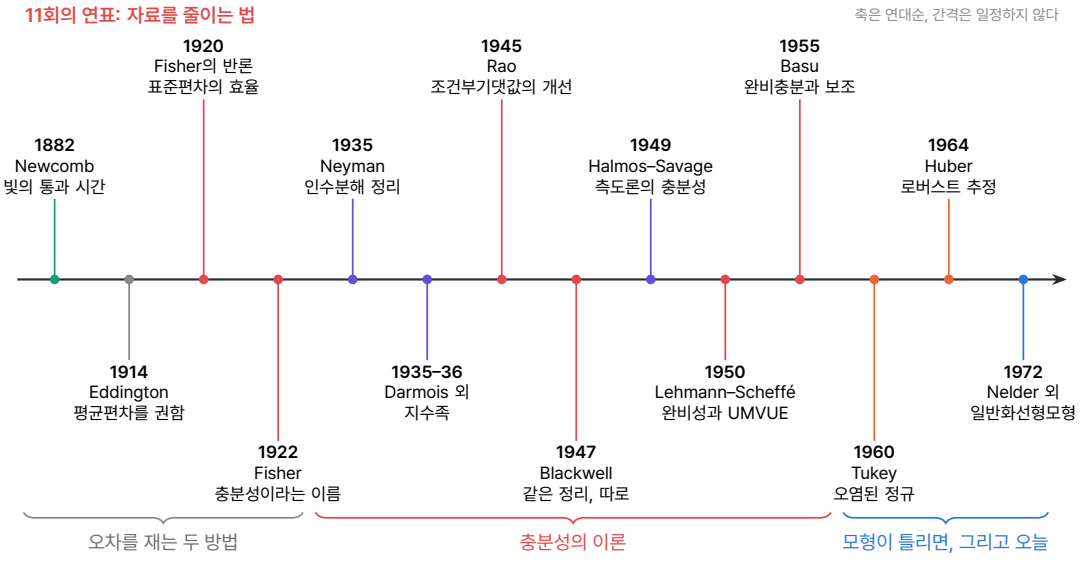
수리통계학 × 통계학의 역사

11회 강의노트

충분성
자료를 줄이는 법

1920 → 1955

1920년 젊은 Fisher는 천문학자 Eddington과 오차의 크기를 재는 두 방법을 놓고 다퉜다. 그 계산에서 Fisher는 표준편차가 자료 속 정보를 하나도 버리지 않는다는 성질을 발견했고, 그것을 충분성이라 불렀다. 이 노트는 인수분해 정리, Rao–Blackwell, Lehmann–Scheffé, 지수족, Basu의 정리를 차례로 증명하고, 가장 좋은 불편추정량이 왜 가장 좋은 추정량은 아닌지 본다.



다루는 것 충분통계량 · 인수분해 정리 · 최소충분 · Rao–Blackwell · 완비성 · Lehmann–Scheffé와 UMVUE · 지수족과 로그분배함수 · Cramér–Rao의 등호 · 보조통계량과 Basu의 정리 · 오염 정규

원자료 Newcomb(1882)의 빛의 통과 시간 66번 · von Bortkiewicz의 프로이센 기병대 · 독일 전차 문제(시뮬레이션)

노트북 `sessions/11_sufficiency/11_sufficiency.ipynb` — 절·정리 번호가 같다

워크북 `11_sufficiency_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`

선수 지식 1–10회 (특히 3회 반복기댓값·전분산 공식, 6회 Student 정리, 8회 독일 전차, 9회 Fisher 정보·Cramér–Rao)

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. 극한과 적분의 교환이나 적분가능성이 필요한 곳은 **조건 확인** 단락에 따로 밝힌다. 나머지 정리는 **증명 전략**만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 이 연습은 건너뛰지 않고 손으로 쓴다.

근거에 붙은 [T1]–[T8]은 별책 **증명 도구 상자**의 도구 번호다. 이 회차에서는 [T1] 원소 추적, [T4] 볼록성, [T6] 극한과 적분의 교환, [T7] 변수변환, [T8] 델타 방법의 o 표기를 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 1–10회에서 가져오는 것 2

0.2 표기 2

1 충분통계량 3

1.1 정의와 인수분해 정리 3

1.2 최소충분통계량 5

1.3 충분통계량은 정보를 보존한다 6

2 Rao–Blackwell — 조건을 걸면 좋아진다 8

2.1 정리 8

2.2 예 9

3 완비성과 Lehmann–Scheffé — 가장 좋은 불편추정량 11

3.1 완비성 11

3.2 Lehmann–Scheffé 정리 12

4 지수족 14

4.1 정의와 로그분배함수 14

4.2 완비충분통계량 16

4.3 Cramér–Rao의 등호는 지수족에서만 17

5 보조통계량과 Basu의 정리 19

6 데이터로 — Fisher의 계산과 Newcomb의 빛 22

6.1 1920년의 계산 22

6.2 Newcomb의 빛의 통과 시간 (1882) 23

6.3 기병대와 독일 전차 24

7 부숴보기 — 큰 오차와 터무니없는 UMVUE 26

7.1 조금 섞인 큰 오차 (Tukey 1960) . . . 26

7.2 터무니없는 UMVUE . . . 26

7.3 불편추정량이 없을 때 . . . 27

8 Beautiful insights — 분산 줄이기와 softmax 29

8.1 Rao–Blackwell화는 분산 줄이기다 . . . 29

8.2 softmax는 로그분배함수의 기울기다 . . . 29

8.3 충분통계량은 합쳐진다 . . . 29

9 정리와 역사 노트 31

9.1 한 장 요약 . . . 31

9.2 Stigler의 명명 법칙 . . . 31

9.3 7 Pillars 지도 . . . 32

9.4 다음 회 — 최적 검정과 결정 (1928 → 1956) . . . 32

A 연습문제 풀이 34

0

들어가며

> **여는 질문.** 관측의 흩어짐을 재는 방법은 둘이다. 평균에서 벗어난 거리의 평균(평균편차)과, 제곱의 평균의 제곱근(표준편차)이다. 천문학자 Eddington은 평균편차가 낫다고 했고, 젊은 Fisher는 표준편차가 낫다고 했다. 누가 옳았나?

1914년 Arthur Eddington은 별의 운동에 관한 책에서 관측 오차의 크기를 "평균편차"로 재라고 권했다. 계산이 쉽고 큰 오차에 덜 휘둘린다는 이유였다. Rothamsted 농업시험장에 막 자리를 잡은 Ronald Fisher는 1920년 왕립천문학회 월보에 반론을 실었다. 오차가 정규분포를 따르면 같은 정밀도를 얻는 데 평균편차는 표준편차보다 관측이 14% 더 필요하다($1/0.876$). 그런데 Fisher는 그 계산에서 더 놀라운 것을 보았다. 표준편차를 알고 나면, 다른 어떤 통계량도 σ에 대해 더 알려 주는 것이 없었다. 표준편차의 값이 주어지면 자료의 나머지 모양은 σ와 무관하게 분포했다. 1922년 그는 이 성질을 **충분성**이라 불렀다.

이 노트는 그 성질을 정리한다. 충분통계량을 정의하고, 그것을 알아보는 인수분해 정리를 증명한다(1절). 3회의 전분산 공식 한 줄로 "충분통계량에 조건을 걸면 불편추정량이 좋아진다"는 Rao–Blackwell 정리를 얻는다(2절). 완비성을 더하면 가장 좋은 불편추정량(UMVUE)을 찾는 방법이 되고, 8회의 독일 전차 추정량이 왜 최선인지 증명된다(3절). 4절은 이 모든 것이 깨끗하게 성립하는 모형인 지수족을 다루고, 9회의 Cramér–Rao 한계에 닿는 모형이 지수족뿐임을 보인다. 5절은 Basu의 정리로 6회의 $\bar{X} \perp S^2$를 다시 증명한다. 6절은 Newcomb의 빛의 속도 자료와 기병대에 쓰고, 7절에서는 Eddington이 이기는 곳(오염된 정규)과 터무니없는 UMVUE를 본다. 8절은 같은 생각이 Monte Carlo의 분산 줄이기와 신경망의 softmax에 있음을 본다.

학습 목표

1. 충분통계량을 정의하고 인수분해 정리로 찾으며, 충분통계량이 Fisher 정보를 보존함을 보인다.
2. Rao–Blackwell 정리를 전분산 공식으로 증명하고, $e^{-\lambda}$와 독일 전차에서 분산이 줄어드는 것을 계산한다.
3. 완비성과 Lehmann–Scheffé 정리로 UMVUE를 구하고, UMVUE의 한계(평균제곱오차, 터무니없는 예)를 안다.

4. 지수족의 로그분배함수로 평균·분산·정보를 구하고, Cramér–Rao의 등호가 지수족에서만 성립함을 증명한다.
5. Basu의 정리를 증명하고 정규 표본에서 $\bar{X} \perp S^2$를 다시 얻는다.

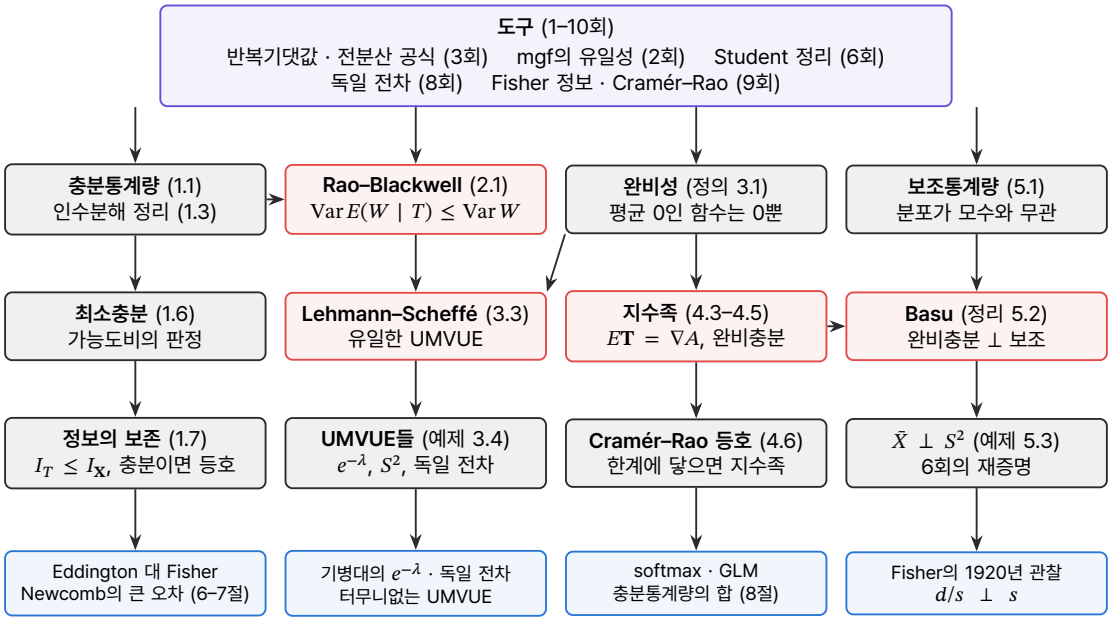


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–10회에서 가져오는 것

앞 회차	내용
3회 정리 2.5 · 2.6	반복기댓값 $E[E(Y \mid X)] = EY$, 전분산 공식 $\operatorname{Var} Y = \operatorname{Var} E(Y \mid X) + E \operatorname{Var}(Y \mid X)$
4회 정리 2.4 · 3.7	Poisson의 합은 Poisson, 감마의 mgf
6회 정리 2.2	Student 정리: 정규 표본에서 $\bar{X} \perp S^2$, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
8회 정리 1.6	독일 전차: $\hat{N} = \frac{k+1}{k} M - 1$은 불편
9회 정리 2.2 · 2.4	Fisher 정보의 두 표현, Cramér–Rao 부등식과 그 등호 조건

0.2 표기

$\mathbf{X} = (X_1, \ldots, X_n)$은 밀도(또는 확률질량함수) $f(\mathbf{x}; \theta)$를 갖는 표본이고, $\theta \in \Omega$는 실수이거나 벡터다. 통계량 $T = T(\mathbf{X})$는 θ를 포함하지 않는 자료의 함수다. "$\tau(\theta)$의 불편추정량"은 모든 θ에서 $E_\theta W = \tau(\theta)$인 통계량 W다. 조건부기댓값 $E(W \mid T)$는 3회 2절의 뜻이다.

1

충분통계량

> **역사** Fisher(1922)는 **통계학의 수학적 기초**에서 통계량이 "자료가 모수에 대해 주는 정보를 모두 담으면" 충분하다고 정의했다. 그가 쓴 판정법은 "통계량을 알면 자료의 나머지 조건부분포가 모수와 무관하다"였다. 이것을 가능도의 모양으로 바꾼 인수분해 정리는 Fisher가 예에서 쓰던 것을 Neyman(1935)이 일반적으로 증명했고, Halmos와 Savage(1949)가 측도론의 언어로 완성했다.

1.1 정의와 인수분해 정리

정의 1.1 (충분통계량) 통계량 $T = T(\mathbf{X})$가 주어질 때 $\mathbf{X}$의 조건부분포가 θ에 의존하지 않으면 T를 θ의 **충분통계량**이라 한다.

정의는 자료를 두 단계로 만들어 보면 읽힌다. 자연은 먼저 θ에 따라 T의 값을 고른다. 그 다음 $\mathbf{X}$의 나머지 모양은 θ와 무관한 확률로, 마치 주사위를 굴리듯 정해진다. 두 번째 단계는 θ에 대해 아무것도 알려 주지 않으므로, T만 보관하고 나머지를 버려도 잃는 것이 없다(그림 1.1).

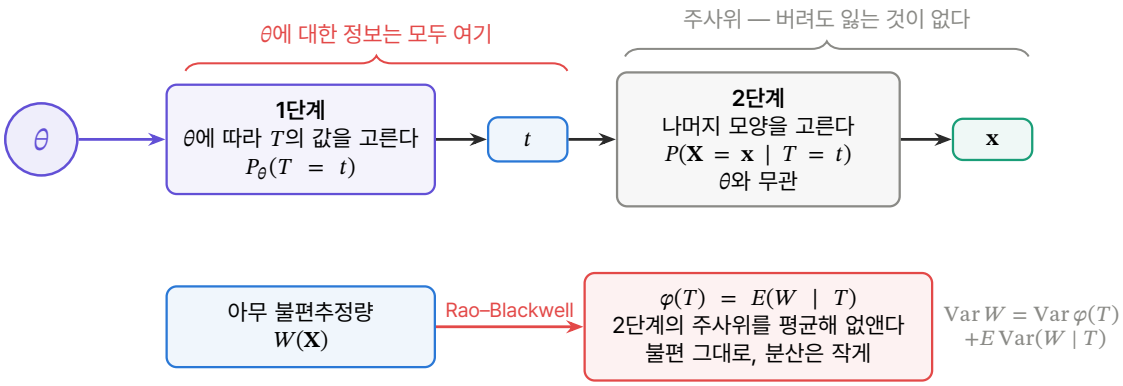


그림 1.1 충분성의 그림. 자료를 두 단계로 만든다. 1단계에서 θ가 T의 값을 정하고, 2단계는 θ와 무관한 주사위로 나머지 모양을 정한다. 아래: Rao–Blackwell(2절)은 2단계의 주사위를 평균해 없애 분산을 줄인다.

예제 1.2 (조건부분포가 모수를 잊는다) (a) $X_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, $T = \sum X_i$이면 $P(\mathbf{X} = \mathbf{x} \mid T = t) = \frac{p^t(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \binom{n}{t}^{-1}$. 성공 t번의 자리는 모든 배치가 같은 확률이다. (b) $X_i \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$, $T = \sum X_i$이면 $\mathbf{X} \mid T = t$는 다항분포 $\text{M}(t; \frac{1}{n}, \dots, \frac{1}{n})$이다(연습 1.1). 사건 t개가 n칸에 고르게 무작위로 떨어진다. 어느 경우든 p나 λ가 사라졌다.

