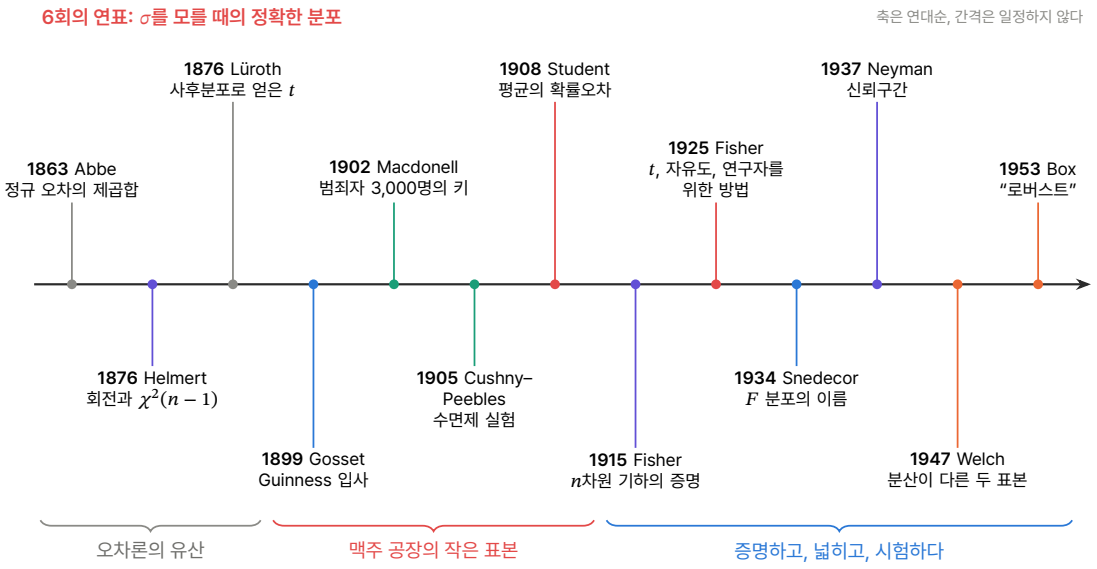
수리통계학 × 통계학의 역사

6회 강의노트

Student의 t 작은 표본의 정확한 분포

1876 → 1925

더블린의 맥주 회사 Guinness의 화학자 Gosset은 표본이 너덧 개뿐인 실험으로 결정을 내려야 했다. 모집단의 표준편차를 모르니 표본의 표준편차로 대신했는데, 그러면 정규분포가 말하는 것보다 자주 틀렸다. 1908년 그는 "Student"라는 이름으로 그 정확한 분포를 내놓았다. 이 노트는 표본평균과 표본분산이 독립이라는 Student의 정리를 증명하고, t와 F 분포를 유도해 작은 표본의 신뢰구간과 검정을 세운다.



다루는 것	표본분산과 자유도 · Helmert 변환 · Student의 정리 · t 분포 · F 분포 · 축량과 t-구간 · 대응·두 표본 t · 비정규에서의 포함확률
원자료	Macdonell(1902)의 범죄자 3,000명 · Cushny–Peebles(1905)의 수면제 · Student(1908)의 카드 실험
노트북	`sessions/06_student_t/06_student_t.ipynb` — 절·정리 번호가 같다
워크북	`06_student_t_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1–5회 (특히 4회 감마·χ^2, 5회 다변량 정규분포의 정리 A·B)

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 앞 회차의 정리)를 단다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 적분의 순서나 극한을 바꾸는 곳에는 **조건 확인** 단락이 붙는다. 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 [T4](더하고 빼기, Cauchy–Schwarz), [T5](꼬리와 비교 판정), [T6](Tonelli), [T7](변수변환), [T8](*O* 표기)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 1–5회에서 가져오는 것 2

0.2 표기 3

1 문제 — σ를 모를 때 4

1.1 큰 표본의 관행과 작은 표본의 함정 4

1.2 표본분산은 왜 $n-1$로 나누나 4

2 Student의 정리 — 평균과 분산은 독립이다 7

2.1 Helmert 변환 7

2.2 Student의 정리 7

3 t 분포 11

3.1 정의와 밀도 11

3.2 성질 12

3.3 Student의 t 13

4 F 분포 — 두 분산의 비 15

5 축량과 t-구간 18

5.1 축량 18

5.2 대응 비교 — Cushny와 Peebles의 수면제 (1905) 19

5.3 두 독립 표본 19

6 데이터로 — Student의 카드 실험 (1908) 22

6.1 모집단 22

6.2 재현 22

6.3 유한 모집단 24

7 부숴보기 — 정규가 아니고, 독립이 아니면 25

7.1 비대칭 모집단 . . . 25

7.2 이상값 하나 . . . 26

7.3 의존하는 관측 . . . 27

7.4 분산의 비교는 정규성에 민감하다 . . . 27

8 Beautiful insights — 작은 표본의 분포, 오늘의 문제 29

8.1 예측구간 — 다음 한 사람 . . . 29

8.2 작은 배치의 BatchNorm . . . 29

8.3 t-SNE — 지도를 그릴 때 꼬리가 필요하다 . . . 31

8.4 꼬리가 두꺼운 가능도 — 로버스트 회귀 . . . 31

8.5 같은 구간, 다른 해석 . . . 32

9 정리와 역사 노트 34

9.1 한 장 요약 . . . 34

9.2 Stigler의 명명 법칙 . . . 34

9.3 7 Pillars 지도 . . . 35

9.4 다음 회 — 극한의 언어: 수렴 · CLT · Monte Carlo (1777 → 1949) . . . 35

A 연습문제 풀이 37

0

들어가며

> **여는 질문.** 맥주 원료인 보리의 품질을 네 번 쟀다. 모집단의 표준편차는 모른다. 네 값만으로 평균의 오차를 말할 수 있는가? 말할 수 있다면, 표준편차 자리에 표본의 표준편차를 넣고 정규분포를 써도 되는가?

19세기의 천문학자와 측지학자들은 관측이 수십, 수백 개였다. 그들은 평균의 오차를 정규분포로 말하면서 모르는 σ 자리에 관측에서 계산한 값을 넣었고, 관측이 많으니 그래도 별 문제가 없었다. 20세기 초 더블린의 맥주 회사 Guinness에 들어간 화학자 William Sealy Gosset의 사정은 달랐다. 보리 품종 시험, 맥아 실험, 효모 세포 수(4회)는 한 번에 몇 개의 측정밖에 얻을 수 없었다. 네 값의 표준편차는 σ에서 크게 빗나갈 수 있고, 그것을 σ처럼 쓰면 평균의 오차를 체계적으로 작게 말하게 된다.

Gosset은 1906–07년 런던의 Karl Pearson 연구실에서 이 문제를 풀었다. 회사가 직원의 실명 발표를 막았으므로 1908년 *Biometrika*에 "Student"라는 이름으로 *The probable error of a mean*을 냈다. 그는 표본평균과 표본분산이 독립이라고 (증명 없이) 놓고 분포를 유도했고, 그 결과를 확인하려고 범죄자 3,000명의 키를 카드에 적어 섞고 넷씩 750묶음을 뽑는 실험을 손으로 했다(6절). 엄밀한 증명은 1915년과 1925년 R. A. Fisher가 했고, 오늘 쓰는 t라는 형태와 이름도 Fisher의 것이다.

이 노트는 그 증명을 따라간다. 표본분산과 자유도(1절), 정규 표본에서 평균과 분산이 독립이라는 Student의 정리(2절), t 분포와 F 분포(3–4절), 축량으로 만드는 신뢰구간과 대응 비교(5절), 그리고 Student의 카드 실험의 재현(6절)이다. 7절은 정규성이 깨질 때 t-구간이 얼마나 틀리는지 보고, 8절은 작은 배치의 정규화(BatchNorm)와 t-SNE에서 같은 분포를 만난다.

학습 목표

1. 표본분산이 σ^2의 불편추정량임을 보이고, 나누는 수가 $n-1$인 이유를 자유도로 설명한다.
2. Helmert 변환으로 정규 표본의 $\bar{X} \perp S^2$와 $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$을 증명한다.
3. t와 F의 밀도를 변수변환으로 유도하고, 꼬리·적률·정규 극한을 보인다.
4. 축량으로 t-구간을 만들고 포함확률이 정확히 $1-\alpha$임을 증명한다. 대응 t와 두 표본 t를 쓴다.
5. Student(1908)의 카드 실험을 원자료로 재현하고, 비정규 모집단에서 t-구간의 실제 포함확

률을 잰다.

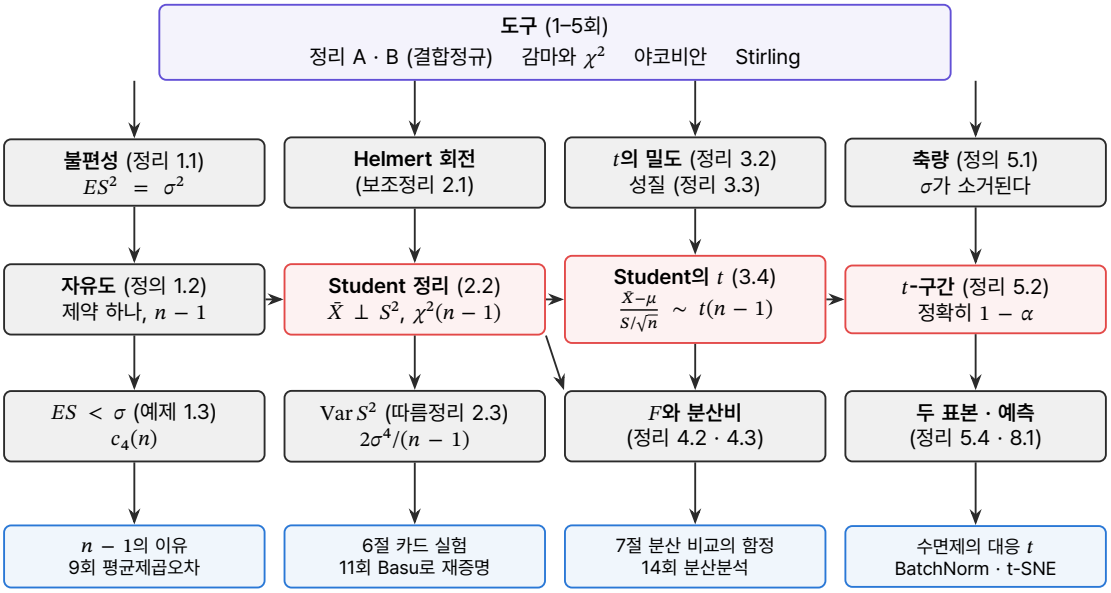


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–5회에서 가져오는 것

앞 회차	내용
2회 정리 2.2 · 3회 정리 4.2	분산의 계산, 공분산의 쌍선형성
3회 정리 5.1 · 정리 3.2	야코비안 변수변환, 인수분해로 독립 판정
3회 정리 5.8	표준정규의 비는 Cauchy
4회 정리 3.5 · 3.7 · 3.9	감마함수, 감마의 합, $Z_1^2 + \cdots + Z_r^2 \sim \chi^2(r)$
4회 정리 4.2 · 연습 4.4	감마의 비 $\frac{X}{X+Y}$는 베타이고 $X+Y$와 독립, X/Y의 밀도
5회 정리 3.8 · 3.11 · 따름정리 3.12	정리 A: 다변량 정규의 선형변환은 정규. 정리 B: 결합정규에서는 무상관 $\iff$ 독립, 공분산행렬이 대각이면 성분이 상호 독립
5회 보조정리 1.1	Stirling 공식 $\Gamma(x+1) \sim \sqrt{2\pi x}\,(x/e)^x$

0.2 표기

$X_1, \dots, X_n$이 표본이면 표본평균과 표본분산을

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

로 쓰고 $S = \sqrt{S^2}$를 표본표준편차라 한다. $t_{r,p}$는 자유도 r인 t 분포의 위쪽 p 분위수, 곧 $P(T > t_{r,p}) = p$인 수다. z_p는 표준정규의 같은 분위수다($z_{0.025} = 1.960$).

1

문제 — σ를 모를 때

1.1 큰 표본의 관행과 작은 표본의 함정

> **역사** Gosset은 1899년 옥스퍼드에서 화학과 수학을 공부하고 Guinness에 들어갔다. 회사는 원료와 공정을 과학적으로 관리하려고 젊은 대학 졸업자들을 뽑고 있었다. 보리 품종을 비교하는 밭 실험, 맥아의 수분, 홉의 수지 함량 같은 측정은 비싸고 느려서 한 조건에 몇 번밖에 잴 수 없었다. 당시의 표준적인 방법은 평균 $\bar{x}$에 "확률오차" $0.6745\,s/\sqrt{n}$을 붙이는 것이었다. 이것은 σ를 아는 정규 이론에서 나온 수이고, 관측이 많을 때만 정당했다.

$X_1, \dots, X_n$이 $\mathcal{N}(\mu, \sigma^2)$에서 나온 독립 표본이면 $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$이므로(5회 정리 1.8(iv))

$$P\left(|\bar{X} - \mu| \le 1.96\,\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

σ를 모르면 자연스럽게 S를 넣는다. 그러면 무엇이 달라지나? 이 노트의 결론(정리 3.4)을 미리 쓰면

$$P\left(|\bar{X} - \mu| \le 1.96\,\frac{S}{\sqrt{n}}\right) = P\big(|T_{n-1}| \le 1.96\big) = \begin{cases} 0.700, & n = 2, \\ 0.855, & n = 4, \\ 0.918, & n = 10, \\ 0.940, & n = 30. \end{cases}$$

네 개로 만든 "95% 구간"은 실제로는 85.5%만 맞는다. 작은 S가 우연히 나오면 구간이 너무 좁아지기 때문이다. σ를 모른다는 불확실성이 하나 더 있는데, 정규 이론은 그것을 무시한다. Student의 기여는 그 불확실성까지 넣은 **정확한** 분포를 준 것이다(그림 1.1).

1.2 표본분산은 왜 $n-1$로 나누나

> **정리 1.1** (표본분산의 기댓값) $X_1, \dots, X_n$이 독립이고 평균 μ, 분산 σ^2이면 (분포에 상관없이)
>
> $$E\sum_{i=1}^{n}(X_i - \bar{X})^2 = (n-1)\sigma^2, \qquad \text{따라서} \quad ES^2 = \sigma^2.$$

증명. **전략** 참평균 μ에서 잰 편차와 $\bar{X}$에서 잰 편차의 차이를 정확히 계산한다. μ를 빼고 $\bar{X}$를 더하고 빼는 것이다.

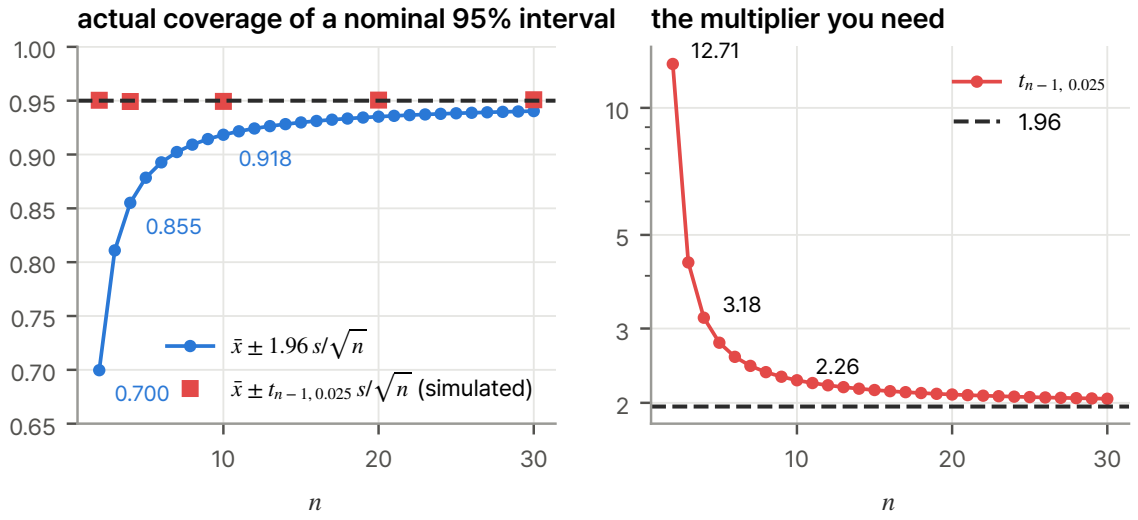


그림 1.1 왼쪽: $\bar{x} \pm 1.96\, s/\sqrt{n}$의 실제 포함확률(파랑)은 95%에 미치지 못한다. t 분위수를 쓰면(빨강) 모든 n에서 정확히 95%다. 오른쪽: 필요한 분위수 $t_{n-1,\,0.025}$는 n이 작을수록 1.96보다 크다.

① $X_i - \mu = (X_i - \bar{X}) + (\bar{X} - \mu)$로 쪼갠다. 더하고 빼기 [T4]

② 제곱해 i에 대해 더한다.

$$\sum_i (X_i - \mu)^2 = \sum_i (X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum_i (X_i - \bar{X}) + n(\bar{X} - \mu)^2.$$

③ 가운데 항은 0이다. $\sum_i (X_i - \bar{X}) = \sum_i X_i - n\bar{X} = 0$이기 때문이다. $\bar{X}$의 정의

④ 따라서 $\sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \mu)^2 - n(\bar{X} - \mu)^2$이다. 2회 정리 2.2(c)를 표본에 쓴 것과 같다. ②③

⑤ 기댓값을 취한다.

$$\begin{aligned} E\sum_i (X_i - \bar{X})^2 &= \sum_i E(X_i - \mu)^2 - n\,E(\bar{X} - \mu)^2 && \text{기댓값의 선형성} \\ &= n\sigma^2 - n \cdot \frac{\sigma^2}{n} && \operatorname{Var}\bar{X} = \sigma^2/n,\ \text{3회 정리 6.3} \\ &= (n-1)\sigma^2. \end{aligned}$$

조건 확인 분산이 유한하므로 ⑤의 기댓값은 모두 유한하다. 정규성은 쓰지 않았다. 독립은 $\operatorname{Var}\bar{X} = \sigma^2/n$에서만 쓰였고, 그 자리에는 서로 무상관이면 충분하다. ■

n으로 나누면 평균적으로 $\frac{n-1}{n}\sigma^2$, 곧 작게 나온다. 참평균 μ 대신 자료에 가장 가깝게 맞춘 $\bar{X}$에서 편차를 쟀기 때문이다. $\bar{X}$는 편차제곱합을 가장 작게 하는 점이므로(2회 정리 2.2(c)) $\sum (X_i - \bar{X})^2 \le \sum (X_i - \mu)^2$이 언제나 성립한다. 그 차이가 평균 σ^2, 정확히 "관측 하나만큼"이다.

