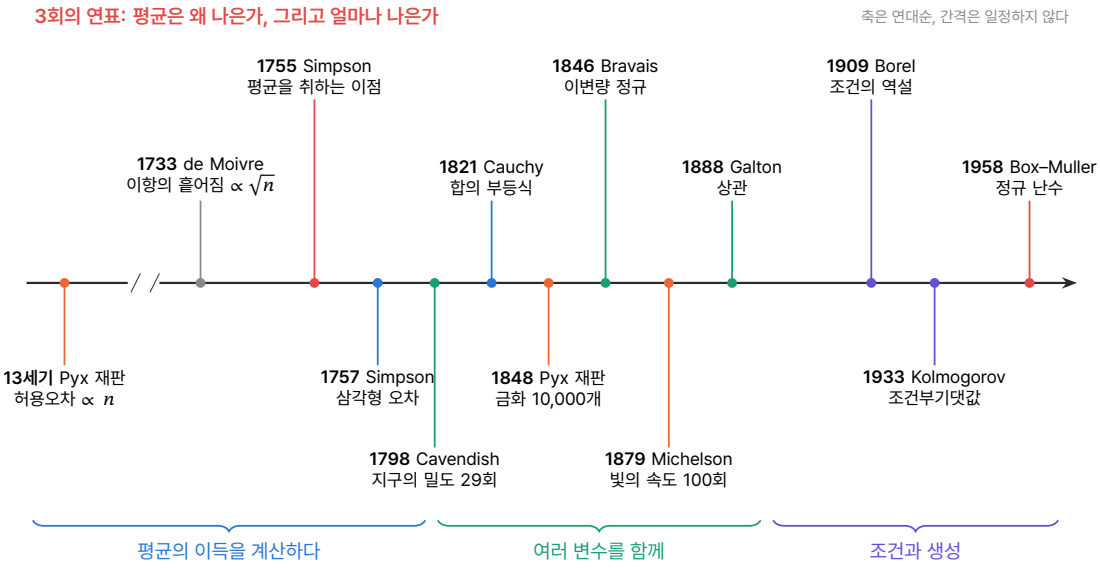
수리통계학 × 통계학의 역사

3회 강의노트

# 왜 평균을 내는가 여러 확률변수와 $\sqrt{n}$

## 1755 → 1958

1755년 Thomas Simpson이 여러 관측을 평균하는 이득을 처음 계산했고, 런던 조폐국은 수백 년 동안 그 이득의 크기를 잘못 알았다. 평균의 오차는 $n$이 아니라 $\sqrt{n}$으로 준다. 이 노트는 결합분포·조건부분포·독립·공분산·야코비안을 세우고, 그것으로 $\sqrt{n}$의 법칙과 그 한계를 증명한다.




**다루는 것** 결합·조건부 분포 · 전분산 공식 · 독립 · 공분산 · 야코비안 · 합성곱 · $\sqrt{n}$ 법칙 · 유효 표본 크기

**원자료** Pyx 재판(1848)의 금화 10,000개 · Michelson(1879) · Cavendish(1798)

**노트북** `sessions/03_averaging/03_averaging.ipynb` — 절·정리 번호가 같다

**워크북** `03_averaging_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`

**선수 지식** 1–2회 전체, 다변수 미적분 (이중적분, 편미분, 행렬식)

## 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 도구)를 단다. 적분의 순서를 바꾸거나 극한과 기댓값을 바꾸는 곳에는 **조건 확인** 단락이 붙는다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 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] 부등식, [T6] 적분 순서 바꾸기(Tonelli·Fubini), [T7] 변수변환과 야코비안을 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 1–2회에서 가져오는 것 2

0.2 행렬과 벡터의 표기 3

**1 결합분포 — 두 확률변수를 함께 4**

1.1 결합분포함수 4

1.2 이산 결합분포 4

1.3 연속 결합분포 5

1.4 함수의 기댓값과 선형성 7

**2 조건부분포와 조건부기댓값 10**

2.1 조건부분포 10

2.2 조건부기댓값 11

2.3 가장 좋은 예측 13

**3 독립 16**

3.1 정의와 판정 16

3.2 독립인 변수의 함수, 곱, 합 17

**4 공분산과 상관 20**

4.1 공분산 20

4.2 Cauchy–Schwarz와 상관의 한계 21

4.3 가장 좋은 직선과 상관의 함정 22

**5 여러 변수의 변수변환 — 야코비안 25**

5.1 변수변환 공식 25

5.2 극좌표 — 가우스 적분과 Box–Muller 26

5.3 합성곱 — 독립합의 밀도 28
5.4 비, 그리고 다시 Cauchy 30
5.5 선형변환 30

**6 평균의 정확도 — $\sqrt{n}$의 법칙 32**

6.1 평균벡터와 공분산행렬 32
6.2 $\sqrt{n}$의 법칙 33
6.3 Simpson의 계산 (1755, 1757) 33
6.4 Pyx 재판 — $n$과 $\sqrt{n}$ 34
6.5 상관된 관측 — 유효 표본 크기 36

**7 부숴보기 — 평균이 통하지 않을 때 38**

7.1 확률 0인 사건에 조건을 걸면 — Borel의 역설 38
7.2 상관이 0이어도 40
7.3 평균으로 줄지 않는 오차 — Michelson과 Cavendish 40
7.4 평균을 내도 좁아지지 않는다 — Cauchy 40

**8 Beautiful insights — 같은 계산, 오늘의 문제 42**

8.1 신경망의 초기화 — 분산을 지키기 42
8.2 재매개화 — Box–Muller의 생각 42
8.3 앙상블 — 상관을 줄이는 것이 이득이다 43
8.4 조건부기댓값으로서의 학습 43

**9 정리와 역사 노트 45**

9.1 한 장 요약 45
9.2 Stigler의 명명 법칙 46
9.3 7 Pillars 지도 46
9.4 다음 회 — 셈하는 분포와 섞인 분포 (1711 → 1898) 46

**A 연습문제 풀이 48**

# 0

# 들어가며

> **여는 질문.** 천문학자가 같은 별의 위치를 여섯 번 쟀다. 여섯 값의 평균을 쓰는 것이 가장 조심스럽게 잰 한 번의 값을 쓰는 것보다 정말 나은가? 낫다면 얼마나 나은가?

지금은 답이 자명해 보인다. 그러나 18세기 중반까지 천문학자들은 이 질문에 쉽게 답하지 못했다. 서로 다른 관측을 평균하면 좋은 관측이 나쁜 관측에 섞여 망가진다고 걱정하는 사람이 많았다. 1755년 Thomas Simpson은 왕립학회 회장에게 보낸 편지에서 처음으로 이것을 **계산했다.** 관측 오차에 확률분포를 주고, 여섯 오차의 합의 분포를 de Moivre의 생성함수로 구해, 평균이 참값 가까이 있을 확률이 한 번의 관측보다 훨씬 크다는 것을 보였다(6절).

얼마나 나은가에 대한 답은 $\sqrt{n}$이다. 독립인 관측 $n$개의 평균은 표준편차가 한 관측의 $1/\sqrt{n}$이다. 2회는 이것을 이산 확률변수에서 증명했다. 그런데 이 법칙이 늘 알려져 있었던 것은 아니다. 런던 조폐국은 13세기부터 금화의 무게를 공개적으로 검사했는데(Pyx 재판), 동전 $n$개의 총무게에 대한 허용오차를 $\sqrt{n}$이 아니라 $n$에 비례하게 잡았다. 그 틈은 수백 년 동안 열려 있었다(6.3절).

$\sqrt{n}$의 법칙을 연속 확률변수까지, 그리고 서로 닮은 관측까지 다루려면 여러 확률변수를 **함께** 다루는 언어가 필요하다. 이 노트는 그 언어를 세운다. 결합분포와 주변분포(1절), 조건부분포와 조건부기댓값(2절), 독립(3절), 공분산과 상관(4절), 여러 변수의 변수변환(5절)이다. 6절에서 그 언어로 $\sqrt{n}$의 법칙과 그 한계(상관된 관측, 유효 표본 크기)를 증명한다. 7절은 확률 0인 사건에 조건을 거는 역설과 평균으로 줄지 않는 오차를, 8절은 신경망 초기화·재매개화·앙상블에서 같은 계산을 본다.

## 학습 목표

1. 결합분포에서 주변분포와 조건부분포를 구하고, 함수의 기댓값을 이중합·이중적분으로 계산한다.
2. 반복기댓값과 전분산 공식을 증명하고, 조건부기댓값이 가장 좋은 예측임을 보인다.
3. 독립을 결합밀도의 인수분해로 판정하고, 2회의 독립합 정리들을 연속 확률변수로 넓힌다.
4. 공분산과 상관계수를 정의하고, Cauchy–Schwarz 부등식으로 $|\rho| \le 1$을 증명한다.
5. 야코비안 변수변환을 증명하고, 그것으로 가우스 적분, Box–Muller 방법, 합성곱, Cauchy 분포를 얻는다.

6. 공분산행렬로 $\mathrm{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a}$를 증명하고, $\sqrt{n}$의 법칙이 상관과 편향 앞에서 어떻게 무너지는지 보인다.

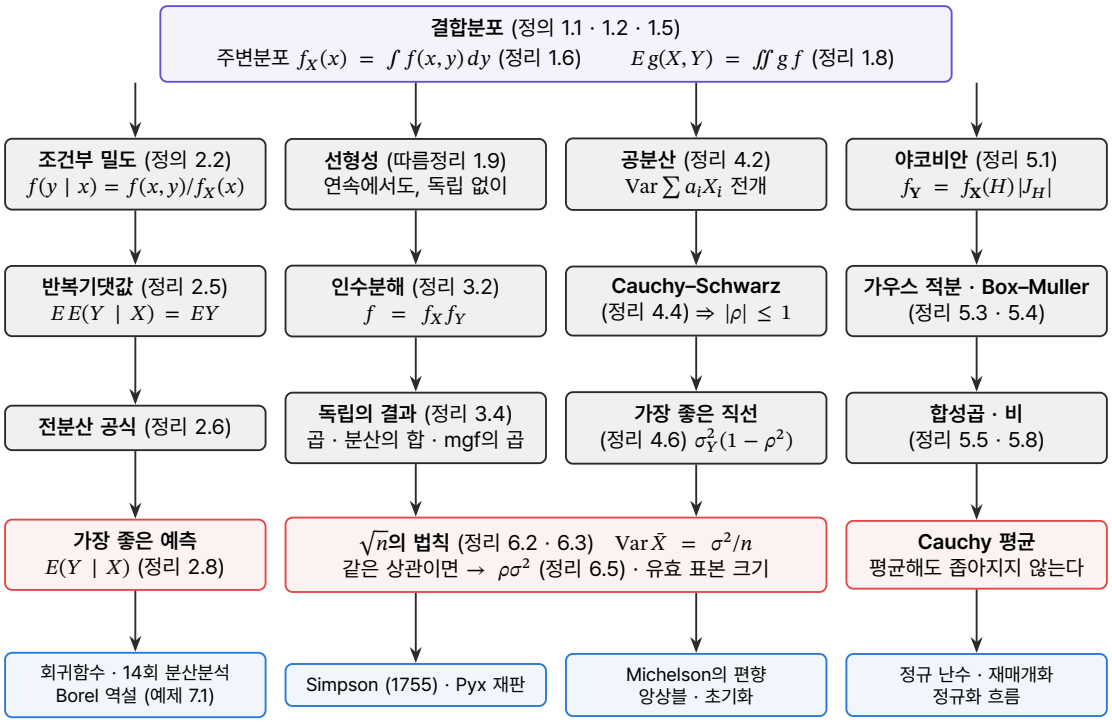



**그림 0.1** 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

## 0.1 1–2회에서 가져오는 것

| 앞 회차 | 내용 |
|---|---|
| 1회 정의 5.1 · 정리 5.4 | 조건부확률 $P(A \mid B) = P(AB)/P(B)$, 전확률 공식 |
| 1회 정의 5.7 · 5.9 | 사건의 독립, 상호 독립 |
| 1회 정의 6.4 · 6.6 | 확률질량함수, 확률밀도함수 |
| 2회 정리 1.9 · 1.11 | 함수의 기댓값, 이산 확률공간에서의 선형성 |
| 2회 1.4절·1.5절의 상자 | 연속형 함수의 기댓값의 계산 규칙, 음이 아닌 함수의 적분 순서 (Tonelli) |
| 2회 정리 2.2 · 2.7 · 2.8 | 분산의 성질, 독립인 이산 변수의 $E(XY) = EX\,EY$, 독립합의 분산 |
| 2회 정리 3.3 · 3.6 · 상자 | mgf와 적률, 아핀 변환, 독립합의 mgf(이산), mgf의 유일성 |
| 2회 정리 4.3 | 단조변환의 밀도 $f_Y(y) = f_X(h(y))\,\lvert h'(y)\rvert$ |
| 2회 정리 5.3 · 6.3 | Chebyshev 부등식, 약한 대수의 법칙 |

2회에는 "이산에서 증명하고, 연속은 3회"라는 꼬리표가 붙은 정리가 셋 있었다. 선형성 $E(X+Y)=EX+EY$(2회 정리 1.11), 독립합의 분산(2회 정리 2.8), 독립합의 mgf(2회 정리 3.6(b))다. 이 노트에서 차례로 그 꼬리표를 뗀다(따름정리 1.9, 정리 3.4).

## 0.2 행렬과 벡터의 표기

6절에서 확률벡터를 쓴다. 벡터는 굵은 글자 $\mathbf{x}=(x_1,\dots,x_n)'$로 쓰고 열벡터로 본다. $'$는 전치다. 두 벡터의 내적은 $\mathbf{a}'\mathbf{x}=\sum_i a_i x_i$이다. 정사각행렬 $A$의 행렬식은 $\det A$, 역행렬은 $A^{-1}$이다. 대칭행렬 $\Sigma$가 모든 $\mathbf{a}$에 대해 $\mathbf{a}'\Sigma\mathbf{a}\geq 0$이면 **양의 준정부호**라 하고 $\Sigma\succeq 0$으로 쓴다.

# 1

# 결합분포 — 두 확률변수를 함께

한 실험에서 두 개 이상의 수를 잰다. 주사위 두 개의 합과 최댓값, 관측 오차와 관측 시각, 키와 몸무게. 각각의 분포만으로는 두 수가 **함께** 어떻게 움직이는지 알 수 없다. 이 절은 두 확률변수의 결합분포를 정의하고, 거기서 각자의 분포(주변분포)를 되찾는 방법과 함수의 기댓값을 구하는 방법을 세운다.

## 1.1 결합분포함수

**정의 1.1** (확률벡터와 결합분포함수) 같은 확률공간 위의 두 확률변수 $X, Y$의 쌍 $(X, Y)$를 **확률벡터**라 하고,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y), \qquad (x, y) \in \mathbb{R}^2$$

를 **결합분포함수**라 한다. 사건 $\{X \le x,\ Y \le y\}$는 $\{X \le x\} \cap \{Y \le y\}$이다.

결합분포함수에서 사각형의 확률이 나온다. 포함배제(1회 정리 4.5)를 두 사건에 쓰면

$$P(a < X \le b,\ c < Y \le d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

이다(연습 1.1). 그리고 $y \to \infty$로 보내면 $\{X \le x,\ Y \le y\} \uparrow \{X \le x\}$이므로 연속성(1회 정리 4.7)으로 $F_{X,Y}(x, y) \to F_X(x)$이다. 한 변수의 분포는 결합분포에 들어 있다.

## 1.2 이산 결합분포

**정의 1.2** (결합 확률질량함수) $X$와 $Y$가 모두 이산이면 $p_{X,Y}(x, y) = P(X = x,\ Y = y)$를 **결합 확률질량함수**라 한다. 모든 $B \subset \mathbb{R}^2$에 대해 $P\big((X, Y) \in B\big) = \sum_{(x,y) \in B} p_{X,Y}(x, y)$이다.

**정리 1.3** (주변 확률질량함수) $p_X(x) = \sum_y p_{X,Y}(x,y)$, $p_Y(y) = \sum_x p_{X,Y}(x,y)$.

**증명.** $\{X = x\}$는 서로소인 사건 $\{X = x,\ Y = y\}$, $y$는 $Y$의 값들의 합집합이다. (P3)을 쓴다. ∎

이름은 표의 **가장자리**(margin)에서 왔다. 결합 확률을 표로 쓰면 행의 합과 열의 합이 표의 가장자리에 적힌다.

**예제 1.4** (두 주사위의 합과 최댓값) 공평한 주사위 두 개를 던져 합 $S$와 최댓값 $M$을 본다. 36가지 결과가 같은 확률이므로 $p_{S,M}(s,m)$은 $\{S = s,\ M = m\}$인 결과의 수를 36으로 나눈 것이다. 예를 들어 $M = m$이면서 $S = s$인 결과는 "하나가 $m$, 다른 하나가 $s - m$"이고 $s - m \le m$이어야 한다. $s - m < m$이면 순서 두 가지, $s - m = m$이면 한 가지이므로

$$p_{S,M}(s,m) = \begin{cases} \dfrac{2}{36}, & 1 \le s - m < m, \\ \dfrac{1}{36}, & s = 2m, \\ 0, & \text{그 밖}. \end{cases}$$

열마다 더하면 $p_M(m) = \frac{2m-1}{36}$이다. $M \le m$일 확률이 $(\frac{m}{6})^2$이므로 $p_M(m) = \frac{m^2-(m-1)^2}{36}$이어서 같다. 그림 1.1에서 두 변수는 함께 움직인다. 최댓값이 크면 합도 크다. 4절에서 이 "함께 움직임"을 공분산으로 잰다.

## 1.3 연속 결합분포

**정의 1.5** (결합 확률밀도함수) 음이 아니고 적분 가능한 함수 $f(x,y)$가 있어 모든 $(x,y)$에서

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)\,ds\,dt$$

이면 $(X,Y)$를 **연속**이라 하고 $f = f_{X,Y}$를 **결합 확률밀도함수**라 한다. 이때 평면의 "보통의" 집합 $B$에 대해 $P\big((X,Y)\in B\big) = \iint_B f(x,y)\,dx\,dy$이다.

마지막 문장은 사각형에서 출발해 1회 6.2절의 상자와 같은 확장으로 얻는다. 작은 사각형으로 읽으면 $P(x < X \le x + dx,\ y < Y \le y + dy) \approx f(x,y)\,dx\,dy$이다. 결합밀도는 평면 위의 확률의 "농도"다.

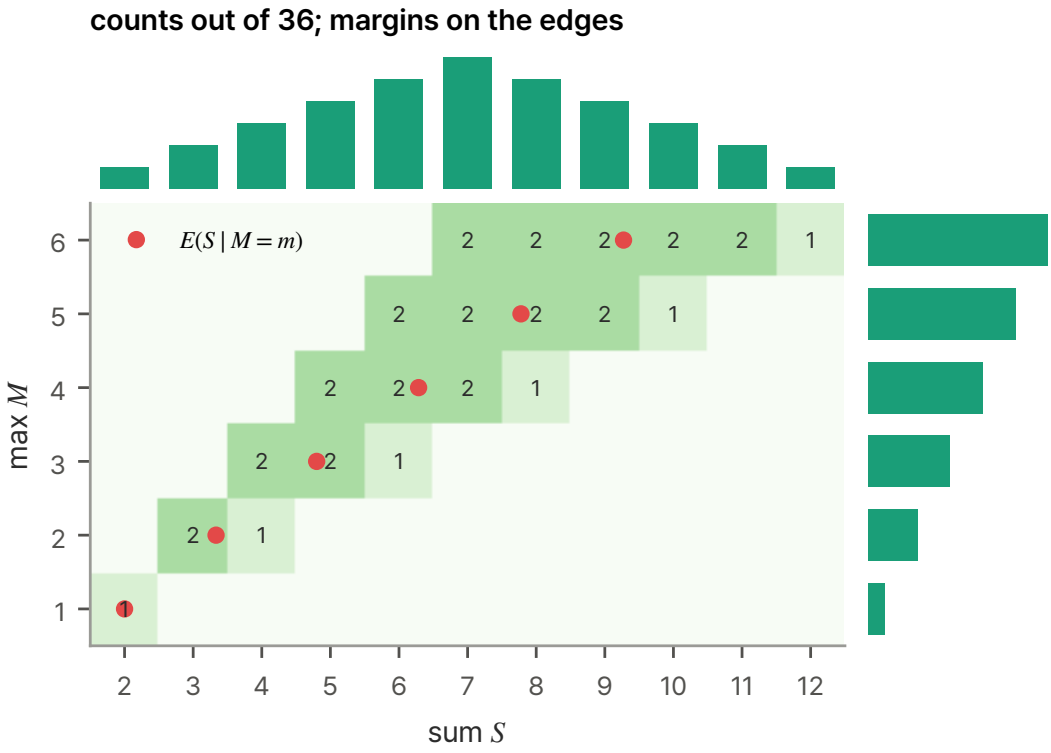



**그림 1.1** 두 주사위의 합 $S$와 최댓값 $M$의 결합 확률질량함수(색의 짙기)와 두 주변분포(가장자리 막대). 빨간 점은 $M=m$일 때 $S$의 조건부 평균이다(2절).

**정리 1.6** (주변 밀도) $(X,Y)$가 결합밀도 $f$를 가지면 $X$와 $Y$는 각각 밀도

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy, \qquad f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$$

를 갖는다.

**증명.** **전략** $F_X(x)$를 반평면 $\{s\le x\}$ 위의 이중적분으로 쓰고, 적분 순서를 바꿔 $\int_{-\infty}^{x}(\cdots)\,ds$의 꼴로 만든다. 음이 아닌 함수이므로 순서를 바꿀 수 있다 [T6]. 단계별 증명은 연습 1.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 49쪽에 있다.