정리 1.3 (인수분해 정리 (Neyman 1935)) $T(\mathbf{X})$가 θ의 충분통계량일 필요충분조건은 음이 아닌 함수 g와 h가 있어 모든 $\mathbf{x}$와 θ에 대해

$$f(\mathbf{x};\theta) = g\big(T(\mathbf{x});\theta\big)\,h(\mathbf{x})$$

인 것이다. h는 θ에 의존하지 않는다.

증명. **전략** 이산인 경우를 증명한다. ($\Rightarrow$)는 결합확률을 "T의 확률 $\times$ T가 주어진 조건부확률"로 쪼개고, 두 인수를 g와 h로 삼는다. ($\Leftarrow$)는 조건부확률을 직접 계산해 g가 분자와 분모에서 약분됨을 보인다.

($\Rightarrow$)

① $\mathbf{x}$를 고정하고 $t = T(\mathbf{x})$라 하자. $\mathbf{X} = \mathbf{x}$이면 $T(\mathbf{X}) = t$이므로 $\{\mathbf{X} = \mathbf{x}\} \subset \{T = t\}$이다.

T는 $\mathbf{X}$의 함수 [T1]

② $P_\theta(T = t) > 0$이면 다음이 성립한다.

$$f(\mathbf{x};\theta) = P_\theta(\mathbf{X} = \mathbf{x},\ T = t) = P_\theta(T = t)\,P_\theta(\mathbf{X} = \mathbf{x} \mid T = t). \qquad \text{①, 1회 정리 5.3}$$

③ 둘째 인수는 θ에 의존하지 않는다. 그것을 $h(\mathbf{x}) = P(\mathbf{X} = \mathbf{x} \mid T = T(\mathbf{x}))$로 쓴다.

충분성, 정의 1.1

④ $g(t;\theta) = P_\theta(T = t)$로 두면 $f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)\,h(\mathbf{x})$이다. ②③

($\Leftarrow$)

① $T(\mathbf{x}) = t$이고 $P_\theta(T = t) > 0$이라 하자. $\{T = t\}$는 서로소인 사건 $\{\mathbf{X} = \mathbf{y}\}$, $T(\mathbf{y}) = t$의 합집합이다.

$$P_\theta(T = t) = \sum_{\mathbf{y}:\,T(\mathbf{y})=t} f(\mathbf{y};\theta) = g(t;\theta) \sum_{\mathbf{y}:\,T(\mathbf{y})=t} h(\mathbf{y}). \qquad \text{가산가법성, 인수분해 가정}$$

② 조건부확률을 계산한다.

$$P_\theta(\mathbf{X} = \mathbf{x} \mid T = t) = \frac{f(\mathbf{x};\theta)}{P_\theta(T = t)} = \frac{g(t;\theta)h(\mathbf{x})}{g(t;\theta)\sum_{\mathbf{y}:\,T(\mathbf{y})=t} h(\mathbf{y})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{y}:\,T(\mathbf{y})=t} h(\mathbf{y})}.$$

1회 정의 5.1, ①

③ 오른쪽에는 θ가 없다. $T(\mathbf{x}) \neq t$이면 조건부확률은 0이다. 따라서 $T = t$가 주어진 $\mathbf{X}$의 조건부분포는 θ에 의존하지 않는다. 정의 1.1

조건 확인 ($\Leftarrow$)의 약분에는 $g(t;\theta) > 0$이 필요하다. $P_\theta(T = t) > 0$이면 ①로 그렇다. ($\Rightarrow$)에서 $P_\theta(T = t) = 0$이면 $f(\mathbf{x};\theta) = 0$이므로 $g(t;\theta) = 0$으로 두면 된다. h는 $P_\theta(T = t) > 0$인 어느 θ로 계산해도 같은 값이다(충분성). ■

> **증명 없이 쓰는 사실** (연속인 경우) 밀도이면 $\{T = t\}$의 확률이 0이므로 조건부분포를 정의하는 데 측도론이 필요하다. 그 언어로 같은 정리가 성립한다(Halmos–Savage 1949). 이 과정에서는 연속 모형에서도 인수분해를 충분성의 판정법으로 쓴다. 변수변환으로 직접 확인할 수 있는 경우가 많다(정규 표본은 6회의 Helmert 변환, 균등 표본은 8회의 순서통계량 밀도).

> **예제 1.4** (충분통계량 찾기)
>
> (a) **Poisson.** $f = \prod \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum x_i} \cdot \frac{1}{\prod x_i!}$이므로 $T = \sum X_i$.
>
> (b) **정규.** $f = (2\pi\sigma^2)^{-n/2}\exp\!\left(-\frac{\sum x_i^2 - 2\mu\sum x_i + n\mu^2}{2\sigma^2}\right)$이므로 $(\sum X_i, \sum X_i^2)$, 같은 말로 $(\bar{X}, S^2)$가 (μ,σ^2)의 충분통계량이다. μ를 알면 $\sum(X_i - \mu)^2$ 하나로 충분하다.
>
> (c) **균등** $\mathrm{U}(0,\theta)$. $f = \theta^{-n}\mathbb{1}\{x_{(n)} \le \theta\}\cdot\mathbb{1}\{x_{(1)} \ge 0\}$이므로 $T = X_{(n)}$. 8회의 독일 전차에서 최댓값이 모든 정보를 담는 이유다.
>
> (d) **감마**(α,β). $f = \Gamma(\alpha)^{-n}\beta^{-n\alpha}\left(\prod x_i\right)^{\alpha-1}e^{-\sum x_i/\beta}$이므로 $(\sum X_i,\ \sum \log X_i)$. 9회 예제 4.4의 MLE가 이 두 수만 쓴 이유다.

1.2 최소충분통계량

자료 전체 $\mathbf{X}$는 언제나 충분하다. 쓸모 있는 것은 **가장 많이** 줄인 것이다.

> **정의 1.5** (최소충분통계량) 충분통계량 T가 다른 모든 충분통계량 U의 함수이면($U(\mathbf{x}) = U(\mathbf{y}) \Rightarrow T(\mathbf{x}) = T(\mathbf{y})$) T를 **최소충분통계량**이라 한다.

> **정리 1.6** (최소충분의 판정 (Lehmann–Scheffé 1950)) 모든 $\mathbf{x},\mathbf{y}$에 대해 "$f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$가 θ에 의존하지 않는다 $\iff$ $T(\mathbf{x}) = T(\mathbf{y})$"이면 T는 최소충분통계량이다. (받침은 θ에 무관하다고 하고, 이산인 경우를 증명한다.)

증명. **전략** 충분성은 T의 값마다 대표 표본을 하나 골라, 가능도를 "대표의 가능도 × 비"로 쓰고 인수분해 정리를 쓴다. 최소성은 다른 충분통계량 U의 인수분해로 가능도비가 θ에 무관함을 보이고 가정의 ($\Rightarrow$) 방향을 쓴다.

충분성

① T의 값 t마다 받침 안에서 $T(\mathbf{x}_t) = t$인 대표 $\mathbf{x}_t$를 하나 고른다. t는 T의 치역의 원소

② $T(\mathbf{x}) = T(\mathbf{x}_{T(\mathbf{x})})$이므로 $h(\mathbf{x}) = f(\mathbf{x};\theta)/f(\mathbf{x}_{T(\mathbf{x})};\theta)$는 θ에 의존하지 않는다.

가정의 ($\Leftarrow$) 방향

③ $f(\mathbf{x};\theta) = f(\mathbf{x}_{T(\mathbf{x})};\theta)\,h(\mathbf{x})$이고, 앞 인수는 $T(\mathbf{x})$와 θ만의 함수다. ②

④ 따라서 T는 충분통계량이다. 정리 1.3

최소성

① U가 충분통계량이고 $U(\mathbf{x}) = U(\mathbf{y})$라 하자. $f(\mathbf{x};\theta) = g(U(\mathbf{x});\theta)h(\mathbf{x})$로 쓸 수 있다.

정리 1.3

② 가능도비에서 g가 약분된다.

$$\frac{f(\mathbf{x};\theta)}{f(\mathbf{y};\theta)} = \frac{g(U(\mathbf{x});\theta)h(\mathbf{x})}{g(U(\mathbf{y});\theta)h(\mathbf{y})} = \frac{h(\mathbf{x})}{h(\mathbf{y})}. \qquad ①,\ U(\mathbf{x}) = U(\mathbf{y})$$

③ 오른쪽은 θ에 의존하지 않으므로 $T(\mathbf{x}) = T(\mathbf{y})$이다. 곧 T는 U의 함수다.

가정의 ($\Rightarrow$) 방향, 정의 1.5

조건 확인 가능도비는 분모가 0이 아닐 때 뜻이 있다. 받침이 θ에 무관하다고 가정했으므로 받침 안의 $\mathbf{x}, \mathbf{y}$에서는 모든 θ에 대해 분모가 양수다. 그래서 ①의 대표도 받침에서 고른다. ■

정규 표본에서 $\frac{f(\mathbf{x};\mu,\sigma^2)}{f(\mathbf{y};\mu,\sigma^2)} = \exp\!\left(-\frac{\sum x_i^2 - \sum y_i^2}{2\sigma^2} + \frac{\mu(\sum x_i - \sum y_i)}{\sigma^2}\right)$이 모수에 무관할 필요충분조건은 $\sum x_i = \sum y_i$이고 $\sum x_i^2 = \sum y_i^2$인 것이다. 그래서 $(\bar{X}, S^2)$는 최소충분이다. 반대로 Cauchy 위치 모형에서는 순서통계량 전체가 최소충분이다(연습 1.4). 자료를 순서 말고는 하나도 줄일 수 없다. 9회에서 Cauchy의 MLE를 Newton으로 구해야 했던 이유다.

1.3 충분통계량은 정보를 보존한다

> **명제 1.7** (통계량의 정보) 9회의 정칙조건 아래, 통계량 T의 분포로 만든 Fisher 정보 $I_T(\theta)$는 자료 전체의 정보 $I_{\mathbf{X}}(\theta) = nI(\theta)$보다 크지 않다. T가 충분하면 등호가 성립한다.

증명. **전략** T의 점수가 자료 전체의 점수의 조건부기댓값임을 보이고, 전분산 공식(3회 정리 2.6)으로 분산을 비교한다. 충분하면 자료의 점수가 이미 T의 함수다. 단계별 증명은 연습 1.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 35쪽에 있다.

Fisher에게 충분성은 처음부터 이 명제였다. 충분통계량은 정보를 하나도 잃지 않고, 다른 통계량은 정보를 잃는다. 1920년의 계산도 이 틀 안에 있다. 평균을 알 때 $\sum X_i^2$는 σ의 충분통계량이고,

평균편차 $\frac{1}{n}\sum|X_i-\mu|$는 아니다. 큰 표본에서 평균편차가 담는 정보는 전체의 $\frac{1}{\pi-2}=0.876$이다(6절).

연습문제

연습 1.1 (Poisson의 조건부분포 ★) $X_i \overset{\text{iid}}{\sim} \mathrm{Poi}(\lambda)$이면 $\mathbf{X}\mid\sum X_i=t \sim \mathrm{M}(t;\frac{1}{n},\dots,\frac{1}{n})$임을 보여라(4회 정리 2.4). 풀이 34쪽

연습 1.2 (인수분해 ★) 인수분해 정리로 충분통계량을 구하라. (a) 베타(a,b), 두 모수 모두 미지. (b) 기하분포 $P(X=x)=p(1-p)^{x-1}$. (c) 지수분포를 c에서 절단한 자료(10회 연습 2.3). 풀이 34쪽

연습 1.3 (이동한 균등분포 ★★) $X_i \overset{\text{iid}}{\sim} \mathrm{U}(\theta,\theta+1)$이면 $(X_{(1)},X_{(n)})$이 최소충분임을 보여라. 모수는 하나인데 최소충분통계량은 2차원이다. 풀이 34쪽

연습 1.4 (Cauchy의 순서통계량 ★★) Cauchy 위치모형에서 $\frac{f(\mathbf{x};\theta)}{f(\mathbf{y};\theta)}=\prod\frac{1+(y_i-\theta)^2}{1+(x_i-\theta)^2}$가 θ에 무관하면 $\mathbf{x}$와 $\mathbf{y}$가 순서만 다름을 보여라. (두 다항식의 근을 비교하라.) 풀이 34쪽

연습 1.5 (부호가 담는 정보 ★★) $X\sim\mathcal{N}(\theta,1)$에서 $T=\mathbb{1}\{X>0\}$의 정보가 $I_T(\theta)=\frac{\varphi(\theta)^2}{\Phi(\theta)(1-\Phi(\theta))}$이고, $\theta=0$에서 $\frac{2}{\pi}=0.637$, 곧 전체 정보의 64%임을 보여라. 풀이 34쪽

연습 1.6 (통계량의 정보: 증명 ★★ 유도형) 이산인 경우에 명제 1.7를 다음 순서로 증명하라. $s_{\mathbf{X}}(\mathbf{x};\theta)=\partial_\theta\log f(\mathbf{x};\theta)$는 자료 전체의 점수다.

(a) $P_\theta(T=t)=\sum_{\mathbf{x}:\,T(\mathbf{x})=t}f(\mathbf{x};\theta)$를 θ로 미분해 T의 점수 $s_T(t)=\partial_\theta\log P_\theta(T=t)$가 $E_\theta\big[s_{\mathbf{X}}(\mathbf{X})\mid T=t\big]$임을 보여라. 합과 미분의 순서를 바꾸는 근거는 무엇인가?

(b) $E_\theta s_T(T)=0$이고 $I_T=\mathrm{Var}_\theta\, s_T(T)$임을 확인하고, 전분산 공식으로 $I_T\le I_{\mathbf{X}}$를 얻어라.

(c) T가 충분하면 인수분해 정리로 $s_{\mathbf{X}}=\partial_\theta\log g(T(\mathbf{x});\theta)$가 T의 함수임을 보이고, 등호를 얻어라.

풀이 35쪽

2

Rao–Blackwell — 조건을 걸면 좋아진다

> **역사** C. R. Rao는 1945년 스물다섯 살에 쓴 논문에서 9회의 Cramér–Rao 부등식과 함께 이 정리를 증명했다. David Blackwell은 1947년 따로 같은 결과를 얻었다. Kolmogorov(1950)도 같은 생각에 이르렀다. 증명은 3회의 전분산 공식 한 줄이다.

2.1 정리

정리 2.1 (Rao–Blackwell) W가 $\tau(\theta)$의 불편추정량이고 $\operatorname{Var}_\theta W < \infty$, T가 충분통계량이면 $\varphi(T) = E(W \mid T)$는

(a) θ를 포함하지 않는 통계량이고,

(b) $\tau(\theta)$의 불편추정량이며,

(c) 모든 θ에서 $\operatorname{Var}_\theta \varphi(T) \le \operatorname{Var}_\theta W$이다. 등호는 $W = \varphi(T)$ (확률 1)일 때뿐이다.

증명. **전략** (a)는 충분성의 정의 그 자체다. (b)는 반복기댓값, (c)는 전분산 공식이다. 충분성은 (a)에서만 쓰인다.

