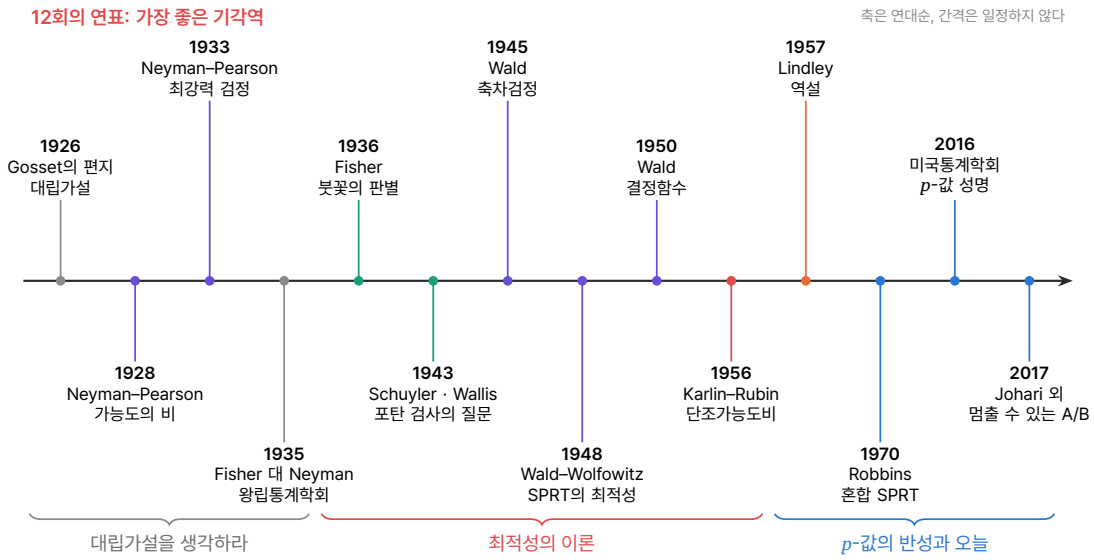
수리통계학 × 통계학의 역사

12회 강의노트

최적 검정과 결정
가장 좋은 기각역을 찾아서

1928 → 1956

Neyman과 Pearson은 1933년 검정을 두 가지 오류의 균형으로 다시 썼고, 크기를 묶었을 때 검정력이 가장 큰 기각역이 가능도비로 정해진다는 것을 증명했다. Fisher는 이 관점을 평생 거부했다. 이 노트는 Neyman–Pearson 보조정리에서 균일최강력 검정과 가능도비 검정으로, Wald의 축차검정과 결정이론으로 나아가, Fisher의 붓꽃에서 판별분석을 유도한다.



다루는 것	Neyman–Pearson 보조정리 · 확률화 검정 · 단조가능도비와 Karlin–Rubin · 가능도비 검정 · 축차확률비 검정(SPRT)과 Wald 경계 · 손실·위험·Bayes·minimax · Bayes 분류와 선형판별분석 · ROC · 효과크기와 Lindley의 역설
원자료	Anderson–Fisher의 붓꽃 150송이(1935–36) · 군수품 검사(Wald의 SPRT) · A/B 테스트(시뮬레이션)
노트북	`sessions/12_optimal_tests/12_optimal_tests.ipynb` — 절·정리 번호가 같다
워크북	`12_optimal_tests_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1–11회 (특히 8회 검정·크기·검정력·p-값, 9회 가능도비·Wilks, 10회 프로파일 가능도, 11회 충분통계량·지수족)

이 노트를 쓰는 법

이 노트는 다른 책 없이 읽도록 썼다. 쓰이는 정의와 정리는 모두 이 노트나 앞 회차의 노트에서 증명한다. 이 과정의 범위를 넘어 증명하지 않는 사실은 **증명 없이 쓰는 사실** 상자에 따로 밝힌다.

각 절은 **역사 → 정의·정리 → 증명 → 그림 → 연습문제** 순서다. 핵심 정리의 증명은 번호 붙은 단계 ①②③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 정의)를 단다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 합과 기댓값, 적분의 순서를 바꾸는 곳은 **조건 확인** 단락에 따로 밝힌다. 나머지 정리는 **증명 전략**만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 이 연습은 건너뛰지 않고 손으로 쓴다.

근거에 붙은 [T1]–[T8]은 별책 **증명 도구 상자**의 도구 번호다. 이 회차에서는 [T1] 지시함수, [T2] 대우, [T4] Cauchy–Schwarz, [T6] 합과 적분의 순서 바꾸기를 쓴다.

연습문제는 소절 끝에 모여 있다. 난이도는 ★ 계산·확인, ★★ 증명·응용, ★★★ 도전이다. 모든 문제의 풀이가 부록 A에 있고, 문제 끝의 **풀이 ○○쪽**을 누르면 그 쪽으로 간다. 손으로 푸는 문제는 따로 **워크북**(계산 드릴, 빈칸 증명, 증명 문제, 퀄형)에 모았다.

상자의 뜻

> **정의** 대상을 처음 정하는 문장. 왼쪽 선이 보라다.

> **정리** 증명할 주장(정리·보조정리·따름정리·명제). 왼쪽 선이 빨강이다.

> **예제** 숫자를 넣어 끝까지 계산해 보는 곳. 왼쪽 선이 파랑이다.

> **가정 주의** 빼면 정리가 무너지는 가정. 왼쪽 선이 주황이다.

> **증명 없이 쓰는 사실** (예) 이 과정에서 증명하지 않고 쓰는 결과. 무엇이 참인지, 왜 그럴듯한지, 어디서 증명되는지를 밝힌다.

> **역사** 그 개념이 태어난 사람과 문제. 회색 선 옆의 작은 글씨다.

> **연습** 소절 끝의 연습문제. 가는 청록 선이고, 끝에 풀이가 있는 쪽 번호가 붙는다.

색의 뜻 수식과 그림에서 보라는 모수, 청록은 데이터·확률변수, 파랑은 통계량, 빨강은 핵심 결론이다. 노트북과 같은 색이다.

차례

0 들어가며 1

0.1 1–11회에서 가져오는 것 2

0.2 표기 2

1 Neyman–Pearson 보조정리 3

1.1 최강력 검정 3

2 균일최강력 검정과 단조가능도비 7

2.1 정의 7

2.2 Karlin–Rubin 정리 8

2.3 양쪽 가설에는 UMP가 없다 9

3 가능도비 검정 11

4 축차확률비 검정 — Wald (1945) 13

4.1 정의와 Wald의 경계 13

4.2 평균 표본수 15

5 결정이론 — 검정과 분류를 하나로 17

5.1 손실, 위험, Bayes 규칙 17

5.2 Bayes 분류와 선형판별분석 18

6 데이터로 — Fisher의 붓꽃 (1936) 22

7 부쉬보기 — p-값이 말하지 않는 것 24

7.1 표본이 거대하면 24

7.2 Lindley의 역설 (1957) 24

8 Beautiful insights — ROC와 멈출 수 있는 검정 26

8.1 ROC 곡선은 Neyman–Pearson 검정들의 모임이다 . . . 26

8.2 보면서 멈출 수 있는 검정 . . . 26

9 정리와 역사 노트 29

9.1 한 장 요약 . . . 29

9.2 Stigler의 명명 법칙 . . . 29

9.3 7 Pillars 지도 . . . 30

9.4 다음 회 — 최소제곱: Legendre–Gauss에서 Gauss–Markov까지 (1749 → 1943) . . . 30

A 연습문제 풀이 32

0

들어가며

> **여는 질문.** p-값이 작다는 것은 무엇을 말하고, 무엇을 말하지 않는가? 같은 자료로 검정을 여러 가지 만들 수 있다면, 그 가운데 "가장 좋은" 검정은 무엇이며 어떤 뜻에서 가장 좋은가?

8회에서 p-값과 검정의 크기, 검정력을 정의했다. 그때는 검정 통계량을 그때그때 골랐다. 이항 검정은 성공 수로, z 검정은 표본평균으로 했다. 왜 그 통계량이어야 하는가? Jerzy Neyman과 Egon Pearson은 1928년부터 이 질문을 붙잡았다. 그들은 검정을 "두 가지 오류의 균형"으로 다시 썼다. 귀무가설이 참인데 기각하는 오류의 확률(크기)을 미리 묶고, 그 안에서 대립가설이 참일 때 기각할 확률(검정력)을 최대로 하자는 것이다. 1933년 그들은 그런 검정이 가능도비로 기각역을 정하는 검정임을 증명했다. 이것이 Neyman–Pearson 보조정리다.

Fisher는 이 관점을 끝까지 거부했다. 그에게 유의성 검정은 한 번의 실험에서 증거의 크기를 재는 **귀납 추론**이었다. 반복되는 결정의 오류율을 관리하는 **귀납 행동**이 아니었다. 두 사람의 다툼은 오늘의 p-값 논쟁에도 그대로 남아 있다. 이 노트는 Neyman–Pearson의 길을 따라간다. 보조정리에서 출발해 한쪽 가설의 균일최강력 검정(Karlin–Rubin, 2절)과 일반화 가능도비 검정(3절)에 이른다. 2차대전의 군수품 검사에서 태어난 Wald의 축차검정(4절)을 지나 결정이론으로 검정과 분류를 하나로 묶는다(5절). 6절에서는 Fisher(1936)의 붓꽃 자료로 판별분석을 유도한다. 7절은 p-값이 말하지 않는 것(효과크기, Lindley의 역설)을 보고, 8절은 같은 수학이 ROC 곡선과 온라인 A/B 테스트에 있음을 본다.

학습 목표

1. Neyman–Pearson 보조정리를 증명하고, 확률화 검정이 왜 필요한지 이산분포에서 계산한다.
2. 단조가능도비를 갖는 족에서 Karlin–Rubin 정리로 한쪽 균일최강력 검정을 구하고, 양쪽 가설에는 그런 검정이 없음을 본다.
3. 축차확률비 검정을 구현하고 Wald 경계와 평균 표본수를 증명·확인한다.
4. 손실·위험·Bayes 규칙으로 검정과 분류를 통일하고, 정규 모형에서 선형판별분석을 유도한다.
5. 붓꽃 자료에 판별분석을 쓰고 ROC 곡선을 그리며, 효과크기와 Lindley의 역설로 p-값의 한계를 안다.

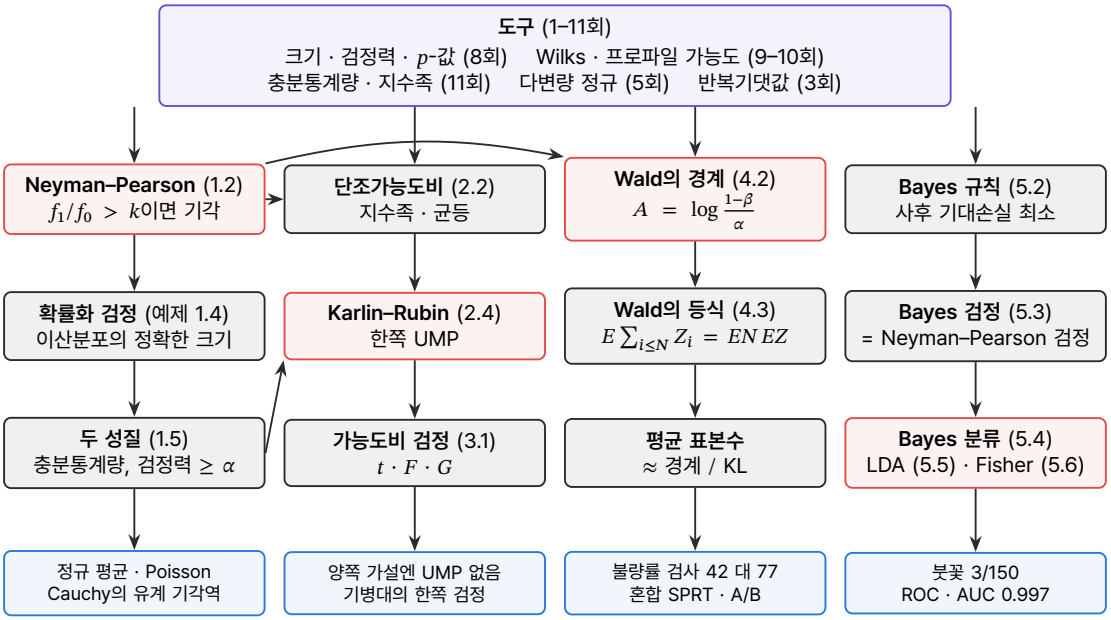


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–11회에서 가져오는 것

앞 회차	내용
5회 정리 A · C	다변량 정규의 선형변환과 조건부분포
8회 정의 3.1 · 3.2, 정리 3.5 · 4.2	검정, 크기와 검정력, 구간과 검정의 쌍대성, p-값의 균등성
8회 예제 7.1	결과를 보며 멈추면 1종 오류가 커진다
9회 정리 5.2 · 10회 정리 1.8	Wilks 정리, 귀찮은 모수가 있는 프로파일 가능도비
11회 정리 1.3 · 4.5	인수분해 정리, 지수족의 완비충분통계량

0.2 표기

자료 $\mathbf{X}$의 밀도는 $f(\mathbf{x};\theta)$이다. 검정은 **검정함수** $\varphi(\mathbf{x}) \in [0,1]$로 쓴다. $\varphi(\mathbf{x})$는 자료가 $\mathbf{x}$일 때 H_0을 기각할 확률이다. 보통의 검정은 0 또는 1의 값만 갖고, 경계에서만 0과 1 사이의 값을 갖는 것을 **확률화 검정**이라 한다. 크기는 $\sup_{\theta\in\Omega_0} E_\theta\varphi(\mathbf{X})$, 검정력 함수는 $\gamma(\theta) = E_\theta\varphi(\mathbf{X})$이다(8회 정의 3.2).

1

Neyman–Pearson 보조정리

> **역사** 1926년 Gosset(Student)은 Egon Pearson에게 보낸 편지에서, 관측이 귀납가설 아래에서 드물다는 것만으로는 기각할 이유가 되지 않는다고 썼다. 다른 설명 아래에서 **더 그럴듯해야** 한다는 것이다. Pearson은 이 생각을 런던에 와 있던 폴란드 수학자 Jerzy Neyman과 함께 발전시켰다. 1928년 그들은 가능도의 비로 검정을 만들자고 제안했고, 1933년 **가장 효율적인 검정의 문제에 대하여**에서 그것이 단순가설 사이의 최강력 검정임을 증명했다.

