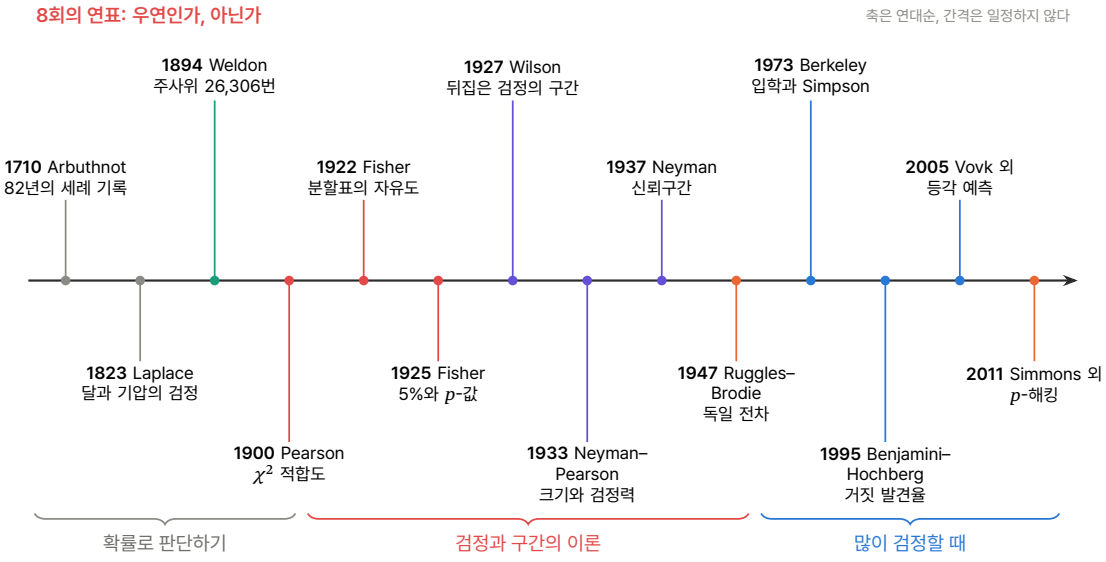
수리통계학 × 통계학의 역사

8회 강의노트

추론의 기초
신뢰구간과 유의성 검정

1710 → 1937

Arbuthnot은 1710년 런던에서 82년 내내 남아가 더 많이 세례받은 기록이 우연일 확률을 $(1/2)^{82}$로 계산했다. 최초의 유의성 검정이다. 이 노트는 순서통계량, 신뢰구간, 검정의 크기·검정력·p-값, Pearson의 χ^2를 증명하고, 원하는 결과가 나올 때까지 검정하면 무엇이 무너지는지 본다.



다루는 것	통계량 · 순서통계량과 표본분위수 · 독일 전차 문제 · 신뢰구간과 축량 · 대표본 구간 · 검정의 크기와 검정력 · 구간과 검정의 쌍대성 · p-값의 균등성 · Pearson의 χ^2 · 분할표 · 다중검정
원자료	Arbuthnot(1710)의 런던 세례 기록 · Weldon(1894)의 주사위 26,306번 · Quetelet(1846)의 가슴둘레 · UC Berkeley 대학원 입학(1973)
노트북	`sessions/08_inference/08_inference.ipynb` — 절·정리 번호가 같다
워크북	`08_inference_workbook.pdf` (손으로 푸는 35문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1–7회 (특히 6회 t-구간과 축량, 7회 중심극한정리·Slutsky·다변량 CLT)

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. 극한·적분을 바꾸거나 조건이 필요한 곳에는 **조건 확인** 단락이 있다. 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](원소 추적), [T5](극한과 하한), [T8](o와 O)을 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 1–7회에서 가져오는 것 2

0.2 표기 2

1 통계량과 순서통계량 — 독일 전차 문제 3

1.1 통계량 3

1.2 순서통계량의 분포 3

1.3 표본분위수의 점근정규성 4

1.4 독일 전차 문제 6

2 신뢰구간 — Neyman (1937) 9

2.1 정의와 축량 9

2.2 대표본 구간 10

2.3 분포를 가정하지 않는 중앙값의 구간 10

3 검정의 언어 — 크기와 검정력 13

3.1 가설과 오류 13

3.2 정규 평균의 검정과 표본 크기 14

3.3 구간과 검정은 같은 것이다 15

4 p-값 — Fisher의 5% (1925) 17

4.1 정의 17

4.2 귀무가설 아래에서 p-값은 균등하다 17

4.3 p-값이 말하지 않는 것 19

5 χ^2 검정 — Pearson (1900)과 Fisher (1922) 20

5.1 Pearson의 정리 20

5.2 모수를 추정하면 — Fisher의 자유도 21
5.3 분할표의 독립성 21

6 데이터로 — 네 가지 검정 24

6.1 Arbuthnot의 82년 (1629–1710) 24
6.2 Weldon의 주사위 (1894) 24
6.3 Quetelet의 가슴둘레 (1846) 25
6.4 UC Berkeley 대학원 입학 (1973) 26

7 부숴보기 — 원하는 결과가 나올 때까지 28

7.1 스무 번 검정하면 28
7.2 결과를 보며 멈추기 28
7.3 비율의 Wald 구간 29
7.4 기대도수가 작은 χ^2 29

8 Beautiful insights — 균등성과 순위 31

8.1 균등성으로 보정을 점검하기 31
8.2 순위와 등각 예측 31
8.3 다중검정 — Bonferroni와 거짓 발견율 32
8.4 시험 자료를 거듭 쓰면 33

9 정리와 역사 노트 35

9.1 한 장 요약 35
9.2 Stigler의 명명 법칙 35
9.3 7 Pillars 지도 36
9.4 다음 회 — 최대가능도의 긴 이야기 (1778 → 1951) 36

A 연습문제 풀이 38

0

들어가며

> **여는 질문.** 1629년부터 1710년까지 82년 동안, 런던에서는 해마다 남자아이가 여자아이보다 많이 세례를 받았다. 한 해도 빠짐없이. 이것은 우연인가, 아니면 무언가가 그렇게 만드는가? 그리고 "우연이라기엔 너무 드물다"는 말을 어떻게 숫자로 할 것인가?

John Arbuthnot은 1710년 이 기록을 왕립학회에 보고하면서 이렇게 논증했다. 남녀가 반반의 확률로 태어난다면 어느 해에 남아가 더 많을 확률은 $\frac{1}{2}$이고, 82년이 모두 그럴 확률은 $(\frac{1}{2})^{82} \approx 2\times10^{-25}$이다. 이렇게 작은 확률의 일이 일어났으니 "우연"이라는 가설은 버려야 한다. 그는 그 대신 신의 섭리를 결론지었다. 결론은 지금 보면 성급하지만, 논증의 형태는 오늘의 유의성 검정과 같다. 가설을 세우고, 그 가설 아래에서 관측의 확률을 계산하고, 그 확률이 너무 작으면 가설을 버린다.

이 형태가 엄밀한 이론이 되기까지 200년이 걸렸다. Pearson(1900)은 Weldon의 주사위로 χ^2 적합도 검정을 만들었고, Fisher(1922, 1925)는 자유도를 바로잡고 "5%"라는 관행을 세웠으며, Neyman(1937)은 신뢰구간을 정의했다. 이 노트는 그 언어를 정리한다. 자료의 함수인 **통계량**과 그 분포(1절), 모수를 덮는 **신뢰구간**(2절), 가설을 판정하는 **검정**과 그 오류(3절), p-**값**의 정체(4절), 그리고 Pearson의 χ^2 **검정**(5절)이다. 6절은 네 원자료에 검정을 쓰고, 7절은 원하는 결과가 나올 때까지 검정하면 무엇이 무너지는지, 8절은 같은 생각이 기계학습의 보정과 등각 예측(conformal prediction)에서 어떻게 쓰이는지 본다.

학습 목표

1. 순서통계량의 결합밀도를 유도하고, 표본분위수의 점근정규성을 증명한다. 독일 전차 문제의 불편추정량을 만든다.
2. 축량과 대표본 근사로 신뢰구간을 만들고, 분포에 의존하지 않는 중앙값의 구간을 증명한다.
3. 검정의 크기·검정력·표본 크기를 계산하고, 구간과 검정이 서로를 정한다는 쌍대성을 증명한다.
4. 귀무가설 아래에서 p-값이 균등분포를 따름을 증명하고, 그 해석의 한계를 말한다.
5. Pearson의 χ^2 통계량이 $\chi^2(k-1)$로 분포수렴함을 증명하고, 적합도와 독립성 검정에 쓴다.

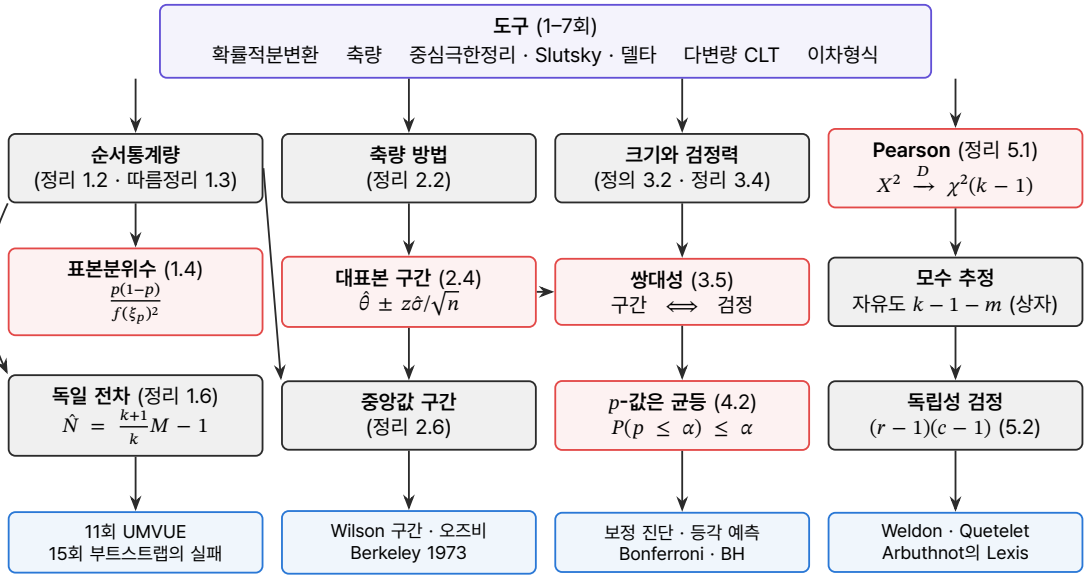


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–7회에서 가져오는 것

앞 회차	내용
2회 정리 4.3 · 4.7	단조변환의 밀도, 확률적분변환($F(X) \sim \mathrm{U}(0,1)$, $F^{-1}(U) \sim F$)
3회 정리 5.1 · 예제 3.5	야코비안 변수변환, iid의 최댓값과 최솟값
4회 정리 1.8 · 예제 4.4	다항분포의 공분산, 균등 순서통계량은 베타분포
5회 정리 3.8 · 4회 정리 3.9	다변량 정규의 선형변환, 표준정규의 제곱합은 χ^2
6회 정의 5.1 · 정리 5.2 · 8.1	축량, t-구간, 예측구간
7회 정리 3.1 · 4.2 · 4.4 · 5.2	중심극한정리, Slutsky, 델타 방법, 다변량 중심극한정리, Berry–Esseen(3.3절의 상자)

0.2 표기

자료 $\mathbf{X} = (X_1, \ldots, X_n)$의 분포는 모수 $\theta \in \Omega$로 정해지고, 확률과 기댓값을 그 모수 아래에서 계산할 때 P_θ, E_θ로 쓴다. z_p, $t_{r,p}$, $\chi^2_{r,p}$, $F_{r_1,r_2,p}$는 각 분포의 위쪽 p 분위수다. 순서통계량은 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$이다.

1

통계량과 순서통계량 — 독일 전차 문제

1.1 통계량

정의 1.1 (통계량과 추정량) 자료 $\mathbf{X}$의 함수로서 모르는 모수에 의존하지 않는 것 $T = T(\mathbf{X})$를 **통계량**이라 한다. 모수 θ(또는 그 함수)를 어림하려고 쓰는 통계량을 **추정량**, 그 관측값을 **추정값**이라 한다. $E_\theta T = \theta$가 모든 θ에서 성립하면 T는 **불편**이고, $T_n \xrightarrow{P} \theta$가 모든 θ에서 성립하면 **일치**한다.

$\bar{X}$, S^2, 표본중앙값은 통계량이다. 6회의 $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$은 μ를 포함하므로 통계량이 아니라 축량이다. 추정량의 좋고 나쁨을 재는 기준(평균제곱오차, 효율)은 9회와 11회에서 다룬다. 이 절은 그 재료인 **순서통계량**을 준비한다.

1.2 순서통계량의 분포

정리 1.2 (순서통계량의 결합밀도) $X_1, \dots, X_n$이 밀도 f를 갖는 iid이면 순서통계량의 벡터 $(X_{(1)}, \dots, X_{(n)})$은 다음 결합밀도를 갖는다.

$$g(y_1, \dots, y_n) = n! \prod_{i=1}^{n} f(y_i), \qquad y_1 < y_2 < \cdots < y_n$$

그 밖에서 $g = 0$이다.

증명. **전략** 순서통계량이 집합 $A \subset \{y_1 < \cdots < y_n\}$에 들어가는 사건을 "어느 X_i가 몇 번째로 작은가"에 따라 $n!$개의 서로소인 조각으로 나눈다. iid이므로 조각마다 확률이 같다.

① $i \neq j$이면 $P(X_i = X_j) = \int P(X_j = x) f(x)\,dx = 0$이다. 따라서 확률 1로 $X_1, \dots, X_n$은 모두 다르다. 독립, 1회 정리 6.7(b)

② 순열 π마다 $E_\pi = \{X_{\pi(1)} < \cdots < X_{\pi(n)}\}$로 두면 E_π들은 쌍마다 서로소이고, 확률 0인 사건을 빼면 합집합이 Ω이다. E_π 위에서 $(X_{(1)}, \dots, X_{(n)}) = (X_{\pi(1)}, \dots, X_{\pi(n)})$이다. ①, 원소 추적 [T1]

③ A의 점은 성분이 증가하므로 $\{(X_{\pi(1)},\dots,X_{\pi(n)})\in A\}\subset E_\pi$이고

$$P\big((X_{(1)},\dots,X_{(n)})\in A\big)=\sum_{\pi}P\big((X_{\pi(1)},\dots,X_{\pi(n)})\in A\big).$$ ②, 유한가법성

④ iid이므로 순서를 바꾼 벡터 $(X_{\pi(1)},\dots,X_{\pi(n)})$도 결합밀도 $\prod_i f(y_i)$를 갖고, 각 항은 $\int_A\prod_i f(y_i)\,d\mathbf{y}$이다. 독립, 곱은 순서와 무관

⑤ 항이 $n!$개이므로 $P\big((X_{(1)},\dots,X_{(n)})\in A\big)=\int_A n!\prod_i f(y_i)\,d\mathbf{y}$이고, 이것이 밀도의 정의다. 밀도의 정의

■

따름정리 1.3 (k번째 순서통계량) 분포함수가 F이면 $X_{(k)}$의 밀도는

$$f_{(k)}(x)=\frac{n!}{(k-1)!\,(n-k)!}F(x)^{k-1}\big(1-F(x)\big)^{n-k}f(x).$$

증명. **전략** $X_{(k)}\le x$는 "x 이하인 관측이 k개 이상"과 같으므로 분포함수가 이항분포의 꼬리다. 그것을 x로 미분하면 합의 항들이 차례로 지워지고 한 항만 남는다. 단계별 증명은 연습 1.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 39쪽에 있다.

밀도의 모양은 셈으로 읽힌다. x 근처에 하나($f(x)$), 그 아래에 $k-1$개(F^{k-1}), 위에 $n-k$개($(1-F)^{n-k}$)가 있고, 누가 어디에 있을지의 경우의 수가 다항계수다. $k=n$이 3회 예제 3.5의 최댓값이다.

1.3 표본분위수의 점근정규성

정리 1.4 (표본분위수) $0<p<1$이고 F가 분위수 $\xi_p=F^{-1}(p)$ 근처에서 연속인 밀도 f를 가지며 $f(\xi_p)>0$이면, $k_n/n\to p$이고 $k_n=np+o(\sqrt{n})$인 k_n에 대해

$$\sqrt{n}\big(X_{(k_n)}-\xi_p\big)\xrightarrow{D}\mathcal{N}\Big(0,\ \frac{p(1-p)}{f(\xi_p)^2}\Big).$$

증명. **전략** 균등분포에서 먼저 증명한다. $U_{(k)}\le u$는 "u 이하인 관측의 수"라는 이항 사건이므로 이항분포의 정규근사로 극한을 구한다. 그다음 $X_{(k)}=F^{-1}(U_{(k)})$에 델타 방법을 쓴다.