(a)

① $W = w(\mathbf{X})$이고, $T = t$가 주어질 때 $\mathbf{X}$의 조건부분포는 θ에 의존하지 않는다. — 정의 1.1

② $\varphi(t) = E(w(\mathbf{X}) \mid T = t)$는 그 조건부분포의 평균이므로 θ 없이 계산된다. — ①, 3회 2절

(b)

① $E_\theta \varphi(T) = E_\theta\big[E(W \mid T)\big] = E_\theta W = \tau(\theta)$이다. — 3회 정리 2.5, W는 불편

(c)

① 전분산 공식을 W와 T에 쓴다.

$$\operatorname{Var}_\theta W = \operatorname{Var}_\theta E(W \mid T) + E_\theta \operatorname{Var}(W \mid T) = \operatorname{Var}_\theta \varphi(T) + E_\theta \operatorname{Var}(W \mid T). \quad \text{3회 정리 2.6}$$

② $\operatorname{Var}(W \mid T) \ge 0$이므로 $\operatorname{Var}_\theta \varphi(T) \le \operatorname{Var}_\theta W$이다. — 조건부분산은 음이 아니다

③ 등호이면 $E_\theta \operatorname{Var}(W \mid T) = E_\theta\big(W - \varphi(T)\big)^2 = 0$이다. ①, 반복기댓값

④ 따라서 $W = \varphi(T)$ (확률 1)이다. 거꾸로 $W = \varphi(T)$이면 등호는 자명하다. 2회 따름정리 5.6

조건 확인 전분산 공식은 $E_\theta W^2 < \infty$일 때 성립하고, 가정 $\operatorname{Var}_\theta W < \infty$가 그것이다. 이 가정은 모든 θ에서 필요하다. 충분하지 않은 T를 쓰면 ①에서 φ가 θ를 포함할 수 있다. (b)와 (c)는 그대로 성립한다. ∎

충분하지 않은 통계량에 조건을 걸면 (b)와 (c)는 여전히 성립하지만 $E(W \mid T)$에 θ가 남아 추정량이 아니다. 정리는 "어떤 불편추정량이든 충분통계량의 함수로 바꾸면 나빠지지 않는다" 고 말한다. 그래서 좋은 추정량은 충분통계량 안에서만 찾으면 된다.

2.2 예

예제 2.2 (한 해에 아무도 죽지 않을 확률) $X_i \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$에서 $\tau = P(X=0) = e^{-\lambda}$를 추정한다. 가장 단순한 불편추정량은 $W = 1\{X_1 = 0\}$이다. 관측 하나만 보고 0이면 1, 아니면 0이다. $T = \sum X_i$가 주어지면 $X_1 \mid T = t \sim \text{b}(t, \tfrac{1}{n})$이므로(연습 1.1)

$$\varphi(T) = P(X_1 = 0 \mid T) = \Big(1 - \frac{1}{n}\Big)^T.$$

$n = 20$, $\lambda = 0.7$이면 $\operatorname{Var} W = e^{-\lambda}(1 - e^{-\lambda}) = 0.250$이 $\operatorname{Var}\varphi(T) = e^{-2\lambda}(e^{\lambda/n} - 1) = 0.0088$로 28분의 1이 된다(연습 2.1). 기병대(280군단–해, 196명)에서는 $(279/280)^{196} = 0.4960$이다. MLE $e^{-\bar{x}} = 0.4966$과 관측 비율 $144/280 = 0.514$와 비교하라.

예제 2.3 (독일 전차의 Rao–Blackwell화) 8회에서 $\{1, \dots, N\}$에서 비복원으로 뽑은 k개의 일련번호로 N을 추정했다. 적률법 $W = 2\bar{X} - 1$은 불편이다. 충분통계량은 최댓값 M이다(예제 1.4(c)의 이산판). $M = m$이 주어지면 나머지 $k-1$개는 $\{1, \dots, m-1\}$에서 비복원으로 뽑은 것이므로 그 평균의 기댓값은 $m/2$이고

$$E(2\bar{X} - 1 \mid M = m) = \frac{2}{k}\Big(m + (k-1)\frac{m}{2}\Big) - 1 = \frac{k+1}{k}m - 1.$$

적률법에 조건을 걸었더니 8회 정리 1.6의 추정량 $\hat{N}$이 나왔다. $N = 300$, $k = 5$에서 표준편차는 77에서 50으로 준다. 3절에서 $\hat{N}$이 모든 불편추정량 가운데 최선임을 보인다.

예제 2.4 (성공 확률의 제곱) Bernoulli 표본에서 p^2의 불편추정량 $W = X_1X_2$에 $T = \sum X_i$로 조건을 걸면 $E(X_1X_2 \mid T = t) = \frac{t(t-1)}{n(n-1)}$(두 자리가 모두 성공일 확률, 비복원)이다. 같은 방법으로 $p(1-p)$의 불편추정량은 $\frac{T(n-T)}{n(n-1)}$이다.

연습문제

연습 2.1 (분산의 계산 ★★) $T \sim \mathrm{Poi}(n\lambda)$일 때 $E\,s^T = e^{n\lambda(s-1)}$(확률생성함수)을 써서 $\mathrm{Var}\big((1-\frac{1}{n})^T\big) = e^{-2\lambda}(e^{\lambda/n}-1)$임을 보여라. Cramér–Rao 한계 $\frac{\lambda e^{-2\lambda}}{n}$(9회 정리 2.4, $k(\lambda)=e^{-\lambda}$)보다 크지만 $n\to\infty$에서 비가 1로 감을 보여라. 풀이 35쪽

연습 2.2 (전차의 조건부 평균 ★★) 예제 2.3에서 $M = m$이 주어질 때 나머지 $k-1$개가 $\{1,\dots,m-1\}$의 크기 $k-1$인 부분집합 가운데 균등하게 뽑힌다는 것을 보이고, 각 원소의 평균이 $m/2$임을 확인하라. 풀이 35쪽

연습 2.3 (정규의 꼬리확률 ★★) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\theta,1)$에서 $\Phi(c-\theta) = P(X \le c)$의 불편추정량 $\mathbb{1}\{X_1 \le c\}$에 $\bar X$로 조건을 걸면 $\Phi\big(\frac{c-\bar X}{\sqrt{(n-1)/n}}\big)$임을 보여라. ($X_1 - \bar X \sim \mathcal{N}(0, \frac{n-1}{n})$이 $\bar X$와 독립이다.) 풀이 35쪽

연습 2.4 (볼록 손실 ★★) 손실 $L(\theta,a)$가 a에 대해 볼록이면 $E_\theta L(\theta, E(W\mid T)) \le E_\theta L(\theta, W)$임을 조건부 Jensen 부등식으로 보여라. 정리 2.1는 제곱손실의 경우다. 풀이 35쪽

3

완비성과 Lehmann–Scheffé — 가장 좋은 불편추정량

Rao–Blackwell은 "충분통계량의 함수로 바꾸면 나빠지지 않는다"고 말했다. 그런데 충분통계량의 함수인 불편추정량이 여럿이면 어느 것이 최선인가? 하나뿐이면 그것이 답이다. 하나뿐이게 하는 성질이 완비성이다.

3.1 완비성

> **정의 3.1** (완비통계량) 모든 $\theta \in \Omega$에서 $E_\theta g(T) = 0$이면 $P_\theta(g(T) = 0) = 1$ (모든 θ)이 되는 통계량 T를 **완비**라 한다.

"평균이 언제나 0인 T의 함수는 0뿐이다." 같은 말로, T의 함수 가운데 $\tau(\theta)$의 불편추정량은 많아야 하나다. 두 개 $\varphi_1(T), \varphi_2(T)$가 있으면 그 차의 평균이 언제나 0이기 때문이다.

예제 3.2 (완비인 것과 아닌 것)

(a) **이항.** $T \sim \mathrm{b}(n,p)$이고 $\sum_{t=0}^{n} g(t)\binom{n}{t}p^t(1-p)^{n-t} = 0$이 모든 $p \in (0,1)$에서 성립하면, $(1-p)^n$으로 나누고 $r = p/(1-p)$로 쓴 $\sum_t g(t)\binom{n}{t}r^t$는 모든 $r>0$에서 0인 다항식이므로 계수가 모두 0, 곧 $g \equiv 0$이다.

(b) **Poisson.** $\sum_t g(t)\frac{e^{-\mu}\mu^t}{t!} = 0$ $(\mu>0)$이면 멱급수 $\sum g(t)\mu^t/t!$이 0이므로 계수가 모두 0이다(연습 3.1).

(c) **균등의 최댓값.** $X_{(n)}$의 밀도는 nt^{n-1}/θ^n $(0<t<\theta)$이다. $\int_0^\theta g(t)t^{n-1}dt = 0$이 모든 θ에서 성립하면 θ로 미분해 $g(\theta)\theta^{n-1}=0$, 곧 $g=0$이다.

(d) **완비가 아닌 것.** $\mathrm{U}(\theta,\theta+1)$에서 범위 $R = X_{(n)} - X_{(1)}$의 분포는 θ에 무관하므로 $g(T) = R - ER$은 평균이 언제나 0이지만 0이 아니다. 최소충분통계량 $(X_{(1)}, X_{(n)})$은 완비가 아니다.

3.2 Lehmann–Scheffé 정리

정리 3.3 (Lehmann–Scheffé (1950)) T가 완비충분통계량이고 $\varphi(T)$가 $\tau(\theta)$의 불편추정량이면, $\varphi(T)$는 $\tau(\theta)$의 **유일한 균등최소분산 불편추정량**(UMVUE)이다. 곧 모든 불편추정량 W와 모든 θ에서 $\mathrm{Var}_\theta\,\varphi(T) \le \mathrm{Var}_\theta\, W$이다.

증명. **전략** 아무 불편추정량 W를 Rao–Blackwell로 T의 함수로 바꾼다. 완비성으로 그 함수는 $\varphi(T)$와 같다. 유일성은 Rao–Blackwell의 등호 조건에서 나온다.

① W가 $\tau(\theta)$의 불편추정량이고 분산이 유한하다고 하자. $\psi(T) = E(W \mid T)$는 T의 함수인 불편추정량이고 $\mathrm{Var}_\theta\,\psi(T) \le \mathrm{Var}_\theta\, W$이다. — 정리 2.1

② $g(T) = \psi(T) - \varphi(T)$이면 모든 θ에서 $E_\theta g(T) = \tau(\theta) - \tau(\theta) = 0$이다. — 둘 다 불편

③ 따라서 모든 θ에서 $\psi(T) = \varphi(T)$ (확률 1)이다. — T의 완비성, 정의 3.1

④ $\mathrm{Var}_\theta\,\varphi(T) = \mathrm{Var}_\theta\,\psi(T) \le \mathrm{Var}_\theta\, W$이다. — ①③

⑤ 유일성. W'도 UMVUE이면 ①–③을 W'에 써서 $E(W' \mid T) = \varphi(T)$이고, UMVUE의 정의와 ④로 $\mathrm{Var}_\theta\, W' = \mathrm{Var}_\theta\,\varphi(T)$이다. — ①–④

⑥ 정리 2.1(c)의 등호 조건으로 $W' = E(W' \mid T) = \varphi(T)$ (확률 1)이다. — 정리 2.1(c)

조건 확인 분산이 무한인 불편추정량에서는 ④의 부등식이 자명하다. 분산이 유한한 불편추정량이 하나라도 있으면 ④로 $\mathrm{Var}_\theta\,\varphi(T)$도 유한하다. 완비성은 모든 θ에서 $E_\theta g(T)=0$일 때 쓰인다. 모수공간의 일부에서만 불편이면 쓸 수 없다. ∎

정리는 UMVUE를 찾는 두 길을 준다. (1) 완비충분통계량의 함수 가운데 불편인 것을 직접 찾는다. (2) 아무 불편추정량에 완비충분통계량으로 조건을 건다(Rao–Blackwell).

예제 3.4 (UMVUE들)

(a) Poisson의 $e^{-\lambda}$: $(1-\frac{1}{n})^T$ (예제 2.2). $T=\sum X_i \sim \mathrm{Poi}(n\lambda)$는 완비충분이다.

(b) 정규의 σ^2: S^2. $(\bar X, S^2)$가 완비충분이고(4절) $ES^2=\sigma^2$이다(6회 정리 1.1). 9회 예제 2.6에서 S^2은 Cramér–Rao 한계에 닿지 않았다(효율 $\frac{n-1}{n}$). 그래도 불편추정량 가운데에서는 최선이다. 한계 자체가 이룰 수 없는 값이었다.

(c) 균등 $\mathrm{U}(0,\theta)$의 θ: $\frac{n+1}{n}X_{(n)}$ (예제 3.2(c)). 9회 연습 2.4에서 Cramér–Rao "한계"보다 작았던 추정량이다. 그 한계는 정칙조건이 깨져서 무의미했지만, Lehmann–Scheffé는 정칙조건이 필요 없다.

(d) 독일 전차: $\hat N=\frac{k+1}{k}M-1$. M은 완비충분이다(연습 3.3). 8회의 약속이 여기서 지켜진다.

가정 주의 UMVUE는 불편추정량 가운데 최선일 뿐이다. 평균제곱오차로 재면 치우친 추정량이 이길 수 있다. 9회 예제 2.6에서 σ^2의 MLE(n으로 나눔)가 S^2을 이겼다. Poisson의 $e^{-\lambda}$도 그렇다. $n=20$, $\lambda=0.7$에서 MLE $e^{-\bar X}$는 위로 0.009만큼 치우쳤지만 평균제곱오차가 0.00872로 UMVUE의 분산 0.00878보다 조금 작다. λ가 0.83보다 크면 순서가 뒤집힌다. 어느 쪽도 모든 λ에서 이기지는 않는다. 7절에서는 UMVUE가 터무니없는 값을 내는 예를 본다.

연습문제

연습 3.1 (Poisson의 완비성 ★★) $T\sim\mathrm{Poi}(\mu)$, $\mu>0$이 완비임을 멱급수의 유일성으로 보여라. 정의역을 $\mu\in\{1,2\}$ 두 점으로 줄이면 완비가 아님을 보여라. 풀이 36쪽

연습 3.2 (균등의 UMVUE ★★) $\mathrm{U}(0,\theta)$에서 θ^2의 UMVUE를 구하라. ($EX_{(n)}^2=\frac{n}{n+2}\theta^2$) 풀이 36쪽

연습 3.3 (전차의 완비성 ★★★) $\{1,\dots,N\}$에서 비복원으로 k개를 뽑은 최댓값 M의 분포 $P(M=m)=\binom{m-1}{k-1}/\binom{N}{k}$에서 M이 $N\in\{k,k+1,\dots\}$에 대해 완비임을 보여라. ($N=k$부터 차례로 $g(k)=g(k+1)=\cdots=0$을 끌어내라.) 풀이 36쪽

연습 3.4 (Bernoulli의 분산 ★) Bernoulli 표본에서 $p(1-p)$의 UMVUE가 $\frac{T(n-T)}{n(n-1)}=\frac{n}{n-1}\hat p(1-\hat p)$임을 보여라. 정규의 S^2에서 $n-1$로 나누는 것과 같은 보정이다. 풀이 36쪽

4

지수족

> **역사** 1935–36년 Darmois(프랑스), Koopman(미국), Pitman(호주)은 서로 모른 채 같은 질문에 같은 답을 냈다. 표본 크기에 관계없이 몇 개의 수로 줄일 수 있는 충분통계량을 갖는 분포족은 어떤 꼴인가? 답은 밀도의 로그가 모수의 함수와 자료의 함수의 곱의 합으로 쓰이는 족, 곧 **지수족**이었다. 정규, Poisson, 이항, 감마, 베타가 모두 여기에 든다.

4.1 정의와 로그분배함수

정의 4.1 (지수족 (자연모수꼴)) 밀도(또는 확률질량함수)가

$$f(x;\boldsymbol{\eta}) = h(x)\exp\!\big(\boldsymbol{\eta}'\mathbf{T}(x) - A(\boldsymbol{\eta})\big), \qquad \boldsymbol{\eta} \in \mathcal{H} \subset \mathbb{R}^k$$

꼴인 분포족을 k-모수 **지수족**이라 한다. $\boldsymbol{\eta}$는 **자연모수**, $\mathbf{T}(x)$는 **자연 충분통계량**, $A(\boldsymbol{\eta}) = \log \int h(x) e^{\boldsymbol{\eta}'\mathbf{T}(x)} dx$는 **로그분배함수**다. $\mathcal{H}$는 이 적분이 유한한 $\boldsymbol{\eta}$의 집합이다.

