpha = 0.05, n = 5$ 인 경우의 검증력 함수입니다. 그리고 오른 쪽 그림은 $\theta_0 = 1, \alpha = 0.05, n = 10$ 인 경우의 검증력 함수입니다. $n = 10$ 인 경우의 검증력이 모든 $\theta (\geq \theta_0)$ 에서 $n = 5$ 인 경우의 검증력에 비하여 큰 것을 볼 수 있습니다. 표본크기가 커짐에 따라 검증력이 커지는 것은 대부분의 검증이 갖는 일반적인 성질입니다.



<그림 3> 지수분포에서의 단측 검증에 대한 검증력 함수 :

왼쪽 그림 $n = 5$ 인 경우, 오른쪽 그림 $n = 10$ 인 경우

$\mathrm{H}_0 : \theta = \theta_0$ 대 $\mathrm{H}_1 : \theta > \theta_0$ 의 검증문제에서 UMP 검증이 존재하는 대표적인 경우를 보기로 하겠습니다. $X_1, \cdots, X_n$ 이 확률(밀도)함수 $f(x;\theta)$ 로부터의 iid 확률변수라고 합시다. 그리고 $T = t(X_1, \cdots, X_n)$ 이 파라미터 θ 에 대한 충분통계량이라고 합시다. 그러면 가능도비는, $\theta_1 > \theta_0$ 에 대하여,

$$\begin{aligned} \lambda(x_1, \cdots, x_n) &= \frac{L(\theta_1; x_1, \cdots, x_n)}{L(\theta_0; x_1, \cdots, x_n)} \\ &= \frac{g(T(x_1, \cdots, x_n); \theta_1) \cdot h(x_1, \cdots, x_n)}{g(T(x_1, \cdots, x_n); \theta_0) \cdot h(x_1, \cdots, x_n)} \\ &= \frac{g(T(x_1, \cdots, x_n); \theta_1)}{g(T(x_1, \cdots, x_n); \theta_0)} \end{aligned}$$

입니다. 어느 $\theta_1 (> \theta_0)$ 에 대하여도 $\lambda = \dfrac{g(T(x_1, \cdots, x_n); \theta_1)}{g(T(x_1, \cdots, x_n); \theta_0)}$ 가 $t = T(x_1, \cdots, x_n)$ 의 증가함수이면

$$T(x_1, \cdots, x_n) \geq c$$

가 균일최강력기각역이 되는 것을 알 수 있습니다. (만약 λ 가 t 의 감소함수이면 균일최강력기각역은 $T(x_1, \cdots, x_n) \leq c$ 의 형태가 됩니다).

예로서, 앞의 지수분포 Exponential(θ) 사례에서 가능도비

$$\lambda(x_1, \cdots, x_n) = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \cdot \sum_{i=1}^n x_i\right\}, \qquad \text{for } \theta_1 > \theta_0$$

는 $t = \sum_{i=1}^n x_i$ 의 단조증가함수입니다. 그러므로 UMP 검증이 $\sum_{i=1}^n x_i \geq c$ 의 형태로 존재하게 되는 것입니다.

다른 한 예를 보기로 하겠습니다. $X_1, \cdots, X_n$ 을 지수분포 Exponential$(\theta, 1)$ 로부터의 iid 확률변수라고 합시다. 즉

$$f(x;\theta) = \exp\{-(x-\theta)\}, \quad x \ge \theta$$
$$= \exp\{-(x-\theta)\} \cdot 1(x \ge \theta)$$

입니다. 이 때

$$H_0 : \theta = \theta_0 \qquad \text{대} \quad H_1 : \theta > \theta_0$$

를 검증하여 보겠습니다. $\theta_1 (> \theta_0)$ 에 대하여 가능도비는

$$\lambda(x_1, \cdots, x_n) = \frac{\exp\left\{-\sum_{i=1}^{n}(x_i - \theta_1)\right\} \cdot 1\left\{\min(x_1, \cdots, x_n) \ge \theta_1\right\}}{\exp\left\{-\sum_{i=1}^{n}(x_i - \theta_0)\right\} \cdot 1\{\min(x_1, \cdots, x_n) \ge \theta_0\}}$$

$$= \begin{cases} 0, & \text{if } \theta_0 \le \min(x_1, \cdots, x_n) < \theta_1 \\ \exp\{n(\theta_1 - \theta_0)\}, & \text{if } \min(x_1, \cdots, x_n) \ge \theta_1 \end{cases}$$

입니다. 따라서 가능도비 $\lambda(x_1, \cdots, x_n)$ 는 $t = \min(x_1, \cdots, x_n)$ 의 계단 꼴 증가함수입니다. 그러므로 UMP 검증의 기각역은

$$t = \min(x_1, \cdots, x_n) \ge c$$

의 형태가 됩니다. 이제 이 검증의 수준 α를 갖기 위해서 c가 얼마여야 하는지를 정해보도록 합시다. 영가설 $H_0 : \theta = \theta_0$ 하에서 $T = \min(X_1, \cdots, X_n)$ 이 취하는 분포는 Exponential$(\theta_0, 1/n)$ 입니다. 즉

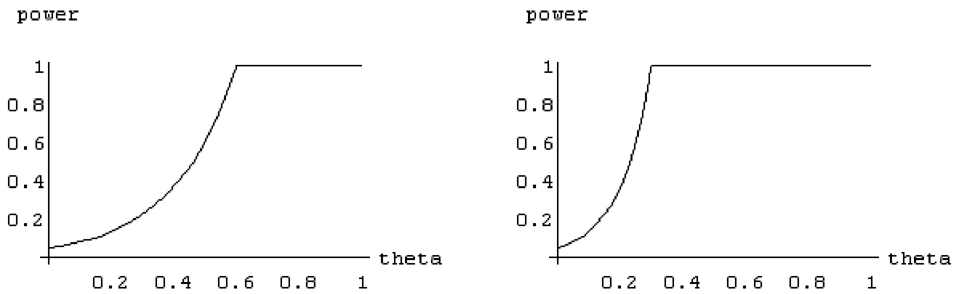
$$f_T(t;\theta_0) = n \exp\{-n(t-\theta_0)\}, \quad t \ge \theta_0$$

입니다. 따라서

$$c = \theta_0 - \frac{1}{n}\log_e \alpha$$

여야 합니다. 예컨대 $\alpha = 0.05$ 인 경우 $c = \theta_0 + \dfrac{2.996}{n}$ 입니다. 한편, 이 UMP 검증의 검증력 함수는 $\theta \ge \theta_0$에 대하여 다음과 같습니다. <그림 4>를 보십시오.

$$\beta(\theta) = P\{\min(X_1, \cdots, X_n) \ge c;\theta\},$$
$$= \begin{cases} \exp\{-n(c-\theta)\} = \alpha \cdot \exp\{n(\theta - \theta_0)\}, & \text{if } \theta_0 \le \theta \le c, \\ 1, & \text{if } \theta > c. \end{cases}$$

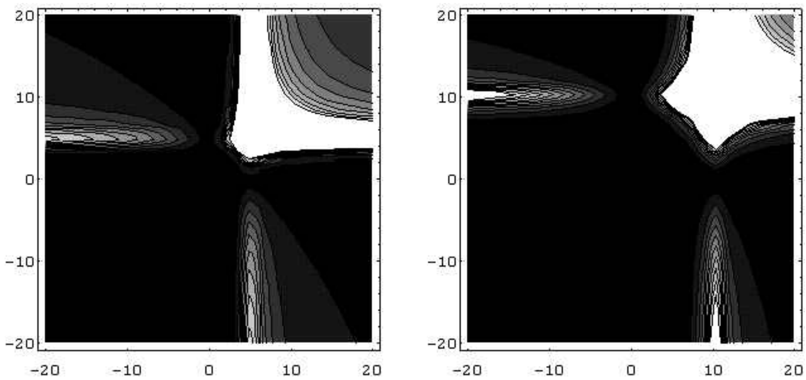


<그림 4> 지수분포에서 위치모수에 대한 검증력 함수 : $\theta_0 = 0$ 에 대하여
왼쪽 그림 $n = 5$ 인 경우, 오른쪽 그림 $n = 10$ 인 경우

$\mathrm{H}_0 : \theta = \theta_0$ 대 $\mathrm{H}_1 : \theta > \theta_0$ 의 검증문제에서 UMP 검증이 존재하지 않는 경우도 있을까요? 물론입니다 (확률함수가 지수족에 속하지 않는 경우 흔한 일입니다). X_1, X_2 가 Cauchy(θ) 분포로부터의 iid 확률변수라고 합시다. $\theta = \theta_0$ 대 $\theta = \theta_1$ 의 가능도비는

$$\lambda(x_1, x_2) = \frac{L(\theta_1 ; x_1, x_2)}{L(\theta_0 ; x_1, x_2)} = \frac{1+(x_1-\theta_0)^2}{1+(x_1-\theta_1)^2} \frac{1+(x_2-\theta_0)^2}{1+(x_2-\theta_1)^2}$$

입니다. <그림 5>는 <그림 2>에 이어 $\theta_1 = 5$ 와 $\theta_1 = 10$ 의 경우에 대한 가능도비 $\lambda(x_1, x_2)$ 에 대한 등고선 그림(contour plot)입니다 ($\theta_0 = 0$). 등고선의 모양이 상당히



<그림 5> 코쉬 가능도비에 대한 등고선 그림 (밝은 색일수록 큰 값을 갖음) :
왼쪽은 $\theta_1 = 5$ 에 대한 것이고 오른쪽 그림은 $\theta_1 = 10$ 에 대한 것

바뀌는 것을 볼 수 있습니다. 이런 경우에는 θ_1의 값에 관계없이 균일한 최강력 검증이 존재하지 않습니다.