1.1 최강력 검정

정의 1.1 (최강력 검정) $H_0 : \theta = \theta_0$ 대 $H_1 : \theta = \theta_1$ (둘 다 **단순가설**)에서, 크기가 α 이하인 모든 검정 φ에 대해 $E_{\theta_1}\varphi^* \geq E_{\theta_1}\varphi$이고 $E_{\theta_0}\varphi^* \leq \alpha$인 검정 φ^*를 크기 α의 **최강력**(MP) 검정이라 한다.

정리 1.2 (Neyman–Pearson 보조정리 (1933)) $f_0 = f(\cdot;\theta_0)$, $f_1 = f(\cdot;\theta_1)$라 하자.

(a) (존재) 모든 $\alpha \in (0,1)$에 대해 상수 $k \geq 0$과 $\gamma \in [0,1]$이 있어

$$\varphi^*(\mathbf{x}) = \begin{cases} 1, & f_1(\mathbf{x}) > k f_0(\mathbf{x}), \\ \gamma, & f_1(\mathbf{x}) = k f_0(\mathbf{x}), \\ 0, & f_1(\mathbf{x}) < k f_0(\mathbf{x}) \end{cases}$$

의 크기가 정확히 α이다.

(b) (충분) 이 꼴의 검정은 크기 $E_{\theta_0}\varphi^*$의 최강력 검정이다.

(c) (필요) 크기 α의 최강력 검정은 $\{f_1 \neq k f_0\}$ 위에서 (거의 어디서나) φ^*와 같다.

증명. **전략** (a)는 가능도비의 생존함수에서 임계값을 고르고 경계에서 확률화한다. (b)는 모든 점에서 성립하는 한 줄의 부등식 $(\varphi^* - \varphi)(f_1 - k f_0) \geq 0$을 적분한다. (c)는 (b)의 부등식이 등호가 되는 조건을 읽는다.

(a)

① $F(k) = P_{\theta_0}\big(f_1(\mathbf{X}) > k f_0(\mathbf{X})\big)$ $(k \ge 0)$는 감소하고 오른쪽 연속이며, $k \to \infty$에서 0으로 간다. — 가능도비의 생존함수, 1회 정리 4.7

② $k = \inf\{t \ge 0 : F(t) \le \alpha\}$로 두면 $F(k) \le \alpha \le F(k-)$이다. 여기서 $F(k-) = P_{\theta_0}(f_1 \ge k f_0)$이고, $k = 0$이면 $F(0-) = 1$로 둔다. — ①, 1회 정리 6.3

③ 경계의 확률은 $P_{\theta_0}(f_1 = k f_0) = F(k-) - F(k)$이므로, $\gamma = \frac{\alpha - F(k)}{F(k-) - F(k)}$로 두면(분모가 0이면 $\gamma = 0$) $E_{\theta_0}\varphi^* = F(k) + \gamma\big(F(k-) - F(k)\big) = \alpha$이다. — ②

(b)

① φ를 크기가 $E_{\theta_0}\varphi^*$ 이하인 아무 검정이라 하자. 모든 $\mathbf{x}$에서 $\big(\varphi^*(\mathbf{x}) - \varphi(\mathbf{x})\big)\big(f_1(\mathbf{x}) - k f_0(\mathbf{x})\big) \ge 0$이다. $f_1 > k f_0$이면 $\varphi^* = 1 \ge \varphi$, $f_1 < k f_0$이면 $\varphi^* = 0 \le \varphi$, 같으면 둘째 인수가 0이다. — $0 \le \varphi \le 1$

② ④를 적분한다.

$$0 \le \int (\varphi^* - \varphi)(f_1 - k f_0)\, d\mathbf{x} \qquad \text{④, 적분의 단조성}$$

$$= \big(E_{\theta_1}\varphi^* - E_{\theta_1}\varphi\big) - k\big(E_{\theta_0}\varphi^* - E_{\theta_0}\varphi\big). \qquad \text{적분의 선형성}$$

③ 따라서 $E_{\theta_1}\varphi^* - E_{\theta_1}\varphi \ge k\big(E_{\theta_0}\varphi^* - E_{\theta_0}\varphi\big) \ge 0$이다. — $k \ge 0$, 크기의 가정

(c)

① φ도 크기 α의 최강력 검정이면 $E_{\theta_1}\varphi = E_{\theta_1}\varphi^*$이고 $E_{\theta_0}\varphi \le \alpha = E_{\theta_0}\varphi^*$이다. ⑤의 오른쪽은 $0 - k(\alpha - E_{\theta_0}\varphi) \le 0$이므로 적분은 정확히 0이다. — 최강력의 정의, ⑤

② 음이 아닌 피적분함수의 적분이 0이므로 $(\varphi^* - \varphi)(f_1 - k f_0) = 0$이 거의 어디서나 성립한다. $f_1 \ne k f_0$인 곳에서는 $\varphi = \varphi^*$이다. — ④

조건 확인 $|(\varphi^* - \varphi)(f_1 - k f_0)| \le f_1 + k f_0$이고 오른쪽의 적분은 $1 + k$로 유한하다. 그래서 ⑤에서 적분을 두 기댓값의 차로 나눌 수 있다. 이산분포이면 적분은 합이다. ■

증명의 핵심은 한 줄의 부등식 $(\varphi^* - \varphi)(f_1 - k f_0) \ge 0$이다(단계 ④). 직관은 "배낭 채우기"다. 크기 α만큼의 H_0 확률을 예산으로 쓸 때, 각 점 $\mathbf{x}$를 기각역에 넣는 "값"은 $f_1(\mathbf{x})$, "비용"은 $f_0(\mathbf{x})$이다. 값 대비 비용이 큰 점, 곧 가능도비 f_1/f_0가 큰 점부터 넣는 것이 최선이다. 기각역의 모양은 가능도비의 모양을 따른다(그림 1.1).

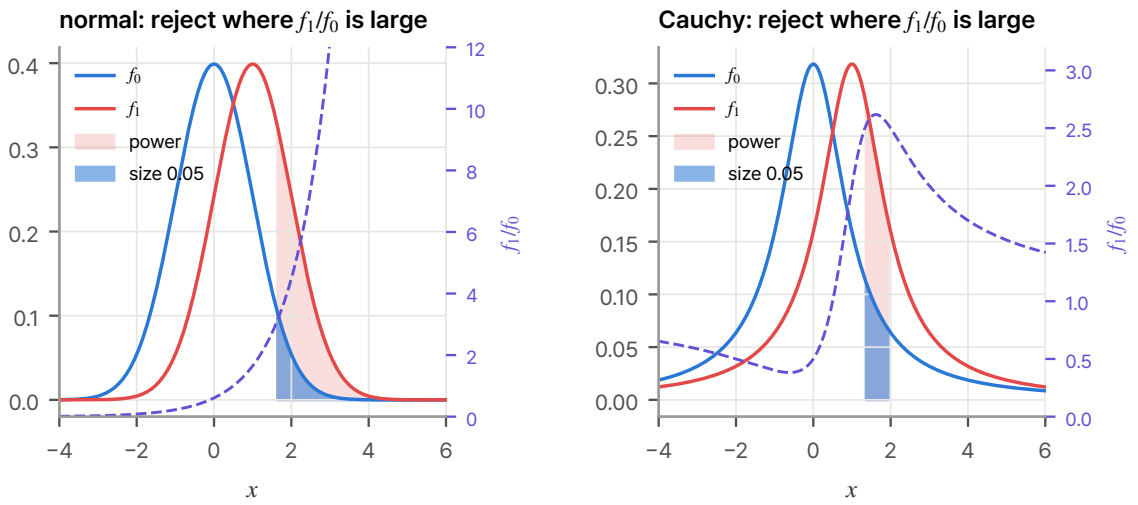


그림 1.1 최강력 기각역은 가능도비가 큰 곳이다. 왼쪽: 정규 $\mathcal{N}(0,1)$ 대 $\mathcal{N}(1,1)$. 가능도비(보라 점선, 오른쪽 축)가 x의 증가함수이므로 기각역은 오른쪽 반직선이다. 기각역에서 f_0 아래 넓이(파랑)가 크기, f_1 아래 넓이(분홍)가 검정력이다. 오른쪽: Cauchy(0) 대 Cauchy(1). 가능도비가 $x \approx 1.6$에서 최대이고 양 끝에서 1로 돌아오므로 크기 0.05의 기각역은 유계인 구간 $[1.37,\ 1.94]$다(연습 1.3).

예제 1.3 (정규 평균) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, $H_0 : \mu = 0$ 대 $H_1 : \mu = \mu_1$ $(\mu_1 > 0)$. 가능도비 $\frac{f_1}{f_0} = \exp\!\left(n\mu_1\bar{x} - \frac{n}{2}\mu_1^2\right)$는 $\bar{x}$의 증가함수이므로, "가능도비 $> k$"는 "$\bar{x} > c$"와 같다. 크기 0.05이면 $c = 1.645/\sqrt{n}$이다. $n = 10$이면 $\bar{x} > 0.520$일 때 기각하고, $\mu_1 = 1$에서의 검정력은 0.935다. 기각역이 μ_1의 값에 의존하지 않는다는 것이 2절의 열쇠다.

예제 1.4 (확률화가 필요한 경우) $X_1, \dots, X_5 \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$, $H_0 : \lambda = 1$ 대 $H_1 : \lambda = 2$. 가능도비는 $T = \sum X_i$의 증가함수이고 H_0 아래 $T \sim \text{Poi}(5)$이다. $P(T \geq 10) = 0.0318$, $P(T \geq 9) = 0.0681$이어서 크기가 정확히 0.05인 비확률화 검정은 없다. $T \geq 10$이면 기각하고 $T = 9$이면 확률 $\gamma = \frac{0.05 - 0.0318}{0.0363} = 0.501$로 기각하면 크기가 0.05가 된다. $\lambda = 2$에서의 검정력은 확률화 검정이 0.605, $T \geq 10$만 쓴 검정(크기 0.032)이 0.542다(연습 1.2).

확률화는 이론을 깨끗하게 만드는 장치다. 실제로 동전을 던져 결론을 정하는 사람은 없다. 실무에서는 p-값(8회 정리 4.2의 이산판)이나 mid-p를 보고한다. 이산분포의 검정은 명목 크기보다 보수적이라는 것만 기억하면 된다.

따름정리 1.5 (최강력 검정의 두 성질) (a) T가 충분통계량이면 최강력 검정은 T의 함수로 쓸 수 있다. (b) 크기 α의 최강력 검정의 검정력은 α 이상이다($E_{\theta_1}\varphi^* \geq \alpha$). 등호는 $f_0 = f_1$일 때뿐이다.

증명. **전략** (a)는 인수분해 정리로 가능도비를 T만의 함수로 쓴다. (b)는 상수 검정 $\varphi \equiv \alpha$와 비교하고, 등호의 경우는 정리 1.2(c)로 읽는다. 단계별 증명은 연습 1.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 33쪽에 있다.

연습문제

연습 1.1 (지수분포 ★) $X_i \overset{\text{iid}}{\sim} \text{Exp}$ (평균 θ)에서 $H_0 : \theta = 1$ 대 $H_1 : \theta = 2$의 크기 α의 최강력 검정이 "$\sum X_i > c$이면 기각" 임을 보이고, c를 χ^2 분포의 분위수로 쓰라. ($\sum X_i \sim$ 감마(n, θ)이고, 4회 정의 3.8로 $\chi^2(2n)$은 감마$(n, 2)$다.) 풀이 32쪽

연습 1.2 (확률화의 검정력 ★) 예제 1.4의 수치를 확인하라. 확률화 검정의 검정력은 $P_2(T \geq 10) + \gamma P_2(T = 9)$이다. 풀이 32쪽

연습 1.3 (Cauchy의 최강력 검정 ★★) $X \sim \text{Cauchy}(\theta)$ 하나로 $H_0 : \theta = 0$ 대 $H_1 : \theta = 1$을 검정한다. 가능도비 $\frac{1+x^2}{1+(x-1)^2}$를 그려 보고, 최강력 검정의 기각역이 x가 큰 쪽의 반직선이 아니라 유계인 구간임을 보여라. 왜 "x가 크면 기각" 이 최선이 아닌가? 풀이 32쪽

연습 1.4 (균등분포 ★★) $X_i \overset{\text{iid}}{\sim} \text{U}(0, \theta)$에서 $H_0 : \theta = 1$ 대 $H_1 : \theta = 2$. 가능도비가 $X_{(n)} \leq 1$이면 2^{-n}, $X_{(n)} > 1$이면 ∞임을 보이고, 크기 α의 최강력 검정이 여럿 있음을 보여라(정리 1.2(c)의 "$f_1 \neq k f_0$ 위에서" 의 뜻). 풀이 32쪽

연습 1.5 (최강력 검정의 두 성질: 증명 ★★ 유도형) 따름정리 1.5를 다음 순서로 증명하라.

(a) 인수분해 정리(11회 정리 1.3)로 $f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)\,h(\mathbf{x})$이다. $h(\mathbf{x}) > 0$인 곳에서 f_1/f_0이 T만의 함수임을 보이고, 정리 1.2(a)의 φ^*를 T의 함수로 쓸 수 있음을 결론지어라. $h(\mathbf{x}) = 0$인 점은 왜 문제가 되지 않는가?

(b) 상수 검정 $\varphi \equiv \alpha$의 크기와 검정력을 구하고, 정리 1.2(b)로 $E_{\theta_1}\varphi^* \geq \alpha$를 얻어라.

(c) 등호이면 $\varphi \equiv \alpha$도 크기 α의 최강력 검정이다. 정리 1.2(c)로 $\{f_1 \neq k f_0\}$ 위에서 $\varphi^* = \alpha$이어야 함을 보이고, 그 위에서 φ^*는 0 또는 1이므로 이 집합이 확률 0임을 얻어라.

(d) $f_1 = k f_0$이 거의 어디서나 성립하면 양변을 적분해 $k = 1$, 곧 $f_0 = f_1$임을 보여라.

풀이 33쪽

2

균일최강력 검정과 단조가능도비

예제 1.3의 기각역 $\bar{x} > 1.645/\sqrt{n}$은 μ_1의 값에 의존하지 않았다. 그러면 이 검정은 $\mu_1 = 0.1$에 대해서도, $\mu_1 = 10$에 대해서도 동시에 최강력이다. 이렇게 대립가설 전체에서 한꺼번에 최선인 검정이 언제 있는가?

2.1 정의

정의 2.1 (균일최강력 검정) $H_0 : \theta \in \Omega_0$ 대 $H_1 : \theta \in \Omega_1$에서, 크기가 α 이하이고 모든 $\theta \in \Omega_1$에서 크기 α 이하인 다른 어떤 검정보다 검정력이 작지 않은 검정을 **균일최강력** (UMP) 검정이라 한다.