예제 4.2 (자연모수로 다시 쓰기)

분포	자연모수 η	$T(x)$	$A(\eta)$
Bernoulli(p)	$\log \frac{p}{1-p}$ (로그 오즈)	x	$\log(1+e^{\eta})$
Poisson(λ)	$\log \lambda$	x	e^{η}
$\mathcal{N}(\mu, \sigma^2)$	$\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$	$(x,\ x^2)$	$-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$
감마(α, β)	$\left(\alpha - 1, -\frac{1}{\beta}\right)$	$(\log x,\ x)$	$\log\Gamma(\eta_1+1) - (\eta_1+1)\log(-\eta_2)$
범주 K개	$(\eta_1, \dots, \eta_K)$ (logit)	원-핫 벡터	$\log \sum_k e^{\eta_k}$

마지막 줄의 ∇A는 softmax다(8절).

정리 4.3 (로그분배함수는 적률을 만든다) $\boldsymbol{\eta}$가 $\mathcal{H}$의 내점이면

$$E_{\boldsymbol{\eta}}\mathbf{T}(X) = \nabla A(\boldsymbol{\eta}), \qquad \mathrm{Cov}_{\boldsymbol{\eta}}\,\mathbf{T}(X) = \nabla^2 A(\boldsymbol{\eta}),$$

이고 $\mathbf{T}$의 mgf는 $M(\mathbf{s}) = \exp\big(A(\boldsymbol{\eta}+\mathbf{s}) - A(\boldsymbol{\eta})\big)$이다. 따라서 A는 볼록함수다.

증명. **전략** $\mathbf{T}$의 mgf를 A로 쓰면, 그 로그(누적생성함수)가 A의 평행이동이다. 누적생성함수의 원점에서의 1계와 2계 미분이 평균과 공분산이다.

① $\boldsymbol{\eta}$가 $\mathcal{H}$의 내점이므로 $\|\mathbf{s}\|$가 작으면 $\boldsymbol{\eta}+\mathbf{s} \in \mathcal{H}$이다. — 내점의 정의

② mgf를 계산한다.

$$M(\mathbf{s}) = E_{\boldsymbol{\eta}} e^{\mathbf{s}'\mathbf{T}} = \int h(x)\, e^{(\boldsymbol{\eta}+\mathbf{s})'\mathbf{T}(x) - A(\boldsymbol{\eta})} dx = e^{A(\boldsymbol{\eta}+\mathbf{s}) - A(\boldsymbol{\eta})}. \qquad \text{정의 4.1, ①}$$

③ 누적생성함수는 $K(\mathbf{s}) = \log M(\mathbf{s}) = A(\boldsymbol{\eta}+\mathbf{s}) - A(\boldsymbol{\eta})$이다. — 2회 따름정리 3.4

④ $\nabla K(\mathbf{0}) = \nabla M(\mathbf{0})/M(\mathbf{0}) = E\mathbf{T}$이고, ③으로 이것은 $\nabla A(\boldsymbol{\eta})$다. — 2회 정리 3.3, $M(\mathbf{0}) = 1$

⑤ $\nabla^2 K(\mathbf{0}) = \nabla^2 M(\mathbf{0}) - \nabla M(\mathbf{0})\nabla M(\mathbf{0})' = E\mathbf{T}\mathbf{T}' - E\mathbf{T}\,E\mathbf{T}' = \mathrm{Cov}\,\mathbf{T}$이고, ③으로 이것은 $\nabla^2 A(\boldsymbol{\eta})$다. — 몫의 미분, 2회 정리 3.3

⑥ 모든 $\mathbf{v}$에서 $\mathbf{v}'\nabla^2 A(\boldsymbol{\eta})\mathbf{v} = \mathrm{Var}(\mathbf{v}'\mathbf{T}) \geq 0$이므로 A는 $\mathcal{H}$의 내부에서 볼록이다. — ⑤, 3회 정리 6.2

조건 확인 ④와 ⑤는 mgf를 원점에서 미분해 적률을 얻는 2회 정리 3.3을 쓴다. 그 정리는 mgf가 원점의 열린 근방에서 유한할 것을 요구하고, ①이 그것을 준다. 미분과 적분의 순서를 바꾸는 근거는 [T6]이다. ∎

따름정리 4.4 (지수족의 가능도) 자연모수에 대해 (a) 한 관측의 Fisher 정보는 $I(\boldsymbol{\eta}) = \nabla^2 A(\boldsymbol{\eta}) = \mathrm{Cov}\,\mathbf{T}$, (b) 로그가능도 $\ell(\boldsymbol{\eta}) = \boldsymbol{\eta}'\sum \mathbf{T}(x_i) - nA(\boldsymbol{\eta}) +$ 상수는 오목이고, (c) 가능도 방정식은 $\frac{1}{n}\sum \mathbf{T}(x_i) = \nabla A(\hat{\boldsymbol{\eta}}) = E_{\hat{\boldsymbol{\eta}}}\mathbf{T}$, 곧 "충분통계량의 평균을 모형의 평균과 같게 놓기"이다.

증명. **전략** 로그밀도를 두 번 미분하면 자료가 사라진다. 정보, 오목성, 가능도 방정식이 모두 정리 4.3의 두 식에서 나온다. 단계별 증명은 연습 4.5(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 37쪽에 있다.

(c)는 지수족에서 최대가능도가 **충분통계량에 대한** 적률법이라는 뜻이다. Poisson의 $\hat{\lambda} = \bar{x}$, 정규의 $(\bar{x}, \hat{\sigma}^2)$, 감마의 MLE 방정식(9회 예제 4.4: $\overline{\log x}$와 $\bar{x}$를 맞춘다)이 모두 이 꼴이다. Pearson의 적률법(10회)이 비효율적이었던 것은 혼합분포가 지수족이 아니어서 맞춘 적률이 충분통계량이 아니었기 때문이다. 그림 4.1은 Bernoulli에서 A, A'(평균), A''(분산이자 정보)의 관계를 보여 준다.

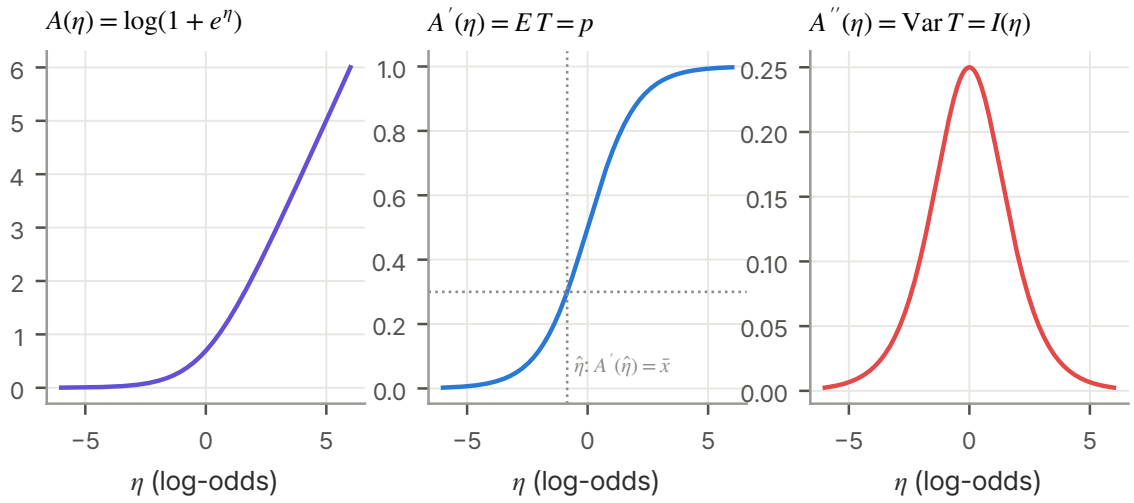


그림 4.1 Bernoulli의 로그분배함수 $A(\eta) = \log(1 + e^{\eta})$. 왼쪽: A는 볼록하다. 가운데: $A'(\eta)$는 평균 p (로지스틱 함수)이고, 가능도 방정식은 $A'(\hat{\eta}) = \bar{x}$다. 오른쪽: $A''(\eta) = p(1-p)$는 T의 분산이자 η의 Fisher 정보다. 로그 오즈가 클수록 정보가 적다.

4.2 완비충분통계량

> **정리 4.5** (지수족의 완비충분통계량) X_i가 k-모수 지수족에서 나왔고 $\mathcal{H}$가 $\mathbb{R}^k$의 열린 집합을 포함하면, $\mathbf{S} = \sum_{i=1}^{n} \mathbf{T}(X_i)$는 $\boldsymbol{\eta}$의 완비충분통계량이다.

증명. **전략** 충분성은 인수분해 정리로 보인다. 완비성은 $\mathbf{S}$의 분포도 지수족이라는 것을 쓰면 "평균이 0" 이 "Laplace 변환이 같다" 로 바뀌고, Laplace 변환의 유일성으로 끝난다.

충분성

① 표본의 결합밀도를 지수족의 꼴로 쓴다.

$$\prod_{i=1}^{n} f(x_i;\boldsymbol{\eta}) = \exp\!\big(\boldsymbol{\eta}'\mathbf{S} - nA(\boldsymbol{\eta})\big)\cdot \prod_{i=1}^{n} h(x_i), \qquad \mathbf{S} = \sum_{i} \mathbf{T}(x_i). \tag{정의 4.1}$$

② 앞 인수는 $\mathbf{S}$와 $\boldsymbol{\eta}$의 함수이고 뒤 인수는 $\boldsymbol{\eta}$에 의존하지 않으므로 $\mathbf{S}$는 충분하다. (정리 1.3)

완비성

① $\mathbf{S}$의 분포는 $h_n(\mathbf{s})\,e^{\boldsymbol{\eta}'\mathbf{s} - nA(\boldsymbol{\eta})}$ 꼴이다. ①의 밀도를 $\{\mathbf{S} = \mathbf{s}\}$ 위에서 더하거나 적분하면 지수 인수는 그대로 남는다. (①, [T7])

② $E_{\boldsymbol{\eta}} g(\mathbf{S}) = 0$이 모든 $\boldsymbol{\eta}$에서 성립한다고 하자. $e^{-nA(\boldsymbol{\eta})} > 0$이므로 열린 집합 위의 모든 $\boldsymbol{\eta}$에서 $\int g(\mathbf{s}) h_n(\mathbf{s}) e^{\boldsymbol{\eta}'\mathbf{s}}\, d\mathbf{s} = 0$이다. (③)

③ $g = g^+ - g^-$로 나누면 $\int e^{\eta' \mathbf{s}} g^+(\mathbf{s}) h_n(\mathbf{s})\, d\mathbf{s} = \int e^{\eta' \mathbf{s}} g^-(\mathbf{s}) h_n(\mathbf{s})\, d\mathbf{s}$가 그 열린 집합 위에서 성립한다.

④ 두 측도 $g^{\pm} h_n\, d\mathbf{s}$는 같다. 따라서 $h_n > 0$인 곳에서 $g^+ = g^-$, 곧 $g(\mathbf{S}) = 0$ (확률 1)이다.

Laplace 변환의 유일성(아래 상자)

조건 확인 아래 상자의 사실은 유한 측도에 대한 것이다. 열린 집합 안의 한 점 η_0을 고르면 $E_{\eta_0}|g(\mathbf{S})| < \infty$이므로 $\mu^{\pm}(d\mathbf{s}) = e^{\eta_0' \mathbf{s}} g^{\pm} h_n\, d\mathbf{s}$는 유한 측도이고, 그 Laplace 변환은 $\eta - \eta_0$이 원점 근방을 움직일 때 같다. ⑥에는 이것을 쓴다. 열린 집합을 포함한다는 가정이 여기서 쓰인다. ■

> **증명 없이 쓰는 사실** (Laplace 변환의 유일성) 두 유한 측도의 Laplace 변환 $\int e^{\eta' \mathbf{s}} \mu(d\mathbf{s})$이 η의 열린 집합 위에서 같으면 두 측도는 같다. 2회 3절의 상자 "mgf의 유일성"과 같은 사실이다. 증명은 해석함수의 유일성(또는 특성함수의 역변환)을 쓴다.

정규 표본의 $(\sum X_i, \sum X_i^2)$, 곧 $(\bar{X}, S^2)$는 완비충분이고(자연모수 공간 $\mathbb{R} \times (-\infty, 0)$이 열린 집합), Poisson의 $\sum X_i$, 이항의 성공 수도 그렇다. 3절의 UMVUE들이 이 정리 위에 선다.

4.3 Cramér–Rao의 등호는 지수족에서만

> **정리 4.6** (한계에 닿는 모형) 9회의 정칙조건 아래, 어떤 통계량 W가 모든 θ에서 $\tau(\theta) = E_\theta W$에 대한 Cramér–Rao 한계 $\dfrac{\tau'(\theta)^2}{nI(\theta)}$에 닿고 $\tau' \neq 0$이면, 표본의 밀도는 1-모수 지수족 $\prod f(x_i;\theta) = H(\mathbf{x}) \exp\big(c(\theta) W(\mathbf{x}) - d(\theta)\big)$이다.

증명. **전략** 9회 정리 2.4의 증명에서 등호가 성립하는 경우, 곧 Cauchy–Schwarz 부등식의 등호 조건을 점수에 쓰고, 얻은 식을 θ에 대해 적분한다. 단계별 증명은 연습 4.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 37쪽에 있다.

9회 2절에서 Cramér–Rao 한계에 닿은 추정량들(Poisson의 $\bar{X}$, Bernoulli의 $\hat{p}$, 정규 평균의 $\bar{X}$)은 모두 지수족의 자연 충분통계량의 평균이었다. 정리는 이것이 우연이 아니라고 말한다. 그리고 지수족에서도 한계에 닿는 것은 $E\,\mathbf{T}$의 일차함수뿐이다. $e^{-\lambda}$의 UMVUE는 한계에 닿지 않는다(연습 2.1).

> **증명 없이 쓰는 사실** (Darmois–Koopman–Pitman) 받침이 모수에 의존하지 않는 분포족에서, 모든 n에 대해 차원이 고정된(연속인) 충분통계량이 있으면 그 족은 지수족이다. 이 과정에서는 증명하지 않는다. 균등분포 $\mathrm{U}(0,\theta)$는 받침이 모수에 의존해서 예외다($X_{(n)}$ 하나로 충분하지만 지수족이 아니다). Cauchy처럼 지수족이 아니면 자료를 줄일 수 없다(연습 1.4).

연습문제

연습 4.1 (감마의 로그분배함수 ★) 예제 4.2의 감마의 $A(\eta_1, \eta_2)$를 미분해 $E\log X = \psi(\alpha)+\log\beta$, $EX = \alpha\beta$를 얻어라. 9회 연습 3.3의 정보 행렬이 $\nabla^2 A$와 어떤 관계인가? 풀이 36쪽

연습 4.2 (Poisson의 누적량 ★) Poisson의 $A(\eta) = e^\eta$의 모든 미분이 λ임을 보이고, Poisson의 모든 누적량이 λ라는 것을 확인하라. 풀이 36쪽

연습 4.3 (한계에 닿는 함수 ★★) Poisson에서 Cramér–Rao 한계에 닿는 불편추정량이 있는 $\tau(\lambda)$는 $a\lambda + b$ 꼴뿐임을 정리 4.6의 증명으로 보여라. 풀이 37쪽

연습 4.4 (softmax ★★) 범주 분포의 $A(\boldsymbol{\eta}) = \log\sum_k e^{\eta_k}$에서 $\nabla A = \mathrm{softmax}(\boldsymbol{\eta})$, $\nabla^2 A = \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'$임을 보여라. 이 행렬은 가역이 아니다. 왜인가? (식별가능성, 9회 연습 1.3)풀이 37쪽

연습 4.5 (지수족의 가능도: 증명 ★ 유도형) 따름정리 4.4를 다음 순서로 증명하라.