**예제 1.7** (원판 위의 균등분포) 단위 원판 $D=\{x^2+y^2\le 1\}$ 위에 고르게 퍼진 점 $(X,Y)$의 결합밀도는 $f(x,y)=\frac{1}{\pi}$ $((x,y)\in D)$이다. 넓이가 $\pi$이기 때문이다. 1회 예제 6.8의 방법 (3)에서 현의 중점이 이렇게 퍼져 있었다. 주변 밀도는 $x$를 고정하고 $y$가 $-\sqrt{1-x^2}$에서 $\sqrt{1-x^2}$까지 움직이는 길이를 재면 된다.

$$f_X(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy=\frac{2}{\pi}\sqrt{1-x^2}, \qquad -1\le x\le 1.$$

결합분포는 균등한데 주변분포는 균등하지 않다. 가운데로 갈수록 세로 방향의 길이가 길기 때문이다(그림 1.2). 이 반원 모양의 밀도는 Wigner(1955)가 큰 무작위 행렬의 고윳값 분포로

다시 찾아낸 “반원 법칙” 의 모양이다.

## 1.4 함수의 기댓값과 선형성

**정리 1.8** (두 변수 함수의 기댓값)

(a) 이산이면, $\sum_{x,y} |g(x,y)|\,p(x,y) < \infty$ 일 때 $E\,g(X,Y) = \sum_{x,y} g(x,y)\,p(x,y)$.

(b) 연속이면, $\iint |g|\,f < \infty$ 일 때 $E\,g(X,Y) = \iint g(x,y)\,f(x,y)\,dx\,dy$.

**증명.** (a)는 $(X,Y)$의 값들을 하나의 셀 수 있는 집합으로 보고 2회 정리 1.9(a)의 증명을 그대로 쓴다. (b)는 2회 1.4절의 상자(연속형 함수의 기댓값)를 평면에서 쓴 것이다. 거기서처럼 이 과정에서는 (b)를 계산 규칙으로 받아들인다. ∎

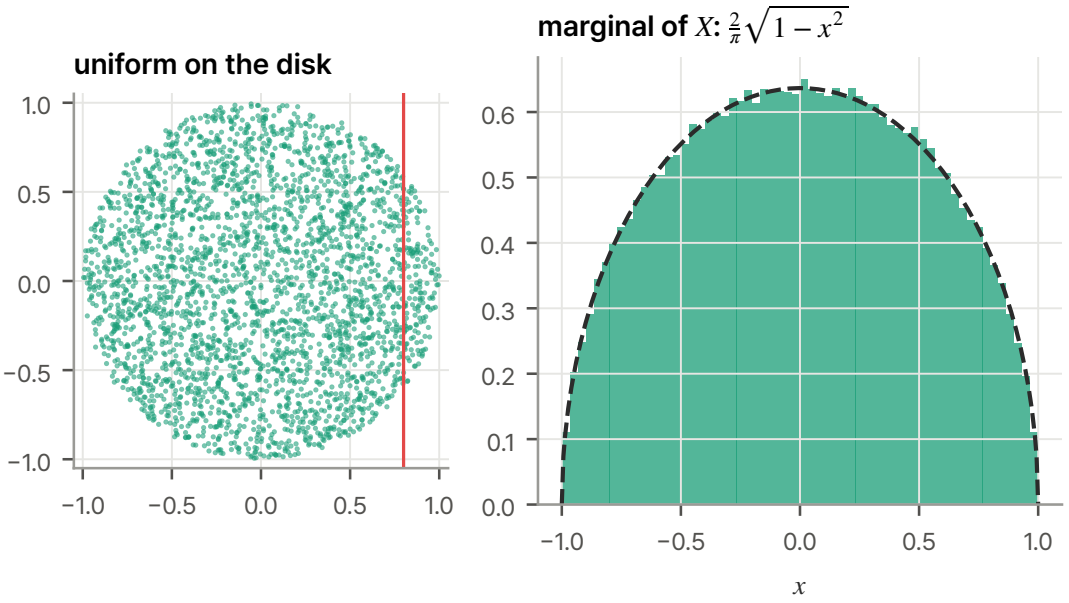



**그림 1.2** 원판 위의 균등한 점 3,000개와 $X$의 주변 밀도 $\frac{2}{\pi}\sqrt{1-x^2}$(점선). 균등한 결합분포의 주변분포는 균등하지 않다.

**따름정리 1.9** (선형성 — 일반적인 경우) $E|X|, E|Y| < \infty$이면 $E(aX+bY) = a\,EX + b\,EY$이다. 독립은 필요 없다.

**증명.** **전략** 정리 1.8(b)로 $E(aX+bY)$를 이중적분으로 쓰고, 두 항을 각각 주변 밀도의 적분으로 줄인다. 적분 순서를 바꾸기 전에 절댓값의 적분이 유한함을 확인한다 [T6]. 단계별 증명은 연습 1.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 49쪽에 있다.

이로써 2회 정리 1.11의 꼬리표가 떨어졌다. 2회 6절의 Monte Carlo 적분과 대수의 법칙에서 미리 쓴 "연속 확률변수의 선형성"이 이것이다.

### 연습문제

**연습 1.1** (사각형의 확률 ★) $P(a < X \le b,\ c < Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$임을 보여라. 이 값이 언제나 0 이상이라는 것이 결합분포함수의 넷째 성질이다. 풀이 48쪽

**연습 1.2** (삼각형 위의 밀도 ★) $f(x,y) = c$ $(0 < x < y < 1)$가 결합밀도가 되는 $c$를 구하고, 두 주변 밀도와 $P(X+Y<1)$을 구하라. 풀이 48쪽

**연습 1.3** (최댓값과 최솟값 ★) 예제 1.4에서 최솟값 $L$의 확률질량함수를 구하고, $EM + EL = 7$임을 선형성 없이, 그리고 선형성으로 확인하라. 풀이 48쪽

**연습 1.4** (원판의 거리 ★★) 예제 1.7의 점에서 원점까지의 거리 $R = \sqrt{X^2+Y^2}$의 분포함수가 $r^2$ $(0 \le r \le 1)$임을 보이고, $ER$과 $E|X|$를 구하라. 풀이 48쪽

**연습 1.5** (결합분포는 주변분포보다 많은 것을 담는다 ★) 주변분포가 모두 Bernoulli$(\tfrac12)$이지만 결합분포가 다른 $(X,Y)$를 두 가지 만들어라. 하나는 $P(X=Y)=1$, 다른 하나는 $P(X=Y)=0$이 되게. 풀이 48쪽

**연습 1.6** (주변 밀도 증명 ★★ 유도형) 정리 1.6를 다음 순서로 증명하라.

(a) 정의 1.5의 마지막 문장을 반평면 $B = \{(s,t) : s \le x\}$에 써서 $F_X(x) = \iint_B f(s,t)\,ds\,dt$임을 보여라.

(b) 적분 순서를 바꿔 $F_X(x) = \int_{-\infty}^{x}\big(\int_{-\infty}^{\infty} f(s,t)\,dt\big)ds$로 써라. 순서를 바꿀 수 있는 근거를 밝혀라.

(c) $g(s) = \int_{-\infty}^{\infty} f(s,t)\,dt$가 1회 정의 6.6의 조건을 만족해 $X$의 밀도임을 결론지어라. $f_Y$도 같은 방법으로 얻어라.

(d) 같은 결론을 결합분포함수의 극한 $F_X(x) = \lim_{n\to\infty} F_{X,Y}(x,n)$에서도 얻어라. 이 등식의 근거는 1회의 어느 정리인가?

풀이 49쪽

**연습 1.7** (선형성 증명 ★★ 유도형) 연속인 경우에 따름정리 1.9를 다음 순서로 증명하라.

(a) $\iint |x|\, f(x,y)\,dx\,dy = E|X|$임을 정리 1.6로 보이고, $\iint |ax+by|\, f < \infty$를 확인해 정리 1.8(b)를 $g(x,y)=ax+by$에 쓸 수 있음을 보여라.

(b) $\iint x\, f(x,y)\,dx\,dy = \int x\, f_X(x)\,dx$를 보여라. 어느 변수로 먼저 적분하는지, 무엇이 순서를 바꾸게 허락하는지 밝혀라.

(c) 결론을 쓰고, 이산인 경우를 같은 순서로 합에 대해 증명하라.

풀이 49쪽

# 2

# 조건부분포와 조건부기댓값

$X$의 값을 알면 $Y$에 대해 무엇을 말할 수 있나? 1회의 조건부확률을 확률변수로 넓힌 것이 조건부분포이고, 그 평균이 조건부기댓값이다. 이 절의 두 정리, 곧 반복기댓값과 전분산 공식은 이후 회차에서 가장 자주 쓰이는 도구 가운데 하나다. 5회의 Galton 회귀, 10회의 EM 알고리즘, 16회의 계층 베이즈 모형이 모두 이 두 줄 위에 선다.

## 2.1 조건부분포

**정의 2.1** (조건부 확률질량함수) 이산이고 $p_X(x) > 0$이면 $X = x$일 때 $Y$의 **조건부 확률질량함수**를

$$p_{Y\mid X}(y \mid x) = P(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

로 정의한다.

1회 정의 5.1을 사건 $\{X = x\}$와 $\{Y = y\}$에 쓴 것이다. $x$를 고정하면 $y \mapsto p(y \mid x)$는 음이 아니고 합이 1이므로 확률질량함수다(1회 정리 5.2). 연속이면 $P(X = x) = 0$이어서 이 정의를 쓸 수 없다. 대신 밀도의 비를 쓴다.

**정의 2.2** (조건부 밀도) 연속이고 $f_X(x) > 0$이면 $X = x$일 때 $Y$의 **조건부 밀도**를

$$f_{Y\mid X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

로 정의한다.

정리 1.6로 $\int f_{Y\mid X}(y \mid x)\,dy = f_X(x)/f_X(x) = 1$이므로 밀도다. 정당화는 극한으로 한다. $h \to 0$이면

$$P(Y \le y \mid x < X \le x + h) = \frac{\int_x^{x+h}\int_{-\infty}^{y} f(s,t)\,dt\,ds}{\int_x^{x+h} f_X(s)\,ds} \longrightarrow \int_{-\infty}^{y} \frac{f(x,t)}{f_X(x)}\,dt$$

($f$가 연속인 점에서). "$X$가 $x$ 근처에 있다"는 확률이 양수인 사건에 조건을 걸고 극한을 취한 것이다. 이 극한을 **어떻게** 취하느냐가 답을 바꿀 수 있다는 것이 7.1절의 Borel 역설이다.

그림으로 읽으면 조건부 밀도는 결합밀도의 **단면**이다. $x$에서 결합밀도를 세로로 자르고, 그 단면의 넓이가 1이 되도록 $f_X(x)$로 나눈 것이다(그림 2.1).

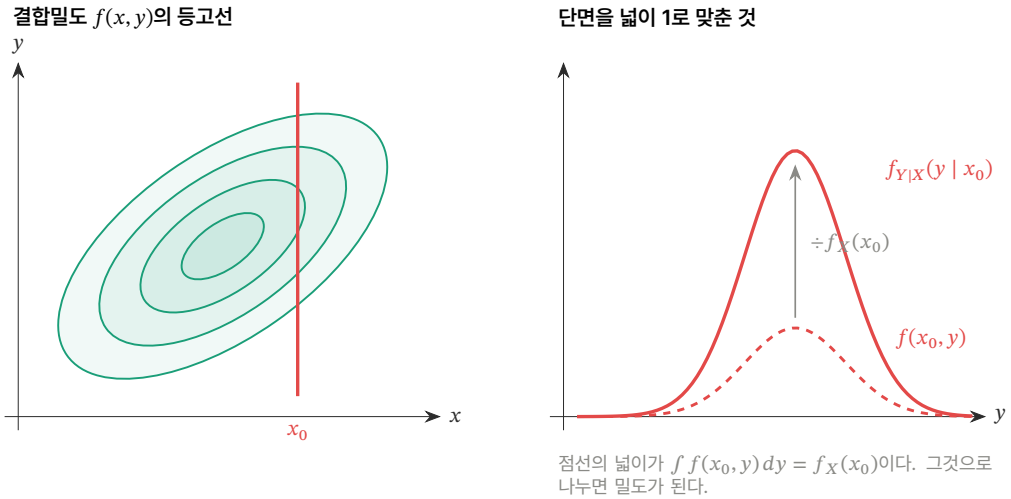



**그림 2.1** 결합밀도의 단면이 조건부 밀도다. $X = x_0$에서 자른 곡선을 그 넓이 $f_X(x_0)$로 나누면 $f_{Y|X}(y \mid x_0)$가 된다.

**예제 2.3** (원판의 단면) 예제 1.7에서 $|x| < 1$이면

$$f_{Y|X}(y \mid x) = \frac{1/\pi}{\frac{2}{\pi}\sqrt{1-x^2}} = \frac{1}{2\sqrt{1-x^2}}, \qquad |y| \le \sqrt{1-x^2}.$$

$X = x$를 알면 $Y$는 구간 $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$ 위의 균등분포다. 결합분포가 균등하면 단면도 균등하다.

## 2.2 조건부기댓값

**정의 2.4** (조건부기댓값) $X = x$일 때 $Y$의 조건부분포의 평균

$$E(Y \mid X = x) = \sum_y y\, p(y \mid x) \quad \text{또는} \quad \int y\, f(y \mid x)\, dy$$

를 **조건부기댓값**이라 한다. 이것은 $x$의 함수 $m(x)$이고, 그 함수에 $X$를 넣은 확률변수 $m(X)$를 $E(Y \mid X)$로 쓴다. 같은 방법으로 $E(g(Y) \mid X)$와 $\mathrm{Var}(Y \mid X) = E\big[(Y - E(Y \mid X))^2 \mid X\big]$를 정의한다.

$E(Y \mid X = x)$는 수이고 $E(Y \mid X)$는 확률변수다. 이 구별이 다음 정리의 핵심이다. 그림 1.1의 빨간 점이 $m(x) = E(S \mid M = m)$이다.

**정리 2.5** (반복기댓값) $E|Y| < \infty$이면

$$E\big[E(Y \mid X)\big] = EY.$$

더 일반적으로 $E|g(X)Y| < \infty$이면 $E\big[g(X)\,E(Y \mid X)\big] = E\big[g(X)\,Y\big]$.

**증명.** **전략** $E[g(X)m(X)]$를 $X$에 대한 적분으로 쓰고, $f(x,y) = f(y \mid x)f_X(x)$로 바꿔 이중적분으로 만든다. 그다음 적분 순서를 바꾼다. 연속인 경우를 쓰고, 이산은 같은 계산을 합으로 한다.

① $m(x) = \int y\, f(y \mid x)\, dy$이고, $f_X(x) > 0$인 곳에서 $f(x,y) = f(y \mid x)\, f_X(x)$이다. 정의 2.2, 2.4

② $g(X)m(X)$는 $X$의 함수이므로

$$E\big[g(X)m(X)\big] = \int g(x)\Big(\int y\, f(y \mid x)\, dy\Big) f_X(x)\, dx. \qquad \text{2회 1.4절의 상자}$$

③ ①을 넣고 적분 순서를 바꾼다.

$$\int g(x)\Big(\int y\, f(y \mid x)\, dy\Big) f_X(x)\, dx = \iint g(x)\, y\, f(x,y)\, dy\, dx. \qquad \text{Fubini [T6]}$$

④ 오른쪽은 $E[g(X)Y]$이다. 정리 1.8(b)

⑤ $g \equiv 1$이면 첫 식 $E[E(Y \mid X)] = EY$가 된다. ④

**조건 확인** 셋째 단계에서 적분 순서를 바꾸려면 $\iint |g(x)\,y|\, f(x,y)\, dy\, dx < \infty$이어야 한다. 이것은 $E|g(X)Y| < \infty$라는 가정 그대로다(Fubini, [T6]). 같은 이유로 거의 모든 $x$에서 $\int |y|\, f(y \mid x)\, dy < \infty$이므로 $m(x)$가 정의된다. $f_X(x) = 0$인 $x$들의 집합은 확률이 0이므로 거기서 $m$을 어떻게 정해도 적분은 바뀌지 않는다. 증명은 $Y$ 대신 $(X,Y)$의 함수 $h(X,Y)$를 넣어도 그대로 된다. $E|h(X,Y)| < \infty$이면 $E\big[E(h(X,Y) \mid X)\big] = E\,h(X,Y)$이다. ∎

정리는 "평균을 두 단계로 낼 수 있다"는 말이다. 먼저 $X$의 값마다 $Y$의 평균을 내고, 그 평균들을 $X$의 분포로 다시 평균한다. 1회 정리 5.4의 전확률 공식을 기댓값으로 쓴 것이기도 하다.

**정리 2.6** (전분산 공식) $EY^2 < \infty$이면

$$\operatorname{Var} Y = E\big[\operatorname{Var}(Y \mid X)\big] + \operatorname{Var}\big(E(Y \mid X)\big).$$

**증명.** **전략** $Y - EY$를 "조건부 평균에서 벗어난 부분"과 "조건부 평균이 전체 평균에서 벗어난 부분"으로 쪼개 제곱한다. 교차항이 0임을 반복기댓값의 둘째 식으로 보인다.

① $m(X) = E(Y \mid X)$로 두면 $Y - EY = \big(Y - m(X)\big) + \big(m(X) - EY\big)$이다. 항등식

② 제곱해 기댓값을 취한다.

$$\operatorname{Var} Y = E\big(Y - m(X)\big)^2 + E\big(m(X) - EY\big)^2 + 2E\big[(Y - m(X))(m(X) - EY)\big]. \quad \text{따름정리 1.9}$$

③ 첫 항은 $E\big[E((Y-m(X))^2 \mid X)\big] = E\big[\operatorname{Var}(Y \mid X)\big]$이다. 정리 2.5를 $(Y-m(X))^2$에, 정의 2.4

④ $E\,m(X) = EY$이므로 둘째 항은 $\operatorname{Var}\big(m(X)\big)$이다. 정리 2.5

⑤ $g(X) = m(X) - EY$로 두면 교차항은

$$E\big[g(X)\,(Y - m(X))\big] = E\big[g(X)\,Y\big] - E\big[g(X)\,m(X)\big] = 0. \quad \text{정리 2.5의 둘째 식}$$

⑥ ③–⑤를 ②에 넣으면 결론이다.

**조건 확인** 다섯째 단계에서 정리 2.5를 쓰려면 $E|g(X)Y| < \infty$이어야 한다. $2|ab| \le a^2 + b^2$ [T4]로 $E|g(X)Y| \le \frac{1}{2}\big(E\,g(X)^2 + EY^2\big)$이고, $E\,g(X)^2 \le 2E\,m(X)^2 + 2(EY)^2$이다. 남은 $E\,m(X)^2 \le EY^2$는 아래 정리 2.8의 첫 단계다. 그 단계는 전분산 공식을 쓰지 않는다. ■

전체의 흩어짐은 "집단 안의 흩어짐의 평균"과 "집단 평균들의 흩어짐"의 합이다. 14회 분산분석의 분해 $SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}$이 이 공식의 표본판이다.

**예제 2.7** (주사위를 던지고 그 수만큼 동전을) 주사위 눈 $N$을 보고 공평한 동전을 $N$번 던져 앞면 수 $K$를 센다. $N = n$이면 $K \sim \mathrm{b}(n, \frac{1}{2})$이므로 $E(K \mid N) = \frac{N}{2}$, $\operatorname{Var}(K \mid N) = \frac{N}{4}$(2회 정리 2.4). 그러면

$$EK = E\Big[\frac{N}{2}\Big] = \frac{3.5}{2} = 1.75,$$
$$\operatorname{Var} K = E\Big[\frac{N}{4}\Big] + \operatorname{Var}\Big(\frac{N}{2}\Big) = \frac{3.5}{4} + \frac{35/12}{4} \approx 0.875 + 0.729 = 1.604.$$

$K$의 분포를 직접 구하지 않고 평균과 분산을 얻었다. 분산의 절반 가까이가 "동전을 몇 번 던지는가"의 불확실성에서 온다.

## 2.3 가장 좋은 예측

2회 정리 2.2(c)는 상수 하나로 $Y$를 예측할 때 제곱오차를 가장 작게 하는 것이 평균임을 보였다. $X$를 보고 예측할 수 있다면 가장 좋은 것은 조건부 평균이다.

**정리 2.8** (조건부기댓값은 가장 좋은 예측) $EY^2 < \infty$이면 $Eg(X)^2 < \infty$인 모든 함수 $g$에 대해

$$E\big(Y - g(X)\big)^2 = E\big(Y - m(X)\big)^2 + E\big(m(X) - g(X)\big)^2 \;\geq\; E\big(Y - m(X)\big)^2.$$

특히 $E\,m(X)^2 \leq EY^2$이다.

**증명.** **전략** $Y - g(X)$를 $(Y - m(X)) + (m(X) - g(X))$로 쪼개 제곱한다. 교차항은 반복기댓값의 둘째 식으로 0이 된다. 전분산 공식과 같은 계산이다. 먼저 모든 항이 유한함을 확인한다.