정의 1.2 (자유도) 편차 $X_1-\bar X,\dots,X_n-\bar X$는 합이 언제나 0이라는 **한 개의 선형 제약**을 받는다. 그래서 n개 가운데 $n-1$개만 자유롭게 움직인다. 편차제곱합이 가진 이 독립적인 방향의 수 $n-1$을 **자유도**라 한다.

정의는 아직 직관이다. 2절에서 정규 표본의 경우 이것이 정확한 뜻을 얻는다. 편차벡터는 $n-1$차원 부분공간에 놓이고, 편차제곱합은 독립인 정규 $n-1$개의 제곱합, 곧 $\chi^2(n-1)$이다. 13회에서는 회귀계수를 p개 맞추면 자유도가 $n-p$가 된다.

예제 1.3 (S는 σ를 작게 어림한다) $ES^2=\sigma^2$이지만 $ES<\sigma$이다. Jensen 부등식(2회 정리 5.9, $\sqrt{\cdot}$는 오목)으로 $ES=E\sqrt{S^2}\le\sqrt{ES^2}=\sigma$이고, S가 상수가 아니므로 등호가 아니다. 정규 표본이면 연습 2.4에서 $ES=c_4(n)\sigma$, $c_4(4)\approx0.921$, $c_4(10)\approx0.973$을 구한다. 네 개로는 σ를 평균 8% 작게 어림한다. 그리고 작게 어림할 때 구간은 짧아지고 틀리기 쉽다.

연습문제

연습 1.1 (표본분산의 다른 꼴 ★) $\sum(X_i-\bar X)^2=\sum X_i^2-n\bar X^2=\frac{1}{2n}\sum_{i,j}(X_i-X_j)^2$임을 보여라. 마지막 식으로 $ES^2=\sigma^2$을 다시 보여라(2회 연습 2.8). 풀이 37쪽

연습 1.2 (표본분산의 분산 ★★★) $\mu_4=E(X-\mu)^4$이 유한하면 $\operatorname{Var}S^2=\frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\sigma^4\right)$임을 보여라. 정규이면 $\mu_4=3\sigma^4$이므로 $\frac{2\sigma^4}{n-1}$이다. 꼬리가 무거운 분포(큰 μ_4)에서 S^2가 크게 흔들리는 이유다. 풀이 37쪽

연습 1.3 (두 관측의 분산 ★) $n=2$이면 $S^2=\frac{1}{2}(X_1-X_2)^2$임을 보여라. $X_1-X_2\sim\mathcal{N}(0,2\sigma^2)$을 써서 $S^2/\sigma^2\sim\chi^2(1)$을 직접 확인하라. 풀이 37쪽

연습 1.4 (n으로 나눈 추정량 ★★) 정규 표본에서 $\hat\sigma_c^2=c\sum(X_i-\bar X)^2$의 평균제곱오차 $E(\hat\sigma_c^2-\sigma^2)^2$을 가장 작게 하는 c는 $\frac{1}{n+1}$임을 보여라($\operatorname{Var}S^2=\frac{2\sigma^4}{n-1}$을 쓴다). 불편성과 평균제곱오차는 다른 기준이다(9회, 16회). 풀이 38쪽

2

Student의 정리 — 평균과 분산은 독립이다

> **역사** 독일의 측지학자 Friedrich Robert Helmert는 1876년 정규 관측오차의 제곱합의 분포를 구하면서, 관측을 직교변환해 $\bar{X}$에 해당하는 성분을 떼어 내는 방법을 썼다. 그는 남은 제곱합이 오늘의 $\chi^2(n-1)$ 분포를 따른다는 것을 보였다. 이 결과는 독일 바깥에 알려지지 않아서, Student(1908)는 같은 사실을 추측으로 놓고 썼고 Fisher(1915)가 n차원 기하로 다시 증명했다.

2.1 Helmert 변환

보조정리 2.1 (Helmert 행렬) $n\times n$ 행렬 H의 첫 행을 $\frac{1}{\sqrt{n}}(1,1,\dots,1)$로, k번째 행$(k=2,\dots,n)$을

$$\frac{1}{\sqrt{k(k-1)}}\big(\underbrace{1,\dots,1}_{k-1},\ -(k-1),\ 0,\dots,0\big)$$

로 두면 H는 직교행렬이다: $HH'=I$.

증명. **전략** 행마다 길이가 1이고 서로 다른 두 행의 내적이 0임을 확인한다. 첫 행과의 내적은 성분의 합에 비례하고, $j<k$인 두 행의 내적은 j번째 행의 성분의 합에 비례한다. 단계별 증명은 연습 2.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 39쪽에 있다.

행렬의 모양보다 중요한 것은 두 사실이다. 첫 행이 $\mathbf{1}/\sqrt{n}$이고, 나머지 행들은 모두 $\mathbf{1}$과 직교한다. 그런 직교행렬이면 무엇이든 아래 증명에 쓸 수 있다.

2.2 Student의 정리

정리 2.2 (Student의 정리 (Helmert 1876, Fisher 1915)) $X_1,\dots,X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma^2)$이면

(a) $\bar{X}\sim\mathcal{N}(\mu,\sigma^2/n)$,

(b) $\bar{X}$와 S^2는 독립이다,

(c) $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

증명. **전략** 표본을 표준화하고 직교행렬로 돌린다. 돌린 좌표에서 첫 성분은 평균이고, 나머지 성분의 제곱합은 편차제곱합이며, 좌표들은 서로 독립인 표준정규다.

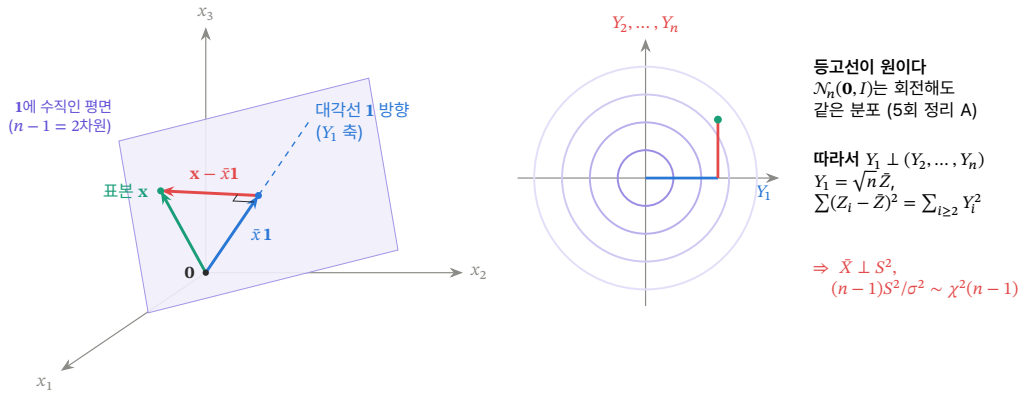


그림 2.1 표본 $\mathbf{x}$를 대각선 방향 $\mathbf{1}$으로 사영하면 $\bar{x}\mathbf{1}$이고, 남는 잔차 $\mathbf{x}-\bar{x}\mathbf{1}$은 대각선에 수직인 $n-1$차원 공간에 놓인다. Helmert 변환은 좌표축을 돌려 첫 축을 대각선에 맞춘 것이다. 정규 표본은 회전해도 독립인 정규이므로, 대각선 성분($\bar{X}$)과 수직 성분(잔차)이 독립이다.

① $Z_i = (X_i-\mu)/\sigma$로 두면 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$이고 $\bar{X} = \mu + \sigma\bar{Z}$, $(n-1)S^2 = \sigma^2 \sum_i (Z_i-\bar{Z})^2$이다. 따라서 $\mathbf{Z}$에 대해 보이면 된다. — 표준화

② 보조정리 2.1의 H로 $\mathbf{Y} = H\mathbf{Z}$라 하면 $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, HIH') = \mathcal{N}_n(\mathbf{0}, I)$이다. — 5회 정리 3.8(정리 A)

③ 공분산행렬이 대각이므로 $Y_1, \dots, Y_n$은 독립인 표준정규다. — 5회 따름정리 3.12

④ 첫 성분은 평균이다: $Y_1 = \frac{1}{\sqrt{n}} \sum_i Z_i = \sqrt{n}\,\bar{Z}$. — H의 첫 행

⑤ 나머지 성분의 제곱합은 편차제곱합이다.

$$\sum_{i=2}^{n} Y_i^2 = \|H\mathbf{Z}\|^2 - Y_1^2 = \|\mathbf{Z}\|^2 - n\bar{Z}^2 \qquad \text{직교변환은 길이를 보존, ④}$$

$$= \sum_i (Z_i - \bar{Z})^2 \qquad \text{정리 1.1의 증명, } \mu = 0$$

⑥ $\bar{Z}$는 Y_1만의 함수이고 $\sum_i (Z_i-\bar{Z})^2$는 $(Y_2, \dots, Y_n)$만의 함수이므로 둘은 독립이다. 이것이 (b)다. — ③, 3회 3.2절의 상자

⑦ $Y_1 \sim \mathcal{N}(0,1)$이므로 $\bar{Z} = Y_1/\sqrt{n} \sim \mathcal{N}(0, \frac{1}{n})$이고, 이것이 (a)다. — ③④

⑧ $\sum_{i\geq 2} Y_i^2$는 독립인 표준정규 $n-1$개의 제곱합이므로 $\chi^2(n-1)$이고, 이것이 (c)다. — ③, 4회 정리 3.9

■

증명이 말하는 것은 기하다(그림 2.1). 표본은 n차원 공간의 한 점이고, $\bar{X}$는 그 점의 대각선 방향 성분, S^2는 대각선에서의 거리의 제곱이다. 표준정규 벡터의 분포는 원점을 중심으로 **회전에 불변**이므로(5회), 서로 수직인 방향의 성분들은 독립이다. 자유도 $n-1$은 대각선에 수직인 공간의 차원이다.

증명 없이 쓰는 사실 (독립은 정규분포만의 성질이다) $\bar X$와 S^2가 독립인 것은 정규 표본에서만 일어난다. 독립이면 모집단이 정규라는 역(Geary 1936, Lukacs 1942)이 성립한다. 증명은 특성함수의 미분방정식을 쓴다. 예를 들어 지수 표본에서는 $\bar X$가 클 때 S^2도 크다(연습 2.3). 11회에서 Basu의 정리로 (b)를 다시 증명한다.

따름정리 2.3 (정규 표본에서 S^2의 적률) 정리 2.2의 조건에서 $ES^2=\sigma^2$, $\operatorname{Var} S^2=\dfrac{2\sigma^4}{n-1}$이다.

증명. $\chi^2(r)$의 평균 r, 분산 $2r$ (4회 정의 3.8)에 $S^2=\dfrac{\sigma^2}{n-1}\chi^2(n-1)$을 넣는다. ∎

S^2의 상대 표준편차는 $\sqrt{2/(n-1)}$이다. $n=4$이면 82%다. 네 개로 잰 분산은 참값의 두 배나 절반이 흔하다. 이 흔들림이 1절의 포함확률 손실의 원인이다.

연습문제

연습 2.1 ($n=3$의 Helmert ★) $n=3$의 Helmert 행렬을 쓰고 $HH'=I$를 확인하라. $Y_2^2+Y_3^2=\sum_i (Z_i-\bar Z)^2$을 직접 전개해 보여라. 풀이 38쪽

연습 2.2 (잔차와 평균의 무상관 ★) 분포와 상관없이 $\operatorname{Cov}(\bar X,\ X_i-\bar X)=0$임을 보여라(3회 정리 4.2). 정규 표본이면 이것과 5회 정리 B로 $\bar X$와 잔차벡터 $(X_1-\bar X,\dots,X_n-\bar X)$가 독립임을 결론지어라. 정리 2.2(b)의 둘째 증명이다. 풀이 38쪽

연습 2.3 (지수 표본에서는 의존한다 ★★) $X_1,X_2\overset{\text{iid}}{\sim}\text{Exp}(1)$이면 $\bar X=\dfrac{X_1+X_2}{2}$, $S^2=\dfrac{(X_1-X_2)^2}{2}$이다. $\operatorname{Cov}(\bar X,S^2)$를 계산해 0이 아님을 보여라. (힌트: $E(X_1^2X_2)=EX_1^2\,EX_2$, $EX^3=6$.) 풀이 38쪽

연습 2.4 (표본표준편차의 평균 ★★) 정규 표본에서 $ES=c_4(n)\sigma$, $c_4(n)=\sqrt{\dfrac{2}{n-1}}\,\dfrac{\Gamma(n/2)}{\Gamma((n-1)/2)}$임을 보여라. (힌트: $V\sim\chi^2(r)$이면 $E\sqrt{V}$를 감마 밀도의 적분으로 구한다. 4회 연습 3.4.) $c_4(4)$와 $c_4(10)$을 계산하라. 풀이 39쪽

연습 2.5 (신뢰구간의 씨앗 — σ^2 ★★) 정리 2.2(c)로 σ^2에 대한 95% 구간 $\left[\dfrac{(n-1)S^2}{\chi^2_{n-1,0.025}},\ \dfrac{(n-1)S^2}{\chi^2_{n-1,0.975}}\right]$을 만들고, 포함확률이 정확히 0.95임을 보여라($\chi^2_{r,p}$는 위쪽 p 분위수). $n=4$, $s^2=1$이면 구간은? 풀이 39쪽

연습 2.6 (Helmert 행렬의 직교성 증명 ★ 유도형) 보조정리 2.1를 다음 순서로 증명하라.

(a) 첫 행과 k번째 행($k\ge 2$)의 길이가 1임을 보여라.

(b) 첫 행과 k번째 행의 내적이 k번째 행의 성분의 합에 비례하고, 그 합이 0임을 보여라.

(c) $2\le j<k$이면 j번째 행의 0이 아닌 성분은 처음 j개이고, 그 자리에서 k번째 행의

성분은 모두 같다. 이것으로 두 행의 내적이 0임을 보여라.

(d) 정사각행렬에서 $HH' = I$이면 $H'H = I$임을 밝히고, 이것으로 모든 벡터 $\mathbf{z}$에서 $\|H\mathbf{z}\| = \|\mathbf{z}\|$임을 보여라. 정리 2.2의 증명에서 이 사실을 어디에 쓰는가?

풀이 39쪽

3

t 분포

3.1 정의와 밀도

정의 3.1 (t 분포) $Z \sim \mathcal{N}(0,1)$과 $V \sim \chi^2(r)$이 독립일 때

$$T = \frac{Z}{\sqrt{V/r}}$$

의 분포를 자유도 r인 t **분포** $t(r)$이라 한다.

분모의 $\sqrt{V/r}$은 평균이 1 근처인 양수이고($E(V/r) = 1$), r이 작으면 크게 흔들린다. 분모가 우연히 작아지면 T가 크게 튄다. 그래서 t의 꼬리가 정규보다 무겁다.

정리 3.2 (t 분포의 밀도)

$$f_r(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)}\left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \qquad t \in \mathbb{R}.$$

증명. **전략** 보조 변수를 하나 붙여 $(Z, V) \mapsto (T, W)$의 변수변환으로 결합밀도를 구하고, 보조 변수를 적분해 없앤다 [T7].

① $Z \perp V$이므로 결합밀도는 두 밀도의 곱 $\phi(z) \cdot \frac{v^{r/2-1}e^{-v/2}}{\Gamma(r/2)2^{r/2}}$이다($z \in \mathbb{R}$, $v > 0$). 독립, 4회 정의 3.8

② 보조 변수 $W = V$를 둔다. $(z, v) \mapsto (t, w) = (z/\sqrt{v/r},\, v)$는 $\mathbb{R} \times (0,\infty)$를 자기 자신 위로 일대일로 보내고, 역변환은 $z = t\sqrt{w/r}$, $v = w$이다. 역변환 [T7]

③ 역변환의 편미분 행렬은 위삼각이고 대각 성분이 $\sqrt{w/r}$와 1이므로 야코비안은 $\sqrt{w/r}$이다. 삼각행렬의 행렬식

④ 변수변환 공식으로 결합밀도를 얻는다.

$$f_{T,W}(t,w) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2 w}{2r}} \cdot \frac{w^{r/2-1}e^{-w/2}}{\Gamma(r/2)2^{r/2}} \cdot \sqrt{\frac{w}{r}}.$$ 3회 정리 5.1

⑤ w로 적분한다. w의 지수는 $\frac{r+1}{2} - 1$이고 지수함수는 $e^{-w(1+t^2/r)/2}$이므로

$$\int_0^\infty w^{\frac{r+1}{2}-1} e^{-w\frac{1+t^2/r}{2}}\, dw = \Gamma\left(\frac{r+1}{2}\right)\left(\frac{2}{1+t^2/r}\right)^{\frac{r+1}{2}}.$$ 감마 적분, 4회 정의 3.4

⑥ 상수를 모으면 $\dfrac{2^{(r+1)/2}}{\sqrt{2\pi}\sqrt{r}\,2^{r/2}} = \dfrac{1}{\sqrt{r\pi}}$이고 결론이 나온다. 계산

조건 확인 $f_{T,W} \geq 0$이므로 w로 적분한 것이 T의 주변밀도라는 것은 Tonelli 정리로 조건 없이 성립한다 [T6]. ⑤의 감마 적분은 $\frac{r+1}{2} > 0$이고 $1 + t^2/r > 0$이므로 모든 t에서 유한하다. ∎

3.2 성질

> **정리 3.3** (t 분포의 성질) $T \sim t(r)$이면
>
> (a) 분포는 0에 대해 대칭이다.
>
> (b) $E|T|^k < \infty \iff k < r$. 따라서 $r > 1$이면 $ET = 0$, $r > 2$이면 $\operatorname{Var} T = \dfrac{r}{r-2}$.
>
> (c) $r = 1$이면 Cauchy 분포다.
>
> (d) $r \to \infty$이면 모든 t에서 $f_r(t) \to \phi(t)$.