정의 2.2 (단조가능도비) 실수 모수 θ의 분포족이 통계량 $T(\mathbf{x})$에 대해 **단조가능도비**(MLR)를 갖는다는 것은, $\theta_1 < \theta_2$이면 가능도비 $f(\mathbf{x};\theta_2)/f(\mathbf{x};\theta_1)$이 $T(\mathbf{x})$의 감소하지 않는 함수라는 것이다.

예제 2.3 (지수족은 단조가능도비를 갖는다) 1-모수 지수족 $f(\mathbf{x};\theta) = h(\mathbf{x})\exp\big(c(\theta)T(\mathbf{x}) - d(\theta)\big)$에서 c가 증가함수이면 $\frac{f(\mathbf{x};\theta_2)}{f(\mathbf{x};\theta_1)} = \exp\big((c(\theta_2) - c(\theta_1))T(\mathbf{x}) - \cdots\big)$이 T의 증가함수다. 정규 평균($T = \sum x_i$), 정규 분산(평균을 알 때, $T = \sum(x_i - \mu)^2$), Poisson과 Bernoulli($T = \sum x_i$), 지수분포의 평균($T = \sum x_i$)이 모두 그렇다. 지수족이 아닌 $\mathrm{U}(0,\theta)$도 $T = X_{(n)}$에 대해 단조가능도비를 갖는다(연습 2.2). Cauchy 위치모형은 갖지 않는다(연습 1.3).

2.2 Karlin–Rubin 정리

정리 2.4 (Karlin–Rubin (1956)) 분포족이 T에 대해 단조가능도비를 가지면, $H_0 : \theta \le \theta_0$ 대 $H_1 : \theta > \theta_0$에서

$$\varphi^*(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > c, \\ \gamma, & T(\mathbf{x}) = c, \\ 0, & T(\mathbf{x}) < c \end{cases} \qquad (E_{\theta_0}\varphi^* = \alpha)$$

는 크기 α의 UMP 검정이다. 그 검정력 함수 $\gamma(\theta) = E_\theta\varphi^*$는 감소하지 않는다.

증명. **전략** 단조가능도비 덕분에 "$T > c$"는 어느 두 모수값 사이에서도 가능도비의 위쪽 집합이다. 그래서 φ^*는 모든 단순 대 단순 문제의 Neyman–Pearson 검정이 되고, 정리 1.2와 따름정리 1.5를 차례로 쓴다.

① $\theta' < \theta''$를 고정한다. 단조가능도비로 $f(\mathbf{x};\theta'')/f(\mathbf{x};\theta') = g\big(T(\mathbf{x})\big)$이고 g는 감소하지 않는다. $k = g(c)$로 둔다. — 정의 2.2

② $f(\mathbf{x};\theta'') > kf(\mathbf{x};\theta')$이면 $g(T) > g(c)$이므로 $T > c$, 곧 $\varphi^* = 1$이다. $f(\mathbf{x};\theta'') < kf(\mathbf{x};\theta')$이면 $T < c$, 곧 $\varphi^* = 0$이다. — g의 단조성 [T2] 대우

③ 따라서 φ^*는 θ' 대 θ''의 Neyman–Pearson 꼴이고, 자기 크기 $E_{\theta'}\varphi^*$의 최강력 검정이다. — ②, 정리 1.2(b)

④ $E_{\theta''}\varphi^* \ge E_{\theta'}\varphi^*$이다. 검정력 함수는 감소하지 않는다. — ③, 따름정리 1.5(b)

⑤ 크기는 $\sup_{\theta\le\theta_0} E_\theta\varphi^* = E_{\theta_0}\varphi^* = \alpha$이다. — ④

⑥ $\theta_1 > \theta_0$과, H_0 전체에서 크기가 α 이하인 검정 φ를 아무거나 고른다. 특히 $E_{\theta_0}\varphi \le \alpha$이다. — 크기의 정의

⑦ ③을 $\theta' = \theta_0$, $\theta'' = \theta_1$에 쓰면 φ^*는 크기 α의 최강력 검정이므로 $E_{\theta_1}\varphi^* \ge E_{\theta_1}\varphi$이다. θ_1이 아무것이었으므로 UMP다. — ⑥, 정의 2.1

조건 확인 가능도비는 ∞를 값으로 가질 수 있다(연습 2.2). $f(\mathbf{x};\theta') = 0 < f(\mathbf{x};\theta'')$인 점에서는 $f'' - kf' > 0$이므로 ②의 분류가 그대로 맞는다. 또 ③에서 쓰는 정리 1.2(b)의 증명은 경계 $\{f'' = kf'\}$ 위의 검정의 값을 쓰지 않으므로, 경계에서 φ^*가 γ와 다른 값을 가져도 된다. ∎

11회 4절의 지수족은 UMP 검정의 무대이기도 하다. 충분통계량 하나(T)가 추정에서도 검정에서도 모든 정보를 담는다.

2.3 양쪽 가설에는 UMP가 없다

예제 2.5 (정규 평균의 양쪽 검정) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, $n = 25$, $H_0 : \mu = 0$ 대 $H_1 : \mu \neq 0$. UMP 검정이 있다면 $\mu_1 > 0$에서는 한쪽 검정 $\bar{x} > 1.645/5$와 같은 검정력을, $\mu_1 < 0$에서는 $\bar{x} < -1.645/5$와 같은 검정력을 가져야 한다(정리 1.2(c)의 유일성). 두 기각역은 다르므로 하나의 검정이 둘 다일 수 없다. 흔히 쓰는 양쪽 검정 $|\bar{x}| > 1.96/5$는 $\mu = 0.4$에서 검정력 0.516으로 한쪽 검정의 0.639보다 낮다. 대신 $\mu = -0.3$에서 0.323을 갖는다. 한쪽 검정은 이쪽에서 0.0008이다(그림 2.1).

증명 없이 쓰는 사실 (균일최강력 불편 검정) 양쪽 가설에서는 기준을 낮춘다. 모든 $\theta \in \Omega_1$에서 검정력이 크기 이상인 검정을 **불편**이라 한다. 불편 검정 가운데서는 UMP가 있을 수 있다. 지수족의 양쪽 검정에서 "$T < c_1$ 또는 $T > c_2$" 꼴의 검정(정규 평균이면 $|z| > 1.96$)이 그것이다(Lehmann–Romano, *Testing Statistical Hypotheses*, 4장). 이 과정에서는 증명하지 않는다.

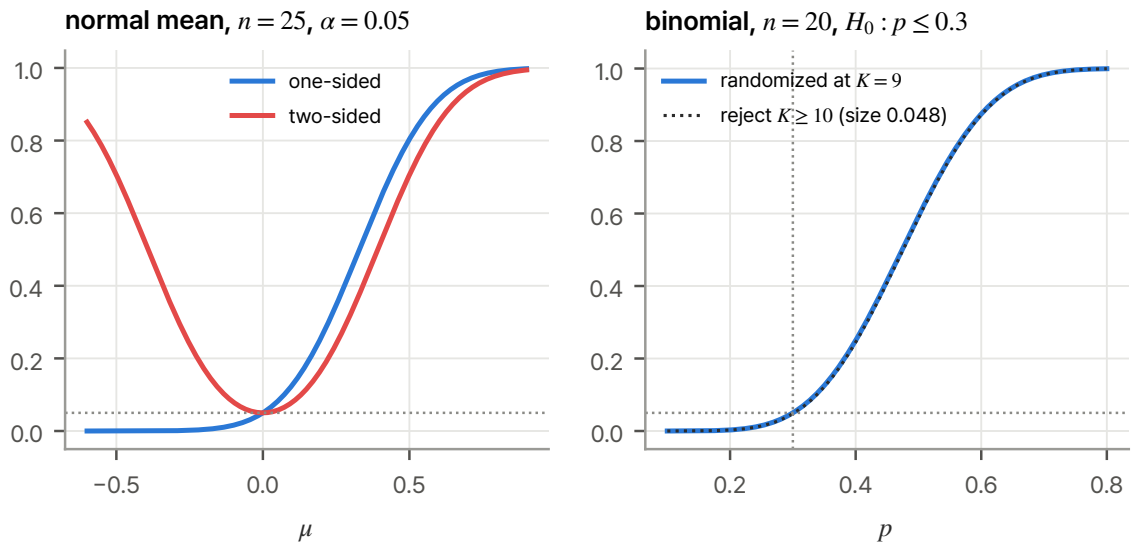


그림 2.1 왼쪽: 정규 평균($n = 25$)에서 한쪽 검정(파랑)과 양쪽 검정(빨강)의 검정력 함수. 오른쪽($\mu > 0$)에서는 한쪽 검정이 낫고, 왼쪽에서는 한쪽 검정의 검정력이 0으로 간다. 오른쪽: 이항 검정($n = 20$, $H_0 : p \le 0.3$)의 검정력 함수는 단조증가한다(정리 2.4). 확률화하지 않은 검정(점선)은 크기가 0.05에 못 미친다.

연습문제

연습 2.1 (지수족의 단조가능도비 ★) 정규 분산(μ를 앎)에서 $H_0 : \sigma^2 \le 1$ 대 $H_1 : \sigma^2 > 1$의 UMP 검정을 구하라. $c(\sigma^2) = -\frac{1}{2\sigma^2}$은 증가함수다. 풀이 33쪽

연습 2.2 (균등분포의 단조가능도비 ★★) $\mathrm{U}(0,\theta)$가 $X_{(n)}$에 대해 단조가능도비를 가짐을 보이고(가능도비는 ∞를 허용한다), $H_0 : \theta \le 1$ 대 $H_1 : \theta > 1$의 UMP 검정 "$X_{(n)} > (1-\alpha)^{1/n}$이면 기각"을 얻어라. 풀이 33쪽

연습 2.3 (Poisson의 UMP ★) 기병대 자료($n = 280$)에서 $H_0 : \lambda \le 0.61$ 대 $H_1 : \lambda > 0.61$의 크기 0.05 UMP 검정(비확률화, 정규근사 없이)의 기각역 $T \ge c$를 구하고, 관측 $T = 196$으로 결론을 내려라(9회 예제 5.3과 비교). 풀이 33쪽

연습 2.4 (불편성 ★★) 예제 2.5의 양쪽 검정 $|\bar{x}| > 1.96/\sqrt{n}$은 불편임을 보여라. 한쪽 검정은 왜 양쪽 가설에 대해 불편이 아닌가? 풀이 33쪽

3

가능도비 검정

UMP 검정은 한쪽 가설과 단조가능도비라는 좁은 조건에서만 있다. 그 밖에서는 Neyman과 Pearson(1928)의 처음 제안으로 돌아간다. 각 가설 아래에서 가장 그럴듯한 값끼리 비교하는 것이다.

정의 3.1 (일반화 가능도비 검정) $H_0 : \theta \in \Omega_0$ 대 $H_1 : \theta \in \Omega \setminus \Omega_0$에서

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Omega} L(\theta; \mathbf{x})} \in [0, 1]$$

가 작으면($\lambda(\mathbf{x}) \le c$) 기각하는 검정을 **가능도비 검정**(LRT)이라 한다. $-2\log\lambda$는 9회 정리 5.2와 10회 정리 1.8의 통계량 Λ이다.

두 가설이 모두 단순이면 $\lambda = f_0/\max(f_0, f_1)$이 f_1/f_0의 감소함수이므로 LRT는 Neyman–Pearson 검정이다. 복합가설에서는 최적성의 보장이 없지만, 대부분의 익숙한 검정이 LRT다.

예제 3.2 (익숙한 검정들은 가능도비 검정이다)

(a) **t 검정.** 정규 표본에서 $H_0 : \mu = \mu_0$ (σ^2 미지)의 LRT는 $-2\log\lambda = n\log\!\left(1 + \frac{t^2}{n-1}\right)$로 $|t|$의 증가함수다(10회 예제 1.9). 임계값을 $t(n-1)$에서 가져오면 크기가 정확히 α이다.

(b) **두 표본 t.** 같은 분산의 두 정규 표본의 평균 비교도 같은 꼴이다(10회 연습 1.5).

(c) **분산의 비.** $H_0 : \sigma_1^2 = \sigma_2^2$의 LRT는 $F = S_1^2/S_2^2$가 1에서 먼지로 정해진다(연습 3.2). 6회 정리 4.3의 F 분포가 정확한 임계값을 준다.

(d) **분할표.** 독립성의 LRT 통계량은 $G = 2\sum O \log \frac{O}{E}$이고 Pearson의 X^2와 점근적으로 같다(9회 연습 5.2).

> **Fisher 대 Neyman** Fisher에게 유의성 검정은 한 번의 실험에서 귀무가설에 반하는 증거를 재는 도구였다. p-값은 그 증거의 크기이고, 대립가설을 미리 정할 필요도, 5%를 기계적으로 지킬 이유도 없었다(**귀납 추론**). Neyman에게 검정은 같은 절차를 되풀이할 때 오류율을 관리하는 규칙이었다. 한 번의 결과가 얼마나 "증거"가 되는지는 묻지 않았다(**귀납 행동**). 두 사람은 1935년 왕립통계학회에서 공개적으로 충돌한 뒤 평생 화해하지 않았다. 오늘의 교과서는 두 관점을 섞어, Neyman의 α와 Fisher의 p-값을 함께 쓴다. 7절에서 이 혼합이 낳는 오해를 본다.

연습문제

연습 3.1 (두 지수분포 ★★) $X_i \overset{\text{iid}}{\sim} \text{Exp}$ (평균 θ_1, $i \le m$), $Y_j \overset{\text{iid}}{\sim} \text{Exp}$ (평균 θ_2, $j \le n$)에서 $H_0 : \theta_1 = \theta_2$의 가능도비가 $\bar{X}/(\bar{X} + \frac{n}{m}\bar{Y})$ 같은 비의 함수임을 보여라. 풀이 34쪽

연습 3.2 (분산의 비 ★★) 두 정규 표본(평균 미지)에서 $H_0 : \sigma_1^2 = \sigma_2^2$의 LRT 통계량이 $F = \hat{\sigma}_1^2/\hat{\sigma}_2^2$의 함수이고, F가 1에서 멀수록 작아짐을 보여라. 풀이 34쪽

연습 3.3 (단순가설의 LRT ★) 두 가설이 모두 단순일 때 LRT가 Neyman–Pearson 검정과 같음을 보여라. 풀이 34쪽

4

축차확률비 검정 — Wald (1945)

> **역사** 1943년 컬럼비아 대학의 통계연구단(SRG)에서 해군 대령 Garret L. Schuyler가 W. Allen Wallis에게 물었다. 포탄의 품질 검사에서 표본을 미리 정한 수만큼 다 쏘지 않아도, 처음 몇 발로 결론이 뻔히 보일 때가 많은데 그때 멈추면 안 되는가? Wallis와 Milton Friedman은 이 질문을 Abraham Wald에게 가져갔다. Wald는 몇 달 만에 **축차확률비 검정**(SPRT)을 만들었다. 같은 오류율을 고정 표본 검정의 절반 정도 관측으로 얻었고, 전쟁이 끝날 때까지 군사 기밀이었다(Wald 1945). 영국의 George Barnard도 같은 시기에 같은 생각에 이르렀다.