6.4 비편향 검증(Unbiased Test)

영가설 $\mathrm{H}_0 : \theta = \theta_0$에 대한 대안가설이 양측인 검증문제가 있습니다. 과학적 주장으로서 양측가설 $\mathrm{H}_1 : \theta \neq \theta_0$ (즉 $\theta > \theta_0$ 또는 $\theta < \theta_0$)은 단측가설인 $\mathrm{H}_{1a} : \theta > \theta_0$이나 $\mathrm{H}_{1b} : \theta < \theta_0$에 비하여 많이 허술합니다. 방향성이 제시되어 있지 않기 때문이죠. 따라서 양측가설은 자료관측에 앞서 설정된 연구가설이기 보다는 사후적으로 (즉, 자료분석 단계에서) 설정되는 가설이기가 쉽습니다.

대안가설이 양측인 경우 비편향 검증이라는 질적 평가기준을 생각해봅니다.

<u>비편향 검증</u>(非偏向檢證, unbiased test) : 정의

영가설 H_0의 기각영역을 C라고 할 때

$$P\{(X_1, \cdots, X_n \in C ; \theta\} \geq \alpha, \text{ for } all\ \theta$$

를 만족하는 검증을 비편향 검증이라고 합니다. 다시 말하자면 검증에서의 비편향성이란 검증력이 모든 θ에서 검증의 수준 α보다 작지 않아야 한다는 것입니다. ■

비편향검증이 아닌 예를 들어 보겠습니다. $X_1, \cdots, X_n$을 지수분포 Exponential(θ) 로부터의 iid 확률변수라고 합시다. 그리고

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{대} \quad \mathrm{H}_1 : \theta \neq \theta_0$$

를 검증하기로 합시다. 이와 같은 대안가설이 양측인 경우에 쉽게 생각할 수 있는 검증은 검증의 수준 α를 둘로 쪼개어 절반은

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{대} \quad \mathrm{H}_{1a} : \theta > \theta_0$$

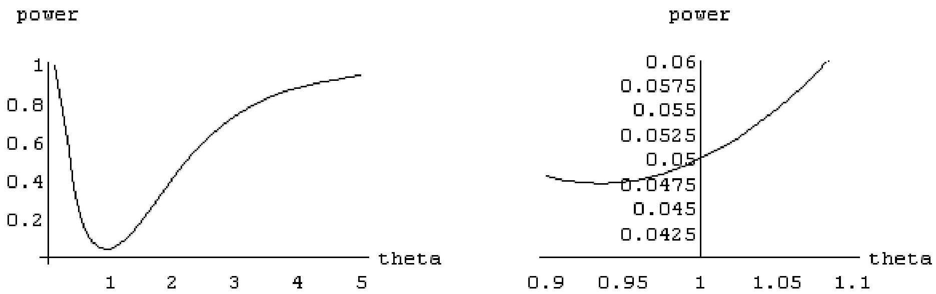
에 배당하고 남은 절반은

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{대} \quad \mathrm{H}_{1b} : \theta < \theta_0$$

에 배당하는 것이겠습니다. 그리고 같은 꼬리확률의 단측검증을 결합하여 기각역

$$C \ : \ \sum_{i=1}^{n} x_i \geq \theta_0\, \Gamma_{n,1,\alpha/2} \quad \text{or} \quad \sum_{i=1}^{n} x_i \leq \theta_0\, \Gamma_{n,1,1-\alpha/2}$$

을 만듭니다 (6.3절 참조). 과연 이 검증이 비편향적일까요? 검증력 함수는



<그림 6> 지수분포에서의 같은 꼬리 양측검증의 검증력 함수 : $n = 5$

오른쪽 그림은 $\theta = 1$ 근처에서의 세부도

$$\beta(\theta) = P\left\{ \sum_{i=1}^{n} X_i \le \theta_0 \, \Gamma_{n,1,1-\alpha/2},\ \sum_{i=1}^{n} X_i \ge \theta_0 \, \Gamma_{n,1,\alpha/2} \ ;\theta \right\}$$

$$= 1 - P\left\{ \frac{\theta_0}{\theta} \, \Gamma_{n,1,1-\alpha/2} \le \frac{1}{\theta} \sum_{i=1}^{n} X_i \le \frac{\theta_0}{\theta} \, \Gamma_{n,1,\alpha/2} \ ;\theta \right\}$$

$$= 1 - P\left\{ \frac{\theta_0}{\theta} \, \Gamma_{n,1,1-\alpha/2} \le G_{n,1} \le \frac{\theta_0}{\theta} \, \Gamma_{n,1,\alpha/2} \ ;\theta \right\}$$

입니다 (여기서 $G_{n,1}$ 은 Gamma$(n, 1)$ 확률변수). $n = 5, \alpha = 0.05$, $\theta_0 = 1$로 놓고 $\beta(\theta)$를 그려보기로 하겠습니다. <그림 6>을 보십시오. 왼쪽 그림에서는 $\theta_0 = 1$이 검증력 함수의 저점인지 아닌지 확실하지 않습니다. 그러나 이 부분을 확대한 오른쪽 그림을 보면 $\theta_0 = 1$이 검증력 함수의 저점이 아님을 알 수 있습니다. 따라서 이 양측검증은 비편향적이 아닙니다.

다른 예를 보겠습니다. $X_1, \cdots, X_n$ 을 정규분포 N(θ, σ^2) (σ^2 旣知)로부터의 iid 확률변수라고 합시다. 그리고

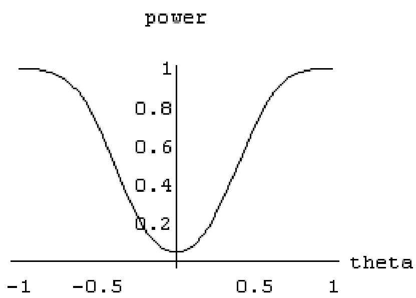
$$\mathrm{H}_0 : \theta = \theta_0 \qquad \text{대} \quad \mathrm{H}_1 : \theta \ne \theta_0$$

을 검증하기로 합시다. 이에 대해 잘 알려져 있는 양측검증은

$$C \ : \ \bar{x} \ge \theta_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \ \text{ 또는 } \bar{x} \le \theta_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

을 기각역으로 합니다. 이 검증의 검증력 함수는

$$\beta(\theta) = 1 - P\left\{ \theta_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \theta_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \ ;\theta \right\}$$



<그림 7> 정규분포에서 양측검증의 검증력 함수

$$= 1 - P\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) - z_{\alpha/2} \le \frac{\sqrt{n}}{\sigma}(\bar{x} - \theta) \le \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + z_{\alpha/2} ; \theta \right\}$$

$$= 1 - P\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) - z_{\alpha/2} \le Z \le \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + z_{\alpha/2} \right\}$$

$$= 1 + \Phi\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) - z_{\alpha/2} \right\} - \Phi\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + z_{\alpha/2} \right\}$$

입니다 (여기서 $\Phi(z)$ 는 표준정규분포 N(0,1) 의 분포함수임). <그림 7>을 보십시오 ($\theta_0 = 0,\ n = 25,\ \sigma = 1$ 인 경우). $\beta(\theta)$ 를 θ 로 미분하면

$$\beta'(\theta) = -\frac{\sqrt{n}}{\sigma} \cdot \phi\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) - z_{\alpha/2} \right\} + \frac{\sqrt{n}}{\sigma} \cdot \phi\left\{ \frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + z_{\alpha/2} \right\}$$

이므로 (여기서 $\phi(z)$ 는 표준정규분포 N(0,1) 의 확률밀도함수임)

$$\beta'(\theta_0) = -\frac{\sqrt{n}}{\sigma} \cdot \phi(-z_{\alpha/2}) + \frac{\sqrt{n}}{\sigma} \cdot \phi(z_{\alpha/2}) = 0$$

입니다. 그러므로 이 검증력 함수는

$$\beta(\theta) \ge \beta(\theta_0) = \alpha \quad \text{for } all \ \theta$$

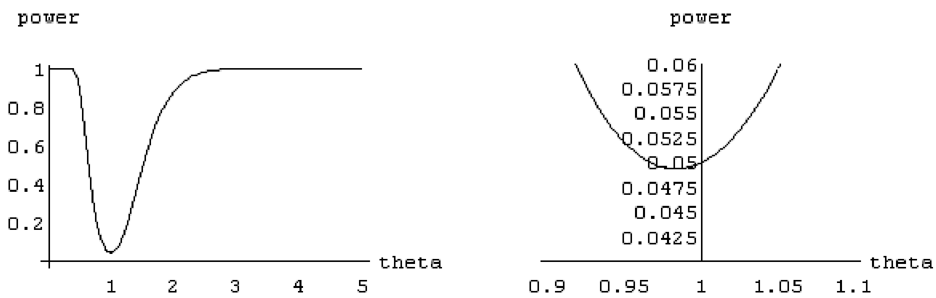
가 됩니다. 다시 말하여, 이 검증은 비편향적입니다.

지수족의 확률분포에 대하여는 다음 정리가 알려져 있습니다 (증명 생략).

<u>지수족과 균일 최강력 비편향 검증</u> : 정리

$X_1, \cdots, X_n$ 이 1-모수 지수족(one-parmeter exponential family)의 확률(밀도)함수

$$f(x;\theta) = \exp\{ c(\theta) P(x) + d(\theta) + Q(x) \} \quad \text{for } x \in A$$



<그림 8> 지수분포에서의 같은 꼬리 양측검증의 검증력 함수 : $n = 20$

오른쪽 그림은 $\theta = 1$ 근처에서의 세부도

로부터의 iid 확률변수여서 $T(X_1, \cdots, X_n) = \sum_{i=1}^{n} P(X_i)$ 가 θ 에 대한 충분통계량이라고 합시다. 이 때 $H_0 : \theta = \theta_0$ 대 $H_1 : \theta \neq \theta_0$ 에 대한 검증을 생각합시다. 이런 경우, 기각역

$$C \ : \ T(x_1, \cdots, x_n) \ \geq \ c_1 \ \text{또는} \ T(x_1, \cdots, x_n) \ \leq \ c_2$$

중에서 비편향적인 것이 있으며 이 검증은 모든 비편향 검증들 중에서 균일하게 최대의 검증력을 갖습니다.