(a) 한 관측의 로그밀도 $\log f(x;\boldsymbol{\eta}) = \log h(x) + \boldsymbol{\eta}'\mathbf{T}(x) - A(\boldsymbol{\eta})$의 헤세 행렬이 $-\nabla^2 A(\boldsymbol{\eta})$로 x에 의존하지 않음을 보이고, 9회 정리 2.2의 둘째 표현으로 $I(\boldsymbol{\eta}) = \nabla^2 A(\boldsymbol{\eta})$를 얻어라.

(b) $\ell(\boldsymbol{\eta})$의 헤세 행렬이 $-n\nabla^2 A(\boldsymbol{\eta})$임을 보이고, 정리 4.3로 ℓ이 오목임을 결론지어라.

(c) $\nabla\ell(\hat{\boldsymbol{\eta}}) = \mathbf{0}$을 풀어 가능도 방정식 $\frac{1}{n}\sum\mathbf{T}(x_i) = \nabla A(\hat{\boldsymbol{\eta}}) = E_{\hat{\eta}}\mathbf{T}$를 얻어라. (b)로 이 방정식의 해는 최대점이다. 왜인가?

풀이 37쪽

연습 4.6 (Cramér–Rao의 등호: 증명 ★★ 유도형) 정리 4.6를 다음 순서로 증명하라. $s(\mathbf{x};\theta) = \partial_\theta \log\prod f(x_i;\theta)$는 표본의 점수다.

(a) 9회 정리 2.4의 증명에서 $\mathrm{Cov}_\theta(W, s) = \tau'(\theta)$이고 한계는 Cauchy–Schwarz 부등식 $\mathrm{Cov}(W,s)^2 \le \mathrm{Var}\,W\ \mathrm{Var}\,s$에서 나왔다. 등호이면 θ마다 상수 $a(\theta)$가 있어 $s(\mathbf{X};\theta) = a(\theta)\big(W(\mathbf{X}) - \tau(\theta)\big)$ (확률 1)임을 보여라. $E_\theta s = 0$을 어디서 쓰는가?

(b) (a)의 식을 θ에 대해 적분해 $\log\prod f(x_i;\theta) = c(\theta)W(\mathbf{x}) - d(\theta) + \log H(\mathbf{x})$ ($c' = a$, $d' = a\tau$)를 얻어라. 적분 상수가 $\mathbf{x}$에만 의존하는 이유는?

(c) Poisson 표본과 $W = \bar{X}$에서 $a(\lambda)$, $c(\lambda)$, $d(\lambda)$, $H(\mathbf{x})$를 구해 (b)의 꼴을 확인하라.

풀이 37쪽

5

보조통계량과 Basu의 정리

> **역사** Fisher(1925, 1934)는 분포가 모수와 무관한 통계량을 **보조**(ancillary)통계량이라 불렀다. 그것 자체는 모수에 대해 아무것도 알려 주지 않지만, 추정의 정밀도가 얼마인지는 알려 줄 수 있다. 인도 통계연구소의 D. Basu(1955)는 완비충분통계량과 보조통계량이 언제나 독립이라는 짧은 정리를 증명했다. 이 정리는 많은 독립성 증명을 한 줄로 만들었다.

정의 5.1 (보조통계량) 분포가 θ에 의존하지 않는 통계량 $V(\mathbf{X})$를 **보조통계량**이라 한다.

위치모형 $f(x-\theta)$에서 $X_i - X_j$나 $X_i - \bar{X}$, 척도모형 $f(x/\sigma)/\sigma$에서 X_i/X_j나 $X_i/\sum X_j$가 보조통계량이다. 위치나 척도를 바꿔도 이 값들은 변하지 않는다.

정리 5.2 (Basu (1955)) T가 완비충분통계량이고 V가 보조통계량이면, 모든 θ에서 T와 V는 독립이다.

증명. **전략** 사건 $\{V \in B\}$의 조건부확률 $\eta(T)$를 만든다. 그 평균은 모든 θ에서 같은 상수 p_B이므로, 완비성으로 $\eta(T)$ 자체가 그 상수다.

① 집합 B를 고정하고 $p_B = P_\theta(V \in B)$라 하자. p_B는 θ에 의존하지 않는다. — V는 보조, 정의 5.1

② $\eta(T) = P(V \in B \mid T)$는 θ를 포함하지 않는 T의 함수다. — T는 충분, 정의 1.1

③ 모든 θ에서 $E_\theta\big[\eta(T) - p_B\big] = P_\theta(V \in B) - p_B = 0$이다. — 3회 정리 2.5, ①

④ 따라서 모든 θ에서 $\eta(T) = p_B$ (확률 1)이다. — T의 완비성, 정의 3.1

⑤ 임의의 집합 C에 대해 다음이 성립하므로 T와 V는 독립이다.

$$P_\theta(T \in C,\, V \in B) = E_\theta\big[\mathbb{1}\{T \in C\}\,\eta(T)\big] = p_B\, P_\theta(T \in C). \qquad \text{반복기댓값, ④}$$

조건 확인 완비성은 평균이 정의된 $g(T)$에 대한 조건이다. $\eta(T) - p_B$는 -1과 1 사이에 있으므로 평균이 언제나 정의된다. B와 C는 확률을 말할 수 있는 모든 집합(1회 7.2절의 Borel 집합)을 움직인다. ∎

예제 5.3 (정규 표본에서 $\bar{X} \perp S^2$, 다시) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$에서 σ^2을 하나 고정하고 μ만 모수로 보자. 그러면 $\bar{X}$는 μ의 완비충분통계량이다(정리 4.5, 1-모수 정규). $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$은 $X_i - \mu$들만의 함수이므로 분포가 μ에 의존하지 않는다(보조). Basu로 $\bar{X} \perp S^2$이다. 이것이 모든 σ^2에서 성립하므로 6회 정리 2.2(b)가 증명되었다. 6회의 Helmert 변환이 직교 행렬의 계산이었다면, 이 증명은 "정보의 구조"에서 나온다. $\bar{X}$는 μ에 대한 모든 것을 갖고, S^2은 μ에 대해 아무것도 갖지 않으니, 둘이 겹칠 수 없다. 정규가 아니면 이 구조가 무너진다. 지수분포 표본에서는 $\bar{X}$가 크면 S^2도 크다(그림 5.1).

예제 5.4 (Fisher의 1920년 관찰) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$이면 $\|\mathbf{X}\|^2 = \sum X_i^2$이 σ^2의 완비충분통계량이고, 방향 $\mathbf{X}/\|\mathbf{X}\|$는 척도와 무관하므로 보조다. Basu로 둘은 독립이다. 평균편차와 표준편차의 비 d/s는 방향만의 함수이므로 s와 독립이다. Fisher가 본 것이 이것이다. s를 알고 나면 d는 σ와 무관하게 흔들리는 잡음일 뿐이어서, d를 함께 써도 σ에 대해 더 알게 되는 것이 없다.

예제 5.5 (감마와 균등) (a) $X_i \overset{\text{iid}}{\sim}$ 감마(α, β) (α는 알고 있음)이면 $\sum X_i$는 β의 완비충분이고 $(X_1/\sum X_j, \ldots, X_n/\sum X_j)$는 보조이므로 독립이다. 4회 정리 4.2의 베타–감마 관계를 다시 얻는다. (b) $\mathrm{U}(0, \theta)$에서 $X_{(n)}$은 완비충분이고 $X_{(1)}/X_{(n)}$은 보조이므로 독립이다. 그래서 $E\frac{X_{(1)}}{X_{(n)}} \cdot EX_{(n)} = EX_{(1)}$로 기댓값의 비를 계산할 수 있다(연습 5.1).

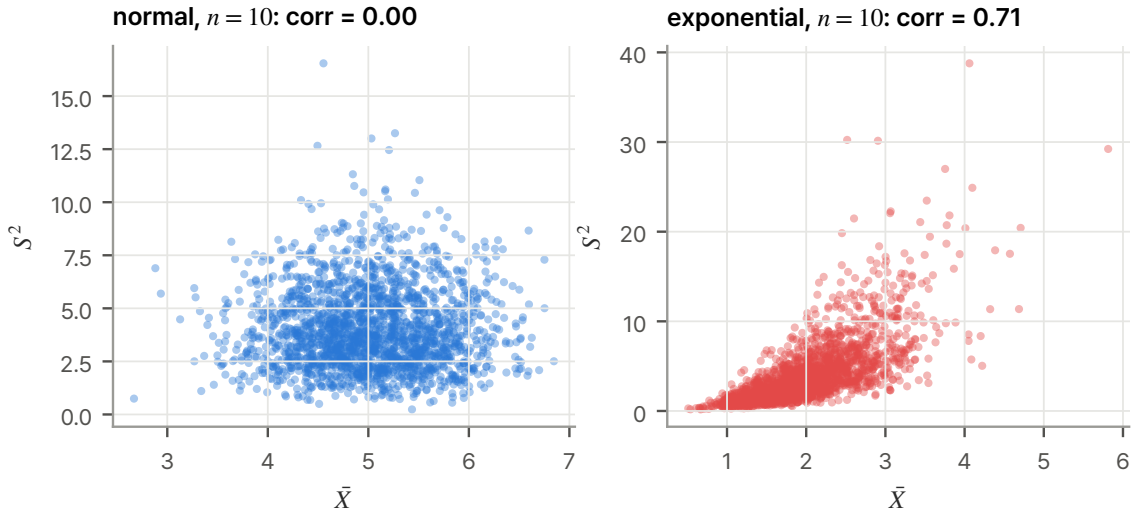


그림 5.1 표본 2,000개($n = 10$)의 $(\bar{X}, S^2)$. 왼쪽: 정규 표본에서는 독립이다(Basu). 오른쪽: 지수분포 표본에서는 $\bar{X}$가 척도의 충분통계량이고 S^2은 보조가 아니어서 강하게 상관된다. 이 경우에도 Basu가 쓰이는 것은 $S^2/\bar{X}^2$ 같은 척도 불변량이다.

가정 주의 완비성을 빼면 정리가 무너진다. $\mathrm{U}(\theta, \theta+1)$에서 $(X_{(1)}, X_{(n)})$은 최소충분이지만 완비가 아니고, 보조통계량인 범위 $X_{(n)} - X_{(1)}$과 독립이 아니다(범위가 크면 $X_{(1)}$은 θ 쪽에 몰린다). 반대 방향도 성립하지 않는다. 독립이라고 해서 하나가 완비충분인 것은 아니다.

연습문제

연습 5.1 (균등분포의 비 ★) $\mathrm{U}(0, \theta)$에서 Basu의 정리로 $E\dfrac{X_{(1)}}{X_{(n)}} = \dfrac{EX_{(1)}}{EX_{(n)}} = \dfrac{1}{n}$임을 보여라. 풀이 37쪽

연습 5.2 (지수분포의 간격 ★★) $X_i \overset{\text{iid}}{\sim} \text{Exp}$ (평균 θ)이면 $\sum X_i$와 $X_{(1)}/\sum X_i$가 독립임을 보이고, 이것으로 $E\left[X_{(1)}/\sum X_i\right] = \dfrac{1}{n^2}$을 구하라. ($EX_{(1)} = \theta/n$, 8회 연습 1.2) 풀이 38쪽

연습 5.3 (정규의 표준화된 잔차 ★★) 정규 표본에서 $(\bar{X}, S)$와 $\left(\dfrac{X_1-\bar{X}}{S}, \ldots, \dfrac{X_n-\bar{X}}{S}\right)$가 독립임을 보여라. 두 모수 (μ, σ)의 완비충분통계량과 위치–척도 보조통계량이다. 풀이 38쪽

6

데이터로 — Fisher의 계산과 Newcomb의 빛

6.1 1920년의 계산

정규 표본 $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$에서 σ를 재는 두 통계량을 비교한다. 표준편차 s와 평균편차 $d = \frac{1}{n}\sum |X_i - \bar{X}|$이다. 둘은 척도가 다르므로$(Ed \approx \sigma\sqrt{2/\pi})$ 로그를 취해 비교한다. $\log s$와 $\log d$의 점근분산의 비가 상대효율이다.

> **증명 없이 쓰는 사실** (중심을 추정해도 극한은 같다) F가 μ에 대해 대칭이고 평균 근처에서 연속이면, $d = \frac{1}{n}\sum |X_i - \bar{X}|$와 $\frac{1}{n}\sum |X_i - \mu|$의 차는 $o_P(n^{-1/2})$이다. 직관은 이렇다. $m \mapsto E|X - m|$의 $m = \mu$에서의 미분은 $2F(\mu) - 1 = 0$이므로, 중심을 $O_P(n^{-1/2})$만큼 잘못 잡아도 평균편차는 그 제곱의 크기로만 변한다. 증명은 경험과정의 논증이 필요해 생략한다. 표준편차에서는 같은 사실이 정확한 계산으로 성립한다(7회 예제 5.5).

명제 6.1 (평균편차의 효율 (Fisher 1920)) 정규 표본에서 $\sqrt{n}(\log d - \log Ed)$와 $\sqrt{n}(\log s - \log\sigma)$의 점근분산은 각각 $\frac{\pi}{2}-1$과 $\frac{1}{2}$이고, 평균편차의 상대효율은

$$\frac{1/2}{\pi/2-1} = \frac{1}{\pi-2} = 0.876$$

이다. 곧 같은 정밀도를 얻으려면 평균편차는 표준편차보다 관측이 $\pi-2 = 1.14$배 필요하다.

증명. **전략** 상자의 사실로 d를 iid 평균 $\frac{1}{n}\sum|X_i-\mu|$로 바꾸고, 중심극한정리와 델타 방법을 두 통계량에 각각 쓴다 [T8]. 단계별 증명은 연습 6.4(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 38쪽에 있다.

$n=100$에서 10만 번 시뮬레이션하면 분산의 비는 약 0.88로 극한값 0.876에 가깝다. Fisher가 옳았다. 다만 그것은 **오차가 정규분포를 따른다면**의 이야기다. Eddington은 실제 관측에는 가끔 큰 오차가 섞인다는 것을 알고 있었다.

6.2 Newcomb의 빛의 통과 시간 (1882)

예제 6.2 (큰 오차 두 개) Simon Newcomb는 1882년 Potomac 강가의 실험실에서 Washington 기념탑 아래의 거울까지 빛이 갔다 오는 시간(약 7.4 km 왕복)을 66번 쟀다(`MASS::newcomb`, 24,800나노초에서 뺀 값). 64개는 16에서 40 사이에 있지만 두 개가 -44와 -2로 동떨어져 있다. 66개 전체의 표준편차는 10.75, 두 값을 빼면 5.08로 절반이 된다. 평균편차로 만든 같은 척도의 추정량 $\sqrt{\pi/2}\,d$는 6.80에서 4.90으로, 덜 변한다. 표본을 다시 뽑아(부트스트랩, 15회) $\log s$와 $\log d$의 분산의 비를 재면, 66개 전체에서 2.48로 평균편차가 훨씬 안정하다. 두 값을 뺀 64개에서는 0.78로, 정규 이론의 0.876 근처로 돌아간다(그림 6.1).

표준편차는 정규 모형 아래에서 충분하다. 모형이 맞지 않으면 충분성은 아무것도 보장하지 않는다. 큰 오차 하나가 $\sum X_i^2$을 지배하고, 표준편차는 그 하나를 정직하게 따라간다. 7절은 "정규에 아주 조금 섞인 큰 오차"만으로 순서가 뒤집힘을 보인다. 15회의 로버스트 통계가 Newcomb의 자료에서 다시 시작한다(Stigler 1977).