① $m(x)^2 \leq E(Y^2 \mid X = x)$이다. 이것을 $X$의 분포로 평균하면 $E\,m(X)^2 \leq E\big[E(Y^2 \mid X)\big] = EY^2 < \infty$이다. 2회 따름정리 5.10(a), 정리 2.5

② $h(X) = m(X) - g(X)$로 두면 $E\,h(X)^2 \leq 2E\,m(X)^2 + 2E\,g(X)^2 < \infty$이다. $(a+b)^2 \leq 2a^2 + 2b^2$ [T4]

③ $Y - g(X) = \big(Y - m(X)\big) + h(X)$를 제곱해 기댓값을 취한다.

$$E\big(Y - g(X)\big)^2 = E\big(Y - m(X)\big)^2 + E\,h(X)^2 + 2E\big[h(X)\,(Y - m(X))\big].$$ 따름정리 1.9

④ 교차항은 $E[h(X)Y] - E[h(X)m(X)] = 0$이다. $E|h(X)Y| < \infty$는 ②와 $2|ab| \leq a^2 + b^2$에서 나온다. 정리 2.5의 둘째 식

⑤ $E\,h(X)^2 \geq 0$이므로 부등식이 나온다. 등호는 $E\big(m(X) - g(X)\big)^2 = 0$일 때뿐이다. ③④

■

이 정리가 **회귀**의 수학적 정의다. $m(x) = E(Y \mid X = x)$를 $X$에 대한 $Y$의 **회귀함수**라 한다. 5회에서 이변량 정규분포의 회귀함수가 직선임을 보이고, 13회에서 회귀함수를 데이터로 추정한다. 기계학습의 제곱손실 회귀가 추정하려는 대상도 바로 이 $m(x)$다.

### 연습문제

**연습 2.1** (조건부 확률질량함수 ★) 예제 1.4에서 $M = 4$일 때 $S$의 조건부분포와 $E(S \mid M = 4)$를 구하라. $E(S \mid M = m)$을 $m$의 식으로 쓰고 반복기댓값으로 $ES = 7$을 확인하라. 풀이 49쪽

**연습 2.2** (삼각형의 조건부 ★) 연습 1.2의 밀도에서 $f_{X|Y}(x \mid y)$와 $E(X \mid Y)$를 구하고, $EX$를 반복기댓값으로 다시 구하라. 풀이 49쪽

**연습 2.3** (무작위 합 ★★) $N$이 음이 아닌 정수값의 확률변수이고, $N$과 독립인 $X_1, X_2, \ldots$가 서로 독립이며 평균 $\mu$, 분산 $\sigma^2$를 갖는다. $S = X_1 + \cdots + X_N$ ($N = 0$이면 0)에 대해 $ES = \mu\,EN$, $\operatorname{Var} S = \sigma^2 EN + \mu^2 \operatorname{Var} N$임을 보여라(Wald의 항등식). 보험사의 한 해 지급액의 분산이 "청구 건수의 불확실성"과 "건당 금액의 불확실성"으로 나뉘는 방식이다. 풀이 50쪽

**연습 2.4** (두 단계의 균등 ★★) $X \sim \mathrm{U}(0,1)$이고, $X = x$일 때 $Y \sim \mathrm{U}(0,x)$이다. $(X,Y)$의 결합밀도, $Y$의 주변 밀도, $EY$, $\mathrm{Var}\,Y$를 구하라. 마지막 둘은 반복기댓값과 전분산 공식으로도 구하라. 풀이 50쪽

**연습 2.5** (조건부기댓값의 성질 ★★) 정의에서 바로 (a) $E(aY_1+bY_2 \mid X) = aE(Y_1 \mid X)+bE(Y_2 \mid X)$, (b) $E(g(X)Y \mid X) = g(X)E(Y \mid X)$를 보여라. (b)를 "아는 것은 상수처럼 밖으로 꺼낸다"고 읽는다. 풀이 50쪽

**연습 2.6** (예측의 개선 ★) 예제 2.7에서 $N$을 보고 $K$를 예측하면 평균제곱오차는 얼마나 줄어드나? $\mathrm{Var}(E(K \mid N))/\mathrm{Var}\,K$를 구하라. 이 비를 "$N$이 설명하는 분산의 비율"이라 부른다. 풀이 50쪽

# 3

# 독립

두 확률변수가 독립이라는 것은 한쪽의 값을 알아도 다른 쪽의 분포가 바뀌지 않는다는 것이다. 2회는 이산 확률변수에서 이것을 정의했다. 이 절은 연속 확률변수를 포함한 일반적인 정의를 세우고, 독립을 판정하는 간단한 기준(인수분해)을 증명한 뒤, 2회에 남긴 꼬리표 둘을 뗀다.

## 3.1 정의와 판정

**정의 3.1** (독립인 확률변수) 확률변수 $X_1, \dots, X_n$이 모든 실수 $x_1, \dots, x_n$에 대해

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

을 만족하면 **독립**이라 한다. 무한히 많은 확률변수는 그 가운데 어떤 유한 개도 독립이면 독립이라 한다. 독립이고 분포가 모두 같으면 **iid**(independent and identically distributed)라 하고 $X_i \overset{\text{iid}}{\sim} F$로 쓴다.

이산 확률변수에서는 이 정의가 2회 정의 2.6($p(x_1, \dots, x_n) = \prod p(x_i)$)과 같다(연습 3.1). 연속에서는 밀도로 판정하는 것이 편하다.

**정리 3.2** (인수분해 판정법) $(X, Y)$가 결합밀도 $f$를 가질 때 다음은 동치다.

(a) $X$와 $Y$는 독립이다.

(b) (거의) 모든 $(x, y)$에서 $f(x, y) = f_X(x)\, f_Y(y)$.

(c) 음이 아닌 함수 $g, h$가 있어 $f(x, y) = g(x)\, h(y)$.

**증명.** **전략** (b)⇒(a)는 이중적분을 두 적분의 곱으로 나눈다. (a)⇒(b)는 결합분포함수를 $x$와 $y$로 한 번씩 미분한다. (c)⇒(b)는 $g$와 $h$의 상수배를 주변 밀도로 맞춘다.

**(b)⇒(a)**

① $F(x, y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_X(s) f_Y(t)\, ds\, dt$이다. 정의 1.5, (b)

② 음이 아닌 곱함수의 적분은 적분의 곱이므로 $F(x, y) = \int_{-\infty}^{x} f_X \cdot \int_{-\infty}^{y} f_Y = F_X(x) F_Y(y)$이다.

Tonelli [T6]

**(a)⇒(b)**

① $F(x,y)=F_X(x)F_Y(y)$를 $x$로, 다시 $y$로 미분하면, $f_X$와 $f_Y$가 연속인 점에서 $\partial^2F/\partial x\,\partial y = f_X(x)f_Y(y)$이다. 1회 정리 6.7(d)

② 정의 1.5의 적분을 두 번 미분하면, $f$가 연속인 점에서 $\partial^2F/\partial x\,\partial y = f(x,y)$이다. 미적분의 기본정리 두 번

③ 두 식이 모두 성립하는 점에서 $f(x,y)=f_X(x)f_Y(y)$이다. 밀도는 넓이 0인 집합에서 바꿔도 같은 분포를 주므로 "(거의) 모든 점"으로 충분하다. ①②

**(b)⇒(c), (c)⇒(b)**

① (b)이면 $g=f_X$, $h=f_Y$로 두면 된다.

② (c)이면 $G=\int g$, $H=\int h$로 두고 $1=\iint f = GH$이다. Tonelli [T6]

③ $f_X(x)=\int g(x)h(y)\,dy = g(x)H$이고 $f_Y(y)=G\,h(y)$이다. 정리 1.6

④ $f_X(x)f_Y(y)=g(x)h(y)\,GH = f(x,y)$이다. ②③

■

**가정 주의** (c)에는 숨은 조건이 있다. $f$가 0이 아닌 영역도 "곱"의 꼴, 곧 사각형 $\{x\in A\}\times\{y\in B\}$여야 한다. 예제 1.7의 원판에서 $f=\frac{1}{\pi}$는 안에서 상수이지만 영역이 원이므로 $f(x,y)=\frac{1}{\pi}\mathbf{1}\{x^2+y^2\le 1\}$은 $g(x)h(y)$로 쓸 수 없다. 실제로 $X$가 1 가까이 있으면 $Y$는 0 근처에 있어야 하므로 둘은 독립이 아니다.

**예제 3.3** (지수분포 두 개와 정사각형) $f(x,y)=e^{-x-y}$ $(x,y>0)$은 $e^{-x}\cdot e^{-y}$이고 영역이 사분면(곱의 꼴)이므로 $X,Y$는 독립인 Exp(1)이다. 정사각형 $[0,1]^2$ 위의 균등분포는 $1=\mathbf{1}\{0<x<1\}\cdot\mathbf{1}\{0<y<1\}$이므로 독립인 두 U(0,1)이다. 컴퓨터가 균등 난수 두 개를 따로 뽑는 것은 정사각형에서 점 하나를 뽑는 것과 같다.

## 3.2 독립인 변수의 함수, 곱, 합

**정리 3.4** (독립의 결과들) $X,Y$가 독립이면

(a) 함수 $g(X)$와 $h(Y)$도 독립이다.

(b) $E|g(X)|, E|h(Y)|<\infty$이면 $E\big[g(X)h(Y)\big]=E\,g(X)\cdot E\,h(Y)$.

(c) 분산이 유한하면 $\mathrm{Var}(X+Y)=\mathrm{Var}\,X+\mathrm{Var}\,Y$.

(d) mgf가 있으면 $M_{X+Y}(t)=M_X(t)\,M_Y(t)$.

여러 개의 독립인 변수에 대해서도 같다.

**증명.** **전략** (b)를 먼저 증명한다. 결합밀도를 인수분해(정리 3.2)하면 이중적분이 두 적분의 곱이 된다. 절댓값으로 먼저 계산해 적분 순서를 바꿀 수 있음을 확인한다 [T6]. (c)와 (d)는 (b)를 알맞은 함수에 쓴 것이다. (a)는 단조함수이면 정의에서 나오고, 일반적인 함수는 아래 상자의 사실이다. 단계별 증명은 연습 3.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 51쪽에 있다.

**증명 없이 쓰는 사실** (독립인 변수의 함수는 독립이다) $X_1, \dots, X_n$이 독립이면, 서로 겹치지 않는 묶음의 ("보통의") 함수들 $g(X_1, X_2)$, $h(X_3)$, … 도 독립이다. 증명은 독립을 사각형에서 Borel 집합 전체로 넓히는 측도론의 $\pi$–$\lambda$ 정리를 쓴다(Billingsley, *Probability and Measure*, 20절). 이 과정에서는 이 사실을 자유롭게 쓴다.

이로써 2회의 꼬리표가 모두 떨어졌다. 독립합의 분산(2회 정리 2.8)과 대수의 법칙(2회 정리 6.3)은 연속 확률변수에서도 그대로 성립하고, 독립합의 mgf가 곱이라는 것(2회 정리 3.6(b))도 그렇다. 4회의 분포들, 5회의 정규분포, 7회의 중심극한정리가 이 결과를 쓴다.

**예제 3.5** (iid의 최댓값과 최솟값) $X_1, \dots, X_n \overset{\text{iid}}{\sim} F$, $M_n = \max X_i$, $L_n = \min X_i$라 하자. $\{M_n \le x\} = \bigcap_i \{X_i \le x\}$이므로 독립으로

$$F_{M_n}(x) = F(x)^n, \qquad P(L_n > x) = \big(1 - F(x)\big)^n.$$

1회 연습 6.6이 예고한 것이 첫 식이다. $\mathrm{U}(0,1)$이면 $EM_n = \int_0^1 (1 - x^n)dx = \frac{n}{n+1}$ (꼬리 공식, 2회 정리 1.13)이다. $\mathrm{Exp}(\lambda)$이면 $P(L_n > x) = e^{-n\lambda x}$이므로 $L_n \sim \mathrm{Exp}(n\lambda)$이다. 전구 $n$개 가운데 처음 꺼지는 시간은 평균 $\frac{1}{n\lambda}$이다.

## 연습문제

**연습 3.1** (이산에서 두 정의는 같다 ★★) $X, Y$가 정수값을 가질 때, 정의 3.1(분포함수의 곱)과 $p(x,y) = p_X(x)p_Y(y)$가 동치임을 보여라. (힌트: $p(x,y)$를 $F$의 네 값으로 쓴다. 연습 1.1.) 풀이 50쪽

**연습 3.2** (독립 판정 ★) 다음 결합밀도에서 $X, Y$가 독립인지 판정하라. (a) $f = 4xy$ $(0 < x, y < 1)$, (b) $f = 8xy$ $(0 < x < y < 1)$, (c) $f = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$ $(\mathbb{R}^2)$. 풀이 50쪽

**연습 3.3** (최댓값의 기댓값 ★) 독립인 $\mathrm{U}(0,1)$ 세 개의 최댓값의 밀도와 평균을 구하라. 주사위 두 개의 최댓값의 평균 $\frac{161}{36}$은 같은 방법으로 어떻게 나오나? 풀이 50쪽

**연습 3.4** (경쟁하는 지수분포 ★★) $X \sim \mathrm{Exp}(\lambda)$와 $Y \sim \mathrm{Exp}(\mu)$가 독립이다. $P(X<Y)=\dfrac{\lambda}{\lambda+\mu}$임을 이중적분으로 보여라. 두 사건 가운데 무엇이 먼저 일어나는가를 결정하는 기본 공식이다.

풀이 51쪽

**연습 3.5** (쌍별 독립의 한계 ★★) 1회의 Bernstein 정사면체에서 $R, G, B$는 쌍별 독립이다. $E(RGB)\neq ER\,EG\,EB$임을 보이고, 정리 3.4(b)를 세 변수로 넓히려면 상호 독립이 필요한 이유를 말하라.

풀이 51쪽

**연습 3.6** (독립의 결과들 증명 ★★ 유도형) $X, Y$가 독립이고 결합밀도를 갖는다. 정리 3.4를 다음 순서로 증명하라.

(a) (정리의 (b)) $\iint |g(x)h(y)|\,f_X(x)f_Y(y)\,dx\,dy = E|g(X)|\,E|h(Y)|$를 Tonelli로 보이고, 같은 계산을 부호를 붙여 다시 해 $E[g(X)h(Y)]=E\,g(X)\,E\,h(Y)$를 얻어라. 두 번째 계산에서 순서를 바꾸는 근거는 무엇인가?

(b) (정리의 (c)) $D_X=X-EX$, $D_Y=Y-EY$로 두고 $\mathrm{Var}(X+Y)=\mathrm{Var}\,X+\mathrm{Var}\,Y+2E(D_XD_Y)$를 보인 뒤 (a)로 교차항이 0임을 보여라.

(c) (정리의 (d)) (a)를 $g(x)=e^{tx}$, $h(y)=e^{ty}$에 써라. $E\,e^{tX}<\infty$가 왜 필요한가?

(d) (정리의 (a)의 특수한 경우) $g$가 증가함수이면 $\{g(X)\le s\}$가 $\varnothing$, $\Omega$, $\{X\le c\}$, $\{X<c\}$ 가운데 하나임을 보여라. 이때 $g(X)$와 $Y$의 결합분포함수가 인수분해됨을 정의 3.1와 1회 정리 4.7로 보여라.

풀이 51쪽

# 4

# 공분산과 상관

독립이 아닌 두 변수가 "얼마나, 어느 방향으로" 함께 움직이는지를 하나의 수로 재고 싶다. 2회 정리 2.8의 증명에 나온 교차항 $E(D_iD_j)$가 그 수다. 이 절은 그것에 이름을 붙이고, Cauchy–Schwarz 부등식으로 그 크기에 한계를 준다.

## 4.1 공분산

**정의 4.1** (공분산과 상관계수) $EX^2, EY^2 < \infty$, $\mu_X = EX$, $\mu_Y = EY$일 때

$$\operatorname{Cov}(X,Y) = E\big[(X-\mu_X)(Y-\mu_Y)\big]$$

을 **공분산**, 두 분산이 양수일 때

$$\rho = \operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

를 **상관계수**라 한다. $\operatorname{Cov}(X,Y)=0$이면 **상관이 없다**고 한다.

**정리 4.2** (공분산의 계산 규칙)

(a) $\operatorname{Cov}(X,Y) = E(XY) - \mu_X\mu_Y$, $\operatorname{Cov}(X,X) = \operatorname{Var}X$, $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$.

(b) (쌍선형) $\operatorname{Cov}(aX+b,\ cY+d) = ac\operatorname{Cov}(X,Y)$, $\operatorname{Cov}(X_1+X_2,Y) = \operatorname{Cov}(X_1,Y) + \operatorname{Cov}(X_2,Y)$.

(c) $\operatorname{Var}\left(\sum_i a_iX_i\right) = \sum_i a_i^2 \operatorname{Var}X_i + 2\sum_{i<j} a_ia_j \operatorname{Cov}(X_i,X_j)$.

(d) 독립이면 상관이 없다.

**증명.** **전략** 모든 규칙을 공분산의 정의와 기댓값의 선형성(따름정리 1.9)으로 옮긴다. (c)는 분산을 자기 자신과의 공분산으로 보고 (b)로 전개한다. 단계별 증명은 연습 4.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 52쪽에 있다.

**예제 4.3** (두 주사위의 합과 최댓값의 공분산) 예제 1.4에서 $S = D_1 + D_2$, $M = \max(D_1, D_2)$이다. 쌍선형으로 $\operatorname{Cov}(S,M) = \operatorname{Cov}(D_1,M) + \operatorname{Cov}(D_2,M) = 2\operatorname{Cov}(D_1,M)$이고, 대칭으로 계산하면 $\operatorname{Cov}(S,M) = \frac{35}{12} \approx 2.917$, $\operatorname{Var} M = \frac{2555}{1296} \approx 1.971$이다(연습 4.1). 따라서

$$\rho(S,M) = \frac{35/12}{\sqrt{35/6}\sqrt{2555/1296}} \approx 0.86.$$

## 4.2 Cauchy–Schwarz와 상관의 한계

**정리 4.4** (Cauchy–Schwarz 부등식) $EX^2, EY^2 < \infty$이면

$$\big(E(XY)\big)^2 \le EX^2\, EY^2.$$

등호는 어떤 상수 $c$에 대해 $P(Y = cX) = 1$이거나 $P(X = 0) = 1$일 때뿐이다.

**증명.** **전략** 모든 실수 $t$에서 $E(Y-tX)^2 \ge 0$이다. 이것을 $t$의 이차식으로 쓰면, 음이 아닌 이차식의 판별식은 0 이하라는 것이 곧 부등식이다 [T4].

① $2|XY| \le X^2 + Y^2$이므로 $E|XY| < \infty$이고 $E(XY)$가 정의된다. [T4]

② $EX^2 = 0$이면 $P(X = 0) = 1$이고 양변이 0이다. 2회 따름정리 5.6

③ $EX^2 > 0$이면 모든 실수 $t$에 대해

$$0 \le E(Y - tX)^2 = EY^2 - 2t\,E(XY) + t^2 EX^2.$$ 따름정리 1.9

④ 음이 아닌 이차식의 판별식은 0 이하이므로 $4\big(E(XY)\big)^2 - 4\,EX^2\,EY^2 \le 0$이다. 이차식의 판별식 [T4]

⑤ 등호이면 판별식이 0이고, 이차식이 0이 되는 $t = c = E(XY)/EX^2$가 있다. 그러면 $E(Y - cX)^2 = 0$이므로 $P(Y = cX) = 1$이다. 2회 따름정리 5.6

⑥ 거꾸로 $P(Y = cX) = 1$이면 양변이 모두 $c^2(EX^2)^2$이다.

■

**따름정리 4.5** (상관계수의 범위) $-1 \le \rho \le 1$이고, $|\rho| = 1$일 필요충분조건은 $P(Y = a+bX) = 1$인 상수 $a$와 $b \ne 0$이 있는 것이다. 이때 $\rho$의 부호는 $b$의 부호다.

**증명.** 정리 4.4를 $X - \mu_X$와 $Y - \mu_Y$에 쓰면 $\operatorname{Cov}(X,Y)^2 \le \sigma_X^2 \sigma_Y^2$. 등호 조건은 $Y - \mu_Y = c(X - \mu_X)$이다. ■

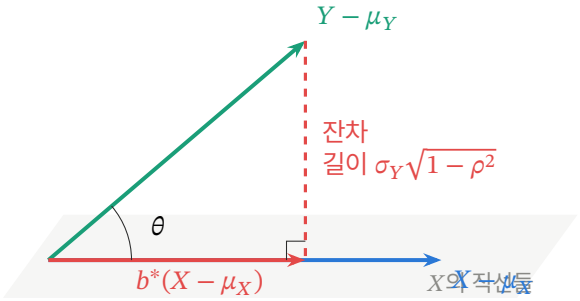



**그림 4.1** 상관계수는 각의 코사인이다. 평균을 뺀 확률변수를 벡터로 보면 공분산이 내적, 표준편차가 길이다. Cauchy–Schwarz는 "코사인은 −1과 1 사이"라는 말이다.

그림 4.1의 기하는 비유가 아니다. 평균이 0이고 분산이 유한한 확률변수들은 내적 $\langle X,Y\rangle = E(XY)$를 갖는 벡터공간을 이룬다. 길이는 표준편차, 각의 코사인은 상관계수, 직교는 상관없음이다. 이 그림으로 보면 정리 2.8의 조건부기댓값은 $Y$를 "$X$의 함수들의 공간"에 **직교사영**한 것이고, 전분산 공식은 피타고라스 정리다. 5회와 13회는 이 기하를 계속 쓴다.

## 4.3 가장 좋은 직선과 상관의 함정

**정리 4.6** (가장 좋은 선형 예측) $E(Y-a-bX)^2$을 가장 작게 하는 $(a,b)$는

$$b^*=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}\,X}=\rho\frac{\sigma_Y}{\sigma_X},\qquad a^*=\mu_Y-b^*\mu_X$$

이고, 그때의 최솟값은 $\sigma_Y^2(1-\rho^2)$이다.

**증명.** **전략** $b$를 고정하고 $a$에 대해 먼저 최소화한다. 그러면 남은 값은 $b$의 이차식이고, 그것을 최소화한다. 단계별 증명은 연습 4.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 53쪽에 있다.