1단계: 균등분포

① $U_i \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$, x를 고정하고 $p_n = p + \frac{x}{\sqrt{n}}$, $B_n = \#\{i : U_i \le p_n\} \sim \mathrm{b}(n, p_n)$으로 두면 $\{U_{(k_n)} \le p_n\} = \{B_n \ge k_n\}$이다. k번째로 작은 값이 t 이하 $\iff$ t 이하가 k개 이상 [T1]

② $q_n = 1 - p_n$, $Z_n = \frac{B_n - np_n}{\sqrt{np_nq_n}}$, $c_n = \frac{k_n - np_n}{\sqrt{np_nq_n}}$이면 $P\big(\sqrt{n}(U_{(k_n)} - p) \le x\big) = P(Z_n \ge c_n)$이다. ①, 표준화

③ $k_n - np_n = (k_n - np) - x\sqrt{n} = -x\sqrt{n} + o(\sqrt{n})$이므로 $c_n \to -x/\sqrt{pq}$이다. $k_n = np + o(\sqrt{n})$ [T8]

④ $p_n \to p \in (0,1)$이므로 $\sup_t |P(Z_n < t) - \Phi(t)| \le C/\sqrt{n}$인 상수 C가 n에 관계없이 있다. Berry–Esseen, 7회 3.2절의 상자

⑤ ③④와 Φ의 연속성으로 $P(Z_n \ge c_n) \to 1 - \Phi(-x/\sqrt{pq}) = \Phi(x/\sqrt{pq})$, 곧 $\sqrt{n}(U_{(k_n)} - p) \xrightarrow{D} \mathcal{N}(0, pq)$이다. 분포수렴의 정의

2단계: 역함수

① F^{-1}은 비감소이므로 순서를 보존하고, $(F^{-1}(U_1), \dots, F^{-1}(U_n))$은 $(X_1, \dots, X_n)$과 분포가 같다. 따라서 $X_{(k_n)}$과 $F^{-1}(U_{(k_n)})$은 분포가 같다. 2회 정리 4.7

② f가 ξ_p 근처에서 연속이고 양수이므로 F는 그 근처에서 순증가하고 미분가능하며, F^{-1}은 p에서 미분계수 $\frac{1}{f(\xi_p)}$를 갖는다. 역함수의 미분

③ 1단계와 ②에 델타 방법을 쓰면 $\sqrt{n}\big(F^{-1}(U_{(k_n)}) - \xi_p\big) \xrightarrow{D} \mathcal{N}\big(0, \frac{pq}{f(\xi_p)^2}\big)$이다. 7회 정리 4.4

조건 확인 Berry–Esseen 한계의 상수는 $E|\xi - p_n|^3/(p_nq_n)^{3/2}$에 비례한다($\xi \sim$ 베르누이(p_n)). $p_nq_n \to pq > 0$이므로 이 양은 n에 대해 유계이고, ④의 고른 한계가 정당하다. 2단계의 델타 방법은 F^{-1}이 p에서 미분가능하기만 하면 된다. ■

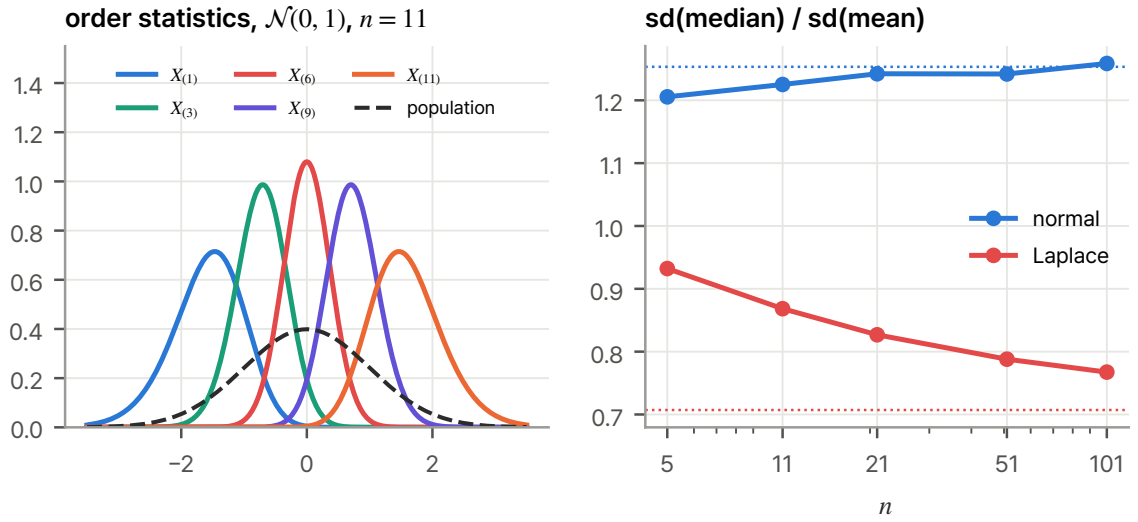


그림 1.1 왼쪽: 정규 표본 $n = 11$의 순서통계량 $X_{(1)}, X_{(3)}, X_{(6)}, X_{(9)}, X_{(11)}$의 밀도(따름정리 1.3). 오른쪽: 표본중앙값과 표본평균의 표준편차의 비 $\times\sqrt{n}$ — 정규(파랑)에서는 $\sqrt{\pi/2} = 1.25$로 평균이 낫고, Laplace(빨강)에서는 $1/\sqrt{2} = 0.71$로 중앙값이 낫다.

예제 1.5 (표본중앙값) $p = \frac{1}{2}$이면 표본중앙값의 점근분산은 $\frac{1}{4nf(m)^2}$이다(7회 연습 7.3에서 받아들였던 식). 정규 $\mathcal{N}(\mu, \sigma^2)$이면 $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$이므로 $\frac{\pi\sigma^2}{2n}$, 평균의 $\frac{\sigma^2}{n}$보다 1.57배 크다. Laplace 분포(밀도 $\frac{1}{2b}e^{-|x-\mu|/b}$, 분산 $2b^2$)이면 $\frac{b^2}{n}$으로 평균의 $\frac{2b^2}{n}$보다 **작다**. 꼬리가 두꺼우면 중앙값이 평균을 이긴다. 5회 1절에서 Laplace 오차의 가능도가 중앙값에서 최대였던 것과 같은 이야기다.

1.4 독일 전차 문제

> **역사** 2차 대전 중 연합군은 노획한 독일 전차의 바퀴와 변속기에 찍힌 일련번호로 독일의 월 생산량을 추정했다. Ruggles와 Brodie(1947)의 보고에 따르면 1940년 6월의 생산량을 정보기관은 1,000대, 통계적 추정은 169대로 보았고, 전후에 확인한 독일 기록은 122대였다. 1941년 6월에는 정보기관 1,550대, 통계 244대, 기록 271대였다. 첩보보다 일련번호의 최댓값이 훨씬 정확했다.

전차에 $1, 2, \ldots, N$의 일련번호가 있고, 그 가운데 k대를 되돌려 넣지 않고 무작위로 노획했다고 하자. N을 어떻게 추정할까? 노획한 번호의 최댓값 M은 N 이하이므로 N을 작게 어림한다. 얼마나 작은지는 계산할 수 있다.

정리 1.6 (최댓값의 기댓값과 불편추정량) $\{1,\dots,N\}$에서 k개를 되돌려 넣지 않고 뽑은 최댓값 M에 대해

$$P(M=m)=\frac{\binom{m-1}{k-1}}{\binom{N}{k}}\quad (m=k,\dots,N),\qquad EM=\frac{k(N+1)}{k+1}.$$

따라서 $\hat N=\dfrac{k+1}{k}M-1$은 N의 불편추정량이다.

증명. **전략** 최댓값이 m인 표본을 세어 분포를 얻는다. 기댓값의 합은 $m\binom{m-1}{k-1}=k\binom{m}{k}$로 바꿔 하키스틱 항등식의 꼴로 만든다.

① $\binom{N}{k}$개의 표본이 같은 확률이다. 최댓값이 m이려면 m을 뽑고 나머지 $k-1$개를 $\{1,\dots,m-1\}$에서 뽑아야 하므로 $P(M=m)=\binom{m-1}{k-1}/\binom{N}{k}$이다. 1회 따름정리 4.10, 정리 1.4

② $m\binom{m-1}{k-1}=\dfrac{m!}{(k-1)!\,(m-k)!}=k\binom{m}{k}$이다. 계승으로 정리

③ 따라서

$$EM=\frac{1}{\binom{N}{k}}\sum_{m=k}^{N}m\binom{m-1}{k-1}=\frac{k}{\binom{N}{k}}\sum_{m=k}^{N}\binom{m}{k}\qquad ①②$$

$$=\frac{k\binom{N+1}{k+1}}{\binom{N}{k}}=\frac{k(N+1)}{k+1}.\qquad \text{하키스틱, 연습 1.3}$$

④ $E\hat N=\dfrac{k+1}{k}EM-1=N$이다. 기댓값의 선형성

■

$\hat N=M+\dfrac{M}{k}-1$은 "관측한 최댓값에 번호 사이의 평균 간격만큼 더한다"로 읽힌다. 같은 문제를 적률법으로 풀면 $E\bar X=\dfrac{N+1}{2}$에서 $2\bar X-1$이 나온다. 둘 다 불편이지만 분산이 다르다. $\operatorname{Var}\hat N=\dfrac{(N+1)(N-k)}{k(k+2)}$이고 $\operatorname{Var}(2\bar X-1)=\dfrac{(N+1)(N-k)}{3k}$이므로(연습 1.4) 최댓값 쪽이 $\dfrac{k+2}{3}$배 정확하다. $N=300$, $k=5$이면 표준편차가 50과 77이다. 11회에서 $\hat N$이 모든 불편추정량 가운데 분산이 가장 작음(UMVUE)을 증명하고, 15회에서는 부트스트랩이 바로 이 최댓값에서 실패함을 본다.

연습문제

연습 1.1 (최솟값과 범위 ★) $\mathrm{U}(0,1)$ 표본에서 $X_{(1)}$과 $X_{(n)}$의 결합밀도를 정리 1.2를 적분해 구하고, 범위 $R=X_{(n)}-X_{(1)}$이 $\mathrm{Beta}(n-1,2)$임을 보여라. 풀이 38쪽

연습 1.2 (지수 표본의 간격 ★★) $\mathrm{Exp}(1)$ 표본의 간격 $D_i=X_{(i)}-X_{(i-1)}$ $(X_{(0)}=0)$에 대해 $(n-i+1)D_i$가 iid $\mathrm{Exp}(1)$임을 보여라. (힌트: 결합밀도 $n!\,e^{-\sum y_i}$에서 $\sum y_i=\sum_i (n-i+1)d_i$.) 따라서 $EX_{(k)}=\sum_{i=1}^{k}\frac{1}{n-i+1}$. 풀이 38쪽

연습 1.3 (하키스틱 항등식 ★) $\sum_{m=k}^{N}\binom{m}{k}=\binom{N+1}{k+1}$을 "$\{1,\dots,N+1\}$에서 $k+1$개를 고를 때 가장 큰 수가 $m+1$"로 나누어 세어 보여라. 풀이 38쪽

연습 1.4 (전차 추정량의 분산 ★★★) $E[M(M+1)]=\frac{k(N+1)(N+2)}{k+2}$을 정리 1.6의 증명과 같은 방법($m(m+1)\binom{m-1}{k-1}=k(k+1)\binom{m+1}{k+1}$)으로 보이고, $\operatorname{Var}\hat{N}=\frac{(N+1)(N-k)}{k(k+2)}$을 유도하라. 4회 정리 1.5(초기하의 분산, 유한 모집단 수정)로 $\operatorname{Var}(2\bar{X}-1)=\frac{(N+1)(N-k)}{3k}$도 보여라. 풀이 38쪽

연습 1.5 (Laplace에서 중앙값 ★★) 정리 1.4로 Laplace 분포에서 표본중앙값의 점근분산이 $\frac{b^2}{n}$임을 보이고, 평균과의 효율비를 구하라. $t(3)$이면? 풀이 39쪽

연습 1.6 (k번째 순서통계량의 밀도 ★★ 유도형) 따름정리 1.3를 다음 순서로 증명하라.

(a) $\{X_{(k)}\le x\}$가 "x 이하인 X_i가 k개 이상"과 같음을 보이고, $P(X_{(k)}\le x)=\sum_{j=k}^{n}\binom{n}{j}F(x)^j\big(1-F(x)\big)^{n-j}$을 얻어라.

(b) $j\binom{n}{j}=n\binom{n-1}{j-1}$, $(n-j)\binom{n}{j}=n\binom{n-1}{j}$를 써서 (a)를 x로 미분한 식이 다음과 같음을 보여라.

$$nf(x)\sum_{j=k}^{n}\Big[\binom{n-1}{j-1}F^{j-1}(1-F)^{n-j}-\binom{n-1}{j}F^{j}(1-F)^{n-1-j}\Big]$$

(c) (b)의 합에서 이웃한 항이 지워져 $j=k$의 첫 항만 남음을 보이고 따름정리의 식을 얻어라.

(d) 다른 길로도 같은 식이 나온다. $U_{(k)}\sim\mathrm{Beta}(k,n-k+1)$(4회 예제 4.4)과 $X_{(k)}=F^{-1}(U_{(k)})$에 단조변환의 밀도(2회 정리 4.3)를 쓰면 같은 식이 나옴을 확인하라.

풀이 39쪽

2

신뢰구간 — Neyman (1937)

> **역사** Jerzy Neyman은 1937년 *Philosophical Transactions*에 실은 논문에서 "신뢰구간"을 정의했다. 핵심은 확률을 **구간을 만드는 방법**에 붙이는 것이었다. 모수는 고정된 미지수이고 구간이 무작위다. 같은 시기 Fisher는 "기준 확률"(fiducial probability, 1930)이라는 다른 해석으로 비슷한 구간을 만들었고, 두 사람은 해석을 두고 오래 다퉜다. 오늘의 교과서는 Neyman의 해석을 따른다.

2.1 정의와 축량

정의 2.1 (신뢰구간) 통계량 $L(\mathbf{X}) \le U(\mathbf{X})$가 모든 $\theta \in \Omega$에서 $P_\theta\big(L(\mathbf{X}) \le \theta \le U(\mathbf{X})\big) \ge 1-\alpha$를 만족하면 $[L, U]$를 θ의 $100(1-\alpha)\%$ **신뢰구간**이라 하고, 왼쪽 확률의 하한 $\inf_\theta P_\theta(\cdot)$을 **신뢰계수**라 한다. 모든 θ에서 등호이면 **정확한** 구간이다.

6회에서 정규 평균의 t-구간과 정규 분산의 χ^2-구간을 축량(6회 정의 5.1)으로 만들었다. 방법은 일반적이다.

정리 2.2 (축량 방법) $Q(\mathbf{X}, \theta)$의 분포가 θ와 무관한 연속분포이고, 그 분위수 $a < b$가 $P(a \le Q \le b) = 1-\alpha$를 만족하며, 각 $\mathbf{x}$에서 $Q(\mathbf{x}, \cdot)$이 θ에 대해 단조이면 $\{\theta : a \le Q(\mathbf{X}, \theta) \le b\}$는 정확한 $100(1-\alpha)\%$ 신뢰구간이다.

증명. **전략** 포함 사건을 축량의 사건으로 바꾸면 확률이 θ와 무관해진다. 단조성은 그 집합이 구간임을 보장한다. 단계별 증명은 연습 2.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

예제 2.3 (지수 수명의 평균) $X_i \overset{\text{iid}}{\sim} \text{Exp}$(평균 θ)이면 $\frac{2}{\theta}\sum X_i \sim \chi^2(2n)$이다(4회 정리 3.7: $\frac{1}{\theta}\sum X_i \sim \Gamma(n,1)$이고 $2\Gamma(n,1) = \Gamma(n,2) = \chi^2(2n)$). 이것을 축량으로 쓰면

$$\left[\frac{2\sum X_i}{\chi^2_{2n,\alpha/2}},\ \frac{2\sum X_i}{\chi^2_{2n,1-\alpha/2}}\right]$$

이 정확한 구간이다. 전구 10개의 평균 수명이 $\bar{x} = 820$시간이면 $\chi^2_{20,0.025} = 34.17$, $\chi^2_{20,0.975} = 9.59$이므로 $[480,\ 1710]$시간이다. 비대칭인 것은 χ^2가 비대칭이기 때문이다.

2.2 대표본 구간

정확한 축량은 드물다. 대부분의 구간은 7회의 극한 정리에서 나온다.

정리 2.4 (대표본 신뢰구간) $\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D}\mathcal{N}(0,\sigma^2(\theta))$이고 $\hat\sigma_n\xrightarrow{P}\sigma(\theta)>0$이면, $\hat\theta_n\pm z_{\alpha/2}\hat\sigma_n/\sqrt{n}$의 포함확률은 $1-\alpha$로 간다.

증명. **전략** 표준오차를 추정값으로 바꾼 통계량을 "참 표준오차로 표준화한 것"과 "1로 가는 비"의 곱으로 쓰고 Slutsky를 쓴다. 단계별 증명은 연습 2.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

이 형태를 **Wald 구간**이라 한다. 비율이면 $\hat p\pm z\sqrt{\hat p(1-\hat p)/n}$이고, 두 집단의 비율의 차이면 분산을 더한 $\hat p_1-\hat p_2\pm z\sqrt{\hat p_1\hat q_1/n_1+\hat p_2\hat q_2/n_2}$이다. 비율의 Wald 구간은 p가 0이나 1에 가까우면 크게 틀린다(7절). 7회의 델타 방법이 척도를 고르는 데 쓰인다.

예제 2.5 (오즈비의 구간) 두 집단에서 성공·실패가 (a,b)와 (c,d)이면 오즈비는 $\widehat{\mathrm{OR}}=\frac{a/b}{c/d}=\frac{ad}{bc}$이다. 7회 예제 4.5로 각 집단의 로그 오즈의 분산은 대략 $\frac{1}{n\hat p\hat q}=\frac{1}{a}+\frac{1}{b}$이고, 두 집단이 독립이므로

$$\log\widehat{\mathrm{OR}}\pm z_{\alpha/2}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

가 로그 오즈비의 대표본 구간이다. 끝점에 지수를 취하면 오즈비의 구간이 된다. 1973년 UC Berkeley 대학원에서 남자 지원자 2,691명 중 1,198명, 여자 1,835명 중 557명이 합격했다. 오즈비는 1.84, 95% 구간은 $[1.62,\ 2.09]$이다. 1을 크게 벗어나지만, 6절에서 이 숫자가 무엇을 말하지 **않는지** 본다.