증명. **전략** (a)와 (c)는 밀도의 모양에서 바로 나온다. (b)는 꼬리의 차수를 비교 판정에 넣고 [T5], 분산은 $T = Z/\sqrt{V/r}$의 독립인 두 인수의 기댓값으로 구한다. (d)는 밀도의 두 인수의 극한을 따로 구한다 [T8]. 단계별 증명은 연습 3.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

(d)의 직관은 이렇다. V/r은 $\chi^2(r)/r$이고, 이것은 독립인 Z_i^2 r개의 평균이므로 대수의 법칙(2회 정리 6.3)으로 1로 간다. 분모가 1로 가면 T는 Z가 된다. 7회의 Slutsky 정리가 이 직관을 정리로 만든다. 자유도 30이면 $t_{30,\,0.025} = 2.04$로 1.96과 거의 같다. "30개 이상이면 정규"라는 경험칙의 근거다. 그러나 30개라도 꼬리 깊은 곳은 다르다. $P(|T| > 4)$는 자유도 30에서 정규의 약 6배,

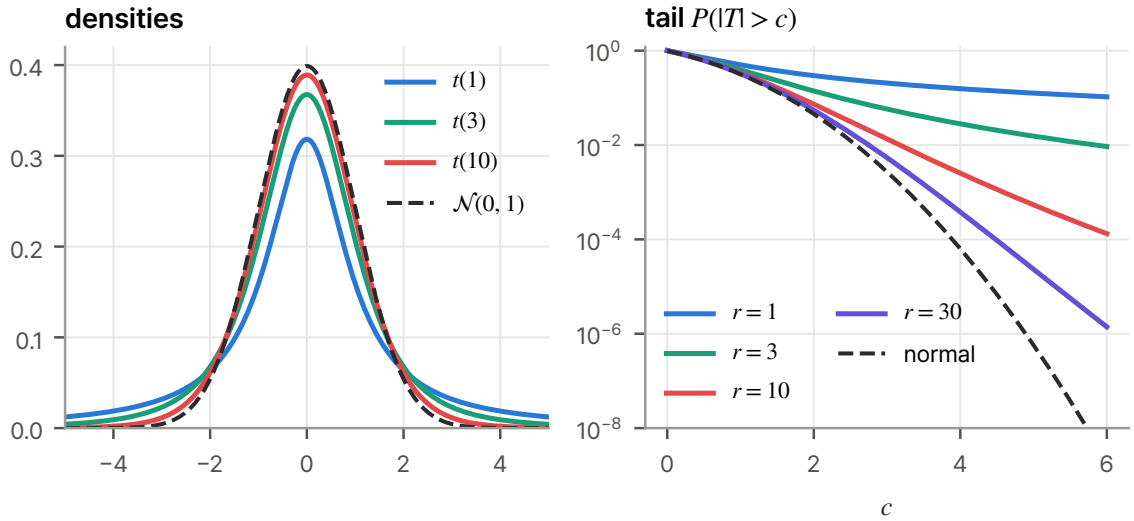


그림 3.1 왼쪽: $t(r)$의 밀도와 표준정규(점선). 오른쪽: 꼬리 확률 $P(|T| > c)$의 로그. $r = 3$이면 $c = 3$에서 정규보다 20배쯤 크다.

자유도 10에서 약 40배다.

3.3 Student의 t

정리 3.4 (Student의 t) $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이면

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

증명. **전략** 분자와 분모를 $\sigma/\sqrt{n}$으로 나눠 정의 3.1의 모양을 만든다. 모르는 σ는 이때 소거된다.

① 분자와 분모를 $\sigma/\sqrt{n}$으로 나눈다.

$$T = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{Z}{\sqrt{V/(n-1)}}, \qquad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}, \quad V = \frac{(n-1)S^2}{\sigma^2}.$$

② $Z \sim \mathcal{N}(0,1)$이다. 정리 2.2(a)

③ $V \sim \chi^2(n-1)$이다. 정리 2.2(c)

④ Z는 $\bar{X}$의 함수이고 V는 S^2의 함수이므로 $Z \perp V$이다. 정리 2.2(b), 3회 정리 3.4(a)

⑤ ①–④로 T는 정의 3.1의 $t(n-1)$이다. 정의 3.1

■

σ가 분자와 분모에서 모두 소거되었다. 그래서 T의 분포는 모르는 σ에 의존하지 않는다. 이것이 Student의 결과가 쓸모 있는 이유이고, 5절의 "축량"의 원형이다.

> **Student의 z와 Fisher의 t** Student(1908)는 $z = (\bar{x}-\mu)/s'$, $s'^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$ (n으로 나눈 표준편차)의 분포를 구해 $n = 10$까지의 표를 만들었다. Fisher(1925)는 $t = z\sqrt{n-1}$로 바꿔, 하나의 표로 평균의 비교, 회귀계수, 두 표본의 차이를 모두 다룰 수 있게 했다. 자유도라는 개념과 t라는 글자가 이때 자리 잡았다.

연습문제

연습 3.1 (감마의 비 ★★) 5회 보조정리 1.1(Stirling)로 $x \to \infty$에서 $\frac{\Gamma(x+\frac{1}{2})}{\Gamma(x)\sqrt{x}} \to 1$임을 보이고, 정리 3.3(d)의 상수가 $\frac{1}{\sqrt{2\pi}}$로 감을 결론지어라. 풀이 39쪽

연습 3.2 (정규의 분산 혼합 ★★★) $W \sim \Gamma(\frac{r}{2}, \frac{2}{r})$ (평균 1)이고 $W = w$일 때 $T \sim \mathcal{N}(0, 1/w)$이면 $T \sim t(r)$임을 보여라. (힌트: $T = Z/\sqrt{W}$로 쓸 수 있다.) t 분포는 "정밀도가 무작위인 정규"들의 혼합이다(4회 5절). 풀이 39쪽

연습 3.3 ($t(2)$의 분포함수 ★★) $f_2(t) = \frac{1}{(2+t^2)^{3/2}}$임을 확인하고 $F_2(t) = \frac{1}{2}\big(1 + \frac{t}{\sqrt{2+t^2}}\big)$임을 보여라. $t_{2,0.025} = 4.303$을 이 식으로 확인하라. 풀이 40쪽

연습 3.4 (꼬리의 비 ★) 노트북으로 $P(|T|>3)$과 $P(|T|>4)$를 $r = 3, 10, 30$과 표준정규에서 구해 비를 표로 만들어라. 풀이 40쪽

연습 3.5 (t 분포의 성질 증명 ★★ 유도형) 정리 3.3를 다음 순서로 증명하라. f_r은 정리 3.2의 밀도다.

(a) 정리의 (a): $f_r(-t) = f_r(t)$를 확인하라.

(b) 정리의 (b) 앞부분: $|t| \to \infty$에서 $|t|^{r+1} f_r(t)$가 양의 상수로 감을 보이고, 비교 판정으로 $\int |t|^k f_r(t)\,dt < \infty \iff k < r$을 결론지어라 [T5].

(c) 정리의 (b) 뒷부분: $r > 2$이면 $ET^2 = EZ^2 \cdot E(r/V)$임을 독립으로 보이고, 감마 적분으로 $E(1/V) = \frac{1}{r-2}$를 계산해 $\operatorname{Var} T = \frac{r}{r-2}$를 얻어라. $r > 1$이면 $ET = 0$인 이유도 밝혀라.

(d) 정리의 (c): $r = 1$을 넣어 $f_1(t) = \frac{1}{\pi(1+t^2)}$를 확인하라.

(e) 정리의 (d): $\log\big(1 + \frac{t^2}{r}\big)^{-(r+1)/2} \to -\frac{t^2}{2}$를 $\log(1+x) = x + O(x^2)$로 보이고 [T8], 상수의 극한은 연습 3.1로 구하라.

풀이 40쪽

4

F 분포 — 두 분산의 비

두 정규 표본의 분산이 같은지 비교하거나(두 기계의 정밀도), 여러 집단의 평균이 같은지 검정할 때(14회 분산분석) 쓰는 것은 두 χ^2의 비다.

정의 4.1 (F 분포) $U \sim \chi^2(r_1)$과 $V \sim \chi^2(r_2)$가 독립일 때

$$F = \frac{U/r_1}{V/r_2}$$

의 분포를 자유도 (r_1, r_2)인 F **분포** $F(r_1, r_2)$라 한다. 이름은 Snedecor(1934)가 Fisher를 기려 붙였다.

정리 4.2 (F 분포의 밀도와 성질)

(a) 밀도는 $x > 0$에서

$$f(x) = \frac{1}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \left(\frac{r_1}{r_2}\right)^{r_1/2} \frac{x^{r_1/2-1}}{\left(1 + \frac{r_1}{r_2}x\right)^{(r_1+r_2)/2}}.$$

(b) $r_2 > 2$이면 $EF = \dfrac{r_2}{r_2 - 2}$.

(c) $F \sim F(r_1, r_2)$이면 $1/F \sim F(r_2, r_1)$이고, $T \sim t(r)$이면 $T^2 \sim F(1, r)$.

(d) $\dfrac{(r_1/r_2)F}{1 + (r_1/r_2)F} \sim \mathrm{Beta}\left(\frac{r_1}{2}, \frac{r_2}{2}\right)$.

증명. **전략** 두 카이제곱을 감마로 쓰면 F는 감마의 비의 상수배다. 비의 밀도(4회 연습 4.4)와 척도 변환에서 (a)가, 감마의 비의 베타 표현(4회 정리 4.2)에서 (d)가, 독립인 두 인수의 기댓값에서 (b)가, 정의에서 (c)가 나온다. 단계별 증명은 연습 4.4(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 41쪽에 있다.

정리 4.3 (두 표본 분산의 비) $X_1, \dots, X_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$과 $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$가 독립이면

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(m-1,\, n-1).$$

증명. 정리 2.2(c)로 $(m-1)S_X^2/\sigma_1^2 \sim \chi^2(m-1)$, $(n-1)S_Y^2/\sigma_2^2 \sim \chi^2(n-1)$이고 두 표본이 독립이므로 둘도 독립이다. 정의 4.1. ∎

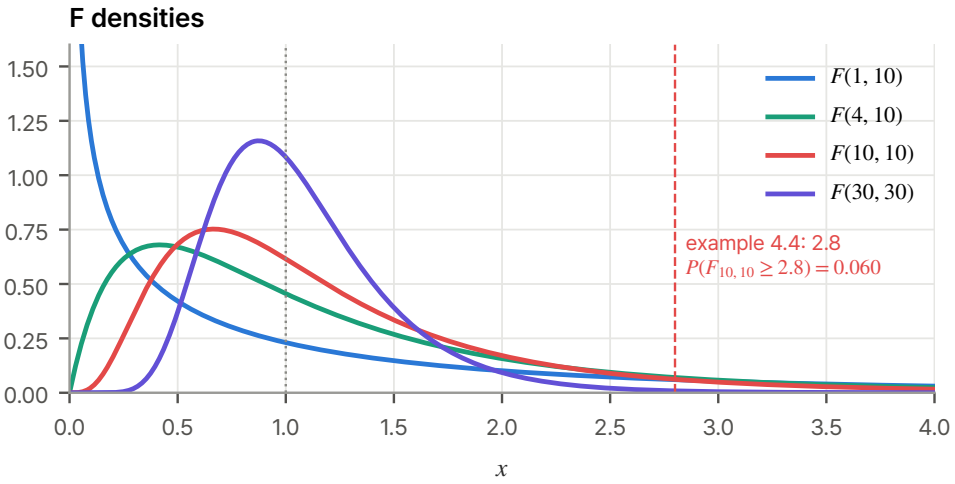


그림 4.1 $F(r_1, r_2)$의 밀도. 자유도가 커지면 1 근처로 몰린다. 두 분산이 같다면 표본분산의 비는 1 근처여야 하고, 얼마나 멀어야 "다르다" 고 할지는 이 분포가 정한다.

예제 4.4 (두 기계의 정밀도) 두 기계로 만든 부품의 지름을 각각 $m = n = 11$개 쟀더니 $s_X^2 = 4.2$, $s_Y^2 = 1.5$였다. 정밀도가 같다면($\sigma_1 = \sigma_2$) $s_X^2/s_Y^2 = 2.8$은 $F(10, 10)$을 따른다. $P(F \geq 2.8) \approx 0.06$이다. 2.8배라는 차이도 표본 11개로는 우연과 쉽게 구별되지 않는다. 분산의 비교는 평균의 비교보다 훨씬 많은 관측이 필요하고, 정규성이 깨지면 크게 틀린다(7절). 8회와 14회에서 이 분포로 검정을 세운다.

연습문제

연습 4.1 (F의 평균과 분산 ★★) $r_2 > 4$이면 $\operatorname{Var} F = \dfrac{2r_2^2(r_1 + r_2 - 2)}{r_1(r_2 - 2)^2(r_2 - 4)}$임을 보여라. (힌트: $E\frac{1}{V^2}$를 4회 연습 3.4로.)

풀이 41쪽

연습 4.2 (분위수의 대칭 ★) $F_{r_1,r_2,p}$를 $F(r_1, r_2)$의 위쪽 p 분위수라 하면 $F_{r_1,r_2,1-p} = 1/F_{r_2,r_1,p}$임을 보여라.

풀이 41쪽

연습 4.3 (σ_1^2/σ_2^2의 구간 ★★) 정리 4.3로 σ_1^2/σ_2^2의 95% 신뢰구간을 만들고, 예제 4.4의 자료로 계산하라. 구간에 1이 들어 있나?

풀이 41쪽

연습 4.4 (F 분포의 성질 증명 ★★ 유도형) 정리 4.2를 다음 순서로 증명하라.

(a) $U \sim \chi^2(r_1)$이면 $X = U/2 \sim \Gamma(\frac{r_1}{2}, 1)$임을 4회 정의 3.8로 확인하라. 같은 방법으로 $Y = V/2$를 두고 $R = U/V = X/Y$의 밀도를 4회 연습 4.4로 적어라.

(b) $F = \frac{r_2}{r_1}R$에 척도 변환(2회 예제 4.5)을 써서 정리의 (a)를 얻어라.

(c) 정리의 (b): $EF = E\frac{U}{r_1} \cdot E\frac{r_2}{V}$를 독립으로 보이고, 연습 3.5(c)의 $E(1/V) = \frac{1}{r_2 - 2}$로 끝내라. $r_2 \leq 2$이면 왜 평균이 없는가?

(d) 정리의 (c): $1/F$와 T^2를 정의 4.1의 꼴로 써라. T^2에서는 $Z^2 \sim \chi^2(1)$(4회 정리 3.9)과 $Z^2 \perp V$를 쓴다.

(e) 정리의 (d): $\dfrac{(r_1/r_2)F}{1+(r_1/r_2)F} = \dfrac{X}{X+Y}$를 보이고 4회 정리 4.2를 써라.

풀이 41쪽

5

축량과 t-구간

5.1 축량

정의 5.1 (축량) 자료 $\mathbf{X}$와 모수 θ의 함수 $Q(\mathbf{X}, \theta)$의 분포가 θ(와 다른 모르는 모수)에 의존하지 않으면 Q를 **축량**(pivot)이라 한다.

정리 3.4의 $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$는 μ와 σ가 무엇이든 $t(n-1)$을 따르므로 축량이다. 축량이 있으면 분포의 분위수로 확률 문장을 쓰고, 그 문장을 θ에 대해 풀어 구간을 얻는다.

정리 5.2 (t-구간) $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 $c = t_{n-1,\alpha/2}$이면, 모든 μ와 σ에서

$$P\left(\bar{X} - c\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + c\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

증명. **전략** 축량 T에 대한 확률 문장을 μ에 대해 푼다. 확률은 T의 분포만으로 정해지므로 μ와 σ에 의존하지 않는다.

① $S > 0$이면 $\bar{X} - c\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + c\frac{S}{\sqrt{n}}$은 $-c \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le c$, 곧 $|T| \le c$와 같은 사건이다. 부등식에 $\sqrt{n}/S > 0$을 곱한다

② $T \sim t(n-1)$이고 t 분포는 연속이므로 $P(|T| \le c) = F_{n-1}(c) - F_{n-1}(-c)$이다. 정리 3.4

③ 대칭으로 $F_{n-1}(-c) = 1 - F_{n-1}(c) = \frac{\alpha}{2}$이고, c의 정의로 $F_{n-1}(c) = 1 - \frac{\alpha}{2}$이다. 정리 3.3(a)

④ 따라서 확률은 $1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$이다. 이 값에는 μ도 σ도 들어 있지 않다. ②③

조건 확인 ①은 $S > 0$을 쓴다. $n \ge 2$이면 $\{S = 0\} = \{X_1 = \cdots = X_n\}$이고, 연속분포에서 $P(X_2 = X_1) = 0$이므로 이 사건의 확률은 0이다. ∎

확률 문장의 주어는 무작위인 **구간**이지 고정된 μ가 아니다. "같은 방법으로 구간을 100번 만들면 약 95번 μ를 포함한다"가 정확한 읽기다. Neyman(1937)이 이 해석에 "신뢰구간"이라는 이름을 붙였고, 8회에서 신뢰구간 일반을 다룬다. 구간의 길이 $2cS/\sqrt{n}$은 무작위다. 평균 길이는 $2c\,c_4(n)\sigma/\sqrt{n}$(연습 2.4)이고, σ를 알 때의 $2 \cdot 1.96\,\sigma/\sqrt{n}$과 비교하면 $n = 4$에서 $\frac{3.182 \cdot 0.921}{1.96} \approx 1.50$배 길다. 모르는 것 하나의 값이다.

5.2 대응 비교 — Cushny와 Peebles의 수면제 (1905)

> **역사** 약리학자 Cushny와 Peebles는 1905년 정신병원 환자 11명에게 hyoscine과 hyoscyamine 계열의 약을 밤을 바꿔 가며 주고, 약을 주지 않은 밤과 수면 시간을 비교했다. Student(1908)는 이 자료로 작은 표본의 비교를 보였고, Fisher(1925)도 같은 자료를 썼다. 두 사람 모두 약의 이름을 잘못 옮겨 적었다는 것이 나중에 밝혀졌다(Senn과 Richardson, 1994). R의 `sleep` 자료는 처음 10명에 대해 L-hyoscyamine과 L-hyoscine이 대조 밤보다 늘린 수면 시간이다.

같은 환자에게 두 처리를 주면 환자마다의 차이 $D_i = X_i - Y_i$를 본다. 환자 사이의 큰 차이(어떤 사람은 원래 잘 잔다)가 빼기에서 사라지므로 비교가 정밀해진다. $D_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$로 놓으면 정리 5.2을 D에 그대로 쓴다.

예제 5.3 (수면제의 대응 t) `sleep` 자료의 두 약의 차이(L-hyoscine − L-hyoscyamine) 10개는 평균 $\bar{d} = 1.58$시간, 표준편차 $s_d = 1.23$이다.

$$t = \frac{1.58}{1.23/\sqrt{10}} = 4.06, \qquad P(|T_9| \ge 4.06) = 0.0028,$$

$$95\%\ \text{구간}: \ 1.58 \pm 2.262 \cdot \frac{1.23}{\sqrt{10}} = [0.70,\ 2.46].$$

L-hyoscine이 L-hyoscyamine보다 수면을 0.7–2.5시간 더 늘린다. 같은 20개의 수를 **대응을 무시하고** 두 독립 표본으로 비교하면(아래 정리 5.4) $t = 1.86$, $p = 0.079$로 약해진다. 환자 사이의 변동이 차이를 가렸기 때문이다. 대조 밤과 비교하면 L-hyoscyamine의 효과는 0.75시간($t = 1.33$, $p = 0.22$)으로 불분명하고, L-hyoscine은 2.33시간($t = 3.68$, $p = 0.005$)이다.