$\rho^2$은 $X$의 직선이 설명하는 $Y$의 분산의 비율이다. 직선 $y=a^*+b^*x$를 $X$에 대한 $Y$의 **회귀직선**이라 한다. 기울기에 $\rho$가 곱해져 있으므로 $|\rho|<1$이면 표준단위로 잰 예측은 평균 쪽으로 "되돌아간다". Galton이 *regression*이라 부른 현상이다(5회).

**예제 4.7** (상관이 없어도 독립이 아니다) $X\sim \mathrm{U}(-1,1)$, $Y=X^2$이면 $Y$는 $X$로 완전히 정해진다. 그런데 $\mathrm{Cov}(X,Y)=EX^3-EX\,EX^2=0$이다(대칭). 상관계수는 **직선** 관계만 잰다. 정리 4.2(d)의 역은 거짓이다.

**예제 4.8** (상관은 인과가 아니다 — 공통 원인) $Z$가 공통 원인이고 $X = Z + E_1$, $Y = Z + E_2$ ($Z, E_1, E_2$ 독립, 분산 $\sigma_Z^2, \sigma^2, \sigma^2$)이면 $\mathrm{Cov}(X,Y) = \mathrm{Var}\,Z$이고 $\rho = \frac{\sigma_Z^2}{\sigma_Z^2+\sigma^2}$이다. $X$는 $Y$에 아무 영향도 주지 않지만 둘은 상관되어 있다. 1회 연습 5.9의 Simpson 역설에서 결석의 크기가 이런 공통 원인이었다.

## 연습문제

**연습 4.1** (합과 최댓값 ★★) $E(D_1 M)$을 직접 세어 $\mathrm{Cov}(D_1, M)$을 구하고, 예제 4.3에서 쓴 $\mathrm{Cov}(S,M)$이 $\frac{35}{12}$임을 확인하라. (힌트: $M = D_1$인 경우와 $M = D_2 > D_1$인 경우로 나눈다.) 풀이 51쪽

**연습 4.2** (합과 차 ★) $X, Y$의 분산이 같으면 $X+Y$와 $X-Y$는 상관이 없음을 보여라. 두 주사위의 합과 차는 독립인가? 풀이 52쪽

**연습 4.3** (포트폴리오 ★★) 수익률 $X, Y$의 표준편차가 $\sigma_X = 0.2$, $\sigma_Y = 0.3$, 상관계수가 $\rho$이다. $wX + (1-w)Y$의 분산을 가장 작게 하는 $w$를 $\rho$의 식으로 구하고, $\rho = -1, 0, 0.5$일 때 최소 분산을 구하라. $\rho = -1$이면 위험을 완전히 없앨 수 있다. 풀이 52쪽

**연습 4.4** (상관계수의 불변성 ★) $a, c > 0$이면 $\mathrm{Corr}(aX+b,\ cY+d) = \mathrm{Corr}(X,Y)$임을 보여라. $a < 0$이면? 풀이 52쪽

**연습 4.5** (부등식의 사촌들 ★★) Cauchy–Schwarz로 (a) $E|X| \le \sqrt{EX^2}$, (b) $\sigma_{X+Y} \le \sigma_X + \sigma_Y$를 보여라. (b)는 표준편차의 삼각부등식이다. 풀이 52쪽

**연습 4.6** (원판의 상관 ★) 예제 1.7의 $(X,Y)$는 상관이 없음을 보여라. 3.1절의 주의 상자에 따르면 둘은 독립이 아니다. 독립이 아님을 확률로 보여라: $P(X > 0.8,\ Y > 0.8)$과 $P(X > 0.8)P(Y > 0.8)$을 비교한다. 풀이 52쪽

**연습 4.7** (공분산의 계산 규칙 증명 ★ 유도형) $EX_i^2 < \infty$라 하자. 정리 4.2를 다음 순서로 증명하라.

(a) $2|xy| \le x^2 + y^2$로 $E|XY| < \infty$임을 보여 공분산이 정의됨을 확인하라.

(b) $(X-\mu_X)(Y-\mu_Y) = XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y$에 선형성을 써서 (a)의 첫 식을 얻어라.

(c) 정의에 $aX+b$와 $cY+d$를 넣어 쌍선형을 보여라. 상수 $b, d$는 어디서 사라지는가?

(d) $\mathrm{Var}(\sum_i a_i X_i) = \mathrm{Cov}(\sum_i a_i X_i, \sum_j a_j X_j) = \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j)$를 보이고, $i = j$인 항과 $i \ne j$인 항으로 나눠 (c)를 얻어라. $i \ne j$인 순서쌍이 $i < j$인 쌍의 두 배인 이유를 말하라.

(e) 정리 3.4(b)로 (d)를 보여라.

풀이 52쪽

**연습 4.8** (가장 좋은 선형 예측 증명 ★★ 유도형) 정리 4.6를 다음 순서로 증명하라.

(a) $b$를 고정하면 $E(Y-bX-a)^2$을 가장 작게 하는 $a$는 $E(Y-bX)$이고 그때의 값은 $\mathrm{Var}(Y-bX)$임을 2회 정리 2.2(c)로 보여라.

(b) $\mathrm{Var}(Y-bX)=\sigma_Y^2-2b\,\mathrm{Cov}(X,Y)+b^2\sigma_X^2$ 임을 정리 4.2로 보여라.

(c) (b)의 이차식을 최소로 하는 $b^*$를 구하고, 최솟값이 $\sigma_Y^2(1-\rho^2)$임을 보여라.

(d) 잔차 $Y-a^*-b^*X$가 $X$와 상관이 없음을 보여라.

풀이 53쪽

# 5

# 여러 변수의 변수변환 — 야코비안

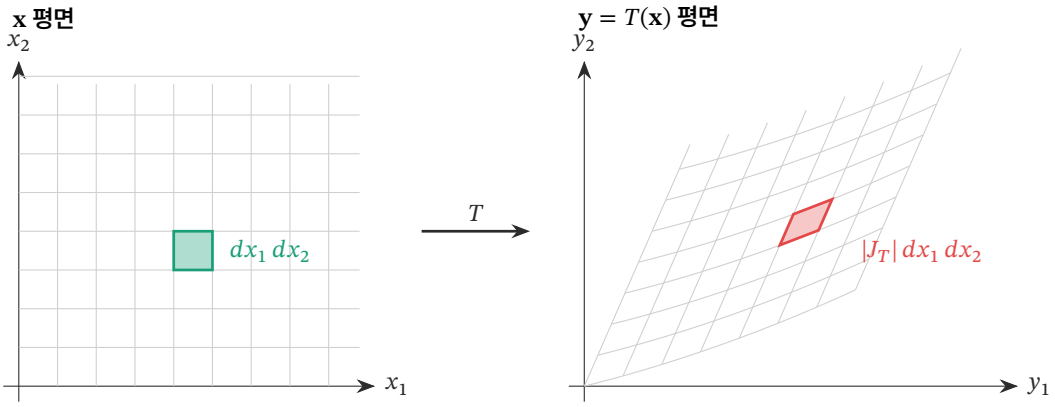
2회 정리 4.3은 한 변수의 단조변환에서 밀도가 $|h'(y)|$배로 바뀐다고 했다. 두 변수 이상에서는 그 배율이 **야코비안 행렬식**이다. 이 절은 그 공식을 세우고, 그것으로 이 과정 전체에서 쓰는 네 결과를 얻는다. 가우스 적분(5회의 정규분포의 상수), Box–Muller 방법(정규 난수), 합성곱(독립합의 밀도), 그리고 정규의 비가 Cauchy라는 것이다.

## 5.1 변수변환 공식

**증명 없이 쓰는 사실** (다중적분의 치환) $T$가 열린 집합 $S \subset \mathbb{R}^n$에서 열린 집합 $T(S)$로 가는 일대일 연속미분가능 변환이고 야코비안 행렬식 $J_T(\mathbf{x}) = \det\left[\partial T_i/\partial x_j\right]$이 0이 아니면, $T(S)$ 위의 적분 가능한 $g$에 대해

$$\int_{T(S)} g(\mathbf{y})\,d\mathbf{y} = \int_{S} g\big(T(\mathbf{x})\big)\,|J_T(\mathbf{x})|\,d\mathbf{x}.$$

다변수 미적분학의 표준 정리다. 작은 정육면체가 $T$로 넓이(부피) $|J_T|$배의 평행육면체로 바뀐다는 것이 핵심이다(그림 5.1).




확률은 보존된다: $f_{\mathbf{X}}(\mathbf{x})\,dx_1dx_2 = f_{\mathbf{Y}}(\mathbf{y})\,|J_T|\,dx_1dx_2$. 그래서 $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})/|J_T(\mathbf{x})| = f_{\mathbf{X}}(H(\mathbf{y}))\,|J_H(\mathbf{y})|$.

**그림 5.1** 변환 $T$는 작은 사각형을 거의 평행사변형으로 보낸다. 넓이의 배율이 야코비안 행렬식의 절댓값이다. 확률은 보존되므로 밀도는 그 역수만큼 바뀐다.

**정리 5.1** (야코비안 변수변환) $\mathbf{X}$가 열린 집합 $S$ 밖에서 0인 밀도 $f_{\mathbf{X}}$를 갖고, $\mathbf{Y} = T(\mathbf{X})$가 위의 조건을 만족하는 변환이며, 역변환 $\mathbf{x} = H(\mathbf{y})$의 야코비안 행렬식이 $J_H(\mathbf{y})$이면

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\big(H(\mathbf{y})\big)\,|J_H(\mathbf{y})|, \qquad \mathbf{y} \in T(S)$$

이고, $T(S)$ 밖에서 0이다.

**증명.** **전략** 어떤 함수가 $\mathbf{Y}$의 밀도라는 것은 모든 보통의 집합 $B$에서 $P(\mathbf{Y} \in B)$가 그 함수의 $B$ 위의 적분이라는 것이다. $\{\mathbf{Y} \in B\}$를 $\mathbf{X}$에 대한 사건으로 바꾸고, 다중적분의 치환을 역변환 $H$에 쓴다 [T7].

① $T(S)$ 안의 보통의 집합 $B$에 대해 $\{\mathbf{Y} \in B\} = \{\mathbf{X} \in H(B)\}$이다. — $T$가 일대일, $H = T^{-1}$ [T1]

② $P(\mathbf{Y} \in B) = \int_{H(B)} f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x}$이다. — 밀도의 정의

③ 위 상자의 치환을 $H$에 쓴다. $H$는 $B$를 $H(B)$로 보낸다.

$$\int_{H(B)} f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x} = \int_{B} f_{\mathbf{X}}\big(H(\mathbf{y})\big)\,|J_H(\mathbf{y})|\,d\mathbf{y}. \qquad \text{다중적분의 치환 [T7]}$$

④ ②③이 모든 $B$에서 성립하므로 적분 안의 함수가 $T(S)$ 위에서 $\mathbf{Y}$의 밀도다. $P(\mathbf{Y} \notin T(S)) = P(\mathbf{X} \notin S) = 0$이므로 $T(S)$ 밖에서는 0으로 둔다. — 1회 정의 6.6의 다변량판

**조건 확인** 셋째 단계에서 상자의 정리를 $H$에 쓰려면 $H$도 연속미분가능하고 야코비안 행렬식이 0이 아니어야 한다. 역함수 정리(다변수 미적분)로 $H$는 연속미분가능하고 $J_H(\mathbf{y}) = 1/J_T\big(H(\mathbf{y})\big) \neq 0$이다. 그래서 $J_T$와 $J_H$ 가운데 계산하기 쉬운 쪽을 구해 역수를 취해도 된다. ∎

$n = 1$이면 2회 정리 4.3이다. 쓰는 순서는 늘 같다. (1) 새 변수 $\mathbf{Y}$를 정하고 필요하면 보조 변수를 더해 변수의 수를 맞춘다. (2) 역변환 $\mathbf{x} = H(\mathbf{y})$를 푼다. (3) 야코비안을 계산한다. (4) 새 영역 $T(S)$를 그린다. (5) 보조 변수를 적분해 없앤다.

## 5.2 극좌표 — 가우스 적분과 Box–Muller

**예제 5.2** (극좌표의 야코비안) $x = r\cos\theta$, $y = r\sin\theta$ $(r > 0,\ 0 < \theta < 2\pi)$이면

$$J = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.$$

극좌표의 넓이 요소 $dx\,dy = r\,dr\,d\theta$가 이것이다.

**정리 5.3** (가우스 적분)

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}.$$

따라서 $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$는 밀도이고, 이 밀도를 갖는 분포를 **표준정규분포** $\mathcal{N}(0,1)$이라 한다.

**증명.** **전략** 한 변수 적분 $I$를 제곱해 평면 위의 이중적분으로 만든다. 극좌표로 바꾸면 넓이 요소의 $r$ 덕분에 원시함수가 생긴다.

① $I = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx$로 두면 $I^2 = \int e^{-x^2/2}dx \int e^{-y^2/2}dy = \iint e^{-(x^2+y^2)/2}\,dx\,dy$이다. Tonelli [T6]

② 극좌표로 바꾼다.

$$\iint e^{-(x^2+y^2)/2}\,dx\,dy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\,r\,dr\,d\theta. \qquad \text{예제 5.2, [T7]}$$

③ 안쪽 적분은 $\int_0^{\infty} e^{-r^2/2} r\,dr = \left[-e^{-r^2/2}\right]_0^{\infty} = 1$이다. 원시함수

④ $I^2 = 2\pi$이고 $I > 0$이므로 $I = \sqrt{2\pi}$이다. ①–③

**조건 확인** 먼저 $I < \infty$이다. $\frac{x^2}{2} \geq |x| - \frac{1}{2}$이므로 $e^{-x^2/2} \leq e^{1/2} e^{-|x|}$이고 오른쪽은 적분 가능하다 [T4]. 극좌표 변환은 $(0,\infty)\times(0,2\pi)$를 평면에서 반직선 $\{y=0,\ x\geq 0\}$을 뺀 집합으로 일대일로 보낸다. 빠진 반직선은 넓이가 0이므로 적분에 영향이 없다. ∎

2회 정리 7.1의 로그정규 밀도가 밀도라는 것($\int \phi = 1$)은 이것으로 증명되었다. 정규분포의 성질은 5회에서 자세히 다룬다. 여기서는 정규 난수를 만드는 방법이 필요하다. 정규분포함수에는 닫힌 꼴의 역함수가 없으므로 2회의 역변환 표집을 바로 쓰기 어렵다. 극좌표가 길을 연다.

**정리 5.4** (Box–Muller (1958)) $U_1, U_2 \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$이고

$$Z_1 = \sqrt{-2\log U_1}\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

이면 $Z_1, Z_2$는 독립인 $\mathcal{N}(0,1)$이다.

**증명.** 거꾸로 간다. $Z_1, Z_2$가 독립인 표준정규이면 결합밀도는 $\frac{1}{2\pi} e^{-(z_1^2+z_2^2)/2}$이다. 극좌표 $(R,\Theta)$로 바꾸면 예제 5.2로

$$f_{R,\Theta}(r,\theta) = \frac{1}{2\pi} e^{-r^2/2}\cdot r = \underbrace{\frac{1}{2\pi}}_{\Theta} \cdot \underbrace{r e^{-r^2/2}}_{R}, \qquad r>0,\ 0<\theta<2\pi$$

이므로 인수분해(정리 3.2)로 $\Theta \sim \mathrm{U}(0,2\pi)$와 $R$은 독립이고 $P(R>r) = e^{-r^2/2}$이다. 2회 정리 4.7로 $R$은 $\sqrt{-2\log U_1}$로, $\Theta$는 $2\pi U_2$로 만들 수 있다($1-U_1$ 대신 $U_1$을 써도 분포가 같다). 이 둘을 다시 직교좌표로 보내면 $(Z_1, Z_2)$이고, 변환이 일대일이므로 분포가 같다. ∎

## 5.3 합성곱 — 독립합의 밀도

**정리 5.5** (합성곱 공식) $X, Y$가 독립이고 밀도 $f_X, f_Y$를 가지면 $Z = X + Y$의 밀도는

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx.$$

이산이면 $p_Z(z) = \sum_x p_X(x)\, p_Y(z - x)$.

**증명.** **전략** 보조 변수 $W = X$를 더해 $(X, Y) \mapsto (W, Z)$를 일대일 변환으로 만든다. 야코비안 공식으로 결합밀도를 구하고 $W$를 적분해 없앤다.

① $W = X$, $Z = X + Y$로 둔다. $T(x, y) = (x,\ x + y)$는 $\mathbb{R}^2$에서 $\mathbb{R}^2$로 가는 일대일 변환이다. — 보조 변수

② 역변환은 $x = w$, $y = z - w$이고 야코비안 행렬식은 $\det\left(\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right) = 1$이다. — 역변환 [T7]

③ $f_{W,Z}(w, z) = f_{X,Y}(w,\ z - w) \cdot 1 = f_X(w)\, f_Y(z - w)$이다. — 정리 5.1, 3.2

④ $w$를 적분해 없애면 $f_Z(z) = \int f_X(w)\, f_Y(z - w)\, dw$이다. — 정리 1.6

⑤ 이산이면 $\{Z = z\} = \bigcup_x \{X = x,\ Y = z - x\}$는 서로소인 합집합이므로 $p_Z(z) = \sum_x p_X(x)\, p_Y(z - x)$이다. — (P3), 독립

■

2회 예제 3.8의 `np.convolve`가 이산 공식이다. 이름은 다항식 곱셈에서 왔다. 확률생성함수의 곱의 계수가 합성곱이다.

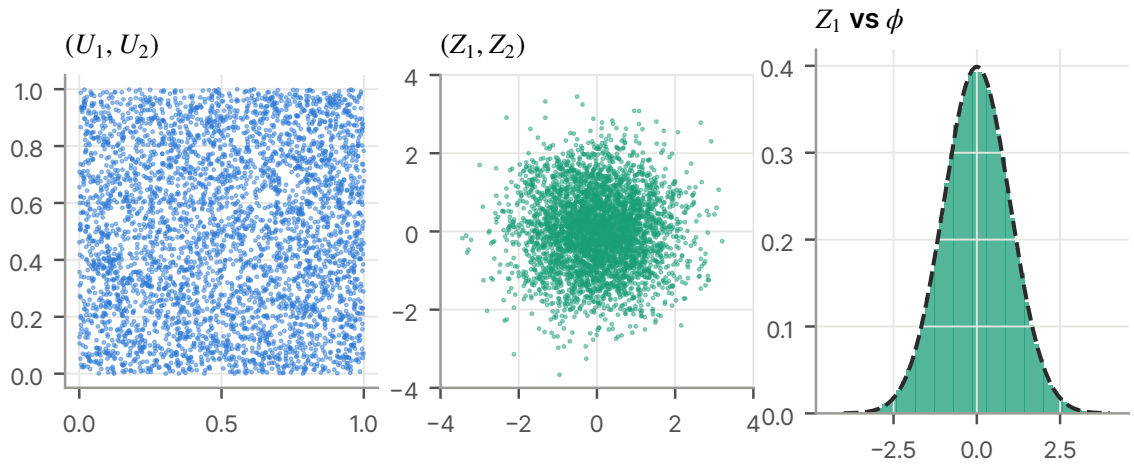



**그림 5.2** Box–Muller. 정사각형의 균등한 점 $(U_1, U_2)$(왼쪽)가 반지름 $\sqrt{-2\log U_1}$, 각 $2\pi U_2$의 점(가운데)으로 가면 두 좌표가 독립인 표준정규다(오른쪽, 점선은 $\phi$).

**예제 5.6** (균등분포의 합 — Simpson의 삼각형) $X, Y \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$이면 $f_X(x) f_Y(z-x) = 1$인 것은 $0 < x < 1$이고 $0 < z - x < 1$일 때뿐이므로

$$f_Z(z) = \int \mathbf{1}\{\max(0, z-1) < x < \min(1, z)\}\, dx = \begin{cases} z, & 0 < z \le 1, \\ 2 - z, & 1 < z < 2. \end{cases}$$

삼각형 분포다. Simpson(1757)이 관측 오차의 모형으로 쓴 것이 이 삼각형이다. 셋을 더하면 2차 다항식 조각으로 이어진 종 모양이 되고, 더할수록 정규분포를 닮아 간다(그림 5.3, 7회 중심극한정리).

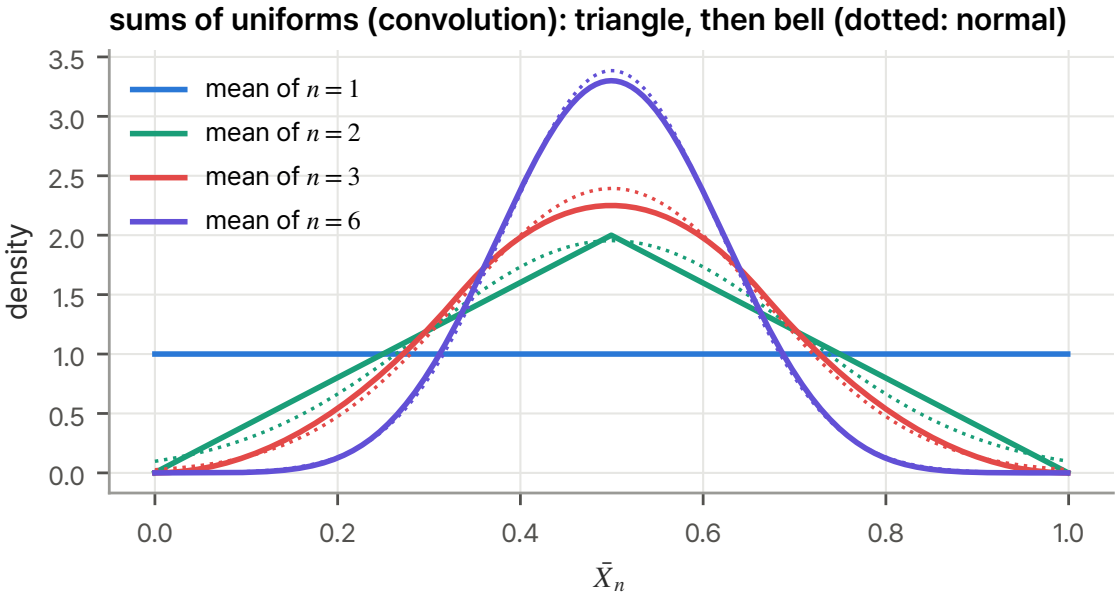



**그림 5.3** 독립인 $\mathrm{U}(0,1)$ $n$개의 합의 밀도(합성곱을 수치로 되풀이). 같은 평균과 분산의 정규 밀도(점선)에 빠르게 다가간다.

**예제 5.7** (지수분포의 합) $X, Y \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda)$이면 $z > 0$에서

$$f_Z(z) = \int_0^z \lambda e^{-\lambda x}\, \lambda e^{-\lambda(z-x)}\, dx = \lambda^2 z e^{-\lambda z}.$$

두 번째 사건을 기다리는 시간의 분포다. $n$개를 더하면 $\frac{\lambda^n z^{n-1}}{(n-1)!} e^{-\lambda z}$이고(연습 5.3), 4회에서 이것을 감마분포라 부른다. mgf로도 확인된다. $\left(\frac{\lambda}{\lambda - t}\right)^n$이 그 mgf이기 때문이다(정리 3.4(d)).