이런 성질을 갖는 검증을 균일최강력비편향검증(均一最强力非偏向檢證 ; uniformly most powerful unbiased test)이라고 합니다 (정의). ■

이 정리에 따라 앞서 다루었던 지수분포의 평균에 대한 양측검증에서도 기각역

$$C \ : \ \sum_{i=1}^{n} x_i \ \geq \ c_1 \quad \text{or} \quad \sum_{i=1}^{n} x_i \ \leq \ c_2$$

중에서 비편향적인 것이 있다는 것을 압니다만 임계값 c_1과 c_2를 찾기가 쉽지 않습니다. 그러나 다행스럽게도 검증의 수준 α를 절반씩 나누어 결합한 같은 꼬리 양측검증도 표본의 크기가 커지면 자연스레 거의 비편향적이 됩니다. <그림 8>을 <그림 6>과 비교해보십시오. 두 경우 모두 $\alpha = 0.05$, $\theta_0 = 1$ 이지만 <그림 8>은 $n = 20$ 인 경우이며 <그림 6>은 $n = 5$ 인 경우입니다. 표본크기 n 이 커짐에 따라 검증력 함수 $\beta(\theta)$ 의 최저점이 $\theta_0 = 1$ 에 가까워 졌음을 볼 수 있습니다.

6.5* 국소 최강력 검증(Locally Most Powerful Test)

다시 단측검증문제를 생각하기로 하겠습니다. 영가설 $H_0 : \theta = \theta_0$에 대한 대안가설이 $H_1 : \theta > \theta_0$라고 합시다. 이 때, 어떤 문제에서는 균일최강력검증(UMP test)이 존재하지 않습니다. 예컨대 코쉬 분포 Cauchy(θ)에서의 검증이 그렇다는 것을 이미 3절에서 본 바 있습니다.

이에 대한 대처 방법 중 하나는 $\theta = \theta_0$ 근처에서 가장 큰 검증력을 갖는 검증을 찾는 것입니다. (물론 이 방법이 가장 좋다고는 말할 수 없습니다. 비유적으로 이것은 달리기 시합에서 스타트를 가장 잘 하는 선수가 우승한다고 생각하는 것과 같기 때문입니다.) $X_1, \cdots, X_n$을 확률(밀도)함수 $f(x;\theta),\ \theta \geq \theta_0$로부터의 iid 확률변수라고 하고 검증의 수준을 α로 둡시다. 그러면, 기각역 C에 대하여

$$\int_C \prod_{i=1}^n f(x_i\,;\theta_0)\,dx_1 \cdots dx_n = \alpha \tag{1}$$

이고 검증력 함수는

$$\beta(\theta) = \int_C \prod_{i=1}^n f(x_i\,;\theta)\,dx_1 \cdots dx_n$$

입니다. $\beta(\theta)$를 θ로 미분하고 $\theta = \theta_0$에서 평가하면

$$\beta'(\theta_0) = \int_C \sum_{i=1}^n \left[\frac{\partial}{\partial\theta} f(x_i\,;\theta) \right]_{\theta=\theta_0} \prod_{i' \neq i} f(x_{i'}\,;\theta_0) \cdot dx_1 \cdots dx_n$$

$$= \int_C \sum_{i=1}^n \left[\frac{\partial}{\partial\theta} \log_e f(x_i\,;\theta) \right]_{\theta=\theta_0} \prod_{i=1}^n f(x_i\,;\theta_0) \cdot dx_1 \cdots dx_n$$

이 됩니다. 이것을 최대화하기 위해서 다음 보조정리를 활용합니다.

<u>일반화 네이만-피어슨 보조정리</u> (generalized Neyman-Pearson lemma) :

(1)의 제약 하에서

$$\int_C h(x_1, \cdots, x_n) \cdot \prod_{i=1}^n f(x_i\,;\theta_0)\,dx_1 \cdots dx_n$$

를 최대로 하는 기각역 C는

$$h(x_1, \cdots, x_n) \geq k, \quad \text{for } (x_1, \cdots, x_n) \in C$$

$$h(x_1, \cdots, x_n) \leq k, \quad \text{for } (x_1, \cdots, x_n) \in \overline{C}$$

로부터 얻어집니다 (단, $k \geq 0$). 증명은 다음과 같습니다.

C^*를 수준 α의 기각역이라고 하면

$$\int_C h(x_1, \cdots, x_n) \cdot \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n - \int_{C^*} h(x_1, \cdots, x_n) \cdot \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n$$

$$= \int_{C \cap \overline{C^*}} \text{'same ftn'} \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n - \int_{\overline{C} \cap C^*} \text{'same ftn'} \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n$$

$$\geq \int_{C \cap \overline{C^*}} k \cdot \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n - \int_{\overline{C} \cap C^*} k \cdot \prod_{i=1}^n f(x_i ; \theta_0)\, dx_1 \cdots dx_n$$

$$\geq k(\alpha - \alpha) = 0 \qquad \blacksquare$$

따라서 $\theta = \theta_0$ 근처에서 가장 큰 검증력을 갖는 국소 최강력 검증(局所最强力檢證, locally most powerful test)의 기각역 C는

$$C \ : \ \sum_{i=1}^n \left[\frac{\partial}{\partial\theta} \log_e f(x_i ; \theta) \right]_{\theta = \theta_0} \geq k$$

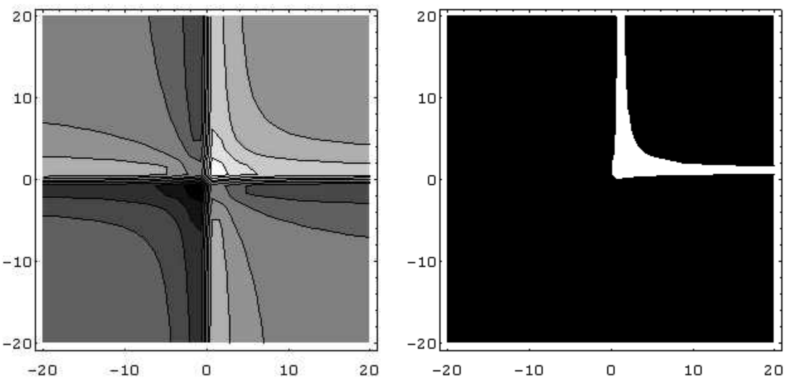
입니다 ($k \geq 0$). 예로서, X_1, X_2 이 코쉬분포 Cauchy(θ) 로 iid 확률변수인 경우

$$f(x ; \theta) = \frac{1}{\pi} \frac{1}{1 + (x_1 - \theta)^2}, \quad \frac{\partial}{\partial\theta} \log_e f(x ; \theta) = \frac{2(x_1 - \theta)}{1 + (x_1 - \theta)^2}$$

입니다. 따라서 $H_0 : \theta = 0$ 대 $H_1 : \theta > 0$ 에 대한 국소최강력검증의 기각역은

$$C \ : \ \frac{2x_1}{1 + x_1^2} + \frac{2x_2}{1 + x_2^2} \geq k, \quad k \geq 0$$

으로 주어집니다. <그림 9>의 등고선 그림을 보십시오.



<그림 9> 코쉬 국소최강력검증 통계량에 대한 등고선 그림 :
오른쪽 그림은 $k = 1$ 에 해당하는 기각역 (밝은 부분)

6.6 일반화 가능도 비 검증과 대표본 이론

이제까지는 파라미터가 1개인 경우를 다루었습니다만 이제부터는 파라미터가 2개 이상인 경우로 확장해보겠습니다. 임의표본 $X_1, \cdots, X_n$의 확률분포가 k개의 파라미터, 즉 $k \times 1$ 벡터 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)^t$에 의하여 표현된다고 합시다. 그리고

$$\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{대} \quad \mathrm{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

에 대한 검증문제를 생각하기로 합시다 (여기서 $\boldsymbol{\theta}_0$은 지정됨). 이런 경우 가능도 비를 일반화하여 검증 방법을 개발하는 데 활용하게 됩니다.

<u>일반화 가능도 비</u> (generalized likelihood ratio, GLR) : 정의

$$\Lambda(x_1, \cdots, x_n) = \frac{\max_{\Omega} L(\boldsymbol{\theta}\ ; x_1, \cdots, x_n)}{\max_{\omega} L(\boldsymbol{\theta}\ ; x_1, \cdots, x_n)} \ ,$$

여기서 $\Omega = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathrm{H}_0 \bigcup \mathrm{H}_1\}$는 파라미터 공간이고 $\omega = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathrm{H}_0\}$는 영가설이 한정하는 Ω의 일부분입니다 (이 경우에는 $\omega = \{\boldsymbol{\theta}_0\}$ 입니다). 따라서 일반화 가능도 비 $\Lambda(x_1, \cdots, x_n)$은

$$\Lambda(x_1, \cdots, x_n) = \frac{L(\hat{\boldsymbol{\theta}}\ ; x_1, \cdots, x_n)}{L(\boldsymbol{\theta}_0\ ; x_1, \cdots, x_n)} = \frac{L(\hat{\boldsymbol{\theta}})}{L(\boldsymbol{\theta}_0)}$$

와 같습니다 (여기서 $\hat{\boldsymbol{\theta}}$은 $\boldsymbol{\theta}$에 대한 mle임). 이것은 $\Lambda(x_1, \cdots, x_n)$이 $\boldsymbol{\theta}_0$와 $\hat{\boldsymbol{\theta}}$에서의 가능도를 상대적으로 비교한 지표임을 말해줍니다.