대응이 도움이 되는 이유는 3회의 공분산이다. $\mathrm{Var}(X-Y) = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$이고, 같은 환자의 두 측정은 ρ가 커서 차이의 분산이 작다. `sleep`에서 두 약의 상관은 0.80이다.

5.3 두 독립 표본

정리 5.4 (두 표본 t (등분산)) $X_1, \dots, X_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \sigma^2)$과 $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \sigma^2)$가 독립이고 분산이 같으면, 합동분산 $S_p^2 = \dfrac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$에 대해

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{m} + \frac{1}{n}}} \ \sim\ t(m+n-2).$$

증명. **전략** 정리 3.4처럼 표준정규를, 그와 독립인 χ^2를 자유도로 나눈 것의 제곱근으로 나눈 꼴을 만든다. 두 표본에 Student의 정리를 따로 쓰고 독립을 합친다.

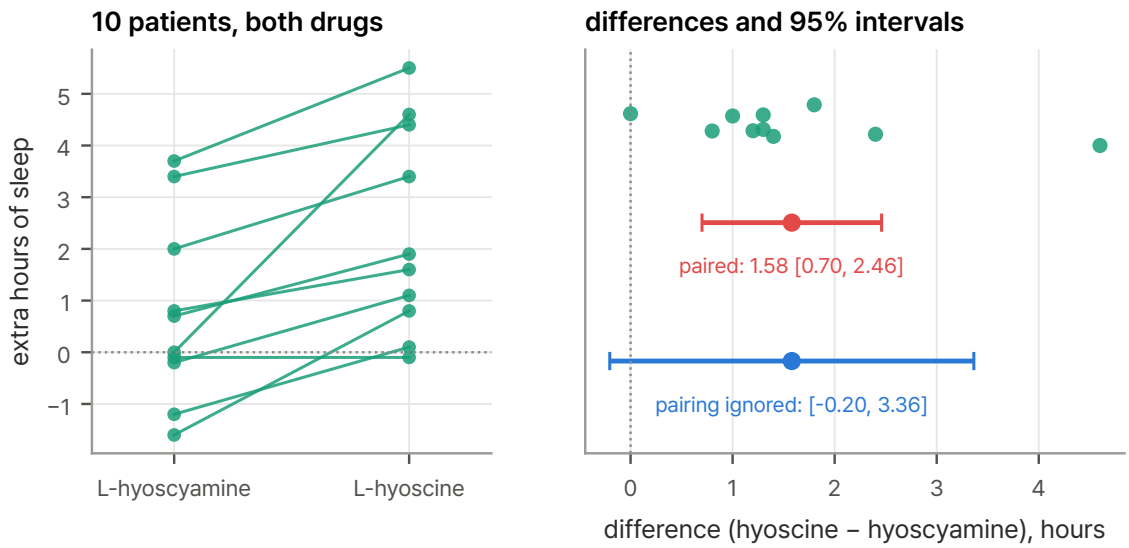


그림 5.1 왼쪽: 환자 10명의 두 약에 대한 수면 증가(선으로 이은 쌍). 거의 모든 선이 오른쪽으로 올라간다. 오른쪽: 차이 10개와 평균의 95% t-구간.

① $\bar{X}-\bar{Y} \sim \mathcal{N}\big(\mu_1-\mu_2,\ \sigma^2(\frac{1}{m}+\frac{1}{n})\big)$이므로 $Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/m+1/n}} \sim \mathcal{N}(0,1)$이다.

정리 2.2(a), 5회 정리 1.8(iv)

② $\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2(m-1)$, $\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(n-1)$이고, 두 표본이 독립이므로 둘도 독립이다.

정리 2.2(c)

③ $V=\frac{(m+n-2)S_p^2}{\sigma^2}$는 ②의 두 변수의 합이므로 $\chi^2(m+n-2)$이다. 4회 정리 3.7

④ $\bar{X} \perp S_X^2$, $\bar{Y} \perp S_Y^2$이고 두 표본이 독립이므로 $\bar{X}, S_X^2, \bar{Y}, S_Y^2$는 상호 독립이다. Z는 $(\bar{X},\bar{Y})$의, V는 (S_X^2, S_Y^2)의 함수이므로 $Z \perp V$이다. 정리 2.2(b), 3회 3.2절의 상자

⑤ $T=\frac{Z}{\sqrt{V/(m+n-2)}}$이다. 분자와 분모의 σ가 소거된다. 정의 3.1로 $T \sim t(m+n-2)$이다.

대수, 정의 3.1

조건 확인 등분산은 ③에서 쓰인다. 두 χ^2를 같은 σ^2로 나눠야 합이 다시 χ^2가 된다. $\sigma_1 \neq \sigma_2$이면 S_p^2는 σ^2 하나의 추정량이 아니고, 아래 주의의 문제가 생긴다. ■

> **가정 주의** 등분산 가정이 틀리고 표본 크기도 다르면 정리 5.4의 구간은 크게 틀린다. 작은 표본 쪽의 분산이 크면 실제 포함확률이 95%보다 훨씬 낮다(연습 5.4). 두 분산이 다를 때의 정확한 축량은 없다(Behrens–Fisher 문제). 실무에서는 분모를 $\sqrt{S_X^2/m+S_Y^2/n}$으로 두고 자유도를 어림하는 Welch(1947)의 근사를 쓴다.

연습문제

연습 5.1 (구간의 평균 길이 ★) $n = 4, 10, 30$에서 t-구간의 평균 길이와 σ를 알 때의 구간 길이의 비 $\dfrac{t_{n-1,0.025}\, c_4(n)}{1.96}$을 구하라. 풀이 42쪽

연습 5.2 (몇 개를 재야 하나 ★★) 평균의 95% 구간의 반폭을 $\sigma/2$ 이하로 **평균적으로** 만들려면 n이 얼마여야 하나? $t_{n-1,0.025}\, c_4(n)/\sqrt{n} \le \frac{1}{2}$을 만족하는 가장 작은 n을 노트북으로 찾고, σ를 알 때의 $n = 16$과 비교하라. 풀이 42쪽

연습 5.3 (대응의 이득 ★) `sleep`에서 두 약의 표본분산과 상관계수로 $s_d^2 = s_X^2 + s_Y^2 - 2 r s_X s_Y$를 확인하라. 대응하지 않은 두 표본 t의 분모와 비교해 대응이 분산을 몇 배 줄였는지 말하라. 풀이 42쪽

연습 5.4 (등분산이 아닐 때 ★★) $m = 5$, $\sigma_1 = 3$과 $n = 20$, $\sigma_2 = 1$에서 정리 5.4의 95% 구간의 실제 포함확률을 시뮬레이션으로 구하라. Welch 구간(`scipy.stats.ttest_ind(..., equal_var=False)`의 방식)은 어떤가? 풀이 42쪽

연습 5.5 (한쪽 구간 ★) μ의 95% 한쪽 구간 $\left(-\infty,\ \bar{X} + t_{n-1,0.05} S/\sqrt{n}\right]$의 포함확률이 0.95임을 보이고, `sleep`의 차이로 계산하라. 풀이 42쪽

6

데이터로 — Student의 카드 실험 (1908)

> **역사** Student는 자기가 유도한 분포를 실험으로 확인했다. 재료는 W. R. Macdonell이 1902년 *Biometrika*에 발표한 잉글랜드와 웨일스의 성인 남자 수감자 3,000명의 키와 왼손 가운뎃손가락 길이였다. 그는 3,000장의 카드에 측정값을 적어 잘 섞은 뒤 넷씩 뽑아 750개의 표본을 만들었다. 표본마다 평균과 표준편차를 손으로 계산하고, 그 분포를 자기 곡선과 비교했다. 오늘 말로 하면 통계학 최초의 체계적인 Monte Carlo 실험이다. 결과는 곡선과 잘 맞았고, Student는 논문의 한 절을 이 "실제 시험"에 썼다.

6.1 모집단

Macdonell의 자료는 R 패키지 `HistData`에 `MacdonellDF`로 남아 있다. 키는 1인치 간격의 계급으로 기록되어 서로 다른 값이 20개뿐이다. 3,000명을 하나의 모집단으로 보면 평균은 $\mu = 65.04$ 인치, 표준편차는 $\sigma = 2.56$ 인치이고, 왜도 0.05와 초과첨도 0.17로 정규에 가깝다(그림 6.1 왼쪽). Student가 아는 것은 이 μ와 σ였으므로, 표본마다 계산한 t가 이론 분포를 따르는지 바로 확인할 수 있었다.

6.2 재현

카드를 섞는 일은 무작위 순열 하나다. 3,000개의 값을 섞어 750×4 배열로 바꾸면 Student의 750개 표본이 된다(노트북, 난수 씨앗 1907).

알고리즘 6.1 (Student의 카드 실험)

1. 모집단 $v_1, \dots, v_{3000}$의 평균 μ를 계산한다.
2. 무작위 순열 π로 $v_{\pi(1)}, \dots, v_{\pi(3000)}$을 넷씩 끊어 표본 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(750)}$을 만든다.
3. 표본마다 $t^{(j)} = \dfrac{\bar{x}^{(j)} - \mu}{s^{(j)}/2}$를 계산한다($s^{(j)} = 0$인 표본은 뺀다).
4. $t^{(j)}$의 분포를 $t(3)$과 $\mathcal{N}(0,1)$에 비교한다.

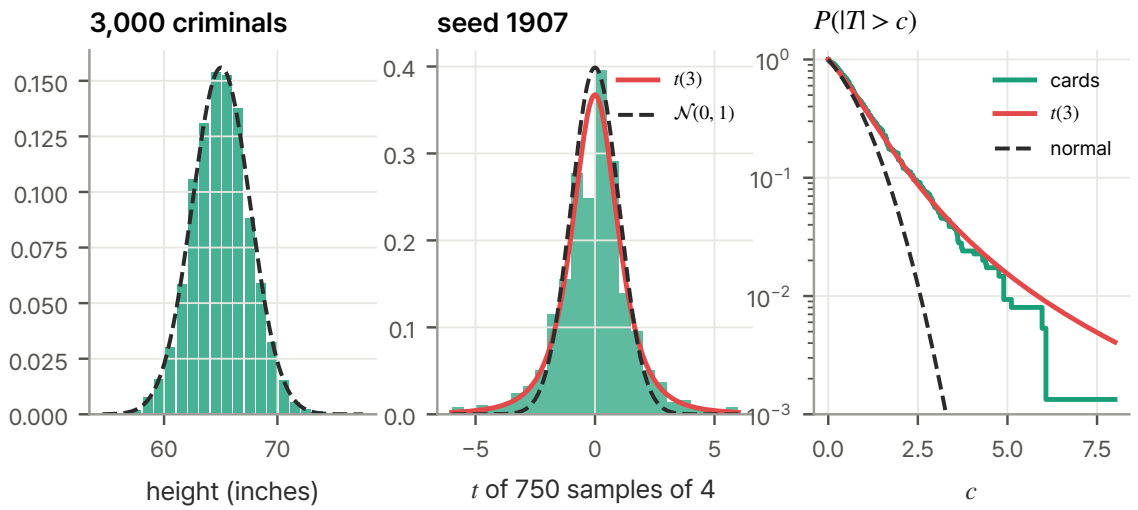


그림 6.1 Student의 카드 실험의 재현. 왼쪽: 범죄자 3,000명의 키(1인치 계급)와 같은 평균·표준편차의 정규 밀도. 가운데: 750개 표본의 t 값의 히스토그램과 $t(3)$(빨강), 표준정규(점선). 오른쪽: 꼬리에서 본 경험 분포 $P(|T| > c)$. 정규는 꼬리를 크게 얕잡는다.

예제 6.2 (750개의 t) 씨앗 1907의 재현에서 네 값이 모두 같은 계급에 들어 $s = 0$인 표본이 1개 나왔다(계급으로 묶인 자료의 흔적이다. 기대 개수는 $750 \sum_k p_k^4 \approx 1.5$). 나머지 749개 가운데

	재현 (씨앗 1907)	섞기 400번의 평균	$t(3)$ 이론
$\lvert t\rvert > 1.96$	14.2%	14.3%	14.5%
$\lvert t\rvert > 3.182$	4.5%	5.0%	5.0%
$\lvert t\rvert > 5.841$	0.8%	0.9%	1.0%

σ를 아는 것처럼 1.96을 쓰면 "5%"가 실제로는 14%다. $t(3)$은 실험과 맞는다. 경험 분포함수와 이론 분포함수의 최대 거리로 비교하는 Kolmogorov–Smirnov 검정에서 $t(3)$과의 차이는 우연으로 설명되고(유의확률 0.30), 표준정규와의 차이는 그렇지 않다(0.005).

손가락 길이로 같은 실험을 하면 결과는 더 좋다(연습 6.1). 계급이 0.1cm로 더 촘촘하기 때문이다. 키에서 s^2의 분포가 $\chi^2(3)$과 조금 어긋나는 것(유의확률 0.03)도 계급의 효과다. Student도 이것을 알아채고 계급 폭을 보정(Sheppard의 수정)한 뒤 비교했다.

실험의 뜻 Student는 증명하지 못한 가정($\bar{X} \perp S^2$) 위에 분포를 세웠고, 그 결과를 **계산 실험으로** 확인했다. 정리가 맞는지 모를 때 시뮬레이션으로 점검하는 습관은 오늘의 통계학과 기계학습의 기본 도구다. Student는 750묶음을 손으로 계산했고, 노트북은 400번의 반복을 몇 초 안에 한다. 7회에서 Monte Carlo의 오차 자체를 정리로 다룬다.

6.3 유한 모집단

카드 실험은 **되돌려 넣지 않는** 뽑기다. 3,000장 가운데 4장을 뽑으면 $\operatorname{Var}\bar{X} = \frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$이고(4회의 초기하 분산과 같은 이유), $\frac{N-n}{N-1} = 0.999$이므로 차이는 무시할 만하다. 그리고 750개의 표본은 서로 겹치지 않으므로 표본들끼리도 약하게 의존한다. 모두 n/N이 작으면 사라지는 효과다.

연습문제

연습 6.1 (손가락으로 ★) 노트북의 실험을 손가락 길이로 되풀이하라. $|t| > 3.182$의 비율과 s^2의 분포를 키의 경우와 비교하라. 풀이 42쪽

연습 6.2 (Student의 z ★★) Student는 $z = (\bar{x}-\mu)/s'$, $s'^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$를 썼다. $z = T/\sqrt{n-1}$임을 보이고, z의 밀도가 $(1+z^2)^{-n/2}$에 비례함을 정리 3.2에서 유도하라. 풀이 43쪽

연습 6.3 (같은 계급에 넷 ★★) 계급 k의 비율이 p_k이면 되돌려 넣는 뽑기에서 네 값이 모두 같은 계급일 확률은 $\sum_k p_k^4$이다. 키 자료에서 이 값을 계산하고, 750개 가운데 $s=0$인 표본의 기대 개수를 구하라. 되돌려 넣지 않는 뽑기에서는 식이 어떻게 바뀌나? 풀이 43쪽

연습 6.4 (섞기의 변동 ★★) 카드를 400번 섞어 매번 $|t| > 3.182$의 비율을 기록하라. 그 비율들의 표준편차를 이항 근사 $\sqrt{0.05\cdot 0.95/750}$와 비교하라. 씨앗 1908로 섞으면 이 비율이 3.2%다. 드문 결과인가? 풀이 43쪽

7

부쉬보기 — 정규가 아니고, 독립이 아니면

정리 5.2의 "정확히 $1-\alpha$"는 세 가정 위에 서 있다. 관측이 **정규**이고, **독립**이고, (두 표본이면) **분산이 같다**. 이 절은 가정을 하나씩 깨고 실제 포함확률을 잰다. 결론을 먼저 쓰면, 평균에 대한 t는 대칭인 비정규에는 강하고, 비대칭에는 작은 n에서 약하며, 의존에는 n이 커져도 약하다. 그리고 분산에 대한 F는 비정규에 아주 약하다.

7.1 비대칭 모집단

모집단	$n=5$	$n=10$	$n=30$	$n=100$	성질
균등 $U(0,1)$	0.934	0.946	0.949	0.948	대칭, 꼬리 짧음
$t(3)$	0.962	0.960	0.956	0.952	대칭, 꼬리 두꺼움
지수 Exp(1)	0.884	0.899	0.927	0.941	오른쪽으로 치우침
로그정규 $e^{\mathcal{N}(0,1)}$	0.822	0.837	0.882	0.918	크게 치우침

명목 95% t-구간의 실제 포함확률이다(각 5만 번 시뮬레이션). 대칭인 분포에서는 $n=5$에서도 거의 맞고, $t(3)$처럼 꼬리가 두꺼우면 오히려 보수적이다. 치우친 분포에서는 $n=30$에서도 모자라다. 원인은 평균과 표준편차의 **상관**이다.

> **명제 7.1** (평균과 표본분산의 공분산) $X_1,\dots,X_n$이 독립이고 같은 분포이며 $\mu_3 = E(X-\mu)^3$이 유한하면
>
> $$\operatorname{Cov}(\bar X, S^2) = \frac{\mu_3}{n}.$$

증명. **전략** $\mu=0$으로 놓고 $(n-1)S^2 = \sum_j X_j^2 - n\bar X^2$에 $\bar X$를 곱해 기댓값을 취한다. 독립이고 평균이 0이면, 한 첨자가 혼자 남는 곱의 기댓값은 0이다. 단계별 증명은 연습 7.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 44쪽에 있다.

정규면 $\mu_3=0$이고, 이것은 $\bar X \perp S^2$와 어긋나지 않는다. 오른쪽으로 치우친 분포($\mu_3>0$)에서는 $\bar X$가 작을 때 S도 작다. 지수 표본 $n=10$에서 $\bar X$와 S의 상관은 0.76이다. 그래서 구간이 아래로 밀렸을 때 하필 짧아진다. 놓치는 방향이 한쪽으로 몰린다.