2.3 분포를 가정하지 않는 중앙값의 구간

정리 2.6 (순서통계량으로 만든 중앙값의 구간) $X_1,\dots,X_n$이 iid이고 분포함수가 연속이며 중앙값이 m이면, $i<j$에 대해

$$P\big(X_{(i)}<m<X_{(j)}\big)=\sum_{k=i}^{j-1}\binom{n}{k}2^{-n}.$$

증명. **전략** m 아래에 있는 관측의 수 K는 이항분포를 따른다. 순서통계량과 m의 대소를 K의 사건으로 바꾼다. 단계별 증명은 연습 2.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

분포에 대해 연속이라는 것 말고는 아무것도 가정하지 않았다. $n=20$이면 $[X_{(6)},X_{(15)}]$의 신뢰계수가 0.959, $[X_{(5)},X_{(16)}]$이 0.988이다. 이산분포의 계단 때문에 정확히 95%는 고를 수 없다. 15

회의 순위 방법의 첫 예다.

연습문제

연습 2.1 (포함확률의 해석 ★) "95% 신뢰구간 $[480, 1710]$에 참 평균 수명이 들어 있을 확률이 95%다" 라는 문장이 Neyman의 해석으로 왜 틀렸는지 설명하고, 맞는 문장으로 고쳐라. 풀이 39쪽

연습 2.2 (균등분포의 상한 ★★) $X_i \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$이면 M/θ $(M = X_{(n)})$가 축량임을 보이고, θ의 95% 구간 $[M,\ M\alpha^{-1/n}]$을 만들어라. 왜 한쪽으로만 넓히는가? 풀이 40쪽

연습 2.3 (두 비율의 차이 ★) Berkeley 자료에서 남녀 합격률의 차이 $0.445 - 0.304$의 95% Wald 구간을 구하라. 풀이 40쪽

연습 2.4 (중앙값 구간 ★★) $n = 20$에서 신뢰계수가 0.95 이상인 가장 좁은 대칭 구간 $[X_{(i)}, X_{(21-i)}]$를 찾아라. n이 크면 $i \approx \frac{n}{2} - z_{\alpha/2}\frac{\sqrt{n}}{2}$임을 정규근사로 보여라. 풀이 40쪽

연습 2.5 (축량 방법의 증명 ★ 유도형) 정리 2.2를 증명하라. $C(\mathbf{X}) = \{\theta : a \le Q(\mathbf{X},\theta) \le b\}$이다.

(a) $\{\theta \in C(\mathbf{X})\} = \{a \le Q(\mathbf{X},\theta) \le b\}$를 원소 추적으로 확인하라. 왼쪽은 $\mathbf{X}$에 대한 사건이다.

(b) $P_\theta(\theta \in C(\mathbf{X})) = 1-\alpha$가 모든 θ에서 성립함을 보여라. "Q의 분포가 θ와 무관하다" 를 어디서 쓰는가?

(c) $Q(\mathbf{x},\cdot)$이 증가하면 $\theta_1 < \theta < \theta_2$이고 $\theta_1, \theta_2 \in C(\mathbf{x})$일 때 $\theta \in C(\mathbf{x})$임을 보여라. 따라서

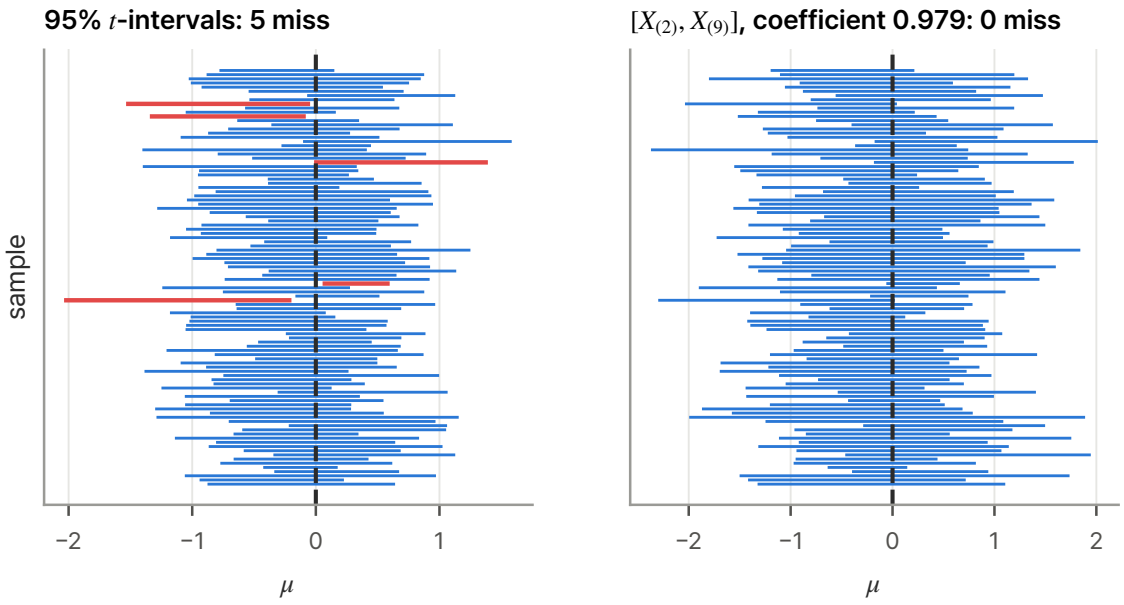


그림 2.1 신뢰구간 100개. 정규 표본 $n = 10$에서 만든 95% t-구간(왼쪽)과 같은 자료의 순서통계량 구간 $[X_{(2)}, X_{(9)}]$(오른쪽, 신뢰계수 0.979). 빨강은 참값 $\mu = 0$(점선)을 놓친 구간이다. 확률은 이 **막대들의 모임**에 대한 말이다.

$C(\mathbf{x})$는 구간이다.

풀이 40쪽

연습 2.6 (대표본 신뢰구간의 증명 ★★ 유도형) 정리 2.4를 증명하라.

(a) $T_n = \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}_n}$을 $\frac{\sigma}{\hat{\sigma}_n}$와 $\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma}$의 곱으로 써라. ($\hat{\sigma}_n = 0$일 확률은 0으로 간다.)

(b) $\frac{\sigma}{\hat{\sigma}_n} \xrightarrow{P} 1$을 보여라. $\sigma > 0$을 어디서 쓰는가?

(c) Slutsky(7회 정리 4.2)로 $T_n \xrightarrow{D} \mathcal{N}(0,1)$을 얻어라.

(d) 포함 사건이 $\{|T_n| \le z_{\alpha/2}\}$임을 보이고, $\pm z_{\alpha/2}$가 Φ의 연속점이므로 그 확률이 $1-\alpha$로 감을 보여라.

풀이 40쪽

연습 2.7 (중앙값 구간의 증명 ★ 유도형) 정리 2.6를 증명하라. $K = \#\{l : X_l < m\}$이다.

(a) $P(X_l < m) = \frac{1}{2}$이고 $P(X_l = m) = 0$임을 보이고 $K \sim \mathrm{b}(n, \frac{1}{2})$을 얻어라. 분포함수의 연속성을 어디서 쓰는가?

(b) $\{X_{(i)} < m\} = \{K \ge i\}$와 $\{m < X_{(j)}\} = \{K \le j-1\}$을 보여라. (확률 0인 사건은 무시한다.)

(c) (a)와 (b)로 정리의 식을 얻어라.

풀이 40쪽

3

검정의 언어 — 크기와 검정력

> **Arbuthnot (1710)** Anne 여왕의 주치의이자 풍자 작가였던 John Arbuthnot은 *An argument for divine providence, taken from the constant regularity observed in the births of both sexes*에서 런던의 82년치 세례 기록을 표로 실었다. 남아가 많은 해를 동전의 앞면으로 보면, 공평한 동전 82번이 모두 앞면일 확률은 1/4 8360 0000 0000 0000 0000 0000이다. 그는 남아가 더 많이 태어나는 것은 남자가 전쟁과 위험한 일로 더 많이 죽기 때문에 결혼할 수 있는 남녀의 수를 맞추려는 섭리라고 결론지었다. 오늘의 말로 그것은 "부호 검정"이다. 귀무가설(남아 확률 $\frac{1}{2}$)을 기각한 것은 옳았다. 기각한 다음 특정한 대립가설(섭리)을 **받아들인** 것은 검정이 말해 주지 않는 일이었다.

3.1 가설과 오류

정의 3.1 (가설, 검정, 기각역) 모수공간을 $\Omega = \Omega_0 \cup \Omega_1$ (서로소)로 나누어 $H_0 : \theta \in \Omega_0$ (**귀무가설**)와 $H_1 : \theta \in \Omega_1$ (**대립가설**)을 세운다. Ω_0가 한 점이면 단순가설, 아니면 복합가설이다. **검정**은 표본공간의 부분집합 C (**기각역**)를 정해 $\mathbf{X} \in C$이면 H_0를 기각하는 규칙이다. 보통 검정통계량 $T(\mathbf{X})$와 기준값 c로 $C = \{T(\mathbf{X}) \geq c\}$의 꼴이다.

정의 3.2 (오류, 크기, 검정력) H_0가 참인데 기각하는 것을 **1종 오류**, H_1이 참인데 기각하지 못하는 것을 **2종 오류**라 한다. **검정력 함수**는 $\gamma(\theta) = P_\theta(\mathbf{X} \in C)$이고, **크기**는 $\alpha = \sup_{\theta \in \Omega_0} \gamma(\theta)$, 곧 1종 오류의 확률의 최댓값이다. $\theta \in \Omega_1$에서 $\gamma(\theta)$를 **검정력**이라 한다.

두 오류는 서로 당긴다. 기각역을 넓히면 1종 오류가 늘고 2종 오류가 준다. Neyman과 Pearson(1933)의 처방은 비대칭이다. 크기를 미리 정한 α 이하로 묶고, 그 안에서 검정력을 가장 크게 하는 기각역을 찾는다. 그 최적 기각역은 12회의 주제다. 이 절은 주어진 검정의 크기와 검정력을 계산한다.

예제 3.3 (이항 검정의 크기와 검정력) 동전이 앞면 쪽으로 치우쳤는지 100번 던져 검정한다. $H_0 : p = \frac{1}{2}$, $H_1 : p > \frac{1}{2}$, 기각역 $\{K \geq 59\}$이면 크기는 $P_{1/2}(K \geq 59) = 0.044$이다. 검정력은

참 p	0.55	0.60	0.65	0.70
$\gamma(p) = P_p(K \geq 59)$	0.24	0.62	0.91	0.99

$p = 0.55$처럼 작은 치우침은 100번으로는 넷에 하나꼴로만 잡는다. "기각하지 못했다"는

> "H_0가 참이다"가 아니다. 검정력이 낮았을 수도 있다.

3.2 정규 평균의 검정과 표본 크기

정리 3.4 (z 검정의 검정력) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (σ 알려짐)에서 $H_0 : \mu \le \mu_0$를 $\bar X \ge \mu_0 + z_\alpha \sigma/\sqrt{n}$일 때 기각하면 크기는 α이고 검정력 함수는

$$\gamma(\mu) = \Phi\!\left(\frac{\sqrt{n}(\mu-\mu_0)}{\sigma} - z_\alpha\right).$$

검정력 $1-\beta$로 차이 $\delta = \mu - \mu_0 > 0$을 잡으려면 $n \ge \left(\frac{(z_\alpha + z_\beta)\sigma}{\delta}\right)^2$이어야 한다.

증명. **전략** 기각 사건을 참 평균 μ로 표준화한 $Z = \frac{\sqrt{n}(\bar X - \mu)}{\sigma} \sim \mathcal{N}(0,1)$의 사건으로 바꾼다. 검정력 함수가 μ에 대해 증가하므로 크기는 경계 μ_0에서 정해진다. 단계별 증명은 연습 3.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 41쪽에 있다.

양쪽 검정이면 z_α 대신 $z_{\alpha/2}$를 쓴다. $\alpha = 0.05$ (양쪽), 검정력 0.8, $\delta = 0.5\sigma$이면 $n \ge \left(\frac{1.96+0.84}{0.5}\right)^2 = 31.4$, 곧 32명이다. 효과가 절반이면 네 배, 128명이 필요하다. 실험을 설계할 때 가장 먼저 하는 계산이다(14회). σ를 모르면 6회의 t 검정을 쓰고, 큰 n에서는 같은 식이 근사적으로 맞는다(7회 따름정리 4.3).

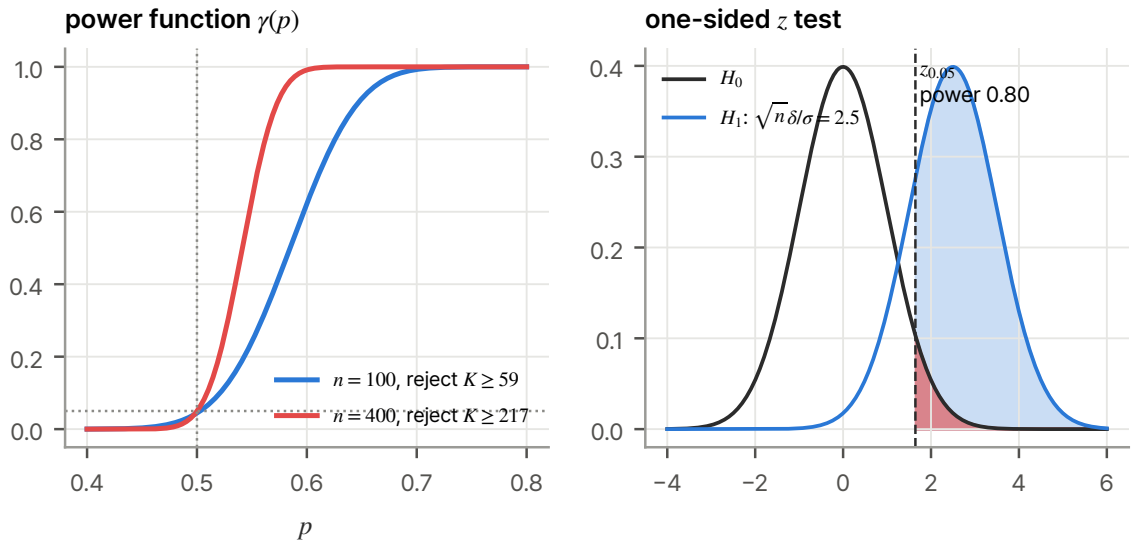


그림 3.1 왼쪽: 예제 3.3의 검정력 함수($n = 100$, 기각역 $K \ge 59$)와 $n = 400$일 때(기각역 $K \ge 217$). 오른쪽: 정규 평균의 한쪽 z 검정에서 H_0와 H_1 아래 검정통계량의 분포. 빨간 넓이가 α, 파란 넓이가 검정력이다.

3.3 구간과 검정은 같은 것이다

정리 3.5 (쌍대성)

(a) 모든 θ_0에 대해 $H_0 : \theta = \theta_0$의 크기 α 검정의 수락역 $A(\theta_0)$가 있으면, $C(\mathbf{X}) = \{\theta_0 : \mathbf{X} \in A(\theta_0)\}$는 $100(1-\alpha)\%$ 신뢰집합이다.

(b) 거꾸로 $100(1-\alpha)\%$ 신뢰집합 $C(\mathbf{X})$가 있으면 "$\theta_0 \notin C(\mathbf{X})$이면 기각"은 $H_0 : \theta = \theta_0$의 크기 α 검정이다.

증명. **전략** 두 방향 모두 사건의 등식 $\{\theta_0 \in C(\mathbf{X})\} = \{\mathbf{X} \in A(\theta_0)\}$ 하나로 귀결된다. 이 등식은 C와 A를 서로의 정의로 삼은 데서 나온다.

(a)

① $C(\mathbf{x}) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\}$이므로 $\theta_0 \in C(\mathbf{x}) \iff \mathbf{x} \in A(\theta_0)$이다. — C의 정의 [T1]

② 크기가 α이므로 $P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) \le \alpha$이다. — 정의 3.2

③ 따라서 모든 θ_0에서 $P_{\theta_0}(\theta_0 \in C(\mathbf{X})) = P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \ge 1-\alpha$이고, C는 $100(1-\alpha)\%$ 신뢰집합이다. — ①②, 정의 2.1

(b)

① $A(\theta_0) = \{\mathbf{x} : \theta_0 \in C(\mathbf{x})\}$로 두면 검정 "$\theta_0 \notin C(\mathbf{X})$이면 기각"의 수락역이 $A(\theta_0)$이다. — A의 정의 [T1]

② $P_{\theta_0}(\text{기각}) = 1 - P_{\theta_0}(\theta_0 \in C(\mathbf{X})) \le \alpha$이므로 크기가 α 이하다. — 정의 2.1, 여사건

■

t-구간이 0을 포함하지 않는 것과 t 검정이 $\mu = 0$을 기각하는 것은 같은 문장이다. 쌍대성은 새로운 구간을 만드는 방법이기도 하다.

예제 3.6 (Wilson 구간 (1927)) $H_0 : p = p_0$를 $\frac{|\hat{p}-p_0|}{\sqrt{p_0(1-p_0)/n}} > z$일 때 기각하는 검정(분산을 **귀무가설의** p_0로 계산한다)을 뒤집는다. 수락역 $(\hat{p}-p_0)^2 \le z^2 p_0(1-p_0)/n$은 p_0에 대한 이차부등식이고, 풀면

$$\frac{\hat{p} + \frac{z^2}{2n}}{1+\frac{z^2}{n}} \pm \frac{z}{1+\frac{z^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}$$

이다(연습 3.4). Wald 구간과 달리 중심이 $\frac{1}{2}$ 쪽으로 당겨지고, $\hat{p} = 0$이어도 폭이 0이 아니다. 7절에서 두 구간의 포함확률을 비교한다.