4.1 정의와 Wald의 경계

정의 4.1 (축차확률비 검정) $X_1, X_2, \dots$ 를 하나씩 관측하며 $H_0 : f = f_0$ 대 $H_1 : f = f_1$을 검정한다. 로그 가능도비의 누적합

$$\Lambda_n = \sum_{i=1}^{n} \log \frac{f_1(X_i)}{f_0(X_i)}$$

에 대해 상수 $B < 0 < A$를 정하고, $B < \Lambda_n < A$이면 관측을 계속한다. 처음으로 $\Lambda_N \geq A$가 되면 H_1을, $\Lambda_N \leq B$가 되면 H_0를 택한다. N은 멈춘 시각이다.

정리 4.2 (Wald의 경계) SPRT가 확률 1로 멈추면, 실제 오류율 $\alpha' = P_0(H_1 \text{ 택}), \beta' = P_1(H_0 \text{ 택})$은

$$\alpha' \le e^{-A}(1-\beta'), \qquad \beta' \le e^{B}(1-\alpha')$$

를 만족한다. 따라서 목표 α, β에 대해 $A = \log\frac{1-\beta}{\alpha}$, $B = \log\frac{\beta}{1-\alpha}$로 두면 $\alpha' \le \frac{\alpha}{1-\beta}$, $\beta' \le \frac{\beta}{1-\alpha}$, 그리고 $\alpha' + \beta' \le \alpha + \beta$이다.

증명. **전략** 멈춘 경로마다 두 가능도의 비가 e^A 이상(또는 e^B 이하)이다. 이 부등식을 H_1을 택하는 경로 전체에 걸쳐 더하면 두 가설 아래의 확률이 비교된다.

① R_1을 H_1을 택하는 경로들의 집합이라 하자. 각 경로는 어떤 n에서 멈춘 $(x_1, \dots, x_n)$이고, 그 위에서 $\prod_{i\le n} f_1(x_i) \ge e^A \prod_{i\le n} f_0(x_i)$이다. 정의 4.1, $\Lambda_n \ge A$

② R_1을 멈춘 시각 n으로 나눠 ①을 적분하고 n에 대해 더한다.

$$1-\beta' = P_1(R_1) = \sum_n \int_{R_1\cap\{N=n\}} \prod_{i\le n} f_1 \;\ge\; e^A \sum_n \int_{R_1\cap\{N=n\}} \prod_{i\le n} f_0 = e^A \alpha'. \quad \text{가산가법성, ①}$$

③ H_0을 택하는 경로들의 집합 R_0 위에서는 $\prod f_1 \le e^B \prod f_0$이므로 같은 방법으로 $\beta' = P_1(R_0) \le e^B P_0(R_0) = e^B(1-\alpha')$이다. ②와 같다, 확률 1로 멈춤

④ $A = \log\frac{1-\beta}{\alpha}$이면 ②는 $\alpha' \le \frac{\alpha}{1-\beta}(1-\beta') \le \frac{\alpha}{1-\beta}$이다. β'도 같다. ②③

⑤ 두 부등식을 $\alpha'(1-\beta) \le \alpha(1-\beta')$, $\beta'(1-\alpha) \le \beta(1-\alpha')$로 쓰고 더하면 $\alpha'\beta$와 $\alpha\beta'$ 항이 양쪽에서 지워져 $\alpha' + \beta' \le \alpha + \beta$이다. ④

조건 확인 사건 $\{N = n\}$은 $(x_1, \dots, x_n)$만으로 정해지므로(멈춘 시각의 정의) 그 위의 확률이 $\prod_{i\le n} f(x_i)$의 n차원 적분이다. 사건 $R_1 \cap \{N = n\}$ $(n = 1, 2, \dots)$은 서로소이다. 확률 1로 멈춘다는 가정은 ③에서 $P_0(R_0) = 1-\alpha'$, $P_1(R_1) = 1-\beta'$에 쓰인다. ■

경계는 "넘침"(멈출 때 Λ_N이 A나 B를 조금 넘어선 양)을 무시한 것이어서 보수적이다. 실제 오류율은 목표보다 조금 작다. 경계는 f_0과 f_1의 모양에 의존하지 않는다. 증명의 ①–③이 두 밀도의 비만 썼기 때문이다.

증명 없이 쓰는 사실 (SPRT는 확률 1로 멈추고, 최적이다) $f_0 \neq f_1$이면 Λ_n은 0이 아닌 분산을 갖는 확률보행이므로 띠 (B, A)를 벗어나지 못할 확률이 기하급수적으로 줄어든다(연습 4.3). 나아가 Wald와 Wolfowitz(1948)는 오류율이 (α', β') 이하인 모든 축차검정 가운데 SPRT가 $E_0 N$과 $E_1 N$을 **동시에** 최소화함을 증명했다. 이 과정에서는 증명하지 않는다.

4.2 평균 표본수

보조정리 4.3 (Wald의 등식) $Z_1, Z_2, \dots$가 iid이고 $E|Z_1| < \infty$이며, N이 멈춤시각(사건 $\{N \geq i\}$가 $Z_1, \dots, Z_{i-1}$로 정해진다)이고 $EN < \infty$이면 $E\sum_{i=1}^{N} Z_i = EN \cdot EZ_1$이다.

증명. **전략** 무작위 개수의 합을 지시함수로 고정된 무한합으로 바꾼다. 멈춤시각이라는 조건이 각 항의 두 인수를 독립으로 만든다.

① $\sum_{i=1}^{N} Z_i = \sum_{i=1}^{\infty} Z_i \, 1\{N \geq i\}$이다. — 지시함수 [T1]

② $\{N \geq i\} = \{N \leq i-1\}^c$은 $Z_1, \dots, Z_{i-1}$로 정해지므로 Z_i와 독립이다. — 멈춤시각의 정의, iid

③ $E\big[Z_i 1\{N \geq i\}\big] = EZ_i \, P(N \geq i) = EZ_1 \, P(N \geq i)$이다. — ②, 2회 정리 2.7

④ ①의 기댓값을 항별로 취하고 ③을 넣는다.

$$E\sum_{i=1}^{N} Z_i = \sum_{i=1}^{\infty} EZ_1 \, P(N \geq i) = EZ_1 \cdot EN.$$

조건 확인 — 2회 정리 1.13

조건 확인 ④의 첫 등호는 합과 기댓값의 순서를 바꾼다. $|Z_i|$에 ①–③을 그대로 쓰면 $\sum_i E\big[|Z_i| 1\{N \geq i\}\big] = E|Z_1| \, EN < \infty$이므로 Fubini–Tonelli로 바꿀 수 있다 [T6]. $E|Z_1| < \infty$와 $EN < \infty$ 가운데 하나라도 없으면 이 단계가 무너진다. ∎

$Z_i = \log \frac{f_1(X_i)}{f_0(X_i)}$에 쓰면 $E\Lambda_N = EN \cdot EZ$이다. 넘침을 무시하고 Λ_N이 A 또는 B와 같다고 보면

$$E_0 N \approx \frac{(1-\alpha)B + \alpha A}{E_0 Z}, \qquad E_1 N \approx \frac{\beta B + (1-\beta)A}{E_1 Z}$$

이다. $E_0 Z = -D(f_0 \| f_1) < 0$이고 $E_1 Z = D(f_1 \| f_0) > 0$이다(9회 보조정리 1.4). 두 분포가 KL 발산으로 멀수록 빨리 끝난다.

예제 4.4 (불량률 검사) 불량률이 $p_0 = 0.05$이면 받아들이고 $p_1 = 0.15$이면 돌려보내는 검사를 $\alpha = 0.05$, $\beta = 0.10$으로 설계한다. 경계는 $A = \log \frac{0.9}{0.05} = 2.890$, $B = \log \frac{0.1}{0.95} = -2.251$이다. 불량품이면 Λ에 $\log 3 = 1.099$를 더하고, 양품이면 $\log \frac{0.85}{0.95} = -0.111$을 더한다. 4만 번의 시뮬레이션에서 평균 표본수는 $p = 0.05$이면 41.9개, $p = 0.15$이면 39.1개다(Wald의 근사는 넘침을 무시해 39.3과 33.8). 실제 오류율은 0.035와 0.096이다. 같은 오류율을 고정

표본의 이항 검정으로 얻으려면 77개가 필요하다($K \geq 8$이면 기각). SPRT는 절반 남짓으로 같은 일을 한다. 평균 표본수가 가장 큰 것은 두 가설의 중간인 $p = 0.1$ 근처다(약 60개, 그림 4.1).

연습문제

연습 4.1 (Wald의 등식 ★★) 보조정리 4.3에서 N이 멈춤시각이 아니면(다음 관측을 엿보고 멈추면) 결론이 틀릴 수 있음을 $Z_i = \pm 1$의 예로 보여라. 풀이 34쪽

연습 4.2 (정규 평균의 SPRT ★★) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, $H_0 : \mu = 0$ 대 $H_1 : \mu = \delta$의 SPRT가 "$\sum_{i \le n} X_i$가 두 직선 $\frac{B}{\delta} + \frac{n\delta}{2}$와 $\frac{A}{\delta} + \frac{n\delta}{2}$ 사이에 있으면 계속" 임을 보여라. 풀이 34쪽

연습 4.3 (멈출 확률 ★★★) $P(Z_1 > 0) > 0$이면 어떤 m과 $q > 0$이 있어 연속된 m개의 Z의 합이 $A - B$를 넘을 확률이 q 이상임을 보이고, 이것으로 $P(N > km) \le (1-q)^k$, 따라서 $EN < \infty$임을 보여라. 풀이 34쪽

연습 4.4 (고정 표본의 크기 ★) 예제 4.4의 고정 표본 검정을 정규근사로 설계하면 $n \approx \left(\frac{z_\alpha \sqrt{p_0 q_0} + z_\beta \sqrt{p_1 q_1}}{p_1 - p_0}\right)^2$임을 보이고 계산하라(8회 연습 3.3). 풀이 35쪽

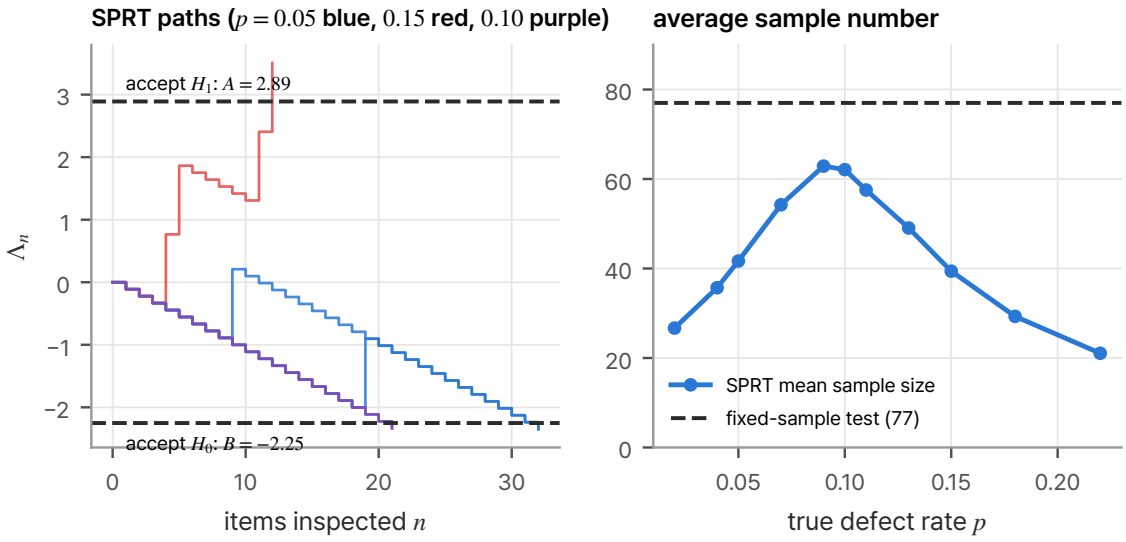


그림 4.1 왼쪽: 불량률 검사의 SPRT 경로 몇 개. 불량품마다 위로 1.10, 양품마다 아래로 0.11 움직이다가 띠(점선)를 벗어나면 멈춘다. 오른쪽: 참 불량률에 따른 평균 표본수(시뮬레이션). 같은 오류율의 고정 표본 검정은 77개(점선)가 필요하다.

5

결정이론 — 검정과 분류를 하나로

> **역사** Wald는 축차검정을 만든 뒤 한 걸음 더 나아갔다. 추정, 검정, 분류는 모두 "자료를 보고 행동을 고르는" 일이고, 행동의 대가를 손실함수로 쓰면 하나의 이론이 된다. 그의 **통계적 결정함수**(1950)는 게임이론(von Neumann–Morgenstern 1944)의 언어로 통계학을 다시 썼다. Bayes 규칙과 minimax 규칙이 그 이론의 두 기둥이다.

5.1 손실, 위험, Bayes 규칙

정의 5.1 (결정 문제) 행동의 집합 $\mathcal{A}$, 손실함수 $L(\theta, a) \geq 0$, 결정규칙 $\delta : \mathbf{x} \mapsto a$가 주어질 때 δ의 **위험**은 $R(\theta, \delta) = E_\theta L(\theta, \delta(\mathbf{X}))$이다. 사전분포 π에 대한 **Bayes 위험**은 $r(\pi, \delta) = \int R(\theta, \delta)\, \pi(d\theta)$이고, 이것을 최소화하는 규칙이 **Bayes 규칙**이다. $\sup_\theta R(\theta, \delta)$를 최소화하는 규칙은 **minimax 규칙**이다.

추정에서는 $\mathcal{A} = \Omega$이고 $L(\theta, a) = (\theta - a)^2$이면 위험이 평균제곱오차다(9회, 11회). 검정에서는 $\mathcal{A} = \{0, 1\}$(받아들임, 기각)이고 L은 두 오류의 비용이다.