예제 7.2 (한쪽으로 놓친다) 지수 표본 $n=10$의 95% t-구간은 90.1%를 포함한다. 놓친 9.9% 가운데 9.5%는 구간 전체가 μ 아래에 있는 경우이고, 위에 있는 경우는 0.4%뿐이다. 양쪽 2.5%씩이 되어야 할 오류가 9.5%와 0.4%로 갈린다. 그래서 μ의 **위쪽 한계**로 $\bar{X}+t_{9,0.025}S/\sqrt{10}$을 쓰면 97.5%가 아니라 90.5%만 맞는다. 로그정규에서는 15.9%와 0.1%로 더 심하다. 로그를 취해 대칭으로 만든 뒤 구간을 만들면(로그정규의 경우 정확히 정규가 되어) 포함확률은 95%로 돌아온다. 그러나 그것은 μ가 아니라 중앙값 $e^{E\log X}$의 구간이다.

n이 커지면 7회의 중심극한정리로 $\bar{X}$가 정규에 가까워지고 $S\to\sigma$이므로 t-구간은 결국 맞는다(그림 7.1 왼쪽). 그러나 수렴의 속도가 왜도 $\gamma=\mu_3/\sigma^3$에 달려 있다. 한쪽으로 놓치는 확률의 어긋남은 대략 $\gamma/\sqrt{n}$에 비례해서 줄어든다. 지수 표본에서 구간이 μ 아래에 있을 확률(이론값 2.5%)은 $n=10$에서 9.5%, $n=100$에서 4.5%다. 양쪽 구간에서는 두 쪽의 어긋남이 일부 상쇄되어 더 빨리 줄지만, 한쪽 한계는 느리다. "$n\ge 30$이면 정규"라는 경험칙은 왜도가 작을 때만의 이야기다. 15회의 부트스트랩-t가 이 왜도를 자료에서 어림해 보정한다.

7.2 이상값 하나

명제 7.3 (이상값은 t를 1로 끌고 간다) $x_2,\dots,x_n$을 고정하고 $x_1=M\to\infty$이면 $t=\dfrac{\sqrt{n}\,\bar{x}}{s}\to 1$.

증명. **전략** $\bar{x}$와 s를 M의 거듭제곱으로 전개해 가장 큰 차수의 항만 비교한다 [T8]. 단계별 증명은 연습 7.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 44쪽에 있다.

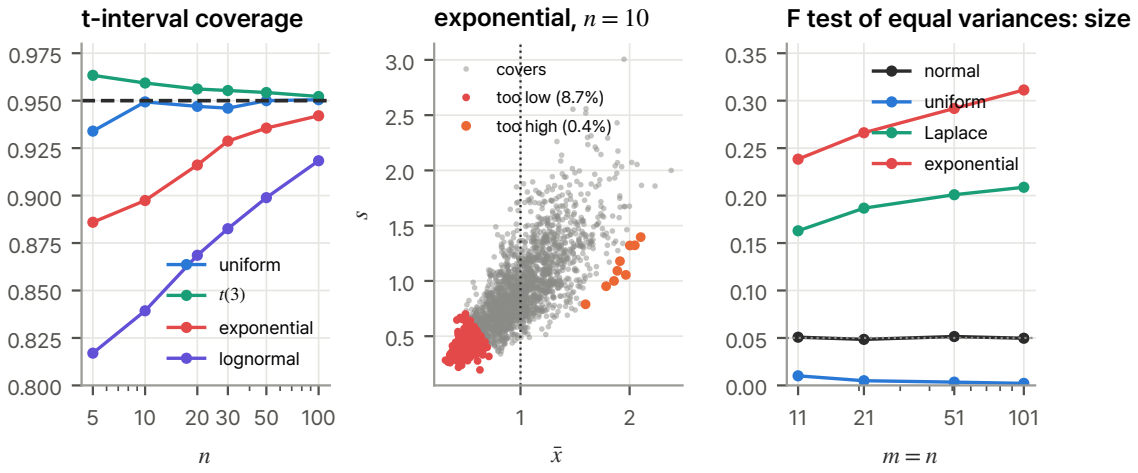


그림 7.1 왼쪽: 모집단별 t-구간의 실제 포함확률. 치우친 분포는 $n=100$에서도 95%에 닿지 못한다. 가운데: 지수 표본 $n=10$의 $(\bar{x},s)$ 2,000쌍. 구간이 μ보다 아래에 있는 경우(빨강)는 거의 모두 $\bar{x}$와 s가 함께 작은 왼쪽 아래에 있다. 오른쪽: 분산의 F 검정(5%)의 실제 기각률. 첨도가 큰 분포에서는 n이 커져도 나아지지 않는다.

큰 이상값은 평균을 끌어올리지만 표준편차를 더 크게 키워서, $\mu = 0$을 검정하는 t가 오히려 작아진다. 평균이 1 근처인 10개의 값에서 $t = 3.58$이었는데, 첫 값을 20으로 바꾸면 $t = 1.46$, 100으로 바꾸면 1.09다. 한 개의 큰 값이 "효과가 있다"는 결론을 지운다. t는 분포의 모양에 대해서는 의외로 튼튼하지만(위 표의 대칭 분포), 소수의 거친 오차에는 무력하다. 15회의 순위 검정과 로버스트 추정이 이 문제에서 출발한다.

7.3 의존하는 관측

관측이 서로 닮으면, 예를 들어 같은 날의 측정이나 시계열처럼 $\operatorname{Corr}(X_i, X_j) = \rho^{|i-j|}$ (AR(1), $0 < \rho < 1$)이면 3회 정리 4.2(공분산의 쌍선형성)로

$$\operatorname{Var}\bar{X} = \frac{\sigma^2}{n}\Big[1 + 2\sum_{k=1}^{n-1}\Big(1 - \frac{k}{n}\Big)\rho^k\Big] \longrightarrow \frac{\sigma^2}{n}\cdot\frac{1+\rho}{1-\rho} \qquad (n \to \infty)$$

이다(연습 7.2). S^2는 여전히 σ^2 근처이므로 t-구간은 $\sqrt{(1+\rho)/(1-\rho)}$배 너무 짧다. $\rho = 0.5$이면 그 배수가 $\sqrt{3}$이고 포함확률은 $n = 10, 30, 100$에서 0.76, 0.74, 0.74로 **n이 커져도 나아지지 않는다.** 극한은 $P(|Z| \le 1.96/\sqrt{3}) = 0.742$다. 비정규성은 표본이 크면 사라지지만 의존은 사라지지 않는다. Student 자신도 1927년의 논문에서 실험실의 일상 분석은 같은 날의 측정끼리 닮아서 오차가 이론보다 크다고 경고했다.

7.4 분산의 비교는 정규성에 민감하다

정리 4.3의 F 검정을 비정규 자료에 쓰면 결과가 크게 틀린다. 두 집단이 같은 분포이고 $m = n = 11$일 때 명목 5% 검정의 실제 기각률은 균등 1.0%, Laplace 16%, 지수 24%이고, $n = 101$에서는 0.3%, 21%, 31%로 오히려 나빠진다. 이유는 연습 1.2에 있다. $\operatorname{Var} S^2 = \frac{\sigma^4}{n}(2 + \kappa) + O(n^{-2})$에서 $\kappa = \mu_4/\sigma^4 - 3$은 초과첨도이고, 정규 이론은 $\kappa = 0$을 가정한다. n이 커져도 이 비율 $1 + \kappa/2$는 그대로 남는다. 반면 같은 자료에서 평균의 t 검정은 지수를 빼면 4.3–5.2%로 거의 정확하다. Box(1953)는 이 대조를 보이며 "로버스트"(robust)라는 말을 통계학에 들였다. 평균의 비교는 로버스트하고, 분산의 비교는 그렇지 않다.

연습문제

연습 7.1 (지수 표본의 공분산 ★) 명제 7.1로 Exp(1) 표본에서 $\operatorname{Cov}(\bar X, S^2) = \frac{2}{n}$임을 확인하라 ($\mu_3 = 2$). 연습 2.3의 $n = 2$의 결과와 맞는가? 풀이 43쪽

연습 7.2 (AR(1)의 분산 ★★) $\operatorname{Var} X_i = \sigma^2$, $\operatorname{Cov}(X_i, X_j) = \sigma^2 \rho^{|i-j|}$이면 본문의 $\operatorname{Var}\bar X$ 식이 성립함을 보이고, $n \to \infty$에서 괄호가 $\frac{1+\rho}{1-\rho}$로 감을 보여라. 음의 상관($\rho < 0$)이면 t-구간은 어떻게 되나? 풀이 43쪽

연습 7.3 (F 검정의 점근 수준 ★★★) 연습 1.2의 $\operatorname{Var} S^2$와 7회의 델타 방법을 미리 빌려 $\log(S_X^2/S_Y^2)$의 분산을 $\frac{2(2+\kappa)}{n}$로 어림하라. 정규 이론의 5% 검정이 실제로는 $P\big(|Z| > 1.96/\sqrt{1+\kappa/2}\big)$로 기각함을 보이고, Laplace($\kappa = 3$)와 지수($\kappa = 6$)에서 계산해 본문의 시뮬레이션과 비교하라. 풀이 44쪽

연습 7.4 (대칭이면 보수적 ★★) $t(3)$ 모집단에서 t-구간이 95%보다 많이 포함하는 이유를 S의 분포로 설명하라. (힌트: 큰 관측 하나가 $|\bar X - \mu|$보다 S를 더 키운다. 명제 7.3.) 풀이 44쪽

연습 7.5 (평균과 분산의 공분산 증명 ★★ 유도형) 명제 7.1를 다음 순서로 증명하라.

(a) X_i를 $X_i - \mu$로 바꿔도 $\operatorname{Cov}(\bar X, S^2)$가 변하지 않음을 보여, $\mu = 0$으로 놓아도 됨을 밝혀라. 그때 $\operatorname{Cov}(\bar X, S^2) = E(\bar X S^2)$이다.

(b) $\mu = 0$이면 $E(X_i X_j^2)$는 $i = j$일 때 μ_3, 아니면 0이고, $E(X_i X_j X_k)$는 $i = j = k$일 때 μ_3, 아니면 0임을 독립으로 보여라.

(c) (b)로 $E\big(\bar X \sum_j X_j^2\big) = \mu_3$, $E(n\bar X^3) = \frac{\mu_3}{n}$을 보여라.

(d) $(n-1)S^2 = \sum_j X_j^2 - n\bar X^2$를 써서 결론을 내려라. 기댓값이 모두 유한한 이유도 밝혀라.

풀이 44쪽

연습 7.6 (이상값의 극한 증명 ★★ 유도형) $x_2, \dots, x_n$을 고정하고 $a = \frac{1}{n}\sum_{i\ge 2} x_i$, $b = \sum_{i\ge 2} x_i^2$로 둔다. 명제 7.3를 다음 순서로 증명하라 [T8].

(a) $\bar x = \frac{M}{n} + a$임을 보여라.

(b) $\sum_i (x_i - \bar x)^2 = \sum_i x_i^2 - n\bar x^2$로 $(n-1)s^2 = \frac{n-1}{n}M^2 - 2aM + b - na^2$을 보여라.

(c) (b)에서 $s = \frac{M}{\sqrt{n}}\big(1 + O(\frac{1}{M})\big)$을 보여라.

(d) $t = \frac{\sqrt{n}\,\bar x}{s} \to 1$을 결론지어라. $x_1 = -M \to -\infty$이면 극한은 무엇인가?

풀이 44쪽

8

Beautiful insights — 작은 표본의 분포, 오늘의 문제

8.1 예측구간 — 다음 한 사람

t-구간은 **평균** μ의 구간이다. 다음 환자 한 사람의 수면 증가를 묻는다면 다른 구간이 필요하다.

정리 8.1 (예측구간) $X_1, \dots, X_n, X_{\text{new}} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이면

$$\frac{X_{\text{new}} - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} \sim t(n-1),$$

따라서 $\bar{X} \pm t_{n-1,\alpha/2}\, S\sqrt{1 + \frac{1}{n}}$은 X_{new}를 확률 $1-\alpha$로 포함한다.

증명. **전략** $X_{\text{new}} - \bar{X}$가 정규이고 S^2와 독립임을 보이면, 정리 3.4의 증명과 같은 꼴이 된다. 단계별 증명은 연습 8.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 46쪽에 있다.

`sleep`에서 다음 환자의 차이에 대한 95% 예측구간은 $1.58 \pm 2.262 \cdot 1.23\sqrt{1.1} = [-1.34,\ 4.50]$이다. 평균의 구간 $[0.70,\ 2.46]$보다 훨씬 넓고, 0을 포함한다. 평균적으로는 L-hyoscine이 낫다는 것이 분명해도 다음 환자 한 사람에게는 그렇지 않을 수 있다. $n \to \infty$이면 평균의 구간은 한 점으로 줄지만 예측구간은 $\mu \pm 1.96\sigma$로 남는다. 모델의 "불확실성"을 말할 때 이 둘(추정의 불확실성과 관측 자체의 흩어짐)을 구별해야 한다.

8.2 작은 배치의 BatchNorm

신경망의 배치 정규화(Ioffe와 Szegedy, 2015)는 미니배치 $x_1, \dots, x_B$의 각 값을 배치 평균과 배치 표준편차로 표준화한다.

$$\hat{x}_i = \frac{x_i - \bar{x}_B}{\hat{\sigma}_B}, \qquad \hat{\sigma}_B^2 = \frac{1}{B}\sum_{j=1}^{B}(x_j - \bar{x}_B)^2.$$

(실제로는 $\hat{\sigma}_B^2$에 작은 ε을 더하고, 뒤에 학습되는 척도와 위치를 곱하고 더한다.) 이것은 이 노트의 표준화와 같은 연산이고, B가 작으면 이 노트의 모든 현상이 일어난다.

명제 8.2 (배치 안의 값은 유계다) $|\hat{x}_i| \le \sqrt{B-1}$이다. 특히 $B=2$이면 $\hat{x}_1 = -\hat{x}_2 = \pm 1$이다.

증명. **전략** 편차의 합이 0이라는 제약으로 d_i를 나머지 편차의 합으로 쓰고 Cauchy–Schwarz 부등식을 쓴다 [T4]. 단계별 증명은 연습 8.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 46쪽에 있다.

$B=2$이면 BatchNorm의 출력은 "두 값 가운데 어느 쪽이 큰가" 라는 부호 하나뿐이다. 크기의 정보가 모두 사라진다. $B=4$이면 출력은 $[-\sqrt{3}, \sqrt{3}]$ 안에 갇힌다. 그런데 학습이 끝난 뒤 추론할 때는 배치 통계량 대신 학습 중에 모은 이동평균, 곧 모집단의 μ와 σ에 가까운 값으로 표준화하므로 출력은 유계가 아니다. **학습 때 본 분포와 추론 때 받는 분포가 다르다.** 그리고 그 이동평균 자체가 흔들린다. 배치 4개의 분산은 따름정리 2.3로 표준편차가 $\sigma^2\sqrt{2/3}$이고, 운동량 0.1의 이동평균으로 다듬어도 σ^2의 약 19%가 남는다(노트북). 작은 배치에서 BatchNorm의 성능이 급격히 떨어지는 이유의 하나이고, 배치 통계량에 의존하지 않는 Layer Norm · Group Norm(Wu와 He, 2018)이나 Batch Renormalization(Ioffe, 2017)이 그 대안으로 나왔다.

반대로 새 값 하나를 작은 배치의 통계량으로 표준화하면 정리 8.1의 t가 나온다. 정규 활성값에서 $\frac{x_{\text{new}}-\bar{x}_B}{\hat{\sigma}_B}$의 분산은 $\frac{B+1}{B-3}$이고(연습 8.2), $B=4$이면 5, $B=32$이면 1.14다. 배치의 표준편차가 우연히 작으면 값이 크게 튀는 것은 Student가 1908년에 본 바로 그 현상이다.

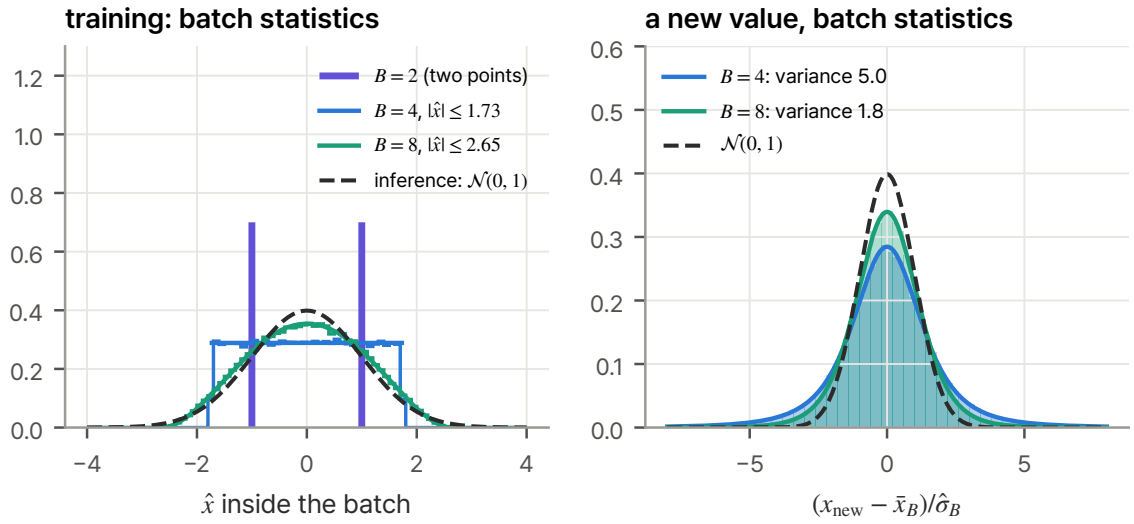


그림 8.1 왼쪽: 배치 안에서 표준화한 값의 분포($B=2,4,8$). $B=2$이면 ± 1 두 점, $B=4$이면 $[-\sqrt{3}, \sqrt{3}]$의 균등분포다(연습 8.3). 점선은 추론 때의 $\mathcal{N}(0,1)$. 오른쪽: 새 값을 배치의 통계량으로 표준화하면 꼬리가 두꺼운 t 꼴이다(실선: 이론 밀도, 히스토그램: 시뮬레이션).

8.3 t-SNE — 지도를 그릴 때 꼬리가 필요하다

t-SNE(van der Maaten과 Hinton, 2008)는 고차원의 점들을 2차원 지도에 그린다. 고차원에서는 점 i와 j의 유사도를 정규 커널 $\exp(-\|x_i - x_j\|^2/2\sigma_i^2)$로 재고, 지도에서는 **자유도 1인 t 커널**

$$q_{ij} \propto \left(1 + \|y_i - y_j\|^2\right)^{-1}$$

로 잰다. 두 유사도 분포가 같아지도록 지도의 점 y_i를 옮긴다. 왜 지도 쪽만 t인가? 고차원에서는 한 점 주위에 "적당히 먼" 이웃이 아주 많을 수 있는데, 2차원에는 그들을 모두 같은 거리에 놓을 자리가 없다(군집 문제, crowding problem). 꼬리가 두꺼운 커널은 적당히 먼 점들을 지도에서 훨씬 멀리 놓아도 유사도가 크게 줄지 않게 해 준다. 최댓값의 1%의 유사도가 정규 커널에서는 거리 3.0, $t(1)$ 커널에서는 거리 9.9에서 나온다(그림 8.2). 그래서 t-SNE의 지도는 군집 사이가 시원하게 벌어진다. 그 대신 지도의 군집 사이 거리는 원래 거리의 척도를 보존하지 않는다.