일반화 가능도 비 검증(generalized likelihood ratio test, GLRT)의 기각역은 다음과 같습니다.

$$C : 2 \log_e \Lambda(x_1, \cdots, x_n) \geq c, \quad c\text{는 상수}.$$

GLRT의 정의에서 일반화 가능도 비 $\Lambda(x_1, \cdots, x_n)$에 왜 2배의 로그 변환을 하는지는 곧 알게 될 것입니다. ■

예를 들어, X_i가 정규분포 $\mathrm{N}(\mu_i, \sigma^2)$을 독립적으로 따르는 확률변수라고 합시다 $(i = 1, \cdots, n)$. 이 때 $\mu_{k+1} = \cdots = \mu_n = 0$으로 정해져 있고 σ^2도 알려져 있다고 합시다. 그리고 영가설과 대안가설이 각각 다음과 같다고 합시다.

$$\mathrm{H}_0 : \mu_1 = 0, \cdots, \mu_k = 0 \quad \text{대} \quad \mathrm{H}_1 : \mu_1 \neq 0 \text{ or } \cdots \text{ or } \mu_k \neq 0 .$$

앞의 표기에 맞추면 $\boldsymbol{\theta} = (\mu_1, \cdots, \mu_k)^t$, $\boldsymbol{\theta}_0 = (0, \cdots, 0)^t$ 입니다. μ_i $(i = 1, \cdots, k)$에 대한 최대가능도추정값(mle)이 $\hat{\mu}_i = x_i$ $(i = 1, \cdots, k)$이므로

$$\Lambda(x_1, \cdots, x_n) = \frac{\left(\dfrac{1}{\sqrt{2\pi}\ \sigma}\right)^n \exp\left\{-\dfrac{1}{2\sigma^2}\sum_{i=k+1}^{n} x_i^2\right\}}{\left(\dfrac{1}{\sqrt{2\pi}\ \sigma}\right)^n \exp\left\{-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\}} = \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{k} x_i^2\right\}$$

이 됩니다. 따라서 GLR 검증의 기각역은

$$C: \quad 2\log_e \Lambda(x_1, \cdots, x_n) = \frac{1}{\sigma^2}\sum_{i=1}^{k} x_i^2 \ \geq\ c$$

가 됩니다. 이 경우에는 영가설 H_0 : $\mu_1 = 0$, $\cdots$, $\mu_k = 0$ 하에서 $\dfrac{1}{\sigma^2}\sum_{i=1}^{k} X_i^2$ 이 자유도 k 의 카이제곱 분포 $\chi^2(k)$를 따르므로 임계값 c는 $\chi^2_{k,\alpha}$로 정해집니다. 여기서 $\chi^2_{k,\alpha}$는 $\chi^2(k)$ 분포의 상위 α 분위수임.

앞의 예에서는 GLR 검증통계량인 $2\log_e \Lambda(X_1, \cdots, X_n)$이 영가설 H_0 하에서 정확히 $\chi^2(k)$ 분포를 따랐습니다. 이 사실은 다음과 같이 일반화됩니다.

<u>영가설 H_0 하에서의 GLR 검증통계량의 분포</u> : 정리

영가설 H_0 하에서 일반화 가능도 비 검증 통계량(generalized likelihood ratio test statistic)인 $2\log_e \Lambda(X_1, \cdots, X_n)$ 은, n이 큰 경우, 카이제곱 분포 $\chi^2(k)$를 *근사적으로* 따릅니다.

증명은 다음과 같습니다 (건너 뛰어도 무방함). 앞으로 로그 가능도를

$$l(\theta) = \log_e L(\theta\ ;\ x_1, \cdots, x_n)$$

으로 표기하겠습니다. 그러면

$$l(\theta_0) \ \simeq\ l(\hat\theta) + \left(\frac{\partial}{\partial\theta}\ l(\hat\theta)\right)^t (\theta_0 - \hat\theta) + \frac{1}{2}(\theta_0 - \hat\theta)^t \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ l(\hat\theta)\right)^t (\theta_0 - \hat\theta)$$

$$\frac{\partial}{\partial\theta}\ l(\hat\theta) = 0$$

이므로 (다변수 해석학에 익숙하지 않다면 $k = 1$인 경우로 환원시켜 봐도 됨),

$$\begin{aligned} 2\log_e \Lambda(x_1, \cdots, x_n) &= 2\left\{l(\hat\theta) - l(\theta_0)\right\} \\ &\simeq -(\theta_0 - \hat\theta)^t \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ l(\hat\theta)\right)^t (\theta_0 - \hat\theta) \end{aligned}$$

가 됩니다. 영가설 H_0 하에서 n이 커짐에 따라 $\hat\theta$이 θ_0로 수렴하므로

$$-\frac{1}{n}\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ l(\hat\theta)\right)^t \ \rightarrow\ I_1(\theta_0)$$

로 확률수렴해 들어갑니다. 그리고

$$n^{1/2}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \;\to\; \mathrm{N}_k\,(\mathbf{0}, I_1(\boldsymbol{\theta}_0)^{-1})$$

로 분포수렴합니다. 따라서

$$-\,(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})^t \frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\, l\,(\hat{\boldsymbol{\theta}})\right)^t (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) \;\to\; \chi^2(k)$$

로 분포수렴합니다. 마찬가지로, $2\log_e \Lambda(X_1, \cdots, X_n)$ 도 $\chi^2(k)$에 분포수렴합니다. ■

한 예를 들기로 하겠습니다. 한 유전학 이론에 따르면 통상적으로 형질 AB, Ab, aB, ab의 발현비율이 9/16, 3/16, 3/16, 1/16이라고 합시다. 그러나 연구자가 그의 연구문제에서 그렇지 않을 것이라는 심증을 갖게 되었다면, 그의 영가설은

$$\mathrm{H}_0 : P\{\mathrm{AB}\} = \frac{9}{16},\; P\{\mathrm{Ab}\} = \frac{3}{16},\; P\{\mathrm{aB}\} = \frac{3}{16},\; P\{\mathrm{ab}\} = \frac{1}{16}$$

이고 대안가설은 영가설을 단순부정한

$$\mathrm{H}_1 : \text{not } \mathrm{H}_0$$

가 됩니다. 이런 대안가설은 아무런 방향성을 제시하지 않기 때문에 연구가설로서의 가치가 있다고는 볼 수 없습니다. 그러나 무방향성인 것을 인정하고서도 검증 결과가 통계적으로 유의하다면 관측자료에 무엇인가 우리가 주목해야 할 패턴이 있다는 메시지가 됩니다.

관측자료가 다음과 같이 얻어졌다고 합시다.

타 입	AB	Ab	aB	ab	합계
관측빈도	76	41	37	6	160
기대빈도	90	30	30	10	160
차이	-14	+11	+7	-4	0

이에 대한 확률모형으로는 다항분포(multinomial distribution)가 좋겠습니다 :

$$(N_1, N_2, N_3, N_4) \sim \mathrm{Multinomial}(n, \theta_1, \theta_2, \theta_3, \theta_4)$$

여기서 $\theta_4 = 1 - \theta_1 - \theta_2 - \theta_3$ 이므로 $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ 이고 $k = 3$ 입니다. 이 경우, 확률함수는 다음과 같습니다.

$$f\,(N_1, N_2, N_3, N_4\,; \theta_1, \theta_2, \theta_3, \theta_4\,) \;=\; \frac{n\,!}{N_1!\,N_2!\,N_3!\,N_4!}\,\theta_1^{N_1}\,\theta_2^{N_2}\,\theta_3^{N_3}\,\theta_4^{N_4}\,,$$

$$단,\; n \;=\; N_1 + N_2 + N_3 + N_4\,, \quad \theta_1 + \theta_2 + \theta_3 + \theta_4 \;=\; 1\,.$$

그리고 영가설은

$$\mathrm{H}_0 \;:\; \theta_1 \,=\, \frac{9}{16},\; \theta_2 \,=\, \frac{3}{16},\; \theta_3 \,=\, \frac{3}{16},\; \theta_4 \,=\, \frac{1}{16}$$

입니다. $\boldsymbol{\theta} \,=\, (\theta_1, \theta_2, \theta_3, \theta_4\,)^t$ 에 대한 mle는

$$\hat{\boldsymbol{\theta}} \,=\, (N_1/n, N_2/n, N_3/n, N_4/n\,)^t$$

입니다. 따라서 GLR 검증통계량은

$$\begin{aligned} G \;&\equiv\; 2\log_e \Lambda\,(N_1,\, \cdots, N_4) \\ &= 2\,\{\,\mathrm{l}\,(\hat{\boldsymbol{\theta}}\,) - \mathrm{l}\,(\boldsymbol{\theta}_0)\} \\ &= 2\,\left[\,\sum_{j=1}^{4} N_j\,\log_e \hat{\theta}_j - \sum_{j=1}^{4} N_j\,\log_e \theta_{j0}\right] \\ &= 2\,\left[\,\sum_{j=1}^{4} N_j\,\log_e \frac{\hat{\theta}_j}{\theta_{j0}}\right] = 2\,\left[\,\sum_{j=1}^{4} N_j\,\log_e \frac{N_j}{n\,\theta_{j0}}\right] \\ &= 2\,\left[\,\sum_{j=1}^{4} N_j\,\log_e \frac{N_j}{E_j}\right] \qquad (여기서\;\; E_j = n\,\theta_{j0};\; 기대빈도) \end{aligned}$$

로 표현됩니다. 앞의 자료(n = 160)에서 GLR 검증통계량은

$$G \,=\, 2\,\left[\,76\log_e \frac{76}{90} + 41\log_e \frac{41}{30} + 37\log_e \frac{37}{30} + 6\log_e \frac{6}{10}\,\right] = 9.30$$

로 산출됩니다. 그리고 앞의 정리에 따르면 이 통계량은 근사적으로 카이제곱 분포 $\chi^2(3)$ 을 따릅니다. p-값은 2.6%이고요. 따라서 유의수준 $\alpha = 0.05$ 에서 영가설을 기각하게 됩니다. 그러므로, 이 자료에는 통상적인 유전학 이론과 합치하지 않는 면이 있음을 알 수 있습니다. 그래서 관측빈도와 기대빈도와의 차이를 살펴보면, A는 b와, a는 B와 기대보다 더 빈번히 결합하는 현상이 있음을 인지할 수 있습니다.

6.7* 그 밖의 대표본 검증

앞 절에서는 θ가 $k\times 1$ 파라미터 벡터일 때 $\mathrm{H}_0 : \theta = \theta_0$ 대 $\mathrm{H}_1 : \theta \neq \theta_0$에 대한 검증에서

$$2\log_e \Lambda(X_1, \cdots, X_n) = 2\left\{l(\hat{\theta}) - l(\theta_0)\right\}$$

에 근거한 일반화 가능도 비 검증(generalized likelihood ratio test)을 다루었습니다만, 이 절에서는 이와 점근적으로 (대표본에서는) 동등한 두 개의 검증 방법을 소개하고자 합니다. $X_1, \cdots, X_n$이 확률(밀도)함수 $f(x;\theta)$로부터의 임의표본임을 가정합니다.