정리 5.2 (Bayes 규칙은 사후 기대손실을 줄인다) 사후분포 $\pi(\theta \mid \mathbf{x})$가 있으면, 각 $\mathbf{x}$에서 사후 기대손실 $\int L(\theta, a)\pi(d\theta \mid \mathbf{x})$를 최소화하는 a를 고르는 규칙은 Bayes 규칙이다.

증명. **전략** Bayes 위험의 적분 순서를 바꿔 $\mathbf{x}$마다의 사후 기대손실을 주변밀도로 평균한 꼴로 쓴다. 그러면 적분을 최소화하는 일이 $\mathbf{x}$마다의 최소화로 나뉜다.

① $r(\pi, \delta) = \int\!\!\int L\big(\theta, \delta(\mathbf{x})\big) f(\mathbf{x}; \theta)\, d\mathbf{x}\, \pi(d\theta)$이다. — 정의 5.1

② $f(\mathbf{x}; \theta)\, \pi(d\theta) = \pi(d\theta \mid \mathbf{x})\, m(\mathbf{x})$ (m은 $\mathbf{x}$의 주변밀도)이므로 적분 순서를 바꾸면

$$r(\pi, \delta) = \int \Big[\int L\big(\theta, \delta(\mathbf{x})\big)\, \pi(d\theta \mid \mathbf{x}) \Big] m(\mathbf{x})\, d\mathbf{x}. \qquad \text{조건 확인, 3회 정리 2.5}$$

③ 사후 기대손실을 최소화하는 규칙을 δ_π라 하면, 모든 $\mathbf{x}$에서 대괄호는 $\delta_\pi(\mathbf{x})$일 때 가장 작다. — δ_π의 정의

④ $m \geq 0$으로 ③의 부등식을 적분하면 모든 δ에 대해 $r(\pi, \delta) \geq r(\pi, \delta_\pi)$이다. — 적분의 단조성

조건 확인 $L \geq 0$이므로 ②의 순서 바꾸기는 적분가능성의 가정 없이 Tonelli로 정당하다 [T6]. ∎

양쪽이 모두 ∞일 수도 있는데, 그때 ④는 뜻이 없다. 그래서 Bayes 규칙을 말할 때는 $r(\pi,\delta)<\infty$인 규칙이 하나는 있다고 가정한다. ■

따름정리 5.3 (Bayes 검정은 가능도비 검정이다) $H_0:f_0$ 대 $H_1:f_1$, 사전확률 π_0,π_1, 1종 오류의 비용 c_{I}, 2종 오류의 비용 c_{II}(맞으면 비용 0)이면 Bayes 규칙은

$$\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > \frac{\pi_0 c_{\mathrm{I}}}{\pi_1 c_{\mathrm{II}}} \text{ 이면 기각}$$

이다. 곧 Neyman–Pearson 검정이다. 거꾸로 모든 Neyman–Pearson 검정은 어떤 사전확률과 비용에 대한 Bayes 검정이다.

증명. **전략** 두 행동(기각, 받아들임)의 사후 기대손실을 각각 쓰고 작은 쪽을 고른다(정리 5.2). 공통 인수 $m(\mathbf{x})$를 지우면 가능도비와 임계값의 비교가 남는다. 단계별 증명은 연습 5.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 35쪽에 있다.

Neyman–Pearson의 α는 사전확률과 비용을 말하지 않고 임계값 k를 정하는 방법이었다. 결정이론은 그 k 뒤에 "두 가설을 얼마나 믿는가 × 두 오류가 얼마나 비싼가"가 숨어 있음을 보여 준다(그림 5.1).

5.2 Bayes 분류와 선형판별분석

정리 5.4 (Bayes 분류기) K개의 무리가 사전확률 π_k로 섞여 있고 무리 k의 밀도가 f_k이면, 오분류 확률을 최소화하는 규칙은 $\mathbf{x}$를 $\arg\max_k \pi_k f_k(\mathbf{x})$에 배정하는 것이다.

증명. **전략** 0–1 손실의 사후 기대손실을 사후확률로 쓰고 정리 5.2을 쓴다. 사후확률은 베이즈 정리로 $\pi_k f_k(\mathbf{x})$에 비례한다. 단계별 증명은 연습 5.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 36쪽에 있다.

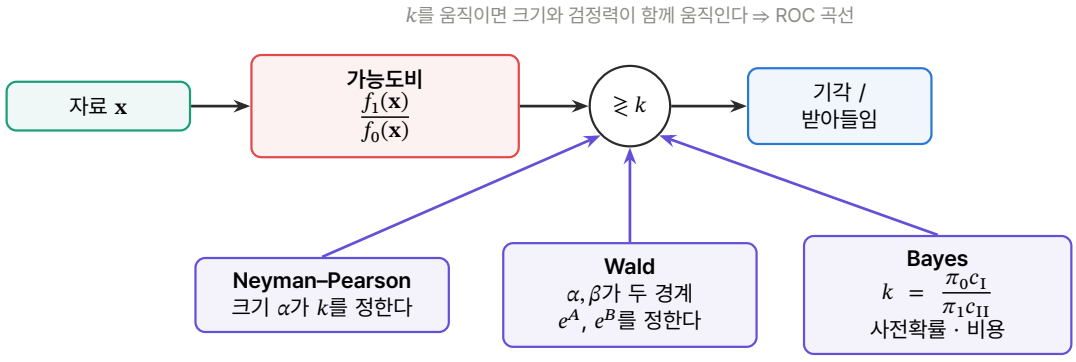


그림 5.1 이번 회차의 세 이론은 모두 가능도비를 임계값과 비교한다. 다른 것은 임계값을 정하는 방법이다. 크기 α로(Neyman–Pearson), 두 오류율의 목표로(Wald), 사전확률과 비용으로(Bayes) 정한다.

따름정리 5.5 (선형판별분석) $f_k = \mathcal{N}_p(\boldsymbol{\mu}_k, \Sigma)$ (공분산이 같음)이면 Bayes 분류기는 판별함수

$$\delta_k(\mathbf{x}) = \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_k - \tfrac{1}{2}\boldsymbol{\mu}_k'\Sigma^{-1}\boldsymbol{\mu}_k + \log \pi_k$$

가 가장 큰 k를 고른다. 두 무리이면 경계는 초평면 $\mathbf{w}'\mathbf{x} = c$, $\mathbf{w} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$이고, $\pi_0 = \pi_1$이면 오분류 확률은 $\Phi(-\Delta/2)$이다. $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$은 **Mahalanobis 거리**의 제곱이다.

증명. **전략** 로그 점수 $\log \pi_k f_k(\mathbf{x})$를 전개하면 이차항이 무리에 무관해 빠진다. 오분류 확률은 판별 점수 $\mathbf{w}'\mathbf{X}$가 정규분포라는 것에서 나온다.

① $\log \pi_k f_k(\mathbf{x}) = \log \pi_k - \tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_k) - \tfrac{1}{2}\log\det(2\pi\Sigma)$이다. — 다변량 정규 밀도, 5회

② 전개하면 $-\tfrac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x} + \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_k - \tfrac{1}{2}\boldsymbol{\mu}_k'\Sigma^{-1}\boldsymbol{\mu}_k$이다. 첫 항과 $\log\det$ 항은 k에 무관하므로 남는 것이 $\delta_k(\mathbf{x})$이다. — Σ^{-1}은 대칭

③ Bayes 분류기는 $\arg\max_k \pi_k f_k(\mathbf{x}) = \arg\max_k \delta_k(\mathbf{x})$를 고른다. — 정리 5.4, log는 증가함수

④ 두 무리이면 $\delta_1(\mathbf{x}) > \delta_0(\mathbf{x})$은 $\mathbf{w}'\mathbf{x} > c$와 같다. $\mathbf{w} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$, $c = \tfrac{1}{2}\mathbf{w}'(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1) + \log\tfrac{\pi_0}{\pi_1}$이다. — ②의 차, $\boldsymbol{\mu}_1'\Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0'\Sigma^{-1}\boldsymbol{\mu}_0 = \mathbf{w}'(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)$

⑤ 무리 0의 $\mathbf{X}$에 대해 $\mathbf{w}'\mathbf{X} \sim \mathcal{N}(\mathbf{w}'\boldsymbol{\mu}_0,\ \mathbf{w}'\Sigma\mathbf{w})$이고 $\mathbf{w}'\Sigma\mathbf{w} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \Delta^2$이다. — 5회 정리 A

⑥ $\pi_0 = \pi_1$이면 $c - \mathbf{w}'\boldsymbol{\mu}_0 = \tfrac{1}{2}\mathbf{w}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \tfrac{1}{2}\Delta^2$이므로 $P_0(\mathbf{w}'\mathbf{X} > c) = 1 - \Phi(\Delta/2) = \Phi(-\Delta/2)$이다. 무리 1에서도 대칭으로 같다. — ⑤, 표준화

■

공분산이 무리마다 다르면 $\mathbf{x}'\Sigma_k^{-1}\mathbf{x}$ 항이 남아 경계가 이차곡면이 된다(이차판별분석, 연습 5.3). Fisher(1936)는 정규분포를 가정하지 않고 같은 방향 $\mathbf{w}$에 이르렀다.

명제 5.6 (Fisher의 판별 기준) $\big(\mathbf{w}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\big)^2 / \mathbf{w}'\Sigma\mathbf{w}$ (무리 사이의 거리 대 무리 안의 흩어짐)를 최대화하는 방향은 $\mathbf{w} \propto \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$이다.

증명. **전략** $\Sigma^{1/2}$로 좌표를 바꿔 분모를 보통의 길이로 만들고 Cauchy–Schwarz 부등식(3회 정리 4.4)을 쓴다 [T4]. 단계별 증명은 연습 5.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 36쪽에 있다.

연습문제

연습 5.1 (Bayes 검정의 임계값 ★) 의료 검사에서 질병의 유병률이 1%이고, 놓친 환자(2종 오류)의 비용이 잘못 경보한 건강인(1종 오류)의 50배라면 가능도비 임계값은 얼마인가? 풀이 35쪽

연습 5.2 (판별의 오류율 ★★) 따름정리 5.5에서 사전확률이 다르면($\pi_0 \neq \pi_1$) 경계가 어떻게 움직이는지 쓰고, 오분류 확률을 구하라. 풀이 35쪽

연습 5.3 (이차판별 ★) 1차원에서 $f_0 = \mathcal{N}(0,1)$, $f_1 = \mathcal{N}(0,4)$, $\pi_0 = \pi_1$이면 Bayes 분류기가 $|x| > c$이면 무리 1로 배정함을 보이고 c를 구하라. 풀이 35쪽

연습 5.4 (minimax 검정 ★★) $H_0 : \mathcal{N}(0,1)$ 대 $H_1 : \mathcal{N}(1,1)$을 관측 하나로 검정할 때 두 오류의 확률의 최댓값을 최소화하는 검정은 $x > \tfrac{1}{2}$이면 기각임을 보이고, 이것이 $\pi_0 = \pi_1$인 Bayes 검정과 같음을 확인하라. 풀이 35쪽

연습 5.5 (Bayes 검정은 가능도비 검정이다: 증명 ★ 유도형) 따름정리 5.3를 다음 순서로 증명하라.

(a) 맞으면 비용이 0이므로, 기각의 사후 기대손실이 $c_{\mathrm{I}}\,\pi(H_0 \mid \mathbf{x})$이고 받아들임의 것이 $c_{\mathrm{II}}\,\pi(H_1 \mid \mathbf{x})$임을 보여라.

(b) 베이즈 정리로 $\pi(H_j \mid \mathbf{x}) = \pi_j f_j(\mathbf{x})/m(\mathbf{x})$를 넣고, 정리 5.2에 따라 기각이 나은 조건이 $f_1(\mathbf{x})/f_0(\mathbf{x}) > \pi_0 c_{\mathrm{I}}/(\pi_1 c_{\mathrm{II}})$임을 얻어라. 두 손실이 같은 점에서는 어느 행동을 골라도 된다.

(c) 거꾸로 임계값 $k > 0$이 주어지면 $\pi_0 c_{\mathrm{I}}/(\pi_1 c_{\mathrm{II}}) = k$가 되는 사전확률과 비용이 있음을 보여라.

풀이 35쪽

연습 5.6 (Bayes 분류기: 증명 ★ 유도형) 정리 5.4를 증명하라.

(a) 0–1 손실 $L(k,a) = 1\{a \neq k\}$에서 행동 a의 사후 기대손실이 $1 - P(\text{무리 } a \mid \mathbf{x})$임을 보여라.

(b) 베이즈 정리로 $P(\text{무리 } a \mid \mathbf{x}) = \pi_a f_a(\mathbf{x}) / \sum_j \pi_j f_j(\mathbf{x})$임을 쓰고, 정리 5.2로 결론을 얻어라.

(c) Bayes 위험이 오분류 확률과 같음을 확인하라.

풀이 36쪽

연습 5.7 (Fisher의 판별 기준: 증명 ★★ 유도형) 명제 5.6를 다음 순서로 증명하라. Σ는 양의 정치다.

(a) $\Sigma = Q\Lambda Q'$ (고유분해)로 $\Sigma^{1/2} = Q\Lambda^{1/2}Q'$를 정의하고, $\Sigma^{1/2}$이 대칭이며 역행렬 $\Sigma^{-1/2}$를 가짐을 보여라.

(b) $\mathbf{v} = \Sigma^{1/2}\mathbf{w}$, $\mathbf{d} = \Sigma^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$로 두면 기준이 $(\mathbf{v}'\mathbf{d})^2/\|\mathbf{v}\|^2$ 임을 보여라.

(c) Cauchy–Schwarz 부등식으로 기준이 $\|\mathbf{d}\|^2 = \Delta^2$ 이하이고, 등호가 $\mathbf{v} \propto \mathbf{d}$, 곧 $\mathbf{w} \propto \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$일 때뿐임을 보여라.

풀이 36쪽

6

데이터로 — Fisher의 붓꽃 (1936)

역사 식물학자 Edgar Anderson은 붓꽃 세 종(*setosa*, *versicolor*, *virginica*)을 50송이씩 모아(그 가운데 둘은 캐나다 Gaspé 반도에서) 꽃받침과 꽃잎의 길이와 너비를 쟀다(1935). 그는 *virginica*와 *versicolor*의 관계를 진화의 관점에서 연구하고 있었다. Fisher(1936)는 이 자료로 "여러 측정값을 하나의 일차결합으로 묶어 종을 가장 잘 가르는 법"을 보였다. 그 일차결합이 명제 5.6의 판별함수다. 같은 해 인도의 Mahalanobis는 무리 사이의 거리 Δ를 정의했다. Fisher의 논문은 *Annals of Eugenics*(우생학 연보)에 실렸다. 통계학의 많은 도구가 우생학의 맥락에서 태어났다는 사실도 이 자료의 역사의 일부다.