연습문제

연습 3.1 (Arbuthnot의 숫자 ★) $2^{82} = 4.836 \times 10^{24}$을 확인하라. Arbuthnot의 "부호 검정"에서 귀무가설과 검정통계량은 무엇이고, 크기 0.05의 기각역은? (힌트: K = 남아가 많은 해의 수.) 풀이 41쪽

연습 3.2 (검정력 함수 ★) 예제 3.3의 검정력 함수가 p에 대해 증가함을 보이고(이항분포의 확률적 순서: $p < p'$이면 $\mathrm{b}(n,p')$를 $\mathrm{b}(n,p)$ 위에 "쌓아" 만들 수 있다), $H_0 : p \le \frac{1}{2}$로 바꿔도 크기가 같음을 설명하라. 풀이 41쪽

연습 3.3 (표본 크기 ★) 두 집단의 평균 차이를 양쪽 5%, 검정력 90%로 $\delta = 0.3\sigma$까지 잡으려면 집단당 몇 명이 필요한가? (두 표본이면 분산이 $2\sigma^2/n$이다.) 풀이 41쪽

연습 3.4 (Wilson 구간 ★★) 예제 3.6의 이차부등식을 풀어 Wilson 구간을 유도하라. $n = 20$, $\hat{p} = 0$이면 Wald 구간과 Wilson 구간은 각각 무엇인가? 풀이 41쪽

연습 3.5 (한쪽 구간과 한쪽 검정 ★★) 정리 3.4의 검정을 모든 μ_0에 대해 뒤집으면 어떤 한쪽 신뢰구간이 나오는가? 풀이 41쪽

연습 3.6 (z 검정의 검정력 증명 ★ 유도형) 정리 3.4를 증명하라.

(a) 기각 사건 $\{\bar{X} \ge \mu_0 + z_\alpha \sigma/\sqrt{n}\}$을 $Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$의 사건으로 바꾸고 $\gamma(\mu) = \Phi\!\left(\frac{\sqrt{n}(\mu-\mu_0)}{\sigma} - z_\alpha\right)$을 얻어라. $Z \sim \mathcal{N}(0,1)$인 이유는?

(b) γ가 μ에 대해 증가함을 보이고, $\Omega_0 = \{\mu \le \mu_0\}$ 위의 최댓값 $\gamma(\mu_0) = \alpha$에서 크기를 얻어라.

(c) $\gamma(\mu_0 + \delta) \ge 1 - \beta$를 n에 대해 풀어 표본 크기 식을 얻어라. $\Phi(z_\beta) = 1-\beta$에 주의한다.

풀이 41쪽

4

p-값 — Fisher의 5% (1925)

> **역사** Fisher는 1925년 *Statistical Methods for Research Workers*에 이렇게 썼다. "$P = 0.05$, 곧 20분의 1에 해당하는 값은 1.96, 거의 2다. 편차가 유의한지 판단하는 기준으로 이 점을 잡는 것이 편리하다." 이 한 문장이 이후 한 세기 동안 과학 논문의 관행이 되었다. Fisher에게 p-값은 증거의 강도를 재는 연속적인 척도였고, Neyman–Pearson에게 검정은 미리 정한 α로 기각 여부를 결정하는 규칙이었다. 오늘의 관행은 둘을 섞어 쓴다.

4.1 정의

정의 4.1 (p-값) T가 클수록 H_0에 불리한 검정통계량이고 관측값이 t이면

$$p(t) = \sup_{\theta \in \Omega_0} P_\theta\big(T(\mathbf{X}) \geq t\big)$$

를 **p-값**이라 한다. "H_0 아래에서 관측한 것만큼, 또는 그보다 더 극단적인 결과가 나올 확률"이다.

크기 α의 검정 $\{T \geq c_\alpha\}$들이 있으면 $p(t) \leq \alpha \iff t \geq c_\alpha$이다. 곧 p-값은 **그 자료로 H_0를 기각하게 되는 가장 작은 α**다. α를 미리 정하지 않고 증거의 크기를 보고하는 방법이다.

4.2 귀무가설 아래에서 p-값은 균등하다

정리 4.2 (p-값의 분포) H_0가 단순가설($\Omega_0 = \{\theta_0\}$)이고 T의 H_0 아래 분포함수 F가 연속이면 $p(T) \sim \mathrm{U}(0,1)$이다. T가 이산이면 모든 α에서 $P_{\theta_0}\big(p(T) \leq \alpha\big) \leq \alpha$이다.

증명. **전략** p-값을 $G(t) = P_{\theta_0}(T \geq t)$로 쓰고 $G(T)$의 분포를 본다. 연속이면 확률적분변환이 그대로 답이다. 일반적으로는 $\{G(T) \leq \alpha\}$가 T의 위쪽 꼬리 사건이라는 것을 쓴다.

(a) F가 연속일 때

① $p(t) = G(t) = 1 - P(T < t) = 1 - F(t-) = 1 - F(t)$이다. — 1회 정리 6.3(d), F 연속

② $F(T) \sim \mathrm{U}(0,1)$이다. — 확률적분변환, 2회 정리 4.7

③ $U \sim \mathrm{U}(0,1)$이면 $P(1-U \le u) = P(U \ge 1-u) = u$이므로 $p(T) = 1 - F(T) \sim \mathrm{U}(0,1)$이다. ①②

(b) 일반적인 T

① G는 비증가이므로 $S_\alpha = \{t : G(t) \le \alpha\}$는 $t_\alpha = \inf S_\alpha$에 대해 $[t_\alpha, \infty)$ 또는 (t_α, ∞)이다. 따라서 $\{G(T) \le \alpha\} = \{T \in S_\alpha\}$는 $\{T \ge t_\alpha\}$ 또는 $\{T > t_\alpha\}$이다. 단조성, 하한 [T5]

② $S_\alpha = [t_\alpha, \infty)$이면 $P(G(T) \le \alpha) = P(T \ge t_\alpha) = G(t_\alpha) \le \alpha$이다. $t_\alpha \in S_\alpha$

③ $S_\alpha = (t_\alpha, \infty)$이면 $\{T \ge t_\alpha + \frac{1}{n}\} \uparrow \{T > t_\alpha\}$이므로 $P(T > t_\alpha) = \lim_n G(t_\alpha + \frac{1}{n}) \le \alpha$이다. 1회 정리 4.7(a)

(b)의 증명은 T가 이산이라는 것을 쓰지 않는다. 어느 분포에서나 $P(p(T) \le \alpha) \le \alpha$이고, 연속일 때만 등호가 모든 α에서 성립한다. ■

정리의 내용은 단순하지만 결과는 크다. 귀무가설이 참인 실험을 많이 하면 그 p-값들은 $[0,1]$에 **고르게** 흩어진다. $p < 0.05$는 그 가운데 5%에서 우연히 일어난다(7절). 그리고 H_1이 참이면 p-값은 0 쪽으로 몰린다. 몰리는 정도가 검정력이다(그림 4.1).

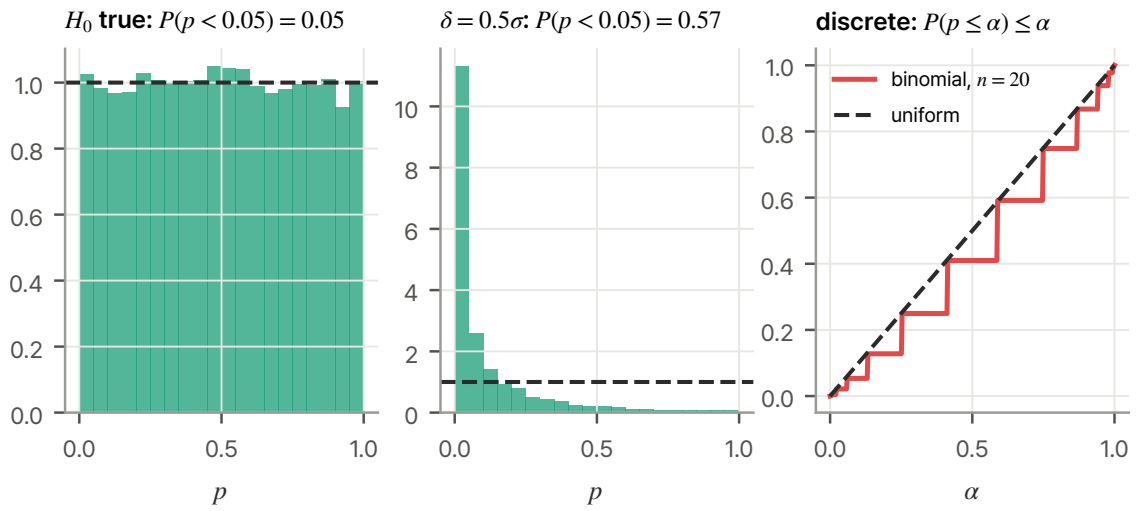


그림 4.1 t 검정 1만 번의 p-값 히스토그램($n=20$). 왼쪽: H_0가 참이면 균등분포. 가운데: 효과 $\delta = 0.5\sigma$이면 0 근처로 몰리고 $P(p<0.05)$가 검정력 0.56이다. 오른쪽: 이산 검정(이항, $n=20$)의 p-값은 계단이고 $P(p \le \alpha) \le \alpha$(보수적).

4.3 p-값이 말하지 않는 것

가정 주의 (1) p-값은 H_0가 참일 확률이 아니다. $P(\text{자료} \mid H_0)$를 $P(H_0 \mid \text{자료})$로 읽는 것은 1회의 조건부확률을 뒤집는 실수다. 후자를 말하려면 사전확률이 필요하다(16회). (2) p-값은 효과의 크기가 아니다. 표본이 크면 의미 없는 작은 차이도 아주 작은 p를 낸다. Arbuthnot의 82년 출생 93만 건에서 남아 비율 0.516과 0.5의 차이에 대한 z는 31.5, p는 10^{-218}이다. 이 p는 "차이가 있다"는 것을 압도적으로 말하지만 차이가 1.6%포인트라는 것은 말하지 않는다. 구간 $[0.5153,\ 0.5173]$이 둘을 함께 말한다. (3) $p > 0.05$는 H_0의 증거가 아니다. 검정력이 낮았을 수 있다(예제 3.3). 미국통계학회(2016)는 이 세 가지를 공식 성명으로 냈다.

연습문제

연습 4.1 (p-값 ★) $Z \sim \mathcal{N}(0,1)$ 아래에서 $|Z|$로 검정하면 $p = 2(1-\Phi(|z|))$임을 보이고, 이것도 H_0 아래 균등함을 보여라. 풀이 42쪽

연습 4.2 (이산 p-값 ★★) $K \sim \mathrm{b}(20, \frac{1}{2})$의 한쪽 p-값 $P(K \ge k)$가 가질 수 있는 값을 나열하고, $P(p(K) \le 0.05)$를 구하라. 정확히 0.05가 아닌 이유는? 풀이 42쪽

연습 4.3 (p-값의 분포와 검정력 ★★) 정리 3.4의 z 검정에서 $\mu = \mu_0 + \delta$일 때 p-값의 분포함수가 $P(p \le u) = \Phi\left(\frac{\sqrt{n}\delta}{\sigma} - z_u\right)$임을 보여라. $u = \alpha$를 넣으면 검정력이다. 풀이 42쪽

연습 4.4 (p-값을 합치기 ★★) 독립인 연구 k개의 p-값 $p_1, \dots, p_k$가 모두 H_0 아래 균등하면 $-2\sum \log p_i \sim \chi^2(2k)$임을 보여라(Fisher, 1932). (힌트: $-2\log U \sim \mathrm{Exp}$, 평균 2.) 풀이 42쪽

5

χ^2 검정 — Pearson (1900)과 Fisher (1922)

> **역사** Karl Pearson은 1900년 *Philosophical Magazine*에 "상관된 변수들의 계에서 관측된 편차가 무작위 표집에서 생겼다고 합리적으로 볼 수 있는지 판정하는 기준"을 발표했다. 관측도수와 기대도수의 차이를 하나의 수 χ^2로 모으고 그 분포를 유도한 것이다. 첫 예가 Weldon의 주사위(2회 2절)였다. 그런데 Pearson은 모수를 자료에서 추정해도 자유도가 줄지 않는다고 보았다. 젊은 Fisher는 1922년 분할표에서, 1924년 일반적으로 자유도가 추정한 모수의 수만큼 준다는 것을 보였다. Pearson은 끝내 받아들이지 않았고, 두 사람의 불화는 평생 이어졌다.

5.1 Pearson의 정리

n번의 독립 시행이 k개의 범주 가운데 하나로 떨어지고 범주 j의 확률이 p_j이면, 범주별 개수 $\mathbf{N} = (N_1, \dots, N_k)$는 다항분포다(4회 1절).

정리 5.1 (Pearson의 χ^2) $H_0 : \mathbf{p} = \mathbf{p}^0$ (모든 $p_j^0 > 0$)가 참이면

$$X^2 = \sum_{j=1}^{k} \frac{(N_j - np_j^0)^2}{np_j^0} \xrightarrow{D} \chi^2(k-1) \qquad (n \to \infty).$$

증명. **전략** X^2를 한 벡터의 길이의 제곱으로 쓴다. 그 벡터의 정규 극한이 $(k-1)$차원 공간으로의 사영임을 보이고, 회전해 $\chi^2(k-1)$을 읽는다.

① $D = \operatorname{diag}(\mathbf{p}^0)$, $\mathbf{V}_n = D^{-1/2}\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0)$로 두면 $X^2 = \|\mathbf{V}_n\|^2$이다. — $\hat{p}_j = N_j/n$을 대입

② $\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0) \xrightarrow{D} \mathcal{N}_k(\mathbf{0},\ D - \mathbf{p}^0\mathbf{p}^{0\prime})$이다. — 다항분포의 CLT, 7회 예제 5.3

③ $\mathbf{u} = \left(\sqrt{p_1^0}, \dots, \sqrt{p_k^0}\right)'$로 두면 $\|\mathbf{u}\| = 1$이고 $D^{-1/2}(D - \mathbf{p}^0\mathbf{p}^{0\prime})D^{-1/2} = I - \mathbf{u}\mathbf{u}'$이므로

$$\mathbf{V}_n \xrightarrow{D} \mathcal{N}_k(\mathbf{0},\ I - \mathbf{u}\mathbf{u}').$$

— 연속사상(7회 2.3절), 5회 정리 3.8

④ $P = I - \mathbf{u}\mathbf{u}'$는 대칭이고 $P^2 = P$인 사영행렬이다. $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, I)$이면 $P\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, PP') = \mathcal{N}_k(\mathbf{0}, P)$이므로 ③의 극한은 $P\mathbf{Z}$의 분포다. — 연습 5.2, 5회 정리 3.8

⑤ 첫 열이 $\mathbf{u}$인 직교행렬 Q를 잡고 $\mathbf{W} = Q'\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, I)$로 두면 $W_1 = \mathbf{u}'\mathbf{Z}$, $\|\mathbf{W}\| = \|\mathbf{Z}\|$이므로

$$\|P\mathbf{Z}\|^2 = \|\mathbf{Z}\|^2 - (\mathbf{u}'\mathbf{Z})^2 = \sum_{i=2}^{k} W_i^2 \sim \chi^2(k-1).$$

— 4회 정리 3.9

⑥ $\mathbf{v} \mapsto \|\mathbf{v}\|^2$은 연속이므로 $X^2 = \|\mathbf{V}_n\|^2 \xrightarrow{D} \|P\mathbf{Z}\|^2 \sim \chi^2(k-1)$이다. 연속사상, 7회 2.3절의 상자

조건 확인 모든 $p_j^0 > 0$이어야 $D^{-1/2}$가 정의된다. ⑤의 회전은 6회 Helmert 회전과 같은 논증이다. 확률이 0인 범주가 있으면 그 범주를 빼고 k를 줄인다. ■

자유도 $k-1$은 "비율의 합이 1"이라는 제약 하나 때문이다. $\mathbf{V}_n$은 언제나 $\mathbf{u}'\mathbf{V}_n = \sum(N_j - np_j^0)/\sqrt{n} = 0$을 만족해 $\mathbf{u}$ 방향의 성분이 없다. 6회의 S^2가 $\sum(X_i - \bar{X}) = 0$ 때문에 자유도 $n-1$을 가졌던 것과 같다.

> **가정 주의** 정리는 극한이다. 기대도수 np_j^0가 작으면 근사가 나빠진다. Cochran(1954)의 경험칙은 "모든 기대도수가 1 이상이고 80% 이상이 5 이상"이다. 작은 칸은 이웃 칸과 합친다. 7절에서 이 규칙을 어기면 크기가 얼마나 틀리는지 본다.

5.2 모수를 추정하면 — Fisher의 자유도

실제로는 $\mathbf{p}^0$가 완전히 정해지지 않고 모수 $\boldsymbol{\eta} \in \mathbb{R}^m$에 의존하는 경우가 많다("이 자료는 **어떤** 정규분포를 따르는가"). 그러면 $\boldsymbol{\eta}$를 자료에서 추정하고 $\hat{p}_j = p_j(\hat{\boldsymbol{\eta}})$로 기대도수를 만든다.

> **증명 없이 쓰는 사실** (추정한 모수만큼 자유도가 준다) $\boldsymbol{\eta}$를 범주별 개수로부터 최대가능도(9회)나 최소 χ^2로 추정하면, 정칙조건 아래 $X^2 = \sum_j \frac{(N_j - n\hat{p}_j)^2}{n\hat{p}_j} \xrightarrow{D} \chi^2(k-1-m)$이다(Fisher 1924; 증명은 Cramér 1946, 30장). 기하학적으로는 추정이 $\mathbf{V}_n$을 m차원 접공간 쪽으로 한 번 더 사영해 그 차원만큼 없앤다. 주의: 모수를 범주로 묶기 전의 **원자료**로 추정하면 극한은 $\chi^2(k-1-m)$과 $\chi^2(k-1)$ 사이에 있다(Chernoff–Lehmann, 1954).