8.4 꼬리가 두꺼운 가능도 — 로버스트 회귀

오차를 정규 대신 $t(\nu)$로 놓으면, 잔차 r의 음의 로그밀도는 $\frac{\nu+1}{2}\log\left(1 + \frac{r^2}{\nu}\right)$이다. 정규의 $\frac{r^2}{2}$와 달리 큰 r에서 로그로만 자란다. 최대가능도(9회)의 방정식을 쓰면 각 관측이 가중치 $w = \frac{\nu+1}{\nu+r^2}$를 받는 가중 최소제곱이 되고, $\nu = 4$이면 잔차 3인 관측의 가중치는 0.38, 잔차 10이면 0.05다. 명제 7.3에서 본 이상값의 지배가 저절로 줄어든다. 이 가중치는 연습 3.2의 "정규의 분산 혼합"에서 숨은 정밀도 W의 조건부 기댓값이고, 10회의 EM 알고리즘이 그것을 반복 계산한다(Lange, Little, Taylor, 1989).

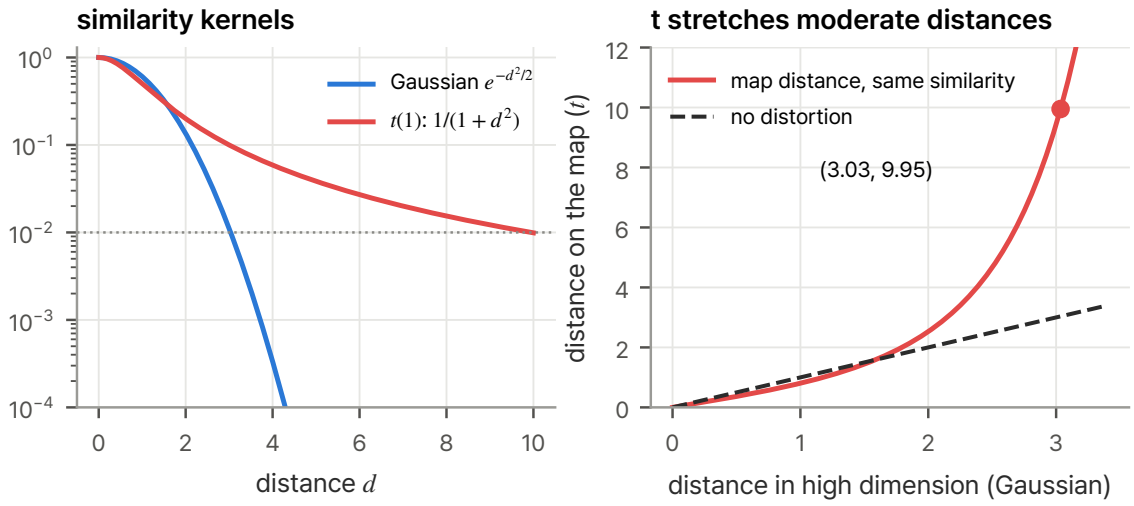


그림 8.2 왼쪽: 정규 커널과 $t(1)$ 커널의 유사도(로그 눈금). 오른쪽: 같은 유사도를 주는 거리의 대응. 고차원에서 거리 3인 이웃은 지도에서 거리 9.9에 놓여도 된다.

8.5 같은 구간, 다른 해석

> **Lüroth (1876)** 독일의 수학자 Jacob Lüroth는 Student보다 32년 먼저 t 분포를 유도했다. 그는 오차론의 논문에서 정규 관측의 평균에 대한 베이즈 구간을 계산하다가, 모르는 정밀도를 적분해 없앤 μ의 사후분포로 t 꼴의 밀도를 얻었다. 이 결과는 측지학 문헌에 묻혀 잊혔다가 Pfanzagl과 Sheynin(1996)이 다시 찾았다.

사전분포를 $\pi(\mu,\sigma)\propto 1/\sigma$로 두면 μ의 사후분포는 정확히

$$\frac{\mu-\bar{x}}{s/\sqrt{n}}\,\Big|\,\text{자료}\;\sim\;t(n-1)$$

이다(연습 8.5). 그러면 95% 사후 구간은 t-구간과 숫자까지 같다. `sleep`에서 사후분포를 시뮬레이션하면 $[0.70,\ 2.46]$이 나온다. 숫자는 같지만 문장은 다르다. 빈도론의 구간은 "이 방법은 95%의 경우에 μ를 잡는다"이고, 베이즈의 구간은 "자료를 본 지금 μ가 이 안에 있을 확률이 95%"다. 16회에서 이 둘이 언제 일치하고 언제 갈라지는지 본다.

연습문제

연습 8.1 (예측구간의 폭 ★) $n=4,10,100$에서 95% 예측구간의 반폭 $t_{n-1,0.025}\sqrt{1+1/n}$을 계산하고 $n\to\infty$의 극한 1.96과 비교하라. 풀이 45쪽

연습 8.2 (새 값의 분산 ★★) 정규 활성값에서 $\frac{x_{\text{new}}-\bar{x}_B}{\hat{\sigma}_B}$의 분산이 $\frac{B+1}{B-3}$임을 보여라($B>3$). (힌트: $\hat{\sigma}_B=S\sqrt{(B-1)/B}$로 정리 8.1의 T로 쓰고, 정리 3.3(b).) 풀이 45쪽

연습 8.3 (배치 안의 값의 분포 ★★★) 정규 표본에서 $\frac{\hat{x}_1^2}{B-1}\sim\text{Beta}\big(\frac12,\frac{B-2}{2}\big)$임을 보여라. (힌트: 정리 2.2의 증명처럼 잔차벡터를 $Y_2,\dots,Y_B$로 쓰면 $x_1-\bar{x}$는 그 가운데 한 방향의 성분이다. 첫 행이 그 방향인 직교행렬로 다시 돌리고 4회 정리 4.2.) $B=4$이면 $\hat{x}_1$이 $[-\sqrt{3},\sqrt{3}]$의 균등분포임을 보여라(그림 8.1의 평평한 선). 풀이 45쪽

연습 8.4 (t-SNE의 거리 ★) 유사도가 최댓값의 c배가 되는 거리를 정규 커널($\sigma=1$)과 $t(1)$ 커널에서 각각 구하라. $c=0.1$과 0.01에서 계산하라. 풀이 45쪽

연습 8.5 (Lüroth의 사후분포 ★★★) $x_1,\dots,x_n\overset{\text{iid}}{\sim}\mathcal{N}(\mu,\sigma^2)$이고 $\pi(\mu,\sigma)\propto 1/\sigma$이면, 결합 사후밀도는 $\sigma^{-(n+1)}\exp\!\big(-\frac{(n-1)s^2+n(\mu-\bar{x})^2}{2\sigma^2}\big)$에 비례한다. σ로 적분해 μ의 사후밀도가 $\big(1+\frac{n(\mu-\bar{x})^2}{(n-1)s^2}\big)^{-n/2}$에 비례함을 보이고 결론을 내려라. (힌트: $u=1/\sigma^2$로 치환하면 감마 적분.) 풀이 45쪽

연습 8.6 (예측구간 증명 ★★ 유도형) 정리 8.1를 다음 순서로 증명하라.

(a) X_{new}, $\bar{X}$, S^2가 상호 독립임을 보여라. 어느 독립을 어디서 가져오는지 밝혀라.

(b) $X_{\text{new}}-\bar{X}\sim\mathcal{N}\big(0,\sigma^2(1+\frac1n)\big)$임을 보이고, 이것이 S^2와 독립인 이유를 말하라.

(c) $T = \dfrac{X_{\text{new}} - \bar{X}}{S\sqrt{1+1/n}}$를 정의 3.1의 꼴 $Z/\sqrt{V/(n-1)}$로 써라.

(d) 정리 5.2의 증명처럼 확률 문장을 X_{new}에 대해 풀어 예측구간을 얻어라. 이 문장의 무작위인 양은 무엇인가?

풀이 46쪽

연습 8.7 (배치 안의 값의 상계 증명 ★★ 유도형) $d_j = x_j - \bar{x}_B$로 둔다. 명제 8.2를 다음 순서로 증명하라 [T4].

(a) $\sum_j d_j = 0$에서 $d_i = -\sum_{j\neq i} d_j$를 얻어라.

(b) Cauchy–Schwarz 부등식(3회 정리 4.4)으로 $d_i^2 \le (B-1)\sum_{j\neq i} d_j^2$를 보여라.

(c) $\sum_{j\neq i} d_j^2 = B\hat{\sigma}_B^2 - d_i^2$를 넣어 $|\hat{x}_i| \le \sqrt{B-1}$을 얻어라.

(d) 등호는 언제 성립하는가? $B=2$이면 $\hat{x}_1 = -\hat{x}_2 = \pm 1$임을 확인하라.

풀이 46쪽

9

정리와 역사 노트

9.1 한 장 요약

통계량	분포 · 성질	조건
$S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$	$ES^2 = \sigma^2$, 자유도 $n-1$	독립, 같은 분포
$\bar{X}$와 S^2	독립 (Helmert 회전)	정규뿐
$\frac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$, $\mathrm{Var}\, S^2 = \frac{2\sigma^4}{n-1}$	정규
$\frac{Z}{\sqrt{V/r}}$	$t(r)$: 꼬리 $\lvert t\rvert^{-(r+1)}$, 분산 $\frac{r}{r-2}$, $r \to \infty$에서 $\mathcal{N}(0,1)$	$Z \perp V$
$\frac{\bar{X}-\mu}{S/\sqrt{n}}$	$t(n-1)$, 축량 → 정확한 t-구간	정규
$\frac{U/r_1}{V/r_2}$, $\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}$	$F(r_1, r_2)$, $F(m-1, n-1)$, $T^2 \sim F(1, r)$	정규, 독립
$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{1/m+1/n}}$	$t(m+n-2)$	등분산
$\frac{X_{\text{new}}-\bar{X}}{S\sqrt{1+1/n}}$	$t(n-1)$, 예측구간	정규

증명은 하나의 그림으로 줄어든다. 정규 표본은 n차원의 회전에 불변이므로, 대각선 방향의 성분($\bar{X}$)과 그에 수직인 성분(잔차, S^2)이 독립이고, 수직 성분의 제곱합은 $n-1$개의 독립 표준정규의 제곱합이다. 나머지는 변수변환(밀도)과 "σ가 분자와 분모에서 소거된다"(축량)이다. 7절은 그 결론이 대칭인 비정규에는 튼튼하고, 치우침·이상값·의존·분산 비교에는 약하다는 것을 보였다.

9.2 Stigler의 명명 법칙

> **Stigler's law of eponymy**
>
> - **Student의 t**: Lüroth(1876)가 베이즈 사후분포로 32년 먼저 유도했다(8.5절). 그리고 오늘의 t라는 형태와 글자는 Fisher(1925)의 것이다. Student가 쓴 것은 $z = t/\sqrt{n-1}$이었다.
> - **Pearson의 χ^2 분포**: Abbe(1863)와 Helmert(1875–76)가 정규 오차의 제곱합의 분포로 먼저 유도했다. Pearson(1900)의 적합도 검정이 이름을 가져갔다(8회).
> - **Fisher의 F**: Fisher가 쓴 것은 $z = \frac{1}{2}\log F$였고, F로 바꿔 Fisher의 이름 첫 글자를 붙인 것은

Snedecor(1934)다.

- **Student라는 이름 자체**: Guinness가 직원의 실명 발표를 막아서 생긴 필명이다. 통계학에서 가장 유명한 분포의 이름은 사람이 아니라 회사의 규칙에서 나왔다.

9.3 7 Pillars 지도

이번 회차는 넷째 기둥 **Intercomparison**의 주춧돌이다. Stigler(2016)는 이 기둥을 "바깥의 기준 없이, 자료 안의 변동만으로 비교할 수 있다"는 생각으로 요약했다. σ를 모르는데도 $\bar{X}$의 오차를 S로 재고, 그 비의 분포를 정확히 아는 것, 그것이 Student의 t다. 두 표본 t와 F, 14회의 분산분석이 모두 "자료 안의 한 변동을 다른 변동과 비교"하는 같은 설계를 따른다. 그리고 S^2는 평균을 맞추고 남은 잔차의 제곱합이므로 일곱째 기둥 **Residual**과도 닿는다. 13회의 회귀에서 자유도 $n-p$로 다시 만난다.

9.4 다음 회 — 극한의 언어: 수렴 · CLT · Monte Carlo (1777 → 1949)

이 노트의 정확한 분포는 모두 정규 가정에서 나왔다. 그러면 7절의 표에서 지수 표본의 포함확률이 $n=100$에서 94%로 올라간 것은 왜인가? 그리고 σ 대신 S를 넣고 1.96을 써도 큰 표본에서는 괜찮다는 19세기의 관행은 어떤 정리로 정당화되나? 7회는 확률수렴과 분포수렴을 정의하고, 중심극한정리와 Slutsky의 정리로 이 두 질문에 답한다. Student의 카드 실험이 그 오차까지 말할 수 있는 Monte Carlo 방법으로 자라는 과정도 따라간다.

참고문헌

- Student. The probable error of a mean. *Biometrika* 6 (1908), 1–25.
- Student. Errors of routine analysis. *Biometrika* 19 (1927), 151–164.
- Fisher, R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10 (1915). · Applications of "Student's" distribution. *Metron* 5 (1925). · *Statistical Methods for Research Workers*. Oliver and Boyd, 1925.
- Helmert, F. R. Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit. *Astronomische Nachrichten* 88 (1876).
- Pfanzagl, J., Sheynin, O. A forerunner of the t-distribution. *Biometrika* 83 (1996), 891–898.
- Macdonell, W. R. On criminal anthropometry and the identification of criminals. *Biometrika* 1 (1902), 177–227.
- Cushny, A. R., Peebles, A. R. The action of optical isomers. II. Hyoscines. *Journal of Physiology* 32 (1905), 501–510.
- Senn, S., Richardson, W. The first t-test. *Statistics in Medicine* 13 (1994), 785–803.

- Zabell, S. L. On Student's 1908 article "The probable error of a mean". *JASA* 103 (2008), 1–7.
- Hanley, J. A., Julien, M., Moodie, E. E. M. Student's z, t, and s: what if Gosset had R? *The American Statistician* 62 (2008), 64–69.
- Snedecor, G. W. *Calculation and Interpretation of Analysis of Variance and Covariance*. Collegiate Press, 1934.
- 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.
- Welch, B. L. The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34 (1947), 28–35.
- Geary, R. C. The distribution of "Student's" ratio for non-normal samples. *JRSS Suppl.* 3 (1936). · Lukacs, E. A characterization of the normal distribution. *Ann. Math. Statist.* 13 (1942), 91–93.
- Box, G. E. P. Non-normality and tests on variances. *Biometrika* 40 (1953), 318–335.
- Ioffe, S., Szegedy, C. Batch normalization. *ICML* (2015). · Ioffe, S. Batch renormalization. *NeurIPS* (2017). · Wu, Y., He, K. Group normalization. *ECCV* (2018).
- van der Maaten, L., Hinton, G. Visualizing data using t-SNE. *JMLR* 9 (2008), 2579–2605.
- Lange, K. L., Little, R. J. A., Taylor, J. M. G. Robust statistical modeling using the t distribution. *JASA* 84 (1989), 881–896.
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016.
- 자료: R 패키지 `HistData`의 `MacdonellDF`, `CushnyPeebles`; `datasets`의 `sleep`.

A

연습문제 풀이

각 풀이의 머리에 있는 문제 ○○쪽을 누르면 문제로 돌아간다. 계산 문제의 수치는 노트북에서 같은 계산으로 확인했다. 유도형 연습의 풀이는 본문에서 연습으로 돌린 정리의 완성된 증명이다.

1절

연습 1.1 문제 6쪽

$\sum(X_i-\bar X)^2=\sum X_i^2-2\bar X\sum X_i+n\bar X^2=\sum X_i^2-n\bar X^2$. 또

$$\sum_{i,j}(X_i-X_j)^2=\sum_{i,j}(X_i^2+X_j^2-2X_iX_j)=2n\sum_i X_i^2-2\Big(\sum_i X_i\Big)^2=2n\Big(\sum X_i^2-n\bar X^2\Big).$$

$i\neq j$이면 $E(X_i-X_j)^2=\operatorname{Var}(X_i-X_j)=2\sigma^2$이고 $i=j$인 항은 0이므로 $E\sum_{i,j}(X_i-X_j)^2=n(n-1)\cdot 2\sigma^2$. $2n$으로 나누면 $(n-1)\sigma^2$이다. 이 식은 μ를 전혀 쓰지 않는다. S^2는 "두 관측의 차이의 제곱의 절반"의 평균이다.

연습 1.2 문제 6쪽

$\mu=0$으로 놓는다. 앞 연습으로 $S^2=\frac{1}{N}\sum_{i<j}h_{ij}$, $h_{ij}=\frac12(X_i-X_j)^2$, $N=\binom{n}{2}$이다. 쌍 (i,j)와 (k,l)이 겹치지 않으면 h_{ij}와 h_{kl}은 독립이다. 한 첨자만 겹치면(예: h_{ij}와 h_{ik}) X_i가 주어질 때 둘은 독립이고, $g(X_i)=E(h_{ij}\mid X_i)=\frac{X_i^2+\sigma^2}{2}$이다. 반복기댓값(3회 정리 2.5)으로 $E(h_{ij}h_{ik})=E\,g(X_i)^2$, $Eh_{ij}=E\,g(X_i)$이므로

$$\operatorname{Cov}(h_{ij},h_{ik})=\operatorname{Var}g(X_i)=\frac14\operatorname{Var}X_i^2=\frac{\mu_4-\sigma^4}{4}=:\zeta_1.$$

같은 쌍이면 $\operatorname{Var}h_{ij}=\frac14\big(E(X_i-X_j)^4-4\sigma^4\big)=\frac14(2\mu_4+6\sigma^4-4\sigma^4)=\frac{\mu_4+\sigma^4}{2}=:\zeta_2$. 쌍 하나마다 한 첨자만 겹치는 쌍이 $2(n-2)$개이므로

$$\operatorname{Var}S^2=\frac{1}{N^2}\cdot N\big(\zeta_2+2(n-2)\zeta_1\big)=\frac{(\mu_4+\sigma^4)+(n-2)(\mu_4-\sigma^4)}{n(n-1)}=\frac1n\Big(\mu_4-\frac{n-3}{n-1}\sigma^4\Big).$$

연습 1.3 문제 6쪽

$\bar X=\frac{X_1+X_2}{2}$이므로 $X_1-\bar X=\frac{X_1-X_2}{2}=-(X_2-\bar X)$이고 $S^2=\frac11\cdot 2\cdot\frac{(X_1-X_2)^2}{4}=\frac{(X_1-X_2)^2}{2}$. $X_1-X_2\sim\mathcal N(0,2\sigma^2)$이므로 $\frac{X_1-X_2}{\sqrt{2}\sigma}\sim\mathcal N(0,1)$이고 S^2/σ^2는 그 제곱, 곧 $\chi^2(1)$(4회 정리 3.9).