## 5.4 비, 그리고 다시 Cauchy

**정리 5.8** (정규의 비는 Cauchy) $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이면 $C = Z_1/Z_2$는 Cauchy 분포 $f(c) = \frac{1}{\pi(1+c^2)}$를 따른다.

**증명.** **전략** 보조 변수 $W = Z_2$를 더해 합성곱 공식(정리 5.5)과 같은 순서로 계산한다. 야코비안이 $|w|$이어서 $w$에 대한 적분에 원시함수가 생긴다. 단계별 증명은 연습 5.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 54쪽에 있다.

2회 예제 1.8에서 기댓값이 없던 Cauchy 분포가 정규분포 두 개의 비로 나타났다. 분모가 0 근처에 갈 확률이 작지 않아서 비가 자주 크게 튄다. 측정한 두 양의 비율(효과 크기의 비, 신호 대 잡음)이 꼬리가 무거운 이유가 여기 있다.

## 5.5 선형변환

**따름정리 5.9** (선형변환의 밀도) $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$이고 $A$가 가역인 $n \times n$ 행렬이면

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det A|}\, f_{\mathbf{X}}\big(A^{-1}(\mathbf{y} - \mathbf{b})\big).$$

**증명.** 역변환 $H(\mathbf{y}) = A^{-1}(\mathbf{y} - \mathbf{b})$의 야코비안 행렬은 $A^{-1}$이고 $\det A^{-1} = 1/\det A$이다. ∎

이것이 5회의 다변량 정규분포의 밀도를 유도하는 데 쓰이는 공식이다. 독립인 표준정규 $\mathbf{Z}$에서 $\mathbf{X} = L\mathbf{Z} + \boldsymbol{\mu}$로 만들면 밀도에 $1/|\det L|$이 붙는다.

## 연습문제

**연습 5.1** (극좌표의 반대 방향 ★) $(R,\Theta)$의 결합밀도가 $re^{-r^2/2}/(2\pi)$이면 $(R\cos\Theta, R\sin\Theta)$가 독립인 표준정규임을 정리 5.1으로 직접 보여라. 풀이 53쪽

**연습 5.2** (곱과 몫 ★★) $X, Y \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$일 때 $V = XY$의 밀도가 $-\log v$ $(0<v<1)$임을 보조변수 $W = X$로 보여라. 풀이 53쪽

**연습 5.3** (감마의 합 ★★) 예제 5.7을 귀납법으로 넓혀 $n$개의 독립인 $\mathrm{Exp}(\lambda)$의 합의 밀도가 $\frac{\lambda^n z^{n-1}}{(n-1)!}e^{-\lambda z}$임을 보여라. 풀이 53쪽

**연습 5.4** (Cauchy의 평균 ★★★) $C_1, C_2$가 독립인 Cauchy이면 $\frac{1}{2}(C_1+C_2)$도 Cauchy임을 합성곱으로 보여라. (힌트: $\frac{1}{(1+x^2)(1+(z-x)^2)}$를 부분분수로 나누거나, 결과 $f_{C_1+C_2}(z) = \frac{2}{\pi(4+z^2)}$를 확인한다.) 평균을 내도 분포가 좁아지지 않는다. 풀이 53쪽

**연습 5.5** (균등 셋의 합 ★★) $\mathrm{U}(0,1)$ 세 개의 합의 밀도를 합성곱으로 구하라. $0<z<1$에서 $\frac{z^2}{2}$, $1\le z<2$에서 $\frac{-2z^2+6z-3}{2}$, $2\le z<3$에서 $\frac{(3-z)^2}{2}$이다. 풀이 53쪽

**연습 5.6** (극좌표에서 거리 ★) Box–Muller의 증명에서 $R^2$의 분포를 구하라. 그것은 평균이 2인 지수분포다. (4회에서 이것이 자유도 2인 $\chi^2$ 분포임을 본다.) 풀이 54쪽

**연습 5.7** (정규의 비 증명 ★★ 유도형) 정리 5.8를 다음 순서로 증명하라.

(a) $S = \{(z_1,z_2) : z_2 \neq 0\}$로 두면 $P\big((Z_1,Z_2)\in S\big) = 1$이다. $T(z_1,z_2) = (z_1/z_2,\ z_2)$가 $S$에서 일대일임을 보이고 역변환과 그 야코비안 행렬식을 구하라.

(b) 정리 5.1으로 $(C,W)$의 결합밀도를 써라. 절댓값 $|w|$는 어디서 오는가?

(c) $w$에 대해 적분해 $f_C(c) = \frac{1}{\pi(1+c^2)}$을 얻어라. (힌트: 적분할 함수가 $w$에 대해 짝함수이므로 $2\int_0^\infty$로 쓰고, $u = \frac{1}{2}(1+c^2)w^2$으로 치환한다.)

(d) $\int_{-\infty}^{\infty} \frac{dc}{\pi(1+c^2)} = 1$을 확인해 (c)가 밀도임을 점검하라.

풀이 54쪽

# 6

# 평균의 정확도 — $\sqrt{n}$의 법칙

이제 여는 질문에 답할 도구가 모두 있다. 이 절은 확률벡터의 평균과 공분산행렬로 선형결합의 분산을 한 번에 계산하는 공식을 세우고, 그것으로 $\sqrt{n}$의 법칙을 증명한다. 그다음 역사로 간다. Simpson이 평균의 이득을 처음 계산한 방법, 그리고 $\sqrt{n}$을 몰랐던 Pyx 재판이다. 마지막으로 관측들이 서로 닮아 있으면 $\sqrt{n}$이 어떻게 무너지는지 본다.

## 6.1 평균벡터와 공분산행렬

**정의 6.1** (평균벡터와 공분산행렬) 확률벡터 $\mathbf{X} = (X_1, \dots, X_n)'$의 성분이 모두 유한한 분산을 가질 때

$$\boldsymbol{\mu} = E\mathbf{X} = (EX_1, \dots, EX_n)', \qquad \Sigma = \operatorname{Cov}\mathbf{X} = E\big[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\big], \quad \Sigma_{ij} = \operatorname{Cov}(X_i, X_j)$$

를 **평균벡터**와 **공분산행렬**이라 한다. 확률행렬의 기댓값은 성분마다 취한다.

**정리 6.2** (선형변환의 평균과 공분산) $A$가 $m \times n$ 상수 행렬, $\mathbf{b}$가 상수 벡터이면

$$E(A\mathbf{X}+\mathbf{b}) = A\boldsymbol{\mu}+\mathbf{b}, \qquad \operatorname{Cov}(A\mathbf{X}+\mathbf{b}) = A\Sigma A'.$$

특히 $\operatorname{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a} \geq 0$이므로 공분산행렬은 대칭이고 양의 준정부호다.

**증명.** **전략** 행렬 곱을 성분으로 풀어 쓰고, 성분마다 기댓값의 선형성(따름정리 1.9)을 쓴다. 단계별 증명은 연습 6.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 55쪽에 있다.

정리 4.2(c)를 행렬로 쓴 것이다. 5회의 다변량 정규분포 전체가 이 한 줄 위에 선다.

## 6.2 $\sqrt{n}$의 법칙

**정리 6.3** (표본평균의 분산) $X_1, \dots, X_n$의 평균이 모두 $\mu$, 분산이 모두 $\sigma^2$이고 **쌍마다 상관이 없으면**

$$E\bar{X} = \mu, \qquad \operatorname{Var}\bar{X} = \frac{\sigma^2}{n}, \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

**증명.** $\bar{X} = \mathbf{a}'\mathbf{X}$, $\mathbf{a} = \frac{1}{n}\mathbf{1}$이고 $\Sigma = \sigma^2 I$이므로 정리 6.2로 $\operatorname{Var}\bar{X} = \frac{1}{n^2}\mathbf{1}'(\sigma^2 I)\mathbf{1} = \frac{\sigma^2}{n}$. ∎

독립이면 상관이 없으므로(정리 4.2(d)) iid 표본에서 언제나 성립한다. 정확도를 두 배로 하려면 관측을 네 배로 해야 한다. $\sigma/\sqrt{n}$을 평균의 **표준오차**라 한다. 6회에서 $\sigma$를 모를 때 이것을 추정하는 법을 배운다.

## 6.3 Simpson의 계산 (1755, 1757)

**역사** 수학자 Thomas Simpson은 1755년 왕립학회 회장 Macclesfield 백작에게 보낸 편지 "실용 천문학에서 여러 관측의 평균을 취하는 이점에 대하여"에서, 관측 오차가 몇 개의 정수 값을 정해진 확률로 갖는다고 놓고 여섯 관측의 평균의 오차 분포를 계산했다. 계산 도구는 de Moivre의 생성함수, 곧 합성곱이었다(2회 예제 3.8). 1757년판에서는 오차의 확률이 0에서 멀어질수록 일정하게 줄어드는 삼각형 모형을 썼고, 그 연속판으로 예제 5.6의 삼각형 분포를 처음 다뤘다. 오차를 확률변수로 본 첫 계산이었다.

**예제 6.4** (삼각형 오차, 여섯 번의 평균) 오차 $E$가 $-5, \dots, 5$(초)의 정수 값을 확률 $\frac{6-|e|}{36}$로 갖는다고 하자(1757년의 모형). 한 번 관측의 오차가 1초 이내일 확률은 $\frac{5+6+5}{36} \approx 0.444$이다. 독립인 여섯 오차의 합의 분포는 합성곱을 다섯 번 하면 나오고(정리 5.5), 평균의 오차가 1초 이내일 확률, 곧 합이 $-6$과 6 사이일 확률은

$$P\big(|\bar{E}_6| \le 1\big) \approx 0.725$$

이다. 2초 이내라면 한 번은 $\frac{24}{36} \approx 0.667$, 평균은 약 0.967이다. $\operatorname{Var} E = \frac{35}{6}$이므로 평균의 표준편차는 정리 6.3로 $\sqrt{35/36} \approx 0.99$초로, 한 번의 2.42초의 $1/\sqrt{6}$이다. Simpson의 결론은 "평균을 취하면 오차가 작을 확률이 크게 늘어난다"였다(그림 6.1).

## 6.4 Pyx 재판 — $n$과 $\sqrt{n}$

**역사** 영국 왕립 조폐국은 늦어도 13세기부터 "Pyx 재판"을 열었다. 일정 기간 만든 금화 가운데 일부를 상자(pyx)에 모아 두었다가, 독립적인 배심원 앞에서 무게와 순도를 검사했다. 조폐국장은 계약으로 정해진 허용오차(remedy) 안에 들면 무사했다. 문제는 그 허용오차의 크기였다. 금화 한 개의 허용오차를 $R$이라 할 때, $n$개를 함께 잰 총무게의 허용오차는 $nR$로 정해졌다. Stigler(1977)는 이것이 수백 년 동안 $\sqrt{n}$이 아니라 $n$에 비례했음을 지적했다.

1848년 재판의 기록(HistData `Pyx`)은 금화 10,000개 하나하나를 표준 무게와 비교해 여덟 구간으로 분류했다. 한 개의 허용오차는 $R = 0.25625$ 그레인이다. 구간 자료에 정규분포를 맞추면 동전 하나의 표준편차는 약 $\sigma \approx 0.13$ 그레인, 곧 $R \approx 2\sigma$이다(노트북). 금화 10,000개 가운데 $\pm R$을 벗어난 것은 4.5%였다.

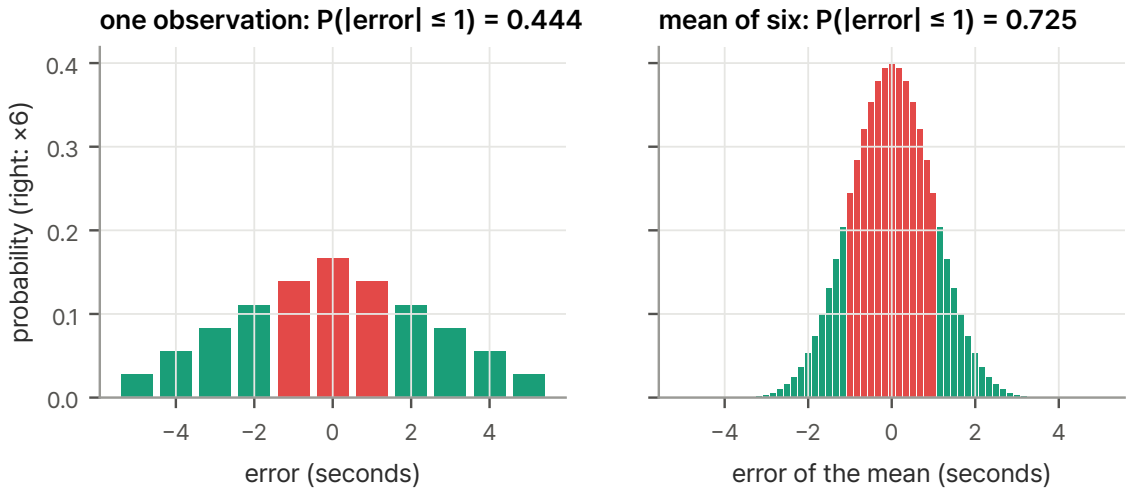



**그림 6.1** Simpson(1757)의 삼각형 오차 모형. 한 번 관측의 오차 분포(왼쪽)와 여섯 번 평균의 오차 분포(오른쪽, 합성곱 다섯 번). 빨간 막대가 1초 이내다.

이제 조폐국장이 동전마다 $\delta$씩 가볍게 만든다고 하자. 총무게의 편차는 평균 $-n\delta$, 표준편차 $\sigma\sqrt{n}$이다(정리 6.3). 허용오차가 $nR$이면 검사를 통과하려면 대략 $n\delta + 3\sigma\sqrt{n} < nR$, 곧

$$\delta < R - \frac{3\sigma}{\sqrt{n}}$$

이면 충분하다. $n = 1{,}000$이면 오른쪽은 $0.256 - 0.012 \approx 0.244$그레인이다. 동전마다 허용오차의 95%를 깎아도 거의 확실히 통과한다.

허용오차가 $\sqrt{n}$에 비례해 $3\sigma\sqrt{n}$이었다면 어떨까? 통과 조건 $|-n\delta + \sigma\sqrt{n}\,Z| \le 3\sigma\sqrt{n}$은 $|Z - \sqrt{n}\,\delta/\sigma| \le 3$과 같으므로, 통과 확률이 절반을 넘으려면 대략 $\delta \le 3\sigma/\sqrt{n} \approx 0.012$그레인이어야 한다. 깎을 수 있는 양이 20분의 1로 준다. $n$에 비례하는 허용오차는 동전 하나의 기준을 $n$개에 그대로 늘린 것이지만, 무작위 오차는 서로 상쇄되어 $\sqrt{n}$으로만 자란다. 상쇄되지 않는 것은 체계적인 $\delta$뿐이다(그림 6.2).

이 틈이 실제로 이용되었는지에 대해 Stigler는 조심스럽다. 1699년부터 1727년까지 조폐국장이었던 Isaac Newton도 이 제도 아래에서 일했다. 분명한 것은, de Moivre(1733)가 이항분포의 흩어짐이 $\sqrt{n}$에 비례함을 보인 뒤에도 이 사실이 검사 제도에 들어가기까지 오랜 시간이 걸렸다는 것이다.

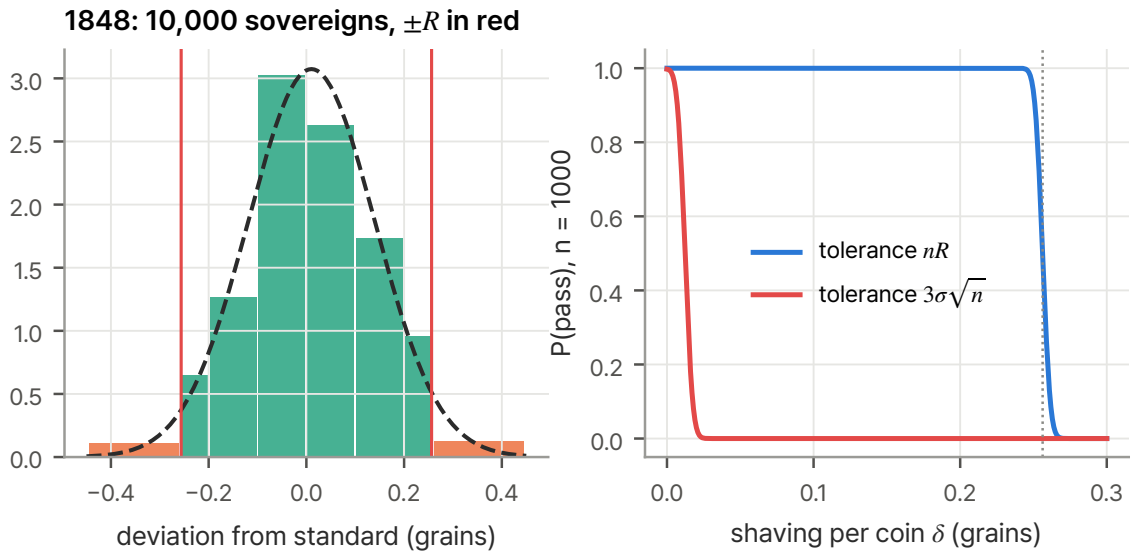



**그림 6.2** 왼쪽: 1848년 Pyx 재판의 금화 10,000개의 무게 편차(구간 자료)와 맞춘 정규 밀도. 오른쪽: 동전마다 $\delta$씩 가볍게 만들 때 $n = 1{,}000$개 총무게가 검사를 통과할 확률. 허용오차가 $nR$이면(파랑) $\delta$가 $R$ 가까이 가도 통과하고, $3\sigma\sqrt{n}$이면(빨강) 조금만 깎아도 걸린다.

## 6.5 상관된 관측 — 유효 표본 크기

정리 6.3은 "상관이 없으면"을 요구한다. 같은 날, 같은 기계, 같은 관측자로 잰 값들은 서로 닮는다.

**정리 6.5** (같은 상관을 갖는 관측의 평균) $\operatorname{Var} X_i = \sigma^2$이고 $i \neq j$이면 $\operatorname{Corr}(X_i, X_j) = \rho$일 때

$$\operatorname{Var} \bar{X} = \frac{\sigma^2}{n}\big(1 + (n-1)\rho\big) = \sigma^2\Big(\rho + \frac{1-\rho}{n}\Big) \longrightarrow \rho\,\sigma^2 \quad (n \to \infty).$$

**증명.** 정리 4.2(c)로 $\operatorname{Var}\sum X_i = n\sigma^2 + n(n-1)\rho\sigma^2$이고 $n^2$으로 나눈다. ∎

$\rho > 0$이면 관측을 아무리 늘려도 평균의 분산은 $\rho\sigma^2$ 아래로 내려가지 않는다. 같은 분산을 독립인 관측으로 얻으려면 몇 개가 필요한가? $\sigma^2/n_{\text{eff}} = \operatorname{Var}\bar{X}$를 풀면

$$n_{\text{eff}} = \frac{n}{1 + (n-1)\rho} \longrightarrow \frac{1}{\rho}.$$

$\rho = 0.1$이면 1,000개의 관측이 독립인 관측 10개만큼의 정보밖에 담지 못한다.

**예제 6.6** (이웃끼리 닮은 관측 — AR(1)) 시간 순서대로 잰 관측에서는 가까운 관측일수록 닮는다. $\operatorname{Corr}(X_i, X_j) = \rho^{|i-j|}$ $(0 < \rho < 1)$이면(**AR(1)** 모형, 2회 7.4절) 정리 4.2(c)로

$$\operatorname{Var}\bar{X} = \frac{\sigma^2}{n^2}\Big(n + 2\sum_{k=1}^{n-1}(n-k)\rho^k\Big) \approx \frac{\sigma^2}{n}\cdot\frac{1+\rho}{1-\rho} \qquad (n\text{이 클 때})$$

이다(연습 6.5). $\rho = 0.9$이면 분산이 19배, 표준오차가 4.4배 커지고, $n_{\text{eff}} \approx n/19$이다. 2회 7.4절의 시뮬레이션에서 본 4.36배가 $\sqrt{19}$이다. MCMC(16회)의 표본이 서로 상관되어 있어 "유효 표본 크기"를 따로 보고하는 이유다.