5회 예제 1.6의 Quetelet 가슴둘레는 칸 13개(기대도수가 작은 양 끝을 합친 뒤)에 정규의 두 모수를 추정했으므로 자유도가 $13 - 1 - 2 = 10$이다.

5.3 분할표의 독립성

두 범주형 변수(성별과 합격 여부, 머리색과 눈색)를 $r \times c$ 표로 센다. 칸 (i,j)의 확률이 p_{ij}이고 H_0는 독립, 곧 $p_{ij} = p_{i\cdot}p_{\cdot j}$이다.

따름정리 5.2 (독립성 검정) H_0 아래에서 $E_{ij} = \frac{N_{i\cdot}N_{\cdot j}}{n}$ (행합과 열합의 곱을 총합으로 나눈 것)에 대해

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \xrightarrow{D} \chi^2\big((r-1)(c-1)\big).$$

증명. 칸은 $k = rc$개이고, 독립 모형의 모수는 행의 주변확률 $r-1$개와 열의 주변확률 $c-1$개다(합이 1이므로 하나씩 덜). 그 최대가능도 추정값이 $\hat{p}_{i\cdot} = N_{i\cdot}/n$, $\hat{p}_{\cdot j} = N_{\cdot j}/n$이다(9회). 위 상자로 자유도는 $rc - 1 - (r-1) - (c-1) = (r-1)(c-1)$. ∎

Pearson은 이 경우에도 자유도를 $rc-1$로 썼다. 2×2 표라면 Pearson의 3과 Fisher의 1의 차이다. 같은 $X^2 = 3.84$를 Pearson의 자유도로 읽으면 $p = 0.28$, Fisher의 자유도로 읽으면 $p = 0.05$이다. 누가 옳은지는 시뮬레이션으로 판정할 수 있다(노트북, 그림 5.1).

연습문제

연습 5.1 (두 범주이면 ★) $k = 2$이면 $X^2 = \frac{(N_1 - np_1)^2}{np_1p_2} = Z^2$, $Z = \frac{\hat{p}_1 - p_1}{\sqrt{p_1p_2/n}}$ 임을 보여라. 비율의 z 검정과 같다.

풀이 42쪽

연습 5.2 (사영행렬 ★) $\|\mathbf{u}\| = 1$이면 $P = I - \mathbf{u}\mathbf{u}'$가 대칭이고 $P^2 = P$이며 $\operatorname{tr} P = k - 1$임을 보여라. P의 고윳값은?

풀이 42쪽

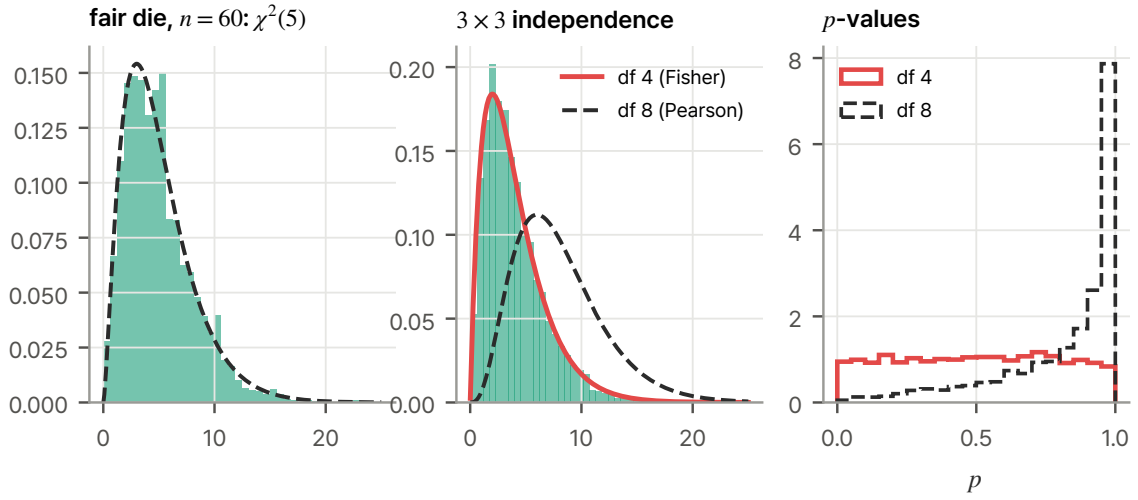


그림 5.1 왼쪽: 공평한 주사위 60번의 X^2 (5천 번 반복)와 $\chi^2(5)$. 가운데: 독립인 3×3 표($n = 300$)에서 주변확률을 추정한 X^2와 $\chi^2(4)$(Fisher, 빨강), $\chi^2(8)$(Pearson, 점선). 오른쪽: 같은 시뮬레이션의 p-값. Fisher의 자유도로 계산하면 균등하고, Pearson의 자유도로 계산하면 1 쪽으로 몰린다(검정이 지나치게 보수적).

연습 5.3 (2×2 표의 X^2 ★★) 2×2 표 $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$에서 $X^2 = \dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 임을 보여라. 풀이 43쪽

연습 5.4 (Poisson 분산 검정 ★★) $X_1,\dots,X_k$가 독립인 $\mathrm{Poi}(\lambda)$이면 합 S가 주어질 때 $(X_1,\dots,X_k)$는 $\mathrm{Mult}(S;\frac{1}{k},\dots,\frac{1}{k})$이다(4회 연습 2.4를 k개로 넓힌 것). 이것으로 $\sum_i \frac{(X_i-\bar X)^2}{\bar X}$이 근사적으로 $\chi^2(k-1)$임을 설명하고, 4회의 14개 군단의 20년 합계에서 27.3, $p=0.011$을 확인하라. 풀이 43쪽

연습 5.5 (Pearson의 자유도 ★) $X^2=3.84$를 자유도 1과 3으로 읽은 p-값을 구하라. 3×3 표에서 Pearson의 8과 Fisher의 4는? 풀이 43쪽

6

데이터로 — 네 가지 검정

6.1 Arbuthnot의 82년 (1629–1710)

예제 6.1 (세 가지 검정) `HistData::Arbuthnot`의 82년 모두에서 남아가 여아보다 많다.

(a) **부호 검정** (Arbuthnot). H_0: 해마다 남아가 많을 확률 $\frac{1}{2}$, 해들은 독립. $K = 82$이므로 $p = 2^{-82} = 2.1 \times 10^{-25}$.

(b) **비율의 검정.** 출생 하나하나가 독립이고 남아 확률이 p라 하자. 82년 합계 938,223명 중 남아 484,382명, $\hat{p} = 0.5163$이다. $H_0 : p = \frac{1}{2}$의 $z = \frac{\hat{p}-0.5}{\sqrt{0.25/n}} = 31.5$, $p \approx 10^{-218}$. 95% 구간은 $[0.5153,\ 0.5173]$이다. 남녀의 비는 약 1.067로 오늘의 출생 성비(약 1.05)와 비슷하다.

(c) **모형의 점검.** (b)의 "출생마다 독립, 해마다 같은 p"는 맞는가? 2회 예제 6.5에서 해마다의 비율의 표준편차가 이항 모형의 1.5배였다(Lexis 비). 82개 연도를 82개의 칸으로 보고 "모든 해의 남아 확률이 같다"를 82×2 분할표의 독립성 검정으로 하면(연습 6.1) $X^2 = 169.7$, 자유도 81, $p = 3 \times 10^{-8}$이다. 비율이 해마다 흔들린다.

검정은 질문을 정확하게 만든다. (a)와 (b)는 "반반인가"에 압도적으로 "아니다"라고 답한다. (c)는 "그러면 하나의 p로 설명되는가"에 다시 "아니다"라고 답한다. 어느 것도 "왜"에는 답하지 않는다. 섭리는 Arbuthnot의 해석이었다.

6.2 Weldon의 주사위 (1894)

예제 6.2 (단순가설과 복합가설) 주사위 12개를 26,306번 굴려 5·6이 나온 개수 K를 셌다 (`vcd::WeldonDice`, 10개 이상은 묶여 11칸).

(a) H_0: 공평한 주사위, $K \sim \mathrm{b}(12, \frac{1}{3})$. 기대도수(가장 작은 것도 14.3)로 $X^2 = 35.5$, 자유도 10, $p = 0.0001$. 기각한다. Pearson(1900)의 결론도 같았다.

(b) H_0: $K \sim \mathrm{b}(12, p)$, p는 모름. $\hat{p} = \bar{K}/12 = 0.3377$을 넣으면 $X^2 = 8.18$, 자유도 $11-1-1 = 9$, $p = 0.52$. 기각하지 않는다.

두 결과를 합치면 이야기가 선명하다. 주사위들은 서로 독립인 "이항" 시행처럼 행동하지만,

> $5\cdot6$이 나올 확률이 $\frac{1}{3}$이 아니라 약 0.338이다. 값싼 주사위의 눈은 파낸 홈이어서, 홈이 많은 $5\cdot6$의 면이 가벼워 위로 오기 쉽다는 설명이 있다. 효과는 $\frac{1}{3}$에 대해 1.3%다. 26,306번이라는 큰 n이 이 작은 차이를 드러냈다.

6.3 Quetelet의 가슴둘레 (1846)

5회 예제 1.6에서 스코틀랜드 병사 5,738명의 가슴둘레에 정규분포를 맞추었다. 기대도수가 5보다 작은 양 끝을 합치면 칸이 13개이고, 모수 둘(평균과 분산)을 칸으로 묶은 자료에서 추정했으므로 Fisher의 규칙으로 자유도는 $13-1-2=10$이다. $X^2=35.0$, $p=0.0001$로 정규를 기각한다(Pearson의 규칙 12로 읽어도 $p=0.0005$). 그런데 이 자료의 왜도는 0.03, 초과첨도는 0.06으로 정규와 거의 같다. 기각을 만든 것은 거의 한 칸이다. 44인치 칸의 관측 92명은 기대 146명보다 훨씬 적고, 표준화한 차이 $(O-E)/\sqrt{E}=-4.4$의 제곱 19.7이 X^2의 절반을 넘는다. 정규 **모양**이 틀린 것이 아니라 한 칸이 이상하다. 옮겨 적기의 실수인지 측정의 습관인지는 검정이 말하지 않는다.

N이 이만큼 크면 검정은 아주 작은 어긋남도 잡는다. 같은 N과 칸으로 시뮬레이션하면 참 분포가 정규일 때 기각 확률은 0.05이고, 왜도가 0.1이면 0.42, 0.2이면 0.99다(연습 6.3). 큰 표본에서 p-값이 작다는 것은 "모형이 정확히 맞지는 않는다"는 뜻일 뿐, "모형이 쓸모없다"는 뜻이 아니다.

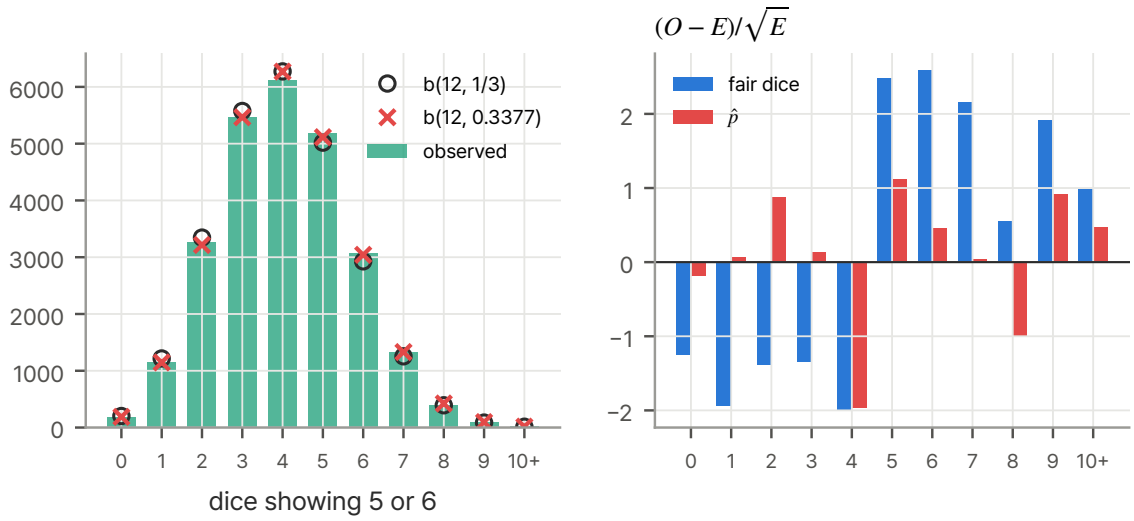


그림 6.1 Weldon의 주사위. 왼쪽: 관측도수(막대)와 b(12, $\frac{1}{3}$)(점선), b(12, 0.3377)(빨강)의 기대도수. 오른쪽: 칸별 기여 $(O-E)/\sqrt{E}$. 공평한 주사위 모형은 많은 $5\cdot6$(오른쪽)을 체계적으로 얕잡는다.

6.4 UC Berkeley 대학원 입학 (1973)

예제 6.3 (큰 X^2와 Simpson의 역설) 1973년 가을 학기 지원자 4,526명을 성별과 합격 여부로 나누면

	합격	불합격	합격률
남	1,198	1,493	44.5%
여	557	1,278	30.4%

독립성 검정의 $X^2 = 92.2$ (자유도 1), $p \approx 10^{-21}$, 오즈비 1.84 [1.62, 2.09]이다(예제 2.5). 그런데 지원자가 가장 많은 여섯 학과(A–F)로 나누면 여섯 중 넷에서 여자의 합격률이 더 높고, 학과별 표를 합친 Mantel–Haenszel 오즈비는 0.90이다. 여자는 합격률이 낮은 학과(C–F)에 많이 지원했고, 남자는 합격률이 높은 학과(A, B)에 많이 지원했다. 성별과 학과의 독립성 검정은 $X^2 = 1068$ (자유도 5)이다. 전체 표의 강한 연관은 "학과"라는 세 번째 변수를 거친 것이었다(Bickel, Hammel, O'Connell 1975). 1회 연습 5.9의 신장결석과 같은 구조다.

검정은 "성별과 합격이 독립인가"에 정확히 답했다. 독립이 아니다. 그러나 그것이 학과 안에서의 차별을 뜻하는지는 검정이 아니라 자료의 구조가 답한다. X^2의 크기는 질문의 크기가 아니다.

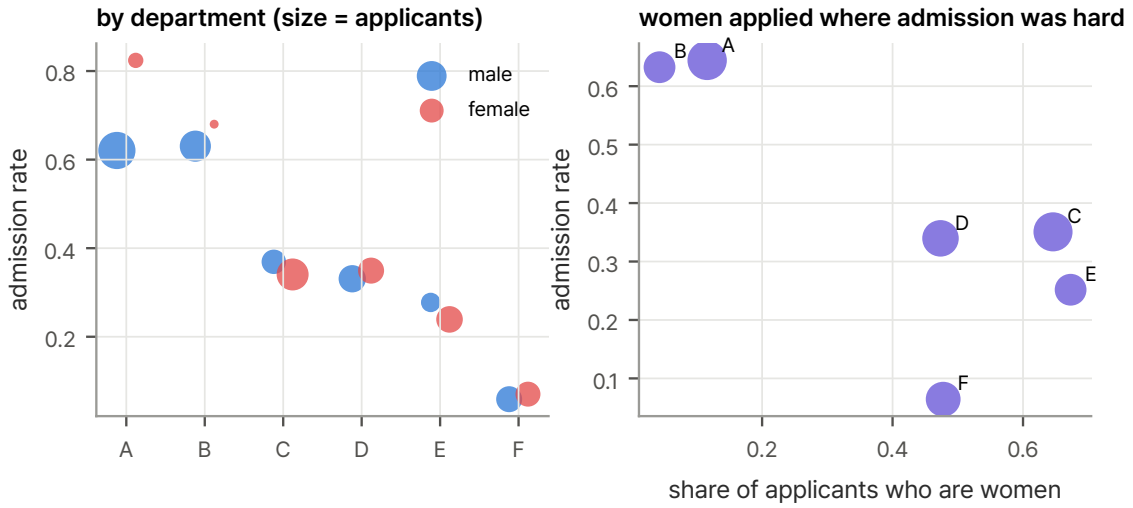


그림 6.2 Berkeley 1973. 왼쪽: 학과별 합격률(점의 크기는 지원자 수). 남녀의 차이는 학과 안에서 작거나 반대이고, 학과 사이의 합격률 차이가 크다. 오른쪽: 학과별 여자 지원자의 비율과 합격률. 여자가 많이 지원한 학과일수록 합격률이 낮다.

연습문제

연습 6.1 (Lexis의 검정 ★★) Arbuthnot 자료를 82×2 표(연도 × 성별)로 보고 독립성 검정의 X^2와 자유도를 구하라(노트북). 이것은 2회 예제 6.5의 Lexis 비와 어떻게 관련되는가? (힌트: $X^2/81$.) 풀이 43쪽

연습 6.2 (주사위의 효과 크기 ★) $\hat{p}=0.3377$의 95% 구간을 구하고(시행 $12\times 26{,}306$번), $\frac{1}{3}$이 들어 있지 않음을 확인하라. 10개 이상을 10으로 묶어 계산한 $\hat{p}$는 참값을 조금 작게 어림한다. 왜인가? 풀이 43쪽

연습 6.3 (Quetelet의 검정력 ★★) 5회의 가슴둘레 자료와 같은 $N=5{,}738$, 같은 칸(1인치로 반올림)으로, 참 분포가 정규일 때와 왜도가 0.05, 0.1, 0.2인 분포(평균과 분산을 맞춘 감마)일 때 정규 적합도 검정이 기각할 확률을 노트북으로 시뮬레이션하라. 44인치 칸을 빼고 다시 검정하면? 풀이 43쪽

연습 6.4 (Mantel–Haenszel ★★) 층 s의 2×2 표 $(a_s,b_s;c_s,d_s)$, 크기 n_s에 대해 $\widehat{\mathrm{OR}}_{\mathrm{MH}}=\dfrac{\sum a_s d_s/n_s}{\sum b_s c_s/n_s}$이다. 모든 층의 오즈비가 같은 ψ이면 분자와 분모의 기댓값의 비가 근사적으로 ψ임을 설명하고, Berkeley 자료에서 0.90을 확인하라. 풀이 44쪽

7

부숴보기 — 원하는 결과가 나올 때까지

정리 4.2는 **미리 정한 하나의** 검정에 대한 정리다. 검정을 여러 번 하거나, 결과를 보면서 언제 멈출지 정하거나, 근사가 맞지 않는 곳에서 쓰면 "5%"는 5%가 아니다.