예제 6.1 (세 종의 선형판별분석) 네 측정값(꽃받침 길이·너비, 꽃잎 길이·너비, cm)으로 따름정리 5.5의 판별함수를 만든다. 평균은 종마다 따로, 공분산은 세 종을 합쳐 추정한다(자유도 $150-3$). 같은 자료로 분류하면 150송이 가운데 3송이를 틀린다. *setosa*는 50송이 모두 맞고, *versicolor* 2송이가 *virginica*로, *virginica* 1송이가 *versicolor*로 간다. 자료를 하나씩 빼고 나머지로 만든 판별함수로 뺀 것을 분류해도(leave-one-out) 3송이를 틀린다. *setosa*는 다른 두 종에서 Mahalanobis 거리로 9.5와 13.4만큼 떨어져 있어 가르기가 쉽고, *versicolor*와 *virginica* 사이는 4.1뿐이다.

예제 6.2 (versicolor 대 virginica) 두 종만으로 명제 5.6의 방향을 구하면

$$\mathbf{w} = (-3.56,\ -5.58,\ 6.97,\ 12.39)$$

(꽃받침 길이, 꽃받침 너비, 꽃잎 길이, 꽃잎 너비)이다. 꽃잎이 길고 넓을수록, 꽃받침이 짧고 좁을수록 *virginica* 쪽이다. 두 종의 Mahalanobis 거리는 $\Delta = 3.77$이므로 정규 모형 아래의 오분류 확률은 $\Phi(-\Delta/2) = 0.030$이다. 실제로 두 평균의 중점을 경계로 쓰면 100송이 가운데 3송이를 틀린다. 경계를 움직이면 "*virginica*를 *virginica*로 맞히는 비율"(검정력)과 "*versicolor*를 *virginica*로 잘못 부르는 비율"(크기)이 함께 움직인다. 그 궤적이 ROC 곡선이고(8절), 곡선 아래 넓이는 0.997이다(그림 6.1).

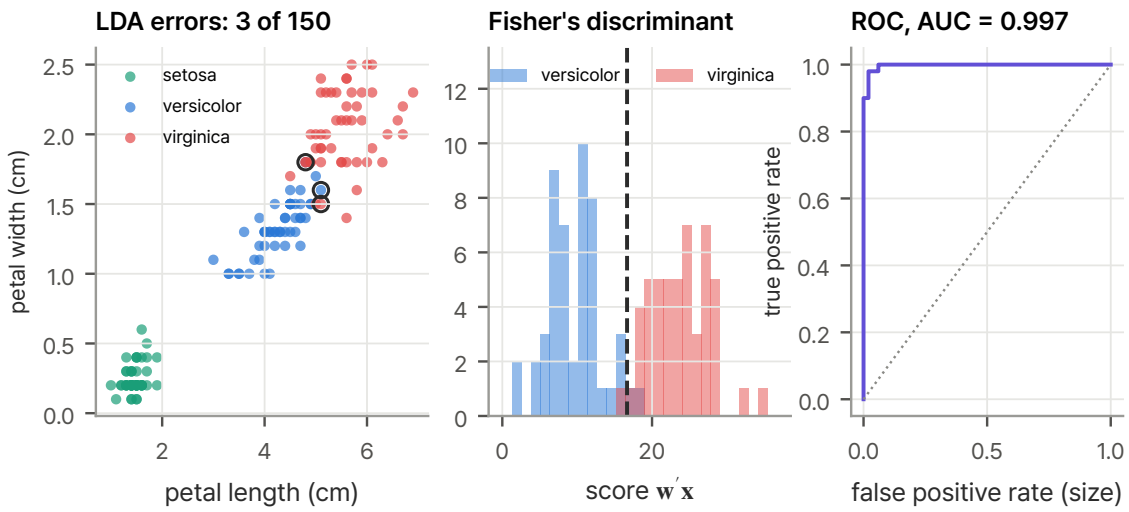


그림 6.1 Fisher의 붓꽃. 왼쪽: 꽃잎의 길이와 너비. 세 종이 색으로, 선형판별분석이 틀린 세 송이가 검은 테두리로 표시되어 있다. 가운데: *versicolor*(파랑)와 *virginica*(빨강)를 Fisher의 방향 $\mathbf{w}$로 투영한 판별 점수. 세로선은 두 평균의 중점. 오른쪽: 경계를 움직여 그린 ROC 곡선(넓이 0.997).

연습문제

연습 6.1 (판별 방향의 해석 ★) 예제 6.2의 $\mathbf{w}$로 만든 점수의 두 종 평균의 차를 점수의 표준편차로 나누면 $\Delta = 3.77$임을 노트북으로 확인하라(명제 5.6의 최댓값). 풀이 36쪽

연습 6.2 (같은 자료로 재면 ★) 같은 자료로 판별함수를 만들고 같은 자료로 오류율을 재면 왜 낙관적인가? leave-one-out이 그 문제를 어떻게 줄이는가? 풀이 36쪽

7

부쉬보기 — p-값이 말하지 않는 것

7.1 표본이 거대하면

검정력은 표본 크기와 함께 1로 간다(8회 정리 3.4). 그 말은 뒤집으면 이렇다. 귀무가설이 **정확히** 참이 아닌 한, 표본이 충분히 크면 아무리 작은 차이도 유의하다.

> **예제 7.1** (A/B 테스트의 0.1%포인트) 웹페이지의 두 판을 사용자 100만 명씩에게 보였더니 클릭률이 10.0%와 10.1%였다. 두 비율의 차의 표준오차는 $\sqrt{2\cdot 0.1\cdot 0.9/10^6}\approx 0.00042$이므로 $z=2.35$, $p=0.019$다. 유의하다. 그러나 95% 구간은 $[0.0002,\ 0.0018]$, 곧 0.02%포인트에서 0.18%포인트다. 이 차이가 페이지를 바꿀 만한지는 p-값이 아니라 효과의 크기와 비용이 정한다. 같은 차이를 사용자 1만 명씩에서 보았다면 $z=0.24$로 전혀 유의하지 않았다. p-값은 효과의 크기와 표본 크기를 한 숫자에 섞는다.

그래서 결과는 p-값 하나가 아니라 효과의 추정값과 신뢰구간으로 보고해야 한다. 구간과 검정은 같은 것이지만(8회 정리 3.5), 구간은 "얼마나"를 함께 말한다.

7.2 Lindley의 역설 (1957)

p-값은 귀무가설이 참일 확률이 아니다. 둘이 얼마나 다를 수 있는지는 Bayes의 관점에서 보면 분명해진다.

> **예제 7.2** (같은 p-값, 다른 증거) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu,1)$에서 $H_0:\mu=0$를 검정하는데 $z=\sqrt{n}\bar{x}=2.0$이 나왔다($p=0.046$). 대립가설 아래 $\mu\sim\mathcal{N}(0,1)$이라는 사전분포를 두면, 두 가설의 Bayes 인수는
>
> $$B_{01}=\frac{f(\mathbf{x}\mid H_0)}{f(\mathbf{x}\mid H_1)}=\sqrt{1+n}\,\exp\!\left(-\frac{z^2}{2}\cdot\frac{n}{n+1}\right)$$
>
> 이다(연습 7.2). z를 2.0으로 고정하면 $n=100$에서 $B_{01}=1.4$, $n=10^4$에서 13.5, $n=10^6$에서 135다. 두 가설의 사전확률이 반반이면 $n=10^6$에서 $P(H_0\mid\mathbf{x})=0.993$이다. 같은 $p=0.046$이 작은 표본에서는 H_0에 약간 반하는 증거, 큰 표본에서는 H_0를 강하게 지지하는 증거다. 큰 표본에서 $z=2$는 $\bar{x}=0.002$, 곧 0에 매우 가까운 값이고, 그것은 "어딘가 넓게 퍼진 μ"보다 "정확히 0"과 더 잘 맞는다.

가정 주의 미국통계학회는 2016년 p-값에 관한 성명에서 여섯 원칙을 밝혔다. 그 가운데 셋이 이 절의 내용이다. p-값은 가설이 참일 확률이 아니다. p-값은 효과의 크기나 결과의 중요성을 재지 않는다. $p < 0.05$ 하나로 과학적 결론이나 정책을 정해서는 안 된다. 8회 7절의 다중검정과 선택적 멈춤까지 더하면, p-값은 정해진 절차 안에서 오류율을 관리하는 도구로 쓸 때만 뜻이 분명하다.

연습문제

연습 7.1 (구간으로 보고하기 ★) 예제 7.1의 95% 구간을 계산하고, 상대적인 개선(10.0% 대비)으로 바꿔 쓰라. 사용자 1만 명씩이었다면 구간은 어떻게 되는가? 풀이 36쪽

연습 7.2 (Lindley의 Bayes 인수 ★★) $\bar{X} \mid \mu \sim \mathcal{N}(\mu, 1/n)$, H_1 아래 $\mu \sim \mathcal{N}(0, \tau^2)$이면 H_1 아래 $\bar{X}$의 주변분포가 $\mathcal{N}(0, \tau^2 + 1/n)$임을 보이고, $B_{01} = \sqrt{1 + n\tau^2}\,\exp\!\left(-\frac{z^2}{2}\,\frac{n\tau^2}{1+n\tau^2}\right)$을 유도하라. 풀이 37쪽

연습 7.3 (p-값의 오독 ★) "$p = 0.03$이므로 귀무가설이 참일 확률은 3%다"와 "$p = 0.03$이므로 우연히 이런 결과가 나올 확률은 3%다"의 무엇이 틀렸는지 설명하라. 정확한 문장은 무엇인가? 풀이 37쪽

8

Beautiful insights — ROC와 멈출 수 있는 검정

8.1 ROC 곡선은 Neyman–Pearson 검정들의 모임이다

분류기는 입력마다 점수 $s(\mathbf{x})$를 내놓고, 임계값 t를 넘으면 "양성"이라 부른다. 임계값을 움직이며 "음성인데 양성이라 부른 비율"(거짓양성률, 검정의 크기)과 "양성을 양성이라 부른 비율"(참양성률, 검정력)을 그린 곡선이 **ROC 곡선**이다. 신호탐지 이론(2차대전의 레이더)에서 나온 이름이다.

Neyman–Pearson 보조정리는 ROC의 언어로 이렇게 읽힌다. 점수가 가능도비 $f_1(\mathbf{x})/f_0(\mathbf{x})$(또는 그 단조함수)이면, 각 임계값의 검정은 그 크기에서 최강력이다. 그러므로 **가능도비의 *ROC* 곡선은 어떤 점수의 *ROC* 곡선보다도 위에 있다**(연습 8.1). 좋은 분류기를 만드는 일은 가능도비, 곧 사후확률(정리 5.4)을 잘 근사하는 일이다. 임계값을 고르는 일은 결정이론이 맡는다(따름정리 5.3). 유병률과 두 오류의 비용이 임계값을 정한다.

ROC 곡선 아래 넓이(AUC)는 "양성 하나와 음성 하나를 무작위로 골랐을 때 양성의 점수가 더 클 확률"이다(연습 8.2). 이것은 15회의 Mann–Whitney 통계량과 같다. 붓꽃의 0.997은 *virginica* 한 송이와 *versicolor* 한 송이를 뽑으면 거의 언제나 판별 점수가 옳은 순서라는 뜻이다.

8.2 보면서 멈출 수 있는 검정

온라인 A/B 테스트에서는 결과가 실시간으로 쌓인다. 분석가는 매일 p-값을 보고, 유의해지면 실험을 멈추고 싶다. 8회 예제 7.1에서 보았듯이 고정 표본의 검정을 이렇게 쓰면 1종 오류가 커진다. 표준정규 자료를 100개마다 보며 $|z| > 1.96$이면 멈추는 일을 1만 개까지 하면, 귀무가설이 참인데도 38%가 "유의" 해진다. 관측마다 보면 59%다(그림 8.1).

Wald의 SPRT는 처음부터 "보면서 멈추는" 검정이었다. 다만 대립가설이 하나의 값(θ_1)이어야 했다. Robbins(1970)의 **혼합 SPRT**는 대립가설의 모수에 분포 π를 두고 가능도비를 평균한다.

$$M_n = \int \prod_{i \le n} \frac{f_\theta(X_i)}{f_0(X_i)} \, \pi(d\theta).$$

H_0 아래 M_n은 평균이 1인 음이 아닌 마틴게일이다. Ville의 부등식은 그런 과정이 한 번이라도

$1/\alpha$를 넘을 확률이 α 이하라고 말한다.

$$P_0\Big(\sup_n M_n \geq \frac{1}{\alpha}\Big) \leq \alpha.$$

그래서 "$M_n \geq 1/\alpha$가 되는 순간 멈추고 기각"하는 검정은 언제 보든, 언제 멈추든 크기가 α 이하다. 정규 평균에서 $\theta \sim \mathcal{N}(0, \tau^2)$로 두면 $M_n = (1+n\tau^2)^{-1/2}\exp\!\big(\frac{\tau^2 S_n^2}{2(1+n\tau^2)}\big)$ ($S_n = \sum X_i$)로 닫힌 꼴이다. 관측마다 보아도 거짓양성률은 약 4%다($\alpha = 0.05$, $\tau^2 = 0.01$, 1만 개까지). 효과가 0.05 표준편차이면 1만 개 안에 98%가 기각되고, 멈추는 시점의 중앙값은 약 2,900번째다. 오늘의 A/B 테스트 플랫폼이 쓰는 "언제나 유효한 p-값"(Johari 외 2017)이 이것이다.

> **증명 없이 쓰는 사실** (Ville의 부등식) $M_n \geq 0$이 마틴게일이고 $EM_0 = 1$이면 $P(\sup_n M_n \geq c) \leq 1/c$이다. 이산시간이면 "처음으로 c를 넘는 시각에서 멈춘 과정"에 마틴게일의 멈춤 정리를 쓰는 짧은 증명이 있다(연습 8.3). 혼합 SPRT의 M_n이 마틴게일인 것은, 각 θ의 가능도비 곱이 H_0 아래 평균 1인 마틴게일이고(3회 반복기댓값) 그 평균도 마틴게일이기 때문이다.