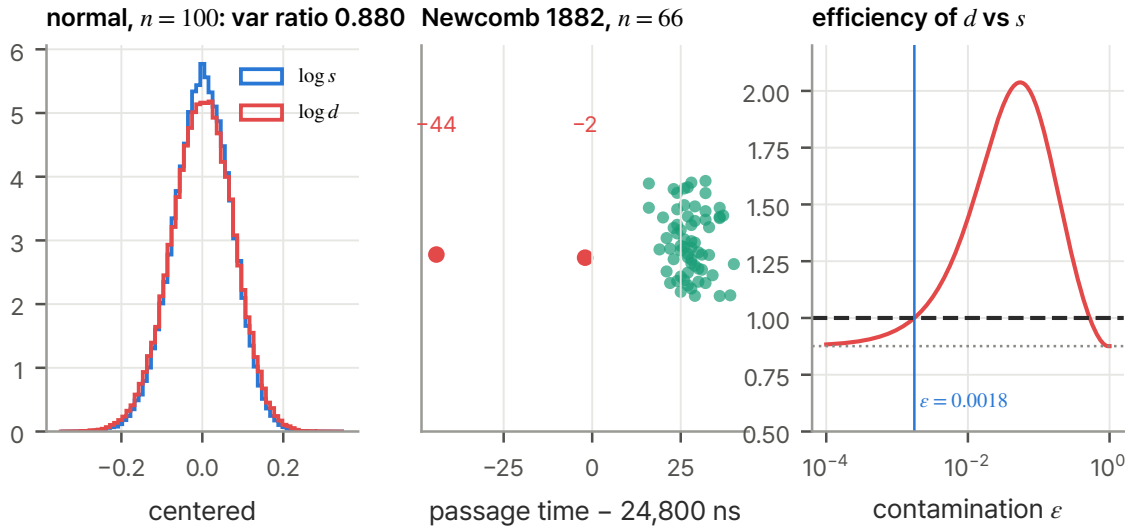


그림 6.1 Eddington 대 Fisher. 왼쪽: 정규 표본($n = 100$)에서 $\log s$와 $\log d$의 표본분포(가운데를 맞춤). 평균편차가 조금 더 넓다(분산의 비 0.876). 가운데: Newcomb의 66번 측정. 빨간 점은 동떨어진 두 값. 오른쪽: 오염 정규 $(1-\varepsilon)\mathcal{N}(0,1)+\varepsilon\mathcal{N}(0,9)$에서 평균편차의 상대효율. $\varepsilon = 0.0018$에서 1을 넘는다(7절).

6.3 기병대와 독일 전차

예제 6.3 (세 추정량의 비교) 기병대 자료의 $P(X=0)$ 추정값은 UMVUE $(279/280)^{196} = 0.4960$, MLE $e^{-0.7} = 0.4966$, 관측 비율 0.514이다. 앞의 둘은 $n = 280$에서 사실상 같고, 셋째(9회 연습 1.4의 비교)는 Poisson 모형을 쓰지 않은 추정이다. 모형을 쓰면 표준오차가 줄지만(관측 비율의 $\sqrt{0.514\cdot 0.486/280} = 0.030$ 대 UMVUE의 $\sqrt{e^{-2\lambda}(e^{\lambda/n}-1)} = 0.025$), 모형이 틀리면 치우친다. 이 자료에서 0.018의 차이는 표준오차보다 작아 우연으로 설명되지만, 군단 사이의 이질성(4회 $\chi^2 = 27.3$, 9회 $G = 26.1$)은 Poisson 모형의 가정 자체를 의심하게 한다.

독일 전차에서 Rao–Blackwell화의 효과는 그림 6.2에 있다. 같은 다섯 개의 일련번호에서, 평균을 쓰는 적률법보다 최댓값을 쓰는 UMVUE가 표준편차 77 대 50으로 훨씬 좁다. 3절의 결론대로 최댓값 쪽은 모든 불편추정량 가운데 최선이다.

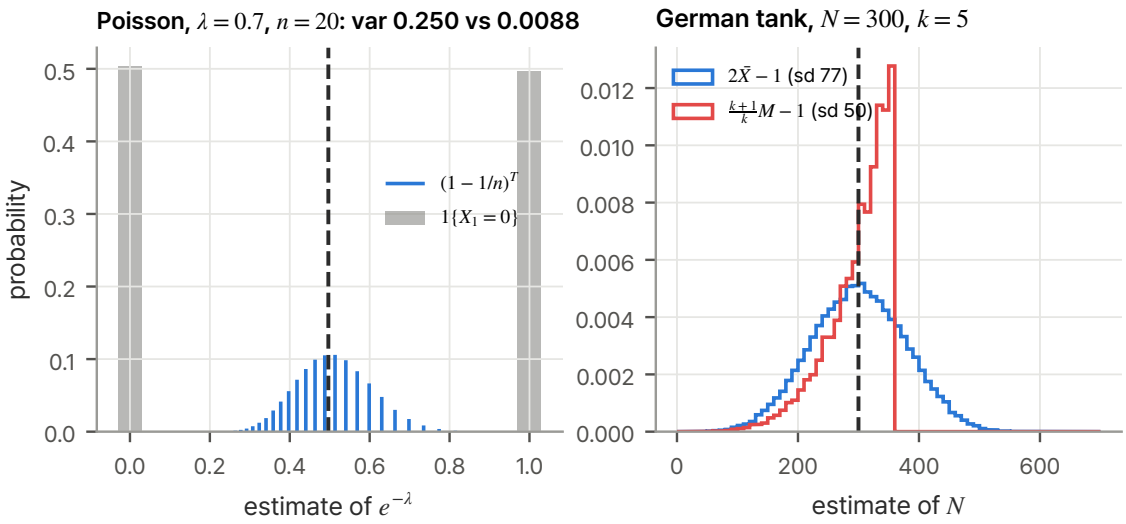


그림 6.2 Rao–Blackwell화. 왼쪽: Poisson($\lambda = 0.7$, $n = 20$)에서 $e^{-\lambda}$의 두 불편추정량. 관측 하나의 지시함수 $1\{X_1 = 0\}$(0과 1, 회색 막대)에 $\sum X_i$로 조건을 건 $(1-1/n)^T$(파랑)는 분산이 28분의 1이다. 점선은 참값. 오른쪽: 독일 전차($N = 300$, $k = 5$). 적률법 $2\bar{X} - 1$(파랑)과 그것에 최댓값으로 조건을 건 UMVUE $\frac{k+1}{k}M - 1$(빨강).

연습문제

연습 6.1 (평균편차의 적률 ★) $Z \sim \mathcal{N}(0,1)$이면 $E|Z| = \sqrt{2/\pi}$, $\mathrm{Var}\,|Z| = 1 - \frac{2}{\pi}$임을 보이고 명제 6.1의 $\frac{\pi}{2} - 1$을 확인하라. 풀이 38쪽

연습 6.2 (Laplace 오차에서는 ★★) 오차가 Laplace(이중지수) 분포이면 평균편차의 상대효율이 1보다 큼을 보여라. ($\mu_4/\sigma^4 = 6$, $E|X| = b$, $EX^2 = 2b^2$.) 4회의 Laplace(1774)는 이 분포에서 중앙값을 권했다. 풀이 38쪽

연습 6.3 (Newcomb의 부트스트랩 ★) 노트북에서 Newcomb 자료의 부트스트랩 분산 비 2.48과 0.78을 재현하라. 두 값만 뺐는데 왜 이렇게 달라지는가? 풀이 38쪽

연습 6.4 (평균편차의 효율: 증명 ★★ 유도형) 명제 6.1를 다음 순서로 증명하라. 상자의 사실을 받아들여 d 대신 $\tilde{d} = \frac{1}{n}\sum |X_i - \mu|$를 쓴다.

(a) $E|X-\mu| = \sigma\sqrt{2/\pi}$, $\mathrm{Var}\,|X-\mu| = \sigma^2(1 - \frac{2}{\pi})$임을 연습 6.1에서 가져오고, 중심극한정리(7회 정리 3.1)로 $\sqrt{n}(\tilde{d} - E\tilde{d})$의 극한분포를 적어라.

(b) 델타 방법(7회 정리 4.4)을 $g(u) = \log u$에 써서 $\sqrt{n}(\log\tilde{d} - \log E\tilde{d})$의 점근분산이 $\frac{\pi}{2} - 1$임을 보여라.

(c) S^2의 점근분산 $\mu_4 - \sigma^4 = 2\sigma^4$(7회 예제 5.5)에 $g(u) = \frac{1}{2}\log u$로 델타 방법을 써서 $\log s$의 점근분산 $\frac{1}{2}$를 얻어라.

(d) 두 점근분산의 비로 상대효율 $\frac{1}{\pi-2}$를 얻고, 같은 정밀도에 필요한 관측 수의 비로 읽어라.

풀이 38쪽

7

부숴보기 — 큰 오차와 터무니없는 UMVUE

7.1 조금 섞인 큰 오차 (Tukey 1960)

John Tukey는 1960년 대부분은 정규이고 가끔 큰 오차가 섞이는 모형을 썼다.

$$X \sim (1-\varepsilon)\mathcal{N}(0,1) + \varepsilon\mathcal{N}(0,9).$$

ε의 확률로 표준편차가 세 배인 관측이 나온다. 그는 이 모형에서 평균편차와 표준편차를 다시 비교했다.

명제 7.1 (오염 정규에서의 효율) 대칭분포에서 $\log s$와 $\log d$의 점근분산은 $\frac{1}{4}\left(\frac{\mu_4}{\mu_2^2}-1\right)$과 $\frac{\mu_2}{\nu_1^2}-1$이다($\mu_k = EX^k$, $\nu_1 = E|X|$). 위 모형에서는 $\mu_2 = 1+8\varepsilon$, $\mu_4 = 3(1+80\varepsilon)$, $\nu_1 = \sqrt{2/\pi}\,(1+2\varepsilon)$이고, 평균편차의 상대효율은

ε	0	0.001	0.002	0.01	0.05	0.1
효율	0.876	0.948	1.016	1.439	2.035	1.903

이다. $\varepsilon = 0.0018$에서 1을 넘는다.

증명. **전략** 명제 6.1의 증명을 일반 대칭분포에서 되풀이하고, 혼합분포의 적률을 성분 적률의 가중평균(4회 정리 5.2)으로 구한다. 단계별 증명은 연습 7.4(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 39쪽에 있다.

1,000개 가운데 두 개만 큰 오차여도 Eddington이 이긴다. 표준편차의 충분성은 정규 모형이 **정확히** 맞을 때의 성질이고, 그 성질은 모형의 작은 어긋남에 아주 민감하다. 이 관찰이 Tukey와 Huber(1964)의 로버스트 통계를 낳았다(15회). 그림 6.1의 오른쪽이 이 곡선이다.

7.2 터무니없는 UMVUE

예제 7.2 (음수가 되는 확률의 추정량) 관측 하나 $X \sim \mathrm{Poi}(\lambda)$로, 두 해 연속 아무도 죽지 않을 확률 $\tau(\lambda) = e^{-2\lambda}$를 추정한다. X는 완비충분이고

$$E_\lambda (-1)^X = e^{-\lambda}\sum_{x} \frac{(-\lambda)^x}{x!} = e^{-2\lambda}$$

> 이므로 Lehmann–Scheffé로 $(-1)^X$가 유일한 UMVUE다. 그런데 이것은 X가 짝수이면 1, 홀수이면 -1이다. $(0,1)$ 사이의 확률을 1이나 -1로 추정한다. 평균제곱오차는 $1-e^{-4\lambda}$로 $\lambda=0.7$에서 0.94인데, 치우친 MLE e^{-2X}는 0.29다(그림 7.1).

이 예는 Lehmann–Scheffé 정리가 틀렸다는 뜻이 아니다. 불편추정량이 하나뿐이면 그것이 최선의 불편추정량이다. 그러나 "불편"이라는 조건 자체가 그 하나를 강요했다. 관측이 n개이면 UMVUE는 $(1-\frac{2}{n})^T$로(연습 7.2), $n\ge 3$이면 0과 1 사이에 머문다. 불편성은 큰 표본에서는 거의 공짜지만 작은 표본에서는 비쌀 수 있다.

7.3 불편추정량이 없을 때

$K\sim \mathrm{b}(n,p)$이면 아무 통계량 $g(K)$의 평균 $\sum_k g(k)\binom{n}{k}p^k(1-p)^{n-k}$은 p의 n차 이하 다항식이다. 따라서 $1/p$나 오즈 $p/(1-p)$처럼 다항식이 아닌 $\tau(p)$에는 불편추정량이 하나도 없다(연습 7.3). 9회의 로그 오즈 $\log\frac{\hat p}{1-\hat p}$는 불편이 아니어도 점근적으로 효율적이었다. 불편성은 출발점일 뿐이다.

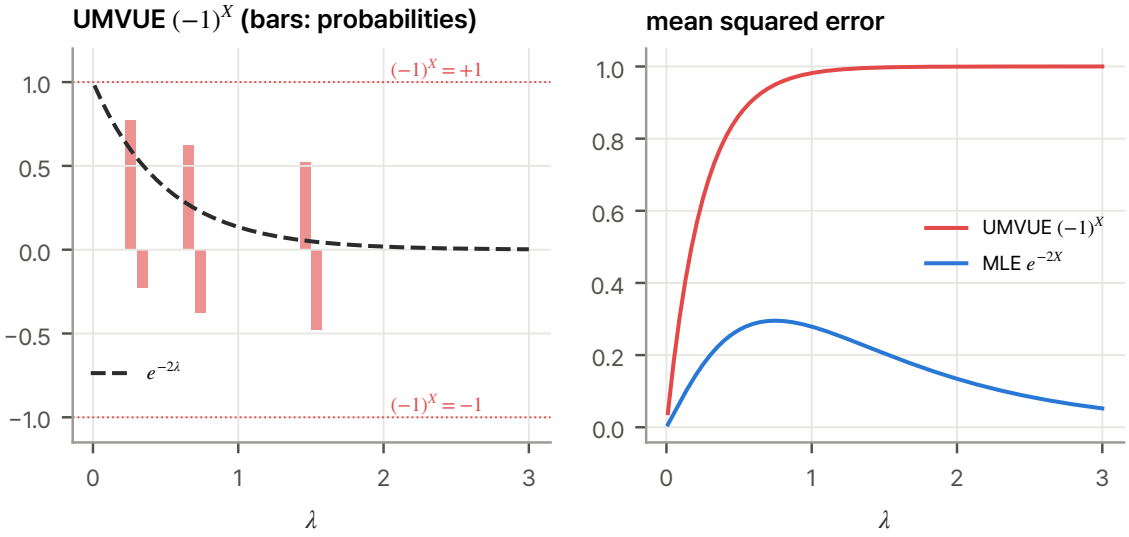


그림 7.1 왼쪽: $X\sim\mathrm{Poi}(\lambda)$ 하나로 추정한 $e^{-2\lambda}$(점선). UMVUE $(-1)^X$의 값은 ±1뿐이고(막대의 높이는 그 확률), MLE e^{-2X}는 $(0,1]$에 있다. 오른쪽: 두 추정량의 평균제곱오차.

연습문제

연습 7.1 (오염 정규의 효율 ★★) 명제 7.1의 적률을 확인하고, 효율이 1이 되는 ε이 0.0018임을 노트북으로 구하라. ε이 커지면 효율이 다시 줄어드는 이유는? 풀이 39쪽

연습 7.2 (관측이 n개이면 ★★) $X_i \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$에서 $e^{-2\lambda}$의 UMVUE가 $(1-\frac{2}{n})^T$임을 보여라. $n=1$과 $n=2$에서 각각 무엇이 되는가? 풀이 39쪽

연습 7.3 (불편추정량이 없다 ★★) $K \sim \text{b}(n,p)$일 때 $1/p$의 불편추정량이 없음을 보여라. 풀이 39쪽

연습 7.4 (오염 정규에서의 효율: 증명 ★★ 유도형) 명제 7.1를 다음 순서로 증명하라. X는 0에 대해 대칭이고 $\mu_4 < \infty$이다.

(a) $\frac{1}{n}\sum X_i^2$의 점근분산이 $\mu_4 - \mu_2^2$임을 보이고, 델타 방법으로 $\log s$의 점근분산 $\frac{1}{4}\left(\frac{\mu_4}{\mu_2^2}-1\right)$을 얻어라. (중심을 추정하는 영향은 6절의 상자처럼 무시한다.)