7.1 스무 번 검정하면

귀무가설이 모두 참인 독립 검정 20개를 각각 5%로 하면, 하나라도 유의할 확률은 $1-0.95^{20} = 0.64$이다. 연구자가 그 하나만 보고하면 독자는 "5%"의 증거를 본다. 결과 변수를 바꿔 보기, 이상값을 빼 보기, 공변량을 넣어 보기 같은 흔한 선택들을 합친 시뮬레이션에서 거짓 양성의 비율은 60%를 넘었다(Simmons, Nelson, Simonsohn 2011). 이것을 p-**해킹**이라 한다. 검정 하나하나는 옳지만, 검정을 고르는 과정이 정리의 가정("미리 정한")을 깬다.

7.2 결과를 보며 멈추기

예제 7.1 (선택적 멈춤) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ (H_0가 참)을 10개씩 모으며 매번 $z = \sqrt{n}\bar{X}_n$을 보고, $|z| > 1.96$이면 "유의"로 멈춘다. 최대 1,000개까지 모으면 거짓 양성의 확률은 0.37이다(노트북). 멈출 기회가 무한하면 그 확률은 1이다.

증명 없이 쓰는 사실 (반복 로그의 법칙) iid, 평균 0, 분산 1이면 확률 1로 다음이 성립한다(Khinchin 1924, Kolmogorov 1929).

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1$$

$\sqrt{2\log\log n}$은 아주 느리지만 끝없이 자라므로 $|z_n| = |S_n|/\sqrt{n}$은 언젠가 반드시 1.96을 넘는다.

7회의 Lazzarini(1901)가 이 함정의 고전적인 예다. "355/113이 나오면 멈춘다"는 규칙 아래에서 보고된 값의 오차를 미리 정한 n의 표준오차로 말하는 것은 틀렸다(7회 연습 4.4). 결과를 보면서 멈추려면 처음부터 그것을 허용하도록 만든 순차 검정(Wald 1945)이나 기준을 쓴다.

7.3 비율의 Wald 구간

예제 7.2 (명목 95%가 22%) $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$의 실제 포함확률은 이항분포의 합으로 정확히 계산된다($\sum_k \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}\{p \in \text{구간}(k)\}$). $n = 50$이면 $p = 0.005$에서 0.22, $p = 0.01$에서 0.39, $p = 0.1$에서 0.88이다. p가 작으면 $\hat{p} = 0$이 자주 나오고, 그때 구간은 폭이 0인 $\{0\}$이 되어 참값을 놓친다. 더 놀라운 것은 p가 크지 않아도 n에 따라 들쭉날쭉하다는 것이다. $p = 0.005$에서 $n = 592$이면 0.79다(Brown, Cai, DasGupta 2001). 같은 자료로 만든 Wilson 구간(예제 3.6)은 대부분의 p에서 0.95 근처다.

대표본 정리(정리 2.4)는 $n \to \infty$에서의 말이다. $n = 50$은 큰가? np가 작으면 아니다. 정리가 말해 주지 않는 것은 **얼마나 커야** 하는가이고, 그것은 계산해 보아야 안다.

7.4 기대도수가 작은 χ^2

칸 10개에 확률이 같고 $n = 20$(기대도수 모두 2)이면 5% 검정의 실제 크기는 0.051로 근사가 놀라울 만큼 잘 맞는다. 그러나 칸 6개 중 다섯이 기대도수 0.6이고 하나가 27이면 크기가 0.11로 두 배다. 작은 기대도수로 나누는 항이 드물게 크게 튀기 때문이다. 이럴 때는 칸을 합치거나, 귀무가설 아래에서 자료를 시뮬레이션해 X^2의 분포를 직접 만든다(**Monte Carlo 검정**). 시뮬레이션한 B개의 통계량 가운데 관측값 이상인 것의 수를 R이라 하면 $\hat{p} = \frac{R+1}{B+1}$은 정확히 타당한 p-값이다(연습 7.4, 8.2절의 순위 논증).

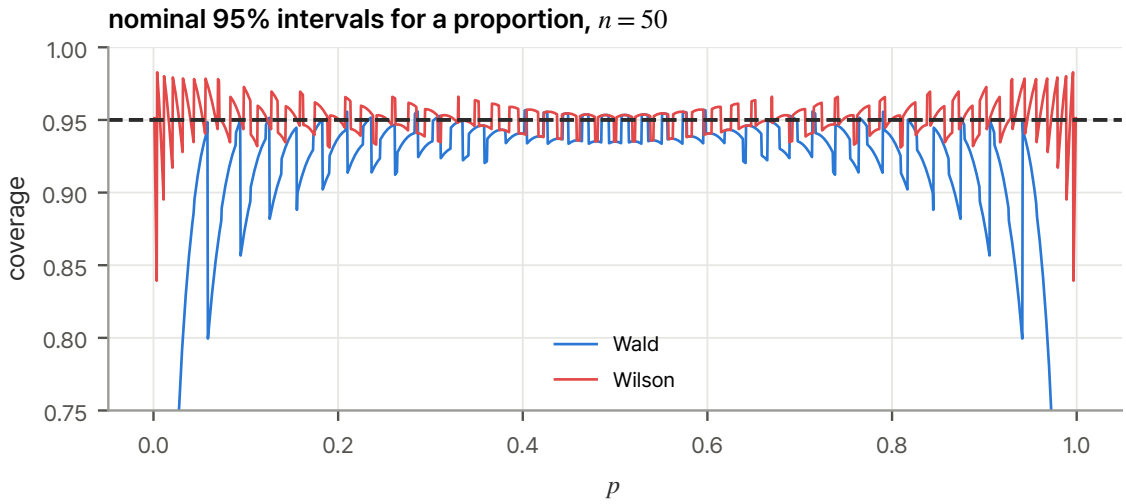


그림 7.1 $n = 50$에서 비율의 95% 구간의 실제 포함확률. Wald(파랑)는 양 끝에서 무너지고 가운데서도 진동한다. Wilson(빨강)은 진동하지만 0.95 근처에 머문다. 진동은 $\hat{p}$가 k/n의 값만 가지는 이산성 때문이다.

연습문제

연습 7.1 (스무 번의 검정 ★) 독립 검정 m개를 각각 α로 할 때 하나라도 거짓 양성이 나올 확률을 구하고, 이것을 0.05 이하로 하려면 각 검정의 α를 얼마로 해야 하는가? ($m = 20$)풀이 44쪽

연습 7.2 (선택적 멈춤 ★★) 예제 7.1을 노트북에서 최대 표본 크기 100, 1,000, 10,000으로 되풀이하라. 거짓 양성의 확률은 어떻게 자라는가? 풀이 44쪽

연습 7.3 (정확한 포함확률 ★★) Wald 구간의 포함확률을 이항 확률의 합으로 쓰고, np가 작으면 포함확률이 거의 $1-(1-p)^n$임을 설명하라. ($\hat{p}=0$이면 구간이 $\{0\}$이어서 놓치고, $\hat{p}=\frac{1}{n}, \frac{2}{n}$이면 작은 p를 대개 포함한다.) $n=50$, $p=0.005$에서 계산해 0.22와 비교하라.풀이 44쪽

연습 7.4 (Monte Carlo 검정 ★★) H_0 아래에서 관측 통계량 T_0와 시뮬레이션한 $T_1,\dots,T_B$가 iid이고 연속이면, T_0의 순위가 $\{1,\dots,B+1\}$ 위에서 균등함을 보이고 $P\big(\frac{R+1}{B+1} \le \alpha\big) \le \alpha$를 결론지어라. 풀이 44쪽

8

Beautiful insights — 균등성과 순위

8.1 균등성으로 보정을 점검하기

정리 4.2의 뒤집은 형태가 쓸모 있다. **계산이 맞다면** p-값은 균등해야 한다. 그러니 균등하지 않으면 계산이 틀렸다. 이 생각이 여러 곳에서 진단 도구가 된다.

- **예측분포의 보정.** 일기예보나 회귀모형이 관측 y_t마다 예측분포 F_t를 내놓는다면, 보정이 맞을 때 $F_t(y_t)$는 균등하다(2회 정리 4.7, PIT). 그 히스토그램이 U자이면 예측이 지나치게 확신하고(분산이 작고), 가운데가 솟으면 지나치게 조심스럽다(Dawid 1984; Gneiting 외 2007).
- **베이즈 계산의 점검.** 사전분포에서 θ를 뽑고, 그 θ로 자료를 만들고, 그 자료로 사후분포의 표본을 뽑으면, 처음의 θ가 사후표본들 사이에서 차지하는 순위는 균등해야 한다(Talts 외 2018, simulation-based calibration). 순위가 치우치면 사후분포를 계산하는 코드가 틀렸다(16회).
- **검정 코드의 점검.** 새 검정을 구현하면 귀무가설 아래에서 자료를 수천 번 만들어 p-값의 히스토그램을 그려 본다. 노트북에서 모든 검정을 이렇게 점검한다.

8.2 순위와 등각 예측

6회 정리 8.1의 예측구간은 정규를 가정했다. 분포를 전혀 가정하지 않고도 예측구간을 만들 수 있다. 열쇠는 순위다.

> **보조정리 8.1** (교환가능한 순위) $S_1, \dots, S_{n+1}$이 교환가능하고(결합분포가 순서를 바꿔도 같고) 같은 값이 나올 확률이 0이면, S_{n+1}의 순위 $R = \#\{i \le n+1 : S_i \le S_{n+1}\}$는 $\{1, \dots, n+1\}$ 위의 균등분포다.

증명. **전략** 교환가능성으로 "S_j가 r번째"의 확률은 j와 무관하다. r번째는 정확히 하나이므로 그 확률들의 합이 1이다. 단계별 증명은 연습 8.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 45쪽에 있다.

정리 8.2 (분할 등각 예측) 아무 예측모형 $\hat{\mu}$를 학습 자료로 만들고, 그것과 독립인 보정 자료 (X_i, Y_i), $i = 1, \dots, n$에서 점수 $S_i = |Y_i - \hat{\mu}(X_i)|$를 계산한다. $k = \lceil (n+1)(1-\alpha) \rceil \le n$일 때 $\hat{q} = S_{(k)}$로 두면, 새 관측 (X_{n+1}, Y_{n+1})이 보정 자료와 교환가능할 때

$$P\big(Y_{n+1} \in [\hat{\mu}(X_{n+1}) - \hat{q},\ \hat{\mu}(X_{n+1}) + \hat{q}]\big) \ge 1 - \alpha.$$

증명. **전략** 학습 자료를 고정하면 점수 $S_1, \dots, S_{n+1}$이 교환가능하다. 포함 사건을 "S_{n+1}의 순위가 k 이하"로 바꾸고 보조정리 8.1를 쓴다.

① 학습 자료가 주어지면 $\hat{\mu}$는 고정된 함수이고, 같은 함수를 교환가능한 (X_i, Y_i)들에 쓴 $S_1, \dots, S_{n+1}$도 교환가능하다. — 학습 자료와 보정 자료의 독립

② $Y_{n+1} \in [\hat{\mu}(X_{n+1}) - \hat{q},\ \hat{\mu}(X_{n+1}) + \hat{q}] \iff S_{n+1} \le \hat{q} = S_{(k)}$이다. 여기서 $S_{(k)}$는 보정 점수 n개의 k번째 순서통계량이다. — 절댓값의 정의

③ 같은 값이 없으면 $S_{n+1} \le S_{(k)}$는 "S_{n+1}보다 작은 보정 점수가 $k-1$개 이하", 곧 S_{n+1}의 $n+1$개 가운데 순위 R이 k 이하인 것과 같다. — 원소 추적 [T1]

④ $P(R \le k) = \dfrac{k}{n+1} \ge 1 - \alpha$이다. 학습 자료에 대해 조건부로 성립하므로 무조건부로도 성립한다. — 보조정리 8.1, $k \ge (n+1)(1-\alpha)$

조건 확인 같은 값이 나올 확률이 양수이면, 같은 값들 사이의 순서를 동전으로 정한 순위 R'은 교환가능성으로 여전히 균등하다. $R' \le k$이면 $S_{n+1} \le S_{(k)}$이므로 포함확률은 $\dfrac{k}{n+1}$ 이상이다. ∎

모형이 신경망이든 무작위 숲이든, 자료가 정규든 아니든 성립한다. 가정은 교환가능성 하나다 (Vovk, Gammerman, Shafer 2005; Lei 외 2018). 같은 논증이 7.4절의 Monte Carlo 검정의 타당성이고, 15회의 순열검정의 원리다. 대가는 구간이 x에 따라 넓어지거나 좁아지지 않는다는 것이고 (점수를 바꿔 고친다), 보장이 **평균적인** 포함이라는 것이다.

8.3 다중검정 — Bonferroni와 거짓 발견율

유전자 2만 개 가운데 어느 것이 질병과 관련 있는지 검정하면 $\alpha = 0.05$에서 거짓 양성이 천 개다. 두 가지 해법이 있다.

명제 8.3 (Bonferroni) m개의 검정을 각각 α/m로 하면, 어떤 의존 구조에서도 하나라도 거짓 양성이 나올 확률(가족별 오류율)은 α 이하다.

증명. 참인 귀무가설의 집합을 I_0라 하면 $P\big(\bigcup_{i \in I_0} \{p_i \le \alpha/m\}\big) \le \sum_{i \in I_0} P(p_i \le \alpha/m) \le |I_0| \dfrac{\alpha}{m} \le \alpha$ (합집합 한계, 1회; 정리 4.2). ∎

> **증명 없이 쓰는 사실** (Benjamini–Hochberg (1995)) p-값을 정렬해 $p_{(i)} \le \frac{i}{m}q$인 가장 큰 i를 k라 하고 $p_{(1)}, \dots, p_{(k)}$를 모두 기각하면, 검정들이 독립일 때 **거짓 발견율**, 곧 기각한 것 가운데 거짓 양성의 기대 비율이 $q \cdot \frac{m_0}{m} \le q$ 이하다. 하나도 틀리지 않으려는 Bonferroni보다 훨씬 많이 찾는다.

1,000개의 검정 가운데 100개에 실제 효과(z의 평균 3)가 있는 시뮬레이션에서 $q = 0.1$의 Benjamini–Hochberg는 평균 79개를, 같은 수준의 Bonferroni는 평균 24개를 찾았다(노트북, 200번 반복). BH가 찾은 것 가운데 거짓은 평균 9%였다. 유전체학, A/B 시험, 뇌영상에서 오늘의 표준이 거짓 발견율인 이유다.

8.4 시험 자료를 거듭 쓰면

기계학습의 시험 자료는 검정과 같은 역할을 한다. 모형이 시험 자료에서 좋은 점수를 받으면 "일반화한다"고 결론짓는다. 그런데 같은 시험 자료로 모형을 수백 번 고르면, 그것은 수백 번의 검정을 하고 가장 좋은 하나를 보고하는 것과 같다(7.1절). 경진대회의 공개 순위표가 비공개 순위표와 어긋나는 것이 그 결과다. Dwork 외(2015)는 시험 자료를 거듭 쓰면서도 이런 적응적 과적합을 막는 방법(재사용 가능한 홀드아웃)을 제안했다. 한편 ImageNet의 시험 자료를 같은 절차로 새로 모은 Recht 외(2019)는 정확도가 11–14%포인트 떨어졌지만 모형들의 순위는 거의 그대로임을 보이고, 그 원인을 적응적 과적합보다 자료 수집의 미묘한 차이로 보았다. 어느 쪽이든 교훈은 7절과 같다. 같은 자료로 질문을 고르고 답을 확인하면 오차의 계산이 무효가 된다.

연습문제

연습 8.1 (PIT 히스토그램 ★) 참 분포가 $\mathcal{N}(0,1)$인데 예측분포를 $\mathcal{N}(0, 0.5^2)$으로 내놓으면 $F(Y)$의 분포가 U자임을 보여라. $\mathcal{N}(0, 2^2)$이면? 풀이 44쪽

연습 8.2 (등각 구간의 위쪽 한계 ★★★) 정리 8.2에서 점수가 연속이면 포함확률이 $1-\alpha+\frac{1}{n+1}$ 이하임도 보여라. $n = 19$, $\alpha = 0.1$이면 포함확률은 정확히 얼마인가? 풀이 45쪽

연습 8.3 (Bonferroni와 독립 ★) 검정들이 독립이면 각각 α/m로 할 때 가족별 오류율이 정확히 $1-(1-\alpha/m)^m$ 임을 보이고, $m = 20$, $\alpha = 0.05$에서 계산하라. Bonferroni 한계와 얼마나 가까운가? 풀이 45쪽

연습 8.4 (BH를 손으로 ★★) $m = 10$개의 p-값 0.001, 0.004, 0.009, 0.02, 0.03, 0.04, 0.2, 0.4, 0.6, 0.9에 $q = 0.05$의 Benjamini–Hochberg를 적용하면 몇 개를 기각하는가? Bonferroni는? 풀이 45쪽

연습 8.5 (순위의 균등성 ★★) $\theta \sim \mathcal{N}(0,1)$, $X \mid \theta \sim \mathcal{N}(\theta,1)$이면 사후분포는 $\mathcal{N}(X/2,\ 1/2)$이다. 노트북에서 θ를 뽑고, X를 만들고, 사후분포에서 표본 19개를 뽑아 θ의 순위를 기록하는 일을 수천 번 하라. 순위가 $\{1,\dots,20\}$에서 균등함을 확인하고, 사후분산을 일부러 $1/4$로 틀리게 하면 히스토그램이 어떻게 되는지 보라. 풀이 45쪽

연습 8.6 (교환가능한 순위의 증명 ★★ 유도형) 보조정리 8.1를 증명하라. $R_j = \#\{i \le n+1 : S_i \le S_j\}$는 S_j의 순위다.