### 연습문제

**연습 6.1** (공분산행렬 ★) $X_1, X_2, X_3$이 독립이고 분산이 1일 때 $\mathbf{Y} = (X_1 + X_2,\ X_2 + X_3)'$의 공분산행렬을 정리 6.2로 구하고, 두 성분의 상관계수를 구하라. 풀이 54쪽

**연습 6.2** (양의 준정부호 ★★) $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$이 공분산행렬이 될 수 있는 $\rho$의 범위를 $\mathbf{a}'\Sigma\mathbf{a} \geq 0$에서 구하라. $3 \times 3$에서 모든 비대각 성분이 $\rho$일 때의 범위는 $-\frac{1}{2} \leq \rho \leq 1$임을 보여라. 풀이 54쪽

**연습 6.3** (Simpson의 1755년 모형 ★) 오차가 $-5, \dots, 5$의 정수를 같은 확률로 가지면 한 번의 오차가 1초 이내일 확률과 여섯 번 평균의 오차가 1초 이내일 확률은 얼마인가? 노트북의 합성곱으로 구하라. 풀이 54쪽

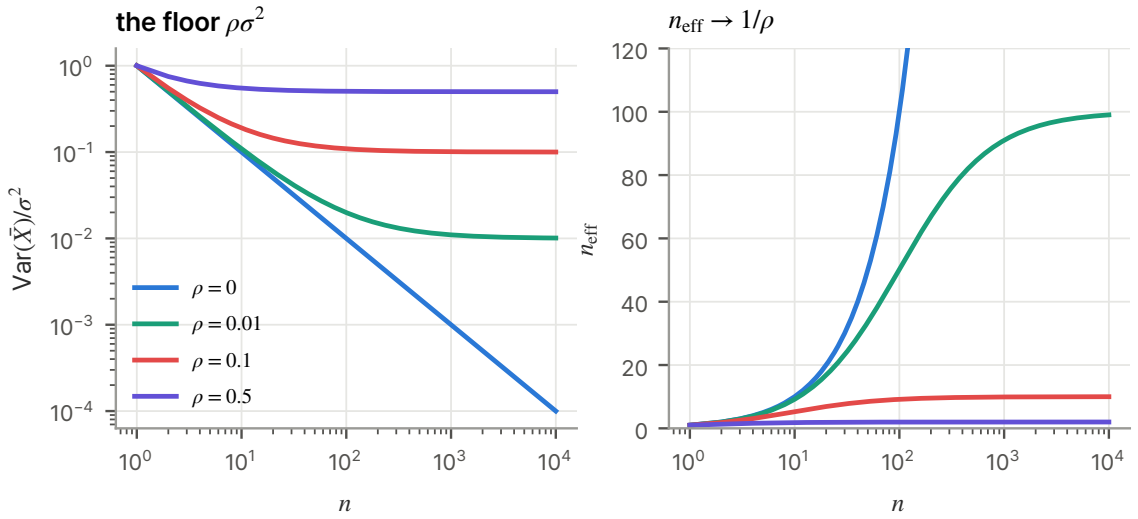



**그림 6.3** 왼쪽: 같은 상관 $\rho$를 갖는 관측의 평균의 분산(정리 6.5). $\rho > 0$이면 $\rho\sigma^2$에서 바닥에 닿는다. 오른쪽: 유효 표본 크기 $n_{\mathrm{eff}}$는 $1/\rho$을 넘지 못한다.

**연습 6.4** (Pyx 재판의 이득 ★★) $\sigma = 0.13$, $R = 0.25625$, $n = 1{,}000$이다. 허용오차 $nR$ 아래에서 통과 확률이 0.99 이상인 가장 큰 $\delta$를 정규근사로 구하라. 1년에 금화 100만 개를 만든다면 조폐국장이 남길 수 있는 금은 몇 그레인인가? 풀이 54쪽

**연습 6.5** (AR(1)의 분산 ★★) 예제 6.6의 식을 유도하고, $n\,\mathrm{Var}\,\bar{X}/\sigma^2 \to \frac{1+\rho}{1-\rho}$임을 보여라. (힌트: $\sum_{k\ge 1}\rho^k = \frac{\rho}{1-\rho}$와 $\frac{1}{n}\sum_{k} k\rho^k \to 0$.) 풀이 55쪽

**연습 6.6** (음의 상관 ★★) $U \sim \mathrm{U}(0,1)$로 $\int_0^1 g$를 추정할 때 $\frac{1}{2}\big(g(U)+g(1-U)\big)$를 쓰는 것을 **대조변량법**(antithetic variates)이라 한다. $g$가 단조이면 $\mathrm{Cov}(g(U), g(1-U)) \le 0$임을 받아들이고, 이 추정량의 분산이 독립인 두 균등 난수를 쓴 것보다 작거나 같음을 보여라. $g(x) = e^x$에서 두 분산을 계산하라. 풀이 55쪽

**연습 6.7** (선형변환의 평균과 공분산 증명 ★ 유도형) 정리 6.2를 다음 순서로 증명하라.

(a) $(A\mathbf{X}+\mathbf{b})_i = \sum_j A_{ij}X_j + b_i$에 선형성을 써서 $E(A\mathbf{X}+\mathbf{b}) = A\boldsymbol{\mu}+\mathbf{b}$를 보여라.

(b) $A\mathbf{X}+\mathbf{b} - E(A\mathbf{X}+\mathbf{b}) = A(\mathbf{X}-\boldsymbol{\mu})$를 확인하라.

(c) 행렬 $A(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'A'$의 $(i,k)$ 성분이 $\sum_j\sum_l A_{ij}(X_j-\mu_j)(X_l-\mu_l)A_{kl}$임을 보이고, 기댓값을 취해 $(A\Sigma A')_{ik}$를 얻어라.

(d) $A = \mathbf{a}'$로 두어 $\mathrm{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a} \ge 0$을 얻고, $\Sigma$가 대칭인 이유를 말하라.

풀이 55쪽

# 7

# 부숴보기 — 평균이 통하지 않을 때

## 7.1 확률 0인 사건에 조건을 걸면 — Borel의 역설

1회 7.3절은 연속 확률변수에서 $P(X = x) = 0$이어서 "$X = x$일 때"를 1회 정의 5.1로 말할 수 없다고 했다. 정의 2.2는 밀도의 비로 이 문제를 피했다. 그러나 그 정의에는 대가가 있다. **어떤 변수로** 조건을 거느냐에 따라, 같은 확률 0인 사건에 대한 조건부분포가 달라질 수 있다.

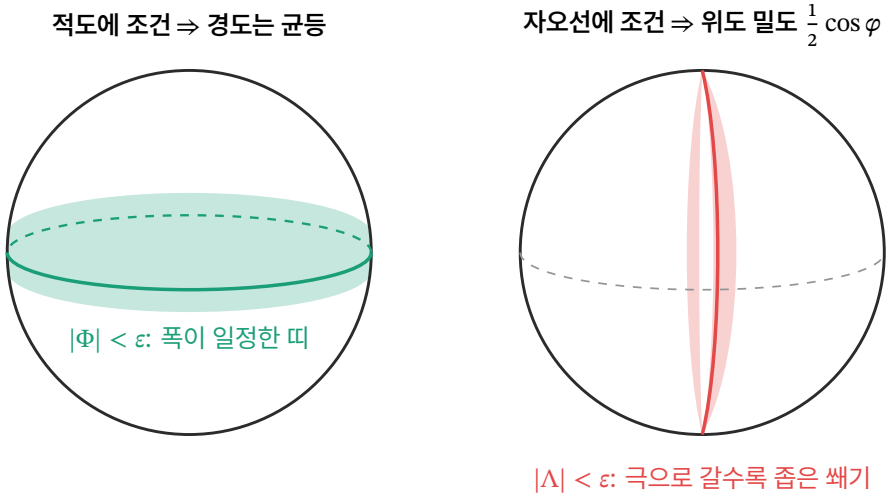



**그림 7.1** Borel의 역설. 적도 근처의 띠 $\{|\Phi| < \varepsilon\}$(왼쪽)는 폭이 일정하지만, 자오선 근처의 쐐기 $\{|\Lambda| < \varepsilon\}$(오른쪽)는 극으로 갈수록 좁아진다. 두 극한이 서로 다른 조건부분포를 준다.

**예제 7.1** (구 위의 점과 대원 (Borel 1909, Kolmogorov 1933)) 구면 위에 고르게 퍼진 점을 경도 $\Lambda \in (-\pi, \pi]$와 위도 $\Phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$로 나타내자. 넓이 요소가 $\cos\varphi \, d\lambda \, d\varphi$이므로(극좌표의 구면판) 결합밀도는

$$f(\lambda, \varphi) = \frac{\cos\varphi}{4\pi},$$

인수분해로 $\Lambda \sim \mathrm{U}(-\pi, \pi)$와 $\Phi$(밀도 $\frac{1}{2}\cos\varphi$)는 독립이다.

**적도에 조건을 걸면.** $\Phi = 0$일 때 $\Lambda$의 조건부 밀도는 $f(\lambda, 0)/f_\Phi(0) = \frac{1}{2\pi}$, 곧 적도 위에서 균등하다. 대칭으로 보아 당연하다.

**자오선에 조건을 걸면.** $\Lambda = 0$(그리고 반대편 $\Lambda = \pi$)일 때 $\Phi$의 조건부 밀도는 $f(0, \varphi)/f_\Lambda(0) = \frac{1}{2}\cos\varphi$, 곧 자오선이라는 대원 위에서 **균등하지 않다.** 극 근처보다 적도 근처에 몰린다.
그런데 적도와 자오선은 회전으로 겹치는 같은 종류의 대원이다. "점이 이 대원 위에 있다는 조건에서의 분포"는 하나여야 할 것 같은데 답이 둘이다.

해결은 2.1절의 극한에 있다. 조건부 밀도는 "확률이 양수인 근처의 사건"에 조건을 걸고 극한을 취한 것이었다. $\Phi = 0$의 근처는 폭이 일정한 띠지만, $\Lambda = 0$의 근처는 극으로 갈수록 좁아지는 쐐기다(그림 7.1). 쐐기에는 적도 근처의 점이 더 많이 들어가므로 극한도 그쪽으로 기운다. **확률 0인 사건에 대한 조건부분포는 그 사건 하나로 정해지지 않고, 그 사건을 어떤 확률변수의 값으로 보느냐로 정해진다.** Kolmogorov(1933)는 이 역설을 근거로 조건부기댓값을 "사건"이 아니라 "확률변수"($\sigma$-대수)에 대해 정의했다. 이 과정에서는 늘 "어떤 변수의 값으로 조건을 거는지"를 밝힌다.

## 7.2 상관이 0이어도

정리 6.3은 상관만 없으면 되었다. 그러나 상관이 없는 것과 독립인 것은 다르다(예제 4.7). 평균의 **분산**은 상관만 보지만, 평균의 **분포**는 그렇지 않다. $X \sim \mathrm{U}(-1,1)$로 $X_1 = X$, $X_2 = X^2 - \frac{1}{3}$을 만들면 둘은 상관이 없고 $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}\,X_1 + \mathrm{Var}\,X_2$이다. 그러나 $X_1$을 알면 $X_2$가 정해지므로 mgf가 곱이 되지 않고(정리 3.4(d)는 독립을 요구한다), 7회의 중심극한정리도 상관이 없다는 것만으로는 성립하지 않는다.

## 7.3 평균으로 줄지 않는 오차 — Michelson과 Cavendish

$\sqrt{n}$의 법칙은 오차의 **흩어짐**을 줄인다. 오차의 **평균**, 곧 측정 방법 자체의 치우침(편향)은 줄이지 못한다. $X_i = \theta + b + \varepsilon_i$이면 $\bar{X} \to \theta + b$이다.

> **예제 7.2** (빛의 속도와 지구의 밀도)
>
> **Michelson (1879).** Albert Michelson은 1879년 6–7월 Annapolis에서 공기 중의 빛의 속도를 100번 쟀다(HistData `Michelson`, km/s에서 299,000을 뺀 값). 평균은 852.4, 표준편차 79.0이므로 평균의 표준오차는 7.9이다. 오늘의 값을 같은 척도로 옮긴 "참값"은 734.5이다(Stigler 1977). 평균은 참값에서 표준오차의 **15배** 떨어져 있다. 100번의 평균이 우연히 이렇게 빗나갈 수는 없다. 측정에 체계적 오차가 있었다.
>
> **Cavendish (1798).** Henry Cavendish는 비틀림 저울로 지구의 평균 밀도를 29번 쟀다(HistData `Cavendish`, 오기 하나를 바로잡은 `density2`). 평균은 5.482, 표준오차는 0.038이고 오늘의 값 5.513과의 차이는 표준오차의 0.8배다. 이쪽은 편향의 증거가 없다.
>
> 두 실험 모두 관측을 늘리면 표준오차는 줄었을 것이다. 그러나 Michelson의 평균은 아무리 관측을 늘려도 734.5로 가지 않는다. 표준오차는 "같은 방법으로 다시 재면 얼마나 달라질까"를 말할 뿐, "참값에서 얼마나 먼가"를 말하지 않는다.

## 7.4 평균을 내도 좁아지지 않는다 — Cauchy

정리 5.8의 Cauchy 분포에는 분산이 없으므로 정리 6.3을 쓸 수 없다. 실제로 두 Cauchy의 평균은 다시 같은 Cauchy다(연습 5.4). 측정 오차가 "두 정규 잡음의 비"처럼 생긴 장치라면, 평균을 내는 것은 아무 소용이 없다. 이럴 때 쓸 수 있는 것은 중앙값이다(15회).

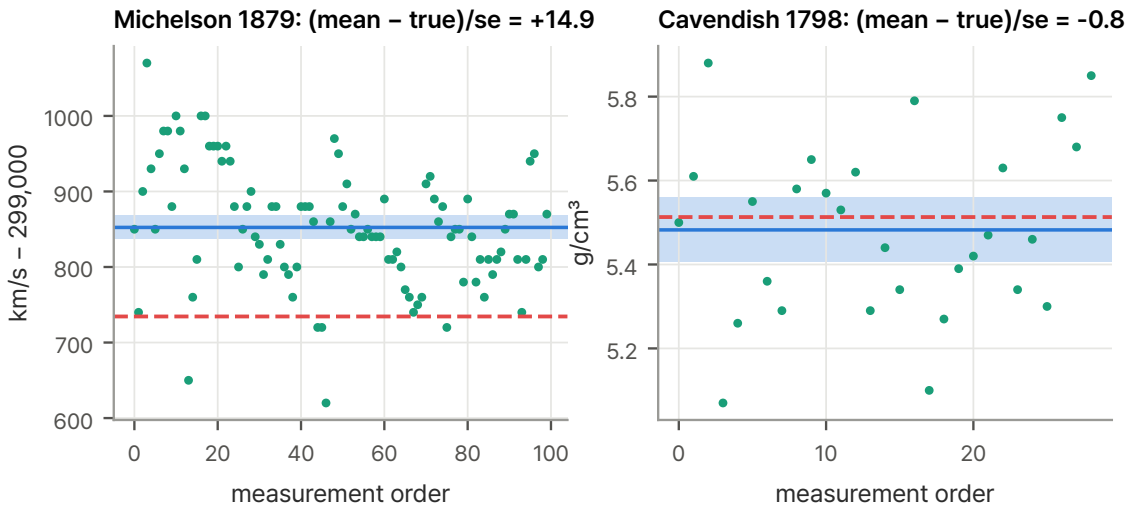



**그림 7.2** Michelson(왼쪽)과 Cavendish(오른쪽)의 측정값, 평균 ±2 표준오차(파랑), 오늘의 값(빨간 점선). 표준오차는 흩어짐만 잰다.

### 연습문제

**연습 7.1** (위도의 분포 ★) 예제 7.1에서 $\Phi$의 주변 밀도가 $\frac{1}{2}\cos\varphi$임을 보이고, 1회 연습 3.3의 "넓이에 고르게"의 답 $P(\Phi \ge 60°) = \frac{1-\sqrt{3}/2}{2}$을 다시 구하라. 풀이 55쪽

**연습 7.2** (같은 직선, 다른 조건 ★★) $X, Y \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$일 때, 사건 $\{X = Y\}$에 대해 (a) $D = X - Y$로 보고 $D = 0$에 조건을 건 $X$의 조건부분포, (b) $Q = Y/X$로 보고 $Q = 1$에 조건을 건 $X$의 조건부 밀도를 구하라. (a)는 $\mathcal{N}(0, \frac{1}{2})$, (b)는 밀도가 $|x|e^{-x^2}$이다. 같은 직선 $\{X = Y\}$에서 답이 다르다. 풀이 55쪽

**연습 7.3** (편향과 분산 ★) $X_i = \theta + b + \varepsilon_i$, $\varepsilon_i$ iid 평균 0 분산 $\sigma^2$일 때 $E(\bar{X} - \theta)^2 = b^2 + \sigma^2/n$임을 보여라. Michelson의 자료에서 $b$와 $\sigma/\sqrt{n}$의 크기를 비교하라. 풀이 55쪽

**연습 7.4** (Michelson의 다섯 묶음 ★★) Stigler는 Michelson의 100개를 시간 순서대로 20개씩 다섯 묶음으로 나눴다(`MichelsonSets`). 묶음 평균의 표준편차를 계산하고, 관측이 독립이라면 기대되는 $79/\sqrt{20} \approx 17.7$과 비교하라. 노트북으로 계산하고 무엇을 결론지을 수 있는지 말하라. 풀이 56쪽

# 8

# Beautiful insights — 같은 계산, 오늘의 문제

## 8.1 신경망의 초기화 — 분산을 지키기

신경망의 한 층은 입력 $x_1, \dots, x_n$에 가중치를 곱해 더한다: $s = \sum_{j=1}^{n} w_j x_j$. 처음에 가중치를 무작위로 뽑는다. $w_j$들이 서로 독립이고 입력과도 독립이며 평균 0, 분산 $\sigma_w^2$이고, 입력이 평균 0, 분산 $v$라 하자. 그러면 정리 3.4와 $E(w_j x_j) = 0$으로

$$\operatorname{Var} s = \sum_{j=1}^{n} \operatorname{Var}(w_j x_j) = \sum_{j=1}^{n} E w_j^2 \, E x_j^2 = n \, \sigma_w^2 \, v.$$

층을 지날 때마다 분산이 $n\sigma_w^2$배가 된다. 층이 50개이면 $(n\sigma_w^2)^{50}$배다. 1보다 조금만 커도 폭발하고 조금만 작아도 사라진다. 그래서 $\sigma_w^2 = \frac{1}{n}$으로 놓는다(LeCun 1998, Glorot–Bengio 2010). 활성화 함수가 ReLU $\max(0, s)$이면 대칭인 $s$에 대해 $E \max(0,s)^2 = \frac{1}{2} E s^2$이므로 절반을 잃는다. 그것을 메우는 $\sigma_w^2 = \frac{2}{n}$이 He 초기화(2015)다. 2회의 선형성과 3회의 독립합의 분산만으로 얻은 처방이다.

## 8.2 재매개화 — Box–Muller의 생각

정리 5.4의 핵심은 "복잡한 분포의 난수를 고정된 단순한 난수의 함수로 쓴다"는 것이다. 정규분포 $\mathcal{N}(\mu, \sigma^2)$의 난수는 $\mu + \sigma Z$, $Z \sim \mathcal{N}(0,1)$이다(2회 예제 4.5). 변분 오토인코더(VAE, Kingma–Welling 2014)는 이것을 **재매개화 기법**이라 부른다. $E_{X \sim \mathcal{N}(\mu,\sigma^2)}[g(X)] = E_Z[g(\mu + \sigma Z)]$이므로, 모수 $\mu, \sigma$에 대한 기울기를 기댓값 안으로 넣어 $Z$의 표본으로 추정할 수 있다. 분포가 모수에 달려 있던 문제가, 함수가 모수에 달린 문제로 바뀐다.

같은 생각을 끝까지 밀면 **정규화 흐름**(normalizing flow)이 된다. 단순한 $\mathbf{Z}$에 학습 가능한 일대일 변환 $T_\theta$를 여러 번 걸어 $\mathbf{X} = T_\theta(\mathbf{Z})$를 만들면, 정리 5.1으로

$$\log f_{\mathbf{X}}(\mathbf{x}) = \log f_{\mathbf{Z}}\big(T_\theta^{-1}(\mathbf{x})\big) + \log \left| \det J_{T_\theta^{-1}}(\mathbf{x}) \right|.$$

변환을 야코비안 행렬식이 쉽게 계산되는 꼴(삼각행렬 등)로 설계하면, 데이터의 정확한 가능도를 계산하며 학습할 수 있다(9회).

## 8.3 앙상블 — 상관을 줄이는 것이 이득이다

$B$개의 모형의 예측을 평균한다. 각 예측의 분산이 $\sigma^2$이고 두 모형의 예측 사이의 상관이 $\rho$이면 정리 6.5로

$$\operatorname{Var}\Big(\frac{1}{B}\sum_{b=1}^{B}\hat{f}_b\Big) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2.$$

모형을 늘리면 둘째 항은 사라지지만 첫째 항 $\rho\sigma^2$는 남는다. 같은 자료로 학습한 모형들은 서로 닮아서 $\rho$가 크다. Breiman의 랜덤 포레스트(2001)는 나무마다 부트스트랩 표본(15회)을 쓰고 분기마다 변수의 일부만 보게 해서 **나무들 사이의 상관 $\rho$를 낮춘다**. 나무 하나는 조금 나빠지지만($\sigma^2$가 커진다), 평균의 분산은 줄어든다. Simpson이 관측을 평균하며 기대한 이득과, 유효 표본 크기가 말한 한계가 같은 식 안에 있다.