연습 1.4 문제 6쪽

$Q=\sum(X_i-\bar X)^2$이면 $EQ=(n-1)\sigma^2$, $\operatorname{Var}Q=2(n-1)\sigma^4$(따름정리 2.3). 평균제곱오차는 분산 더하기 편의의 제곱이므로

$$\operatorname{MSE}(c)=c^2\cdot 2(n-1)\sigma^4+\big(c(n-1)-1\big)^2\sigma^4.$$

c로 미분해 0으로 놓으면 $4c(n-1)+2(n-1)\big(c(n-1)-1\big)=0$, 곧 $c(n+1)=1$. 그때 $\operatorname{MSE}=\frac{2\sigma^4}{n+1}$이고, $c=\frac{1}{n-1}$이면 $\frac{2\sigma^4}{n-1}$, $c=\frac{1}{n}$이면 $\frac{(2n-1)\sigma^4}{n^2}$이다. 불편인 S^2가 평균제곱오차로는 가장 나쁘다. 조금 작게 어림하는 대신 흔들림을 줄이는 쪽이 이득이다.

2절

연습 2.1 문제 9쪽

$$H=\begin{pmatrix}\frac{1}{\sqrt3} & \frac{1}{\sqrt3} & \frac{1}{\sqrt3}\\ \frac{1}{\sqrt2} & -\frac{1}{\sqrt2} & 0\\ \frac{1}{\sqrt6} & \frac{1}{\sqrt6} & -\frac{2}{\sqrt6}\end{pmatrix}.$$

각 행의 제곱합은 1이고 서로 다른 행의 내적은 $\frac{1}{\sqrt6}(1-1)$, $\frac{1}{\sqrt{18}}(1+1-2)$, $\frac{1}{\sqrt{12}}(1-1)$로 모두 0이다. $Y_2=\frac{Z_1-Z_2}{\sqrt2}$, $Y_3=\frac{Z_1+Z_2-2Z_3}{\sqrt6}$이므로

$$Y_2^2+Y_3^2=\frac{3(Z_1-Z_2)^2+(Z_1+Z_2-2Z_3)^2}{6}=\frac{2}{3}\Big(\sum_i Z_i^2-Z_1Z_2-Z_1Z_3-Z_2Z_3\Big).$$

한편 $\sum(Z_i-\bar Z)^2=\sum Z_i^2-\frac{1}{3}\big(\sum Z_i\big)^2=\frac{2}{3}\sum Z_i^2-\frac{2}{3}\sum_{i<j}Z_iZ_j$로 같다.

연습 2.2 문제 9쪽

$\operatorname{Cov}(\bar X,X_i)=\frac{1}{n}\sum_j\operatorname{Cov}(X_j,X_i)=\frac{\sigma^2}{n}$이므로 $\operatorname{Cov}(\bar X,X_i-\bar X)=\frac{\sigma^2}{n}-\operatorname{Var}\bar X=0$. 정규 표본이면 $(\bar X,X_1-\bar X,\dots,X_n-\bar X)'$는 $\mathbf{X}$의 선형변환이므로 결합정규(5회 정리 A)이고, $\bar X$와 잔차벡터 사이의 공분산 블록이 0이므로 5회 정리 B로 독립이다. S^2는 잔차벡터의 함수이므로 $\bar X\perp S^2$.

연습 2.3 문제 9쪽

$EX=1$, $EX^2=2$, $EX^3=6$. $(X_1+X_2)(X_1-X_2)^2=X_1^3+X_2^3-X_1^2X_2-X_1X_2^2$이므로 독립으로 기댓값은 $6+6-2-2=8$. $\bar XS^2=\frac{(X_1+X_2)(X_1-X_2)^2}{4}$이므로 $E(\bar XS^2)=2$. $E\bar X=1$, $ES^2=\operatorname{Var}X=1$이므로 $\operatorname{Cov}(\bar X,S^2)=2-1=1\neq0$. 큰 평균은 큰 흩어짐과 함께 온다(7절 명제 7.1의 $\mu_3/n=2/2$).

연습 2.4 문제 9쪽

$V=(n-1)S^2/\sigma^2\sim\chi^2(r)$, $r=n-1$이면 $S=\sigma\sqrt{V/r}$이고

$$E\sqrt{V}=\int_0^\infty v^{1/2}\frac{v^{r/2-1}e^{-v/2}}{\Gamma(r/2)2^{r/2}}\,dv=\frac{\Gamma\left(\frac{r+1}{2}\right)2^{(r+1)/2}}{\Gamma(r/2)2^{r/2}}=\sqrt{2}\,\frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)}.$$

$r=n-1$을 넣으면 $ES=\sigma\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$. $c_4(4)=\sqrt{\frac{2}{3}}\frac{\Gamma(2)}{\Gamma(3/2)}=\sqrt{\frac{2}{3}}\cdot\frac{2}{\sqrt{\pi}}=0.921$, $c_4(10)=\sqrt{\frac{2}{9}}\frac{\Gamma(5)}{\Gamma(9/2)}=0.973$.

연습 2.5 문제 9쪽

정리 2.2(c)로 $P\left(\chi^2_{n-1,0.975}\le\frac{(n-1)S^2}{\sigma^2}\le\chi^2_{n-1,0.025}\right)=0.95$이고, 괄호 안을 σ^2에 대해 풀면 주어진 구간이다. μ와 σ에 상관없이 성립한다. $n=4$이면 $\chi^2_{3,0.975}=0.216$, $\chi^2_{3,0.025}=9.348$이므로 구간은 $[3/9.348,\ 3/0.216]=[0.321,\ 13.90]$. 네 관측으로는 분산을 40배 넘는 범위로밖에 말할 수 없다.

연습 2.6 문제 9쪽

(a) 첫 행의 길이의 제곱은 $n\cdot\frac{1}{n}=1$이다. k번째 행은 1이 $k-1$개, $-(k-1)$이 하나이므로 $\frac{(k-1)+(k-1)^2}{k(k-1)}=\frac{(k-1)k}{k(k-1)}=1$이다. (b) 첫 행의 성분은 모두 $\frac{1}{\sqrt{n}}$이므로 내적은 $\frac{1}{\sqrt{n}}\cdot\frac{1}{\sqrt{k(k-1)}}\big[(k-1)-(k-1)\big]=0$이다. (c) j번째 행은 처음 j자리에만 0이 아닌 성분이 있다. $j\le k-1$이므로 그 자리에서 k번째 행의 성분은 모두 $\frac{1}{\sqrt{k(k-1)}}$이다. 내적은 이 상수와 j번째 행의 성분의 합의 곱이고, 그 합은 $\frac{1}{\sqrt{j(j-1)}}\big[(j-1)-(j-1)\big]=0$이다. (d) (a)–(c)는 $HH'=I$이다. $\det H\cdot\det H'=1$이므로 H는 가역이고, H'는 H의 오른쪽 역행렬이므로 곧 역행렬이다. 따라서 $H'H=I$이고 $\|H\mathbf{z}\|^2=\mathbf{z}'H'H\mathbf{z}=\|\mathbf{z}\|^2$이다. 정리 2.2의 증명 ⑤에서 $\|H\mathbf{Z}\|^2=\|\mathbf{Z}\|^2$로 쓴다.

3절

연습 3.1 문제 14쪽

Stirling을 $\Gamma(x)=\Gamma(x+1)/x$와 $\Gamma(x+\frac{1}{2})=\Gamma\big((x-\frac{1}{2})+1\big)$에 쓰면

$$\Gamma(x)\sqrt{x}\sim\frac{\sqrt{2\pi x}\,x^xe^{-x}}{x}\sqrt{x}=\sqrt{2\pi}\,x^xe^{-x},$$
$$\Gamma(x+\tfrac{1}{2})\sim\sqrt{2\pi(x-\tfrac{1}{2})}\,(x-\tfrac{1}{2})^{x-\frac{1}{2}}e^{-(x-\frac{1}{2})}=\sqrt{2\pi}\,(x-\tfrac{1}{2})^xe^{-x+\frac{1}{2}}.$$

비는 $\left(1-\frac{1}{2x}\right)^xe^{1/2}\to e^{-1/2}e^{1/2}=1$이다. $x=r/2$로 두면 $t(r)$의 상수는 $\frac{\Gamma(x+1/2)}{\Gamma(x)\sqrt{2\pi x}}\to\frac{1}{\sqrt{2\pi}}$.

연습 3.2 문제 14쪽

$\Gamma(\frac{r}{2},\frac{2}{r})$는 $\Gamma(\frac{r}{2},2)/r$, 곧 V/r ($V\sim\chi^2(r)$)이다. $W=w$일 때 $T\sim\mathcal{N}(0,1/w)$라는 것은 $T=Z/\sqrt{W}$

$(Z\sim\mathcal{N}(0,1),\,Z\perp W)$와 같은 결합분포를 준다. 따라서 $T=Z/\sqrt{V/r}\sim t(r)$. 직접 적분해도 된다: $\int_0^\infty \sqrt{\frac{w}{2\pi}}e^{-wt^2/2}f_W(w)\,dw$는 감마 적분이 되어 정리 3.2의 밀도를 준다.

연습 3.3 문제 14쪽

$r=2$이면 상수는 $\frac{\Gamma(3/2)}{\sqrt{2\pi}\Gamma(1)}=\frac{\sqrt{\pi}/2}{\sqrt{2\pi}}=\frac{1}{2\sqrt{2}}$이고 $\left(1+\frac{t^2}{2}\right)^{-3/2}=\frac{2^{3/2}}{(2+t^2)^{3/2}}$이므로 $f_2(t)=(2+t^2)^{-3/2}$. $\frac{d}{dt}\frac{t}{\sqrt{2+t^2}}=\frac{2}{(2+t^2)^{3/2}}$이고 $t\to-\infty$에서 $\frac{t}{\sqrt{2+t^2}}\to-1$이므로 $F_2(t)=\frac{1}{2}\left(1+\frac{t}{\sqrt{2+t^2}}\right)$. $t=4.303$이면 $\frac{4.303}{\sqrt{20.52}}=0.950$이므로 $F_2=0.975$.

연습 3.4 문제 14쪽

	$r=3$	$r=10$	$r=30$	$\mathcal{N}(0,1)$
$P(\lvert T\rvert>3)$	0.0577 (21배)	0.0133 (4.9배)	0.0054 (2.0배)	0.0027
$P(\lvert T\rvert>4)$	0.0280 (442배)	0.0025 (40배)	0.00038 (6.0배)	0.000063

꼬리로 갈수록 비가 커진다. t의 꼬리는 거듭제곱 $|t|^{-(r+1)}$로, 정규의 꼬리는 $e^{-t^2/2}$로 줄기 때문이다.

연습 3.5 문제 14쪽

상수를 $c_r=\frac{\Gamma((r+1)/2)}{\sqrt{r\pi}\,\Gamma(r/2)}$로 쓴다. (a) $f_r(t)=c_r(1+t^2/r)^{-(r+1)/2}$는 t^2에만 의존하므로 $f_r(-t)=f_r(t)$이고 분포는 0에 대해 대칭이다. (b) $|t|^{r+1}f_r(t)=c_r\left(\frac{t^2}{1+t^2/r}\right)^{(r+1)/2}\to c_r\,r^{(r+1)/2}=:L>0$이다. 따라서 어떤 t_0가 있어 $|t|\ge t_0$에서 $\frac{L}{2}|t|^{-(r+1)}\le f_r(t)\le 2L|t|^{-(r+1)}$이다. $|t|\le t_0$ 부분의 적분은 유한하고, $\int_{t_0}^\infty t^{k-r-1}\,dt$는 $k-r-1<-1$, 곧 $k<r$일 때만 유한하다. 비교 판정으로 $E|T|^k<\infty\iff k<r$이다. (c) $T^2=Z^2\cdot\frac{r}{V}$이고 $Z^2\perp\frac{r}{V}$이므로 3회 정리 3.4(b)로 $ET^2=EZ^2\cdot E\frac{r}{V}$이다(두 인수가 음이 아니므로 값이 무한대여도 등식이 성립한다). 감마 적분과 $\Gamma(x+1)=x\Gamma(x)$로

$$E\frac{1}{V}=\int_0^\infty \frac{1}{v}\cdot\frac{v^{r/2-1}e^{-v/2}}{\Gamma(r/2)2^{r/2}}\,dv=\frac{\Gamma(r/2-1)\,2^{r/2-1}}{\Gamma(r/2)\,2^{r/2}}=\frac{1}{2(r/2-1)}=\frac{1}{r-2}\qquad(r>2).$$

따라서 $ET^2=\frac{r}{r-2}$이다. $r>1$이면 (b)로 $E|T|<\infty$이고 (a)의 대칭으로 $ET=0$이므로 $\operatorname{Var}T=ET^2=\frac{r}{r-2}$이다. (d) $r=1$이면 $c_1=\frac{\Gamma(1)}{\sqrt{\pi}\,\Gamma(1/2)}=\frac{1}{\pi}$이고 $f_1(t)=\frac{1}{\pi(1+t^2)}$, 곧 Cauchy 밀도다. (e) $\log(1+x)=x+O(x^2)$이므로 $-\frac{r+1}{2}\log\left(1+\frac{t^2}{r}\right)=-\frac{r+1}{2}\left(\frac{t^2}{r}+O(r^{-2})\right)\to-\frac{t^2}{2}$이다. 상수는 $x=r/2$로 두면 $c_r=\frac{\Gamma(x+1/2)}{\Gamma(x)\sqrt{x}}\cdot\frac{1}{\sqrt{2\pi}}$이고, 연습 3.1로 첫 인수는 1로 간다. 따라서 모든 t에서 $f_r(t)\to\frac{1}{\sqrt{2\pi}}e^{-t^2/2}=\phi(t)$이다.

4절

연습 4.1 문제 16쪽

$EF^2 = \left(\frac{r_2}{r_1}\right)^2 EU^2\, E\frac{1}{V^2}$. $EU^2 = \operatorname{Var} U + (EU)^2 = r_1(r_1+2)$이고, 4회 연습 3.4($k=-2$)로 $E\frac{1}{V^2} = \frac{\Gamma(r_2/2-2)}{4\Gamma(r_2/2)} = \frac{1}{(r_2-2)(r_2-4)}$. 따라서

$$\operatorname{Var} F = \frac{r_2^2(r_1+2)}{r_1(r_2-2)(r_2-4)} - \frac{r_2^2}{(r_2-2)^2} = \frac{r_2^2\big[(r_1+2)(r_2-2) - r_1(r_2-4)\big]}{r_1(r_2-2)^2(r_2-4)} = \frac{2r_2^2(r_1+r_2-2)}{r_1(r_2-2)^2(r_2-4)}.$$

연습 4.2 문제 16쪽

$F \sim F(r_1, r_2)$이면 $1/F \sim F(r_2, r_1)$이다. $x = 1/F_{r_2,r_1,p}$로 두면 $P(F \le x) = P(1/F \ge F_{r_2,r_1,p}) = p$, 곧 $P(F > x) = 1-p$이므로 $x = F_{r_1,r_2,1-p}$.

연습 4.3 문제 16쪽

$R = \frac{S_X^2/S_Y^2}{\sigma_1^2/\sigma_2^2} \sim F(m-1, n-1)$이므로 $P(F_{m-1,n-1,0.975} \le R \le F_{m-1,n-1,0.025}) = 0.95$. σ_1^2/σ_2^2에 대해 풀면

$$\left[\frac{S_X^2/S_Y^2}{F_{m-1,n-1,0.025}},\ \frac{S_X^2/S_Y^2}{F_{m-1,n-1,0.975}}\right].$$

예제의 자료에서 $F_{10,10,0.025} = 3.717$이고 연습 4.2로 $1/F_{10,10,0.975} = 3.717$이므로 구간은 $[2.8/3.717,\ 2.8\times 3.717] = [0.753,\ 10.41]$. 1을 포함한다. 두 기계의 정밀도가 같다는 것과 한쪽이 10배 나쁘다는 것이 모두 자료와 맞는다.

연습 4.4 문제 16쪽

$a = \frac{r_1}{2}$, $b = \frac{r_2}{2}$로 쓴다. (a) 4회 정의 3.8로 $\chi^2(r_1) = \Gamma(a, 2)$이므로 척도를 바꾸면 $X = U/2 \sim \Gamma(a,1)$이고, 같은 방법으로 $Y = V/2 \sim \Gamma(b,1)$이다. $R = U/V = X/Y$의 밀도는 4회 연습 4.4로 $f_R(x) = \frac{x^{a-1}}{B(a,b)(1+x)^{a+b}}$ $(x>0)$이다. (b) $F = \frac{r_2}{r_1}R$이므로 척도 변환으로

$$f_F(x) = \frac{r_1}{r_2} f_R\!\left(\frac{r_1}{r_2}x\right) = \frac{1}{B(a,b)}\left(\frac{r_1}{r_2}\right)^a \frac{x^{a-1}}{\left(1+\frac{r_1}{r_2}x\right)^{a+b}}.$$

(c) $U \perp V$이므로 $EF = E\frac{U}{r_1}\cdot E\frac{r_2}{V} = 1\cdot\frac{r_2}{r_2-2}$이다(연습 3.5(c)). $r_2 \le 2$이면 $E\frac{1}{V} = \int_0^\infty v^{r_2/2-2}e^{-v/2}\,dv/(\cdots)$의 피적분함수가 0 근처에서 v^{-1}보다 느리지 않게 커지므로 적분이 발산하고, $EF = \infty$이다. (d) $1/F = \frac{V/r_2}{U/r_1}$는 정의 4.1에서 두 자리를 바꾼 꼴이므로 $F(r_2, r_1)$이다. $T = Z/\sqrt{V/r}$이면 $T^2 = \frac{Z^2/1}{V/r}$이고, $Z^2 \sim \chi^2(1)$(4회 정리 3.9)과 V는 독립이므로 $F(1,r)$이다. (e) $\frac{r_1}{r_2}F = R = \frac{X}{Y}$이므로 $\frac{R}{1+R} = \frac{X}{X+Y}$이고, 4회 정리 4.2로 $\operatorname{Beta}(a,b)$이다.

5절

연습 5.1 문제 21쪽

비 $t_{n-1,0.025}c_4(n)/1.96$은 $n=4$에서 $3.182\times 0.921/1.96 = 1.50$, $n=10$에서 $2.262\times 0.973/1.96 = 1.12$, $n=30$에서 1.035다. σ를 모르는 값은 $n=4$에서 구간 길이의 50%, $n=30$에서 3.5%다.