(a) j와 $n+1$을 맞바꾸는 순열을 $(S_1,\dots,S_{n+1})$에 쓰면 결합분포가 같다는 것에서 $P(R_j = r) = P(R_{n+1} = r)$을 보여라.

(b) 같은 값이 없으면 각 r에 대해 $R_j = r$인 j가 정확히 하나임을 보이고 $\sum_{j=1}^{n+1} P(R_j = r) = 1$을 얻어라.

(c) (a)와 (b)로 $P(R_{n+1} = r) = \dfrac{1}{n+1}$을 결론지어라.

(d) 같은 값이 나올 수 있으면 (b)의 어느 부분이 틀리는가? S_i가 모두 같은 극단적인 경우로 확인하라.

풀이 45쪽

9

정리와 역사 노트

9.1 한 장 요약

정리	내용	조건
순서통계량 (1.2, 1.3)	결합밀도 $n!\prod f(y_i)$, $X_{(k)}$의 밀도	연속, iid
표본분위수 (1.4)	$\sqrt{n}(X_{(k)}-\xi_p)\xrightarrow{D}\mathcal{N}\left(0,\frac{p(1-p)}{f(\xi_p)^2}\right)$	$f(\xi_p)>0$
독일 전차 (1.6)	$EM=\frac{k(N+1)}{k+1}$, $\hat{N}=\frac{k+1}{k}M-1$ 불편	비복원 추출
축량 · 대표본 구간 (2.2, 2.4)	축량의 분위수를 뒤집는다; $\hat{\theta}\pm z\hat{\sigma}/\sqrt{n}$	정확 / 근사
중앙값 구간 (2.6)	$P(X_{(i)}<m<X_{(j)})=\sum_{k=i}^{j-1}\binom{n}{k}2^{-n}$	연속
검정력 · 쌍대성 (3.4, 3.5)	$\gamma(\mu)=\Phi(\sqrt{n}\delta/\sigma-z_\alpha)$; 구간 $\iff$ 검정	—
p-값 (4.2)	H_0 아래 $\mathrm{U}(0,1)$, 이산이면 $P(p\le\alpha)\le\alpha$	단순 H_0
Pearson (5.1, 5.2)	$X^2\xrightarrow{D}\chi^2(k-1)$, 모수 m개를 추정하면 $k-1-m$, 분할표 $(r-1)(c-1)$	큰 기대도수
등각 예측 (8.2) · Bonferroni (8.3)	순위의 균등성으로 포함확률 $\ge 1-\alpha$; α/m로 가족별 오류율 $\le\alpha$	교환가능 / —

이번 회차의 정리들은 세 가지 도구로 증명되었다. **확률적분변환**($F(X)$가 균등하다: 순서통계량, p-값), **뒤집기**(축량의 분위수를 모수에 대해 풀기, 검정의 수락역을 구간으로 모으기), 그리고 7회의 **극한 정리**(대표본 구간, 표본분위수, χ^2)다. 7절은 그 정리들이 "미리 정한 하나의 절차"에 대한 것임을 보였다.

9.2 Stigler의 명명 법칙

> Stigler's law of eponymy
>
> - **Simpson의 역설**: Simpson(1951)보다 반세기 먼저 Pearson(1899)과 Yule(1903)이 같은 현상을 기술했다. 명명 법칙의 교과서적인 예다.
> - **Bonferroni 부등식**: 합집합의 확률이 확률의 합 이하라는 것은 Boole(1854)의 것이다. Bonfer-

roni(1936)는 그것을 다듬었고, 다중비교에 쓰는 법을 정리한 것은 Dunn(1961)이었다.

- **유의성 검정**: 형태는 Arbuthnot(1710)의 것이고, 't Gravesande(1712)가 곧 이항분포로 다시 계산했다. Laplace(1823)는 달이 기압에 미치는 영향을 같은 방식으로 검정했다. "유의" 라는 말과 5% 는 Edgeworth와 Fisher의 것이다.

9.3 7 Pillars 지도

이번 회차는 셋째 기둥 **Likelihood**, 곧 "확률의 척도로 증거를 잰다" 의 시작이다. Stigler(2016)는 이 기둥의 장을 바로 Arbuthnot의 1710년 논증으로 연다. 관측의 확률을 가설 아래에서 계산하고, 그 작음을 증거로 읽는 것이다. 9회의 최대가능도와 12회의 Neyman–Pearson 보조정리가 이 기둥을 완성한다. 6절의 Berkeley 자료는 일곱째 기둥 **Residual**(설명되지 않고 남는 것을 보라)과 다섯째 기둥 **Regression**(조건부로 보라)이 왜 필요한지도 보였다. 학과로 조건을 걸자 결론이 뒤집혔다.

9.4 다음 회 — 최대가능도의 긴 이야기 (1778 → 1951)

이 노트에서는 모수를 "그럴듯하게" 추정하는 일반적인 방법 없이 표본평균, 최댓값, 순서통계량을 그때그때 썼다. 5절의 Fisher의 자유도도 "최대가능도로 추정하면" 이라는 조건이 붙어 있었다. 9회는 Daniel Bernoulli(1778)의 질문, 곧 "관측들이 서로 어긋날 때 가장 그럴듯한 값은 무엇인가" 에서 출발해 가능도함수, 최대가능도 추정량, Fisher 정보, Cramér–Rao 한계, 그리고 가능도비 검정을 증명한다. 7회의 델타 방법과 중심극한정리가 최대가능도 추정량의 점근정규성이 된다.

참고문헌

- Arbuthnot, J. An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions* 27 (1710), 186–190.
- Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* 50 (1900), 157–175.
- Fisher, R. A. On the interpretation of χ^2 from contingency tables, and the calculation of P. *JRSS* 85 (1922), 87–94. · The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *JRSS* 87 (1924). · *Statistical Methods for Research Workers*. Oliver and Boyd, 1925.
- Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. R. Soc. A* 236 (1937), 333–380. · Neyman, J., Pearson, E. S. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. A* 231 (1933).

- Wilson, E. B. Probable inference, the law of succession, and statistical inference. *JASA* 22 (1927), 209–212.
- Brown, L. D., Cai, T. T., DasGupta, A. Interval estimation for a binomial proportion. *Statistical Science* 16 (2001), 101–133.
- Ruggles, R., Brodie, H. An empirical approach to economic intelligence in World War II. *JASA* 42 (1947), 72–91.
- Cochran, W. G. Some methods for strengthening the common χ^2 tests. *Biometrics* 10 (1954), 417–451. · Chernoff, H., Lehmann, E. L. The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* 25 (1954), 579–586.
- Bickel, P. J., Hammel, E. A., O'Connell, J. W. Sex bias in graduate admissions: data from Berkeley. *Science* 187 (1975), 398–404.
- Wald, A. Sequential tests of statistical hypotheses. *Ann. Math. Statist.* 16 (1945), 117–186.
- Simmons, J. P., Nelson, L. D., Simonsohn, U. False-positive psychology. *Psychological Science* 22 (2011), 1359–1366. · Wasserstein, R. L., Lazar, N. A. The ASA statement on *p*-values. *The American Statistician* 70 (2016), 129–133.
- Benjamini, Y., Hochberg, Y. Controlling the false discovery rate. *JRSS B* 57 (1995), 289–300.
- Vovk, V., Gammerman, A., Shafer, G. *Algorithmic Learning in a Random World*. Springer, 2005. · Lei, J. 외. Distribution-free predictive inference for regression. *JASA* 113 (2018), 1094–1111.
- Dawid, A. P. Statistical theory: the prequential approach. *JRSS A* 147 (1984). · Gneiting, T., Balabdaoui, F., Raftery, A. E. Probabilistic forecasts, calibration and sharpness. *JRSS B* 69 (2007). · Talts, S. 외. Validating Bayesian inference algorithms with simulation-based calibration. *arXiv:1804.06788* (2018).
- Dwork, C. 외. The reusable holdout. *Science* 349 (2015), 636–638. · Recht, B. 외. Do ImageNet classifiers generalize to ImageNet? *ICML* (2019).
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016.
- 자료: `HistData`의 `Arbuthnot`, `ChestSizes`; `vcd`의 `WeldonDice`; `datasets`의 `UCBAdmissions`.

A

연습문제 풀이

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

1절

연습 1.1 문제 7쪽

정리 1.2의 밀도 $n!$을 가운데 $n-2$개의 좌표에 대해 $y_1 < y_2 < \cdots < y_{n-1} < y_n$ 위에서 적분하면 $(X_{(1)}, X_{(n)})$의 결합밀도는 $n(n-1)(v-u)^{n-2}$ $(0 < u < v < 1)$이다(가운데 $n-2$개가 (u,v)에 있는 부피 $\frac{(v-u)^{n-2}}{(n-2)!}$). $R = V - U$, $W = U$로 바꾸면(야코비안 1) $f(r,w) = n(n-1)r^{n-2}$, $0 < w < 1-r$이고, w로 적분하면 $f_R(r) = n(n-1)r^{n-2}(1-r)$, 곧 $\mathrm{Beta}(n-1, 2)$.

연습 1.2 문제 7쪽

$y_i = \sum_{j \le i} d_j$로 바꾸면 야코비안이 1이고 $\sum_i y_i = \sum_i (n-i+1)d_i$이므로 결합밀도는 $n! \exp\!\left(-\sum (n-i+1)d_i\right) = \prod_i (n-i+1) e^{-(n-i+1)d_i}$ $(d_i > 0)$이다. 인수분해되므로 D_i는 독립이고 $D_i \sim \mathrm{Exp}$ (비율 $n-i+1$), 곧 $(n-i+1)D_i \sim \mathrm{Exp}(1)$. 3회 예제 3.5의 "최솟값은 비율 $n\lambda$의 지수"를 되풀이한 것이다. $EX_{(k)} = \sum_{i \le k} ED_i$.

연습 1.3 문제 8쪽

$\{1, \dots, N+1\}$에서 $k+1$개를 고르는 방법 $\binom{N+1}{k+1}$가지를 가장 큰 수 $m+1$ $(m = k, \dots, N)$로 나누면, 나머지 k개를 $\{1, \dots, m\}$에서 고르는 $\binom{m}{k}$가지씩이다.

연습 1.4 문제 8쪽

$m(m+1)\binom{m-1}{k-1} = \frac{(m+1)!}{(k-1)!(m-k)!} = k(k+1)\binom{m+1}{k+1}$이고 하키스틱으로 $\sum_{m=k}^{N} \binom{m+1}{k+1} = \binom{N+2}{k+2}$이므로

$E[M(M+1)] = \frac{k(k+1)\binom{N+2}{k+2}}{\binom{N}{k}} = \frac{k(N+1)(N+2)}{k+2}$. 따라서

$$\operatorname{Var} M = E[M(M+1)] - EM - (EM)^2 = \frac{k(N+1)(N-k)}{(k+1)^2(k+2)},$$
$$\operatorname{Var} \hat{N} = \left(\frac{k+1}{k}\right)^2 \operatorname{Var} M = \frac{(N+1)(N-k)}{k(k+2)}.$$

한편 $\{1, \dots, N\}$의 분산은 $\frac{N^2-1}{12}$이고, 비복원 표본평균의 분산은 4회 정리 1.5의 유한 모집단 수정으로 $\frac{N^2-1}{12k} \cdot \frac{N-k}{N-1} = \frac{(N+1)(N-k)}{12k}$. 4배 하면 $\operatorname{Var}(2\bar{X} - 1)$. $N = 300$, $k = 5$이면 표준편차 50.4와 76.9(시뮬레이션 50.3, 77.0).

연습 1.5 문제 8쪽

Laplace의 중앙값 μ에서 $f(\mu) = \frac{1}{2b}$이므로 점근분산 $\frac{1}{4nf^2} = \frac{b^2}{n}$. 평균의 분산 $\frac{2b^2}{n}$의 절반이다. 중앙값이 평균보다 두 배 효율적이다. $t(3)$이면 $f(0) = \frac{\Gamma(2)}{\sqrt{3\pi}\,\Gamma(3/2)} = 0.368$이므로 중앙값의 분산은 $\frac{1.85}{n}$, 평균의 분산은 $\frac{\operatorname{Var} T}{n} = \frac{3}{n}$이다. 중앙값이 1.62배 효율적이다.

연습 1.6 문제 8쪽

(a) $X_{(k)} \le x$이면 $X_{(1)}, \dots, X_{(k)}$이 모두 x 이하이므로 x 이하인 관측이 k개 이상이다. 거꾸로 x 이하인 관측이 k개 이상이면 k번째로 작은 값은 x 이하다. x 이하인 관측의 수는 $\mathrm{b}(n, F(x))$이므로 $P(X_{(k)} \le x) = \sum_{j=k}^{n} \binom{n}{j} F^j (1-F)^{n-j}$이다.

(b) F가 미분가능한 점에서 $\frac{d}{dx} F^j (1-F)^{n-j} = f\left[jF^{j-1}(1-F)^{n-j} - (n-j)F^j(1-F)^{n-j-1}\right]$이다. 여기에 $\binom{n}{j}$를 곱하고 $j\binom{n}{j} = n\binom{n-1}{j-1}$, $(n-j)\binom{n}{j} = n\binom{n-1}{j}$를 쓰면 주어진 식이다. ($j = n$이면 둘째 항은 $\binom{n-1}{n} = 0$이다.)

(c) $a_j = \binom{n-1}{j-1} F^{j-1}(1-F)^{n-j}$로 두면 둘째 항은 a_{j+1}이다. 합은 $\sum_{j=k}^{n} (a_j - a_{j+1}) = a_k - a_{n+1} = a_k$ ($a_{n+1} = 0$)이므로

$$f_{(k)}(x) = n\binom{n-1}{k-1} F^{k-1}(1-F)^{n-k} f = \frac{n!}{(k-1)!\,(n-k)!} F(x)^{k-1}\big(1-F(x)\big)^{n-k} f(x).$$

(d) $U_i = F(X_i)$가 아니라 $X_i = F^{-1}(U_i)$로 표본을 만들면(2회 정리 4.7) F^{-1}이 비감소이므로 $X_{(k)} = F^{-1}(U_{(k)})$이다. $U_{(k)}$의 밀도 $\frac{n!}{(k-1)!(n-k)!} u^{k-1}(1-u)^{n-k}$에 단조변환의 밀도 공식을 쓰면 $f_{(k)}(x) = f_{\mathrm{Beta}}(F(x))\, f(x)$로 같은 식이다.

2절

연습 2.1 문제 11쪽

계산이 끝난 구간 $[480, 1710]$은 고정된 두 수이고 참 평균도 고정된 수이므로, 그 안에 있을 "확률"

은 0이나 1이다(모를 뿐이다). 맞는 문장: "이 방법으로 만든 구간은 반복된 표본의 95%에서 참 평균을 포함한다. $[480, 1710]$은 그 방법으로 이 표본에서 만든 구간이다." 확률을 모수에 붙이려면 사전분포가 필요하다(16회).

연습 2.2 문제 11쪽

$P(M/\theta \le x) = x^n$ $(0 \le x \le 1)$이 θ와 무관하므로 축량이다. $P(\alpha^{1/n} \le M/\theta \le 1) = 1-\alpha$를 θ에 대해 풀면 $M \le \theta \le M\alpha^{-1/n}$. θ는 언제나 M 이상이므로 아래쪽으로 넓힐 이유가 없고, 축량의 분포가 1 근처에 몰려 있어 위쪽 꼬리만 자르는 것이 가장 짧다.

연습 2.3 문제 11쪽

$0.4452 - 0.3035 = 0.1417$, 표준오차 $\sqrt{\frac{0.4452\cdot 0.5548}{2691} + \frac{0.3035\cdot 0.6965}{1835}} = 0.0144$이므로 $[0.114,\ 0.170]$.

연습 2.4 문제 11쪽

정리 2.6로 $[X_{(i)}, X_{(21-i)}]$의 신뢰계수는 $i=5$에서 0.988, $i=6$에서 0.959, $i=7$에서 0.885이므로 $i=6$. 정규근사로 $P(i \le K \le n-i) \approx 1-\alpha$이고 $K \approx \mathcal{N}(\frac{n}{2}, \frac{n}{4})$이므로 $\frac{n}{2} - i \approx z_{\alpha/2}\frac{\sqrt{n}}{2}$. $n=20$이면 $10 - 4.4 = 5.6$이다.

연습 2.5 문제 11쪽

(a) $\theta \in C(\mathbf{X})$는 C의 정의로 $a \le Q(\mathbf{X},\theta) \le b$와 같다. θ를 고정하면 오른쪽은 $\mathbf{X}$에 대한 사건이다. (b) (a)로 $P_\theta(\theta \in C(\mathbf{X})) = P_\theta(a \le Q(\mathbf{X},\theta) \le b)$이다. $\mathbf{X}$가 P_θ를 따를 때 $Q(\mathbf{X},\theta)$의 분포는 θ와 무관한 분포이고, a, b는 그 분포의 분위수이므로 이 확률은 모든 θ에서 $1-\alpha$이다. (c) $Q(\mathbf{x},\cdot)$이 증가하고 $\theta_1 < \theta < \theta_2$이면 $a \le Q(\mathbf{x},\theta_1) \le Q(\mathbf{x},\theta) \le Q(\mathbf{x},\theta_2) \le b$이므로 $\theta \in C(\mathbf{x})$이다. 두 점 사이를 모두 포함하는 실수의 집합은 구간이다. 감소이면 부등호가 뒤집힐 뿐 같다.