## 8.4 조건부기댓값으로서의 학습

정리 2.8는 제곱손실을 쓰는 모든 회귀 모형이 결국 $E(Y \mid X = x)$를 추정한다는 말이다. 분류에서 $Y$가 0과 1이면 $E(Y \mid X = x) = P(Y = 1 \mid X = x)$이므로, 제곱손실이든 교차 엔트로피(2회 8.2절)든 잘 학습된 분류기의 출력은 조건부확률이어야 한다. 2회 4.3절의 보정(calibration) 검사가 묻는 것이 바로 이것이다. 그리고 전분산 공식은 "모형이 아무리 좋아도 줄일 수 없는 오차" $E\,\operatorname{Var}(Y \mid X)$의 존재를 말한다. 기계학습에서 이것을 "줄일 수 없는 오차"(irreducible error)라 부른다.

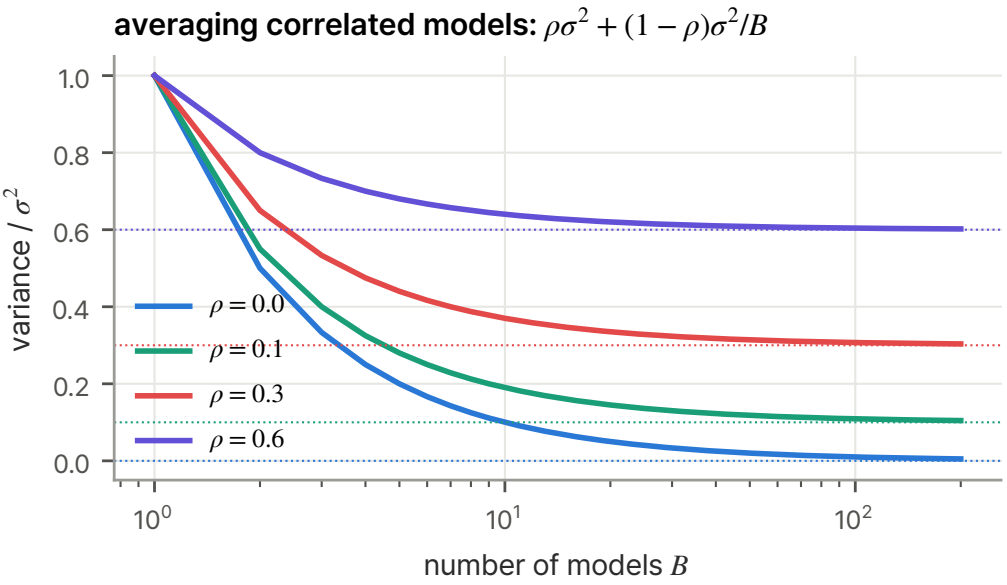



**그림 8.1** $B$개 예측의 평균의 분산(단위 $\sigma^2$). 상관 $\rho$가 바닥을 정한다. 앙상블의 기술은 $B$를 늘리는 것보다 $\rho$를 줄이는 것이다.

### 연습문제

**연습 8.1** (ReLU의 절반 ★) $s$의 분포가 0에 대해 대칭이고 $Es^2 < \infty$이면 $E\max(0,s)^2 = \frac{1}{2}Es^2$임을 보여라. He 초기화의 2는 여기서 온다. 풀이 56쪽

**연습 8.2** (재매개화의 기울기 ★★) $L(\mu) = E\big[(X-3)^2\big]$, $X \sim \mathcal{N}(\mu, 1)$이다. (a) $L(\mu) = (\mu-3)^2 + 1$임을 보이고 $L'(\mu)$를 구하라. (b) $X = \mu + Z$로 쓰면 $L'(\mu) = E[2(\mu + Z - 3)]$이다. 표본 $Z_1, \dots, Z_n$으로 이 기울기를 추정하는 식을 쓰고, 그 분산이 $\frac{4}{n}$임을 보여라. 풀이 56쪽

**연습 8.3** (앙상블의 한계 ★) 나무 하나의 예측 분산이 $\sigma^2 = 1$, 나무 사이의 상관이 $\rho = 0.3$이다. 나무 100개 평균의 분산은? 상관을 0.1로 낮추는 대신 나무 하나의 분산이 1.5로 커진다면 어느 쪽이 나은가? 풀이 56쪽

# 9

# 정리와 역사 노트

## 9.1 한 장 요약

| | 내용 | 증명의 핵심 |
|---|---|---|
| **주변분포** | $f_X(x) = \int f(x,y)\,dy$ | Tonelli |
| **선형성** | $E(aX+bY) = aEX + bEY$ (연속 포함) | 이중적분 |
| **조건부 밀도** | $f(y \mid x) = f(x,y)/f_X(x)$ | 근처의 사건의 극한 |
| **반복기댓값** | $E\,E(Y \mid X) = EY$ | 적분 순서 |
| **전분산** | $\operatorname{Var} Y = E \operatorname{Var}(Y \mid X) + \operatorname{Var} E(Y \mid X)$ | 교차항이 0 |
| **가장 좋은 예측** | $E(Y \mid X)$가 제곱오차 최소 | 피타고라스 |
| **독립** | $f = f_X f_Y \iff f = g(x)h(y)$ (곱의 영역) | 미분·적분 |
| **독립의 결과** | $E[gh] = Eg\,Eh$, 분산의 합, mgf의 곱 | 인수분해 |
| **공분산** | 쌍선형, $\operatorname{Var}\sum a_i X_i$ 전개 | 선형성 |
| **Cauchy–Schwarz** | $(EXY)^2 \le EX^2 EY^2$, $\lvert\rho\rvert \le 1$ | 판별식 |
| **가장 좋은 직선** | $b^* = \rho\sigma_Y/\sigma_X$, 잔차 분산 $\sigma_Y^2(1-\rho^2)$ | 이차식 최소 |
| **야코비안** | $f_{\mathbf{Y}} = f_{\mathbf{X}}(H(\mathbf{y}))\lvert J_H\rvert$ | 다중적분 치환 |
| **가우스 적분** | $\int e^{-x^2/2} = \sqrt{2\pi}$ | 극좌표 |
| **Box–Muller** | $\sqrt{-2\log U_1}(\cos,\sin)(2\pi U_2)$ | 인수분해 + 역변환 |
| **합성곱 · 비** | $f_{X+Y} = \int f_X f_Y(z-\cdot)$, 정규/정규 = Cauchy | 보조 변수 |
| **공분산행렬** | $\operatorname{Cov}(A\mathbf{X}+\mathbf{b}) = A\Sigma A' \succeq 0$ | 성분마다 선형성 |
| **$\sqrt{n}$ 법칙** | $\operatorname{Var}\bar{X} = \sigma^2/n$; 상관 $\rho$이면 $\to \rho\sigma^2$ | $\mathbf{1}'\Sigma\mathbf{1}$ |
| **부숴보기** | Borel 역설, 상관 0 ≠ 독립, 편향, Cauchy | 조건을 하나씩 뺀다 |

## 9.2 Stigler의 명명 법칙

> **Stigler's law of eponymy** "어떤 과학 법칙도 처음 발견한 사람의 이름을 달고 있지 않다."
>
> - **Cauchy–Schwarz 부등식**: 합의 형태는 Cauchy(1821), 적분의 형태는 Bunyakovsky(1859)가 먼저 증명했고, Schwarz(1888)는 그 뒤다. 러시아에서는 Cauchy–Bunyakovsky 부등식이라 부른다.
> - **Pearson의 상관계수**: 이변량 정규분포의 식은 Bravais(1846)가 먼저 썼고, 상관(co-relation)이라는 개념과 이름은 Galton(1888)의 것이다. Pearson(1896)은 표본 공식을 정리했다. Pearson 자신은 한때 이것을 Bravais의 공으로 돌렸다가 뒤에(1920) 철회했다.
> - **Simpson의 역설**(1회)과 이 회차의 **Thomas Simpson**은 다른 사람이다. 역설의 이름은 Edward Simpson(1951)에게서 왔고, 그 현상은 Yule(1903)이 먼저 기술했다.

## 9.3 7 Pillars 지도

Stigler(2016)의 둘째 기둥은 **Information**이다. "정보는 관측 수의 제곱근으로만 쌓인다"는 것, 곧 $\sqrt{n}$의 법칙이다. 첫째 기둥 Aggregation(2회)은 평균이 참값으로 간다고 말했고, 둘째 기둥은 그 속도를 말한다. Stigler는 이 기둥의 첫 증거로 바로 Pyx 재판을 든다. 정보가 $n$이 아니라 $\sqrt{n}$으로 쌓인다는 사실은 제도 설계, 실험 설계, 여론조사의 표본 크기를 정하는 출발점이다. 그리고 7절은 이 기둥의 그늘을 보였다. 관측이 서로 닮으면 정보는 $\sqrt{n}$보다 느리게 쌓이고, 편향은 전혀 줄지 않는다.

## 9.4 다음 회 — 셈하는 분포와 섞인 분포 (1711 → 1898)

이번 회차에서 지수분포의 합이 감마분포가 되는 것(예제 5.7)과, 동전을 던지는 횟수가 무작위일 때의 분산(예제 2.7)을 보았다. 4회는 이 둘을 분포의 가족으로 정리한다. 이항분포의 극한으로서의 Poisson 분포(de Moivre 1711, Poisson 1837), 그 대기시간인 감마분포, Bayes의 당구대에서 나온 베타분포, 그리고 여러 집단이 섞인 혼합분포다. 역사의 질문은 Bortkiewicz(1898)의 것이다. 프로이센 기병대에서 말에 차여 죽은 병사의 수는 어떤 분포를 따르는가?

## 참고문헌


- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard, 1986. 2장 (Simpson과 오차의 분포).
- Stigler, S. M. Eight centuries of sampling inspection: the trial of the Pyx. *JASA* 72 (1977), 493–500.
- Stigler, S. M. *Statistics on the Table*. Harvard, 1999. 21장 (Pyx 1848의 자료).
- Stigler, S. M. Do robust estimators deal with real data? *Annals of Statistics* 5 (1977), 1055–1098.

- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016. 2장 (Information).
- Simpson, T. A letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the advantage of taking the mean of a number of observations, in practical astronomy. *Philosophical Transactions* 49 (1755).
- Box, G. E. P., Muller, M. E. A note on the generation of random normal deviates. *Annals of Mathematical Statistics* 29 (1958).
- Borel, É. *Éléments de la théorie des probabilités*. Hermann, 1909.
- Michelson, A. A. Experimental determination of the velocity of light. *Astronomical Papers* 1 (1882). · Cavendish, H. Experiments to determine the density of the earth. *Philosophical Transactions* 88 (1798).
- He, K., Zhang, X., Ren, S., Sun, J. Delving deep into rectifiers. *ICCV* (2015). · Breiman, L. Random forests. *Machine Learning* 45 (2001).
- 자료: R 패키지 `HistData`의 `Pyx`, `Michelson`, `MichelsonSets`, `Cavendish`.

# 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 8쪽

$A=\{X\le b,\,Y\le d\}$에서 $\{X\le a\}$나 $\{Y\le c\}$인 부분을 뺀 것이 사각형 사건이다. $B=\{X\le a,\,Y\le d\}$, $C=\{X\le b,\,Y\le c\}$는 $A$의 부분집합이고 $B\cap C=\{X\le a,\,Y\le c\}$이므로 포함배제로 $P(B\cup C)=F(a,d)+F(b,c)-F(a,c)$. 따라서 $P(A\setminus(B\cup C))=F(b,d)-F(a,d)-F(b,c)+F(a,c)$.

### 연습 1.2 문제 8쪽

영역의 넓이가 $\frac{1}{2}$이므로 $c=2$. $f_X(x)=\int_x^1 2\,dy=2(1-x)$, $f_Y(y)=\int_0^y 2\,dx=2y$. $\{X+Y<1\}$과 영역의 교집합은 $0<x<\frac{1}{2}$, $x<y<1-x$이고 넓이가 $\int_0^{1/2}(1-2x)dx=\frac{1}{4}$이므로 확률은 $2\cdot\frac{1}{4}=\frac{1}{2}$.

### 연습 1.3 문제 8쪽

$L=l$인 결과는 "하나가 $l$, 다른 하나가 $l$보다 큼"(순서 2가지씩 $6-l$개)과 $(l,l)$이므로 $p_L(l)=\frac{2(6-l)+1}{36}=\frac{13-2l}{36}$. $EL=\frac{13\cdot21-2\cdot91}{36}=\frac{91}{36}$이고 $EM=\frac{161}{36}$이므로 합은 $\frac{252}{36}=7$. 선형성으로는 $M+L=D_1+D_2$이므로 바로 7이다.

### 연습 1.4 문제 8쪽

$P(R\le r)$는 반지름 $r$인 원의 넓이를 $\pi$로 나눈 $r^2$이다. 밀도 $2r$이므로 $ER=\int_0^1 2r^2dr=\frac{2}{3}$. $E|X|=\frac{4}{\pi}\int_0^1 x\sqrt{1-x^2}\,dx=\frac{4}{\pi}\cdot\frac{1}{3}=\frac{4}{3\pi}\approx0.424$.

### 연습 1.5 문제 8쪽

$X\sim\text{Bernoulli}(\frac{1}{2})$로 (i) $Y=X$, (ii) $Y=1-X$로 두면 두 경우 모두 $Y\sim\text{Bernoulli}(\frac{1}{2})$이지만 결합분포는 (i) $p(0,0)=p(1,1)=\frac{1}{2}$, (ii) $p(0,1)=p(1,0)=\frac{1}{2}$이다. 독립인 경우(네 칸 모두 $\frac{1}{4}$)

까지 세 가지가 같은 주변분포를 갖는다.

### 연습 1.6 문제 8쪽

(a) $\{X \le x\} = \{(X,Y) \in B\}$이고 $B = \{(s,t) : s \le x\}$는 보통의 집합이므로 정의 1.5의 마지막 문장으로 $F_X(x) = \iint_B f(s,t)\,ds\,dt$이다. (b) $f \ge 0$이므로 Tonelli 정리로 반복적분의 순서를 마음대로 바꿀 수 있다(값이 무한대일 수도 있지만 여기서는 1 이하다). $s$를 바깥에 두면 $F_X(x) = \int_{-\infty}^{x}\big(\int_{-\infty}^{\infty} f(s,t)\,dt\big)ds$이다. (c) $g(s) = \int f(s,t)\,dt$는 음이 아니고, (b)로 모든 $x$에서 $F_X(x) = \int_{-\infty}^{x} g$이다. 이것이 1회 정의 6.6의 조건이므로 $g$는 $X$의 밀도다. $t$를 바깥에 두면 같은 방법으로 $f_Y(y) = \int f(s,y)\,ds$이다. (d) $\{X \le x,\ Y \le n\}$은 $n$에 대해 증가하고 합집합이 $\{X \le x\}$이므로 1회 정리 4.7(a)로 $F_X(x) = \lim_n F_{X,Y}(x,n)$이다. $F_{X,Y}(x,n) = \int_{-\infty}^{x}\big(\int_{-\infty}^{n} f(s,t)\,dt\big)ds$ (Tonelli)이고, 안쪽 적분은 $n$에 대해 증가하며 $g(s)$로 가므로 단조수렴 정리 [T6]로 극한이 $\int_{-\infty}^{x} g$이다.

### 연습 1.7 문제 9쪽

(a) $|x|f(x,y) \ge 0$이므로 Tonelli로 $y$를 먼저 적분하면 $\iint |x|f = \int |x|\big(\int f(x,y)\,dy\big)dx = \int |x| f_X(x)\,dx = E|X|$이다(정리 1.6). 같은 방법으로 $\iint |y|f = E|Y|$이다. $|ax+by| \le |a||x|+|b||y|$이므로 $\iint |ax+by|f \le |a|E|X|+|b|E|Y| < \infty$이고, 정리 1.8(b)를 쓸 수 있다. (b) (a)로 $\iint |x|f < \infty$이므로 Fubini 정리로 순서를 바꿔 $y$를 먼저 적분할 수 있고, $\iint xf = \int x\big(\int f(x,y)\,dy\big)dx = \int x f_X(x)\,dx = EX$이다. (c) 적분의 선형성으로 $E(aX+bY) = a\iint xf + b\iint yf = aEX + bEY$이다. 이산이면 $\sum_{x,y}|x|p(x,y) = \sum_x |x| p_X(x) = E|X|$(음이 아닌 급수의 재배열, 1회 4.6절의 상자)로 절대수렴을 확인하고, 같은 재배열로 $\sum_{x,y} x\,p(x,y) = EX$를 얻는다.

## 2절

### 연습 2.1 문제 14쪽

$M=4$이면 $S$는 $5,6,7$(각각 $\frac{2}{36}$)과 $8$($\frac{1}{36}$)이고 $p_M(4) = \frac{7}{36}$이므로 조건부 확률은 $\frac{2}{7},\frac{2}{7},\frac{2}{7},\frac{1}{7}$, $E(S \mid M=4) = \frac{2(5+6+7)+8}{7} = \frac{44}{7}$. 일반적으로

$$E(S \mid M=m) = \frac{\sum_{k=1}^{m-1} 2(m+k)+2m}{2m-1} = \frac{m(3m-1)}{2m-1},$$

$ES = \sum_m \frac{2m-1}{36}\cdot\frac{m(3m-1)}{2m-1} = \frac{1}{36}\sum_m (3m^2-m) = \frac{273-21}{36} = 7.$

### 연습 2.2 문제 14쪽

$f_{X|Y}(x \mid y) = \frac{2}{2y} = \frac{1}{y}$ $(0<x<y)$, 곧 $\mathrm{U}(0,y)$이므로 $E(X \mid Y) = \frac{Y}{2}$. $EX = E\frac{Y}{2} = \frac{1}{2}\cdot\frac{2}{3} = \frac{1}{3}$이고, 직접 계산 $\int_0^1 x\cdot 2(1-x)dx = \frac{1}{3}$과 같다.

**연습 2.3** 문제 14쪽

$N=n$이면 $S$는 독립인 $n$개의 합이므로(그리고 $N$과 $X_i$가 독립이므로) $E(S\mid N)=N\mu$, $\operatorname{Var}(S\mid N)=N\sigma^2$. 반복기댓값으로 $ES=\mu EN$, 전분산 공식으로 $\operatorname{Var}S=E(N\sigma^2)+\operatorname{Var}(N\mu)=\sigma^2 EN+\mu^2\operatorname{Var}N$.

**연습 2.4** 문제 15쪽

$f(x,y)=1\cdot\frac{1}{x}$ $(0<y<x<1)$. $f_Y(y)=\int_y^1\frac{dx}{x}=-\log y$ $(0<y<1)$. $EY=E\frac{X}{2}=\frac{1}{4}$. $\operatorname{Var}Y=E\frac{X^2}{12}+\operatorname{Var}\frac{X}{2}=\frac{1}{36}+\frac{1}{48}=\frac{7}{144}$. 직접: $EY^2=\int_0^1 y^2(-\log y)dy=\frac{1}{9}$이므로 $\operatorname{Var}Y=\frac{1}{9}-\frac{1}{16}=\frac{7}{144}$.

**연습 2.5** 문제 15쪽

(a) 조건부분포에 대한 기댓값도 합·적분이므로 선형이다. (b) $X=x$로 고정하면 $g(X)=g(x)$는 상수이므로 $E(g(x)Y\mid X=x)=g(x)E(Y\mid X=x)$. 이것이 모든 $x$에서 성립하므로 확률변수로 쓰면 $g(X)E(Y\mid X)$.

**연습 2.6** 문제 15쪽

$\operatorname{Var}K=\frac{42}{48}+\frac{35}{48}=\frac{77}{48}$이고 $\operatorname{Var}E(K\mid N)=\frac{35}{48}$이므로 비는 $\frac{35}{77}=\frac{5}{11}\approx 0.45$. 평균제곱오차는 $\operatorname{Var}K=1.604$에서 $E\operatorname{Var}(K\mid N)=0.875$로 줄어든다.

## 3절

**연습 3.1** 문제 18쪽

정수값이면 $p(x,y)=F(x,y)-F(x-1,y)-F(x,y-1)+F(x-1,y-1)$이다(연습 1.1). $F=F_XF_Y$이면 이것이 $\big(F_X(x)-F_X(x-1)\big)\big(F_Y(y)-F_Y(y-1)\big)=p_X(x)p_Y(y)$로 인수분해된다. 거꾸로 $p=p_Xp_Y$이면 $F(x,y)=\sum_{s\le x}\sum_{t\le y}p_X(s)p_Y(t)=F_X(x)F_Y(y)$.

**연습 3.2** 문제 18쪽

(a) $4xy=2x\cdot 2y$이고 영역이 정사각형이므로 독립. (b) 영역 $0<x<y<1$이 곱의 꼴이 아니므로 독립이 아니다($y$가 작으면 $x$도 작아야 한다). (c) $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-y^2/2}$이고 영역이 평면 전체이므로 독립.

**연습 3.3** 문제 18쪽

$F_M(x)=x^3$, 밀도 $3x^2$, $EM=\frac{3}{4}$. 주사위 두 개: 꼬리 공식으로 $EM=\sum_{k=1}^{6}P(M\ge k)=$

$\sum_{k=1}^{6}\left(1-\left(\frac{k-1}{6}\right)^2\right) = 6 - \frac{55}{36} = \frac{161}{36}$.

### 연습 3.4 문제 19쪽

$P(X<Y) = \int_0^\infty \lambda e^{-\lambda x} P(Y>x)\,dx = \int_0^\infty \lambda e^{-(\lambda+\mu)x}dx = \frac{\lambda}{\lambda+\mu}$. 첫 등호는 결합밀도 $\lambda e^{-\lambda x}\mu e^{-\mu y}$를 $y>x$에서 먼저 적분한 것이다.