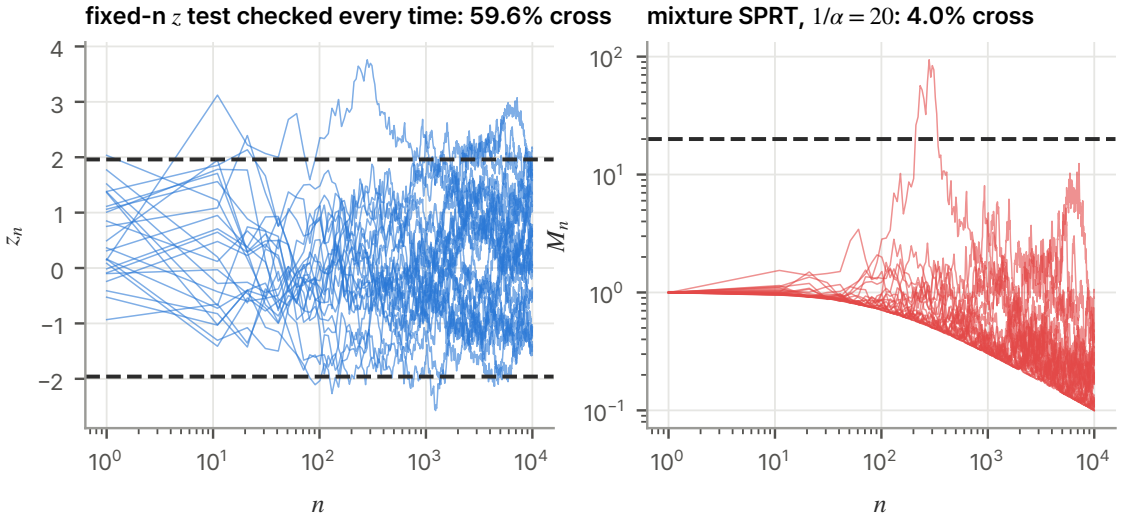


그림 8.1 귀무가설이 참일 때(표준정규 자료) 관측마다 결과를 보는 경우. 왼쪽: 고정 표본 z 검정을 매번 적용하면 경로가 ±1.96을 한 번이라도 넘을 확률이 표본과 함께 커진다. 오른쪽: 혼합 SPRT의 M_n이 $1/\alpha = 20$을 넘은 경로는 약 4%뿐이다.

연습문제

연습 8.1 (가능도비의 ROC가 위에 있다 ★★) 아무 점수 s의 ROC 곡선 위의 점 (α, γ)에 대해, 가능도비의 ROC 곡선의 α에서의 높이가 γ 이상임을 정리 1.2로 보여라. 풀이 37쪽

연습 8.2 (AUC와 순위 ★) 양성 점수 S_1과 음성 점수 S_0가 독립이고 연속이면 ROC 곡선 아래 넓이가 $P(S_1 > S_0)$임을 보여라. 풀이 37쪽

연습 8.3 (Ville의 부등식 ★★★) $M_n \geq 0$이 평균 1인 마틴게일이고 $\tau = \min\{n : M_n \geq c\}$이면, N을 고정하고 $\tau \wedge N$에서 멈춘 값의 평균이 1임(멈춤 정리)을 써서 $P(\max_{n \leq N} M_n \geq c) \leq 1/c$를 보이고, $N \to \infty$로 보내라. 풀이 37쪽

9

정리와 역사 노트

9.1 한 장 요약

정리	내용	조건
Neyman–Pearson (1.2)	가능도비 $> k$에서 기각하는 검정이 최강력	단순 대 단순
최강력 검정의 성질 (1.5)	충분통계량의 함수, 검정력 $\geq$ 크기	—
Karlin–Rubin (2.4)	$T > c$ 꼴이 한쪽 UMP, 검정력 함수는 단조	단조가능도비
Wald의 경계 (4.2)	$\alpha' \leq e^{-A}(1-\beta'), \beta' \leq e^{B}(1-\alpha')$	SPRT가 멈춤
Wald의 등식 (4.3)	$E\sum_{i\leq N} Z_i = EN \cdot EZ$	N 멈춤시각
Bayes 규칙 (5.2, 5.3)	사후 기대손실 최소, Bayes 검정 = 가능도비 검정	—
Bayes 분류 · LDA (5.4, 5.5)	$\arg\max \pi_k f_k$, 정규·같은 공분산이면 선형, 오류 $\Phi(-\Delta/2)$	—
Fisher의 기준 (5.6)	$\mathbf{w} \propto \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$	—

이번 회차의 모든 최적성은 가능도비에서 나왔다. 단순가설의 최강력 검정(Neyman–Pearson), 한쪽 가설의 균일최강력 검정(Karlin–Rubin), 축차검정(Wald), Bayes 검정과 분류기가 모두 "가능도비를 임계값과 비교하라" 였다. 다른 것은 임계값을 정하는 방법뿐이다. 크기로 정하거나(Neyman–Pearson), 오류율의 목표로 정하거나(Wald), 사전확률과 비용으로 정한다(Bayes).

9.2 Stigler의 명명 법칙

> **이름과 발견**
>
> - **Neyman–Pearson 보조정리**: 대립가설을 생각하라는 착상은 Gosset의 1926년 편지에 있었다.
> - **SPRT**: Wald(1945)와 같은 시기에 영국의 Barnard(1946)가 따로 같은 검정에 이르렀다. 질문을 처음 던진 사람은 해군 장교 Schuyler와 경제학자 Wallis·Friedman이었다.
> - **선형판별분석**: Fisher(1936)의 판별함수는 정규분포를 가정하지 않은 기준(명제 5.6)에서 나왔다. 정규 모형의 Bayes 분류기로서의 유도는 Welch(1939)와 Wald(1944)가 했다. 자료는 Anderson의

> 것이다.
> - **Karlin–Rubin**: 단조가능도비의 UMP 정리는 Karlin과 Rubin(1956)의 이름으로 불리지만, 정규평균처럼 단순한 경우의 "모든 대립가설에 공통인 최량 기각역"은 Neyman–Pearson(1933)이 이미 다뤘다.

9.3 7 Pillars 지도

이번 회차는 셋째 기둥 **Likelihood**의 검정 쪽 완성이다. 9회가 가능도를 추정에 썼다면, 12회는 가능도비가 가장 좋은 검정을 만든다는 것을 보였다. 다섯째 기둥 **Design**의 씨앗도 보였다. Wald의 축차검정은 "몇 개를 관측할지"를 자료가 정하게 한 최초의 설계다. 14회의 실험설계가 그 이야기를 잇는다.

9.4 다음 회 — 최소제곱: Legendre–Gauss에서 Gauss–Markov까지 (1749 → 1943)

지금까지의 모형은 대부분 한 모집단의 모수 몇 개였다. 13회는 관측이 다른 변수의 일차함수에 오차가 더해진 것이라는 모형, 곧 선형모형으로 넘어간다. Euler(1749)가 풀지 못한 달의 운동, Mayer(1750)의 묶음 평균, Legendre(1805)의 최소제곱과 Gauss의 우선권 주장에서 출발한다. 사영의 기하로 최소제곱을 다시 세우고, Gauss–Markov 정리와 $t \cdot F$ 검정을 증명한다. 이번 회차의 가능도비 검정이 거기서 F 검정이 된다.

참고문헌

- Neyman, J., Pearson, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 20A (1928), 175–240, 263–294. · On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. A* 231 (1933), 289–337.
- Neyman, J. Statistical problems in agricultural experimentation. *JRSS Suppl.* 2 (1935), 107–180 (Fisher의 토론 포함).
- Wald, A. Sequential tests of statistical hypotheses. *Ann. Math. Statist.* 16 (1945), 117–186. · *Statistical Decision Functions*. Wiley, 1950. · Wald, A., Wolfowitz, J. Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* 19 (1948), 326–339.
- Barnard, G. A. Sequential tests in industrial statistics. *JRSS Suppl.* 8 (1946), 1–26. · Wallis, W. A. The Statistical Research Group, 1942–1945. *JASA* 75 (1980), 320–330.
- Karlin, S., Rubin, H. The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* 27 (1956), 272–299.

- Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7 (1936), 179–188. · Anderson, E. The irises of the Gaspé Peninsula. *Bull. Amer. Iris Soc.* 59 (1935), 2–5. · Mahalanobis, P. C. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2 (1936), 49–55.
- Lindley, D. V. A statistical paradox. *Biometrika* 44 (1957), 187–192. · Wasserstein, R. L., Lazar, N. A. The ASA statement on *p*-values. *American Statistician* 70 (2016), 129–133.
- Robbins, H. Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* 41 (1970), 1397–1409. · Johari, R., Koomen, P., Pekelis, L., Walsh, D. Peeking at A/B tests. *KDD* (2017).
- Lehmann, E. L., Romano, J. P. *Testing Statistical Hypotheses*, 4th ed. Springer, 2022.
- 자료: `datasets::iris`.

A

연습문제 풀이

각 풀이의 머리에 있는 문제 ○○쪽을 누르면 문제로 돌아간다. 계산 문제의 수치는 노트북에서 같은 계산으로 확인했다.

1절

연습 1.1 문제 6쪽

$\frac{f_1}{f_0} = \frac{2^{-n}e^{-\sum x_i/2}}{e^{-\sum x_i}} = 2^{-n}e^{\sum x_i/2}$는 $\sum x_i$의 증가함수이므로 최강력 검정은 $\sum X_i > c$에서 기각한다. H_0 아래 $\sum X_i \sim$ 감마$(n,1)$이므로 $2\sum X_i \sim$ 감마$(n,2) = \chi^2(2n)$이고 $c = \frac{1}{2}\chi^2_{2n,\alpha}$이다.

연습 1.2 문제 6쪽

H_0 아래 $P(T \geq 10) = 0.0318$, $P(T = 9) = 0.0363$이므로 $\gamma = (0.05 - 0.0318)/0.0363 = 0.501$. H_1 아래 $T \sim \text{Poi}(10)$이고 $P(T \geq 10) = 0.542$, $P(T = 9) = 0.125$이므로 검정력은 $0.542 + 0.501 \cdot 0.125 = 0.605$.

연습 1.3 문제 6쪽

가능도비 $r(x) = \frac{1+x^2}{1+(x-1)^2}$는 $x \to \pm\infty$에서 1로 가고, $x = \frac{1+\sqrt{5}}{2} \approx 1.62$ 근처에서 최대(약 2.6)다. 그래서 $\{r > k\}$ $(k > 1)$는 유계인 구간이다. 크기 0.05이면 $k = 2.53$, 기각역은 $[1.37,\ 1.94]$, 검정력은 0.13이다(노트북). 아주 큰 x는 $\theta = 0$에서도 $\theta = 1$에서도 비슷하게 드문 꼬리 관측이어서 두 가설을 가르지 못한다.

연습 1.4 문제 6쪽

$X_{(n)} \leq 1$이면 $f_1/f_0 = 2^{-n}/1$, $X_{(n)} > 1$이면 $f_0 = 0$이므로 ∞이다. $k = 2^{-n}$으로 두면 $\{f_1 > kf_0\} = \{X_{(n)} > 1\}$은 H_0 아래 확률 0이고, 경계 $\{f_1 = kf_0\} = \{X_{(n)} \leq 1\}$ 위에서는 크기가 α가 되도록 아무렇게나 기각해도 된다. 예를 들어 "$X_{(n)} > (1-\alpha)^{1/n}$이면 기각"과 "$X_{(n)} > 1$ 또는 $X_{(1)} > 1 - \alpha^{1/n}$이면 기각"은 둘 다 크기가 α이고 검정력이 $1 - (1-\alpha)2^{-n}$으로 같다. ($\theta = 2$이면 $X_{(n)} \leq 1$일 확률이 2^{-n}이고, 그 조건 아래 표본은 U$(0,1)$이므로 어느 검정이든 확률 α로 기각한다.)

연습 1.5 문제 6쪽

(a) $h(\mathbf{x})>0$이면 $\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}=\frac{g(T(\mathbf{x});\theta_1)h(\mathbf{x})}{g(T(\mathbf{x});\theta_0)h(\mathbf{x})}=\frac{g(T(\mathbf{x});\theta_1)}{g(T(\mathbf{x});\theta_0)}$로 T만의 함수다. 그러면 $\{f_1>kf_0\}$, $\{f_1=kf_0\}$, $\{f_1<kf_0\}$가 $\{h>0\}$ 위에서 T의 값만으로 정해지므로 φ^*는 T의 함수다. $h(\mathbf{x})=0$이면 $f_0(\mathbf{x})=f_1(\mathbf{x})=0$이고, 그런 점들의 집합은 두 분포 아래 모두 확률 0이므로 거기서 φ^*를 어떻게 정해도 크기와 검정력이 바뀌지 않는다. (b) $\varphi\equiv\alpha$의 크기는 $E_{\theta_0}\alpha=\alpha$이고 검정력도 α이다. 정리 1.2(b)로 $E_{\theta_1}\varphi^*\geq E_{\theta_1}\alpha=\alpha$이다. (c) 등호이면 $\varphi\equiv\alpha$는 크기 α에서 최강력 검정과 같은 검정력을 가지므로 그 자체로 최강력이다. 정리 1.2(c)로 $\{f_1\neq kf_0\}$ 위에서 거의 어디서나 $\alpha=\varphi^*$이다. 그런데 그 위에서 $\varphi^*\in\{0,1\}$이고 $0<\alpha<1$이므로, 이 집합은 측도 0이다. (d) $f_1=kf_0$이 거의 어디서나 성립하면 $1=\int f_1=k\int f_0=k$이므로 $k=1$이고 $f_1=f_0$ (거의 어디서나)이다. 두 분포가 같다.

2절

연습 2.1 문제 10쪽

$f=(2\pi\sigma^2)^{-n/2}\exp\!\big(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\big)$에서 $c(\sigma^2)=-\frac{1}{2\sigma^2}$이 증가하므로 $T=\sum(X_i-\mu)^2$에 대해 단조가능도비. UMP 검정은 $T>c$이고 H_0의 경계 $\sigma^2=1$에서 $T\sim\chi^2(n)$이므로 $c=\chi^2_{n,\alpha}$.

연습 2.2 문제 10쪽

$\theta_1<\theta_2$이면 $\frac{f(\mathbf{x};\theta_2)}{f(\mathbf{x};\theta_1)}=(\theta_1/\theta_2)^n$ $(x_{(n)}\leq\theta_1)$, ∞ $(\theta_1<x_{(n)}\leq\theta_2)$로 $x_{(n)}$의 감소하지 않는 함수다. 정리 2.4으로 $X_{(n)}>c$가 UMP이고, $P_1(X_{(n)}>c)=1-c^n=\alpha$에서 $c=(1-\alpha)^{1/n}$.