연습 2.6 문제 12쪽

(a) $\hat\sigma_n > 0$이면 $T_n = \frac{\sigma}{\hat\sigma_n}\cdot\frac{\sqrt{n}(\hat\theta_n-\theta)}{\sigma}$이다. $\hat\sigma_n \xrightarrow{P} \sigma > 0$이므로 $P(\hat\sigma_n \le \sigma/2) \to 0$이고, 그 사건은 극한에 영향을 주지 않는다. (b) $g(s) = \sigma/s$는 $s=\sigma>0$에서 연속이므로 연속사상으로 $\sigma/\hat\sigma_n \xrightarrow{P} 1$이다. $\sigma = 0$이면 g가 정의되지 않는다. (c) $\frac{\sqrt{n}(\hat\theta_n-\theta)}{\sigma} \xrightarrow{D} \mathcal{N}(0,1)$이고 (b)이므로 Slutsky로 곱은 $\mathcal{N}(0,1)$로 분포수렴한다. (d) $\theta \in [\hat\theta_n \pm z_{\alpha/2}\hat\sigma_n/\sqrt{n}] \iff |T_n| \le z_{\alpha/2}$이다. Φ는 모든 점에서 연속이므로 $P(|T_n| \le z_{\alpha/2}) \to \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1-\alpha$이다.

연습 2.7 문제 12쪽

(a) 분포함수 F가 연속이고 m이 중앙값이면 $F(m) = \frac{1}{2}$이고 $P(X_l = m) = F(m) - F(m-) = 0$이므로 $P(X_l < m) = \frac{1}{2}$이다. 관측들이 독립이므로 $K \sim \mathrm{b}(n, \frac{1}{2})$이다. (b) $X_{(i)} < m$이면 가장

작은 i개가 모두 m보다 작으므로 $K \geq i$이고, 거꾸로 $K \geq i$이면 i번째로 작은 값이 m보다 작다. $m < X_{(j)}$이면 m보다 작은 관측은 $X_{(1)}, \dots, X_{(j-1)}$ 가운데에만 있으므로 $K \leq j-1$이고, 거꾸로 $K \leq j-1$이면 $X_{(j)} \geq m$이며 $X_{(j)} = m$의 확률은 0이다. (c) $P(X_{(i)} < m < X_{(j)}) = P(i \leq K \leq j-1) = \sum_{k=i}^{j-1} \binom{n}{k} 2^{-n}$이다.

3절

연습 3.1 문제 16쪽

$2^{82} = 2^{80} \cdot 4 = (1024)^8 \cdot 4 \approx 4.836 \times 10^{24}$. H_0: 해마다 남아가 더 많을 확률이 $\frac{1}{2}$이고 해들이 독립. 검정통계량은 그런 해의 수 $K \sim \mathrm{b}(82, \frac{1}{2})$. 크기 0.05의 한쪽 기각역은 $\{K \geq 49\}$이다 ($P(K \geq 49) = 0.049$, $P(K \geq 48) = 0.075$). 관측은 $K = 82$.

연습 3.2 문제 16쪽

$U_i \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$로 $K_p = \sum 1\{U_i \leq p\}$를 만들면 $K_p \sim \mathrm{b}(n, p)$이고, $p < p'$이면 같은 U로 만든 $K_p \leq K_{p'}$이다. 따라서 $P(K_p \geq 59) \leq P(K_{p'} \geq 59)$. 검정력이 증가하므로 $\Omega_0 = \{p \leq \frac{1}{2}\}$ 위의 최댓값은 $p = \frac{1}{2}$에서이고 크기는 그대로 0.044다.

연습 3.3 문제 16쪽

$\bar{X} - \bar{Y}$의 분산이 $2\sigma^2/n$이므로 정리 3.4의 식에서 σ^2를 $2\sigma^2$로 바꿔 $n \geq 2\left(\frac{1.960+1.282}{0.3}\right)^2 = 233.5$, 집단당 234명.

연습 3.4 문제 16쪽

$(\hat{p} - p_0)^2 \leq \frac{z^2}{n} p_0(1-p_0)$을 p_0에 대해 정리하면 $(1 + \frac{z^2}{n})p_0^2 - (2\hat{p} + \frac{z^2}{n})p_0 + \hat{p}^2 \leq 0$. 근의 공식의 두 근이 예제 3.6의 끝점이다. $n = 20$, $\hat{p} = 0$이면 Wald는 $\{0\}$ (폭 0), Wilson은 $[0,\ 0.161]$.

연습 3.5 문제 16쪽

$H_0 : \mu \leq \mu_0$를 기각하지 않는 μ_0의 집합은 $\bar{X} < \mu_0 + z_\alpha \sigma/\sqrt{n}$, 곧 $\mu_0 > \bar{X} - z_\alpha \sigma/\sqrt{n}$이다. 따라서 $[\bar{X} - z_\alpha \sigma/\sqrt{n},\ \infty)$가 μ의 $100(1-\alpha)\%$ 아래쪽 한계다.

연습 3.6 문제 16쪽

(a) $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$이므로(6회 정리 2.2) $Z \sim \mathcal{N}(0,1)$이다. 기각 사건은 $\bar{X} - \mu \geq \mu_0 - \mu + z_\alpha \sigma/\sqrt{n}$, 곧 $Z \geq z_\alpha - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}$이므로 $\gamma(\mu) = 1 - \Phi\left(z_\alpha - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) = \Phi\left(\frac{\sqrt{n}(\mu-\mu_0)}{\sigma} - z_\alpha\right)$이다. (b) Φ가 증가함수이고 그 안의 식이 μ에 대해 증가하므로 γ는 증가한다. $\Omega_0 = \{\mu \leq \mu_0\}$에서 최댓값은

$\gamma(\mu_0) = \Phi(-z_\alpha) = \alpha$이다. (c) $\gamma(\mu_0+\delta) \geq 1-\beta = \Phi(z_\beta)$은 $\frac{\sqrt{n}\delta}{\sigma} - z_\alpha \geq z_\beta$와 같고, n에 대해 풀면 $n \geq \left(\frac{(z_\alpha+z_\beta)\sigma}{\delta}\right)^2$이다.

4절

연습 4.1 문제 19쪽

$p = P(|Z| \geq |z|) = 2(1-\Phi(|z|))$. H_0 아래 $|Z|$의 분포함수 $G(t) = 2\Phi(t)-1$은 연속이므로 정리 4.2로 $p = 1-G(|Z|)$는 균등하다.

연습 4.2 문제 19쪽

$p(k) = P(K \geq k)$는 $k = 0, \ldots, 20$에서 $1,\ 0.99999, \ldots, 0.0577\ (k=14),\ 0.0207\ (k=15),\ 0.0059, \ldots$의 21개 값만 가진다. $p \leq 0.05$는 $K \geq 15$와 같으므로 $P(p \leq 0.05) = 0.0207$. 0.05와 0.0207 사이의 값이 없기 때문에 정확히 0.05가 될 수 없다.

연습 4.3 문제 19쪽

$p = 1-\Phi(Z_{\text{obs}})$, $Z_{\text{obs}} = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim \mathcal{N}(\frac{\sqrt{n}\delta}{\sigma}, 1)$이다. $p \leq u \iff Z_{\text{obs}} \geq z_u$이므로 $P(p \leq u) = P\big(\mathcal{N}(0,1) \geq z_u - \frac{\sqrt{n}\delta}{\sigma}\big) = \Phi\big(\frac{\sqrt{n}\delta}{\sigma} - z_u\big)$.

연습 4.4 문제 19쪽

$p_i \sim \mathrm{U}(0,1)$이면 $P(-2\log p_i > x) = P(p_i < e^{-x/2}) = e^{-x/2}$이므로 $-2\log p_i \sim$ Exp(평균 2)$= \chi^2(2)$. 독립인 $\chi^2(2)$ k개의 합은 $\chi^2(2k)$(4회 정리 3.7).

5절

연습 5.1 문제 22쪽

$N_2 - np_2 = -(N_1 - np_1)$이므로 $X^2 = (N_1-np_1)^2\big(\frac{1}{np_1}+\frac{1}{np_2}\big) = \frac{(N_1-np_1)^2}{np_1p_2}$이고, 이것은 $\big(\frac{\hat{p}_1-p_1}{\sqrt{p_1p_2/n}}\big)^2$이다.

연습 5.2 문제 22쪽

$P' = I - \mathbf{u}\mathbf{u}' = P$. $P^2 = I - 2\mathbf{u}\mathbf{u}' + \mathbf{u}(\mathbf{u}'\mathbf{u})\mathbf{u}' = P$. $\mathrm{tr}P = k - \mathbf{u}'\mathbf{u} = k-1$. 고윳값은 $\mathbf{u}$ 방향에서 0, 수직인 $k-1$개 방향에서 1이다.

연습 5.3 문제 23쪽

$E_{11} = \frac{(a+b)(a+c)}{n}$이고 $a - E_{11} = \frac{ad-bc}{n}$이다. 네 칸 모두 차이의 절댓값이 $\frac{|ad-bc|}{n}$이므로 $X^2 = \frac{(ad-bc)^2}{n^2}\sum \frac{1}{E_{ij}}$이고, $\sum \frac{1}{E_{ij}} = \frac{n^3}{(a+b)(c+d)(a+c)(b+d)}$ (네 분수를 통분하면 분자가 $n \cdot n^2$). 정리하면 식이다. Berkeley 자료에서 92.2.

연습 5.4 문제 23쪽

$S = s$가 주어지면 X_i들은 확률이 같은 k개 범주에 s번을 나눈 다항분포이므로 Pearson 통계량 $\sum \frac{(X_i - s/k)^2}{s/k} = \sum \frac{(X_i - \bar{X})^2}{\bar{X}}$는 s가 클 때 $\chi^2(k-1)$이다. 조건부 분포가 λ와 무관하므로 무조건부로도 그렇다. 군단의 20년 합계 14개로 27.3, 자유도 13, $p = 0.011$이다.

연습 5.5 문제 23쪽

$X^2 = 3.84$: 자유도 1이면 $p = 0.050$, 3이면 0.279, 4이면 0.428, 8이면 0.871. 자유도를 크게 잡으면 같은 증거가 훨씬 약해 보인다.

6절

연습 6.1 문제 27쪽

$X^2 = 169.7$, 자유도 $(82-1)(2-1) = 81$, $p = 3 \times 10^{-8}$. $X^2/81 = 2.09$는 연도별 남아 비율의 분산이 이항 모형의 약 2배라는 뜻이고, 2회의 Lexis 비(표준편차의 비 1.5, 분산으로 2.25)와 같은 것을 다른 방법으로 잰 것이다.

연습 6.2 문제 27쪽

$0.3377 \pm 1.96\sqrt{0.3377 \cdot 0.6623/315{,}672} = [0.3360,\ 0.3393]$이고 $\frac{1}{3} = 0.3333$은 밖에 있다. 10개 이상을 10으로 묶으면 11개와 12개가 10으로 세어지므로 합계, 따라서 $\hat{p}$가 조금 작아진다(그런 굴림은 18번뿐이어서 영향은 작다).

연습 6.3 문제 27쪽

1천 번의 시뮬레이션에서 기각 확률은 정규 0.051, 왜도 0.05에서 0.12, 0.1에서 0.42, 0.2에서 0.99다. $N = 5{,}738$의 검정은 왜도 0.1 정도의 어긋남을 절반 가까이 잡는다. 44인치 칸의 기여 19.7을 빼면 $X^2 \approx 15.3$, 남은 12칸의 자유도 9에서 $p \approx 0.08$로 기각하지 않는다. 기각은 거의 그 한 칸 때문이다.

연습 6.4 문제 27쪽

층 s에서 $a_s d_s/n_s$와 $b_s c_s/n_s$의 기댓값의 비가 대략 그 층의 오즈비 ψ이므로, 합의 비도 ψ의 가중평균, 곧 ψ에 가깝다. 층마다 표본이 작아도 분자와 분모를 따로 더하므로 안정적이다. Berkeley 자료에서 $\frac{\sum a_s d_s/n_s}{\sum b_s c_s/n_s} = 0.90$.

7절

연습 7.1 문제 30쪽

$1-(1-\alpha)^m$. $m=20$에서 0.05 이하로 하려면 $\alpha \le 1-0.95^{1/20} = 0.00256$(Šidák). Bonferroni의 $0.05/20 = 0.0025$와 거의 같다.

연습 7.2 문제 30쪽

10개마다 들여다보며 최대 100, 1,000, 10,000개까지 모으면 거짓 양성의 확률은 0.19, 0.37, 0.53이다. 들여다보는 횟수의 로그에 대략 비례해 자라고, 반복 로그의 법칙으로 결국 1이 된다.

연습 7.3 문제 30쪽

포함확률은 $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}\{p \in [\hat{p}_k \pm 1.96\sqrt{\hat{p}_k(1-\hat{p}_k)/n}]\}$이다. np가 작으면 K는 대개 0, 1, 2다. $K=0$이면 구간이 $\{0\}$이어서 놓치고, $K=1,2$이면 구간의 위쪽 끝 $\frac{k}{n} + 1.96\sqrt{k(1-k/n)}/n$이 작은 p보다 크고 아래쪽 끝은 0 근처이므로 대개 포함한다. 그래서 포함확률은 $P(K \ge 1) = 1-(1-p)^n$ 근처다. $n=50$, $p=0.005$이면 $1-0.995^{50} = 0.222$로 정확한 값 0.22와 같다.

연습 7.4 문제 30쪽

$T_0, \dots, T_B$가 iid 연속이면 교환가능하므로 보조정리 8.1로 T_0의 위에서부터의 순위 $R+1$ ($R = T_0$보다 큰 T_b의 수)은 $\{1, \dots, B+1\}$에서 균등하다. $P\left(\frac{R+1}{B+1} \le \alpha\right) = \frac{\lfloor \alpha(B+1) \rfloor}{B+1} \le \alpha$.

8절

연습 8.1 문제 33쪽

$F(Y) = \Phi(Y/0.5) = \Phi(2Y)$이고 $2Y \sim \mathcal{N}(0,4)$이므로 $F(Y)$는 0과 1 근처에 몰린다(U자). 예측분포가 너무 좁으면 관측이 예측의 꼬리에 자주 떨어진다. $\mathcal{N}(0, 2^2)$이면 $\Phi(Y/2)$가 0.5 근처에 몰린다(혹 모양).

연습 8.2 문제 33쪽

점수가 연속이면 순위가 정확히 균등하므로 포함확률은 정확히 $\frac{k}{n+1}$이고, $k = \lceil (n+1)(1-\alpha) \rceil < (n+1)(1-\alpha)+1$이므로 $\frac{k}{n+1} < 1-\alpha+\frac{1}{n+1}$. $n=19$, $\alpha=0.1$이면 $k=18$이고 포함확률은 정확히 0.9다.

연습 8.3 문제 33쪽

독립이면 하나도 기각하지 않을 확률이 $(1-\alpha/m)^m$이므로 가족별 오류율은 $1-(1-0.0025)^{20} = 0.0488$. Bonferroni 한계 0.05와 거의 같다. 독립에서도 한계가 거의 정확하다.

연습 8.4 문제 33쪽

기준선 $\frac{i}{10}\cdot 0.05 = 0.005, 0.010, 0.015, 0.020, 0.025, \ldots$와 비교하면 $p_{(4)} = 0.02 \le 0.02$이고 $i \ge 5$에서는 모두 넘으므로 $k=4$, 네 개를 기각한다. Bonferroni(0.005)는 둘(0.001, 0.004)이다.

연습 8.5 문제 34쪽

올바른 사후분포 $\mathcal{N}(X/2, \frac{1}{2})$이면 θ와 사후표본 19개는 (X가 주어질 때) 같은 분포에서 나온 교환가능한 20개이므로 순위가 균등하다(보조정리 8.1). 사후분산을 $\frac{1}{4}$로 줄이면 사후표본이 너무 모여 있어 θ가 양 끝 순위(1이나 20)에 자주 떨어진다. 히스토그램이 U자가 된다.

연습 8.6 문제 34쪽

(a) σ를 j와 $n+1$을 맞바꾸는 순열이라 하자. 교환가능성으로 $(S_{\sigma(1)}, \ldots, S_{\sigma(n+1)})$은 $(S_1, \ldots, S_{n+1})$과 분포가 같다. R_j는 $(S_1, \ldots, S_{n+1})$의 함수이고, 맞바꾼 벡터에 같은 함수를 쓰면 R_{n+1}이 된다. 따라서 R_j와 R_{n+1}은 분포가 같다. (b) 값이 모두 다르면 $S_1, \ldots, S_{n+1}$을 작은 것부터 늘어놓을 때 r번째 자리에 정확히 하나가 있고, 그 하나만 $R_j = r$이다. 사건 $\{R_j = r\}$ $(j = 1, \ldots, n+1)$은 확률 0인 사건을 빼면 쌍마다 서로소이고 합집합이 전체이므로 확률의 합이 1이다. (c) (a)로 $n+1$개의 확률이 모두 $P(R_{n+1}=r)$이므로 $(n+1)P(R_{n+1}=r) = 1$이다. (d) 같은 값이 있으면 r번째 자리에 둘 이상이 있거나 아무도 없을 수 있다. 모든 S_i가 같으면 모든 $R_j = n+1$이고, R_{n+1}은 균등하지 않다.