한 방법은 왈드 검증(Wald test)인데, 최대가능도추정량 $\hat{\theta}$이 대표본에서

$$\sqrt{n}\,(\hat{\theta} - \theta) \;\sim\; \mathrm{N}(0, I_1^{-1}(\theta))$$

를 근사적으로 따르고 $\hat{\theta}$가 θ로 수렴하므로 영가설 $\mathrm{H}_0 : \theta = \theta_0$ 하에서

$$W \equiv n\,(\hat{\theta} - \theta_0)^t I_1(\hat{\theta})(\hat{\theta} - \theta_0) \;\sim\; \chi^2(k)\text{, 근사적으로}$$

가 되므로 W를 검증 통계량으로 잡습니다 (따라서 임계값은 $\chi^2_{k,\alpha}$임). 또는

$$W' \equiv (\hat{\theta} - \theta_0)^t \left[-\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta} l(\hat{\theta})\right)^t\right](\hat{\theta} - \theta_0) \;\sim\; \chi^2(k)\text{, 근사적으로}$$

를 검증통계량으로 합니다.

또 하나의 방법은 점수 검증(score test)입니다. 이것은

$$\frac{\partial}{\partial\theta} l(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log_e f(x_i;\theta),$$

$$E\left\{\frac{\partial}{\partial\theta}\log_e f(X_i;\theta)\right\} = 0,\; Var\left\{\frac{\partial}{\partial\theta}\log_e f(X_i;\theta)\right\} = I_1(\theta),\; i = 1, \cdots, n$$

이므로, 중심극한정리에 의하여

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta} l(\theta) \;\sim\; \mathrm{N}(0, I_1(\theta))$$

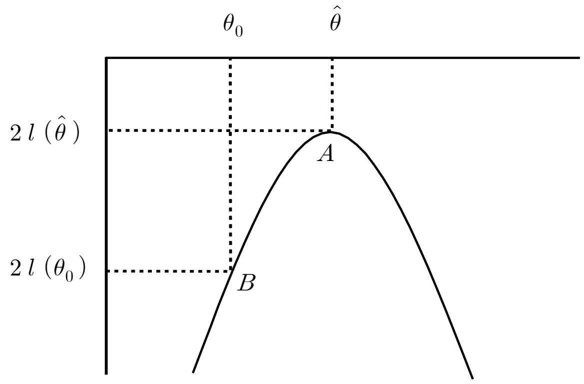
를 근사적으로 따릅니다. 따라서, 영가설 $\mathrm{H}_0 : \theta = \theta_0$ 하에서

$$S \equiv \frac{1}{n}\left(\frac{\partial}{\partial\theta} l(\theta_0)\right)^t I_1(\theta_0)^{-1}\left(\frac{\partial}{\partial\theta} l(\theta_0)\right) \;\sim\; \chi^2(k)\text{, 근사적으로}$$

가 되므로, S를 검증 통계량으로 합니다 (역시 임계값은 $\chi^2_{k,\alpha}$임). 또는

$$S' \equiv \left(\frac{\partial}{\partial\theta} l(\theta_0)\right)^t \left[-\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta} l(\hat{\theta})\right)^t\right]^{-1}\left(\frac{\partial}{\partial\theta} l(\theta_0)\right) \;\sim\; \chi^2(k)\text{, 근사적으로}$$

를 검증 통계량으로 합니다.



<그림 10> 대표본 검증을 위한 기하적 설명

이 두 방법과 일반화 가능도 비 검증 사이의 상호적 관계를 기하적으로 보기 위하여 이하 k를 1로 간주하겠습니다. <그림 10>에서 일반화 가능도 비 검증 통계량은 점 A와 점 B의 y 좌표값 차이, 즉 $2l(\hat{\theta}) - 2l(\theta_0)$로 나타내집니다.

이것을 점 A에서의 국소적 정보인 y 좌표값 $2l(\hat{\theta})$과 1계 미분값 0 및 2계 미분값 $2\frac{\partial^2}{\partial\theta^2}l(\hat{\theta})$만을 활용하여 근사해 봅시다. A에서의 2차 근사식이

$$y_A(\theta) = 2l(\hat{\theta}) + \frac{1}{2}\cdot 2\frac{\partial^2}{\partial\theta^2}l(\hat{\theta})\cdot(\theta-\hat{\theta})^2$$

이므로 $\theta = \theta_0$ 에서는 y값이

$$y_A(\theta_0) = 2l(\hat{\theta}) + \frac{\partial^2}{\partial\theta^2}l(\hat{\theta})\cdot(\theta_0-\hat{\theta})^2$$

입니다. 따라서 $2l(\hat{\theta}) - 2l(\theta_0)$는

$$W' \equiv y_A(\hat{\theta}) - y_A(\theta_0) = -\frac{\partial^2}{\partial\theta^2}l(\hat{\theta})\cdot(\theta_0-\hat{\theta})^2$$

으로 근사됩니다. $-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}l(\theta)$이 $I_1(\theta)$에 수렴하므로 W'에서 $-\frac{\partial^2}{\partial\theta^2}l(\hat{\theta})$을 $nI_1(\hat{\theta})$로 대치한 것이 바로 왈드 검증 통계량인

$$W = nI_1(\hat{\theta})\cdot(\hat{\theta}-\theta_0)^2$$

인 것입니다. W'도 W 못지 않게 훌륭한 또 하나의 왈드 검증 통계량입니다.

점 A와 점 B의 y 좌표값 차이 $2l(\hat{\theta}) - 2l(\theta_0)$에 대한 또 하나의 근사는 점 B에서의 국소적 정보를 활용하여 얻어집니다. 즉 $\theta = \theta_0$에서의 y 좌표값 $2l(\theta_0)$과 1

계 미분값 $2\frac{\partial}{\partial\theta}l(\theta_0)$ 및 2계 미분값 $2\frac{\partial^2}{\partial\theta^2}l(\theta_0)$를 써보자는 것입니다. B에서의 2차 근사식이

$$y_B(\theta) = 2l(\theta_0) + 2\frac{\partial}{\partial\theta}l(\theta_0)\cdot(\theta-\theta_0) + \frac{1}{2}\cdot 2\frac{\partial^2}{\partial\theta^2}l(\theta_0)\cdot(\theta-\theta_0)^2$$

이므로 이 함수의 정점은 $(\theta_{\max}, y_B(\theta_{\max}))$가 됩니다. 여기서

$$\theta_{\max} = \theta_0 - \frac{\frac{\partial}{\partial\theta}l(\theta_0)}{\frac{\partial^2}{\partial\theta^2}l(\theta_0)}, \qquad y_B(\theta_{\max}) = 2l(\theta_0) - \frac{\left(\frac{\partial}{\partial\theta}l(\theta_0)\right)^2}{\frac{\partial^2}{\partial\theta^2}l(\theta_0)}.$$

따라서 $2l(\hat{\theta}) - 2l(\theta_0)$는

$$S' \equiv y_B(\theta_{\max}) - y_B(\theta_0) = -\frac{\left(\frac{\partial}{\partial\theta}l(\theta_0)\right)^2}{\frac{\partial^2}{\partial\theta^2}l(\theta_0)}$$

로 근사됩니다. 그런데 $-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}l(\theta)$이 $I_1(\theta)$에 수렴하므로 $-\frac{\partial^2}{\partial\theta^2}l(\theta_0)$을 $nI_1(\theta_0)$로 대치한 것이 앞서의 점수 검증 통계량인

$$S = \frac{\left(\frac{\partial}{\partial\theta}l(\theta_0)\right)^2}{n\cdot I_1(\theta_0)}$$

입니다. S'도 S와 마찬가지로 또 하나의 점수 검증 통계량입니다.

점수 검증의 특징은 최대가능도 추정값 $\hat{\boldsymbol{\theta}}$을 명시적으로 구할 필요가 없다는 데 있습니다. 그러므로 $\hat{\boldsymbol{\theta}}$을 필요로 하는 다른 검증 방법에 비해서는 계산이 간단합니다. 계량경제학 분야에서는 라그랑지 승수 검증(Lagrange multiplier test)이라고 합니다.

예를 들어 (N_1, N_2, N_3, N_4)가 다항분포 Multinomial$(n, \theta_1, \theta_2, \theta_3, \theta_4)$를 따른다고 합시다 $(n = N_1+N_2+N_3+N_4, \theta_1+\theta_2+\theta_3+\theta_4 = 1, \boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^t)$. 즉 $k = 3$입니다.

$$l(\boldsymbol{\theta}) = constant + \sum_{j=1}^{4} N_j \log_e \theta_j, \quad \theta_4 = 1 - \theta_1 - \theta_2 - \theta_3$$

이므로, $\hat{\theta}_j = \frac{N_j}{n}$ 이고$(j = 1, 2, 3, 4)$

$$\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = \begin{pmatrix} \frac{N_1}{\theta_1} - \frac{N_4}{\theta_4} \\ \frac{N_2}{\theta_2} - \frac{N_4}{\theta_4} \\ \frac{N_3}{\theta_3} - \frac{N_4}{\theta_4} \end{pmatrix},$$

$$\frac{\partial^2}{\partial \boldsymbol{\theta}^2} l(\boldsymbol{\theta}) = \begin{pmatrix} -\frac{N_1}{\theta_1^2} - \frac{N_4}{\theta_4^2}, & -\frac{N_4}{\theta_4^2}, & -\frac{N_4}{\theta_4^2} \\ -\frac{N_4}{\theta_4^2}, & -\frac{N_2}{\theta_2^2} - \frac{N_4}{\theta_4^2}, & -\frac{N_4}{\theta_4^2} \\ -\frac{N_4}{\theta_4^2}, & -\frac{N_4}{\theta_4^2}, & -\frac{N_3}{\theta_3^2} - \frac{N_4}{\theta_4^2} \end{pmatrix}$$

$$= -\begin{pmatrix} \frac{N_1}{\theta_1^2}, & 0, & 0 \\ 0, & \frac{N_2}{\theta_2^2}, & 0 \\ 0, & 0, & \frac{N_3}{\theta_3^2} \end{pmatrix} - \frac{N_4}{\theta_4^2} \begin{pmatrix} 1,1,1 \\ 1,1,1 \\ 1,1,1 \end{pmatrix}$$