연습 5.2 문제 21쪽

$t_{n-1,0.025}c_4(n)/\sqrt{n}$은 $n=17$에서 0.506, $n=18$에서 0.490이므로 $n=18$. σ를 알면 $1.96/\sqrt{n}\le \frac{1}{2}$에서 $n=16$이다. σ를 모르는 값은 관측 두 개다.

연습 5.3 문제 21쪽

$s_X^2 = 3.201$, $s_Y^2 = 4.009$, $r = 0.795$이므로 $s_X^2+s_Y^2-2rs_Xs_Y = 3.201+4.009-2(0.795)(1.789)(2.002) = 1.513 = s_d^2$. 대응하지 않으면 차이의 분산을 $s_X^2+s_Y^2 = 7.21$로 어림하므로, 대응이 분산을 $7.21/1.513 = 4.8$배 줄였다. 표준오차로는 0.849에서 0.389로 2.2배다.

연습 5.4 문제 21쪽

5만 번 이상의 시뮬레이션에서 합동 t의 포함확률은 약 0.73, Welch 구간은 0.95다. 합동분산은 큰 표본(분산 1)에 무게를 두어 $S_p^2 \approx \frac{4\cdot 9+19\cdot 1}{23} = 2.4$가 되고, 이것으로 어림한 $\mathrm{Var}(\bar X-\bar Y)$는 $2.4(\frac{1}{5}+\frac{1}{20}) = 0.60$이지만 참값은 $\frac{9}{5}+\frac{1}{20} = 1.85$다. 구간이 $\sqrt{1.85/0.60}\approx 1.76$배 너무 짧다. 거꾸로 큰 표본 쪽의 분산이 크면 포함확률이 0.998로 지나치게 보수적이다.

연습 5.5 문제 21쪽

$P\big(\mu \le \bar X + t_{n-1,0.05}S/\sqrt{n}\big) = P(T \ge -t_{n-1,0.05}) = 1-0.05$ (t의 대칭). `sleep`의 차이에서 $t_{9,0.05} = 1.833$이므로 위쪽 한계는 $1.58 + 1.833\cdot 1.23/\sqrt{10} = 2.29$, 같은 방법의 아래쪽 한계는 0.87이다. “L-hyoscine은 수면을 적어도 0.87시간 더 늘린다”가 95% 한쪽 문장이다.

6절

연습 6.1 문제 24쪽

씨앗 1907에서 손가락 길이의 $|t| > 3.182$ 비율은 6.3%(섞기 400번의 평균 4.7%), $t(3)$과의 Kolmogorov–Smirnov 유의확률은 0.94다. $3s^2/\sigma^2$와 $\chi^2(3)$의 비교에서 유의확률은 0.31로, 키

의 0.03보다 훨씬 잘 맞는다. 400번 평균으로도 키 0.08, 손가락 0.48이다. 계급이 촘촘할수록(손가락 0.1cm, 키 1인치) 연속 분포의 이론과 잘 맞는다. $s=0$인 표본도 손가락에서는 거의 나오지 않는다(평균 0.12개).

연습 6.2 문제 24쪽

$s'^2=\frac{n-1}{n}S^2$이므로 $z=\frac{\bar{x}-\mu}{S\sqrt{(n-1)/n}}=\frac{\sqrt{n}(\bar{x}-\mu)}{S}\cdot\frac{1}{\sqrt{n-1}}=\frac{T}{\sqrt{n-1}}$. $r=n-1$로 두면 2회 예제 4.5(척도)로 $f_z(z)=\sqrt{r}\,f_r(\sqrt{r}z)\propto\left(1+\frac{rz^2}{r}\right)^{-(r+1)/2}=(1+z^2)^{-n/2}$.

연습 6.3 문제 24쪽

키 자료의 20개 계급 비율로 $\sum_k p_k^4=0.00197$이고, 750개 가운데 기대 개수는 1.48이다. 되돌려 넣지 않으면 계급 k의 개수를 c_k라 할 때 $\sum_k \frac{c_k(c_k-1)(c_k-2)(c_k-3)}{N(N-1)(N-2)(N-3)}=0.00194$로 거의 같다. 실제로 섞기 400번에서 $s=0$인 표본은 평균 1.50개였다.

연습 6.4 문제 24쪽

400번의 비율들의 표준편차는 0.0077로 이항 근사 $\sqrt{0.05\cdot0.95/750}=0.0080$과 거의 같다. 750개의 표본이 거의 독립인 시행처럼 행동한다. 씨앗 1908의 3.2%는 평균 5.0%보다 표준편차의 2.3배 아래이고, 400번 가운데 이만큼 낮은 것은 1번(0.25%)뿐이었다. 드물지만 일어난다. 한 번의 Monte Carlo 실험은 그 자체가 오차를 가진 표본이다. Student의 실험도 마찬가지였고, 7회에서 이 오차를 정리로 다룬다.

7절

연습 7.1 문제 28쪽

Exp(1)에서 $\mu_3=E(X-1)^3=EX^3-3EX^2+3EX-1=6-6+3-1=2$. 명제 7.1로 $\operatorname{Cov}(\bar{X},S^2)=\frac{2}{n}$이고 $n=2$이면 1로 연습 2.3와 같다.

연습 7.2 문제 28쪽

3회 정리 4.2로 $\operatorname{Var}\bar{X}=\frac{1}{n^2}\sum_{i,j}\sigma^2\rho^{|i-j|}$. $|i-j|=k\ge1$인 순서쌍이 $2(n-k)$개이므로

$$\operatorname{Var}\bar{X}=\frac{\sigma^2}{n^2}\Big[n+2\sum_{k=1}^{n-1}(n-k)\rho^k\Big]=\frac{\sigma^2}{n}\Big[1+2\sum_{k=1}^{n-1}\Big(1-\frac{k}{n}\Big)\rho^k\Big].$$

$\sum_{k\ge1}\rho^k=\frac{\rho}{1-\rho}$이고 $\frac{1}{n}\sum_k k|\rho|^k\le\frac{|\rho|}{n(1-|\rho|)^2}\to0$이므로 괄호는 $1+\frac{2\rho}{1-\rho}=\frac{1+\rho}{1-\rho}$로 간다. $\rho<0$이면 이 값이 1보다 작아 $\operatorname{Var}\bar{X}$를 크게 어림하므로 구간은 지나치게 넓다(보수적).

연습 7.3 문제 28쪽

델타 방법: $\log S^2 \approx \log\sigma^2 + \frac{S^2-\sigma^2}{\sigma^2}$ 이므로 $\operatorname{Var}\log S^2 \approx \frac{\operatorname{Var}S^2}{\sigma^4} \approx \frac{2+\kappa}{n}$. 두 독립 표본의 차 $\log F = \log S_X^2 - \log S_Y^2$ 의 분산은 $\frac{2(2+\kappa)}{n}$ 이다. 정규 이론의 검정은 이 분산을 $\frac{4}{n}$ 으로 보고 $|\log F| > 1.96\sqrt{4/n}$ 이면 기각하므로 실제 기각률은 $P\big(|Z|\sqrt{2(2+\kappa)/n} > 1.96\sqrt{4/n}\big) = P\big(|Z| > 1.96/\sqrt{1+\kappa/2}\big)$. Laplace는 0.215, 지수는 0.327 이고, 시뮬레이션의 $n = 101$ 값 0.21 과 0.31 에 가깝다. 균등($\kappa = -1.2$)이면 0.002 로 거의 기각하지 못한다.

연습 7.4 문제 28쪽

꼬리가 두꺼우면 가끔 아주 큰 관측이 나온다. 그 관측은 $\bar X$ 를 M/n 만큼, S 를 $M/\sqrt{n}$ 만큼 움직이므로 $|T|$ 는 1 근처로 끌려간다(명제 7.3). 정규 이론에서 $|T|$ 가 큰 경우는 "$\bar X$ 가 우연히 멀고 S 가 우연히 작은" 경우인데, 두꺼운 꼬리에서는 $\bar X$ 가 멀 때 대개 S 도 크다. 그래서 $|T| > t_{n-1,0.025}$ 인 경우가 줄어든다. 대칭이므로 7.1 절의 한쪽 쏠림도 없다.

연습 7.5 문제 28쪽

(a) X_i 를 $X_i - \mu$ 로 바꾸면 편차 $X_i - \bar X$ 는 그대로이므로 S^2 는 변하지 않고, $\bar X$ 는 상수 μ 만큼 움직인다. 공분산은 상수를 더해도 변하지 않으므로 $\mu = 0$ 으로 놓아도 된다. 그때 $E\bar X = 0$ 이므로 $\operatorname{Cov}(\bar X, S^2) = E(\bar X S^2)$ 이다. (b) $i \neq j$ 이면 $X_i \perp X_j^2$ 이므로 $E(X_i X_j^2) = EX_i \cdot EX_j^2 = 0$ 이고, $i = j$ 이면 $EX_i^3 = \mu_3$ 이다. $E(X_iX_jX_k)$ 에서 세 첨자 가운데 한 번만 나오는 첨자가 있으면, 그 변수는 나머지와 독립이고 평균이 0 이므로 기댓값은 0 이다. 세 첨자가 모두 같을 때만 그런 첨자가 없고, 그때 값은 μ_3 이다. (c) $E\big(\bar X \sum_j X_j^2\big) = \frac{1}{n}\sum_{i,j} E(X_iX_j^2) = \frac{1}{n}\cdot n\mu_3 = \mu_3$ 이고, $E(n\bar X^3) = \frac{n}{n^3}\sum_{i,j,k} E(X_iX_jX_k) = \frac{n\mu_3}{n^2} = \frac{\mu_3}{n}$ 이다. (d) $(n-1)E(\bar X S^2) = E\big(\bar X \sum_j X_j^2\big) - E(n\bar X^3) = \mu_3\big(1 - \frac{1}{n}\big)$ 이므로 $\operatorname{Cov}(\bar X, S^2) = \frac{\mu_3}{n}$ 이다. 기댓값이 유한한 것은 AM–GM [T4]로 $|X_iX_jX_k| \le \frac{1}{3}(|X_i|^3 + |X_j|^3 + |X_k|^3)$ 이고 $E|X|^3 < \infty$ 이기 때문이다.

연습 7.6 문제 28쪽

(a) $\bar x = \frac{1}{n}(M + \sum_{i\ge 2} x_i) = \frac{M}{n} + a$. (b) $\sum_i x_i^2 = M^2 + b$ 이고 $n\bar x^2 = \frac{M^2}{n} + 2aM + na^2$ 이므로 $(n-1)s^2 = \frac{n-1}{n}M^2 - 2aM + b - na^2$ 이다. (c) $s^2 = \frac{M^2}{n}\big(1 - \frac{2an}{(n-1)M} + O(M^{-2})\big)$ 이고 $\sqrt{1+u} = 1 + O(u)$ 이므로 $s = \frac{M}{\sqrt{n}}\big(1 + O(\frac{1}{M})\big)$ 이다 [T8]. (d) $t = \frac{\sqrt{n}(M/n + a)}{(M/\sqrt{n})(1+O(1/M))} = \frac{1 + an/M}{1+O(1/M)} \to 1$ 이다. $x_1 = -M$ 이면 $\bar x = -\frac{M}{n} + a$ 이고 s 의 주된 항은 그대로이므로 $t \to -1$ 이다. 이상값 하나는 부호만 남기고 크기를 지운다.

8절

연습 8.1 문제 32쪽

$n=4$: $3.182\sqrt{1.25}=3.56$, $n=10$: $2.262\sqrt{1.1}=2.37$, $n=100$: $1.984\sqrt{1.01}=1.99$. 극한은 1.96이다. 평균의 구간은 0으로 줄지만 예측구간의 반폭은 1.96σ 아래로 내려가지 않는다.

연습 8.2 문제 32쪽

$\hat\sigma_B = S\sqrt{(B-1)/B}$이고 $T=\frac{x_{\text{new}}-\bar x_B}{S\sqrt{1+1/B}}\sim t(B-1)$이므로

$$\frac{x_{\text{new}}-\bar x_B}{\hat\sigma_B}=T\sqrt{1+\tfrac1B}\sqrt{\tfrac{B}{B-1}}=T\sqrt{\tfrac{B+1}{B-1}}.$$

정리 3.3(b)로 $\operatorname{Var}T=\frac{B-1}{B-3}$이므로 분산은 $\frac{B+1}{B-3}$.

연습 8.3 문제 32쪽

잔차벡터 $\mathbf{R}=(x_1-\bar x,\dots,x_B-\bar x)$는 $\mathbf{1}$에 수직인 $(B-1)$차원 공간 L에 있다. e_1을 L에 사영한 벡터 $e_1-\frac1B\mathbf{1}$의 길이는 $\sqrt{(B-1)/B}$이고, 그 방향의 단위벡터 u를 첫째로 하는 L의 정규직교기저를 잡으면(Helmert와 같은 방식) 그 좌표 $W_1,\dots,W_{B-1}$은 정리 2.2의 증명처럼 독립인 $\mathcal N(0,\sigma^2)$이다. 그러면 $x_1-\bar x=\langle\mathbf{R},e_1\rangle=\sqrt{(B-1)/B}\,W_1$이고 $B\hat\sigma_B^2=\|\mathbf{R}\|^2=\sum_k W_k^2$이므로

$$\frac{\hat x_1^2}{B-1}=\frac{W_1^2}{W_1^2+\sum_{k\ge2}W_k^2}.$$

$W_1^2/2\sigma^2\sim\Gamma(\frac12,1)$과 $\sum_{k\ge2}W_k^2/2\sigma^2\sim\Gamma(\frac{B-2}{2},1)$이 독립이므로 4회 정리 4.2로 $\operatorname{Beta}(\frac12,\frac{B-2}{2})$. $B=2$이면 분모의 둘째 항이 없어 값이 1, 곧 $\hat x_1=\pm1$이다. $B=4$이면 $U=\hat x_1^2/3\sim\operatorname{Beta}(\frac12,1)$, 곧 $P(U\le u)=\sqrt u$이므로 $P(|\hat x_1|\le z)=P(U\le z^2/3)=z/\sqrt3$ $(0\le z\le\sqrt3)$. 대칭과 합치면 $\hat x_1\sim U(-\sqrt3,\sqrt3)$이다. 네 개로 정규화한 값은 종 모양이 아니라 평평하다.

연습 8.4 문제 32쪽

정규 커널 $e^{-d^2/2}=c$에서 $d=\sqrt{-2\log c}$: $c=0.1$이면 2.15, $c=0.01$이면 3.03. $t(1)$ 커널 $\frac{1}{1+d^2}=c$에서 $d=\sqrt{1/c-1}$: 3.0과 9.95. 유사도가 작을수록 $t(1)$의 거리가 훨씬 빨리 커진다.

연습 8.5 문제 32쪽

가능도와 사전분포의 곱은 $\sigma^{-1}\cdot\sigma^{-n}\exp\!\big(-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\big)$이고 $\sum(x_i-\mu)^2=(n-1)s^2+n(\mu-\bar x)^2=:A$이다. $u=1/\sigma^2$로 치환하면 $d\sigma=-\frac12u^{-3/2}du$이므로

$$\int_0^\infty\sigma^{-(n+1)}e^{-A/(2\sigma^2)}\,d\sigma=\frac12\int_0^\infty u^{n/2-1}e^{-Au/2}\,du=\frac12\Gamma\!\Big(\frac n2\Big)\Big(\frac2A\Big)^{n/2}\propto A^{-n/2}.$$

따라서 $\pi(\mu \mid \mathbf{x}) \propto \left(1+\frac{n(\mu-\bar{x})^2}{(n-1)s^2}\right)^{-n/2}$이고, $t=\frac{\mu-\bar{x}}{s/\sqrt{n}}$로 쓰면 $\left(1+\frac{t^2}{n-1}\right)^{-n/2}$, 곧 $t(n-1)$의 밀도 꼴이다.

연습 8.6 문제 32쪽

(a) X_{new}는 표본 $(X_1,\dots,X_n)$과 독립이므로 그 함수인 $(\bar{X},S^2)$와 독립이다(3회 3.2절의 상자). 정리 2.2(b)로 $\bar{X}\perp S^2$이다. 따라서 $P(X_{\text{new}}\in A,\bar{X}\in B,S^2\in C)=P(X_{\text{new}}\in A)P(\bar{X}\in B)P(S^2\in C)$이고, 셋은 상호 독립이다. (b) $X_{\text{new}}\sim\mathcal{N}(\mu,\sigma^2)$과 $\bar{X}\sim\mathcal{N}(\mu,\sigma^2/n)$이 독립이므로 5회 정리 1.8(iv)로 $X_{\text{new}}-\bar{X}\sim\mathcal{N}\big(0,\sigma^2(1+\frac{1}{n})\big)$이다. 이것은 $(X_{\text{new}},\bar{X})$의 함수이고 (a)로 S^2와 독립이다. (c) $Z=\frac{X_{\text{new}}-\bar{X}}{\sigma\sqrt{1+1/n}}\sim\mathcal{N}(0,1)$, $V=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$, $Z\perp V$이고 $T=Z/\sqrt{V/(n-1)}$이다. σ가 소거되므로 $T\sim t(n-1)$이다. (d) $c=t_{n-1,\alpha/2}$이면 $P(|T|\le c)=1-\alpha$이고, $|T|\le c$는 $\bar{X}-cS\sqrt{1+1/n}\le X_{\text{new}}\le\bar{X}+cS\sqrt{1+1/n}$과 같다. 신뢰구간에서는 μ가 고정되고 구간만 무작위였지만, 여기서는 구간의 양 끝과 X_{new}가 모두 무작위다.

연습 8.7 문제 33쪽

(a) $\sum_j d_j=\sum_j x_j-B\bar{x}_B=0$이므로 $d_i=-\sum_{j\neq i}d_j$이다. (b) Cauchy–Schwarz로 $\left(\sum_{j\neq i}1\cdot d_j\right)^2\le(B-1)\sum_{j\neq i}d_j^2$이다. (c) $\sum_j d_j^2=B\hat{\sigma}_B^2$이므로 $d_i^2\le(B-1)(B\hat{\sigma}_B^2-d_i^2)$, 곧 $Bd_i^2\le B(B-1)\hat{\sigma}_B^2$이다. 양변을 $B\hat{\sigma}_B^2$로 나누면 $\hat{x}_i^2\le B-1$이다. (d) Cauchy–Schwarz의 등호는 $(d_j)_{j\neq i}$가 $(1,\dots,1)$에 비례할 때, 곧 x_i를 뺀 나머지 값이 모두 같을 때 성립한다. $B=2$이면 나머지가 하나뿐이므로 항상 등호다. $d_1=-d_2$이고 $\hat{\sigma}_B^2=\frac{d_1^2+d_2^2}{2}=d_1^2$이므로 $\hat{x}_1=d_1/|d_1|=\pm1$, $\hat{x}_2=-\hat{x}_1$이다.