(b) $\frac{1}{n}\sum |X_i|$에 같은 방법을 써서 $\log d$의 점근분산 $\frac{\text{Var}\,|X|}{\nu_1^2} = \frac{\mu_2}{\nu_1^2} - 1$을 얻어라.

(c) 혼합 $(1-\varepsilon)\mathcal{N}(0,1)+\varepsilon\mathcal{N}(0,9)$의 μ_2, μ_4, ν_1을 4회 정리 5.2로 구하라. $\mathcal{N}(0,\sigma^2)$의 $EX^4 = 3\sigma^4$, $E|X| = \sigma\sqrt{2/\pi}$를 쓴다.

(d) $\varepsilon = 0$에서 명제 6.1의 0.876이 다시 나옴을 확인하라.

풀이 39쪽

8

Beautiful insights — 분산 줄이기와 softmax

8.1 Rao–Blackwell화는 분산 줄이기다

Monte Carlo로 $\mu = E\,h(X,Y)$를 추정할 때(7회 6절), Y가 주어진 조건부기댓값 $m(Y) = E[h(X,Y)\mid Y]$를 손으로 계산할 수 있으면 $h(X_i,Y_i)$ 대신 $m(Y_i)$를 평균한다. 정리 2.1의 (b)와 (c)와 같은 이유로 평균은 같고 분산은 줄어든다. 충분성은 필요 없다. m을 계산하는 데 모수가 들어가지 않기 때문이다. 예를 들어 $Z_1,Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$에서 $P(Z_1+Z_2>3)=0.0169$를 추정할 때, 지시함수 $1\{Z_1+Z_2>3\}$ 대신 $P(Z_2>3-Z_1\mid Z_1)=1-\Phi(3-Z_1)$를 평균하면 분산이 7분의 1이 된다(노트북). 16회의 Gibbs 표집기에서 사후평균을 추정할 때 같은 방법을 쓴다. 강화학습의 정책 기울기에서 기댓값의 일부를 해석적으로 적분하는 것도 같은 생각이다.

8.2 softmax는 로그분배함수의 기울기다

분류 신경망의 마지막 층은 logit $\boldsymbol{\eta} = W\mathbf{h}$를 내놓고, softmax로 범주 확률 $\mathbf{p} = \mathrm{softmax}(\boldsymbol{\eta}) = \nabla A(\boldsymbol{\eta})$를 만든다. 범주 분포를 지수족으로 쓴 것이다(예제 4.2). 그러면 교차엔트로피 $-\log p_y = A(\boldsymbol{\eta}) - \eta_y$의 logit에 대한 기울기는

$$\nabla_{\eta}\big(A(\boldsymbol{\eta}) - \boldsymbol{\eta}'\mathbf{e}_y\big) = \mathbf{p} - \mathbf{e}_y = E_{\eta}\mathbf{T} - \mathbf{T}(y)$$

($\mathbf{e}_y$는 원-핫)로, "모형이 예측하는 충분통계량의 평균 − 관측한 충분통계량"이다. 따름정리 4.4(c)의 가능도 방정식이 기울기가 0이 되는 조건이고, 헤세 행렬 $\nabla^2 A = \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'$가 Fisher 정보다. 9회 8절의 자연기울기와 온도 보정이 모두 이 구조 위에 있다. 로지스틱 회귀와 Poisson 회귀(일반화선형모형)도 자연모수를 입력의 일차함수로 놓은 지수족이다.

8.3 충분통계량은 합쳐진다

자료가 여러 기기에 흩어져 있거나(연합학습) 끝없이 흘러 들어오면(스트리밍), 자료를 모두 모으지 않고 요약만 주고받고 싶다. 지수족이면 충분통계량 $\sum \mathbf{T}(x_i)$는 **더해진다**. 기기마다 자기 합을 보내면 중앙에서 더해 정확한 MLE를 얻는다. 정규 선형회귀에서는 $X'X$와 $X'\mathbf{y}$가 그 합이다(13회). 그 반대도 4절의 Darmois–Koopman–Pitman 상자가 말한다. 크기가 자료 수에 무관한

요약으로 정보를 잃지 않으려면 모형이 지수족이어야 한다. 10회의 EM이 "충분통계량의 기댓값"을 주고받는 알고리즘인 것도, 혼합분포의 완전 자료 모형이 지수족이기 때문이다.

연습문제

연습 8.1 (조건부 Monte Carlo ★) $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$에서 $1\{Z_1 + Z_2 > 3\}$와 $1 - \Phi(3 - Z_1)$의 평균이 같고 후자의 분산이 더 작음을 정리 2.1의 증명으로 설명하라. 풀이 39쪽

연습 8.2 (교차엔트로피의 기울기 ★) $K = 3$, $\boldsymbol{\eta} = (1, 0, -1)$, 정답 $y = 2$일 때 교차엔트로피의 logit에 대한 기울기를 계산하고, 그 성분의 합이 0인 이유를 설명하라. 풀이 40쪽

9

정리와 역사 노트

9.1 한 장 요약

정리	내용	조건
인수분해 (1.3)	T 충분 $\iff f(\mathbf{x};\theta)=g(T(\mathbf{x});\theta)h(\mathbf{x})$	—
최소충분 (1.6)	가능도비가 θ에 무관 $\iff T(\mathbf{x})=T(\mathbf{y})$	받침 고정
정보의 보존 (1.7)	$I_T \le I_{\mathbf{X}}$, 충분이면 등호	정칙조건
Rao–Blackwell (2.1)	$\operatorname{Var} E(W\mid T) \le \operatorname{Var} W$, 불편성 유지	T 충분
Lehmann–Scheffé (3.3)	완비충분통계량의 불편 함수 = 유일한 UMVUE	T 완비충분
지수족 (4.3–4.5)	$E\mathbf{T}=\nabla A$, $\operatorname{Cov}\mathbf{T}=\nabla^2 A$, $\sum \mathbf{T}(X_i)$ 완비충분	자연모수 공간이 열린 집합 포함
Cramér–Rao 등호 (4.6)	한계에 닿으면 지수족	정칙조건
Basu (5.2)	완비충분 $\perp$ 보조	—

이번 회차의 중심은 "조건부기댓값은 분산을 줄인다" (3회 전분산 공식)와 "평균이 언제나 0인 함수는 0뿐이다" (완비성) 두 줄이다. 앞의 것이 Rao–Blackwell을, 둘을 합친 것이 Lehmann–Scheffé와 Basu를 준다. 그리고 6–7절은 이 모든 최적성이 모형 안에서의 이야기임을 보였다. 정규에 큰 오차가 0.2%만 섞여도 Eddington이 이긴다.

9.2 Stigler의 명명 법칙

> **이름과 발견**
>
> - **Rao–Blackwell**: Rao(1945)와 Blackwell(1947)이 따로 증명했고, Kolmogorov(1950)도 같은 결과를 냈다. 러시아 문헌에서는 Kolmogorov의 이름이 함께 붙는다.
> - **인수분해 정리**: Fisher(1922)가 예에서 쓰던 판정법을 Neyman(1935)이 일반화했고, Halmos–Savage(1949)가 측도론으로 완성했다. 이름이 "Fisher–Neyman"으로 불리기도 한다.
> - **지수족**: Darmois, Koopman, Pitman이 1935–36년 거의 동시에 같은 정리를 냈다. 정규, 이항,

> Poisson은 그보다 한 세기 앞서 있었다.
> - **Eddington과 Fisher**: 1920년의 논쟁에서 이긴 쪽은 이론에서는 Fisher, 실제 자료에서는 (Tukey 1960이 보였듯) Eddington이었다.

9.3 7 Pillars 지도

이번 회차는 첫째 기둥 **Aggregation**(모으기)의 이론적 완성이다. 1회부터 평균을 내 왔지만, 그렇게 모아도 잃는 것이 없는지는 묻지 않았다. 충분성은 "언제 모아도 잃지 않는가"에 대한 정확한 답이다. 모형이 지수족이면 몇 개의 합으로 모든 것을 모을 수 있다. 둘째 기둥 **Information**도 함께 정리되었다. 충분통계량은 Fisher 정보를 보존하고, Cramér–Rao 한계는 지수족에서만 닿는다.

9.4 다음 회 — 최적 검정과 결정 (1928 → 1956)

이번 회차는 "가장 좋은 추정량"을 찾았다. 12회는 "가장 좋은 검정"을 찾는다. Neyman과 Pearson(1933)은 크기를 고정하고 검정력을 최대화하는 검정이 가능도비로 기각하는 검정임을 증명했다. 지수족처럼 가능도비가 충분통계량에 대해 단조이면 그 검정이 모든 대립가설에서 동시에 최선이 된다(Karlin–Rubin 1956). Fisher는 이 "귀납적 행동"의 관점을 평생 거부했다. Wald(1945)의 축차검정과 결정이론이 그 논쟁 위에 섰다.

참고문헌

- Eddington, A. S. *Stellar Movements and the Structure of the Universe*. Macmillan, 1914.
- Fisher, R. A. A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices R. Astron. Soc.* 80 (1920), 758–770. · On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. A* 222 (1922), 309–368.
- Neyman, J. Sur un teorema concernente le cosiddette statistiche sufficienti. *Giorn. Ist. Ital. Attuari* 6 (1935), 320–334. · Halmos, P. R., Savage, L. J. Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* 20 (1949), 225–241.
- Rao, C. R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* 37 (1945), 81–91. · Blackwell, D. Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* 18 (1947), 105–110.
- Lehmann, E. L., Scheffé, H. Completeness, similar regions, and unbiased estimation. *Sankhyā* 10 (1950), 305–340.
- Koopman, B. O. On distributions admitting a sufficient statistic. *Trans. AMS* 39 (1936), 399–409. ·

Pitman, E. J. G. Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.* 32 (1936), 567–579. · Darmois, G. Sur les lois de probabilité à estimation exhaustive. *C. R. Acad. Sci. Paris* 200 (1935), 1265–1266.

- Basu, D. On statistics independent of a complete sufficient statistic. *Sankhyā* 15 (1955), 377–380.
- Tukey, J. W. A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (Olkin 외 편), Stanford, 1960, 448–485.
- Stigler, S. M. Do robust estimators deal with real data? *Ann. Statist.* 5 (1977), 1055–1098. · 자료: `MASS::newcomb`.
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016.

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

$\sum x_i = t$이면 $P(\mathbf{X}=\mathbf{x} \mid T=t) = \dfrac{\prod e^{-\lambda}\lambda^{x_i}/x_i!}{e^{-n\lambda}(n\lambda)^t/t!} = \dfrac{t!}{\prod x_i!}\left(\dfrac{1}{n}\right)^t$ (4회 정리 2.4로 $T \sim \mathrm{Poi}(n\lambda)$). 이것이 $\mathrm{M}(t; \frac{1}{n}, \dots, \frac{1}{n})$의 확률질량함수다.

연습 1.2 문제 7쪽

(a) $\prod \frac{x_i^{a-1}(1-x_i)^{b-1}}{B(a,b)}$ 이므로 $(\sum \log X_i,\ \sum \log(1-X_i))$. (b) $p^n(1-p)^{\sum x_i - n}$ 이므로 $\sum X_i$. (c) $\theta^{-r}e^{-\sum y_i/\theta}$ 이므로 (절단되지 않은 수 r, $\sum y_i$)의 쌍.

연습 1.3 문제 7쪽

$f(\mathbf{x};\theta) = 1\{x_{(n)} - 1 \le \theta \le x_{(1)}\}$이다. 비 $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$가 θ에 무관하려면(0/0을 제외한 뜻에서) 두 구간 $[x_{(n)}-1, x_{(1)}]$과 $[y_{(n)}-1, y_{(1)}]$이 같아야 하고, 그것은 $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$과 같다. 정리 1.6로 최소충분이다.

연습 1.4 문제 7쪽

비가 상수 c이면 $\prod(1+(y_i-\theta)^2) = c\prod(1+(x_i-\theta)^2)$가 θ의 $2n$차 다항식으로서 같다. 최고차 계수가 둘 다 1이므로 $c=1$이고, 복소근의 모임 $\{x_i \pm \mathrm{i}\}$와 $\{y_i \pm \mathrm{i}\}$가 같다. 따라서 $\{x_i\}$와 $\{y_i\}$는 순서만 다르다.

연습 1.5 문제 7쪽

$T \sim \mathrm{Bernoulli}(p(\theta))$, $p = \Phi(\theta)$이므로 $I_T = \dfrac{p'(\theta)^2}{p(1-p)} = \dfrac{\varphi(\theta)^2}{\Phi(\theta)(1-\Phi(\theta))}$ (9회 연습 2.2의 변환 규칙). $\theta = 0$에서 $\dfrac{1/(2\pi)}{1/4} = \dfrac{2}{\pi}$.

연습 1.6 문제 7쪽

(a) 9회의 정칙조건은 합(적분)과 θ-미분의 순서를 바꿀 수 있게 한다 [T6]. $P_\theta(T=t)>0$이면

$$s_T(t)=\frac{\partial_\theta P_\theta(T=t)}{P_\theta(T=t)}=\sum_{\mathbf{x}:\,T(\mathbf{x})=t}\frac{\partial_\theta f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)}\cdot\frac{f(\mathbf{x};\theta)}{P_\theta(T=t)}=\sum_{\mathbf{x}:\,T(\mathbf{x})=t}s_{\mathbf{X}}(\mathbf{x};\theta)\,P_\theta(\mathbf{X}=\mathbf{x}\mid T=t),$$

곧 $s_T(t)=E_\theta[s_{\mathbf{X}}(\mathbf{X})\mid T=t]$이다. (b) 반복기댓값으로 $E_\theta s_T(T)=E_\theta s_{\mathbf{X}}=0$이다(9회 정리 2.2). 따라서 $I_T=E_\theta s_T(T)^2=\operatorname{Var}_\theta s_T(T)$이다. 전분산 공식으로 $I_{\mathbf{X}}=\operatorname{Var}s_{\mathbf{X}}=\operatorname{Var}E[s_{\mathbf{X}}\mid T]+E\operatorname{Var}(s_{\mathbf{X}}\mid T)\ge\operatorname{Var}s_T(T)=I_T$이다. (c) 인수분해 정리로 $\log f(\mathbf{x};\theta)=\log g(T(\mathbf{x});\theta)+\log h(\mathbf{x})$이고, h는 θ에 의존하지 않으므로 $s_{\mathbf{X}}=\partial_\theta\log g(T(\mathbf{x});\theta)$는 T의 함수다. 그러면 $E[s_{\mathbf{X}}\mid T]=s_{\mathbf{X}}$이고 $\operatorname{Var}(s_{\mathbf{X}}\mid T)=0$이므로 (b)의 부등식이 등호다.

2절

연습 2.1 문제 10쪽

$E(1-\frac{1}{n})^T=e^{n\lambda(-1/n)}=e^{-\lambda}$, $E(1-\frac{1}{n})^{2T}=e^{n\lambda((1-1/n)^2-1)}=e^{-2\lambda+\lambda/n}$이므로 분산은 $e^{-2\lambda}(e^{\lambda/n}-1)$. 한계는 $k'(\lambda)^2/(n/\lambda)=\lambda e^{-2\lambda}/n$이고 비는 $\frac{e^{\lambda/n}-1}{\lambda/n}>1$, $n\to\infty$에서 1로 간다.

연습 2.2 문제 10쪽

최댓값이 m인 표본(크기 k인 부분집합)은 m과 $\{1,\dots,m-1\}$의 크기 $k-1$인 부분집합으로 일대일 대응하고, 각 표본의 확률은 $1/\binom{N}{k}$로 같다. 그러므로 $M=m$이 주어지면 나머지는 $\binom{m-1}{k-1}$개의 부분집합 가운데 균등하다. 대칭으로 각 원소의 주변분포는 $\{1,\dots,m-1\}$ 위의 균등분포이고 평균은 $m/2$이다.