연습 2.3 문제 10쪽

H_0의 경계 $\lambda=0.61$에서 $T\sim\mathrm{Poi}(170.8)$이고 $P(T\geq c)\leq0.05$인 가장 작은 정수 c는 194다(실제 크기 0.043, 노트북). $T=196\geq194$이므로 크기 0.05에서 기각한다. 한쪽 p-값은 $P(T\geq196)=0.03$ 정도다. 9회의 양쪽 검정($p=0.05$–0.07)은 기각하지 못했다. 한쪽 가설을 미리 정했다면 결론이 달라진다.

연습 2.4 문제 10쪽

검정력 $\gamma(\mu)=\Phi(-1.96-\sqrt{n}\mu)+1-\Phi(1.96-\sqrt{n}\mu)$는 $\mu=0$에서 최솟값 0.05를 갖는다($\gamma'(0)=0$이고 볼록). 따라서 모든 $\mu\neq0$에서 0.05 이상이다. 한쪽 검정은 $\mu<0$에서 검정력이 0.05보다 작아 불편이 아니다.

3절

연습 3.1 문제 12쪽

H_1 아래 MLE는 $\bar{x}$, $\bar{y}$이고 H_0 아래는 $\hat{\theta} = \frac{m\bar{x}+n\bar{y}}{m+n}$이다. $\lambda = \frac{\bar{x}^m \bar{y}^n}{\hat{\theta}^{m+n}} = \frac{(m+n)^{m+n}}{m^m n^n} u^m (1-u)^n$, $u = \frac{m\bar{x}}{m\bar{x}+n\bar{y}}$이므로 u의 함수다. u가 $\frac{m}{m+n}$에서 멀수록 작다.

연습 3.2 문제 12쪽

H_1 아래 $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ (n으로 나눔), H_0 아래 $\hat{\sigma}^2 = \frac{m\hat{\sigma}_1^2 + n\hat{\sigma}_2^2}{m+n}$이고 $\lambda = \frac{(\hat{\sigma}_1^2)^{m/2}(\hat{\sigma}_2^2)^{n/2}}{(\hat{\sigma}^2)^{(m+n)/2}}$이다. $F = \hat{\sigma}_1^2/\hat{\sigma}_2^2$로 쓰면 $\lambda \propto \frac{F^{m/2}}{(mF+n)^{(m+n)/2}}$이고, 이것은 $F = 1$에서 최대이며 F가 1에서 멀어질수록 작아진다.

연습 3.3 문제 12쪽

$\lambda = \frac{f_0}{\max(f_0, f_1)} = \min\left(1, \frac{f_0}{f_1}\right)$이므로 $\lambda \le c < 1$은 $f_1/f_0 \ge 1/c$와 같다.

4절

연습 4.1 문제 16쪽

$Z_i = \pm 1$ (각 확률 $\frac{1}{2}$)이고 $Z_2 = -1$이면 $N = 1$, $Z_2 = +1$이면 $N = 2$로 하자. $\{N \ge 2\} = \{Z_2 = +1\}$이 Z_1만으로 정해지지 않으므로 멈춤시각이 아니다(아직 보지 않은 Z_2를 엿보고 멈춘다). $\sum_{i \le N} Z_i$는 $Z_2 = -1$이면 Z_1, $Z_2 = +1$이면 $Z_1 + 1$이므로 평균이 $\frac{1}{2}$이다. 그런데 $EN \cdot EZ = \frac{3}{2} \cdot 0 = 0$이다.

연습 4.2 문제 16쪽

$\log \frac{f_1(x)}{f_0(x)} = \delta x - \frac{\delta^2}{2}$이므로 $\Lambda_n = \delta S_n - \frac{n\delta^2}{2}$. $B < \Lambda_n < A$는 $\frac{B}{\delta} + \frac{n\delta}{2} < S_n < \frac{A}{\delta} + \frac{n\delta}{2}$와 같다. 두 평행선 사이의 띠다.

연습 4.3 문제 16쪽

$P(Z > 0) > 0$이면 $P(Z > \varepsilon) = q_0 > 0$인 $\varepsilon > 0$이 있다. $m > (A-B)/\varepsilon$로 잡으면 연속된 m개가 모두 ε보다 클 확률 $q = q_0^m$에서 합이 $A - B$를 넘어 띠를 벗어난다. 서로 겹치지 않는 k개의 구간에서 한 번도 그런 일이 없을 확률은 $(1-q)^k$이므로 $P(N > km) \le (1-q)^k$이고, $EN = \sum P(N > j) \le m \sum_k (1-q)^k < \infty$.

연습 4.4 문제 16쪽

$\hat{p}$의 정규근사로 기각역 $\hat{p} > p_0 + z_\alpha\sqrt{p_0q_0/n}$의 검정력이 $1-\beta$가 되려면 $p_0 + z_\alpha\sqrt{p_0q_0/n} = p_1 - z_\beta\sqrt{p_1q_1/n}$이다. 풀면 $n = \left(\frac{1.645\cdot 0.218 + 1.282\cdot 0.357}{0.10}\right)^2 \approx 67$. 정확한 이항 계산(77)보다 작다. 불량률이 작아 정규근사가 낙관적이다.

5절

연습 5.1 문제 20쪽

$k = \frac{\pi_0 c_{\mathrm{I}}}{\pi_1 c_{\mathrm{II}}} = \frac{0.99\cdot 1}{0.01\cdot 50} = 1.98$. 가능도비가 약 2를 넘으면 양성으로 부른다. 유병률이 낮으면 비용이 커도 임계값이 1보다 크다.

연습 5.2 문제 20쪽

경계는 $\mathbf{w}'\mathbf{x} = \frac{1}{2}\mathbf{w}'(\boldsymbol{\mu}_0+\boldsymbol{\mu}_1) + \log\frac{\pi_0}{\pi_1}$로 사전확률이 작은 쪽으로 밀린다. $\eta = \log(\pi_0/\pi_1)$이면 오분류 확률은 $\pi_0\Phi\!\left(-\frac{\Delta}{2} - \frac{\eta}{\Delta}\right) + \pi_1\Phi\!\left(-\frac{\Delta}{2} + \frac{\eta}{\Delta}\right)$이다.

연습 5.3 문제 20쪽

$\log\frac{f_1}{f_0} = -\log 2 - \frac{x^2}{8} + \frac{x^2}{2} = \frac{3x^2}{8} - \log 2 > 0$이면 무리 1이므로 $|x| > c = \sqrt{\frac{8}{3}\log 2} = 1.36$.

연습 5.4 문제 20쪽

$x > c$에서 기각하면 두 오류는 $1-\Phi(c)$와 $\Phi(c-1)$이다. 앞은 감소, 뒤는 증가하므로 최댓값은 두 값이 같은 $c = \frac{1}{2}$에서 최소(둘 다 0.309)다. $\pi_0 = \pi_1$, 같은 비용의 Bayes 검정은 $f_1/f_0 = e^{x-1/2} > 1$, 곧 $x > \frac{1}{2}$로 같다.

연습 5.5 문제 20쪽

(a) 기각($a=1$)하면 H_0이 참일 때만 비용 c_{I}가 생기므로 사후 기대손실은 $c_{\mathrm{I}}\pi(H_0\mid\mathbf{x})$이다. 받아들이면($a=0$) H_1이 참일 때만 비용 c_{II}가 생기므로 $c_{\mathrm{II}}\pi(H_1\mid\mathbf{x})$이다. (b) $\pi(H_j\mid\mathbf{x}) = \pi_j f_j(\mathbf{x})/m(\mathbf{x})$, $m = \pi_0 f_0 + \pi_1 f_1$이다. 정리 5.2로 Bayes 규칙은 작은 쪽을 고르므로 $c_{\mathrm{I}}\pi_0 f_0(\mathbf{x}) < c_{\mathrm{II}}\pi_1 f_1(\mathbf{x})$, 곧 $f_1(\mathbf{x})/f_0(\mathbf{x}) > \pi_0 c_{\mathrm{I}}/(\pi_1 c_{\mathrm{II}})$일 때 기각한다. $m(\mathbf{x}) > 0$인 공통 인수는 비교에 영향이 없다. 두 값이 같으면 어느 쪽을 골라도 사후 기대손실이 같으므로 경계에서는 확률화해도 Bayes 규칙이다. (c) 예를 들어 $c_{\mathrm{I}} = c_{\mathrm{II}} = 1$로 두고 $\pi_0/\pi_1 = k$, 곧 $\pi_0 = \frac{k}{1+k}$, $\pi_1 = \frac{1}{1+k}$로 고르면 된다. 따라서 임계값 $k>0$의 Neyman–Pearson 검정은 이 사전확률에 대한 Bayes 검정이다.

연습 5.6 문제 20쪽

(a) 행동 a의 손실은 참인 무리가 a가 아닐 때 1이므로 사후 기대손실은 $\sum_{k\neq a} P(\text{무리 } k \mid \mathbf{x}) = 1 - P(\text{무리 } a \mid \mathbf{x})$이다. (b) 베이즈 정리(1회 정리 5.5)로 $P(\text{무리 } a \mid \mathbf{x}) = \pi_a f_a(\mathbf{x}) / \sum_j \pi_j f_j(\mathbf{x})$이고 분모는 a에 무관하다. 사후 기대손실을 최소화하는 것은 $\pi_a f_a(\mathbf{x})$를 최대화하는 것이므로, 정리 5.2로 $\arg\max_k \pi_k f_k(\mathbf{x})$가 Bayes 규칙이다. (c) 0–1 손실의 위험은 $R(k,\delta) = P_k(\delta(\mathbf{X}) \neq k)$이고, Bayes 위험 $\sum_k \pi_k R(k,\delta)$는 무리를 무작위로 뽑고 분류할 때 틀릴 확률이다.

연습 5.7 문제 20쪽

(a) Σ가 대칭 양의 정치이므로 $\Sigma = Q\Lambda Q'$ (Q 직교, Λ 대각, 대각원소 양수)이다. $\Sigma^{1/2} = Q\Lambda^{1/2}Q'$는 대칭이고 $\Sigma^{1/2}\Sigma^{1/2} = Q\Lambda Q' = \Sigma$이며, $\Sigma^{-1/2} = Q\Lambda^{-1/2}Q'$가 역행렬이다. (b) $\mathbf{w}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \mathbf{w}'\Sigma^{1/2}\Sigma^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \mathbf{v}'\mathbf{d}$이고 $\mathbf{w}'\Sigma\mathbf{w} = \mathbf{w}'\Sigma^{1/2}\Sigma^{1/2}\mathbf{w} = \|\mathbf{v}\|^2$이다. (c) Cauchy–Schwarz로 $(\mathbf{v}'\mathbf{d})^2 \le \|\mathbf{v}\|^2\|\mathbf{d}\|^2$이므로 기준은 $\|\mathbf{d}\|^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \Delta^2$ 이하이다. 등호는 $\mathbf{v} = t\mathbf{d}$일 때뿐이고, 그때 $\mathbf{w} = \Sigma^{-1/2}\mathbf{v} = t\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$이다. 기준은 $\mathbf{w}$의 길이에 무관하므로 방향만 정해진다.

6절

연습 6.1 문제 23쪽

점수 $\mathbf{w}'\mathbf{x}$의 두 평균의 차는 $\mathbf{w}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \Delta^2$, 무리 안의 분산은 $\mathbf{w}'\Sigma\mathbf{w} = \Delta^2$이므로 표준화한 차는 $\Delta = 3.77$이다.

연습 6.2 문제 23쪽

판별함수는 같은 자료의 오류를 줄이도록 맞춰졌으므로, 같은 자료로 잰 오류율은 새 자료에서의 오류율보다 작게 나온다(과적합). leave-one-out은 각 송이를 그것이 빠진 자료로 만든 규칙으로 분류하므로 그 낙관을 거의 없앤다. 붓꽃처럼 모수(평균 12개, 공분산 10개)에 비해 자료가 많으면 차이가 작다.

7절

연습 7.1 문제 25쪽

차 0.001, 표준오차 0.000424이므로 구간은 $0.001 \pm 0.00083 = [0.00017,\ 0.00183]$, 상대적으로 0.2%에서 1.8%의 개선이다. 1만 명씩이면 표준오차가 0.00424로 열 배가 되어 구간 $[-0.0073,\ 0.0093]$

은 0을 넉넉히 포함한다.

연습 7.2 문제 25쪽

H_1 아래 $\bar X = \mu + (\bar X - \mu)$는 독립인 두 정규의 합이므로 $\mathcal{N}(0, \tau^2 + 1/n)$. $B_{01} = \frac{\varphi(\bar x;0,1/n)}{\varphi(\bar x;0,\tau^2+1/n)} = \sqrt{\frac{\tau^2+1/n}{1/n}}\exp\!\left(-\frac{\bar x^2}{2}\left(n - \frac{1}{\tau^2+1/n}\right)\right)$이고, $\bar x^2 = z^2/n$을 넣으면 $\sqrt{1+n\tau^2}\exp\!\left(-\frac{z^2}{2}\frac{n\tau^2}{1+n\tau^2}\right)$.

연습 7.3 문제 25쪽

첫 문장은 $P(\text{자료} \mid H_0)$를 $P(H_0 \mid \text{자료})$로 바꿔 읽었다(Lindley의 예가 보이듯 둘은 크게 다를 수 있다). 둘째 문장은 "우연"이 이미 H_0을 가정한 말이므로 결국 같은 오독이다. 정확한 문장: "H_0이 참이라면, 지금만큼 또는 이보다 더 극단적인 검정 통계량이 나올 확률이 0.03이다."

8절

연습 8.1 문제 28쪽

점수 s의 임계값 검정이 크기 α, 검정력 γ를 가지면, 크기 α의 Neyman–Pearson 검정(가능도비 임계값, 필요하면 확률화)의 검정력은 정리 1.2로 γ 이상이다. 그 검정이 가능도비 ROC 곡선의 α에서의 점이다.

연습 8.2 문제 28쪽

임계값 t에서 $\mathrm{FPR}(t) = P(S_0 > t)$, $\mathrm{TPR}(t) = P(S_1 > t)$이므로 넓이는 $\int \mathrm{TPR}\, d\mathrm{FPR} = \int P(S_1 > t) f_{S_0}(t)\, dt = P(S_1 > S_0)$ (독립).

연습 8.3 문제 28쪽

$\tau \wedge N$은 유계인 멈춤시각이므로 $EM_{\tau\wedge N} = EM_0 = 1$이다. $\{\max_{n\le N} M_n \ge c\} = \{\tau \le N\}$ 위에서 $M_{\tau\wedge N} = M_\tau \ge c$이고 다른 곳에서는 ≥ 0이므로 $1 \ge cP(\tau \le N)$. $N \to \infty$로 보내면 $P(\sup_n M_n \ge c) \le 1/c$.