### 연습 3.5 문제 19쪽

$E(RGB) = P(\text{세 색 모두}) = \frac{1}{4}$이지만 $ER\,EG\,EB = \frac{1}{8}$이다. 정리 3.4(b)를 셋으로 넓히는 증명은 결합분포가 세 주변분포의 곱으로 인수분해된다는 것을 쓰는데, 쌍별 독립은 두 개씩의 결합분포만 인수분해한다.

### 연습 3.6 문제 19쪽

(a) 정리 3.2로 $f(x,y) = f_X(x)f_Y(y)$이다. $|g(x)h(y)|f_X(x)f_Y(y) \geq 0$이므로 Tonelli로

$$\iint |g(x)h(y)|\,f_X(x)f_Y(y)\,dx\,dy = \int |g|f_X \cdot \int |h|f_Y = E|g(X)|\,E|h(Y)| < \infty.$$

따라서 정리 1.8(b)를 쓸 수 있고, 절대적분이 유한하므로 Fubini 정리로 부호가 있는 함수에도 같은 순서 바꾸기가 허락된다. $E[g(X)h(Y)] = \int g\,f_X \cdot \int h\,f_Y = E\,g(X)\,E\,h(Y)$이다. (b) $X+Y-E(X+Y) = D_X + D_Y$를 제곱해 선형성을 쓰면 $\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2E(D_XD_Y)$이다. $D_X$는 $X$의 함수, $D_Y$는 $Y$의 함수이고 $E|D_X| \leq \sqrt{\mathrm{Var}\,X} < \infty$이므로 (a)로 $E(D_XD_Y) = ED_X\,ED_Y = 0$이다. (c) $g(x) = e^{tx}$, $h(y) = e^{ty}$이면 $e^{t(X+Y)} = g(X)h(Y)$이고 (a)로 $M_{X+Y}(t) = M_X(t)M_Y(t)$이다. (a)는 $E|g(X)| < \infty$를 요구하므로 $M_X(t)$와 $M_Y(t)$가 유한한 $t$에서만 성립한다. (d) $g$가 증가하면 $A_s = \{x : g(x) \leq s\}$는 $x \in A_s$, $x' < x$이면 $x' \in A_s$인 집합이다. 따라서 $c = \sup A_s$로 두면 $A_s$는 $\varnothing$, $\mathbb{R}$, $(-\infty, c]$, $(-\infty, c)$ 가운데 하나다. $\{g(X) \leq s\} = \{X \leq c\}$이면 정의 3.1로 $P(g(X)\leq s,\, Y\leq y) = F_X(c)F_Y(y)$이다. $\{X<c\}$이면 $\{X \leq c - \frac{1}{n},\, Y\leq y\} \uparrow \{X<c,\, Y\leq y\}$이므로 1회 정리 4.7로 $P(X<c,\, Y\leq y) = \lim_n F_X(c-\frac{1}{n})F_Y(y) = P(X<c)F_Y(y)$이다. 어느 경우든 $g(X)$와 $Y$의 결합분포함수는 주변분포함수의 곱이다.

## 4절

### 연습 4.1 문제 23쪽

$M = D_1$인 결과($D_1 \geq D_2$, 21가지)에서 $D_1M = D_1^2$, 나머지 15가지에서 $D_1M = D_1D_2$이다. 세면 $\sum D_1M = \sum_{d=1}^{6} d^2\cdot d + \sum_{a<b} ab = 441 + 175 = 616$이므로 $E(D_1M) = \frac{616}{36}$. $\mathrm{Cov}(D_1, M) = \frac{616}{36} - 3.5\cdot\frac{161}{36} = \frac{35}{24}$이고 $\mathrm{Cov}(S,M) = 2\cdot\frac{35}{24} = \frac{35}{12}$. ($\sum_{a<b} ab = \frac{1}{2}\left((\sum a)^2 - \sum a^2\right) = \frac{441-91}{2} = 175$.)

### 연습 4.2 문제 23쪽

$\operatorname{Cov}(X+Y, X-Y) = \operatorname{Var} X - \operatorname{Var} Y = 0$. 두 주사위의 합과 차는 상관이 없지만 독립이 아니다. $S = 2$이면 차는 반드시 0이다.

### 연습 4.3 문제 23쪽

$V(w) = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\rho\sigma_X\sigma_Y$를 미분하면

$$w^* = \frac{\sigma_Y^2 - \rho\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}.$$

$\rho = -1$: $w^* = 0.6$, $V = 0$. $\rho = 0$: $w^* \approx 0.692$, $V \approx 0.0277$ (표준편차 0.166). $\rho = 0.5$: $w^* \approx 0.857$, $V \approx 0.0386$ (표준편차 0.196). 상관이 낮을수록 섞어서 얻는 이득이 크다.

### 연습 4.4 문제 23쪽

$\operatorname{Cov}(aX+b, cY+d) = ac\operatorname{Cov}(X,Y)$, $\sigma_{aX+b} = |a|\sigma_X$이므로 상관계수는 $\frac{ac}{|a||c|}\rho$. $a, c > 0$이면 $\rho$, $a < 0 < c$이면 $-\rho$.

### 연습 4.5 문제 23쪽

(a) Cauchy–Schwarz를 $|X|$와 1에 쓰면 $(E|X|)^2 \le EX^2$. (b) $\operatorname{Var}(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\operatorname{Cov}(X,Y) \le \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y = (\sigma_X + \sigma_Y)^2$.

### 연습 4.6 문제 23쪽

$(x,y) \mapsto (x,-y)$로 밀도가 변하지 않으므로 $E(XY) = 0$, $EX = EY = 0$이고 상관이 없다. $0.8^2 + 0.8^2 > 1$이므로 $P(X > 0.8,\ Y > 0.8) = 0$이지만 $P(X > 0.8) = \frac{2}{\pi}\int_{0.8}^{1}\sqrt{1-x^2}\,dx \approx 0.052$이어서 곱은 약 $0.0027 > 0$이다. 독립이 아니다.

### 연습 4.7 문제 23쪽

(a) $2|XY| \le X^2 + Y^2$이므로 $E|XY| \le \frac{1}{2}(EX^2 + EY^2) < \infty$이고, $E|X| \le \frac{1}{2}(1 + EX^2) < \infty$도 같은 방법으로 나온다. 따라서 $(X-\mu_X)(Y-\mu_Y)$의 기댓값이 정의된다. (b) 선형성으로 $E(XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y) = E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$이다. $Y = X$이면 $\operatorname{Var} X$이고, 정의가 $X$와 $Y$에 대칭이므로 $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$이다. (c) $aX + b - E(aX+b) = a(X-\mu_X)$이고 $cY + d - E(cY+d) = c(Y-\mu_Y)$이므로 $\operatorname{Cov}(aX+b, cY+d) = E[ac(X-\mu_X)(Y-\mu_Y)] = ac\operatorname{Cov}(X,Y)$이다. 상수 $b, d$는 평균을 뺄 때 사라진다. 합에 대해서는 $X_1 + X_2 - E(X_1+X_2) = (X_1-\mu_1) + (X_2-\mu_2)$를 정의에 넣고 선형성을 쓴다. (d) (c)를 되풀이하면 $\operatorname{Cov}(\sum_i a_iX_i, \sum_j a_jX_j) = \sum_i\sum_j a_ia_j\operatorname{Cov}(X_i,X_j)$이다. $i = j$인 항은 $a_i^2\operatorname{Var}X_i$이고, $i \ne j$인 순서쌍은 $(i,j)$와 $(j,i)$가 같은 값 $a_ia_j\operatorname{Cov}(X_i,X_j)$를 주므로 $i < j$인 쌍마다 두 번 나온다. (e) 정리 3.4(b)를 $g(x) = x$, $h(y) = y$에 쓰면 $E(XY) = EX\,EY$

이고, (b)로 $\mathrm{Cov}(X,Y)=0$이다.

### 연습 4.8 문제 24쪽

(a) $W=Y-bX$로 두면 $E(W-a)^2$을 가장 작게 하는 상수는 $a=EW$이고 그 값은 $\mathrm{Var}\,W$이다(2회 정리 2.2(c)). (b) 정리 4.2(c)를 $a_1=1$, $a_2=-b$에 쓴다. (c) (b)를 $b$로 미분해 0으로 두면 $-2\,\mathrm{Cov}(X,Y)+2b\sigma_X^2=0$, 곧 $b^*=\mathrm{Cov}(X,Y)/\sigma_X^2=\rho\sigma_Y/\sigma_X$이다. 이차항의 계수 $\sigma_X^2>0$이므로 최소다. 넣으면 $\sigma_Y^2-2\,\mathrm{Cov}^2/\sigma_X^2+\mathrm{Cov}^2/\sigma_X^2=\sigma_Y^2-\mathrm{Cov}^2/\sigma_X^2=\sigma_Y^2(1-\rho^2)$이다. 이때 $a^*=E(Y-b^*X)=\mu_Y-b^*\mu_X$이다. (d) $\mathrm{Cov}(X,\,Y-a^*-b^*X)=\mathrm{Cov}(X,Y)-b^*\sigma_X^2=0$이다.

## 5절

### 연습 5.1 문제 31쪽

$(r,\theta)\mapsto(z_1,z_2)=(r\cos\theta,r\sin\theta)$의 역변환의 야코비안은 예제 5.2의 역수 $\frac{1}{r}$이다. 따라서 $f(z_1,z_2)=\frac{re^{-r^2/2}}{2\pi}\cdot\frac{1}{r}=\frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2}$, 독립인 표준정규의 곱이다.

### 연습 5.2 문제 31쪽

$V=XY$, $W=X$의 역변환은 $x=w$, $y=v/w$이고 야코비안은 $\det\left(\begin{smallmatrix}0 & 1\\ 1/w & -v/w^2\end{smallmatrix}\right)=-\frac{1}{w}$. 영역은 $0<v<w<1$이고 $f_{V,W}=\frac{1}{w}$이므로 $f_V(v)=\int_v^1\frac{dw}{w}=-\log v$.

### 연습 5.3 문제 31쪽

$n$개의 합의 밀도가 $g_n(z)=\frac{\lambda^n z^{n-1}}{(n-1)!}e^{-\lambda z}$라 하면, 합성곱으로

$$g_{n+1}(z)=\int_0^z g_n(x)\,\lambda e^{-\lambda(z-x)}dx=\frac{\lambda^{n+1}e^{-\lambda z}}{(n-1)!}\int_0^z x^{n-1}dx=\frac{\lambda^{n+1}z^n}{n!}e^{-\lambda z}.$$

### 연습 5.4 문제 31쪽

부분분수(또는 잔류 계산)로 $\int\frac{dx}{\pi^2(1+x^2)(1+(z-x)^2)}=\frac{2}{\pi(4+z^2)}$이다. 노트북에서 수치적분으로 확인한다. 그러면 $\frac{1}{2}(C_1+C_2)$의 밀도는 $2\cdot\frac{2}{\pi(4+4y^2)}=\frac{1}{\pi(1+y^2)}$, 표준 Cauchy다.

### 연습 5.5 문제 31쪽

삼각형 밀도 $f_2$와 $\mathrm{U}(0,1)$의 합성곱 $f_3(z)=\int_{z-1}^{z}f_2(x)\,dx$를 구간별로 계산한다. $0<z<1$: $\int_0^z x\,dx=\frac{z^2}{2}$. $1\le z<2$: $\int_{z-1}^1 x\,dx+\int_1^z(2-x)dx=\frac{-2z^2+6z-3}{2}$. $2\le z<3$: $\int_{z-1}^2(2-x)dx=\frac{(3-z)^2}{2}$.

### 연습 5.6 문제 31쪽

$P(R^2 > s) = P(R > \sqrt{s}) = e^{-s/2}$이므로 $R^2 \sim \mathrm{Exp}(\frac{1}{2})$, 평균 2.

### 연습 5.7 문제 31쪽

(a) $S$는 열린 집합이고 $P(Z_2 = 0) = 0$이므로 확률이 1이다. $T(z_1, z_2) = (c, w)$이면 $w = z_2$, $z_1 = cw$로 하나로 정해지므로 일대일이고 $T(S) = \{(c, w) : w \neq 0\}$이다. 역변환 $H(c, w) = (cw,\, w)$의 야코비안 행렬식은 $\det\left(\begin{smallmatrix} w & c \\ 0 & 1 \end{smallmatrix}\right) = w$이다. (b) 정리 5.1으로 $f_{C,W}(c, w) = f_{Z_1,Z_2}(cw, w)\,|w| = \frac{1}{2\pi} e^{-(c^2w^2+w^2)/2}|w|$이다. 밀도의 배율은 넓이의 배율이고, 넓이는 음수가 될 수 없으므로 행렬식의 절댓값이 붙는다. (c) 적분할 함수가 $w$에 대해 짝함수이므로

$$f_C(c) = 2\int_0^\infty \frac{1}{2\pi} e^{-(1+c^2)w^2/2}\, w\, dw = \frac{1}{\pi}\cdot\frac{1}{1+c^2}\int_0^\infty e^{-u}\, du = \frac{1}{\pi(1+c^2)}.$$

$u = \frac{1}{2}(1+c^2)w^2$이면 $du = (1+c^2)w\,dw$이다. (d) $\int \frac{dc}{1+c^2} = [\arctan c]_{-\infty}^{\infty} = \pi$이므로 적분은 1이다.

## 6절

### 연습 6.1 문제 36쪽

$A = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{smallmatrix}\right)$, $\Sigma = I$이므로 $AA' = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$. 상관계수는 $\frac{1}{2}$. 두 성분이 $X_2$를 공유하기 때문이다.

### 연습 6.2 문제 36쪽

$\mathbf{a} = (a_1, a_2)$에 대해 $\mathbf{a}'\Sigma\mathbf{a} = a_1^2 + 2\rho a_1 a_2 + a_2^2 \geq 0$이 모든 $\mathbf{a}$에서 성립할 필요충분조건은 $|\rho| \leq 1$이다. $3 \times 3$에서는 $\Sigma = (1-\rho)I + \rho\mathbf{1}\mathbf{1}'$로 쓴다. $\mathbf{a} = \mathbf{1}$이면 $\Sigma\mathbf{1} = (1+2\rho)\mathbf{1}$이고, $\mathbf{1}$에 직교하는 $\mathbf{a}(\sum a_i = 0)$이면 $\Sigma\mathbf{a} = (1-\rho)\mathbf{a}$이다. 고윳값이 $1+2\rho$와 $1-\rho$(두 번)이므로 양의 준정부호일 조건은 $-\frac{1}{2} \leq \rho \leq 1$이다. 음의 상관은 변수가 많을수록 강하게 걸 수 없다.

### 연습 6.3 문제 36쪽

한 번: $\frac{3}{11} \approx 0.273$. 여섯 번 평균(합이 $-6$과 6 사이): 합성곱으로 약 0.589. 삼각형 모형(예제 6.4)보다 둘 다 작지만 평균의 이득은 더 크다.

### 연습 6.4 문제 37쪽

통과 확률은 근사적으로 $\Phi\big(\sqrt{n}(R-\delta)/\sigma\big)$이다(아래쪽 경계는 무시할 만하다). 0.99 이상이려면 $\sqrt{n}(R-\delta)/\sigma \geq 2.326$, 곧 $\delta \leq R - 2.326\sigma/\sqrt{n} = 0.25625 - 0.0096 \approx 0.2467$그레인. 금화 100만 개면 약 $2.47 \times 10^5$그레인, 트로이 파운드(5,760그레인)로 약 43파운드의 금이다.

### 연습 6.5 문제 37쪽

$\operatorname{Var}\sum X_i = \sigma^2\sum_{i,j}\rho^{|i-j|} = \sigma^2\bigl(n + 2\sum_{k=1}^{n-1}(n-k)\rho^k\bigr)$ ($|i-j|=k$인 쌍은 $2(n-k)$개). $n$으로 나누면 $1 + 2\sum_{k<n}\rho^k - \frac{2}{n}\sum_{k<n}k\rho^k \to 1 + \frac{2\rho}{1-\rho} = \frac{1+\rho}{1-\rho}$.

### 연습 6.6 문제 37쪽

$\operatorname{Var}\frac{1}{2}(g(U)+g(1-U)) = \frac{1}{2}\operatorname{Var}g(U) + \frac{1}{2}\operatorname{Cov}(g(U),g(1-U)) \le \frac{1}{2}\operatorname{Var}g(U)$이고, 오른쪽이 독립인 두 난수의 평균의 분산이다. $g = e^x$: $\operatorname{Var}e^U = \frac{e^2-1}{2} - (e-1)^2 \approx 0.2420$, $\operatorname{Cov}(e^U, e^{1-U}) = e - (e-1)^2 \approx -0.2342$. 독립 둘의 평균은 0.1210, 대조변량은 $\frac{1}{2}(0.2420 - 0.2342) \approx 0.0039$로 약 31배 작다.

### 연습 6.7 문제 37쪽

(a) 선형성으로 $E\bigl[\sum_j A_{ij}X_j + b_i\bigr] = \sum_j A_{ij}\mu_j + b_i = (A\boldsymbol{\mu}+\mathbf{b})_i$이다. (b) (a)를 빼면 $A\mathbf{X}+\mathbf{b} - A\boldsymbol{\mu} - \mathbf{b} = A(\mathbf{X}-\boldsymbol{\mu})$이다. (c) $\mathbf{v} = \mathbf{X}-\boldsymbol{\mu}$로 두면 $(A\mathbf{v})_i = \sum_j A_{ij}v_j$이고 $(\mathbf{v}'A')_k = \sum_l v_l A_{kl}$이므로 $(A\mathbf{v}\mathbf{v}'A')_{ik} = \sum_j\sum_l A_{ij}v_j v_l A_{kl}$이다. 유한합이므로 기댓값이 안으로 들어가고 $E(v_j v_l) = \Sigma_{jl}$이므로 $\sum_j\sum_l A_{ij}\Sigma_{jl}A_{kl} = (A\Sigma A')_{ik}$이다. (d) $A = \mathbf{a}'$이면 $A\mathbf{X}$는 수 하나이고 그 분산이 $\mathbf{a}'\Sigma\mathbf{a}$이다. 분산은 음이 아니므로 모든 $\mathbf{a}$에서 $\mathbf{a}'\Sigma\mathbf{a} \ge 0$이다. $\Sigma_{ij} = \operatorname{Cov}(X_i,X_j) = \operatorname{Cov}(X_j,X_i) = \Sigma_{ji}$이므로 대칭이다.

## 7절

### 연습 7.1 문제 41쪽

$f_\Phi(\varphi) = \int_{-\pi}^{\pi}\frac{\cos\varphi}{4\pi}d\lambda = \frac{1}{2}\cos\varphi$. $P(\Phi \ge \frac{\pi}{3}) = \frac{1}{2}\bigl(1 - \sin\frac{\pi}{3}\bigr) = \frac{1-\sqrt{3}/2}{2} \approx 0.067$.

### 연습 7.2 문제 41쪽

(a) $(X,D)$, $D = X - Y$의 역변환 $y = x - d$의 야코비안은 1이므로 $f_{X,D}(x,0) = \phi(x)^2 \propto e^{-x^2}$, 곧 $\mathcal{N}(0,\frac{1}{2})$. (b) $(X,Q)$, $Q = Y/X$의 역변환 $y = qx$의 야코비안은 $|x|$이므로 $f_{X,Q}(x,1) = \phi(x)^2|x| \propto |x|e^{-x^2}$이고, $\int |x|e^{-x^2}dx = 1$이므로 밀도는 $|x|e^{-x^2}$. 같은 직선 $\{X=Y\}$인데 (b)는 0 근처에서 0이다. $Q$로 보면 원점 근처의 "쐐기"가 좁기 때문이다.

### 연습 7.3 문제 41쪽

$\bar{X} - \theta = b + \bar{\varepsilon}$이고 $E\bar{\varepsilon} = 0$, $\operatorname{Var}\bar{\varepsilon} = \sigma^2/n$이므로 $E(\bar{X}-\theta)^2 = b^2 + \sigma^2/n$(2회 정리 2.2(c)). Michelson: $b \approx 852.4 - 734.5 = 117.9$, $\sigma/\sqrt{n} \approx 7.9$. 평균제곱오차의 99% 이상이 편향이다.

### 연습 7.4 문제 41쪽

묶음 평균은 852.5, 854.5, 858.5, 857.0, 839.5이고 표준편차는 약 7.6으로, 독립이라면 기대되는 17.7보다 오히려 작다(다섯 개로 잰 표준편차는 그 자체로 많이 흔들린다는 점을 감안해야 한다). 묶음 사이의 흔들림이 크다는 증거는 없다. 15 표준오차의 어긋남은 날마다 달라지는 오차가 아니라 다섯 묶음 모두에 같이 들어 있는 상수, 곧 편향이었다.

## 8절

### 연습 8.1 문제 44쪽

$\max(0, s)^2 = s^2 \mathbf{1}\{s > 0\}$이고, 대칭이므로 $Es^2\mathbf{1}\{s > 0\} = Es^2\mathbf{1}\{s < 0\}$이며 $s = 0$에서는 $s^2 = 0$이다. 두 값의 합이 $Es^2$이므로 각각 $\frac{1}{2}Es^2$.

### 연습 8.2 문제 44쪽

(a) $X - 3 = (\mu - 3) + Z$이므로 $L = (\mu - 3)^2 + 1$, $L' = 2(\mu - 3)$. (b) $\hat{g} = \frac{1}{n}\sum_i 2(\mu + Z_i - 3)$이고 기댓값이 $L'(\mu)$, 분산은 $4 \operatorname{Var} \bar{Z} = \frac{4}{n}$.

### 연습 8.3 문제 44쪽

$0.3 + \frac{0.7}{100} = 0.307$. 다른 쪽은 $1.5(0.1 + \frac{0.9}{100}) = 0.1635$로 절반 가까이 작다. 상관을 낮추는 쪽이 낫다.