연습 2.3 문제 10쪽

$(X_1-\bar X,\bar X)$는 결합정규이고 $\operatorname{Cov}(X_1-\bar X,\bar X)=\frac{1}{n}-\frac{1}{n}=0$이므로 독립이다(5회 정리 B). $X_1-\bar X\sim\mathcal{N}(0,\frac{n-1}{n})$이므로 $P(X_1\le c\mid\bar X)=P(X_1-\bar X\le c-\bar X\mid\bar X)=\Phi\big(\frac{c-\bar X}{\sqrt{(n-1)/n}}\big)$.

연습 2.4 문제 10쪽

조건부 Jensen으로 $L(\theta,E(W\mid T))\le E[L(\theta,W)\mid T]$이고, 양변의 기댓값을 취한다.

3절

연습 3.1 문제 13쪽

$\sum_t g(t)\mu^t/t! = 0$ (모든 $\mu > 0$)이면 이 멱급수가 0이므로 계수 $g(t)/t!$가 모두 0이다. $\mu \in \{1, 2\}$이면 g를 $t = 0, 1, 2$에서만 0이 아니게 두어 $g(0) + g(1) + \frac{1}{2}g(2) = 0$, $g(0) + 2g(1) + 2g(2) = 0$을 풀면 $g = (1, -\frac{3}{2}, 1)$이 나온다. 0이 아닌데 평균이 두 점에서 모두 0이다.

연습 3.2 문제 13쪽

$EX_{(n)}^2 = \int_0^\theta t^2 \frac{nt^{n-1}}{\theta^n} dt = \frac{n}{n+2}\theta^2$이므로 $\frac{n+2}{n}X_{(n)}^2$은 완비충분통계량의 불편 함수, 곧 UMVUE다.

연습 3.3 문제 13쪽

$\sum_{m=k}^N g(m)\binom{m-1}{k-1} = 0$이 모든 $N \geq k$에서 성립한다. $N = k$이면 $g(k) = 0$. $N = k+1$이면 $g(k) \cdot 1 + g(k+1)\binom{k}{k-1} = 0$에서 $g(k+1) = 0$. 이렇게 차례로 $g \equiv 0$이다. $\hat{N} = \frac{k+1}{k}M - 1$은 불편이므로(8회 정리 1.6) UMVUE다.

연습 3.4 문제 13쪽

$E[T(n-T)] = nET - ET^2 = n^2 p - \big(np(1-p) + n^2p^2\big) = n(n-1)p(1-p)$이므로 $\frac{T(n-T)}{n(n-1)}$은 불편이고 완비충분통계량 T의 함수다.

4절

연습 4.1 문제 18쪽

$\partial_{\eta_1} A = \psi(\eta_1 + 1) - \log(-\eta_2) = \psi(\alpha) + \log\beta = E\log X$, $\partial_{\eta_2} A = -\frac{\eta_1+1}{\eta_2} = \alpha\beta = EX$. (α, β)로 바꾸면 야코비 $\mathrm{diag}(1, 1/\beta^2)$를 양쪽에 곱해 9회 연습 3.3의 행렬 $\left(\begin{smallmatrix} \psi'(\alpha) & 1/\beta \\ 1/\beta & \alpha/\beta^2 \end{smallmatrix}\right)$가 나온다(9회 연습 2.2의 다변수판).

연습 4.2 문제 18쪽

$A(\eta) = e^\eta$의 k차 미분은 모두 $e^\eta = \lambda$이고, 누적생성함수 $A(\eta + s) - A(\eta)$의 $s = 0$에서의 k차 미분이 k차 누적량이다.

연습 4.3 문제 18쪽

등호이면 점수 $\frac{n}{\lambda}(\bar{x}-\lambda)=a(\lambda)(W-\tau(\lambda))$가 모든 λ에서 성립하므로 $W=\tau(\lambda)+\frac{n}{\lambda a(\lambda)}(\bar{x}-\lambda)$이다. 왼쪽은 λ에 무관하므로 $\bar{x}$의 계수 $c=\frac{n}{\lambda a(\lambda)}$가 상수이고, $\tau(\lambda)=c\lambda+(W-c\bar{x})$에서 $W-c\bar{x}$도 상수다. 곧 $\tau(\lambda)=a\lambda+b$.

연습 4.4 문제 18쪽

$\partial_k A=\frac{e^{\eta_k}}{\sum_j e^{\eta_j}}=p_k$, $\partial_l\partial_k A=p_k(\delta_{kl}-p_l)$. $(\operatorname{diag}\mathbf{p}-\mathbf{p}\mathbf{p}')\mathbf{1}=\mathbf{p}-\mathbf{p}=\mathbf{0}$이므로 가역이 아니다. 모든 logit에 같은 수를 더해도 분포가 같아서($\boldsymbol{\eta}$가 식별되지 않아서) 그 방향의 정보가 0이다.

연습 4.5 문제 18쪽

(a) $\nabla_\eta \log f=\mathbf{T}(x)-\nabla A(\boldsymbol{\eta})$이고 $\nabla^2_\eta \log f=-\nabla^2 A(\boldsymbol{\eta})$이다. 이것은 x를 포함하지 않으므로 9회 정리 2.2의 둘째 표현 $I=-E\nabla^2\log f$는 $\nabla^2A(\boldsymbol{\eta})$ 자신이다. 정리 4.3로 이것은 $\operatorname{Cov}\mathbf{T}$이다. (b) $\ell(\boldsymbol{\eta})=\sum_i \log h(x_i)+\boldsymbol{\eta}'\sum_i\mathbf{T}(x_i)-nA(\boldsymbol{\eta})$의 헤세 행렬은 $-n\nabla^2A(\boldsymbol{\eta})$이고, 정리 4.3로 ∇^2A는 양의 준정부호이므로 ℓ은 오목이다. (c) $\nabla\ell(\boldsymbol{\eta})=\sum_i\mathbf{T}(x_i)-n\nabla A(\boldsymbol{\eta})=\mathbf{0}$을 n으로 나누면 가능도 방정식이고, 정리 4.3로 $\nabla A(\hat{\boldsymbol{\eta}})=E_{\hat\eta}\mathbf{T}$이다. 오목함수는 기울기가 0인 점에서 최대가 되므로 해는 최대점이다. 열린 정의역에서 오목함수의 정류점은 전역 최대점이기 때문이다 [T4].

연습 4.6 문제 18쪽

(a) $E_\theta s=0$이므로 $\operatorname{Cov}_\theta(W,s)=E_\theta\big[(W-\tau(\theta))\,s\big]=\tau'(\theta)$이다. Cauchy–Schwarz의 등호는 두 중심화된 변수가 일차종속일 때, 곧 $s-Es=s$가 $W-\tau(\theta)$의 상수배일 때다(3회 정리 4.4의 등호 조건). $\tau'\neq0$이므로 s는 0이 아니고 상수 $a(\theta)$는 0이 아니다. (b) $\partial_\theta\log\prod f(x_i;\theta)=a(\theta)W(\mathbf{x})-a(\theta)\tau(\theta)$를 θ에 대해 적분한다. $c(\theta)=\int a$, $d(\theta)=\int a\tau$로 두면 $\log\prod f(x_i;\theta)=c(\theta)W(\mathbf{x})-d(\theta)+K(\mathbf{x})$이다. 적분은 $\mathbf{x}$를 고정하고 θ에 대해 했으므로 적분 상수 K는 $\mathbf{x}$에만 의존하고, $H(\mathbf{x})=e^{K(\mathbf{x})}$이다. (c) $s=\sum_i(x_i/\lambda-1)=\frac{n}{\lambda}(\bar{x}-\lambda)$이므로 $a(\lambda)=\frac{n}{\lambda}$, $\tau(\lambda)=\lambda$이다. $c(\lambda)=n\log\lambda$, $d(\lambda)=n\lambda$이고 $\prod f=\exp(n\bar{x}\log\lambda-n\lambda)\cdot\frac{1}{\prod x_i!}$이므로 $H(\mathbf{x})=1/\prod x_i!$이다.

5절

연습 5.1 문제 21쪽

$X_{(1)}/X_{(n)}$은 척도를 바꿔도 같으므로 보조이고 $X_{(n)}$은 완비충분이다. Basu로 독립이므로 $EX_{(1)}=E\big[\frac{X_{(1)}}{X_{(n)}}X_{(n)}\big]=E\frac{X_{(1)}}{X_{(n)}}\,EX_{(n)}$이고 $\frac{\theta/(n+1)}{n\theta/(n+1)}=\frac{1}{n}$.

연습 5.2 문제 21쪽

$\sum X_i$는 1-모수 지수족의 완비충분통계량이고 $X_{(1)}/\sum X_i$는 척도 불변이므로 보조다. 독립이므로 $E\left[\frac{X_{(1)}}{\sum X_i}\right] = \frac{EX_{(1)}}{E\sum X_i} = \frac{\theta/n}{n\theta} = \frac{1}{n^2}$.

연습 5.3 문제 21쪽

$(\bar{X}, S)$는 2-모수 정규 지수족의 완비충분통계량이다(정리 4.5). 표준화된 잔차는 $X_i \mapsto a + bX_i$ $(b>0)$에 대해 변하지 않으므로 그 분포가 (μ,σ)에 무관하다. Basu로 독립이다.

6절

연습 6.1 문제 25쪽

$E|Z| = 2\int_0^\infty z\varphi(z)dz = 2\varphi(0) = \sqrt{2/\pi}$, $\operatorname{Var}|Z| = EZ^2 - (E|Z|)^2 = 1 - \frac{2}{\pi}$. 비는 $\frac{1-2/\pi}{2/\pi} = \frac{\pi}{2} - 1$.

연습 6.2 문제 25쪽

명제 7.1의 식으로 $\frac{1}{4}(6-1) = 1.25$와 $\frac{2b^2}{b^2} - 1 = 1$이므로 효율은 1.25다. Laplace 오차에서는 평균편차가 표준편차보다 낫다. 꼬리가 두꺼우면 제곱이 큰 오차에 너무 큰 무게를 준다.

연습 6.3 문제 25쪽

66개의 제곱합 $\sum(x_i - \bar{x})^2 = 65 \cdot 10.75^2 \approx 7505$ 가운데 -44와 -2 두 값이 약 5,700(76%)을 차지한다. 부트스트랩 표본마다 이 두 값이 뽑히는 횟수가 $0, 1, 2, \cdots$로 달라지므로 s가 크게 흔들린다. 평균편차는 거리를 제곱하지 않아 덜 흔들린다. 두 값을 빼면 자료가 정규에 가까워 0.78(정규 이론 0.876)이 된다.

연습 6.4 문제 25쪽

(a) $|X_i - \mu|$들은 iid이고 평균 $\sigma\sqrt{2/\pi}$, 분산 $\sigma^2(1-\frac{2}{\pi})$이다(연습 6.1에 $Z = (X-\mu)/\sigma$). 중심극한정리로 $\sqrt{n}(\tilde{d} - \sigma\sqrt{2/\pi}) \to \mathcal{N}\big(0, \sigma^2(1-\frac{2}{\pi})\big)$이다. (b) $g(u) = \log u$, $g'(u) = 1/u$이므로 델타 방법으로 점근분산은 $\frac{\sigma^2(1-2/\pi)}{\sigma^2 \cdot 2/\pi} = \frac{\pi}{2} - 1$이다. 척도 σ가 약분되므로 log를 취해 비교하는 것이 공정하다. (c) $\sqrt{n}(S^2 - \sigma^2) \to \mathcal{N}(0, 2\sigma^4)$이고 $g(u) = \frac{1}{2}\log u$, $g'(\sigma^2) = \frac{1}{2\sigma^2}$이므로 점근분산은 $\frac{1}{4\sigma^4} \cdot 2\sigma^4 = \frac{1}{2}$이다. (d) 상대효율은 $\frac{1/2}{\pi/2 - 1} = \frac{1}{\pi - 2} = 0.876$이다. 점근분산이 v/n이면 같은 분산을 얻는 데 드는 관측 수는 v에 비례하므로, 평균편차는 표준편차보다 관측이 $\pi - 2 = 1.14$배 필요하다.

7절

연습 7.1 문제 28쪽

혼합의 적률은 성분 적률의 가중평균이다: $\mu_2 = 1 + 8\varepsilon$, $\mu_4 = 3 + 240\varepsilon$, $\nu_1 = \sqrt{2/\pi}(1 + 2\varepsilon)$. 이 식에서 효율 $= 1$을 풀면 $\varepsilon = 0.0018$이다. ε이 크면 큰 오차가 더 이상 드물지 않아 분포가 다시 "넓은 정규"에 가까워지므로, $\varepsilon \to 1$에서 효율이 0.876으로 돌아간다. 그 사이 $\varepsilon \approx 0.05$ 근처에서 가장 크다.

연습 7.2 문제 28쪽

$T \sim \mathrm{Poi}(n\lambda)$이므로 $E(1 - \frac{2}{n})^T = e^{n\lambda(-2/n)} = e^{-2\lambda}$이고, 완비충분통계량의 함수이므로 UMVUE다. $n = 1$이면 $(-1)^X$, $n = 2$이면 $0^T = 1\{T = 0\}$(두 관측이 모두 0이면 1)이다.

연습 7.3 문제 28쪽

$E_p g(K)$는 p의 n차 이하 다항식이므로 $(0, 1)$에서 유계다. $1/p$는 $p \to 0$에서 무한대로 가므로 어떤 다항식과도 같을 수 없다.

연습 7.4 문제 28쪽

(a) 대칭이면 $EX = 0$이고, X_i^2들은 iid로 평균 μ_2, 분산 $\mu_4 - \mu_2^2$이다. 중심극한정리로 $\sqrt{n}(\frac{1}{n}\sum X_i^2 - \mu_2) \to \mathcal{N}(0, \mu_4 - \mu_2^2)$이고, $g(u) = \frac{1}{2}\log u$로 델타 방법을 쓰면 점근분산은 $\frac{1}{4\mu_2^2}(\mu_4 - \mu_2^2) = \frac{1}{4}(\frac{\mu_4}{\mu_2^2} - 1)$이다. (b) $|X_i|$들은 평균 ν_1, 분산 $\mu_2 - \nu_1^2$이므로 $g(u) = \log u$로 점근분산은 $\frac{\mu_2 - \nu_1^2}{\nu_1^2} = \frac{\mu_2}{\nu_1^2} - 1$이다. (c) 혼합의 기댓값은 성분 기댓값의 가중평균이다. $\mathcal{N}(0, 1)$과 $\mathcal{N}(0, 9)$에서 EX^2은 1과 9, EX^4는 3과 $3 \cdot 81$, $E|X|$는 $\sqrt{2/\pi}$와 $3\sqrt{2/\pi}$이므로

$$\mu_2 = 1 + 8\varepsilon, \qquad \mu_4 = 3(1 - \varepsilon) + 243\varepsilon = 3(1 + 80\varepsilon), \qquad \nu_1 = \sqrt{2/\pi}\,(1 + 2\varepsilon).$$

(d) $\varepsilon = 0$이면 (a)는 $\frac{1}{4}(3 - 1) = \frac{1}{2}$, (b)는 $\frac{1}{2/\pi} - 1 = \frac{\pi}{2} - 1$이므로 비는 $\frac{1}{\pi - 2} = 0.876$이다.

8절

연습 8.1 문제 30쪽

$E[1\{Z_1 + Z_2 > 3\} \mid Z_1] = P(Z_2 > 3 - Z_1) = 1 - \Phi(3 - Z_1)$이므로 반복기댓값으로 평균이 같고, 전분산 공식으로 분산이 작다. 노트북에서 분산의 비는 약 7이다.

연습 8.2 문제 30쪽

$\mathbf{p} = \text{softmax}(1, 0, -1) = (0.665,\ 0.245,\ 0.090)$이고 기울기는 $\mathbf{p} - \mathbf{e}_2 = (0.665,\ -0.755,\ 0.090)$이다. $\sum p_k = 1 = \sum(\mathbf{e}_2)_k$이므로 성분의 합은 0이다. 모든 logit에 같은 수를 더하는 방향으로는 손실이 변하지 않는다(연습 4.4).