입니다. 따라서

$$I_1(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\theta_1}, & 0, & 0 \\ 0, & \frac{1}{\theta_2}, & 0 \\ 0, & 0, & \frac{1}{\theta_3} \end{pmatrix} + \frac{1}{\theta_4} \begin{pmatrix} 1,1,1 \\ 1,1,1 \\ 1,1,1 \end{pmatrix},$$

$$[I_1(\boldsymbol{\theta})]^{-1} = \begin{pmatrix} \theta_1, & 0, & 0 \\ 0, & \theta_2, & 0 \\ 0, & 0, & \theta_3 \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} (\theta_1, \theta_2, \theta_3)$$

가 됩니다. 그러므로 영가설 H_0 : $\theta_1 = \theta_{10}$, $\theta_2 = \theta_{20}$, $\theta_3 = \theta_{30}$, $\theta_4 = \theta_{40}$ 에 대한 왈드 검증 통계량과 점수 검증 통계량이 각각

$$W = \sum_{j=1}^{4} \frac{(N_j - n\,\theta_{j0})^2}{N_j} = \sum_{j=1}^{4} \frac{(N_j - E_j)^2}{N_j},$$

$$S = \sum_{j=1}^{4} \frac{(N_j - n\,\theta_{j0})^2}{n\,\theta_{j0}} = \sum_{j=1}^{4} \frac{(N_j - E_j)^2}{E_j} \quad : \text{Pearson's chi-square}$$

로 얻어집니다. 또한 $W = W'$, $S = S'$ 임을 보일 수 있습니다. 영가설 H_0 하에서 W, W', S, S' 은 모두 자유도 3의 카이제곱 분포 $\chi^2(3)$ 을 따릅니다.

앞 절의 자료와 영가설에 대하여 W와 S 통계값을 구하여 보면 각각 9.52와 9.44가 나옵니다. 그리고 이에 해당하는 대표본 p-값은 각각 2.3%와 2.4%입니다. (앞 절에서 보았듯이 GLR 검증 통계값은 9.30이고 대표본 p-값은 2.6%입니다.) 표본 크기 $n = 160$에서의 몬테칼로 근사 p-값을 구하기 위해 본 결과는 GLR 검증이 2.9%, W 검증이 4.4%, S 검증(피어슨 카이제곱 검증)이 2.4%를 기록하였습니다 (몬테칼로 반복수 = 1000). <표 1>의 SAS 프로그램을 보십시오.

<표 1> 대표본 검증 통계량의 몬테칼로 근사 p-값 계산을 위한 SAS/IML 프로그램

```
/* GLRT, Wald and Score Testing for Multinomial Distribution */
/* glrt.iml                                                   */

proc iml;
   Nrepeat = 1000;
   E  = {90,30,30,10};
   n  = sum(E);
   count0=0;  count1=0;  count2=0;   count3=0;
   do repeat=1 to Nrepeat;
       F  = j(4,1,0);
       do i=1 to n;
          temp = uniform(0);
          if temp <= 9/16 then F[1,1]=F[1,1]+1;
             else if temp <= 12/16 then F[2,1]=F[2,1]+1;
                else if temp <= 15/16 then F[3,1]=F[3,1]+1;
                   else F[4,1]=F[4,1]+1;
       end;
       if min(F) ^= 0 then do;
           count0=count0 +1;
           G = 2*sum(F#log(F/E));
           W = sum((F-E)#(F-E)/F);
           S = sum((F-E)#(F-E)/E);
           if G >= 9.30 then count1=count1+1;
           if W >= 9.52 then count2=count2+1;
           if S >= 9.44 then count3=count3+1;
       end;
   end;
   pvalue1 = count1/count0;
   pvalue2 = count2/count0;
   pvalue3 = count3/count0;
   pvalue1a = 1-probchi(9.30,3);
   pvalue2a = 1-probchi(9.52,3);
   pvalue3a = 1-probchi(9.44,3);
   print "Monte-Carlo"  count0 Nrepeat pvalue1 pvalue2 pvalue3;
   print "Large-Sample" pvalue1a[format=8.3] pvalue2a[format=8.3] pvalue3a[format=8.3];
quit;
```

6.8 복합 영가설에 대한 대표본 검증

$X_1, \cdots, X_n$이 확률(밀도)함수 $f(x;\boldsymbol{\theta},\boldsymbol{\lambda})$로부터의 임의표본이라고 합시다. 여기서 $\boldsymbol{\theta}$가 $k\times 1$ 파라미터 벡터이고 $\boldsymbol{\lambda}$를 $s\times 1$ 파라미터 벡터라고 하겠습니다. 이 때

$$\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0,\ \boldsymbol{\lambda}\ \text{非지정} \quad \text{대} \quad \mathrm{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \boldsymbol{\lambda}\ \text{非지정}$$

에 대한 가설 검증 문제를 생각하기로 합시다. 다시 말하여 이 문제에서 $\boldsymbol{\lambda}$는 관심 밖의 파라미터입니다. 이런 파라미터를 장애 모수(障碍母數, nuisance parameter)라고 하지요. 예컨대 정규분포 $\mathrm{N}(\theta,\sigma^2)$, σ^2 未知(unknown)로부터의 임의표본에서

$$\mathrm{H}_0 : \theta = \theta_0, \sigma^2\ \text{非지정} \quad \text{대} \quad \mathrm{H}_1 : \theta \neq \theta_0, \sigma^2\ \text{非지정}$$

을 검증하는 경우 σ^2은 장애 모수가 됩니다.

이런 경우에서는 영가설이 완전히 명시되어 있지 않습니다. 이런 가설을 복합가설(composite hypothesis)이라고 합니다. 이와 같이 영가설이 복합가설인 경우에도 일반화 가능도 비가 큰 역할을 합니다.

<u>일반화 가능도 비</u> (generalized likelihood ratio, GLR) : 정의

$$\Lambda(x_1, \cdots, x_n) = \frac{\max_{\Omega} L(\boldsymbol{\theta}, \boldsymbol{\lambda}\ ; x_1, \cdots, x_n)}{\max_{\omega} L(\boldsymbol{\theta}, \boldsymbol{\lambda}\ ; x_1, \cdots, x_n)},$$

여기서 $\Omega = \{ (\boldsymbol{\theta}, \boldsymbol{\lambda}) \}$는 전체 파라미터 공간이고 $\omega = \{ (\boldsymbol{\theta}, \boldsymbol{\lambda}) : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \}$는 영가설 H_0가 한정하는 Ω의 일부분입니다. 따라서 $\Lambda(x_1, \cdots, x_n)$은

$$\Lambda(x_1, \cdots, x_n) = \frac{L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}\ ; x_1, \cdots, x_n)}{L(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0\ ; x_1, \cdots, x_n)} = \frac{L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})}{L(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0)}$$

과 같습니다 (여기서 $\hat{\boldsymbol{\theta}}$ 과 $\hat{\boldsymbol{\lambda}}$ 은 $\boldsymbol{\theta}$ $(k\times 1)$과 $\boldsymbol{\lambda}$ $(s\times 1)$에 대한 최대가능도추정값(mle)이고 $\hat{\boldsymbol{\lambda}}_0$은 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$라는 제약 하에서 $\boldsymbol{\lambda}$ 에 대한 mle입니다).

일반화 가능도 비 검증(generalized likelihood ratio test, GLRT)의 기각역은 다음과 같습니다.

$$C : 2\log_e \Lambda(x_1, \cdots, x_n) \geq c, \quad c\text{는 상수}.$$

즉,

$$2\log_e \Lambda(X_1, \cdots, X_n) = 2\left\{ l(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - l(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0) \right\}$$

을 검증통계량으로 합니다. 여기서 $l(\boldsymbol{\theta}, \boldsymbol{\lambda})$은 로그 가능도 함수입니다. ■

예를 들어, X_i 가 정규분포 $\mathrm{N}(\mu_i, \sigma^2)$ 을 독립적으로 따르는 확률변수라고 합시다 $(i = 1, \cdots, n)$. 그런데, $\mu_{k+1} = \cdots = \mu_n = 0$ 으로 정해져 있으나 σ^2 이 未知라고 합시다. 그리고 영가설과 대안가설이 각각 다음과 같다고 합시다.

$$\mathrm{H}_0 : \mu_1 = 0, \cdots, \mu_k = 0 \quad \text{대} \quad \mathrm{H}_1 : \mu_1 \neq 0 \ \text{or} \ \cdots \ \text{or} \ \mu_k \neq 0 .$$

앞의 표기에 맞추면 $\boldsymbol{\theta} = (\mu_1, \cdots, \mu_k)^t$, $\boldsymbol{\theta}_0 = (0, \cdots, 0)^t$, $\lambda = \sigma^2$ (즉 s =1)입니다.
μ_i $(i = 1, \cdots, k)$ 에 대한 최대가능도추정값(mle)이 $\hat{\mu}_i = x_i$ $(i = 1, \cdots, k)$ 이고 σ^2 에 대한 mle가 $\hat{\sigma}^2 = \sum_{i=k+1}^{n} x_i^2 / n$, $\hat{\sigma}_0^2 = \sum_{i=1}^{n} x_i^2 / n$ 이므로

$$\Lambda(x_1, \cdots, x_n) = \frac{\left(\frac{1}{\sqrt{2\pi}\ \hat{\sigma}}\right)^n \exp\left\{-\frac{1}{2\ \hat{\sigma}^2}\sum_{i=k+1}^{n} x_i^2\right\}}{\left(\frac{1}{\sqrt{2\pi}\ \hat{\sigma}_0}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{n} x_i^2\right\}} = \frac{\left(\frac{1}{\sqrt{2\pi}\ \hat{\sigma}}\right)^n \exp\left\{-\frac{n}{2}\right\}}{\left(\frac{1}{\sqrt{2\pi}\ \hat{\sigma}_0}\right)^n \exp\left\{-\frac{n}{2}\right\}}$$

$$= \left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i^2}\right)^{n/2} = \left(1 + \frac{\sum_{i=1}^{k} x_i^2}{\sum_{i=k+1}^{n} x_i^2}\right)^{n/2}$$

이 됩니다. 따라서 GLR 검증의 기각역은

$$C: \quad 2\log_e \Lambda(x_1, \cdots, x_n) = n \log_e \left\{1 + \frac{\sum_{i=1}^{k} x_i^2}{\sum_{i=k+1}^{n} x_i^2}\right\} \ \geq \ c$$

즉,
$$C: \quad F_{k,n-k} \ \equiv \ \frac{\sum_{i=1}^{k} x_i^2 / k}{\sum_{i=k+1}^{n} x_i^2 / (n-k)} \ \geq \ c'$$

가 되며, 영가설 $\mathrm{H}_0 : \mu_1 = 0, \cdots, \mu_k = 0$ 하에서 임계값 c' 이 $F_{k,n-k,\alpha}$ 로 정해집니다 (여기서 $F_{k,n-k,\alpha}$ 는 $F(k, n-k)$ 분포의 상위 α 분위수임). 또는

$$C: \quad \chi_k^2 \ \equiv \ \frac{\sum_{i=1}^{k} x_i^2}{\sum_{i=k+1}^{n} x_i^2 / (n-k)} \ \geq \ c''$$

로 표현할 수 있겠는데 n 이 무한히 커짐에 따라 $\sum_{i=k+1}^{n} x_i^2 / (n-k) \to \sigma^2$ 로 확률수렴하므로 앞의 χ_k^2 은 영가설 H_0 하에서 근사적으로 $\chi^2(k)$ 분포를 따를 것입니다. 그러므로 임계값 c'' 이 $\chi_{k,\alpha}^2$ 로 근사됩니다. $2\log_e \Lambda(x_1, \cdots, x_n)$ 의 임계값 c 도 이와 같습니다. 즉 $c \simeq \chi_{k,\alpha}^2$ (연습문제 6.11).

복합 영가설 H_0 하에서의 GLR 검증통계량의 분포 : 정리

복합적인 영가설 H_0 하에서 일반화 가능도 비 검증 통계량(generalized likelihood ratio test statistic)인 $2\log_e \Lambda(X_1, \cdots, X_n)$ 은, n 이 큰 경우, 카이제곱 분포 $\chi^2(k)$ 를 *근사적으로* 따릅니다. 증명은 생략하겠습니다. ■

한 예를 보도록 하겠습니다. 다음 2원 분할표에서 행은 4개 수준(교육정도)이고 열은 3항 반응(후보자 선택)입니다.

	반응 1	반응 2	반응 3	합 계
수준 1	10	5	5	20
수준 2	25	15	10	50
수준 3	30	40	10	80
수준 4	10	30	10	50
합 계	75	90	35	200

반응이 수준별로 서로 다른 데가 있다고 할 수 있을까를 검증해 봅시다. 이를 위해서 자료에 대하여 모형으로 다항분포를 생각하기로 하겠습니다. 구체적으로

$$\begin{pmatrix} N_{11}, N_{12}, N_{13} \\ N_{21}, N_{22}, N_{23} \\ N_{31}, N_{32}, N_{33} \\ N_{41}, N_{42}, N_{43} \end{pmatrix} \sim \text{Multinomial}\left(n, \begin{pmatrix} p_{11}, p_{12}, p_{13} \\ p_{21}, p_{22}, p_{23} \\ p_{31}, p_{32}, p_{33} \\ p_{41}, p_{42}, p_{43} \end{pmatrix} \right)$$

을 가정하기로 하지요 (여기서 $n = \sum_{i=1}^{4} \sum_{j=1}^{3} N_{ij}$, $\sum_{i=1}^{4} \sum_{j=1}^{3} p_{ij} = 1$). 그러면 영가설은

$$H_0 : p_{ij} = p_{i+} \cdot p_{+j}, \quad (\text{단, } p_{i+} = \sum_{j=1}^{3} p_{ij},\ p_{+j} = \sum_{i=1}^{4} p_{ij})$$

로 나타내집니다. 그러므로 앞의 표기에 맞춘다면 장애모수 λ 는 실질적으로

$$(p_{1+}, p_{2+}, p_{3+}, p_{+1}, p_{+2})$$

의 5개이고 ($p_{4+} = 1 - (p_{1+} + p_{2+} + p_{3+})$, $p_{+3} = 1 - (p_{+1} + p_{+2})$ 은 위의 장애모수들로 표현 가능하기 때문에 λ 에 포함시킬 필요가 없음), 관심모수 θ 는 사실상

$$\begin{pmatrix} p_{11}, p_{12} \\ p_{21}, p_{22} \\ p_{31}, p_{32} \end{pmatrix}$$

의 6개입니다 (나머지 p_{ij}들은 θ 와 λ 로 표현가능하기 때문에). 그리고 가능도는

$$l(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(p_{ij}, p_{i+}, p_{+j}) = constant + \sum_{i=1}^{4} \sum_{j=1}^{3} N_{ij} \log_e p_{ij}$$

입니다. 따라서 $\boldsymbol{\theta}$ 와 $\boldsymbol{\lambda}$ 에 대한 mle는

$$\hat{p}_{ij} = N_{ij}/n,\ \hat{p}_{i+} = N_{i+}/n,\ \hat{p}_{+j} = N_{+j}/n$$

이며, 영가설 H_0 하에서 $\boldsymbol{\lambda}$ 에 대한 mle를 구해보면

$$\hat{p}_{i+,0} = N_{i+}/n,\ \hat{p}_{+j,0} = N_{+j}/n$$

입니다. 따라서 일반화 가능도 비 검증 통계량은

$$\begin{aligned} G &\equiv 2\left\{\sum_{i=1}^{4}\sum_{j=1}^{3} N_{ij}\log_e \hat{p}_{ij} - \sum_{i=1}^{4}\sum_{j=1}^{3} N_{ij}\log_e \hat{p}_{i+,0}\hat{p}_{+j,0}\right\} \\ &= 2\left\{\sum_{i=1}^{4}\sum_{j=1}^{3} N_{ij}\log_e N_{ij}/n - \sum_{i=1}^{4}\sum_{j=1}^{3} N_{ij}\log_e N_{i+}N_{+j}/n^2\right\} \\ &= 2\left\{\sum_{i=1}^{4}\sum_{j=1}^{3} N_{ij}\log_e \frac{N_{ij}}{E_{ij}}\right\} \end{aligned}$$

가 됩니다. 여기서 $E_{ij} = N_{i+}N_{+j}/n$. G는 앞의 정리에 따라 영가설 H_0 하에서 자유도 6 (=dim($\boldsymbol{\theta}$))의 카이제곱분포 $\chi^2(6)$을 따릅니다.

<u>복합 영가설 H_0 하에서의 왈드 검증(Wald test)과 점수 검증(score test)</u> : 정리

영가설 $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0,\ \boldsymbol{\lambda}$ 非지정 대 대안가설 $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,\ \boldsymbol{\lambda}$ 非지정에 대한 왈드 검증 통계량 W와 점수 검증 통계량 S는 다음과 같습니다.

$$W = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\ I_{1,11}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

$$S = \frac{1}{n}\left(\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0)\right)^t\ \Sigma_{1,11}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0)\ \left(\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}_0, \hat{\boldsymbol{\lambda}}_0)\right).$$

여기서

$$I_1(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{array}{cc} \begin{pmatrix} I_{1,11}(\boldsymbol{\theta}, \boldsymbol{\lambda}) & I_{1,12}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \\ I_{1,21}(\boldsymbol{\theta}, \boldsymbol{\lambda}) & I_{1,22}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \end{pmatrix} & \begin{matrix} k \\ s \end{matrix} \\ \begin{matrix} \quad k \qquad\qquad\qquad s \end{matrix} & \end{array}$$

는 피셔 단위정보행렬이고 이에 대한 역행렬이

$$\Sigma_1(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{array}{cc} \begin{pmatrix} \Sigma_{1,11}(\boldsymbol{\theta}, \boldsymbol{\lambda}) & \Sigma_{1,12}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \\ \Sigma_{1,21}(\boldsymbol{\theta}, \boldsymbol{\lambda}) & \Sigma_{1,22}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \end{pmatrix} & \begin{matrix} k \\ s \end{matrix} \\ \begin{matrix} \quad k \qquad\qquad\qquad s \end{matrix} & \end{array}$$

입니다.

영가설 H_0 하에서 왈드 검증 통계량 W와 점수 검증 통계량 S는 GLR 검증 통계량 G와 마찬가지로, n이 큰 경우, 카이제곱 분포 $\chi^2(k)$를 *근사적으로* 따릅니다. 증명은 생략합니다. ■

2원 분할표에서 행과 열의 독립성에 관한 가설 $H_0 : p_{ij} = p_{i+} \cdot p_{+j}$ 에 대한 왈드 검증 통계량과 점수 검증 통계량은 각각

$$W = \sum_{i=1}^{4} \sum_{j=1}^{3} \frac{\{N_{ij} - E_{ij}\}^2}{N_{ij}},$$

$$S = \sum_{i=1}^{4} \sum_{j=1}^{3} \frac{\{N_{ij} - E_{ij}\}^2}{E_{ij}}$$

가 됩니다 (증명 생략). 4×3 분할표에서 대하여 이들은 큰 n에 대하여 영가설 하에서 카이제곱 분포 $\chi^2(6)$을 근사적으로 따릅니다. 앞의 자료에 대하여 $G = 17.39$, $W = 21.43$, $S = 16.37$ 이고 해당하는 p-값은 각각 0.8%, 0.2%, 1.2%입니다.

6.A 연습문제

6.1 여덟 잔의 티를 가지고 우유가 먼저였는지 아니면 찻물이 먼저였는지를 알아보는 실험에서 네 잔에는 우유를 먼저 넣고 나머지 네 잔에는 찻물을 먼저 넣는다고 합시다. 그리고 피험자에게 그렇게 실험이 설계된다는 사실을 알리고 어느 네 잔이 우유가 먼저인 것인지를 말하도록 한다고 합시다. 결과적으로 피험자가 6잔을 맞춘 경우에 그의 능력에 관한 유의성 검증을 하세요.

답 : θ를 그의 능력이라고 할 때, $H_0 : \theta = 1/2$ 대 $H_1 : \theta > 1/2$에 대한 p-값은 24.3%로 나옵니다 (초기하분포).

6.2 $X_1, \cdots, X_n$을 오염정규분포

$$0.9\,\mathrm{N}(\theta, \sigma^2) + 0.1\,\mathrm{N}(\theta, k^2\sigma^2), \quad k \geq 1$$

로부터의 임의표본이고 $\sigma^2 = 1$, $k = 3$, $n = 10$이라고 합시다. 이 때,

$$H_0 : \theta = 10 \quad \text{vs.} \quad H_1 : \theta > 10$$

에 대한 검증통계량으로 표본중위수 $\tilde{X}$를 활용하고자 합니다. 중위수가 10.5로 나온 표본에서 몬테칼로 방법을 활용하여 p-값을 계산해 보세요.

답 : 대략 10.2% (6.1절 본문참조.).

6.3 확률변수 X의 분포에 대하여 다음 두 가설이 있다고 합시다.

$$H_0 : X \sim \text{Geometric}(1/51) \quad 대 \quad H_1 : X \sim \text{Poisson}(50).$$

따라서 두 분포의 평균이 같습니다. 가능도 비가 2 이상인 경우에 H_0를 기각한다고 합시다. H_0 하에서의 H_0의 기각 확률을 구해보세요.

답 : $P\{40 \leq X \leq 62 \,;\, H_0\} = 0.166$

6.4 X_1과 X_2를 Cauchy(θ) 분포로부터의 iid 확률변수라고 합시다.

$$H_0 : \theta = 0 \quad 대 H_1 : \theta = 1$$

에 대하여 검증으로서 가능도비가 2를 넘어갈 때 영가설을 기각한다고 합시다 (6.2절의 <그림 2> 참조). 검증의 수준이 얼마입니까? 그리고 대안가설 하에서 검증력은 얼마입니까? 몬테칼로 방법을 활용해보세요.

6.5 X_1과 X_2를 확률밀도함수 $f(x\,;\theta) = \theta\, x^{\theta-1}$, $0 < x \leqq 1$로부터의 iid 확률변수라고 합시다. $H_0 : \theta = 1$ 대 $H_1 : \theta > 1$에 대하여 검증으로서 기각역 C를

$$C = \{(x_1, x_2) : x_1 \geq 0.75 \text{ and } x_2 \geq 0.75\}$$

로 하였을 때, 검증력 함수 $\beta(\theta), \theta \geq 1$ 을 구하세요.

6.6 앞 문제에서 균일최강력(UMP) 검증이 존재하는지를 살펴보고 만약 존재한다면 앞 문제와 동일한 검증수준의 UMP 검증을 만드세요.

6.7 $X_1, \cdots, X_{10}$ 을 포아송 분포 Poisson(θ) 로부터의 iid 확률변수라고 할 때,

$$\mathrm{H}_0 : \theta = 1 \text{ 대 } \mathrm{H}_1 : \theta > 1$$

에 대한 수준 5%의 검증을 만드세요. 그 검증이 UMP 검증입니까?

6.8 $X_1, \cdots, X_{10}$ 을 균일분포 Uniform$(0, \theta)$ 로부터의 iid 확률변수라고 할 때,

$$\mathrm{H}_0 : \theta = 1 \text{ 대 } \mathrm{H}_1 : \theta > 1$$

에 대한 수준 5%의 검증을 만드세요. 그 검증이 UMP 검증입니까?

6.9 $X_1, \cdots, X_m$ 이 정규분포 $\mathrm{N}(\theta, \theta^2)$ 으로부터의 iid 확률변수라고 할 때,

$$\mathrm{H}_0 : \theta = 1 \text{ 대 } \mathrm{H}_1 : \theta > 1$$

에 대한 국소최강력 검증통계량을 제시하세요.

6.10 $X_1, \cdots, X_m$ 이 지수분포 Exponential(θ_1) 으로부터의 iid 확률변수이고, 이와 독립적으로 $Y_1, \cdots, Y_n$ 이 Exponential(θ_2) 로부터의 iid 확률변수라고 합시다. 이 때

$$H_0 : \theta_1 = \theta_2 \quad \text{대} \quad H_1 : \theta_1 \neq \theta_2$$

에 대한 일반화 가능도 비 검증을 만드세요.

힌트 : $T = \sum_{i=1}^{m} X_i \Big/ \left\{ \sum_{i=1}^{m} X_i + \sum_{j=1}^{n} Y_j \right\}$ 를 검증통계량으로 하세요.

6.11 (6.8절 참조). X_i 가 정규분포 $\mathrm{N}(\mu_i, \sigma^2)$ 을 독립적으로 따르는 확률변수라고 하고 $(i = 1, \cdots, n)$, $\mu_{k+1} = \cdots = \mu_n = 0$, σ^2 이 未知라고 할 때,

$$\mathrm{H}_0 : \mu_1 = 0, \cdots, \mu_k = 0 \quad \text{대} \quad \mathrm{H}_1 : \mu_1 \neq 0 \text{ or } \cdots \text{ or } \mu_k \neq 0$$

에 대하여 일반화 가능도 비 검증통계량 $2\log_e \Lambda(x_1, \cdots, x_n)$ 의 임계값이 근사적으로 $\chi^2_{k,\alpha}$ 임을 *직접* 보이세요 (정리를 사용하지 말고).

6.12 유전학 모형에 의하면 형질 A1, A2, A3의 비율이 θ^2, $2\theta(1-\theta)$, $(1-\theta)^2$ 이라고 합니다. 실제로 $n = 40$ 의 표본에서 형질 A1, A2, A3이 8, 24, 8로 관측되었을 때, 이 모형에 대한 유의성 검증을 해보세요.

답 : 피어슨 $\chi^2 = 1.6$ (자유도 1), p-값 = 0.206.

6.13 $X_1, \cdots, X_n$ 을 정규분포 $N(\theta, \sigma^2)$ (σ^2 未知)으로부터의 iid 확률변수라고 할 때

$$H_0 : \theta = 0 \quad \text{대} \quad H_1 : \theta \neq 0$$

에 대한 일반화 가능도 비 검증을 만드세요.

6.14 $X_1, \cdots, X_m$ 을 정규분포 $N(\theta_1, \sigma^2)$ 으로부터의 iid 확률변수라고 하고, 이와 독립적으로 $Y_1, \cdots, Y_n$ 을 $N(\theta_2, \sigma^2)$ 로부터의 iid 확률변수라고 할 때 (σ^2 未知)

$$H_0 : \theta_1 = \theta_2 \quad \text{대} \quad H_1 : \theta_1 \neq \theta_2$$

에 대한 일반화 가능도 비 검증을 만드세요.

6.15 $X_1, \cdots, X_m$ 이 정규분포 $N(0, \sigma_1^2)$ 으로부터의 iid 확률변수라고 하고, 이와 독립적으로 $Y_1, \cdots, Y_n$ 이 $N(0, \sigma_2^2)$ 로부터의 iid 확률변수라고 합시다. 이 때

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{대} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

에 대한 일반화 가능도 비 검증을 만드세요.

6.16 다음은 어떤 방사성 물질로부터 방출되는 단위시간당 알파입자 수에 대한 관측자료입니다 (총 1207 사례, 평균 8.392). 연구자는 이것을 포아송 분포로 모형화하고자 합니다. 연구자의 의도를 반박할 만한 통계적 증거를 찾을 수 있습니까?

입자 수	사례 수	입자 수	사례 수
0-2	18	10	123
3	28	11	101
4	56	12	74
5	105	13	53
6	126	14	23
7	146	15	15
8	164	16	9
9	161	17+	5

탐구문제 1: 문제 6.4와 같은 맥락에서

$$H_0 : \theta = 0 \text{ 대 } H_1 : \theta > 0$$

에 대한 국소최강력 검증을 같은 수준에서 만들어 보세요. 그리고 $\theta = 1$ 에서의 검증력을 비교해보세요.

탐구문제 2: 문제 6.16에서 대표본 이론에 의존하지 않고 다른 방법으로 p-값을 구할 수 있겠습니까? 다음과 같은 모수적 붓스트랩(parametric Bootstrap)은 어떨까요?

1) 1207개의 관측을 평균이 8.392인 포아송 분포로부터 임의 발생시킵니다.
2) 1)의 모의생성자료로부터 카이제곱 통계량을 구합니다.
3) 1)과 2)의 과정을 많은 회수 반복하여 (1,000번쯤) 카이제곱 통계량의 분포를 만듭니다. 이 분포의 평균을 구해보세요. 대표본 이론에 따르면 이것은 14여야 합니다. 비슷합니까?
4) 3)의 분포에서 원관측자료로부터의 카이제곱값의 유의성을 평가합니다.

6.B 읽을만한 책

가설검증론에 대하여도 일반적인 수리통계학 책들이 잘 다루고 있습니다. 다음 중에서 아무 책이나 보기 바랍니다.

- Hogg, R.V. and Craig, A.T. (1995) *Introduction to Mathematical Statistics*, 5th Edition. Prentice Hall. (Chapters 6 and 9)
- Bickel, P.J. and Duksum, K.A. (1977) *Mathematical Statistics.*. Holden-Day. (Chapters 5 and 6)
- Rice, J.A. (1995) *Mathematical Statistics and Data Analysis*, 2nd Edition. Duxbury Press. (Chapter 9)
- Casella, G. and Berger, R.L. (1990) *Statistical Inference*. Duxbury. (Chapter 8)
- Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall. (Chapters 3, 4, and 5)

반증주의와 관련하여 과학철학에 관심이 있다면 다음 책에서 시작하길 권합니다.

▪ 앨런 찰머스 (신일철 · 신중섭 옮김) (1985)「현대의 과학철학」 서광사.